EVALUATION OF ESTIMATED HAMMERSTEIN MODELS VIA NORMALIZED PROJECTION MISALIGNMENT OF LINEAR AND NONLINEAR SUBSYSTEMS

Gerald Enzner, Tim C. Kranemann¹, and Philipp Thüne

Institute of Communication Acoustics, Ruhr-Universität Bochum, D-44780 Bochum, Germany Email: {gerald.enzner,tim.kranemann,philipp.thuene}@rub.de

ABSTRACT

In linear system identification, the coexistence of parametermisadjustment and output-error metrics has turned out very practical and their relation is well understood. In nonlinear system identification, however, such tools for performance evaluation are far less developed and each nonlinear type may need its own treatment. This paper focuses on the Hammerstein model as an instance of nonlinear systems. Irrespective of particular identification algorithms, we generalize the framework of parameter- and output-based performance metrics known from linear systems. An ambiguity in system parameters is resolved via the projection misalignment technique.

Index Terms— system modeling and identification, nonlinear systems, adaptive signal processing, system performance

1. INTRODUCTION & RELATION TO PRIOR WORK

Many system identification algorithms rely on a linear plant [1], but every physical system presumably shows a nonlinear behavior over a certain excitation [2]. Seminal work in [3,4] and extensions in [5] thus outlined the concept and the need for nonlinear adaptive filters in a range of applications. Volterra filter structures [6] have been frequently considered, based on the argument that Volterra models can serve as universal approximators to a large variety of nonlinear systems, including the Wiener, Hammerstein, Wiener-Hammerstein, or even more general model architectures [7,8].

A nice property of the Volterra model is its linearity in the parameters. Linear LMS- or RLS-type [1] adaptation algorithms can hence be applied [7]. The price for this "quasi-linearity" of the general Volterra model is, however, its vast computational demand related to the typically huge number of parameters (just reflecting the high dimensionality of the universal nonlinear space [7]). Due to this curse of complexity, many practical applications revert to more compact and specific nonlinear models, such as the Hammerstein system [3] in the center of this contribution.

The Hammerstein model is widespread [9] and recognized as one of the simplest extensions of linear filters into the nonlinear domain. It is created by arranging a memoryless nonlinear subsystem ahead of an ordinary linear, for instance, FIR system, cf. Fig. 1 in Sec. 2. The parametric Hammerstein model, as a special case of the Volterra series [7], exhibits linearity in all coefficients too. Applications are found in audio and acoustic signal processing [10, 11], sound and vibration [12], medical ultrasound [13], biological [14, 15] or chemical modeling [16], to name just a few.

More specifically, the Hammerstein model is frequently envisioned as a key component for realtime nonlinear acoustic echo control [17]. Adaptation algorithms with LMS- or RLS-type identification of its polynomial subsystem and time-domain LMS-type adap-

¹Now with the Chair of Medical Engineering, Ruhr-Universität Bochum.

tation of its FIR filter were presented early on in [18] and a recursive Bayesian algorithm with coupled estimation of both subsystems was recently proposed as an update [19]. Upon merging nonlinear and linear coefficients of the parametric Hammerstein model, an alternative multichannel-linear representation is obtained quickly and corresponding adaptation algorithms were again formulated with LMStype [20] and recursive Bayesian estimation [21].

The Hammerstein model is further recognized for excellent behavior in terms of its accuracy/complexity tradeoff with respect to nonlinear power amplifier modeling in communications [22]. The utility of the Hammerstein model has thus found a lot of appreciation very recently in signal processing for communications, when inband full-duplex wireless transmission is desired [23–26]. In this application, various performance measures for models and algorithms were explicitly considered in [27, 28].

