CYRO SCARANO HEMSI

# Digital Signal Processing Techniques for the Compensation of Analog Circuit Impairments in Impedance Mismatched Radio-Frequency Transmitters

São Paulo
2021

CYRO SCARANO HEMSI

# Digital Signal Processing Techniques for the Compensation of Analog Circuit Impairments in Impedance Mismatched Radio-Frequency Transmitters

Tese apresentada à Escola Politécnica da Universidade de São Paulo para obtenção do título de Doutor em Ciências.

São Paulo
2021

CYRO SCARANO HEMSI

# Digital Signal Processing Techniques for the Compensation of Analog Circuit Impairments in Impedance Mismatched Radio-Frequency Transmitters

Versão Corrigida

Tese apresentada à Escola Politécnica da Universidade de São Paulo para obtenção do título de Doutor em Ciências.

Área de Concentração: Sistemas Eletrônicos

Orientador: Prof. Dr. Cristiano M. Panazio

São Paulo

2021

Este exemplar foi revisado e corrigido em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, _____ de _____ de _____

Assinatura do autor: _____

Assinatura do orientador: _____

Catalogação-na-publicação

Hemsi, Cyro Scarano

Técnicas de processamento digital de sinais para a compensação de imperfeições em circuitos analógicos de transmissores de radiofrequência com descasamento de impedância / C. S. Hemsi -- versão corr. -- São Paulo, 2021.

171 p.

Tese (Doutorado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Sistemas Eletrônicos.

1.Amplificador de potência 2.Modelo de Volterra 3.Pré-distorção digital 4.LASSO 5.Modelo comportamental I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Sistemas Eletrônicos II.t.

Este trabalho é dedicado a Bianca e Manuela.

# ACKNOWLEDGEMENTS

After working for 20 years as a systems engineer, product manager and applications engineer at Promon, Ericsson, ZTE and Keysight, I decided to dedicate myself to an academic career.

Today, as I complete my doctorate, I think back to both the enjoyment and the harsh times that I had, the lessons of persistence, discipline and self-knowledge I learnt and would like to express my gratitude to the following people:

first of all, my advisor Prof. Cristiano Magalhães Panazio, for his ongoing support, patience, friendship and valuable advice

my thesis committee members, Profa. Cynthia C. M. Junqueira, Prof. Ivan R. S. Casella, Prof. Eduardo L. Ortiz Batista and Prof. Phillip M. S. Burt

my wife, Bianca Sabbag Hemsi, and my daughter, Manuela Sabbag Hemsi

my parents, Lilia A. M. Scarano Hemsi and Ibrahim Hemsi

my brother, Paulo S. Hemsi

Prof. Paul Jean E. Jenszensky and Prof. José Sidnei Colombo Martini

Profa. Fatima S. Correra and Antonio Sandro Verri from the Laboratório de Microondas (LME) for supporting the experiments.

I am also grateful to the Laboratório de Comunicações e Sinais (LCS) from the Departamento de Engenharia de Telecomunicações e Controle (PTC) - POLI - USP and to the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for granting me a doctorate scholarship during the development of this work.

*"Work is love made visible".*
*(The Prophet, Khalil Gibran)*

# RESUMO

Esta tese investiga e propõe novos modelos para a compensação digital de imperfeições de rádio frequência (RF - *radio frequency*) em transmissores sem fio de banda larga, mais especificamente para amplificadores de potência (PA - *power amplifiers*) não-lineares e sujeitos a descasamento de impedância de carga (LMM - *load mismatch*). Tais imperfeições em transmissores de RF, juntamente com o desbalanceamento entre as componentes em fase e quadratura (IQ - *in-phase and quadrature*) em moduladores digitais, são responsáveis por degradar o desempenho do transmissor, tanto em termos de pureza espectral, quanto em qualidade da modulação e taxa de erro de bit (BER - *bit error rate*).

Diversos cenários práticos em que PAs estão sujeitos a LMM motivam a pesquisa por modelos comportamentais não-lineares e com memória voltados a predistorção digital (DPD - *digital pre-distortion*) mais avançados, capazes de superar as limitações relatadas na literatura de modelos tradicionais e que sejam menos complexos do que as abordagens existentes para PA LMM.

Nesta tese é proposta a aplicação do modelo polinomial Wiener-Hammerstein com realimentação (WHFB - *Wiener-Hammerstein with feedback*) como modelo comportamental simplificado para DPD no contexto de PAs sujeitos a LMM.

A elevada dimensionalidade da estrutura WHFB proposta pode ser reduzida através de técnicas esparsas de estimação, tais como o operador de encolhimento e seleção absoluto mínimo (LASSO - *least absolute shrinkage and selection operator*) e extensões de LASSO em blocos, que são capazes de diminuir significativamente o número de coeficientes necessário, reduzindo, assim, o comprimento do filtro de DPD e, proporcionalmente, o custo da filtragem.

Além disso, extensões de LASSO voltadas a blocos, como *Group-LASSO* e *Sparse-group LASSO*, são aplicadas no contexto de dimensionamento do modelo, ou seja, na tarefa de determinar valores adequados para seus parâmetros, o que, tradicionalmente, requer uma busca exaustiva.

Os modelos WHFB denso e esparso são validados experimentalmente através de medições em um *set-up* de testes experimental e também comparados a outros, incluindo modelos baseados na decomposição em fatores paralelos (PARAFAC - *parallel factors*) e na expansão de Laguerre, demonstrando, assim, sua capacidade de compensar adequadamente PAs sujeitos a LMM.

Finalmente, é proposta uma nova estratégia de redução do modelo de Volterra, que resulta em um polinômio de memória mais flexível e modular, em que os parâmetros são escolhidos independentemente para cada ordem de não linearidade. Esta abordagem flexível é capaz de descrever com precisão uma gama ampla de condições operacionais/ ambientais de PAs.

**Palavras-chave**: Amplificador de potência. Descasamento de impedância de carga. Modelo comportamental. Modelo de Volterra. Pré-distorção digital. LASSO.

# ABSTRACT

This thesis investigates and proposes new models for the digital compensation of radio frequency (RF) imperfections in broadband wireless transmitters, more specifically for non-linear power amplifiers (PA) subject to load impedance mismatch (LMM). Such imperfections in RF transmitters, together with the in-phase and quadrature (IQ) imbalance in digital modulators, are responsible for degrading the transmitter's performance, in terms of spectral purity, modulation quality and bit error rate (BER).

Several practical scenarios in which PAs are subject to LMM motivate the research for more advanced non-linear behavioral models with memory for digital predistortion (DPD), capable of overcoming the limitations reported in the literature of traditional polynomial models, while being less complex than existing approaches to PA LMM.

This thesis proposes the application of the Wiener-Hammerstein with feedback (WHFB) polynomial model as a simplified behavioral model for DPD in the context of PAs subject to LMM.

The high dimensionality of the proposed WHFB structure can be reduced through sparse estimation techniques, such as the least absolute shrinkage and selection operator (LASSO) and group LASSO extensions, which are able to significantly decrease the number of coefficients needed, thus reducing the length of the DPD filter and, proportionally, the cost of filtering.

In addition, block-oriented LASSO extensions, such as Group-LASSO and Sparse-group LASSO, are applied in the context of model sizing, that is, in the task of determining appropriate values for the model parameters, which traditionally requires an exhaustive search.

Dense and sparse WHFB models are experimentally validated through measurements from an experimental test set-up and also compared to others, including models based on parallel factors (PARAFAC) decomposition and on Laguerre expansion, thus demonstrating their ability to adequately compensate for subject PAs the LMM.

Finally, a new strategy for reducing the Volterra model is proposed, which results in a more flexible and modular memory polynomial, in which the parameters are chosen independently for each order of nonlinearity. This flexible approach is able to accurately describe a wide range of operating/environmental conditions of PAs.

**Keywords**: Power amplifier. Load impedance mismatch. Behavioural model. Volterra model. Digital pre-distortion. LASSO.

# LIST OF FIGURES

# LIST OF TABLES

# ACRONYMS

| | |
|---|---|
| 3GPP | $3^{rd}$ generation partnership project |
| 4G | $4^{th}$ generation |
| 5G | $5^{th}$ generation |
| ACPR | adjacent channel power ratio |
| ADC | analog-to-digital converter |
| AIC | Akaike's information criterion |
| ALS | alternating least squares |
| AM/AM | amplitude-to-amplitude |
| AM/PM | amplitude-to-phase |
| ANN | artificial neural network |
| AWG | additive white Gaussian |
| BER | bit error rate |
| BN | bias network |
| BPF | bandpass filter |
| CD | coordinate descent |
| CLMS | complex least mean squares |
| CR | cognitive radio |
| CS | compressed sensing |
| DAC | digital-to-analog converter |
| DC | direct-current |
| DDR | dynamic deviation reduction |
| DIDO | double input double output |
| DLA | direct learning architecture |
| DPD | digital pre-distortion |
| DSP | digital signal processing |
| EVM | error vector magnitude |
| FET | field-effect transistor |
| FIR | finite impulse response |
| FPET | fixed pole expansion technique |
| FSK | frequency-shift keying |
| FV | full-Volterra |
| GaAs | Gallium Arsenide |
| GMP | generalized memory polynomial |
| I | in-phase |

| | |
|---|---|
| I/O | input-output |
| IBO | input power back-off |
| IF | intermediate frequency |
| IIR | infinite impulse response |
| ILA | indirect learning architecture |
| IM3 | $3^{rd}$-order intermodulation |
| IMD | intermodulation distortion |
| IQ | in-phase and quadrature |
| IQM | IQ modulator |
| ISI | inter-symbol interference |
| KKT | Karush-Kuhn-Tucker |
| LASSO | least absolute shrinkage and selection operator |
| LM | Levenberg-Marquardt |
| LMM | load mismatch |
| LO | local oscillator |
| LPE | low-pass equivalent |
| LPF | lowpass filter |
| LTI | linear time-invariant |
| LUT | look-up table |
| LV | Laguerre-Volterra |
| MEMS | micro-electromechanical system |
| MESFET | metal–semiconductor field-effect transistor |
| MIMO | multi-input multi-output |
| MN | matching network |
| MP | memory polynomial |
| mset | multiset |
| NL | non-linear |
| NLS | non-linear least squares |
| NMSE | normalized mean square error |
| NP-hard | non-deterministic polynomial time |
| NVNA | non-linear vector network analyzer |
| OBF | orthonormal basis function |
| OFDM | orthogonal frequency-division multiplexing |
| OLS | ordinary least squares |
| OMP | orthogonal matching pursuit |
| PA | power amplifier |
| PAPR | peak to average power ratio |

| | |
|---|---|
| PARAFAC | parallel factors |
| PCA | principal component analysis |
| PH | parallel Hammerstein |
| PHD | poly-harmonic distortion |
| PRSS | penalized residual sum of squares |
| PSD | power spectral density |
| PSK | phase-shift keying |
| PSO | particle swarm optimization |
| Q | quadrature-phase |
| QAM | quadrature amplitude modulation |
| RF | radio frequency |
| RSS | residual sum of squares |
| SDR | software defined radio |
| SGL | sparse-group LASSO |
| SNL | static non-linear |
| SoC | system-on-a-chip |
| SSPA | solid state power amplifier |
| SVD | singular value decomposition |
| TF | transfer function |
| TMN | tunable matching network |
| TX | transmitter |
| VP | Volterra-PARAFAC |
| VSWR | voltage standing wave ratio |
| WCDMA | wideband code division multiple access |
| WHFB | Wiener-Hammerstein with feedback |

# LIST OF SYMBOLS

| | |
|---|---|
| $\tilde{x}(n)$ | PA input signal in baseband notation |
| $\tilde{x}_I(n)$ | in-phase component of $\tilde{x}(n)$ |
| $\tilde{x}_Q(n)$ | quadrature component of $\tilde{x}(n)$ |
| $\hat{x}(n)$ | DPD training modelled output signal using ILA |
| $x_{RF}(t)$ | PA RF input signal |
| $\hat{y}(n)$ | PA baseband modelled output signal |
| $y_{RF}(t)$ | PA RF output signal |
| $\tilde{u}(n)$ | normalized PA output or ILA DPD training input signal |
| $\hat{u}(n)$ | DPD training modelled output signal using DLA |
| $\hat{e}(n)$ | DPD training error signal |
| $\tilde{s}(n)$ | IQM baseband input signal before DAC |
| $\tilde{s}(t)$ | IQM baseband input signal after DAC |
| $\tilde{w}(n)$ | block model intermediate signal |
| $\tilde{g}(m)$ | block model LTI filter impulse response |
| $\tilde{\gamma}(l)$ | feedback LTI filter impulse response |
| $\tilde{h}_k(m_{1:k})$ | $k^{th}$-order baseband Volterra kernel |
| $\boldsymbol{\alpha}$ | PA model coefficients vector |
| $\alpha$ | SGL group-sparsity parameter |
| $\boldsymbol{\beta}$ | DPD model coefficients vector |
| $\epsilon$ | model residual or AWGN noise |
| $\lambda$ | LASSO sparsity parameter |
| $\boldsymbol{\phi}$ | DPD model regressor |
| $\boldsymbol{\Phi}$ | DPD model regression matrix |
| $\Gamma$ | load reflection coefficient |
| $\mu$ | amplitude deviation |
| $\rho$ | phase deviation |
| $\omega_c$ | RF carrier frequency |
| $G$ | PA intended linear power gain |
| $K$ | model non-linear order |
| $M$ | model memory depth |
| $Q$ | model cross-term non-linear order |
| $L$ | model cross-term memory span |
| $N$ | number of samples |

# NOTATIONS

| | |
|---|---|
| $a$ | scalar $a$ |
| $\boldsymbol{x}$ | vector $\boldsymbol{x}$ |
| $\boldsymbol{X}$ | matrix $\boldsymbol{X}$ |
| $\mathcal{X}$ | tensor $\mathcal{X}$ |
| $\mathsf{X}$ | set or multiset $\mathsf{X}$ |
| $\mathbb{R}$ | set of real numbers |
| $\mathbb{C}$ | set of complex numbers |
| $\mathbb{C}^{n_1 \times n_2 \times \ldots \times n_m}$ | set of complex-valued $n_1 \times n_2 \times \ldots \times n_m$ dimensional tensors |
| $\Re\{\cdot\}$ | real part |
| $\Im\{\cdot\}$ | imaginary part |
| $(\cdot)^*$ | conjugate operator |
| $(\cdot)^T$ | transpose operator |
| $(\cdot)^H$ | conjugate or Hermitian transpose operator |
| $(\cdot)^{-1}$ | matrix inverse operator |
| $(\cdot)^\dagger$ | matrix Moore-Penrose pseudo-inverse operator |
| $sgn(\cdot)$ | signum function |
| $\|\cdot\|_0$ | $\ell_0$ pseudo-norm of vector |
| $\|\cdot\|_1$ | $\ell_1$ norm of vector |
| $\|\cdot\|_2$ | $\ell_2$ or Euclidean norm of vector |
| $\|\cdot\|_F$ | Frobenius norm of matrix |
| $\langle\ \rangle$ | inner (dot) product of vectors |
| $\circ$ | outer product of vectors |
| $(*)$ | Hadamard (element-wise) product of matrix |
| $*$ | convolution of vectors |
| $\otimes$ | Kronecker product |
| $\oslash$ | truncated Kronecker product |
| $\oslash_{\_}$ | modified truncated Kronecker product |
| $\odot$ | Khatri-Rao (column-wise Kronecker) product |
| $\binom{n}{k}$ | binomial coefficient $n$ choose $k$ |
| $\left(\!\binom{n}{k}\!\right)$ | $n$ multichoose $k$ |
| $supp(\cdot)$ | support of a vector or matrix |
| $card(\cdot)$ | cardinality or the number of elements |
| $vec(\cdot)$ | vectorization operator that stacks the columns of a matrix |
| $diag(\cdot)$ | diagonalization operator of a vector into a diagonal matrix |

| | |
|---|---|
| $rank(\cdot)$ | rank of matrix or tensor |
| $colspace(\cdot)$ | column space of matrix |
| $nullspace(\cdot)$ | null space of matrix |
| $cond(\cdot)$ | condition number of matrix |
| $det(\cdot)$ | determinant of matrix |
| $\frac{\partial f}{\partial \theta}$ | partial derivative of $f$ w.r.t. variable $\theta$ |
| $\mathcal{O}(\cdot)$ | complexity in Ordo-notation |

# CONTENTS

# 1 INTRODUCTION

This chapter introduces the context of modern wireless communications transceivers and the importance of mitigating radio frequency (RF) analog impairments that compromise the system performance. A summary of state-of-the-art research on major sources of impairments, as well as on digital compensation techniques, is briefly presented. This chapter also motivates for the use of digital signal processing (DSP) techniques for combating the major sources of RF impairments, specially the effective compensation for power amplifiers under load impedance mismatch.

## 1.1 CONTEXT AND MOTIVATION

Modern wireless communications systems need to handle wideband signals carrying high data-rates with elevated spectral efficiency. Therefore, the majority of them employs digital modulation formats with non-constant envelope and high peak to average power ratio (PAPR), such as high-order quadrature amplitude modulation (QAM), wideband code division multiple access (WCDMA) and orthogonal frequency-division multiplexing (OFDM). These systems also need to guarantee high linearity and power efficiency, limiting in-band and out-of-band distortions and, at the same time, avoiding the waste of supply power and environmental resources (LEUNG; ZHU, 2014).

The radio transceivers for such systems need to be low cost and support multi-band and multi-standard operation, according to the software defined radio (SDR) architecture, in which the RF front-end is reconfigured through software (ZHU; LEUNG; HUANG, 2013). In addition, their designs should also foresee the evolution to cognitive radio (CR) (MITOLA; MAGUIRE, 1999) architecture and monolithic integration, in a system-on-a-chip (SoC). In this challenging context, the direct conversion transceiver topology, also known as zero-intermediate frequency (IF) or homodyne (MAK; U; MARTINS, 2007), shown in Figure 1.1, is considered the most promising architecture, due to its single stage of frequency conversion from baseband to RF. Figure 1.1 shows the direct conversion transmitter (TX), including the in-phase (I) and quadrature-phase (Q) signal paths with the digital-to-analog converter (DAC), lowpass filter (LPF) and the IQ modulator (IQM) mixer pair, fed by the high-frequency local oscillator (LO). The signals at the output of the IQM mixers are combined and sent to the RF power amplifier (PA), bandpass filter (BPF) and load antenna. Note that using this topology, fewer analog components are needed, such as mixing and filtering stages, as compared to the conventional dual- or multi-stage heterodyne transceivers (ABIDI, 1995), (MAK; U; MARTINS, 2007).

Analog components in the RF front-end always introduce circuit imperfections or

non-idealities, limiting the system performance (FETTWEIS et al., 2007), (VALKAMA; RENFORS; KOIVUNEN, 2001a). The major sources of RF impairments in the transmitter side are the non-linear (NL) distortions caused by the PA (PEDRO; MAAS, 2005), (GHANNOUCHI; HAMMI, 2009), gain and phase imbalance caused by the IQM (VALKAMA; RENFORS; KOIVUNEN, 2001a), (HUANG; CARON, 2002), PA load impedance mismatch (BEZOOIJEN; MAHMOUDI; ROERMUND, 2005), frequency-dependent gain and group delay of analog BPFs (MYOUNG; KIM; YOOK, 2005) and carrier frequency offset and phase noise of the LOs (ARMADA, 2001). These impairments inevitably introduce distortions to the envelope signal to be transmitted and, as the operating frequency and data bandwidth increase – for example, in $4^{th}$ generation (4G) or $5^{th}$ generation (5G) wireless communications standards from the $3^{rd}$ generation Partnership Project (3GPP) – more stringent constraints should be met on the imperfections of transceiver circuits.

This thesis focuses on DSP techniques for combating the following sources of impairments on wireless transmitters (SMAINI, 2012):

- The PA non-linearity that introduces NL distortions to the output signal, degrading both the in-band performance, measured by the error vector magnitude (EVM) and bit error rate (BER), as well as the out-of-band performance, measured by the adjacent channel power ratio (ACPR). Moreover, NL distortions may also cause adjacent channel interference, also known as spectrum regrowth, so ACPR levels are strictly limited by spectrum emission masks imposed by the regulatory bodies (HORLIN; BOURDOUX, 2008), (CRIPPS, 2006).

- The PA load impedance mismatch, or load mismatch (LMM), in which the PA is not perfectly terminated, so only part of the outgoing power is delivered to the load and signal reflections return back to the PA. This RF impairment reduces the maximum achievable output power, causes voltage clipping and additional signal distortions, worsening the performance metrics mentioned above, as well as reinforcing PA memory effects (ZENTENO; ISAKSSON; HANDEL, 2015).

- The gain and phase imbalance between the I and Q paths of the IQM, which breaks the orthogonality between the two components and reduces the attenuation of the image signals at mirror frequencies, resulting in signal constellation errors that degrade the EVM and BER performance (HORLIN; BOURDOUX, 2008), (VALKAMA; RENFORS; KOIVUNEN, 2001b), (ANTTILA; VALKAMA; RENFORS, 2008b).

In the next sections, we discuss in further details these non-idealities and their compensation techniques.