Performance measures of nonlinear adaptive systems have unfortunately not received sufficient attention yet. In most of the cases, merely the output error between the actual system and its estimate is evaluated, since the identification algorithms already utilize this output error to adapt the model. In system identification, however, one is naturally interested in system parameters and one refers to perfect identification if the estimated parameters perfectly match the actual ones. Since this has not necessarily been achieved when the output error vanishes, "one has to distinguish between the methods used to obtain and to evaluate the estimate" [29]. This forms the basis of the coexistence of output-signal and system distances as known from linear systems [1] where both types of metrics are equal for broadband system input [30]. A particular Hammerstein system distance was proposed in [18], but it lacks a clear relationship with the output error. We therefore introduce Hammerstein distance metrics that can a) achieve dedicated inspection of both subsystems and b) deliver a prediction of the output error in case of sufficient excitation of the system. In this way, procedures that have been appreciated in linear adaptive systems will be generalized to nonlinear systems.

The remainder of the paper is organized as follows: Sec. 2 first recalls the parametric Hammerstein model and currently available performance measures in order to motivate the need for a Hammerstein specific treatment. Sec. 3 proposes the normalized projection misalignment (NPM) [29] for the evaluation of the linear subsystem and, with some generalization, of the nonlinear subsystem of the Hammerstein model. We then derive how those two NPMs jointly predict the output-error performance. Sec. 4 eventually demonstrates the utility of our evaluation framework on simulation data.

2. SIGNAL MODEL & PERFORMANCE MEASURES

2.1. Parametric Hammerstein Model

The Hammerstein model in a system identification setup is depicted in Fig. 1. Therein, the upper signal path represents an actual non-



Fig. 1: Hammerstein model in a system identification scenario.

linear function f(x) followed by a linear subsystem with impulse response w_k , while the lower signal path comprises estimates $\hat{f}(x)$ and \hat{w}_k of the actual subsystems. Actual and estimated Hammerstein system share the input signal x[k] at discrete time k, but their output can be different due to observation noise n[k] or imperfect adaptation of the subsystems. The output error $e[k] = d[k] - \hat{y}[k]$ is typically used to control the adaptation in each time-step.

A parametric representation of the nonlinear subsystem f(x)can be defined via basis functions $\Phi(x) = [\phi_1(x), \dots, \phi_P(x)]^T$ of order P, such that the output of the nonlinear subsystem reads

$$u[k] = f(x[k]) = \mathbf{\Phi}^{T}(x[k]) \mathbf{a}, \qquad (1)$$

with coefficient vector $\mathbf{a} = [a_1, \dots, a_P]^T$. The unobserved signal u[k] is then fed to the FIR filter to form the overall output signal

$$y[k] = w_{\kappa} * u[k] = \sum_{\kappa=0}^{N-1} w_{\kappa} u[k-\kappa] = \mathbf{u}^{T}[k] \mathbf{w}, \qquad (2)$$

via convolution. Here, $\mathbf{u}[k] = [u[k], \dots, u[k - N + 1]]^T$ and $\mathbf{w} = [w_0, \dots, w_{N-1}]^T$ are employed as short-hand notation.

In this paper, we do not focus on particular algorithms for the identification or adaptation of $\hat{\mathbf{w}}$ and $\hat{\mathbf{a}}$. We rather look into the methodology how to evaluate system parameters and output signals delivered by any algorithm. Generally, coefficient vectors \mathbf{w} and \mathbf{a} , although not explicitly denoted, could be time-varying. Estimated coefficients $\hat{\mathbf{w}}$ and $\hat{\mathbf{a}}$ will typically adapt with time, but the time index k is omitted for the sake of brevity, too. Actual and estimated subsystems are assumed to be of the same structure and model order, while model order mismatches could be resolved by zero padding.

2.2. Performance Measures for Hammerstein Identification

For the assessment of the quality of a particular identification process, especially with systems of relatively high model order, one is frequently interested in the evolution of single-number metrics related to either $\hat{\mathbf{w}}$, $\hat{\mathbf{a}}$, or in most of the cases both quantities.

On the one hand, output-signal distances can be formed on the basis of the output error e[k], typically square-error metrics, that are supposed to vanish in case of perfect system identification. Those measures are often normalized to a reference signal, see for instance the normalized mean-square error (NMSE) as shown by [1] or the echo return loss enhancement (ERLE) known in acoustic signal processing [31]. In what follows, those types of measures are together referred to as normalized output error (NOE) metrics denoted by

$$\eta = \frac{\mathrm{E}\{e^2[k]\}}{\mathrm{E}\{d^2[k]\}} , \qquad (3)$$

with $E\{\cdot\}$ being the statistical expectation operator.