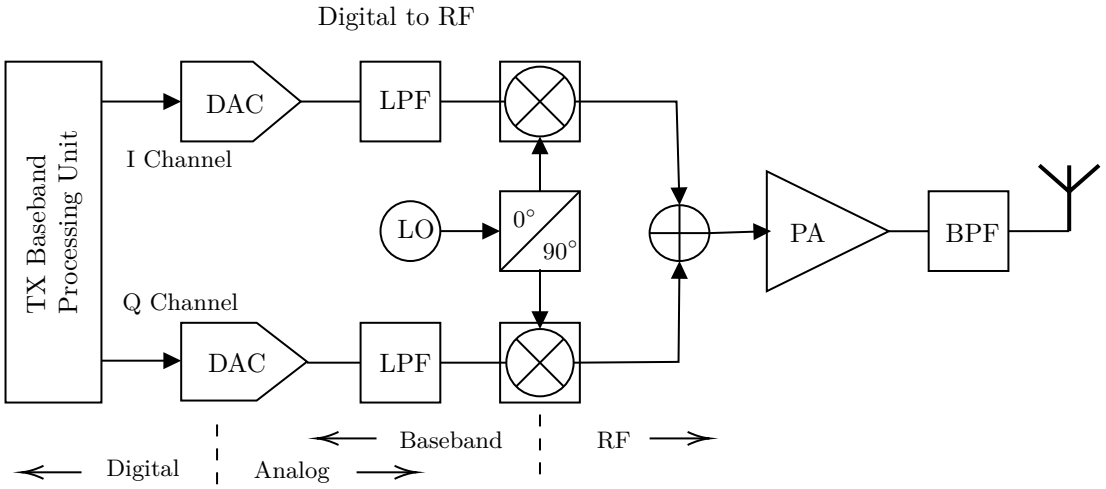



Figure 1.1 – Block diagram of the direct conversion TX.

## 1.2 TRANSCEIVER RF IMPAIRMENTS

This section presents a summary of state-of-the-art approaches for modelling major sources of RF impairments in wireless transmitters, as well as motivates for their compensation using digital pre-distortion (DPD) techniques.

### 1.2.1 Power Amplifier Non-linearity

Ideally, amplification is a purely linear operation, described by a linear power gain. However, in practice, typical PAs achieve their highest power efficiency when operating well beyond the maximum linear output power, thus exhibiting a trade-off between efficiency and linearity. In this analysis, the PA power efficiency is defined by $\eta = \frac{P_{RF}}{P_{DC}}$, in which $P_{RF}$ is the RF output power at the fundamental band and $P_{DC}$ is the power drawn from the direct-current (DC) source. PA non-linearity distorts the complex envelope of the transmitted signal, creating constellation distortion and inter-symbol interference (ISI), whereas low PA efficiency leads to short battery life in mobile devices and high operating costs in base stations. PAs in wireless communications are classified by their quiescent (or DC biasing) point of operation (CRIPPS, 2006) into the classic Class A, AB, B, C, etc. (CRIPPS, 2006), ranging from highly linear, but inefficient, towards non-linear, highly efficient PA operation. Indeed, the Class A quiescent point is high enough so that the PA conducts current during the entire excursion of the input signal, whereas $P_{DC}$ is reduced in the classes AB, B and C and the PA only conducts during a portion of the input signal. Traditionally, constant envelope modulation schemes, such as frequency-shift keying (FSK), allow the linear PA to be biased more efficiently. However, wideband signals require bandwidth efficient amplitude and phase modulation schemes, such as QAM, WCDMA and OFDM, which produce non-constant envelopes and high PAPR that are specially vulnerable to PA non-linearities. Back-off strategies could be used to reduce the dynamic

range of the input signals and avoid severe output distortions, but at the expense of reducing the power efficiency. In fact, there is a considerable effort in the literature to achieve linear transmitters with high efficiency by compensating for PA non-linearities in wideband wireless communication systems.

In order to better understand and mitigate the PA NL distortions discussed so far, models have been developed following two major approaches: physical-based (also known as circuit-based) models and behavioural (also known as empirical or black-box) models. In the physical-based models, analytical expressions are derived from the so-called PA equivalent circuit, based on circuit theory and physical knowledge of the transistor (active device), biasing and matching conditions (COLANTONIO; GIANNINI; LIMITI, 2009), (PEDRO; PEREZ, 1994). Behavioural models, on their turn, are black-box descriptions of the PA's NL input to output relation, in which no physical knowledge is required, being entirely based on fitting mathematical functions to sampled input/output data (PEDRO; MAAS, 2005) and (ISAKSSON; WISELL; RONNOW, 2006). The ever increasing bandwidth of wireless communications systems makes the modelling and compensation of non-linear PAs with memory effects a relevant topic and leads to sophisticated PA and DPD modelling frameworks, such as the Volterra series model (SCHETZEN, 2006) and artificial neural networks (ANN) (NASKAS; PAPANANOS, 2004), (ISAKSSON; WISELL; RONNOW, 2005), (LIU; BOUMAIZA; GHANNOUCHI, 2004), (XU et al., 2002). Comparing these two approaches, one important aspect is that Volterra-based models are often linear in the coefficients, whereas ANNs are not (PEDRO; MAAS, 2005), (ISAKSSON; WISELL; RONNOW, 2006). Because Volterra-based models are linear in the coefficients, linear least squares estimation methods are applicable. These methods are very desirable for DPD; for this reason, this thesis focuses on Volterra-based behavioural models (DING et al., 2004).

### 1.2.2 PA Load Impedance Mismatch (LMM)

The PA load comprises the external impedances associated to the transistor's output impedance through the output matching network (MN) and DC biasing network (BN). Thus, in this work, we consider LMM as the combined effect of impedance variations that occur at baseband (modulation band), fundamental and harmonic frequencies and appear at the interface between the transistor output port and the external networks above (VUOLEVI; RAHKONEN, 2003), (CARVALHO; PEDRO, 2002), (GHANNOUCHI; HAMMI, 2009). These variations may occur due to several reasons, such as when the PA is subject to load antenna impedance deviations, antenna crosstalk, DC power supply and temperature fluctuations, MN and BN transfer function that depend on the PA operating conditions, for example, power and frequency (BEZOOIJEN; MAHMOUDI; ROERMUND, 2011). The PA LMM creates multiple signal reflections at the PA output port, generating

frequency-dependent interactions between the forward and reflected RF waves, responsible for worsening the PA's linearity and memory effects (VUOLEVI; RAHKONEN, 2003). As described in (CRIPPS, 2006), LMM reduces the maximum PA output power, as well as the PA power efficiency (BEZOOIJEN; MAHMOUDI; ROERMUND, 2005), (SANCHEZ-PEREZ et al., 2011). Moreover, LMM at the fundamental band accentuates the PA output signal distortions (YE et al., 2010), whereas load impedance fluctuations at baseband and harmonics are the main responsible for the so-called NL memory effects (MOON; KIM, 2011), i.e., PA memory effects that only manifest under the NL regime. Finally, the impedance mismatch of the BN accounts for most of the NL memory effects, due to operational conditions, such as battery de-charging, temperature, power and frequency variations (BEZOOIJEN; MAHMOUDI; ROERMUND, 2011).

The operation of radiating devices, i.e., antennas, may be affected by nearby objects, such as the user body and metallic structures (BOYLE; YUAN; LIGTHART, 2007), (BERG; SONKKI; SALONEN, 2009). These proximity (or body) effects are responsible for degrading the performance of some types of antennas, as discussed in the literature (KESHMIRI; CRAEYE, 2009), (MORISHITA; KIM; FUJIMOTO, 2001), (ILVONEN et al., 2011) which includes antenna gain reduction, center frequency detuning and input impedance fluctuations. Depending on the magnitude and phase of the antenna input impedance deviations, the PA load-line and biasing point are displaced, so the output signal becomes subject to voltage clipping (CRIPPS, 2006), an effect that is known to generate odd-order NL distortion terms that degrade the linearity and the performance of the transmitter.

In several practical cases, the PAs are expected to suffer from LMM, thus justifying the research for more robust DPD solutions, such as in:

- Portable/ mobile devices in the presence of obstacles and proximity-effects that induce time-varying antenna input impedance fluctuations (BEZOOIJEN; MAHMOUDI; ROERMUND, 2011), (MESSAOUDI et al., 2007),

- Wideband RF transmitters, in which the design of the MN and BN are subject to technical limitations that prevent a perfect impedance matching both at RF and baseband (AKMAL et al., 2010), (BRINKHOFF; PARKER, 2003),

- Multi-band, multi-mode transmitters, such as SDRs, where the MN are expected to operate in multiple frequency bands, signal bandwidths and power levels (RAWAT; HASHMI; GHANNOUCHI, 2012),

- Multi-antenna transceivers, such as multi-input multi-output (MIMO) in 4G and 5G wireless standards and phased-arrays in radar applications, due to mutual coupling

and crosstalk between the antenna elements (HAUSMAIR et al., 2017), (KORPI; ANTTILA; VALKAMA, 2017),

- Millimeter-wave compact transceivers in 5G wireless standards, where frequency-dependent transistor parasitics become extremely difficult to eliminate (ALSHARIF; NORDIN, 2017), as well as the electromagnetic coupling inside the chip (BARRADAS et al., 2017).

This thesis addresses the digital compensation of NL PAs subject to LMM by proposing the application of a Volterra-based behavioural model, namely, the Wiener-Hammerstein with feedback (WHFB) model (MKADEM et al., 2014), that offers a good trade-off between model performance and complexity.

### 1.2.3 IQ Modulator Imbalance

In the ideal IQM case, the local oscillators feeding the I and Q signal components have identical amplitude and exactly 90° of phase shift, as well as the frequency responses of the two independent I and Q branches are identical. The IQM imbalance occurs when at least one of these conditions is not met. In general, IQM are subject to both the frequency-independent amplitude and phase imbalance and the frequency-dependent imbalance. This last impairment may arise due to the frequency responses of LPFs, DACs and mixing stages at the I and Q branches (RAZAVI, 1997). Any IQM amplitude or phase imbalance degrades the transmitter mirror-image suppression capability and, consequently, the system performance (ANTTILA; VALKAMA; RENFORS, 2008b). Depending on the transmitter architecture, the residual mirror-image produces either self-interference (zero-IF TX) or adjacent channel interference (low-IF TX) (MAK; U; MARTINS, 2007). The major part of the models in the literature address the narrowband case, in which only the frequency-independent IQM imbalance is considered. Wideband IQM imbalance models, also accounting for the frequency-dependent effects, can be found in some papers (VALKAMA; RENFORS; KOIVUNEN, 2001b), (TUTHILL; CANTONI, 2005).

The DPD model proposed in this thesis for NL PAs subject to LMM is extended to the joint compensation of IQM imbalance, including frequency-dependent effects.

In the next section, we describe some important compensation techniques targeted to the RF impairments discussed so far.

## 1.3 DIGITAL COMPENSATION OF RF IMPAIRMENTS

Modern wireless communication systems face the challenge of implementing low cost analog RF front-ends that are capable of operating with high performance in compliance with strict spectral regulations. Conceptually, the most straightforward approach would

be to optimize the quality of the analog components in the front-end, such that the RF impairments are made by design at an acceptable level. However, it is very expensive to improve the processes of design, fabrication and calibration of RF analog components (NAGATA, 1989). Besides, this can limit the transceiver flexibility, since optimal designs are often frequency specific and cannot track changes over time, such as temperature variations and component ageing. An alternative approach is to model the RF analog front-end hardware limitations and apply DSP techniques in the digital baseband domain to improve the key performance metrics, such as the in-band and out-of-band spectral quality and the system BER (FETTWEIS et al., 2007). Therefore, DSP compensation techniques are a consolidated approach for mitigating RF analog front-end impairments, being one of the most cost-effective solutions and providing significant accuracy and flexibility (HORLIN; BOURDOUX, 2008).

The next subsections briefly discusses existing DSP techniques to compensate for the impairments described in the previous section.

### 1.3.1 PA Digital Pre-distortion

DPD is a well-known technique for compensating for NL PAs. This technique allows the linearized PA to be utilized up to its saturation point, while still providing linear amplification, thus allowing the PA power efficiency to increase. As shown in Figure 1.2, DPD is applied in the digital baseband domain, since the DPD predistorter precedes the "Digital to RF" conversion block. The predistorter is responsible for applying to the incoming baseband signal a NL function with memory that is approximately the inverse of the normalized NL characteristic of the PA, assuming that a DPD NL model is properly chosen and that its coefficients are effectively estimated. The process of identifying or estimating the NL model coefficients from measured input/output data is performed by the so-called DPD training function and is presented in more details in section 2.5.

The first DPD technique was proposed by (NAGATA, 1989), (CAVERS, 1990) for memoryless PAs, using a two-dimensional look-up table (LUT) (MUHONEN; KAVEHRAD; KRISHNAMOORTHY, 2000), (BARRADAS et al., 2014) that provided a complex gain correction based on the real and imaginary parts of the input signal. When considering wideband signals, memory effects cannot be neglected, so the complex gain corrections should also vary with frequency for an effective compensation. Among the NL modelling approaches proposed in the literature for wideband PA and DPD modelling, we focus on behavioural models derived from the Volterra series (SCHETZEN, 2006), such as memory polynomials and block-oriented models (PEDRO; MAAS, 2005), (ISAKSSON; WISELL; RONNOW, 2006), because most of them are linear in the coefficients, allowing linear least squares estimation (DING et al., 2004), very advantageous for DPD.

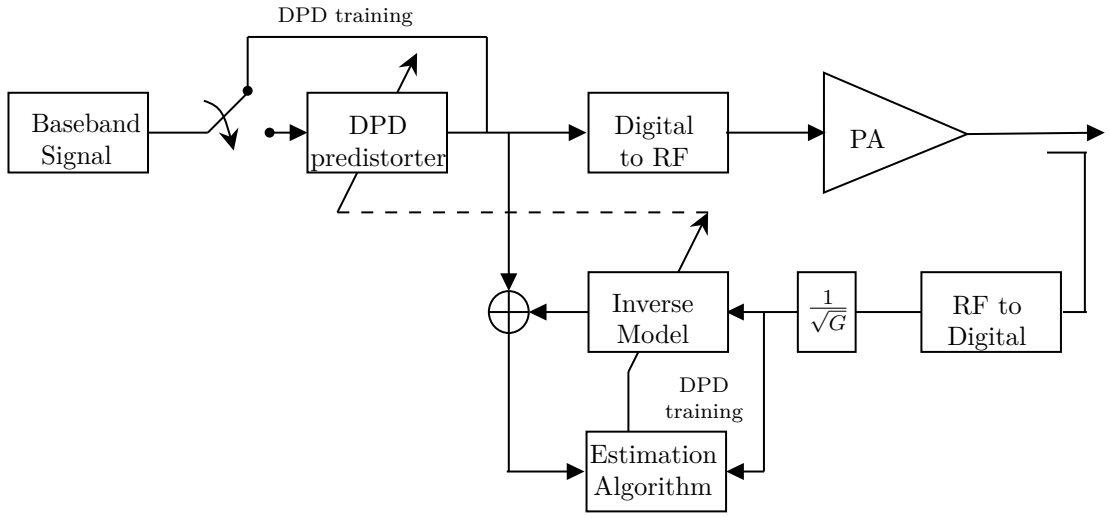



Figure 1.2 – Conceptual DPD block-diagram.

The high dimensionality of the Volterra series model has motivated sparse, parsimonious approximate solutions (STANKOVIC et al., 2018), (BRAITHWAITE, 2017), also used in the context of compressed sensing (CS), to the original model. The more prominent ones have been applied to PA and DPD, such as the least absolute shrinkage and selection operator (LASSO) (ZENTENO et al., 2015), (KEKATOS; GIANNAKIS, 2011), (WISELL; JALDEN; HANDEL, 2008), ridge regression (GUAN; ZHU, 2012) and orthogonal matching pursuit (OMP) (ABDELHAFIZ et al., 2014), (TOSINA et al., 2015), (YAO et al., 2014). In this thesis, firstly, the LASSO technique is applied to the WHFB DPD model in chapter 6 and to pruned-Volterra models in chapter 7. In both cases, the DPD performance of the sparse models is superior than with more traditional models, such as the generalized memory polynomial (GMP) (MORGAN et al., 2006), for the same running complexity. Moreover, this thesis proposes a block-wise modular structure for the WHFB and pruned-Volterra models, which allows group LASSO techniques to be applied to select the most relevant model blocks. Based on this, we propose the sparse sizing of the models, i.e., setting appropriate values for the model parameters. Finally, also based on group-wise LASSO, we propose an approximate technique in which, first, the most relevant model blocks are selected, then, LASSO is run only over them, reducing the model estimation cost.

### 1.3.2 PA LMM Compensation

A traditional way to prevent power reflections caused by LMM from reaching the PA is to place an isolator at the PA's output port that absorbs the reflected power. This approach, however, is not power efficient and is very bulky for integrated transmitters. One widely employed, hardware-based compensation technique uses a tunable matching network (TMN) to adaptively adjust the PA load impedance (BEZOOIJEN; MAHMOUDI; ROERMUND, 2011), employing switchable or tunable circuit elements

such as micro-electromechanical systems (MEMS) components or switchable capacitors (MINGO; CRESPO; VALDOVINOS, 2002), (FU; MORTAZAWI, 2008). Some other hardware-based techniques have been proposed in the literature for compensating for PA LMM. The technique presented in (BEZOOIJEN; MAHMOUDI; ROERMUND, 2005) adaptively decreases the PA input power, depending on the measured output reflection coefficient, i.e. applies a variable input power back-off (IBO) to avoid the output voltage clipping. However, this technique also significantly reduces the transmitted power and the PA power efficiency. In (KEERTI; PHAM, 2007), it was experimentally observed that the PA distortion levels reached a minimum at a specific load impedance phase, thus, the authors proposed an adaptive circuit that only adjusts the phase seen by the PA. Still some other hardware-based compensation techniques employ adaptive analog pre-equalizers, such as in (PAVAN; SHIVAPPA, 2006).

Although a TMN could be employed for the load impedance adjustment (JI; JEON; KIM, 2015), (MINGO et al., 2004), some recent papers, as for example, (DHAR et al., 2018), (DUNN et al., 2017), (CAI; GONÇALVES; PEDRO, 2017), suppress the TMN and propose sophisticated load-aware DPD approaches, to be discussed in section 2.3. The main advantages of the digital approach, comparing to TMN, are the cost, size and flexibility. Furthermore, DPD can simultaneously account for RF and baseband load mismatch (BRINKHOFF; PARKER; LEUNG, 2003) and extend the linear range of the PA. However, the load-aware DPD approaches mentioned above require additional measurements, such as the knowledge of the load reflection coefficient, are prone to numerical instabilities and suffer from high estimation complexity (QI; BENEDIKT; TASKER, 2009), (ZARGAR; BANAI; PEDRO, 2015a). This thesis proposes the WHFB model as a simplified DPD behavioural model for LMM PAs. We justify this choice using the Volterra series analysis. Also, the performance of the model is verified experimentally. Moreover, the implementation of the proposed model is less complex than with DPD models for PA LMM proposed in the literature and discussed in section 5.2.

### 1.3.3 IQM Imbalance Compensation

When applying DPD to compensate for NL PAs, one difficulty is that IQM imbalance might compromise the performance of the DPD methods (DING et al., 2008), (CAVERS, 1997), thus motivating the joint compensation of IQM and PA imperfections (ANTTILA; HANDEL; VALKAMA, 2010), (KIM; JEONG; LEE, 2007). In fact, the separate compensation of them is typically an inefficient solution, because both impairments interact to each other. Furthermore, extra hardware, such as an additional coupler and RF switch, is required to sample both the IQM and the PA output signals (ANTTILA; VALKAMA; RENFORS, 2008b). (ANTTILA; HANDEL; VALKAMA, 2010) proposes the joint compensation of wideband NL PA and frequency-dependent IQM imbalance, using the

parallel Hammerstein DPD model, discussed in section 2.2, and simultaneously estimating all the joint DPD model coefficients through least squares regression. In turn, (CAO et al., 2008) uses the dual-input NL model based on two real-valued Volterra series models to jointly compensate for the IQM imbalance and NL PA. However, the complexity of the full-Volterra model limited the practical implementation of this proposal; in (SAFFAR et al., 2013), the dual-input Volterra model is replaced by the dual-input memory polynomial (MP) model. These models are presented and discussed in more details in the appendix A, where we also extend the DPD behavioural model proposed for PA LMM for the joint and simultaneous compensation of IQM imbalance and PA LMM.

## 1.4 THESIS CONTRIBUTIONS

The following are the main contributions of this thesis to the state-of-the-art PA and DPD modelling:

- In chapter 4, two reduced-complexity representations of the baseband full-Volterra model, using the PARAllel FACtors (PARAFAC) decomposition and the Laguerre expansion, are discussed and compared. The corresponding Volterra-PARAFAC (VP) and Laguerre-Volterra (LV) models are also compared to the WHFB and sparse WHFB models, to be discussed in chapters 5 and 6, and to the pruned-Volterra structure proposed in chapter 7.