On the other hand, in system identification one is naturally interested in the misalignment of system parameters, e.g., FIR filter coefficients. The filter misalignment is frequently expressed in linear adaptive systems in terms of the *normalized system distance* (NSD)

$$\zeta_w = \frac{\|\mathbf{w} - \widehat{\mathbf{w}}\|^2}{\|\mathbf{w}\|^2} , \qquad (4)$$

with $\|\cdot\|$ denoting the l^2 vector norm (Euclidean norm).

In linear systems, where $f(x) = \hat{f}(x) = x$, and when n[k] = 0, an equality $\eta = \zeta_w$ can be proven for broadband white-noise excitation x[k] via $E\{d^2[k]\} = \|\mathbf{w}\|^2 \sigma_x^2$ and $E\{e^2[k]\} = \|\mathbf{w} - \widehat{\mathbf{w}}\|^2 \sigma_x^2$, e.g., [30]. In case of non-white excitation both measures are different since NOE can only judge the system identification at frequencies that are actually excited (hence, NOE has to be used with care regarding the actual system identification regardless of the input signals. NSD thus is an independent and unfailing metric regarding system identification, while the clear link between NOE and NSD under the full-excitation condition, however, makes both of their definitions particularly reasonable.

Considering the Hammerstein model, where generally $f(x) \neq \hat{f}(x) \neq x$, the equality of NOE and NSD is lost. $\zeta_w = 0$, for instance, is never sufficient for $\eta = 0$. Thus, the meaning of the plain NSD applied to the nonlinear system is already very limited. Moreover, due to the cascade of linear and nonlinear subsystems, a Hammerstein inherent *gain ambiguity* exists, i.e., a linear factor can be interchanged and mutually compensated between both subsystems. This results in an infinite solution space regarding a quasi perfect system identification status and may even lead to numerical problems in practice [32]. Therefore both, the presence of the nonlinearity, and the gain ambiguity related to the cascaded model, have to be addressed in a suitable definition of system distances for the subsystems and for the overall Hammerstein cascade.

A previous instance of an overall Hammerstein system distance, which supposedly absorbs the issue of gain ambiguity, was already stated in [18],

$$\zeta_{\text{joint}} = \frac{\|\mathbf{w} \otimes \mathbf{a} - \widehat{\mathbf{w}} \otimes \widehat{\mathbf{a}}\|^2}{\|\mathbf{w} \otimes \mathbf{a}\|^2}, \qquad (5)$$

where \otimes is the Kronecker product. However, a desired relationship with the output error does not exist according to the authors of [18]. In other words, the measure does not support a clear prediction from the parameter estimation error to the output error and *vice versa*.

Our aspiration for the remainder of this paper, hence, is the development of Hammerstein evaluation tools

- to comprise output-error-based and parameter-based system distance measures known from linear systems,
- such that individual inspection of linear and nonlinear subsystems can be achieved,
- thereby absorbing the aforementioned gain ambiguity of the cascaded model arrangement,
- and supporting an output-error prediction from the parameter distances in case of sufficient excitation of the system.

3. PROPOSED HAMMERSTEIN SYSTEM DISTANCES

3.1. Application of NPM to the Linear Subsystem

The normalized projection misalignment (NPM),

$$\xi_w = \min_{\beta} \frac{\|\mathbf{w} - \beta \,\widehat{\mathbf{w}}\|^2}{\|\mathbf{w}\|^2} = \min_{\beta} \frac{\|\Delta \mathbf{w}\|^2}{\|\mathbf{w}\|^2} \,, \tag{6}$$

first introduced in [29], is a system distance measure which absorbs a gain ambiguity of linear subsystems. The optimal factor β that minimizes (6) can easily be computed as

$$\beta = \frac{\widehat{\mathbf{w}}^T \mathbf{w}}{\|\widehat{\mathbf{w}}\|^2} = \frac{\langle \widehat{\mathbf{w}}, \mathbf{w} \rangle}{\|\widehat{\mathbf{w}}\|^2}, \qquad (7)$$

where $\langle \cdot, \cdot \rangle$ denotes the inner vector product. In (6), a system model is implied in which the actual system is split into a scaled estimate and a residual, i.e., $\mathbf{w} = \beta \hat{\mathbf{w}} + \Delta \mathbf{w}$, and the components are orthogonalized through β , i.e., $\langle \Delta \mathbf{w}, \hat{\mathbf{w}} \rangle = 0$.