- In chapter 5, the application of a physically-inspired pruned-Volterra model, namely, the WHFB model (MKADEM et al., 2014), is proposed as a simplified DPD behavioural model for PAs subject to LMM. This chapter combines the behavioural modelling approach with some prior knowledge about the operational/ environmental conditions of the PA, as discussed therein, to derive a simplified DPD model that is linear in the coefficients. The application of the WHFB DPD model for PA LMM is justified using the Volterra series analysis in this chapter and is also experimentally validated in chapter 6. Besides, the WHFB model is extended in the appendix A to an unified approach that simultaneously compensates for IQM imbalance and PA LMM, although it is not experimentally validated in this thesis.

- In chapter 6, the parsimonious, sparse estimation of the WHFB model is proposed, using LASSO and group-wise LASSO techniques. This additional step of CS-based unstructured pruning is able to further reduce the required number of degrees of freedom of the model, thus reducing the DPD filter length and, proportionally, the DPD running cost. This chapter also proposes a block-wise structure for the WHFB model, enabling its sparse estimation in blocks using group-LASSO (KRONVALL et al., 2016), (MEIER; GEER; BUHLMANN, 2008), (YUAN; LIN, 2006) and sparse-

group LASSO (SIMON et al., 2013) to obtain the sparse sizing of the model. Finally, this chapter also develops a simplified, approximate technique in which the most relevant WHFB blocks are selected prior to the LASSO estimation, resulting in lower estimation cost than with LASSO. The sparse WHFB model is compared to its dense counterpart from chapter 5, as well as to the VP and LV models.

- Finally, in chapter 7, a Volterra pruning strategy is proposed that leads to a flexible and modular memory polynomial, whose model parameters can be controlled independently for each NL order, thus avoiding the curse of dimensionality, while exploring higher polynomial dimensions than the MP and GMP models. The Group-wise LASSO technique is then applied to select the most relevant regressor blocks in the model and obtain its model sizing. This flexible approach is expected to accurately describe a wide range of PA operational/ environmental conditions.

## 1.5 PUBLICATIONS

The following papers were published during the doctorate:

- C. S. Hemsi, C. M. Panazio, "Adaptive baseband frequency-equalization for RF impedance matching correction", IEEE International Symposium on Circuits and Systems (ISCAS), May, 2017, Baltimore, MD, USA.
- C. S. Hemsi, C. M. Panazio, "Sparse Estimation Technique for Digital Pre-distortion of Impedance-Mismatched Power Amplifiers" Circuits, Systems, and Signal Processing, p. 1-29, 02-2021.

## 1.6 TEXT OUTLINE

The thesis is organized as follows. Chapter 2 presents state-of-the-art discrete-time baseband models to describe the major sources of RF impairments and their digital compensation. Chapter 3 introduces the figures of merit for evaluating the performance of PA and DPD models, also presenting a brief simulation study and the experimental test set-up employed in this thesis. Chapter 4 investigates the Volterra-PARAFAC and Laguerre-Volterra techniques for reducing the complexity of the baseband FV model. Chapter 5 proposes the use of the WHFB DPD model for PAs under LMM, including an extension to the joint compensation of PA LMM and IQM imbalance in the appendix A. In chapter 6, the WHFB model complexity is reduced using the parsimonious, sparse estimation of the model coefficients through LASSO and block-wise LASSO. Finally, in chapter 7, a flexible Volterra pruning technique is proposed and evaluated against other Volterra-based models. Chapter 8 presents the conclusions and perspectives.

# 2 SYSTEM MODEL

This chapter presents state-of-the-art discrete-time baseband models used in the literature for describing major sources of RF analog impairments in wireless communications transmitters, including the NL PA with memory effects, the PA LMM and IQM imbalance. In chapters 4 to 7, the models presented here are considered on the formulation of DPD approaches. In the next sections, we present baseband models for NL PAs, PA LMM and imbalanced IQMs.

## 2.1 BACKGROUND

Behavioural models (also known as black-box models) relate sampled input and output signals extracted from the system to be modelled. One advantage of such models is that they can be used for describing systems that are too complex for physical or circuit level modelling. Theoretically, in low-power and/or narrowband wireless communications systems, the PA output only depends on its instantaneous input signal, thus memoryless black-box PA models provide a precise description. A PA is strictly memoryless (BOSCH; GATTI, 1989), (RAICH; ZHOU, 2002), if it can be characterized by a real gain versus the input signal level, i.e. the PA does not exhibit any frequency response or phase variation inside its operating band. In practice, low-power and/or narrowband PAs are quasi-memoryless (BOSCH; GATTI, 1989), (RAICH; ZHOU, 2002), i.e., characterized by a complex gain versus the input signal level, or equivalently, by their amplitude-to-amplitude (AM/AM) and amplitude-to-phase (AM/PM) conversion functions. Well known quasi-memoryless PA models have been proposed for different types of PAs, e.g., the Saleh model (SALEH, 1981), for travelling-wave tube amplifiers, the Rapp model (RAPP, 1991), for solid state power amplifiers (SSPAs) and the Ghorbani model, for field-effect transistor SSPAs (CRIPPS, 2006). As communications systems evolve towards wideband and/or high-power, the PA models should include memory effects (KU; MCKINLEY; KENNEY, 2002), i.e. the PA output not only depends on the instantaneous input, but also on its history. The so-called electrical memory effects are mainly attributed to the PA's MN and BN and their impedance variations over the baseband, fundamental and harmonic frequency ranges. On the other hand, the thermal memory effects are mainly attributed to temperature changes in the transistor junction, as a function of present and past input signal levels (VUOLEVI; RAHKONEN, 2003). Memory effects are experimentally observable as dispersion around the AM/AM and AM/PM curves.

In this work, we consider polynomial behavioural models based on the full-Volterra

(FV) series, a very powerful modelling framework for NL systems with memory in general (SCHETZEN, 2006). Although PAs are passband devices, low-pass equivalent (LPE) polynomials, derived from the baseband version of the FV model, are widely used as approximate models for PAs and DPD, due to major advantages comparing to passband counterparts, as shown in the literature (NGOYA; MONS, 2014), (FEHRI; BOUMAIZA, 2014). The disadvantage of the FV series model is the large number of coefficients that need to be estimated, making this model not very useful in many practical cases. To overcome this problem, reduced complexity models are preferred, as discussed in (PEDRO; MAAS, 2005), (ISAKSSON; WISELL; RONNOW, 2006) and (RAAB et al., 2002), such as the memory polynomial (MP), (KIM; KONSTANTINOU, 2001), (DING et al., 2004), where only the diagonal terms of the FV series model are kept, the generalized memory polynomial (GMP) (MORGAN et al., 2006), where some cross-product terms of the input samples are allowed, and block-oriented models, which represent the NL model by interconnections of dynamic linear and static (memoryless) NL building blocks. The major advantage of such models is that they are easy to implement and estimate, using linear least squares.

Throughout this chapter, up to section 2.3, the PA notation is adopted for describing the models, e.g., $\hat{y}(n) = f\{\tilde{x}(n), \tilde{x}(n-1), \dots\}$, where $\tilde{x}(n)$ and $\hat{y}(n)$ are, respectively, the PA input and modelled output signals. In section 2.4, the IQM notation is adopted, i.e., $\hat{x}(n) = f\{\tilde{s}(n)\}$, where $\tilde{s}(n)$ is the IQM input signal. Starting from section 2.5, the DPD training notation is adopted, e.g., $\hat{x}(n) = f^{-1}\{\tilde{u}(n), \tilde{u}(n-1), \dots\}$, where $\tilde{u}(n)$ and $\hat{x}(n)$ are, respectively, the DPD training input and modelled output signals; unless the contrary is expressly stated.

## 2.2 VOLTERRA SERIES MODELS

This section presents the FV series model (SCHETZEN, 2006) and pruned-Volterra, polynomial models with memory, developed in the literature, that are very useful for wideband PA and DPD modelling. Major advantages of the Volterra-based polynomial models, comparing to other modelling frameworks, are a solid theoretical foundation and, in general, NL models that are linear in the parameters, thus allowing the model coefficients to be recovered by ordinary least squares (OLS) estimation. Figure 2.1 shows the direct-conversion transmitter block diagram, including the digital and analog signals used on the model expressions.

The NL, time-invariant, passband FV series model, truncated to NL order $K$, is given by a linear combination of NL, multidimensional basis functions:

$$y_{RF}(t) = \sum_{k=1}^{K} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h_k(\tau_{1:k}) \prod_{i=1}^{k} x_{RF}(t-\tau_i)\ d\tau_1 \dots d\tau_k, \tag{2.1}$$

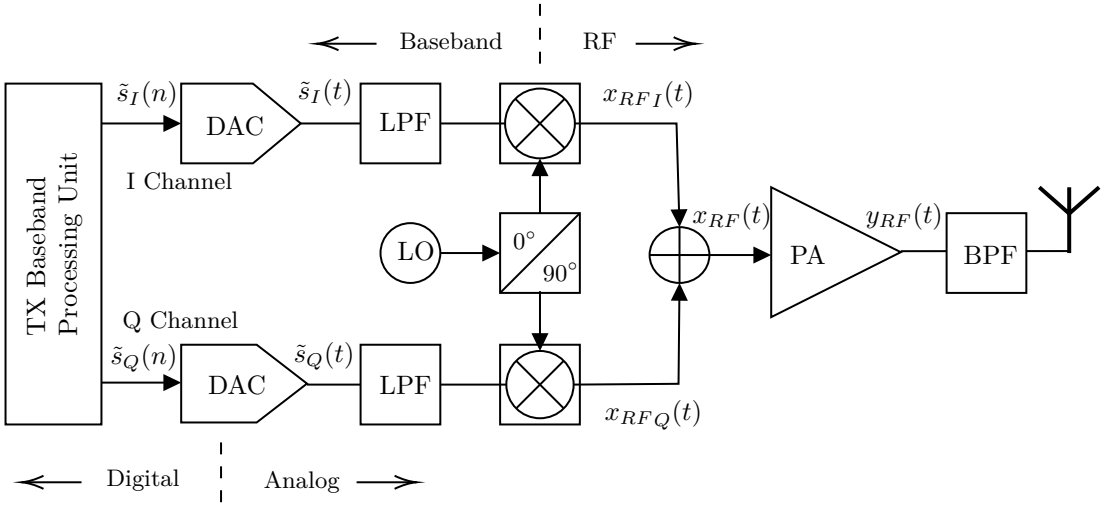



Figure 2.1 – Block diagram of the TX for impairment modelling.

where $x_{RF}(t)$ is the RF input signal and $h_k(\tau_{1:k})$, denoting $(\tau_{1:k}) = (\tau_1, \ldots, \tau_k)$, is the $k$-dimensional Volterra kernel.

### 2.2.1 Baseband Full-Volterra Models

The corresponding discrete-time baseband FV series model, truncated to memory $M+1$ and NL order $K$, is given by (FEHRI; BOUMAIZA, 2014) and references therein:

$$\begin{aligned}\hat{y}_{FV}(n) &= \sum_{\substack{k=1\\ k\ odd}}^{K} \tilde{y}_k(n) \\ &= \sum_{\substack{k=1\\ k\ odd}}^{K} \sum_{m_1=0}^{M} \cdots \sum_{m_k=0}^{M} \tilde{h}_k(m_{1:k}) \prod_{i=1}^{\frac{k+1}{2}} \tilde{x}(n-m_i) \prod_{i=\frac{k+3}{2}}^{k} \tilde{x}^*(n-m_i),\end{aligned} \tag{2.2}$$

where $\tilde{x}(n)$ and $\hat{y}_{FV}(n)$ are, respectively, the complex-valued, discrete-time, LPE PA input and modelled output signals at sample $n$, $(\cdot)^*$ is the conjugate operator, the time lags $\tau_i$ in eq. (2.1) are replaced by the sample lags $m_i$ and $\tilde{h}_k(m_{1:k})$ is the $k^{th}$-order full-Volterra kernel, denoting $(m_{1:k}) = (m_1, \ldots, m_k)$. Note that the RF signals and their LPE complex envelopes are related by $x_{RF}(t) = \Re\{\tilde{x}(t)e^{j\omega_c t}\}$, where $\omega_c$ is the RF carrier frequency. In LPE models, we are interested in frequency components close to $\omega_c$, implying that $k$ is odd, i.e., even-order NL kernels can be omitted (ISAKSSON; WISELL; RONNOW, 2006), and the number of $\tilde{x}$ terms in eq. (2.2) differs from the number of $\tilde{x}^*$ terms by exactly one. For example, for $k = 3$, eq. (2.2) is composed of the product of two $\tilde{x}$ terms and one $\tilde{x}^*$ term. Also note that in LPE models, a sampling frequency equal to a multiple of the envelope bandwidth is adopted (BONFIM; LIMA, 2016). Finally, note that a truncated Volterra series (finite dimensional), such as in eq. (2.2), is obtained by limiting the infinite Volterra series both up to NL order $K$ and memory depth $M+1$ and assuming that beyond

these limits the series is null. From the Stone–Weierstrass theorem (BISHOP, 1961) the truncated Volterra model is able to approximate any continuous non-linearity, by choosing appropriate values for $K$ and $M$. The model output is linear in the parameters and its total number of coefficients is given by:

$$N_{FV} = \sum_{\substack{k=1 \\ k\ odd}}^{K} (M+1)^k. \tag{2.3}$$

In eq. (2.2), the baseband Volterra kernels for $k \geq 3$ present doubly-symmetry, i.e., all permutations of indexes in each of the products $\prod_{i=1}^{\frac{k+1}{2}} \tilde{x}(n-m_i)$ and $\prod_{i=\frac{k+3}{2}}^{k} \tilde{x}^*(n-m_i)$ correspond to the same polynomials. Let $S_{\frac{k+1}{2}}$ be the finite group of permutations of $\frac{k+1}{2}$ elements. Then, if any arbitrary permutation $\pi \in S_{\frac{k+1}{2}}$ is applied to the sample delays of the non-conjugate terms in eq. (2.2), i.e., $\tilde{x}(n-m_i) \rightarrow \tilde{x}(n-m_{\pi(i)})$, the output signal $\hat{y}_{FV}(n)$ is unchanged. Likewise, the same applies for any arbitrary permutation $\pi_* \in S_{\frac{k-1}{2}}$ of the time delays of the conjugate terms. Thus, the triangular baseband Volterra model ($FV_{\triangle}$), more economical in its number of coefficients, can be obtained by exploiting the symmetry of the permutations and adjusting the sample delay indexes, as follows:

$$\begin{aligned}\hat{y}_{FV_{\triangle}}(n) = \sum_{\substack{k=1 \\ k\ odd}}^{K} \sum_{m_1=0}^{M} \sum_{m_2=m_1}^{M} \cdots \sum_{m_{\frac{k+1}{2}}=m_{\frac{k-1}{2}}}^{M} \cdots \sum_{m_{\frac{k+3}{2}}=0}^{M} \cdots \sum_{m_k=m_{k-1}}^{M} \tilde{h}_{k,\triangle}(m_{1:k}) \\ \prod_{i=1}^{\frac{k+1}{2}} \tilde{x}(n-m_i) \prod_{i=\frac{k+3}{2}}^{k} \tilde{x}^*(n-m_i).\end{aligned} \tag{2.4}$$

The $k^{th}$-order triangular kernel $\tilde{h}_{k,\triangle}(m_{1:k})$ in eq. (2.4) relates to the double-symmetric kernel $\tilde{h}_k(m_{1:k})$ in eq. (2.2) by:

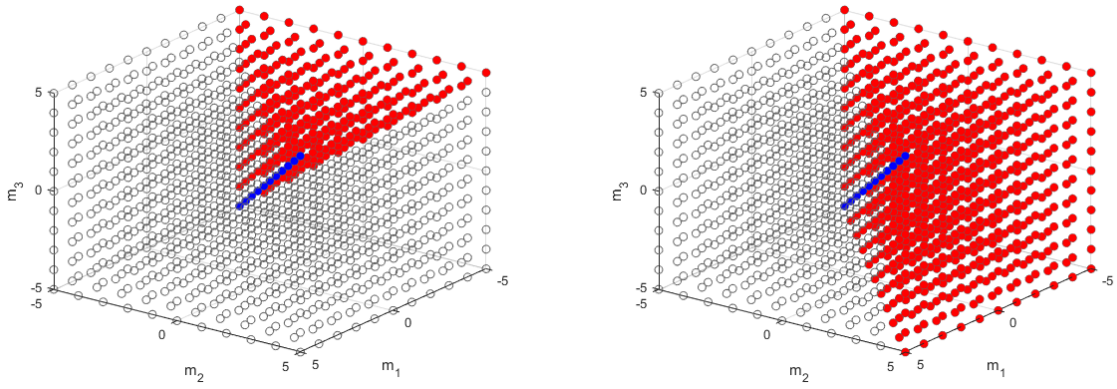
$$\tilde{h}_{k,\triangle}(m_{1:k}) = n_{\pi} n_{\pi_*} \tilde{h}_k(m_{1:k}), \tag{2.5}$$

for $m_1 \leq m_2 \leq \ldots \leq m_{\frac{k+1}{2}}$ and $m_{\frac{k+3}{2}} \leq m_{\frac{k+5}{2}} \leq \ldots \leq m_k$, otherwise $\tilde{h}_{k,\triangle}(m_{1:k})$ is zero. In the expression, $n_{\pi}$ and $n_{\pi_*}$ are the numbers of index permutations in the sets $\{m_1, \ldots, m_{\frac{k+1}{2}}\}$ and $\{m_{\frac{k+3}{2}}, \ldots, m_k\}$, respectively.

The number of model coefficients in the $FV_{\triangle}$ model reduces to:

$$N_{FV_{\triangle}} = \sum_{\substack{k=1 \\ k\ odd}}^{K} \binom{M+\frac{k+1}{2}}{\frac{k+1}{2}} \binom{M+\frac{k-1}{2}}{\frac{k-1}{2}}, \tag{2.6}$$

which still grows fast with the NL order and memory length.




(a) Passband symmetry (b) Baseband doubly-symmetry

Figure 2.2 – $3^{rd}$-order full-Volterra triangular kernels.

Figure 2.2 shows the elements in the $3^{rd}$-order kernels of the passband and baseband triangular Volterra models. The kernels are represented in a cubic space of the indexes $m_1$, $m_2$ and $m_3$ ranging from $-5$ to 5. The diagonal elements of each kernel ($m_1 = m_2 = m_3$) are highlighted in blue, whereas the remaining are in red. The symmetry in the passband case implies that the $3^{rd}$-order triangular kernel elements are such that $m_1 \leq m_2 \leq m_3$. In the baseband case, the dual-symmetry only implies $m_1 \leq m_2$, allowing all $m_3$. The distinct subsets of points in the cubic space illustrate the difference between symmetry and doubly-symmetry in this context.

In (STAUDINGER; NANAN; WOOD, 2010), the modified triangular Volterra ($MV_{\triangle}$) model is proposed, in which the memory depth for each NL order is chosen independently. In this way, the model dimensionality can be controlled, by setting larger memory spans for the lower non-linear orders of the model, as for example applied in (ZENTENO et al., 2015). The $MV_{\triangle}$ model is given by:

$$\hat{y}_{MV_{\triangle}}(n) = \sum_{\substack{k=1 \\ k\ odd}}^{K} \sum_{m_1=0}^{M_k} \sum_{m_2=m_1}^{M_k} \cdots \sum_{m_{\frac{k+1}{2}}=m_{\frac{k-1}{2}}}^{M_k} \cdots \sum_{m_{\frac{k+3}{2}}=0}^{M_k} \cdots \sum_{m_k=m_{k-1}}^{M_k} \tilde{h}_{k,\triangle}(m_{1:k}) \prod_{i=1}^{\frac{k+1}{2}} \tilde{x}(n-m_i) \prod_{i=\frac{k+3}{2}}^{k} \tilde{x}^*(n-m_i), \tag{2.7}$$

where $K$ is the NL order and $M_k + 1$ is the memory depth of the $k^{th}$ order polynomials and also denoting $(m_{1:k}) = (m_1, \ldots, m_k)$. The required number of coefficients is given by:

$$N_{MV_{\triangle}} = \sum_{\substack{k=1 \\ k\ odd}}^{K} \binom{M_k + \frac{k+1}{2}}{\frac{k+1}{2}} \binom{M_k + \frac{k-1}{2}}{\frac{k-1}{2}}. \tag{2.8}$$

However, the number of $MV_{\triangle}$ coefficients in eq. (2.8) is still given by a doubly-combinatorial expression, resulting in large numbers of coefficients to be estimated and high computational cost, as increase the NL order $K$ and the memory depths in the set $\{M_k\}$, $k = 1, \ldots, K$.

### 2.2.2 Memory Polynomial Models

Compact versions of the FV models, with lower computational complexity, have been widely applied to behavioural modelling and DPD. In this subsection, we present the baseband MP (KIM; KONSTANTINOU, 2001), (DING et al., 2004), GMP (MORGAN et al., 2006) and dynamic deviation reduction (DDR)(ZHU; DOOLEY; BRAZIL, 2006) models, all of them are very well accepted reduced-complexity versions of the FV model. The MP model corresponds to retaining only the diagonal terms of the Volterra model, while ignoring the off-diagonal terms (PEDRO; MAAS, 2005), as follows:

$$\hat{y}_{MP}(n) = \sum_{m=0}^{M} \sum_{\substack{k=1 \\ k\ odd}}^{K} \alpha_{m,k} \tilde{x}(n-m) \left|\tilde{x}(n-m)\right|^{k-1}, \tag{2.9}$$

where $\tilde{x}(n)$ and $\hat{y}_{MP}(n)$ are, respectively, the discrete-time baseband PA input and modelled output signals, $M$ is the memory depth, $K$ is the NL polynomial order and the vector $\boldsymbol{\alpha}$ is the model coefficients vector with size $(M+1)\frac{K+1}{2} \times 1$. From eq. (2.9), the PA coefficients vector has the following elements:

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_{0,1} & \ldots & \alpha_{0,K} & \ldots & \alpha_{M,1} & \ldots & \alpha_{M,K} \end{bmatrix}^T. \tag{2.10}$$

Note that the baseband model in eq. (2.9) only considers the terms that model the PA output at the fundamental frequency band (FEHRI; BOUMAIZA, 2014). As depicted in Figure 2.3, the MP model can also be seen as the summation of memoryless polynomials $f_m(n) = \sum_{k=1}^{K} \alpha_{m,k} \tilde{x}(n-m) \left|\tilde{x}(n-m)\right|^{k-1}$, where $m = 0, \ldots, M$ is the sample delay, also represented by $z^{-1}$ (DING et al., 2004). Besides, the MP model structure corresponds to the reduced parallel Hammerstein model, to be described in the next subsection.

In order to compute the MP model output signal, $\hat{y}_{MP}(n)$, at sample $n$, we assemble the vector of powered delayed input samples, $\tilde{\boldsymbol{x}}(n)$, as follows:

$$\tilde{\boldsymbol{x}}(n) = \begin{bmatrix} \tilde{x}(n) \\ \vdots \\ \tilde{x}(n)\,|\tilde{x}(n)|^{K-1} \\ \vdots \\ \tilde{x}(n-M) \\ \vdots \\ \tilde{x}(n-M)\,|\tilde{x}(n-M)|^{K-1} \end{bmatrix}^T, \tag{2.11}$$

where $(\cdot)^T$ is the transpose operator, $M$ is the memory depth, $K$ is the NL order and the vector $\tilde{\boldsymbol{x}}(n)$ has size $1 \times (M+1)\frac{K+1}{2}$. Thus, given the PA coefficients vector $\boldsymbol{\alpha}$, the modelled output signal is calculated as:

$$\hat{y}_{MP}(n) = \tilde{\boldsymbol{x}}(n)\boldsymbol{\alpha}. \tag{2.12}$$

Alternatively, the MP model coefficients can be identified by OLS from a batch of $N$ samples of the PA input and output signals. First, the regression matrix of the measured PA input samples is assembled, for $n = 0, \ldots, N-1$, as follows:

$$\tilde{\boldsymbol{X}} = \begin{bmatrix} \tilde{\boldsymbol{x}}(0) \\ \vdots \\ \tilde{\boldsymbol{x}}(N-1) \end{bmatrix}, \tag{2.13}$$

where $\tilde{\boldsymbol{x}}(n)$ is a row vector defined by eq.(2.11) and the matrix $\tilde{\boldsymbol{X}}$ has size $N \times (M+1)\frac{K+1}{2}$. Each column of $\tilde{\boldsymbol{X}}$ is a regressor of the model.

Also, the measured PA output samples are collected in the following vector:

$$\tilde{\boldsymbol{y}} = \begin{bmatrix} \tilde{y}(0) \\ \vdots \\ \tilde{y}(N-1) \end{bmatrix}. \tag{2.14}$$

The OLS estimation of the coefficients is performed by minimizing the cost function $J(\boldsymbol{\alpha})$ that is the mean squared error $\tilde{\boldsymbol{e}}$ of approximating the observed PA output signal $\tilde{\boldsymbol{y}}$ by the modelled output signal $\hat{\boldsymbol{y}}_{MP} = \tilde{\boldsymbol{X}}\boldsymbol{\alpha}$, for $N$ samples, as defined below:

$$\begin{aligned} J(\boldsymbol{\alpha}) &= \sum_{n=0}^{N-1} |\tilde{e}(n)|^2 \\ &= \|\tilde{\boldsymbol{y}} - \hat{\boldsymbol{y}}_{MP}\|_2^2 \\ &= \|\tilde{\boldsymbol{y}} - \tilde{\boldsymbol{X}}\boldsymbol{\alpha}\|_2^2, \end{aligned} \tag{2.15}$$

where $\|\cdot\|_2$ is the Euclidean $\ell_2$ norm. Then, the OLS solution is given by:

$$\hat{\boldsymbol{\alpha}} = (\tilde{\boldsymbol{X}}^H \tilde{\boldsymbol{X}})^{-1} \tilde{\boldsymbol{X}}^H \tilde{\boldsymbol{y}}, \tag{2.16}$$

where $(\cdot)^H$ denotes the complex conjugate transpose, $\hat{\boldsymbol{\alpha}}$ is the estimated coefficients vector and $\tilde{\boldsymbol{X}}^\dagger = (\tilde{\boldsymbol{X}}^H \tilde{\boldsymbol{X}})^{-1} \tilde{\boldsymbol{X}}^H$ is the Moore-Penrose pseudo-inverse of the matrix $\tilde{\boldsymbol{X}}$, assuming $\tilde{\boldsymbol{X}}^H \tilde{\boldsymbol{X}}$ is invertible. This is discussed in more details in section 2.6.

In the GMP model (MORGAN et al., 2006), some cross-product terms of the signal and lagging/leading even-order envelope powers from the FV model are preserved, as follows:

$$\begin{aligned} \hat{y}_{GMP}(n) = & \sum_{m=0}^{M_1} \sum_{\substack{k=1 \\ k\ odd}}^{K_1} \alpha_{m,k}^{(1)} \tilde{x}(n-m) \left|\tilde{x}(n-m)\right|^{k-1} \\ & + \sum_{m=0}^{M_2} \sum_{\substack{k=3 \\ k\ odd}}^{K_2} \sum_{l=1}^{L_2} \alpha_{m,k,l}^{(2)} \tilde{x}(n-m) \left|\tilde{x}(n-m-l)\right|^{k-1} \\ & + \sum_{m=0}^{M_3} \sum_{\substack{k=3 \\ k\ odd}}^{K_3} \sum_{l=1}^{L_3} \alpha_{m,k,l}^{(3)} \tilde{x}(n-m) \left|\tilde{x}(n-m+l)\right|^{k-1}, \end{aligned} \tag{2.17}$$

where parameters $M_1$ and $K_1$ are, respectively, the memory depth and power order of the terms between signal and aligned powered envelope, $M_2$, $L_2$ and $K_2$ are, respectively, the memory depth, envelope time-shift and power order for the terms between signal and lagging powered envelope and $M_3$, $L_3$ and $K_3$ are, respectively, the memory depth, envelope time-shift and power order for the terms between signal and leading powered envelope. Note that the GMP model reduces to the MP when the cross-terms are discarded.

In the DDR model (ZHU; DOOLEY; BRAZIL, 2006), (ZHU; PEDRO; BRAZIL, 2006), first, the passband FV model is separated into its static and dynamic parts, as

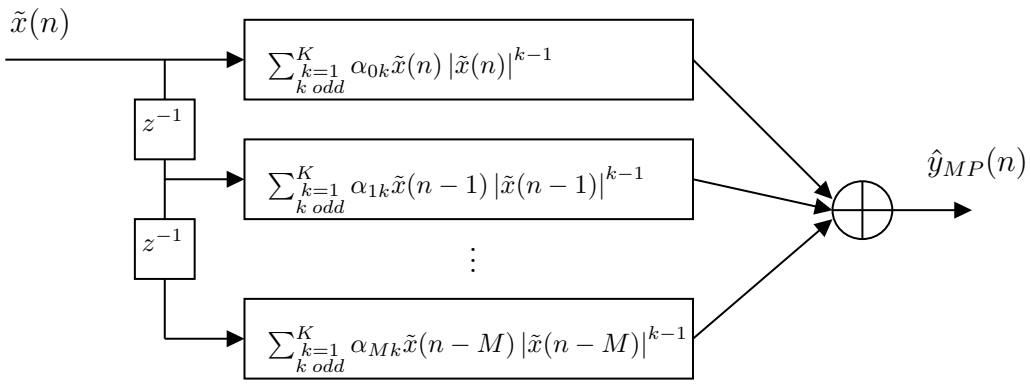



Figure 2.3 – PA baseband MP model.

follows:

$$\hat{y}_{RF}(n) = y_{RF}^{(S)}(n) + y_{RF}^{(D)}(n) = \sum_{\substack{k=1 \\ k\ odd}}^{K} h_k x_{RF}^k(n) + \sum_{\substack{k=1 \\ k\ odd}}^{K} \sum_{r=1}^{k} x_{RF}^{k-r}(n) \sum_{m_1=1}^{M} \dots \sum_{m_r=m_{r-1}}^{M} h_{k,r}(m_{1:r}) \prod_{i=1}^{r} x_{RF}(n-m_i), \tag{2.18}$$

where $y_{RF}^{(S)}(n)$ and $y_{RF}^{(D)}(n)$ are, respectively, the static and dynamic model polynomials. Next, assuming that the dynamic effects tend to fade with increasing NL orders, the higher orders of dynamics are removed from the model (ZHU; PEDRO; BRAZIL, 2006). Therefore, the pruned-Volterra DDR model is obtained by limiting $r$ on the $y_{RF}^{(D)}(n)$ term. In LPE notation, the first order DDR model ($r = 1$) is given by:

$$\hat{y}_{DDR-1}(n) = \sum_{\substack{k=1 \\ k\ odd}}^{K} \sum_{m=0}^{M} \alpha_k^{(1)}(m)\tilde{x}(n-m)|\tilde{x}(n)|^{k-1} + \sum_{\substack{k=3 \\ k\ odd}}^{K} \sum_{m=1}^{M} \alpha_k^{(2)}(m)\tilde{x}^2(n-m)\tilde{x}^*(n-m)|\tilde{x}(n)|^{k-3}. \tag{2.19}$$

Even more sophisticated PA models have been proposed by pruning the FV model, such as in (CHEANG; MAK; MARTINS, 2018), (LANDIN et al., 2015), (DU et al., 2013), (RAHATI; MOTAMEDI; SHARIFIAN, 2017), (PEDRO; MAAS, 2005). One such approach, block-oriented models, is described next.

### 2.2.3 Block-oriented Models

Block-oriented NL models are constructed by the series, parallel and feedback interconnections of two basic building blocks: a dynamic linear time-invariant (LTI) block

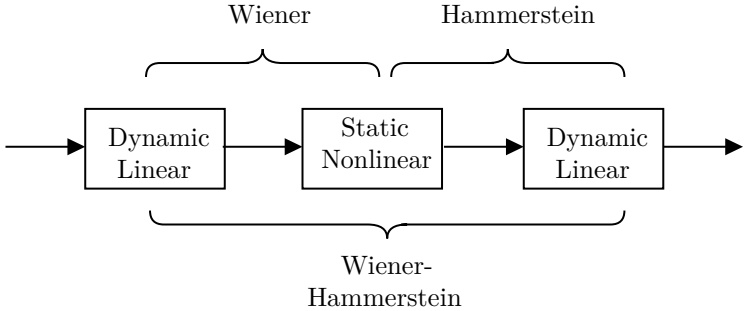



Figure 2.4 – PA Baseband block-oriented models.

and a static (memoryless) non-linear (SNL) block. The series (or cascade) Wiener and Hammerstein models are the simplest structures (PEDRO; MAAS, 2005), (ISAKSSON; WISELL; RONNOW, 2006). A Wiener model is composed of the cascade connection of an LTI block followed by an SNL block, as in Figure 2.4, and describes memory effects at the input of a NL system. Popular choices are power series for the SNL block and digital filters, mostly finite impulse response (FIR) filters, for the LTI block. Similarly, a Hammerstein model is composed of the cascade connection of an SNL block followed by an LTI block (also in Figure 2.4) and describes memory effects at the output of the system. The parallel Wiener and parallel Hammerstein models are obtained by connecting, respectively, several Wiener and Hammerstein branches in parallel.

The Wiener series model is described by the cascade connection of the two subsystems below:

$$\tilde{w}(n) = \sum_{m=0}^{M} \tilde{g}_W(m)\tilde{x}\,(n-m)\,, \tag{2.20}$$

where $\tilde{g}_W(0), \ldots, \tilde{g}_W(M)$ are the the impulse response coefficients of the LTI subsystem, and

$$\hat{y}(n) = \sum_{\substack{k=1 \\ k\ odd}}^{K} \alpha_k \tilde{w}(n)\,|\tilde{w}(n)|^{k-1}\,, \tag{2.21}$$

where $\tilde{x}(n)$, $\tilde{w}(n)$ and $\hat{y}(n)$ are, respectively, the discrete-time baseband input, intermediate and modelled output signals of the Wiener series model and $\alpha_k$ are the memoryless non-linearity coefficients. Note that the chosen non-linear expression includes only odd-order NL terms, known to be largely responsible for in-band and adjacent channel distortions (ISAKSSON; WISELL; RONNOW, 2006). The resulting expression is given by:

$$\hat{y}_W\,(n) = \sum_{\substack{k=1 \\ k\ odd}}^{K} \alpha_k \sum_{m=0}^{M} \tilde{g}_W(m)\tilde{x}\,(n-m) \left| \sum_{m=0}^{M} \tilde{g}_W(m)\tilde{x}\,(n-m) \right|^{k-1}, \tag{2.22}$$

where the powered term above is equivalent to:

$$\left| \sum_{m=0}^{M} \tilde{g}_W(m)\tilde{x}\,(n-m) \right|^{k-1} \equiv \left| \sum_{m_1=0}^{M} \cdots \sum_{m_{k-1}=0}^{M} \tilde{g}_W(m_1) \ldots \tilde{g}_W(m_{k-1})\tilde{x}\,(n-m_1) \ldots \tilde{x}\,(n-m_{k-1}) \right|. \tag{2.23}$$

The Wiener model is a special case of the Volterra model, in which:

$$\tilde{h}_k(m_{1:k}) = \alpha_k \prod_{\substack{i=1 \\ k\ odd}}^{k} \tilde{g}_W(m_i), \tag{2.24}$$

where each $m_i = 0, \ldots, M$. Note that this model depends non-linearly on the coefficients $\tilde{g}_W$.

In the parallel Wiener model, the overall output is the sum of the parallel outputs, resulting in:

$$\hat{y}_{PW}(n) = \sum_{p=1}^{P} \left\{ \sum_{\substack{k=1 \\ k\ odd}}^{K} \alpha_k^{(p)} \sum_{m=0}^{M_p} \tilde{g}_W^{(p)}(m) \tilde{x}(n-m) \left| \sum_{m=0}^{M_p} \tilde{g}_W^{(p)}(m) \tilde{x}(n-m) \right|^{k-1} \right\}, \tag{2.25}$$

where $P$ is the number of distinct branches. The parallel Wiener model is a special case of the Volterra model, in which: $\tilde{h}_k(m_{1:k}) = \sum_{p=1}^{P} \alpha_k^{(p)} \prod_{\substack{i=1 \\ k\ odd}}^{k} \tilde{g}_W^{(p)}(m_i)$.

Likewise, the Hammerstein model is described by the cascade connection of the two subsystems below:

$$\tilde{w}(n) = \sum_{\substack{k=1 \\ k\ odd}}^{K} \alpha_k \tilde{x}(n) \left| \tilde{x}(n) \right|^{k-1}, \tag{2.26}$$

where $\alpha_k$ are the memoryless non-linearity coefficients, and

$$\hat{y}(n) = \sum_{m=0}^{M} \tilde{g}_H(m) \tilde{w}(n-m), \tag{2.27}$$

where $\tilde{x}(n)$, $\tilde{w}(n)$ and $\hat{y}(n)$ are, respectively, the discrete-time baseband input, intermediate and modelled output signals of the Hammerstein series model and $\tilde{g}_H(0), \ldots, \tilde{g}_H(M)$ are the the impulse response coefficients of the LTI subsystem. The resulting expression is given by:

$$\hat{y}_H(n) = \sum_{m=0}^{M} \tilde{g}_H(m) \sum_{\substack{k=1 \\ k\ odd}}^{K} \alpha_k \tilde{x}(n-m) \left| \tilde{x}(n-m) \right|^{k-1}. \tag{2.28}$$

The Hammerstein model is a special case of the Volterra model, in which:

$$\tilde{h}_k(m_{1:k}) = \begin{cases} \alpha_k \tilde{g}_H(m), & \text{if } m_1 = m_2 =, \ldots, = m_k = m, \\ 0, & \text{otherwise.} \end{cases} \tag{2.29}$$

In this case, only the diagonal elements of the Volterra kernels are non-zero.

In the parallel Hammerstein (PH) model, the overall output is the sum of the parallel outputs, resulting in:

$$\hat{y}_{PH}(n) = \sum_{p=1}^{P} \left\{ \sum_{m=0}^{M_p} \tilde{g}_H^{(p)}(m) \sum_{\substack{k=1 \\ k\ odd}}^{K} \alpha_k^{(p)} \tilde{x}(n-m) \left| \tilde{x}(n-m) \right|^{k-1} \right\}, \tag{2.30}$$

where $P$ is the number of distinct branches. The parallel Hammerstein model is a special case of the Volterra model, in which: $\tilde{h}_k(m_{1:k}) = \sum_{p=1}^{P} \alpha_k^{(p)} \tilde{g}_H^{(p)}(m)$, if $m_1 = m_2 =, \ldots, = m_k = m$ and zero, otherwise.

To avoid any redundancy that results from considering all NL orders in each of the branches, a reduced parallel Hammerstein model can be also defined by:

$$\hat{y}_{PH}(n) = \sum_{\substack{k=1 \\ k\ odd}}^{K} \sum_{m=0}^{M_k} \alpha_k \tilde{g}_H^{(k)}(m) \tilde{x}(n-m) \left|\tilde{x}(n-m)\right|^{k-1}, \tag{2.31}$$

i.e., in this case the $(\frac{k+1}{2})^{th}$-branch only has the $k^{th}$-order non-linearity and the model is linear in the coefficients $\alpha_k \tilde{g}_H^{(k)}(m)$, for $m = 0, \ldots, M_k$.

Another common block structure is the Wiener-Hammerstein (LTI-SNL-LTI), whose model output is given by:

$$\hat{y}_{WH}(n) = \sum_{l=0}^{M_g} \tilde{g}_H(l) \sum_{\substack{k=1 \\ k\ odd}}^{K} \alpha_k \sum_{m=0}^{M_h} \tilde{g}_W(m) \tilde{x}(n-m-l) \left| \sum_{m=0}^{M_h} \tilde{g}_W(m) \tilde{x}(n-m-l) \right|^{k-1}. \tag{2.32}$$

The Wiener-Hammerstein model is a special case of the Volterra model with:

$$\tilde{h}_k(m_{1:k}) = \alpha_k \sum_{l=0}^{M_g} \tilde{g}_H(l) \prod_{\substack{k=1 \\ k\ odd}}^{k} \tilde{g}_W(m_i - l). \tag{2.33}$$

### 2.2.4 The Wiener-Hammerstein with Feedback (WHFB) Model

Finally, another way to increase the model capabilities is by introducing feedback to the block-oriented model structure, as in the physically-inspired feedback topology proposed in (PEDRO; CARVALHO; LAVRADOR, 2003), (ZHU; PEDRO; CUNHA, 2007). The Wiener-Hammerstein with feedback (WHFB) model, developed in (MKADEM et al., 2014), is obtained by simplifying the feedback structure presented in Figure 2.5. In the figure, the discrete-time low-pass equivalent model is composed of two FIR filters in series with the static (memoryless) NL block and a feedback FIR filter around the NL block. The two FIR filters in series model the linear memory effects, whereas the feedback FIR filter models nonlinear memory effects. The simplification adopted in the following model development is proposed by (MKADEM et al., 2014) and leads to the simplified block diagram shown in Figure 2.6. Comparing to the MP and GMP models, the WHFB includes important interactions of the input signal, envelope powers and lagging/leading envelope powers that are based on physical characteristics of the PA.

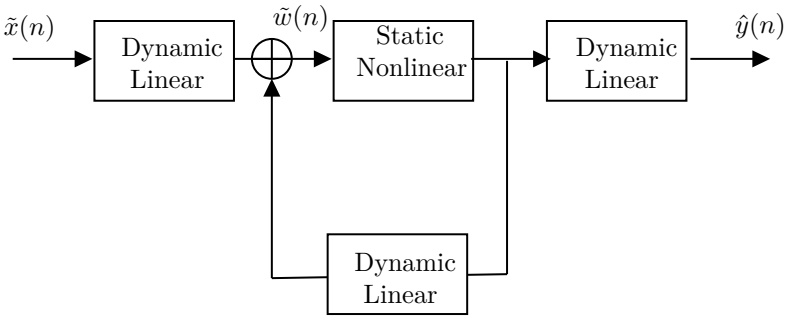



Figure 2.5 – PA baseband WHFB model.

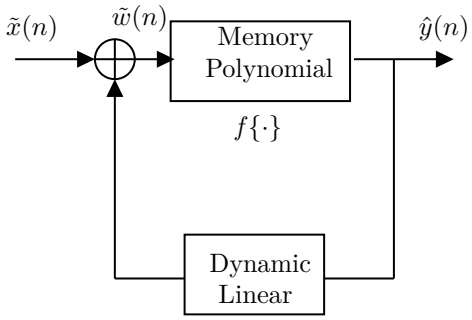



Figure 2.6 – PA simplified baseband WHFB model.