A straightforward alternative to the previous NPM definition is

$$\widetilde{\xi}_{w} = \min_{\widetilde{\beta}} \frac{\|\widehat{\mathbf{w}} - \widetilde{\beta} \,\mathbf{w}\|^{2}}{\|\widehat{\mathbf{w}}\|^{2}} = \min_{\widetilde{\beta}} \frac{\|\widetilde{\Delta} \mathbf{w}\|^{2}}{\|\widehat{\mathbf{w}}\|^{2}}, \quad (8)$$

where the estimated system is now split according to the alternative model $\widehat{\mathbf{w}} = \widetilde{\beta} \mathbf{w} + \widetilde{\Delta} \mathbf{w}$. We find a new minimization factor

$$\widetilde{\beta} = \frac{\mathbf{w}^T \widehat{\mathbf{w}}}{\|\mathbf{w}\|^2} = \frac{\langle \mathbf{w}, \widehat{\mathbf{w}} \rangle}{\|\mathbf{w}\|^2}$$
(9)

and by analogy with the previous NPM definition the different orthogonality $\langle \widetilde{\Delta} \mathbf{w}, \mathbf{w} \rangle = 0$. The main attention, however, should be devoted to the fact that (8) and (9) eventually yield the same NPM value $\widetilde{\xi}_w = \xi_w$ as the previous pair of equations (6) and (7). We find the alternative representation of the same NPM more intuitive regarding further utilization and generalization.

By rearranging (6) and (8), i.e., for either model, and by exploiting orthogonalities, we find expressions for the filter-norm ratio

$$\frac{\|\widehat{\mathbf{w}}\|^2}{\|\mathbf{w}\|^2} = \frac{(1-\xi_w)}{\beta^2} = \frac{\widetilde{\beta}^2}{(1-\xi_w)},$$
(10)

each of which confirms the effect of gain ambiguity even in a wellestimated system, i.e., when ξ_w small. Eq. (10) will turn out useful.

3.2. Application of NPM to the Nonlinear Subsystem

In order to account for arbitrary nonlinear basis functions, our definition of a performance measure for the nonlinear subsystem will be accomplished in the global f(x) domain, rather than evaluating performance on the expansion parameter level. This leads to an applicability of the proposed metric with any type of nonlinear expansion used in the adaptation algorithm, such as polynomial [5], Fourier [33,34], or spline [35]. Furthermore we can achieve an analogy in notation with equations (8) and (9) of the linear subsystem. In contrast to the l^2 vector-norm used before, we rely on the L^2 function-norm $\|\cdot\|$ in this context, as shown by the definition

$$\xi_f = \min_{\widetilde{\alpha}} \frac{\int_a^b \left(\widehat{f}(x) - \widetilde{\alpha}f(x)\right)^2 \mathrm{d}x}{\int_a^b \widehat{f}^2(x) \mathrm{d}x} = \min_{\widetilde{\alpha}} \frac{\|\widetilde{\Delta}f\|^2}{\|\widehat{f}\|^2} , \quad (11)$$

where a system model of the form $\hat{f}(x) = \tilde{\alpha}f(x) + \tilde{\Delta}f(x)$ is implied, and the integral is evaluated only over the support of the input signal $x[k] \in [a, b]$. The scale factor for minimization is then obtained as

$$\widetilde{\alpha} = \frac{\int_{a}^{b} f(x)\widehat{f}(x)dx}{\int_{a}^{b} f^{2}(x)dx} = \frac{\langle f, \widehat{f} \rangle}{\|f\|^{2}}$$
(12)

based on the inner product $\langle \cdot, \cdot \rangle$ in the L^2 -norm regime. The factor $\widetilde{\alpha}$ is the counterpart of the scale constant $\widetilde{\beta}$ for the linear subsystem. Hence, we expect $\widetilde{\alpha}\widetilde{\beta} \approx 1$ in case of successful adaptation of