Thus, we define the transistor polynomial dynamic non-linearity, $f\{\cdot\}$, based on the MP model, as:

$$f\{\tilde{w}(n)\} = \sum_{m=0}^{M} \sum_{\substack{k=1 \\ k\ odd}}^{K} \alpha_{m,k} \tilde{w}^k (n-m), \tag{2.34}$$

where $M+1$ and $K$ are the transistor block memory depth and NL order and $\boldsymbol{\alpha}$ is the MP model coefficients vector. From Figure 2.6, the output signal $\tilde{y}(n)$ can be expressed as:

$$\tilde{y}(n) = f\Big\{\tilde{x}(n) + \sum_{l=1}^{L} \tilde{\gamma}(l)\tilde{y}(n-l)\Big\}, \tag{2.35}$$

where $\tilde{\gamma}(l)$ is the impulse response of the FIR filter that models the feedback mechanism, whose length is $L$. Notice that the summation terms in eq. (2.35) are output samples passing through the feedback block i.e., being transferred back to input. These output samples are then remixed with input samples in the transistor's NL block.

Thus, the output signal $\tilde{y}(n)$ can be written as:

$$\tilde{y}(n) = \sum_{m=0}^{M} \sum_{\substack{k=1 \\ k\ odd}}^{K} \alpha_{m,k} \Big(\tilde{x}(n-m) + \sum_{l=1}^{L} \tilde{\gamma}(l)\tilde{y}(n-m-l)\Big)^k. \tag{2.36}$$

Firstly, in order to simplify this equation, we approximate the reflected (delayed) output signal terms $\tilde{\gamma}(l)\tilde{y}(n-m-l)$. Using eqs. (2.34) and (2.35), we express the term

$\tilde{\gamma}(l_0)\tilde{y}\,(n-j_0)$ for any particular delay index $j_0 = m_0 + l_0$, such that $0 \leq m_0 \leq M$ and $1 \leq l_0 \leq L$, as:

$$\tilde{\gamma}(l_0)\tilde{y}\,(n-j_0) = \sum_{m=0}^{M} \sum_{\substack{k=1 \\ k\ odd}}^{K} \alpha_{m,k}\tilde{\gamma}(l_0)\Big(\tilde{x}(n-j_0-m) + \sum_{l=1}^{L}\tilde{\gamma}(l)\tilde{y}\,(n-j_0-m-l)\Big)^k. \tag{2.37}$$

In eq. (2.37), all the terms containing $\tilde{\gamma}^i(l)$ for $i = 2, \ldots, k$ represent multiple cycles through the feedback path and are assumed negligible for power orders higher than one, since $|\tilde{\gamma}(l)| < 1$, for $l = 1, \ldots, L$, given that the LTI filter emulates a passive network. Thus, such terms can be omitted from eq. (2.37), which simplifies to:

$$\begin{aligned}\tilde{\gamma}(l_0)\tilde{y}\,(n-j_0) &\approx \tilde{\gamma}(l_0)\sum_{m=0}^{M}\sum_{\substack{k=1 \\ k\ odd}}^{K}\alpha_{m,k}\tilde{x}^k(n-j_0-m) \\ &= \tilde{\gamma}(l_0)f\Big\{\tilde{x}(n-j_0)\Big\}.\end{aligned} \tag{2.38}$$

Next, in order to simplify eq. (2.36), we develop the following binomial expansion, using the result from eq. (2.38):

$$\begin{aligned}\Big(\tilde{x}(n-m) + \sum_{l=1}^{L}\tilde{\gamma}(l)\tilde{y}\,(n-m-l)\Big)^k &= \sum_{i=0}^{k}\binom{k}{i}\tilde{x}^{k-i}(n-m)\Big(\sum_{l=1}^{L}\tilde{\gamma}(l)f\{\tilde{x}(n-m-l)\}\Big)^i \\ &= \tilde{x}^k(n-m) + (k-1)\tilde{x}^{k-1}(n-m) \\ &\quad \sum_{l=1}^{L}\tilde{\gamma}(l)f\{\tilde{x}(n-m-l)\} \\ &\quad + \sum_{i=2}^{k}\binom{k}{i}\tilde{x}^{k-i}(n-m)\Big(\sum_{l=1}^{L}\tilde{\gamma}(l)f\{\tilde{x}(n-m-l)\}\Big)^i,\end{aligned} \tag{2.39}$$

where the terms containing $\tilde{\gamma}^i(l)$ are assumed as negligible for $i = 2, \ldots, q$. Thus, eq. (2.36) can be simplified to:

$$\tilde{y}(n) \approx \sum_{m=0}^{M}\sum_{\substack{k=1 \\ k\ odd}}^{K}\alpha_{m,k}^{(1)}\tilde{x}^k(n-m) + \sum_{m=0}^{M}\sum_{\substack{k=1 \\ k\ odd}}^{K}(k-1)\alpha_{m,k}^{(2)}\tilde{x}^{k-1}(n-m)\sum_{l=1}^{L}\tilde{\gamma}(l)f\{\tilde{x}(n-m-l)\}. \tag{2.40}$$

Eq. (2.40) can also be written as:

$$\begin{aligned}\tilde{y}(n) \approx & \sum_{m=0}^{M}\sum_{\substack{k=1\\ k\ odd}}^{K}\alpha_{m,k}^{(1)}\tilde{x}^{k}(n-m) \\ & + \sum_{m_1=0}^{M}\sum_{\substack{k_1=1\\ k_1\ odd}}^{K}(k_1-1)\alpha_{m_1,k_1}^{(2,1)}\tilde{x}^{k_1-1}(n-m_1)\sum_{l=1}^{L}\tilde{\gamma}(l)\sum_{m_2=0}^{M}\sum_{\substack{k_2=1\\ k_2\ odd}}^{K}\alpha_{m_2,k_2}^{(2,2)} \\ & \tilde{x}^{k_2}(n-m_1-m_2-l).\end{aligned} \tag{2.41}$$

By limiting the cross-term memory span to $L$, removing redundant cross-terms and limiting the NL order, eq. (2.41) simplifies to:

$$\begin{aligned}\tilde{y}(n) = & \sum_{m=0}^{M}\sum_{\substack{k=1\\ k\ odd}}^{K}\alpha_{m,k}^{(1)}\tilde{x}^{k}(n-m) + \sum_{m=0}^{M}\sum_{\substack{k=1\\ k\ odd}}^{K}\sum_{l=1}^{L}\sum_{\substack{q=1\\ q\ odd}}^{Q}(k-1)\alpha_{m,k}^{(2,1)}\alpha_{m,q}^{(2,2)}\tilde{\gamma}(l)\tilde{x}^{k}(n-m) \\ & \tilde{x}^{q}(n-m-l).\end{aligned} \tag{2.42}$$

Finally, we keep only terms that contribute to the in-band (fundamental) response (LIMA; CUNHA; PEDRO, 2011) and collect the parameters into the coefficients vectors $\boldsymbol{\alpha}^{(1)}$, $\boldsymbol{\alpha}^{(2)}$ and $\boldsymbol{\alpha}^{(3)}$. The resulting discrete-time baseband WHFB expression is given by:

$$\begin{aligned}\hat{y}_{WHFB}(n) = & \sum_{m=0}^{M_1}\sum_{\substack{k=1\\ k\ odd}}^{K_1}\alpha_{m,k}^{(1)}\tilde{x}(n-m)\left|\tilde{x}(n-m)\right|^{k-1} + \sum_{m=0}^{M_2}\sum_{l=1}^{L_2}\sum_{\substack{k=1\\ k\ odd}}^{K_2}\sum_{\substack{q=2\\ q\ even}}^{Q_2}\alpha_{m,l,k,q}^{(2)}\tilde{x}(n-m) \\ & \left|\tilde{x}(n-m)\right|^{k-1}\left|\tilde{x}(n-m-l)\right|^{q} + \sum_{m=0}^{M_3}\sum_{l=1}^{L_3}\sum_{\substack{k=0\\ k\ even}}^{K_3}\sum_{\substack{q=1\\ q\ odd}}^{Q_3}\alpha_{m,l,k,q}^{(3)}\tilde{x}(n-m-l) \\ & \left|\tilde{x}(n-m-l)\right|^{q-1}\left|\tilde{x}(n-m)\right|^{k},\end{aligned} \tag{2.43}$$

where $\hat{y}_{WHFB}(n)$ is the modelled baseband PA output signal, $\tilde{x}(n)$ is the input PA signal, parameters $M_1$ and $K_1$ are, respectively, memory depth and power order for the terms in the first summation. $M_2$, $L_2$, $K_2$ and $Q_2$ are, respectively, the memory depth, envelope time-shift, aligned envelope power order and time-shifted envelope power order for the terms in the second summation. $M_3$, $L_3$, $K_3$ and $Q_3$ are, respectively, the memory depth, envelope time-shift, time-shifted envelope power order and aligned envelope power order for the terms in the third summation. $\boldsymbol{\alpha}^{(1)}$, $\boldsymbol{\alpha}^{(2)}$ and $\boldsymbol{\alpha}^{(3)}$ are the PA model coefficients vectors. Note that a similar model is also derived in (BONFIM; LIMA, 2016).

### 2.2.5 Discussion

NL PA and DPD models based on the $FV_{\triangle}$ model in eq. (2.4) require a large number of coefficients to be estimated. In reality, practical PA and DPD systems can be

described by only a subset of the Volterra regressors (ENZINGER, 2018), (TOSINA et al., 2015). However, choosing the pruned-Volterra structure that accurately describes an unknown PA or DPD is not an easy problem. There is no guarantee that any of the memory polynomial or block-oriented alternatives presented in section 2.2 accurately models an unknown NL system, especially when no physical knowledge is available (ISAKSSON; WISELL; RONNOW, 2006). In addition, several other models have also been proposed, each taking into account some additional non-linearities (GHANNOUCHI; HAMMI, 2009), (CRESPO-CADENAS et al., 2010). In fact, the search for the best regression basis set is a combinatorial optimization task. This is the major disadvantage of models such as the MP, GMP and DDR, compared to those developed based on physical knowledge of a specific system. In contrast, models that take into account physical knowledge of the PA, such as the operating/ environmental conditions, are difficult to obtain, more complex and less general (ISAKSSON; WISELL; RONNOW, 2006).

Sections 2.3 and 2.4 present pruned-Volterra models developed in the literature specifically for the PA LMM and IQ imbalance conditions, respectively.

## 2.3 PA LMM MODELS

As described in (PEDRO; MAAS, 2005), (SCHREURS et al., 2008), the PA's transistor generates at its output an amplified version of the input signal that contains NL distortions whose spectra spans from baseband to harmonics of the operating frequency. The MN and BN placed at the PA output, as shown in Figure 2.7, are designed to provide an appropriate impedance termination for each component of the output signal at the envelope, fundamental and harmonic frequency bands. At the fundamental band, deviations on the load impedance produce a non-constant frequency response of the output MN, which causes frequency-dependent gain and phase responses, i.e., memory effects. At the baseband and harmonic bands, MN and BN not properly terminated due to deviations in the load impedance, DC supply voltage, temperature and other operating conditions, such as the instantaneous power and frequency (CRIPPS, 2006) contribute to NL memory effects. In both cases, distortion terms of the output signal not properly terminated reflect back and remix with the input signal inside the transistor, increasing the transistor memory effects (VUOLEVI; RAHKONEN, 2003), (CARVALHO; PEDRO, 2002), (MAZIERE et al., 2005), (NIELSEN et al., 2005), (MESSAOUDI et al., 2007), (ZHU; PEDRO; CUNHA, 2007), (LANDIN; BENGTSSON; ISAKSSON, 2009), (ZENTENO; ISAKSSON; HANDEL, 2015). The increased memory effects caused by LMM were experimentally verified by (WILLIAMS; LECKEY; TASKER, 2002), (CABRAL; PEDRO; CARVALHO, 2006) and others. The BN impedance mismatch (CARVALHO; PEDRO, 2000) also manifests as asymmetry between the two sides of the $3^{rd}$ order intermodulation distortion (IMD)

spectrum, as experimentally verified in (BOSCH; GATTI, 1989), (SEVIC; BURGER; STEER, 1998). Several works have concluded that the MN and BN frequency dependencies are the main sources of gain and phase non-linearities and electrical memory effects in PAs (PEDRO; MAAS, 2005).

The major load impedance dependent PA models proposed in the literature can be grouped into:

- Physically-inspired models, such as in (LANDIN et al., 2015), (PEDRO; CARVALHO; LAVRADOR, 2003), (ZHU; PEDRO; CUNHA, 2007).

- Models based on non-linear vector network analyzer (NVNA) measurements, such as X-Parameters (VERSPECHT et al., 2005), (SIMPSON et al., 2008), S-functions (VERBEYST; BOSSCHE, 2005), the Cardiff (WOODINGTON et al., 2008) models and load-pull models (PEDRO; NUNES; CABRAL, 2015).

- Double input double output (DIDO) behavioural models that rely on measuring the transmitted and reflected signals at both PA ports, such as in (CAI; GONÇALVES; PEDRO, 2017), (ZARGAR; BANAI; PEDRO, 2015b), (GIBIINO et al., 2018). When the PA is not perfectly matched at the output port due to small deviations in the load impedance, there are reflections at the baseband, fundamental and harmonics back into the PA, as shown in Figure 2.8. A general DIDO model is proposed in (ZARGAR; BANAI; PEDRO, 2015b):

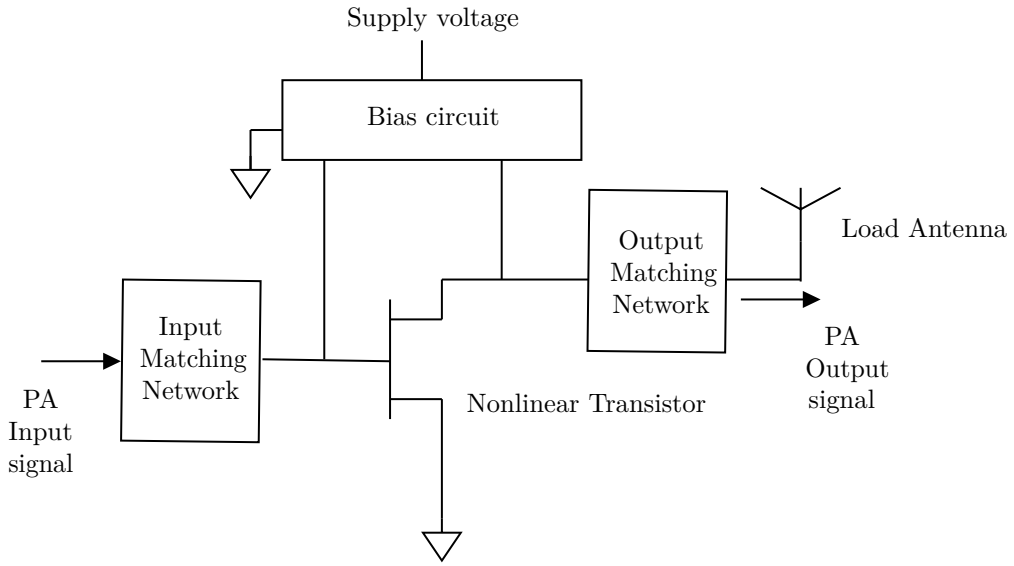



Figure 2.7 – Simplified PA block-diagram.

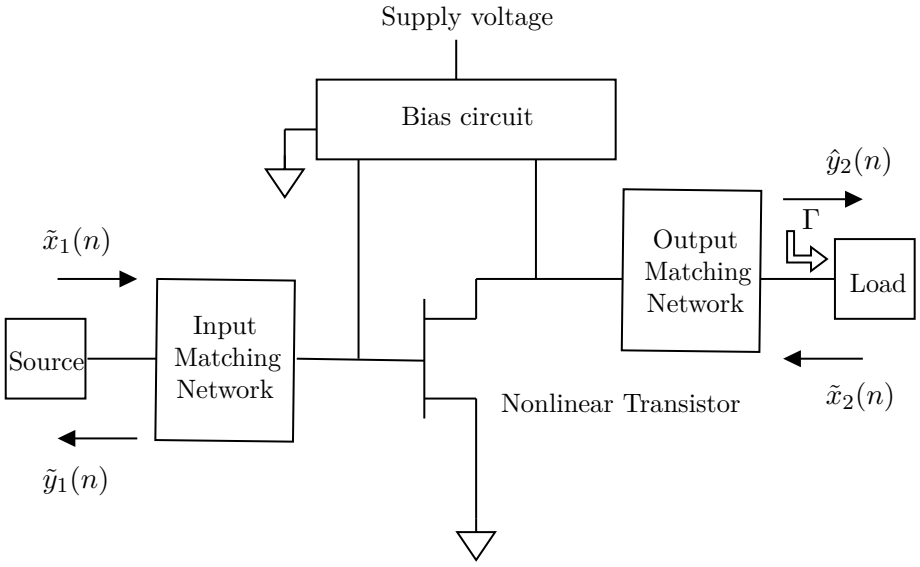



Figure 2.8 – Simplified PA block-diagram with DIDO signals.

$$\hat{y}_2(n) = \sum_{\substack{k=1 \\ k\ odd}}^{K} \sum_{m=0}^{M} \sum_{p=0}^{\frac{k+1}{2}} \sum_{q=0}^{\frac{k-1}{2}} \alpha_{kpq,m} \tilde{x}_1^{\frac{k+1}{2}-p}(n-m)\, \tilde{x}_1^{*\frac{k-1}{2}-q}(n-m) \tag{2.44}$$

$$\tilde{x}_2^{p}(n-m)\, \tilde{x}_2^{*q}(n-m),$$

where $K$ is odd, $\tilde{x}_1(n)$ is the input signal, $\tilde{x}_2(n)$ is the reflected signal at the output port, $\tilde{y}_1(n)$ is the reflected signal at the input port and $\tilde{y}_2(n)$ is the output signal. This formulation considers $\frac{k+1}{2}$ products of the complex envelope signals times $\frac{k-1}{2}$ products of the complex conjugate of the same signals. In the DIDO approach, the reflected signal $\tilde{x}_2(n)$ is considered as important as the input signal, $\tilde{x}_1(n)$, increasing the model capabilities, but at the expense of a high complexity that increases fast with the NL order and memory depth.

- PHD-based behavioural models, such as in (FAGER et al., 2014), (CAI; GONÇALVES; PEDRO, 2017). The PHD assumes that the reflections are small enough, such that their contributions appear only as linear perturbations. The dual-PHD model proposed in (FAGER et al., 2014), considering only odd-order terms, is given by:

$$\begin{aligned}\hat{y}_2(n) = &\sum_{m=0}^{M_1}\sum_{\substack{k=1\\ k\ odd}}^{K_1}\alpha_{m,k}^{(1)}\tilde{x}_1(n-m)\left|\tilde{x}_1(n-m)\right|^{k-1}\\ &+\sum_{m_2=0}^{M_2}\sum_{m_1=0}^{M_1}\sum_{\substack{k=1\\ k\ odd}}^{K_2}\alpha_{m_1,m_2,k}^{(2)}\tilde{x}_2(n-m_2)\left|\tilde{x}_1(n-m_1)\right|^{k-1}\\ &+\sum_{m_2=0}^{M_2}\sum_{m_1=0}^{M_1}\sum_{\substack{k=1\\ k\ odd}}^{K_2-2}\alpha_{m_1,m_2,k}^{(3)}\tilde{x}_1^2(n-m_1)\,\tilde{x}_2^*(n-m_2)\left|\tilde{x}_1(n-m_1)\right|^{k-1}.\end{aligned} \tag{2.45}$$

Note that only the first order of $\tilde{x}_2(n)$ is included. To overcome this limitation, a so-called dual-PHD model is also proposed in (CAI; GONÇALVES; PEDRO, 2017):

$$\begin{aligned}\hat{y}_2(n) = &\sum_{m=0}^{M_1}\sum_{\substack{k=1\\ k\ odd}}^{K_1}\alpha_{m,k}^{(1)}\tilde{x}_1(n-m)\left|\tilde{x}_1(n-m)\right|^{k-1}\\ &+\sum_{m_2=0}^{M_2}\sum_{m_1=0}^{M_1}\sum_{\substack{k=1\\ k\ odd}}^{K_2}\alpha_{m_1,m_2,k}^{(2)}\tilde{x}_2(n-m_2)\left|\tilde{x}_1(n-m_1)\right|^{k-1}\\ &+\sum_{m_2=0}^{M_2}\sum_{m_1=0}^{M_1}\sum_{\substack{k=1\\ k\ odd}}^{K_2-2}\alpha_{m_1,m_2,k}^{(3)}\tilde{x}_1^2(n-m_1)\,\tilde{x}_2^*(n-m_2)\left|\tilde{x}_1(n-m_1)\right|^{k-1}\\ &+\sum_{m_2=0}^{M_2}\sum_{m_1=0}^{M_1}\sum_{\substack{k=1\\ k\ odd}}^{K_2-2}\alpha_{m_1,m_2,k}^{(4)}\tilde{x}_1^*(n-m_1)\,\tilde{x}_2^2(n-m_2)\left|\tilde{x}_1(n-m_1)\right|^{k-1}\\ &+\sum_{m_2=0}^{M_2}\sum_{m_1=0}^{M_1}\sum_{\substack{k=1\\ k\ odd}}^{K_2-2}\alpha_{m_1,m_2,k}^{(5)}\tilde{x}_1(n-m_1)\left|\tilde{x}_2(n-m_2)\right|^2\left|\tilde{x}_1(n-m_1)\right|^{k-1}\\ &+\sum_{m_2=0}^{M_2}\sum_{m_1=0}^{M_1}\sum_{\substack{k=1\\ k\ odd}}^{K_2-4}\alpha_{m_1,m_2,k}^{(6)}\tilde{x}_2^{*^2}(n-m_2)\,\tilde{x}_1^3(n-m_1)\left|\tilde{x}_1(n-m_1)\right|^{k-1}.\end{aligned} \tag{2.46}$$

Comparing to the DIDO model in eq. (2.44), the dual-PHD model suppresses the terms of $\tilde{x}_2(n)$ whose NL orders are higher than two, thus reducing the model complexity.