Fig. 2: Decomposition of the Hammerstein model in an NPM-sense.

both subsystems. By analogy with the linear subsystem, we can now prove an orthogonality $\langle \widetilde{\Delta}f, f \rangle = 0$ and a function-norm ratio

$$\frac{|\hat{f}||^2}{|f||^2} = \frac{\tilde{\alpha}^2}{(1-\xi_f)} \,. \tag{13}$$

Fig. 2 revisits the estimated Hammerstein model, but substitutes \hat{f} and $\hat{\mathbf{w}}$ entities of the previous Fig. 1 with the misalignment models of linear and nonlinear subsystems introduced here. The upper signal path in Fig. 2 will then cancel with the upper signal path in Fig. 1, when $\tilde{\alpha}\tilde{\beta} = 1$. Fig. 2 further clarifies that generally four misalignment paths extend from the input x[k] to the output error e[k] between the actual and the estimated Hammerstein system. It can thus be expected that the individual NPMs of linear and nonlinear subsystems will contribute to the NOE of the Hammerstein system.

3.3. Expressing NOE with System Parameters and NPM

In order to establish the formal relationship of NOE and NPM, we rewrite both numerator and denominator of (3). When n[k] = 0,

$$E\{d^{2}[k]\} = E\{(w_{\kappa} * f(x[k]))^{2}\}$$

=
$$E\{\sum_{\kappa=0}^{N-1} \sum_{\nu=0}^{N-1} w_{\kappa} w_{\nu} f(x[k-\kappa]) f(x[k-\nu])\}.$$
 (14)

Then reusing the previously made assumption of a stationary whitenoise excitation x[k], i.e., $E\{x[k]x[k-\lambda]\}=0, \forall \lambda \neq 0$,

$$E\{d^{2}[k]\} = \sum_{\kappa=0}^{N-1} \sum_{\nu=0}^{N-1} w_{\kappa} w_{\nu} E\{f(x[k-\kappa])f(x[k-\nu])\}$$

$$= \sum_{\kappa=0}^{N-1} w_{\kappa}^{2} \int_{-\infty}^{-\infty} p(x)f^{2}(x[k])dx,$$
 (15)

where the second moment $E\{f^2(x)\} = \int_{-\infty}^{\infty} p(x)f^2(x)dx$ based on the probability density function (PDF) p(x) of x[k] has been used. When x[k] is uniformly distributed, the PDF is given as $p(x) = \frac{1}{b-a}$ for $x[k] \in [a, b]$ and zero otherwise, while the limits can still be adjusted. In practice, limiting $x[k] \in [-1, 1]$, without loss of generality, is reasonable for a polynomial expansion model since the powers of x[k] are getting large for inputs |x[k]| > 1. Note that uniform amplitude distribution is a straightforward counterpart and extension of the previous, say "uniform", white-noise excitation of all frequencies of a linear system. Then using l^2 -vector and L^2 -function norms,

$$\mathbf{E}\{d^{2}[k]\} = \frac{1}{2} \int_{-1}^{1} f^{2}(x) \mathrm{d}x \, \sum_{\kappa=0}^{N-1} w_{\kappa}^{2} = \frac{1}{2} \, \|f\|^{2} \, \|\mathbf{w}\|^{2} \, . \tag{16}$$

Since $e^2[k] = (d[k] - \hat{y}[k])^2 = d^2[k] + \hat{y}^2[k] - 2 d[k] \hat{y}[k]$, see Fig. 1, the NOE numerator is similarly found, when n[k] = 0, as

$$\mathbb{E}\{e^{2}[k]\} = \frac{1}{2} \|f\|^{2} \|\mathbf{w}\|^{2} + \frac{1}{2} \|\widehat{f}\|^{2} \|\widehat{\mathbf{w}}\|^{2} - \langle \mathbf{w}, \widehat{\mathbf{w}} \rangle \langle f, \widehat{f} \rangle .$$
(17)