- Load-dependent behavioural models, such as in (CAI et al., 2015), (DHAR et al., 2018), (DUNN et al., 2017), in which the model coefficients are parametrized by the complex-valued load reflection coefficient, $\Gamma$. In such models, the model coefficients are previously estimated for a batch of $\Gamma$ values, then are interpolated during the DPD training phase, based on the reflection coefficient that is actually measured at

the PA/ antenna interface. In (DHAR et al., 2018), the model is given by:

$$\begin{aligned}\hat{y}_2(n,\Gamma) = &\sum_{m_2=0}^{M_2}\sum_{m_1=0}^{M_1}\sum_{\substack{k=1\\ k\ odd}}^{K_1}\alpha^{(1)}_{m_1,m_2,k}(\Gamma)\tilde{x}_1(n-m_1)\left|\tilde{x}_1(n-m_1)\right|^{\frac{k-1}{2}}\left|\tilde{x}_2(n-m_2)\right|^{\frac{k-1}{2}}\\ &+\sum_{m_2=0}^{M_2}\sum_{m_1=0}^{M_1}\sum_{\substack{k=1\\ k\ odd}}^{K_2}\alpha^{(2)}_{m_1,m_2,k}(\Gamma)\tilde{x}_2(n-m_2)\left|\tilde{x}_1(n-m_1)\right|^{\frac{k-1}{2}}\\ &\left|\tilde{x}_2(n-m_2)\right|^{\frac{k-1}{2}},\end{aligned} \tag{2.47}$$

where $\tilde{x}_1(n)$ and $\tilde{x}_2(n)$ are, respectively, the PA input signal and the reflected signal at the PA output. $\alpha^{(1)}_{m_1,m_2,k}(\Gamma)$ and $\alpha^{(2)}_{m_1,m_2,k}(\Gamma)$ are the interpolated $\Gamma$-dependent coefficients.

Alternatively, (DUNN et al., 2017) proposed a load-dependent model that includes a polynomial expression of $\Gamma$ in the model, as follows:

$$\hat{y}_2(n,\Gamma) = \sum_{m=0}^{M}\sum_{\substack{k=1\\ k\ odd}}^{K}\sum_{q=0}^{Q}\alpha_{m,k,q}\tilde{x}_1(n-m)\left|\tilde{x}_1(n-m)\right|^{k-1}\Gamma\left|\Gamma\right|^{q}, \tag{2.48}$$

where $\alpha_{m,k,q}$ are the model coefficients, $\Gamma$ is the measured reflection coefficient and $Q+1$ is the order of the load impedance mismatch dependency.

## 2.4 IQM IMBALANCE MODELS

In the frequency-independent IQM imbalance case, the LO is the main responsible for the IQM imbalance. The LO imbalance model takes the I branch LO signal as the reference and models the Q branch LO signal with deviations in amplitude $\mu$ and phase $\rho$ (in radians) from its nominal values. The IQM RF bandpass output signal is given by:

$$\begin{aligned}\hat{x}_{RF}(t) &= \tilde{s}_I(t)\cos(2\pi f_0 t) - \mu\tilde{s}_Q(t)\sin(2\pi f_0 t-\rho)\\ &= [\tilde{s}_I(t)+\mu\sin(\rho)\tilde{s}_Q(t)]\cos(2\pi f_0 t) - \mu\cos(\rho)\tilde{s}_Q(t)\sin(2\pi f_0 t),\end{aligned} \tag{2.49}$$

where $\hat{x}_{RF}(t)$ is the bandpass IQM output signal at RF center frequency $f_0$, $\tilde{s}_I(t)$ and $\tilde{s}_Q(t)$ are, respectively, the real and imaginary parts of the IQM complex-valued baseband input signal, $\mu$ and $\rho$ are, respectively, the LO amplitude and phase deviations. In complex baseband, the equivalent IQM output signal, $\hat{x}(t)$, has real and imaginary parts given by:

$$\hat{x}_I(t) = \tilde{s}_I(t) + \mu\sin(\rho)\tilde{s}_Q(t), \tag{2.50}$$

and

$$\hat{x}_Q(t) = -\mu \cos(\rho)\tilde{s}_Q(t). \tag{2.51}$$

Alternatively, the IQM baseband output signal can be written in terms of the input complex baseband signal and its complex conjugate, as:

$$\hat{x}(t) = \frac{1+\mu e^{j\rho}}{2}\tilde{s}(t) + \frac{1-\mu e^{j\rho}}{2}\tilde{s}^*(t), \tag{2.52}$$

where $\tilde{s}(t)$ and $\hat{x}(t)$ are, respectively, the IQM complex baseband input and modelled output signals and $(\cdot)^*$ is the complex conjugate operator. Since the spectrum of $\tilde{s}^*(t)$ is a conjugated and mirrored-image version of the spectrum of $\tilde{s}(t)$, eq. (2.52) explicitly shows that IQM imbalance adds mirror-image interference to the original complex-valued signal (VALKAMA, 2010), (ANTTILA; VALKAMA; RENFORS, 2008b).

In wideband applications, the IQM imbalance model should also include frequency-dependent effects, i.e., the imbalance between the frequency responses of LPFs and mixers on the I and Q branches. The IQM RF bandpass output signal is then given by:

$$\begin{aligned}\hat{x}_{RF}(t) &= \tilde{s}_I(t)\cos(2\pi f_0 t) - \mu\tilde{s}_Q(t) * \tilde{h}(t)\sin(2\pi f_0 t - \rho) \\ &= [\tilde{s}_I(t) + \mu\sin(\rho)\tilde{s}_Q(t) * \tilde{h}(t)]\cos(2\pi f_0 t) - \mu\cos(\rho)\tilde{s}_Q(t) * \tilde{h}(t)\sin(2\pi f_0 t),\end{aligned} \tag{2.53}$$

where $\hat{x}_{RF}(t)$ is the bandpass IQM output signal at RF frequency $f_0$, $\tilde{s}_I(t)$ and $\tilde{s}_Q(t)$ are, respectively, the real and imaginary parts of the IQM complex-valued baseband input signal, $*$ is the convolution operator and $\tilde{h}(t)$ is the impulse response of the baseband filter that models the frequency-dependent IQM imbalance between I and Q paths. In this case, the equivalent complex baseband IQM output signal, $\hat{x}(t)$, can be written in terms of the baseband input signal, $\tilde{h}(t)$, and its complex conjugate, as:

$$\hat{x}(t) = \frac{\delta(t) + \tilde{h}(t)\mu e^{j\rho}}{2} * \tilde{s}(t) + \frac{\delta(t) - \tilde{h}(t)\mu e^{j\rho}}{2} * \tilde{s}^*(t), \tag{2.54}$$

where $\tilde{s}(t)$ and $\hat{x}(t)$ are, respectively, the IQM complex baseband input and modelled output signals, $\delta(t)$ is the Dirac delta function, $*$ is the convolution operator and $(\cdot)^*$ is the complex conjugate operator. Figure 2.9 represents the frequency-dependent IQM imbalance in eq. (2.54) with $\tilde{h}_1(t) = \frac{\delta(t)+\tilde{h}(t)\mu e^{j\rho}}{2}$ and $\tilde{h}_2(t) = \frac{\delta(t)-\tilde{h}(t)\mu e^{j\rho}}{2}$.

Equation (2.54) captures the widely-linear mechanism of IQM imbalance and can also be equivalently represented by Figure 2.10 (ANTTILA; HANDEL; VALKAMA, 2010). This symmetrical discrete-time baseband IQM imbalance model has four finite impulse response (FIR) filters to model the IQ impairments in the general frequency-dependent case (SAFFAR et al., 2013), (DING et al., 2003), (DING et al., 2008).

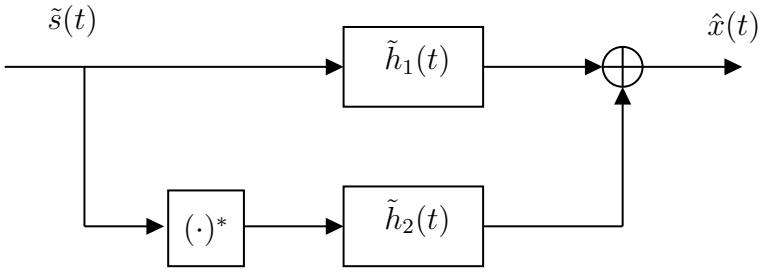



Figure 2.9 – IQM imbalance model with two complex filters.

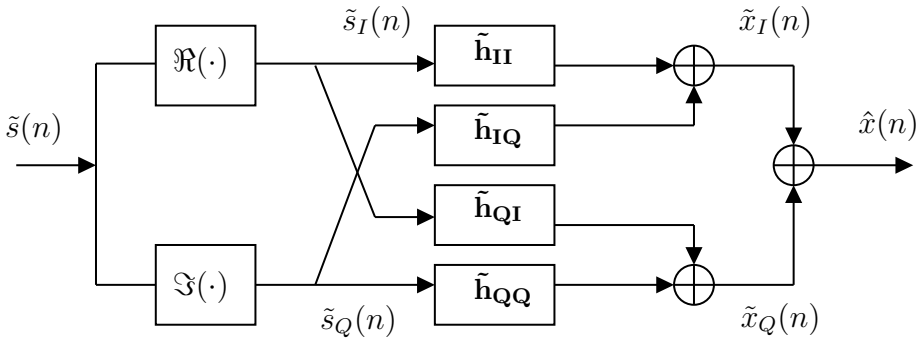



Figure 2.10 – IQM imbalance model with four real filters.

In Figure 2.10, the FIR filters $\tilde{\boldsymbol{h}}_{II}$, $\tilde{\boldsymbol{h}}_{IQ}$, $\tilde{\boldsymbol{h}}_{QI}$ and $\tilde{\boldsymbol{h}}_{QQ}$ are linear time-invariant (LTI) with real-valued coefficients and memory length $L$, scaling and delaying the I and Q components of the complex baseband input signal. In order to compute the modelled output signal $\hat{x}(n)$ at sample $n$, we collect the vectors of delayed input samples, $\tilde{s}_I(n)$ and $\tilde{s}_Q(n)$ that are given by:

$$\tilde{\boldsymbol{s}}_I(n) = \begin{bmatrix} \Re\{\tilde{s}(n)\} \\ \vdots \\ \Re\{\tilde{s}(n-L)\} \end{bmatrix}^T \tag{2.55}$$

and

$$\tilde{\boldsymbol{s}}_Q(n) = \begin{bmatrix} \Im\{\tilde{s}(n)\} \\ \vdots \\ \Im\{\tilde{s}(n-L)\} \end{bmatrix}^T, \tag{2.56}$$

where $(\cdot)^T$ is the transpose operator and $L$ is the length of the FIR filters $\tilde{\boldsymbol{h}}_{II}$, $\tilde{\boldsymbol{h}}_{IQ}$, $\tilde{\boldsymbol{h}}_{QI}$ and $\tilde{\boldsymbol{h}}_{QQ}$. The modelled output complex baseband signal at time instant $n$ can be written as:

$$\hat{x}(n) = \tilde{\boldsymbol{s}}_I(n)\tilde{\boldsymbol{h}}_{II} + \tilde{\boldsymbol{s}}_Q(n)\tilde{\boldsymbol{h}}_{IQ} + j\{\tilde{\boldsymbol{s}}_Q(n)\tilde{\boldsymbol{h}}_{QQ} + \tilde{\boldsymbol{s}}_I(n)\tilde{\boldsymbol{h}}_{QI}\}, \tag{2.57}$$

where $\tilde{\boldsymbol{h}}_{II}$, $\tilde{\boldsymbol{h}}_{IQ}$, $\tilde{\boldsymbol{h}}_{QI}$ and $\tilde{\boldsymbol{h}}_{QQ}$ are column vectors, $\hat{x}(n)$ is the modelled IQM output signal at time instant $n$, $\tilde{\boldsymbol{s}}_I(n)$ and $\tilde{\boldsymbol{s}}_Q(n)$ are, respectively, the real and imaginary parts of the input vector with length $L$. Equivalently, by representing the four FIR filters by two

complex filters:

$$\tilde{\boldsymbol{h}}^I = \tilde{\boldsymbol{h}}_{II} + j\tilde{\boldsymbol{h}}_{QI} \tag{2.58}$$

and

$$\tilde{\boldsymbol{h}}^Q = \tilde{\boldsymbol{h}}_{IQ} + j\tilde{\boldsymbol{h}}_{QQ}. \tag{2.59}$$

Finally, eq. (2.57) can also be expressed as:

$$\hat{x}(n) = \tilde{\boldsymbol{s}}_I(n)\tilde{\boldsymbol{h}}^I + \tilde{\boldsymbol{s}}_Q(n)\tilde{\boldsymbol{h}}^Q. \tag{2.60}$$

In the frequency-independent case, the $\tilde{\boldsymbol{h}}^I$ and $\tilde{\boldsymbol{h}}^Q$ filters are reduced to complex scalars, corresponding to amplitude and phase imbalances between the LO signals in the I and Q branches.

## 2.5 DPD LEARNING ARCHITECTURES

Having introduced a number of PA models, the next objective is to apply these models to DPD. As shown in (SCHETZEN, 2006), inverse models can also be implemented using the same structures presented in the previous sections. Therefore, DPD models in this thesis are also based on the Volterra series models. DPD is one of the most cost-effective techniques for improving the performance of NL PAs, by both enhancing its linearity and enabling its operation at higher power levels. Significant improvements have been achieved, (KIM; KONSTANTINOU, 2001), (MORGAN et al., 2006), (DING et al., 2004), (LEE; FIGUEIREDO, 2006), including for mobile device PAs (PRESTI; KIMBALL; ASBECK, 2012), (NORRIS et al., 2007). Moreover, DPD is entirely implemented in the digital baseband domain, being more flexible and robust than hardware compensation counterparts. The DPD predistorter is the functional block that precedes the PA in the forward path and, typically, produces an expanding NL output, assuming that the PA has an amplitude compressing behaviour. During the DPD training phase, in which the DPD model is estimated, the predistorter is bypassed and the PA input and output signals are measured, down-converted and converted to the discrete domain, if required. Two different schemes exist for estimating the DPD inverse model coefficients, namely, the direct and the indirect learning architectures, as shown in Figures 2.11 and 2.12, respectively. The DPD model is estimated using $N$ samples of the discrete-time baseband signals $\tilde{\boldsymbol{x}} = \{\tilde{x}(n)\}_{n=0}^{N-1}$ and $\tilde{\boldsymbol{u}} = \{\tilde{u}(n)\}_{n=0}^{N-1}$, respectively the PA input and normalized output signals, i.e., $\tilde{\boldsymbol{u}} = \tilde{\boldsymbol{y}}/\sqrt{G}$ and $G$ is the PA intended linear power gain.

In the direct learning architecture (DLA), firstly the coefficients of the PA NL model are estimated from measured PA input/output data; next, an approximate inverse model is derived for the DPD, using the $p^{th}$-order inverse technique (SCHETZEN, 2006), as shown in the DLA block diagram in Figure 2.11. During the training phase, the coefficients of the PA model are estimated using the (undistorted) input signal $\tilde{x}(n)$ as the model

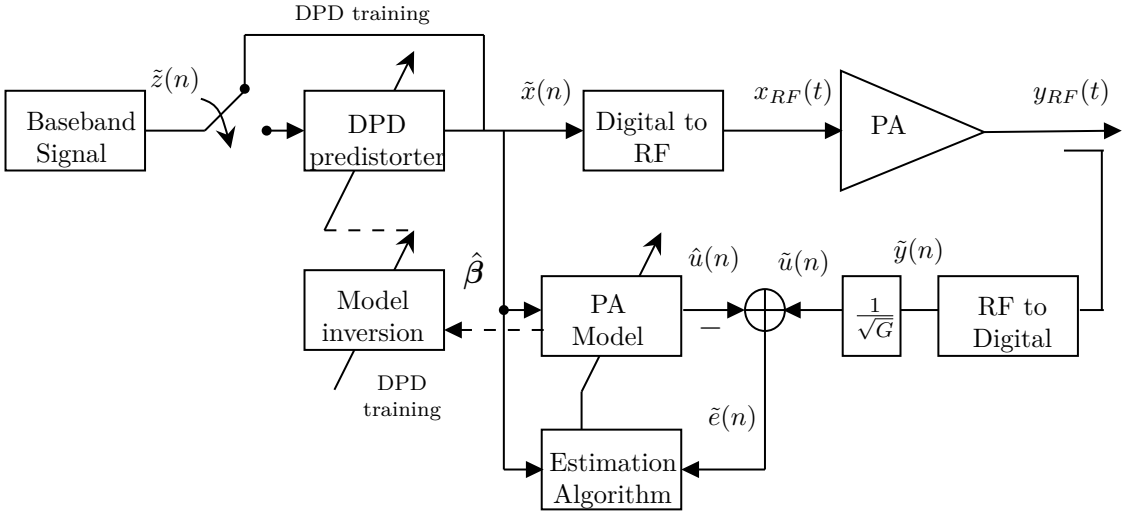



Figure 2.11 – DLA DPD block-diagram.

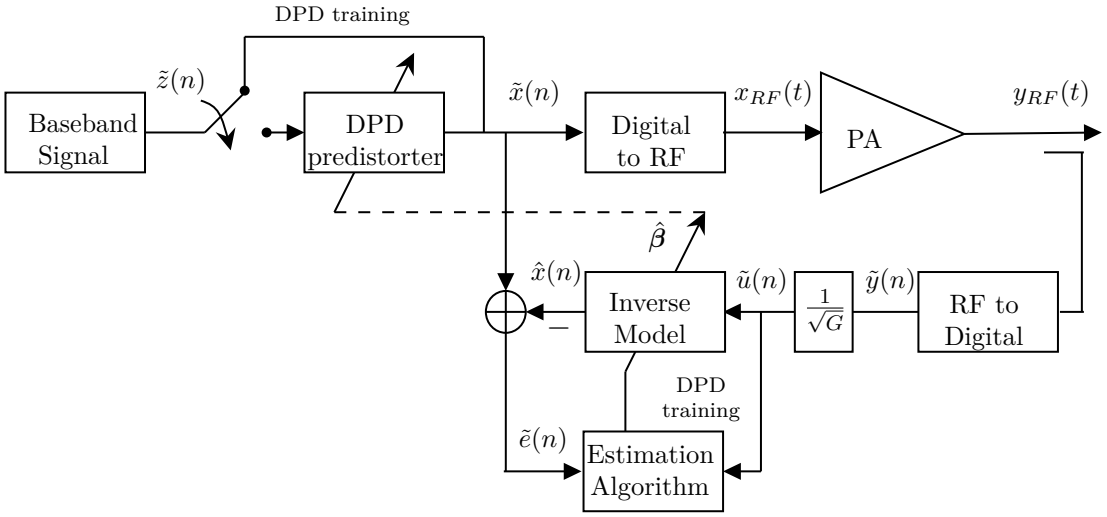



Figure 2.12 – ILA DPD block-diagram.

input and minimizing the error between the modelled output $\hat{u}(n)$ and the PA normalized output signal $\tilde{u}(n)$, in the least squares sense. Next, the $p^{th}$-order inverse technique is used to invert the estimated model. Since this step is not always accurate and is highly demanding in terms of computational cost, the DLA becomes unattractive for wideband applications.

In turn, in the indirect learning architecture (ILA) (EUN; POWERS, 1997), a postdistorter NL model, placed after the PA, is estimated directly, bypassing the predistorter, by inverting the roles of the PA input and output signals. Then, the estimated postdistorter inverse model is copied to the DPD predistorter block (SCHETZEN, 2006). Figure 2.12 shows a simplified block diagram of the ILA DPD scheme, including the signals considered for the DPD formulation. During the training phase, the NL PA is fed with the original (undistorted) input signal, i.e., $\tilde{x}(n) = \tilde{z}(n)$. The DPD training phase is carried out by

the blocks in the feedback path, responsible for estimating/updating the inverse model coefficients, using the PA normalized output $\tilde{u}(n)$ as the post-inverse input and minimizing the error $\tilde{e}(n)$ between the modelled post-inverse output $\hat{x}(n)$ and the (undistorted) PA input signal $\tilde{x}(n)$. The inverse model is typically linear with respect to the coefficients, thus linear least squares is employed, either using a batch of input/output samples, as detailed in the next section, or adaptively. Once the post-inverse coefficients have been estimated and the training phase is complete, the estimated coefficients are copied to the predistorter block in the forward path and the DPD system is run in open loop.