Now forming the NOE quotient according to (3), and immediately substituting filter- and function-norm ratios, (10) and (13), we arrive at the desired relationship between the Hammerstein normalized output-error and our projection-misalignment quantities,

$$\widehat{\eta} = 1 + \frac{\widetilde{\alpha}^2}{(1 - \xi_f)} \frac{\beta^2}{(1 - \xi_w)} - 2\widetilde{\alpha}\widetilde{\beta} = \frac{(\widetilde{\alpha}\widetilde{\beta} - 1)^2 + (2\widetilde{\alpha}\widetilde{\beta} - 1)(\xi_w + \xi_f - \xi_w\xi_f)}{(1 - \xi_f)(1 - \xi_w)}, \quad (18)$$

where symbol $\hat{\eta}$ is used to depict a prediction of NOE from the involved NPMs and their associated scale factors, when looking at the reference case with *uniformly-distributed white-noise excitation*. Since $0 < \xi_{w/f} < 1$, we always have $\xi_w + \xi_f - \xi_w \xi_f > 0$. Thus, towards successful Hammerstein identification in the NPM-sense, i.e., with $\tilde{\alpha}\tilde{\beta} \rightarrow 1$ and both ξ_w and ξ_f small, we can further approximate

$$\widehat{\eta} \approx (\widetilde{\alpha}\beta - 1)^2 + \xi_w + \xi_f .$$
(19)

This expression cannot be simplified further, unless trivial, and signifies the effect of residual scale error $\widetilde{\alpha}\widetilde{\beta} \neq 1$ and individual projection misalignments ξ_w and ξ_f on the Hammerstein output error η .

4. EXPERIMENTAL RESULTS

We use uniformly distributed white excitation noise $x[k] \in [-1, 1]$, and, if not stated otherwise, the nonlinear function in the experiments is $f(x) = \operatorname{atan}(2x)/2$. The linear subsystem w_k is given by N = 256 normally distributed random coefficients. In order to depict the utility of the Hammerstein system distance measures, we rely on the NLMS/LMS configuration of the adaptation algorithm (without Gram-Schmidt orthogonalization) in [18]. Regarding the identification of the linear subsystem, the FIR model order is matched to the Hammerstein system, while the identification of the nonlinear subsystem utilizes polynomial basis functions $\phi_i(x) = x^{2i-1}$, i = 1, 2, ..., P with only odd powers and model order P = 5.

Fig. 3 shows the evolution of established and proposed distance metrics as a function of time, where a sampling frequency of 16 kHz was assumed to render the continuous time axis. From the top, the plain system distance ζ_w indicates moderate identification of the linear part of the Hammerstein system in the order of -10 dB, while the normalized output error η suggests much more accurate identification in the order of -20 dB. The new tools of this paper resolve the contradiction and reveal a contribution of different effects over time. In the beginning of the adaptation process, the initial status of the nonlinear subsystem is well accurate according to the projectionmisalignment ξ_f , while high projection-misalignment ξ_w of the initial linear subsystem explains the high Hammerstein output-error. As time goes, the algorithm adapts the linear subsystem very accurately, i.e., with ξ_w in the order of -30 dB, while ξ_f depicts the limiting factor for the output error η in the nonlinear subsystem. A snapshot of the estimated nonlinear function $\widehat{f}(x) = \mathbf{\Phi}^T(x) \,\widehat{\mathbf{a}}$ at the end of the adaptation is found for illustration in Fig. 4. The graph also signifies the expected alignment of f(x) and $\hat{f}(x)$ via the NPM-based scale factor $\tilde{\alpha}$. Our expectation of $\tilde{\alpha}\tilde{\beta} \approx 1$ is confirmed in Fig. 3.

Relationship (19) between subsystem misalignments and output error is eventually confirmed in Fig. 5. Here we present for two different nonlinear functions (model-match and model mismatch) that the actual output error η is well predicted by the $\hat{\eta}$ compound.

5. CONCLUSIONS

A set of advanced system distance measures was proposed for deep inspection of estimated Hammerstein systems. We found that



Fig. 3: Hammerstein normalized output-error and system distances.



Fig. 4: Hammerstein nonlinear subsystem and its identification.



Fig. 5: Hammerstein normalized output-error and its prediction.

both the established and our slight redefinition of the normalized projection-misalignment (the latter preferred) will provide accurate insight into the estimation quality of the subsystems, which has not been achieved with conventional mean-square error distances before. The utilization of the proposed output-error and systemmisalignment metrics for nonlinear Hammerstein systems is eventually simple, intuitive, and nicely compatible with the procedures known from the evaluation of estimated linear systems.

6. REFERENCES

- S. Haykin, Adaptive Filter Theory, Pearson, Essex, England, 5th edition, 2014.
- [2] S.A. Billings, "Identification of nonlinear systems a survey," *IEE Proc. D, Control Theory and Applications*, vol. 127, pp. 272–285, Nov. 1980.
- [3] K.S. Narendra and P.G. Gallmann, "An iterative method for identification of nonlinear systems using a Hammerstein model," *IEEE Trans. Automatic Control*, vol. 11, pp. 546–550, July 1966.
- [4] V.J. Mathews, "Adaptive polynomial filters," *IEEE Signal Process. Mag.*, vol. 8, pp. 10–26, Nov. 1991.
- [5] V. J. Mathews and G.L. Sicuranza, *Polynomial Signal Process*ing, Wiley, New York, 2000.
- [6] M. Schetzen, *The Volterra & Wiener Theories of Nonlinear Systems*, Krieger, Malabar, 2006.
- [7] O. Nelles, *Nonlinear System Identification*, Springer, Berlin, Germany, 2001.
- [8] D.R. Morgan, Z. Ma, and J. Kim *et al.*, "A generalized memory polynomial model for digital predistortion of RF power amplifiers," *IEEE Trans. Signal Proc.*, vol. 54, pp. 3852–3860, Oct. 2006.
- [9] A.E. Nordsjo and L.H. Zetterberg, "Identification of certain time-varying nonlinear Wiener and Hammerstein systems," *IEEE Trans. Signal Proc.*, vol. 49, pp. 577–592, Mar. 2001.
- [10] A. Novak, L. Simon, F. Kadlec, and P. Lotton, "Nonlinear system identification using exponential swept-sine signal," *IEEE Trans. Instrum. Meas.*, vol. 59, pp. 2220–2229, Aug. 2010.
- [11] L. Tronchin, "The emulation of nonlinear time-invariant audio systems with memory by means of Volterra series," *Jrnl. Audio Eng. Soc.*, vol. 60, pp. 984–996, Dec. 2012.
- [12] M. Rébillat, R. Hennequin, É. Corteel, and B.F.G. Katz, "Identification of cascade of Hammerstein models for the description of nonlinearities in vibrating devices," *Jrnl. of Sound and Vibration*, vol. 330, no. 5, pp. 1018–1038, Feb. 2011.
- [13] F. Sbeity, S. Ménigot, J. Charara, and J. Girault, "Contrast improvement in sub- and ultraharmonic ultrasound contrast imaging by combining several Hammerstein models," *Intl. Jrnl. Biomedical Imaging*, 2013, Article ID 270523.
- [14] I.W. Hunter and M.J. Korenberg, "The identification of nonlinear biological systems: Wiener and Hammerstein cascade models.," *Biological Cybernetics*, vol. 55, no. 2-3, pp. 135– 144, 1986.
- [15] K. Jalaleddini and R.E. Kearney, "Subspace identification of SISO Hammerstein systems: Application to stretch reflex identification," *IEEE Trans. Bio-Med. Eng.*, vol. 60, pp. 2725– 2734, May 2013.
- [16] E. Eskinat, S.H. Johnson, and W.L. Luyben, "Use of Hammerstein models in identification of nonlinear systems," *AIChE Journal*, vol. 37, pp. 255–268, Feb. 1991.
- [17] G. Enzner, H. Buchner, A. Favrot, and F. Kuech, "Acoustic Echo Control," in *Academic Press Library*, R. Chellappa and S. Theodoridis, Eds., vol. 4, pp. 807–878. Elsevier, 2014.
- [18] A. Stenger and W. Kellermann, "Adaptation of a memoryless preprocessor for nonlinear acoustic echo cancelling," *Signal Processing, Elsevier*, vol. 80, pp. 1747–1760, Sept. 2000.