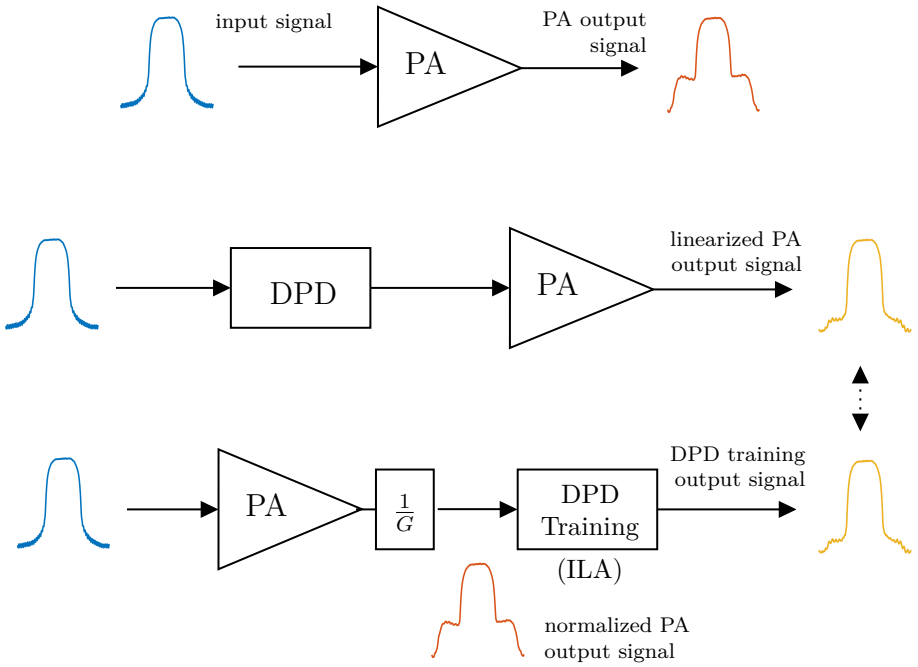



Figure 2.13 – Equivalence of the DPD training and linearized PA output signals.

Comparing to the DLA, the main advantage of the ILA is to avoid the explicit inversion of the PA model (DING et al., 2002), i.e., the DPD coefficients are extracted directly. Therefore, ILA DPD is more popular due to its estimation simplicity and lower computational complexity. Nevertheless, the ILA approach assumes that the estimated post-inverse and the pre-inverse models (EUN; POWERS, 1997) are equivalent, which allows the DPD (predistorter) to be a copy of the estimated DPD training postdistorter. In this thesis, the identification of the DPD coefficients is performed through the ILA approach, assuming a perfect feedback path (DING et al., 2002). In this case, as shown in Figure 2.13, the DPD training output signal, in the feedback path, and the linearized PA output signal (normalized) are equal, both correspond to the PA output signal compensated for PA non-linearities and memory effects.

## 2.6 LEAST SQUARES ESTIMATION

The terms estimation and identification are used in general for the task of obtaining the kernels or coefficients of a model from measured input/output data. The behavioural modelling of a dynamic NL system consists, first, in selecting an appropriate model structure to describe the system, then, setting the model parameters, e.g., memory depth and NL order, and, finally, applying an identification method to estimate the model coefficients, by processing noisy input/output data from measurements. Volterra-based models are widely used in various fields of applications for representing dynamic NL systems and their coefficients can be identified using various methods. This section presents the OLS estimation technique. In chapter 4, non-linear extensions of the OLS estimation are required for the estimation of Volterra-PARAFAC and Laguerre-Volterra models. In chapter 6, CS-based least squares techniques are employed in the context of sparse models.

Recalling from the previous section, under the ILA framework shown in Figure 2.12, the DPD training block estimates the DPD model coefficients directly by interchanging the roles of the measured, length $N$, PA input $\tilde{\boldsymbol{x}} = \{\tilde{x}(n)\}_{n=0}^{N-1}$ and normalized output $\tilde{\boldsymbol{u}} = \{\tilde{u}(n)\}_{n=0}^{N-1}$ signals (NGOYA; MONS, 2014).

Let us consider the baseband $FV_{\triangle}$ model in eq. (2.4) as the DPD model, with parameters $M$ and $K$. Given measured input/output training data $\{\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{x}}\}_{n=0}^{N-1}$, the $N \times R$ regression matrix of the model is composed of the regressors $\boldsymbol{\phi}_r$, with $r = 1, \ldots, R$, i.e., polynomial functions of the input signal, as follows:

$$\boldsymbol{\Phi}(\tilde{\boldsymbol{u}}) = \{\boldsymbol{\phi}_r(\tilde{\boldsymbol{u}})\}_{r=1}^{R} \tag{2.61}$$

where $\boldsymbol{\phi}_r \in \mathbb{C}^{(N\times 1)}$ are polynomial regressors and $R$, the number of regressors in the model. In practice, the triangular Volterra regression matrix $\boldsymbol{\Phi}$ can be computed using the truncated Kronecker operator (FERNANDES; MOTA; FAVIER, 2010) and (FAVIER; KIBANGOU; CAMPELLO, 2003), as follows. Firstly, the regression matrix is expressed as the concatenation of the sub-matrices $\boldsymbol{\Phi}^{(k)}$, for $k = 1, \ldots, K$ ($k$ odd):

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\Phi}^{(1)} & \boldsymbol{\Phi}^{(3)} & \ldots & \boldsymbol{\Phi}^{(k)} & \ldots & \boldsymbol{\Phi}^{(K)} \end{bmatrix}, \tag{2.62}$$

where the $k^{th}$-order sub-matrix $\boldsymbol{\Phi}^{(k)}$ is expressed as:

$$\boldsymbol{\Phi}^{(k)} = \begin{bmatrix} \boldsymbol{\phi}^{(k)}(0) \\ \vdots \\ \boldsymbol{\phi}^{(k)}(N-1) \end{bmatrix}. \tag{2.63}$$

The row vectors $\boldsymbol{\phi}^{(k)}(n)$ in eq. (2.63), for $n = 0, \ldots, N-1$, i.e., the rows of the sub-matrix $\boldsymbol{\Phi}^{(k)}$, can be computed as:

$$\boldsymbol{\phi}^{(k)}(n) \ = (\oslash^{\frac{k+1}{2}} \tilde{\boldsymbol{u}}(n)) \otimes (\oslash^{\frac{k-1}{2}} \tilde{\boldsymbol{u}}^*(n)), \tag{2.64}$$

where $\otimes$ is the Kronecker product and $\oslash^q$ is the $q^{th}$-order truncated Kronecker product, such that $\oslash^q \tilde{\boldsymbol{u}}(n) \equiv \underbrace{\tilde{\boldsymbol{u}}(n) \oslash \ldots \oslash \tilde{\boldsymbol{u}}(n)}_{\mathrm{q}}$, that is obtained recursively, as follows:

$$\oslash^q \tilde{\boldsymbol{u}}(n) \ = \begin{bmatrix} \tilde{u}(n) \oslash^{q-1} \tilde{\boldsymbol{u}}(n) \\ \tilde{u}(n-1) \oslash^{q-1} \tilde{\boldsymbol{u}}(n-1) \\ \vdots \\ \tilde{u}(n-M) \oslash^{q-1} \tilde{\boldsymbol{u}}(n-M) \end{bmatrix} \tag{2.65}$$

with $\oslash^1 \tilde{\boldsymbol{u}}(n) = \tilde{\boldsymbol{u}}(n)$ and:

$$\tilde{\boldsymbol{u}}(n-i) = \begin{bmatrix} \tilde{u}(n-i) & \tilde{u}(n-i-1) & \ldots & \tilde{u}(n-M) \end{bmatrix}. \tag{2.66}$$

The Kronecker products are also described in the Appendix A. Note that in eq.(2.64) any redundant cross-terms are suppressed.

The triangular Volterra models are linear with respect to the coefficients; thus, the unknown model coefficients can be found by solving a system of linear equations:

$$\tilde{\boldsymbol{x}} = \boldsymbol{\Phi}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{2.67}$$

where the $R \times 1$ coefficients vector $\boldsymbol{\beta}$ corresponds to the Volterra kernels $\tilde{h}_{k,\triangle}(m_{1:k})$, stacked for $k = 1, 3, \ldots, K$, and $\tilde{\boldsymbol{x}}$ is the $N \times 1$ vector of measured DPD reference outputs. As discussed above, the regression matrix $\boldsymbol{\Phi}$ is composed of the sub-matrices $\boldsymbol{\Phi}^{(k)}$, with $k = 1, 3, \ldots, K$, whose columns are the $k^{th}$-order model-specific regressors (also known as features or predictors) computed using $\tilde{\boldsymbol{u}}$, the input signal (explanatory or independent variable). The vector $\boldsymbol{\epsilon}$ accounts for un-modelled contributions and additive white Gaussian (AWG) measurement noise.

Linear regression techniques allow to express the output of the model $\hat{\boldsymbol{x}}$ (also known as response, or dependent variable), as a linear combination of polynomial functions of the input signal. The linear combination is learnt from the training data, parametrized by a vector of coefficients $\boldsymbol{\beta} \in \mathbb{C}^{R\times 1}$ (sometimes, also an intercept $\beta_0$) and is used for accurately predicting future outputs.

In order to find the coefficients vector $\boldsymbol{\beta}$, the OLS batch estimator can be formulated as the following optimization problem:

$$\hat{\boldsymbol{\beta}}_{OLS} = \underset{\boldsymbol{\beta}\in\mathbb{C}^{R\times 1}}{\arg\min} \ \ \frac{1}{2N} RSS(\boldsymbol{\beta}) = \underset{\boldsymbol{\beta}\in\mathbb{C}^{R\times 1}}{\arg\min} \ \ \frac{1}{2N} \|\tilde{\boldsymbol{x}} - \boldsymbol{\Phi}\boldsymbol{\beta}\|_2^2, \tag{2.68}$$

where the cost function RSS is the residual sum of squares, given by $RSS(\boldsymbol{\beta}) = \|\tilde{\boldsymbol{x}} - \boldsymbol{\Phi}\boldsymbol{\beta}\|_2^2$, whose squared Euclidean norm is minimized between the noisy $\tilde{\boldsymbol{x}}$ and $\boldsymbol{\Phi}\boldsymbol{\beta}$. Differentiating $(\tilde{\boldsymbol{x}} - \boldsymbol{\Phi}\boldsymbol{\beta})^H(\tilde{\boldsymbol{x}} - \boldsymbol{\Phi}\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ and equating to zero, leads to the well-known normal equations:

$$\boldsymbol{\Phi}^H\boldsymbol{\Phi}\boldsymbol{\beta} = \boldsymbol{\Phi}^H\tilde{\boldsymbol{x}}. \tag{2.69}$$

Note that, under the assumption of AWG output error, it can be proved that the OLS estimate is equal to the maximum likelihood estimate and that the OLS estimate is unbiased.

Considering the existence and uniqueness of solutions for $\boldsymbol{\Phi}\boldsymbol{\beta} = \tilde{\boldsymbol{x}}$:

- A unique solution exists if $rank(\boldsymbol{\Phi}) = rank(\boldsymbol{\Phi}|\tilde{\boldsymbol{x}}) = R$, where the augmented matrix $(\boldsymbol{\Phi}|\tilde{\boldsymbol{x}})$ is obtained by appending the column vector $\tilde{\boldsymbol{x}}$ to the matrix $\boldsymbol{\Phi}$. This means that both $\boldsymbol{\Phi}$ is full column rank and $\tilde{\boldsymbol{x}} \in colspace(\boldsymbol{\Phi})$ ($\tilde{\boldsymbol{x}}$ is in the span of the columns of $\boldsymbol{\Phi}$). In this case, the least squares solution for each $\tilde{\boldsymbol{x}}$ is exact and given by $\hat{\boldsymbol{\beta}}_{OLS} = (\boldsymbol{\Phi}^H\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^H\tilde{\boldsymbol{x}}$, where $\boldsymbol{\Phi}^H\boldsymbol{\Phi}$ is non-singular. The Moore-Penrose pseudo-inverse of $\boldsymbol{\Phi}$ is defined as $\boldsymbol{\Phi}^\dagger = (\boldsymbol{\Phi}^H\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^H$ and, when $\boldsymbol{\Phi}$ is square ($N = R$) and invertible ($det(\boldsymbol{\Phi}) \neq 0$), $\boldsymbol{\Phi}^\dagger = \boldsymbol{\Phi}^{-1}$ ($\boldsymbol{\Phi}^{-1}$ is the inverse matrix of $\boldsymbol{\Phi}$). Note that in practice $\boldsymbol{\Phi}^\dagger$ is computed using either the QR factorization or singular value decomposition (SVD).

- Infinitely many solutions exist if $rank(\boldsymbol{\Phi}) = rank(\boldsymbol{\Phi}|\tilde{\boldsymbol{x}}) < R$, i.e., the number of coefficients is larger than the rank of the matrix. In this case, $\boldsymbol{\Phi}$ is rank-deficient and, if $\boldsymbol{\Phi}$ is a square matrix, it is singular. Most likely, $\boldsymbol{\Phi}$ is a *short* and *wide* matrix ($N < R$), the system is under-determined (or over-complete), $\boldsymbol{\Phi}^H\boldsymbol{\Phi}$ is singular and the problem is ill-posed. This means that the $N \times 1$ measurement vector $\tilde{\boldsymbol{x}}$ is unable to uniquely determine the higher-dimensional $R \times 1$ vector $\boldsymbol{\beta}$ and the under-determined system is likely to have infinitely many solutions $\hat{\boldsymbol{\beta}}$ for each $\tilde{\boldsymbol{x}}$. Considering that $\tilde{\boldsymbol{x}} \in colspace(\boldsymbol{\Phi})$, among an infinite number of exact least squares solutions, often the minimum $\ell_2$-norm of $\beta$ solution is chosen. The minimum $\|\boldsymbol{\beta}\|_2^2$ least squares solution (exact) is given by: $\hat{\boldsymbol{\beta}}_{LN} = \boldsymbol{\Phi}^H(\boldsymbol{\Phi}\boldsymbol{\Phi}^H)^{-1}\tilde{\boldsymbol{x}}$ and is always unique. Note that, in this case, $\boldsymbol{\Phi}\boldsymbol{\Phi}^H$ is non-singular. Alternatively, one can search for sparse solutions for the system, in many cases the sparse solution is unique, as discussed latter.

- No exact solution exists if the system of equations is inconsistent , i.e., $rank(\boldsymbol{\Phi}) < rank(\boldsymbol{\Phi}|\tilde{\boldsymbol{x}})$, corresponding to $\tilde{\boldsymbol{x}} \notin colspace(\boldsymbol{\Phi})$. Most likely, $\boldsymbol{\Phi}$ is a *tall* and *thin* matrix ($N > R$), the system is over-determined and $\boldsymbol{\Phi}^H\boldsymbol{\Phi}$ is non-singular. This way, there are too many measurements in the $N \times 1$ vector $\tilde{\boldsymbol{x}}$ to determine a solution for the given

$R \times 1$ vector $\boldsymbol{\beta}$ and the over-determined system is likely to have no exact solution. An inconsistent system of equations also arises due to the noisy or stochastic nature of measured data $\tilde{\boldsymbol{x}}$. If the system is over-determined, the least squares (approximate) solution is given by the orthogonal projection of the higher dimensional $\tilde{\boldsymbol{x}}$ into the lower dimensional column space of $\boldsymbol{\Phi}$, i.e., $\boldsymbol{\Phi}\hat{\boldsymbol{\beta}} = proj_{colspace(\boldsymbol{\Phi})}\tilde{\boldsymbol{x}}$ and corresponds to $\hat{\boldsymbol{\beta}}_{OLS} = (\boldsymbol{\Phi}^H\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^H\tilde{\boldsymbol{x}}$ for each $\tilde{\boldsymbol{x}}$.

In practice, however, as the model dimensions increase, frequently the problems above become ill-posed, due to the high correlation among the columns of the rank-deficient regression matrix $\boldsymbol{\Phi}$ and given that $rank(\boldsymbol{\Phi}^H\boldsymbol{\Phi}) = rank(\boldsymbol{\Phi}\boldsymbol{\Phi}^H) = rank(\boldsymbol{\Phi})$. In this case, the accuracy and stability of the estimation is compromised, as indicated by the exponential growth of the condition number $cond(\boldsymbol{\Phi})$, i.e., the ratio of largest to smallest singular values of $\boldsymbol{\Phi}$ (RAICH; QIAN; ZHOU, 2004). The concept of regularization has been introduced by Tikhonov to stabilize ill-posed problems (FRANKLIN, 1974), leading to the ridge regression estimation, as follows:

$$\underset{\boldsymbol{\beta}\in\mathbb{C}^{R\times 1}}{\text{minimize}} \quad \frac{1}{2N}\|\tilde{\boldsymbol{x}} - \boldsymbol{\Phi}\boldsymbol{\beta}\|_2^2 \quad \text{subject to} \quad \|\boldsymbol{\beta}\|_2^2 \leq t, \tag{2.70}$$

where $\|\boldsymbol{\beta}\|_2^2 = \sum_{i=1}^{R}|\beta_i|^2$. Without loss of generality, assume that the matrix $\boldsymbol{\Phi}$ is standardized, such that each of its columns $\boldsymbol{\phi}_j$ is centered, i.e., $\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{\phi}_j)_i = 0$ and normalized, i.e., $\frac{1}{N}\boldsymbol{\phi}_j^H\boldsymbol{\phi}_j = 1$, for all $j = 1, \ldots, R$. Also $\tilde{\boldsymbol{x}}$ is centered, i.e., $\frac{1}{N}\sum_{i=1}^{N}\tilde{x}_i = 0$. In this case the cost function is the penalized residual sum of squares (PRSS), given by $PRSS(\boldsymbol{\beta}) = \|\boldsymbol{\Phi}\boldsymbol{\beta} - \tilde{\boldsymbol{x}}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_2^2$, leading to the following closed-form solution:

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{RR}(\lambda) &= (\boldsymbol{\Phi}^H\boldsymbol{\Phi} + \lambda\boldsymbol{I}_R)^{-1}\boldsymbol{\Phi}^H\tilde{\boldsymbol{x}} \\ &= \boldsymbol{\Phi}^H(\boldsymbol{\Phi}\boldsymbol{\Phi}^H + \lambda\boldsymbol{I}_N)^{-1}\tilde{\boldsymbol{x}},\end{aligned} \tag{2.71}$$

where $\lambda \geq 0$ is the tuning or regularization parameter that controls the shrinkage of coefficients. In this case, the biased estimator $\hat{\boldsymbol{\beta}}_{RR}$ allows the OLS variance to be reduced, often achieving better prediction accuracy. Also note that both regularized matrices $(\boldsymbol{\Phi}^H\boldsymbol{\Phi} + \lambda\boldsymbol{I}_R)$ and $(\boldsymbol{\Phi}\boldsymbol{\Phi}^H + \lambda\boldsymbol{I}_N)$ are non-singular. Note that the ridge regression does not shrink any of the coefficients to zero and, hence, does not generate sparse solutions.

## 2.7 CONCLUSIONS

This chapter presented several state-of-the-art Volterra-based models from the literature for modelling and compensating for RF analog impairments in wireless transmitters. Firstly, the FV, triangular FV and MV models were introduced, then reduced-complexity models obtained in the literature by imposing pruning strategies to the FV series were

also discussed. In the next chapters, these models are applied in the context of DPD as benchmarks for the ones to be derived in the thesis.

# 3 EXPERIMENTAL TEST SET-UP AND PRELIMINARY STUDY

In this chapter, the experimental test set-up used in this thesis for the experimental validation of the proposed DPD models and estimation techniques is presented. Moreover, the figures of merit considered for evaluating the performance achieved by each of the proposed techniques are defined. Finally, this chapter also presents a preliminary simulation study, in which RF impairments and their compensation are modelled and simulated, and the corresponding figures of merit are evaluated.

The identification and experimental validation of behavioural models for PAs and DPD depend on the availability of high-quality data. Captured datasets must provide a highly dynamic description of the PA signals over the whole operating bandwidth, in order to allow the DPD training phase to be completed satisfactorily.

As discussed in (FAVIER, 2010), Figure 3.1 illustrates the behavioural modelling methodology, consisting of an experimental test set-up that enables the extraction of input-output (I/O) data from the non-linear system, followed by the choice of the model structure, often based on try-and-error or *a priori* information about the system, then the sizing of model parameters and the estimation of the model coefficients. After choosing a set of candidate DPD models and sizing the corresponding model parameters, the DPD training block, responsible for identification, estimates the coefficients of each model from training data. Next, the best among the candidate models is chosen, based on information criteria, such as the Akaike's information criterion (AIC), or cross-validation, as discussed in subsection 6.1.4. The predictive performance of the model is accessed using new measured data, i.e., not used to build any of the models. Deciding which DPD behavioural model to use implies a compromise among different aspects, such as model accuracy, DPD running cost and estimation complexity. Note that this procedure is generally iterative, i.e., the sub-problems in the figure are iteratively addressed, with the need to revise some choices until the procedure is complete.