- [19] S. Malik and G. Enzner, "A variational Bayesian learning approach for nonlinear acoustic echo control," *IEEE Trans. Signal Proc.*, vol. 61, pp. 5853–5867, Dec. 2013.
- [20] F. Kuech and W. Kellermann, "Orthogonalized power filters for nonlinear acoustic echo cancellation," *Signal Processing*, *Elsevier*, vol. 86, pp. 1168–1181, June 2006.
- [21] S. Malik and G. Enzner, "State-space frequency-domain adaptive filtering for nonlinear acoustic echo cancellation," *IEEE Trans. Audio, Speech and Language Process.*, vol. 20, pp. 2065–2079, Sept. 2012.
- [22] A. Tehrani, C. Haiying, and S. Afsardoost *et al.*, "A comparative analysis of the complexity/accuracy tradeoff in power amplifier behavioral models," *IEEE Trans. Microw. Theory Tech.*, vol. 58, pp. 1510–1520, June 2010.
- [23] L. Anttila, D. Korpi, V. Syrälä, and M. Valkama, "Cancellation of power amplifier induced nonlinear self-interference in fullduplex tranceivers," in *Proc. of Asilomar Conf. on Signals, Systems & and Computers*, Nov. 2013, pp. 1–6.
- [24] D. Bharadia, E. McMilin, and S. Katti, "Full-duplex radios," in Proc. of IEEE Int. Symp. Image and Signal Processing and Analysis (ISPA), Aug. 2013, pp. 375–386.
- [25] A. Sabharwal, P. Schniter, and D. Guo *et al.*, "In-band fullduplex wireless: challanges and opportunities," *IEEE J. Sel. Areas Commun.*, vol. 32, pp. 1637–1652, Oct. 2014.
- [26] M. Heino, D. Korpi, and T. Huusari *et al.*, "Recent advances in antenna design and interference cancellation algorithms for in-band full duplex relays," *IEEE Commun. Mag.*, vol. 53, pp. 91–101, May 2015.
- [27] D. Wisell, M. Isaksson, and N. Keskitalo, "A general evaluation criteria for behavioral power amplifier modeling," in *ARFTG Microw. Meas. Conf. (ARFTG)*, June 2007, pp. 1–5.
- [28] P. Landin, M. Isaksson, and P. Handel, "Comparison of evaluation criteria for power amplifier behavioral modeling," in *Proc.* of *IEEE MTT-S Int. Microw. Symp. Dig. (IMS)*, June 2008, pp. 1441–1444.
- [29] D.R. Morgan, J. Benesty, and M.M. Sondhi, "On the evaluation of estimated impulse responses," *IEEE Signal Process. Lett.*, vol. 5, pp. 174–176, July 1998.
- [30] P. Vary and R. Martin, Digital Speech Transmission Enhancement, Coding and Error Concealment, Wiley, Chichester, England, 2006.
- [31] E. Hänsler and G. Schmidt, *Acoustic Echo and Noise Control: A Practical Approach*, Wiley, Hoboken, 2004.
- [32] E.L.O. Batista and R. Seara, "A new perspective on the convergence and stability of NLMS Hammerstein filters.," in *Proc. IEEE Intl. Conf. Acoust., Speech, Signal Proc. (ICASSP)*, Sept. 2013, pp. 343–348.
- [33] S. Malik and G. Enzner, "Fourier expansion of Hammerstein models for nonlinear acoustic system identification," in *IEEE Intl. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 85–88.
- [34] A. Carini and G.L. Sicuranza, "Fourier nonlinear filters," Signal Processing, Elsevier, vol. 94, pp. 183–194, Jan. 2014.
- [35] M. Scarpiniti, D. Comminiello, R. Parisi, and A. Uncini, "Hammerstein uniform cubic spline adaptive filters: Learning and convergence properties," *Signal Processing, Elsevier*, vol. 100, pp. 112–123, July 2014.