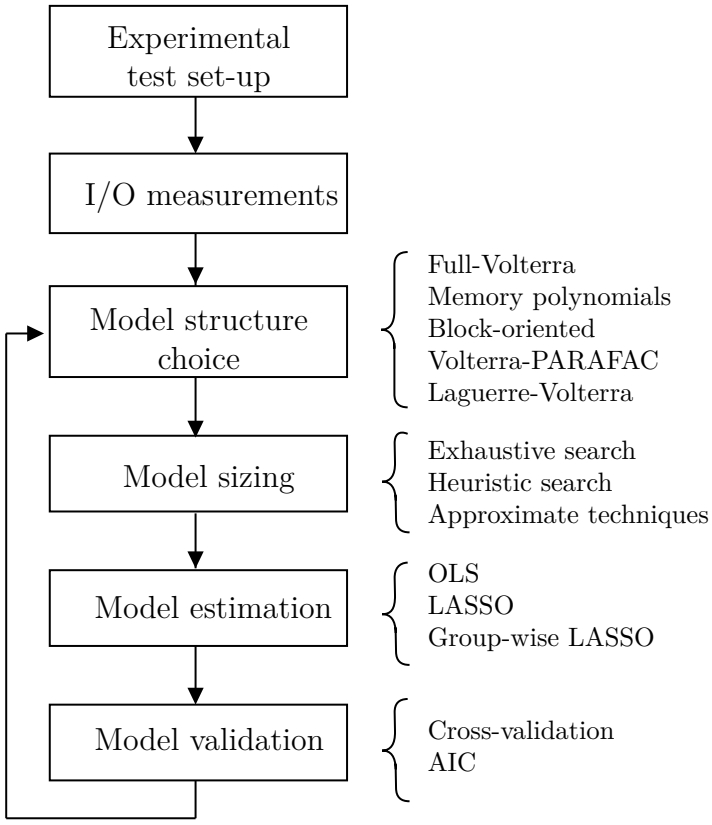



Figure 3.1 – Overview of behavioural system modelling.

## 3.1 EXPERIMENTAL TEST SET-UP

On the experimental phase, to validate the proposed DPD estimation techniques, a 10 W Gallium Arsenide (GaAs) metal–semiconductor field-effect transistor (MESFET) RF PA with 44 dB of nominal (linear) gain, designed at LME/EPUSP, is measured in Class AB configuration, with the test set-up shown in Figures 3.2 and 3.3. As detailed in the block diagram in Figure 3.3, an N5182B (MXG) vector signal generator from Keysight Technologies is used to synthesize, at the carrier frequency of 5.9 GHz, the PA input signal for this experiment, an orthogonal frequency division multiplexing (OFDM) waveform with 1,024 sub-carriers modulated with 64-quadrature amplitude modulation (64-QAM) and total bandwidth of 20 MHz. An 8045D load tuner from Maury Microwaves is used to vary the load impedance at the PA output, causing reflections back to the (isolator-free) PA output port. Load impedance values are manually adjusted at the 8045D to obtain VSWR of 4:1, a typical value considered in (BEZOOIJEN; MAHMOUDI; ROERMUND, 2011) and (ZARGAR; BANAI; PEDRO, 2015b). The complex impedance values are confirmed using a network analyzer. Table 3.1 summarizes the test set-up parameters. An N9020B (MXA) vector signal analyzer from Keysight Technologies is used to digitize and capture the complex I/Q data streams of the PA input and output waveforms. Both the MXG and MXA are connected to the PC via LAN. DPD training is processed off-line by MATLAB running in the PC.

Table 3.1 – Test set-up parameters.

| | | |
|---|---|---|
| PA | Transistor: | GaAs MESFET |
| | Biasing: | Class AB |
| | Max. power: | 10 W |
| | Lin. gain: | 44 dB |
| | Center frequency: | 5.9 GHz |
| | VSWR: | 1:1 (Matched)<br>4:1 (Mismatched) |
| Input signal | Waveform: | OFDM |
| | Sub-carriers: | 1024 |
| | Modulation format: | 64-QAM |
| | Bandwidth: | 20MHz |
| | Input power: | -8 dBm |

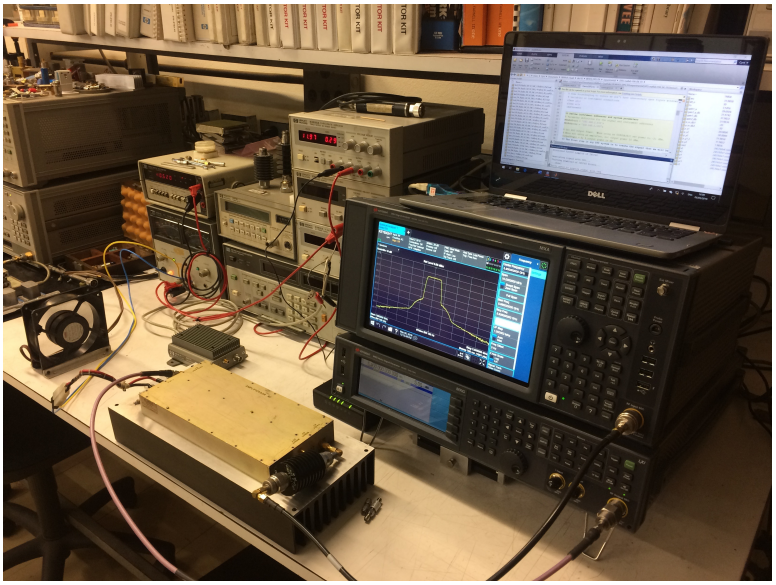


Figure 3.2 – Picture of test set-up during experimental measurements.

For each of the behavioural models and estimation techniques in this thesis, the experimental PA input and output signals, captured with the test set-up described above, are processed by the ILA DPD training block implemented in MATLAB. Following the ILA approach, the feedback path composed by the "PA" and "DPD training" blocks is equivalent to the forward path ("DPD" and "PA" blocks) (EUN; POWERS, 1997). Therefore, after completing the steps in Figure 3.1, the DPD training output signal and the linearized PA output signal (normalized) are equal, as shown in Figure 2.13.

In order to choose the input signal power level for the validation of the DPD techniques, firstly, the PA was driven with modulated input signal at four average power levels, $P_{in}(Avg) = -17$, $-12$, $-8$ and $-5$ dBm and the corresponding PA output complex envelopes were captured with the PA load impedance matched. The AM-AM and AM-PM

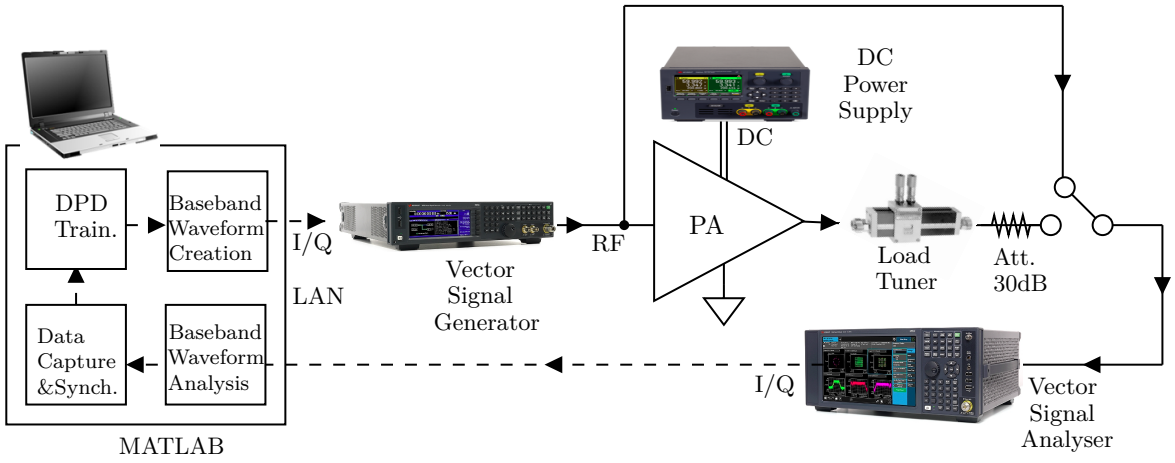
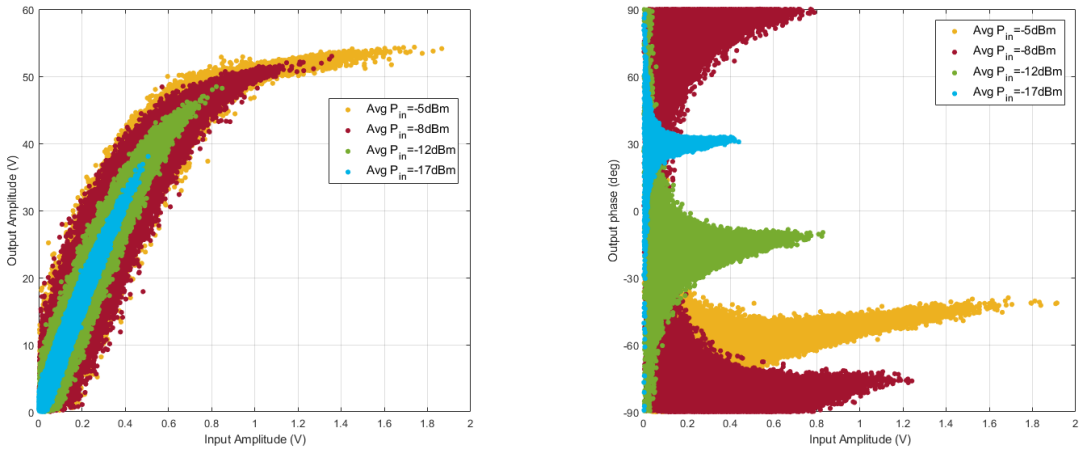



Figure 3.3 – Block diagram of the experimental test set-up.

plots were obtained by comparing the amplitudes and phases of the PA input and output waveforms on a sample-by-sample basis. Note that to compare the PA input and output data, proper time alignment between the waveforms is required. In Fig. 3.4a, the AM/AM plot shows that the PA output levels increase linearly with the input signal voltage until the PA reaches its gain compression region, where the output signal level saturates and only slightly increases with the input. This effect is observed specially for input powers of $P_{in}(Avg) = -8$ and $-5$ dBm, as shown in the figure. For input powers of $-17$ and $-12$ dBm, the PA is not in the non-linear region and very limited amplitude distortions occur, only at the extreme peak envelope excursions. In Fig. 3.4b, the AM/PM plot shows that the PA insertion phase remains constant until the PA reaches gain compression, where phase distortion is observable, specially for $P_{in}(Avg) = -8$ and $-5$ dBm. In both figures, the dispersion of the AM/AM and AM/PM data points is caused by PA memory effects. The PA increasingly exhibits non-linear behaviour and memory effects, as the input average power level increases and, based on these results, the OFDM input signal is set to average power $P_{in}(Avg) = -8$ dBm. This power level is able to drive the PA into non-linear operation and cause spectral regrowth and distortion and is chosen for both matched and mismatched conditions. Next, Figs. 3.5 and 3.6 illustrate, respectively, the PA gain versus the instantaneous input signal power and the power spectral density (PSD), for the average power levels of the input signal defined above. These figures also validate the choice of the input average power level for evaluating PA non-linear behaviour compensation techniques.

For $P_{in}(Avg) = -8$ dBm, the input and output signals are digitized, captured and processed off-line by the DPD training function in MATLAB. The ability of the behavioural models and estimation techniques to compensate for spectral regrowth and distortion is evaluated by computing the DPD training output signal, as discussed in section 2.5, and the figures of merit discussed in the next section.




(a) AM/AM (b) AM/PM

Figure 3.4 – PA measurements versus $P_{in}(Avg)$ (AM/AM and AM/PM).

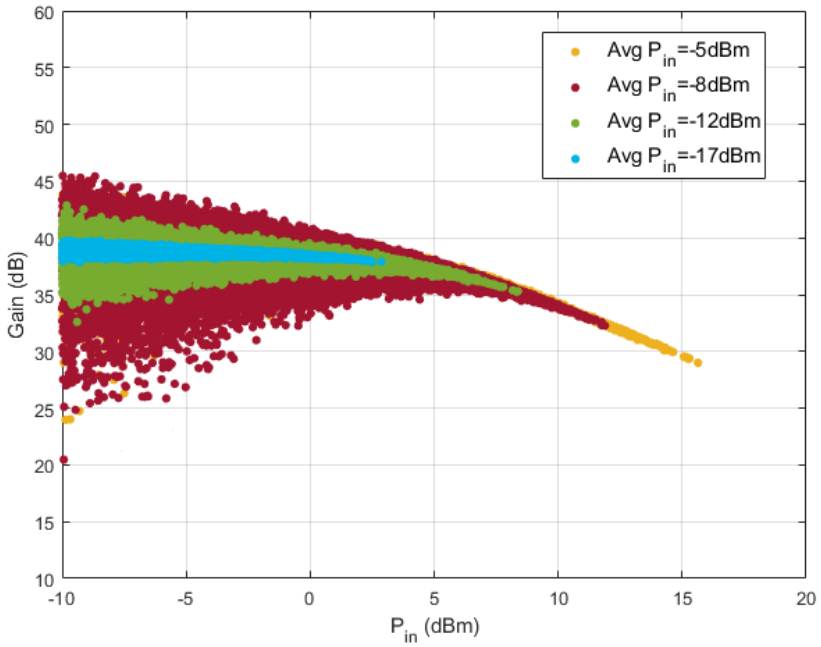



Figure 3.5 – PA measurements versus $P_{in}(Avg)$ (gain).

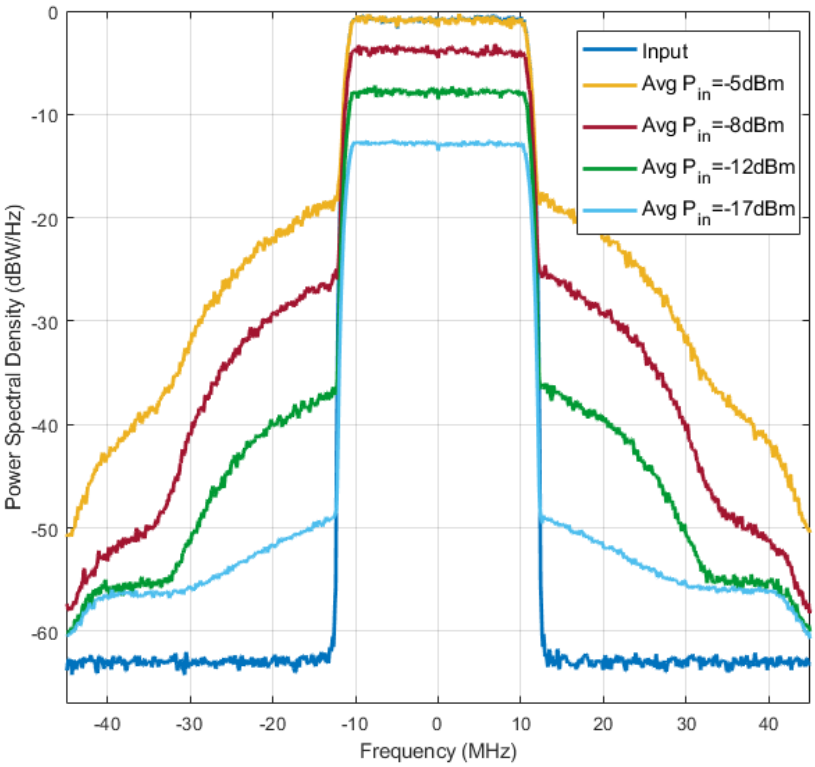



Figure 3.6 – Spectral regrowth versus $P_{in}(Avg)$.

## 3.2 FIGURES OF MERIT

The following figures of merit are required for evaluating the DPD performance.

The linearization accuracy evaluates how close the inverse model modelled and measured output signals are, being computed by the normalized mean square error (NMSE), in dB, given as follows:

$$NMSE(dB) = 10\log\frac{\|\tilde{\boldsymbol{x}} - \hat{\boldsymbol{x}}\|_2^2}{\|\hat{\boldsymbol{x}}\|_2^2}, \tag{3.1}$$

where $\hat{x}(n)$ and $\tilde{x}(n)$ are, respectively, the modelled and the measured output signals.

The out-of-band DPD performance is evaluated by the adjacent channel power ratio (ACPR), which is defined, in the frequency domain, as follows:

$$ACPR_{(dBc)} = \max_{i=1,2}\Big\{10\log\frac{\int_{(adj)_i} PSD(\omega)d\omega}{\int_{ch.} PSD(\omega)d\omega}\Big\}, \tag{3.2}$$

where $PSD(\omega)$ is the power spectral density of the DPD training output signal $\hat{x}(n)$ and $ch.$ and $(adj.)_i$ correspond to, respectively, the central and the (lower and upper) adjacent channels.

The in-band performance after demodulation is addressed by the error-vector magnitude (EVM), which is calculated in percentage by the following expression:

$$EVM_{RMS(\%)} = \sqrt{\frac{\frac{1}{N}\sum_{n=1}^{N}(\tilde{I}_{err}^2(n) + \tilde{Q}_{err}^2(n))}{\frac{1}{N}\sum_{n=1}^{N}(\tilde{I}_{ref}^2(n) + \tilde{Q}_{ref}^2(n))}} \times 100\%, \tag{3.3}$$

where $N$ is the sample size, $\tilde{I}_{err}(n) = \tilde{I}_{ref}(n) - \tilde{I}_{mea}(n)$ and $\tilde{Q}_{err}(n) = \tilde{Q}_{ref}(n) - \tilde{Q}_{mea}(n)$. The signals $\tilde{I}_{ref}(n)$ and $\tilde{Q}_{ref}(n)$ are, respectively, the $I$ and $Q$ components of the $n^{th}$ reference (ideal) symbol and $\tilde{I}_{mea}(n)$ and $\tilde{Q}_{mea}(n)$ are, respectively, the $I$ and $Q$ components of the $n^{th}$ measured (received) symbol.

In eq. (3.5), the NMSE does not take into account the number of coefficients required in the model, i.e., its running complexity. When the intention is to compare distinct models that fit a measured dataset, it is desirable to consider not only the performance, but also the running cost (in terms of number of coefficients), in this case, a mean squared error criterion that adds a penalty term can be used, e.g. Akaike's information criterion (AIC)(AKAIKE, 1974):

$$AIC = N\ln\Big(\frac{\|\tilde{\boldsymbol{x}} - \hat{\boldsymbol{x}}\|_2^2}{N}\Big) + 2\|\hat{\boldsymbol{\beta}}\|_0, \tag{3.4}$$

where $\hat{\boldsymbol{\beta}}$ is the candidate model coefficients vector and $\|\cdot\|_0$ is the $\ell_0$ norm of the vector. The AIC is a widely used selection criterion for choosing the best among several competing models, not only by its fitting capability, but also based on the number of model parameters required. Note that the AIC trades-off between the residual sum-of-squares and the number of estimated coefficients (BANKS; JOYNER, 2017). As the model accuracy increases, the logarithmic term becomes more negative.

## 3.3 PRELIMINARY SIMULATION STUDY

The objective of this preliminary simulation study is to briefly illustrate the impacts of some major sources of RF impairments on the TX output signal, then address their compensation by running the PA and DPD models described hereafter. MATLAB is used to implement the block diagrams in Figures 1.1 and 1.2, as well as the input signal with 16-QAM modulation and the RF impairments described below. The simulated PA in this study is described by the well-known MP model, with odd-order terms, as given in eq. (2.9). The MP model parameters are NL order $K = 5$, memory depth $M = 2$ and model coefficients provided in (DING et al., 2004) and (RAICH; ZHOU, 2002), as follows:

$$\begin{aligned}
c_{10} &= 1.0513 + 0.0904j \quad & c_{30} &= -0.0542 - 0.2900j \\
c_{50} &= -0.9657 - 0.7028j \quad & c_{11} &= -0.0680 - 0.0023j \\
c_{31} &= 0.2234 + 0.2317j \quad & c_{51} &= -0.2451 - 0.3735j \\
c_{12} &= 0.0289 - 0.0054j \quad & c_{32} &= -0.0621 - 0.0932j \\
c_{52} &= 0.1229 + 0.1508j
\end{aligned} \tag{3.5}$$

Figure 3.7 shows the PSD plots of the PA input and output signals, highlighting the spectral regrowth on the adjacent channels, due to the PA NL distortions. Figure 3.8 shows the 16-QAM constellation diagram at the output of the NL PA, where we observe the amplitude compression of some demodulated symbols, specially those with larger amplitudes. This is caused by the PA gain saturation and deteriorates the TX modulation quality. In the plot, the red circles denote the ideal symbol grid.

Additionally, the IQM imbalance is included in the simulation, by extending the MP model with the inclusion of an IQM imbalance filter (CRIPPS, 2006). Figure 3.9 illustrates the impact of IQM imbalance on the constellation diagram at the IQM output, considering the frequency-independent and frequency-dependent cases. In the frequency-dependent case, the memory effects are observable as dispersion around each symbol.

Finally, PA LMM is also included in the simulation, by adding a feedback FIR filter around the MP model, to describe in a simplified the reflections of the output signal (CRIPPS, 2006). The next figures present the simulated BER vs. SNR curves, obtained for the system under AWGN channel and PA NL, IQM imbalance and PA LMM impairments.

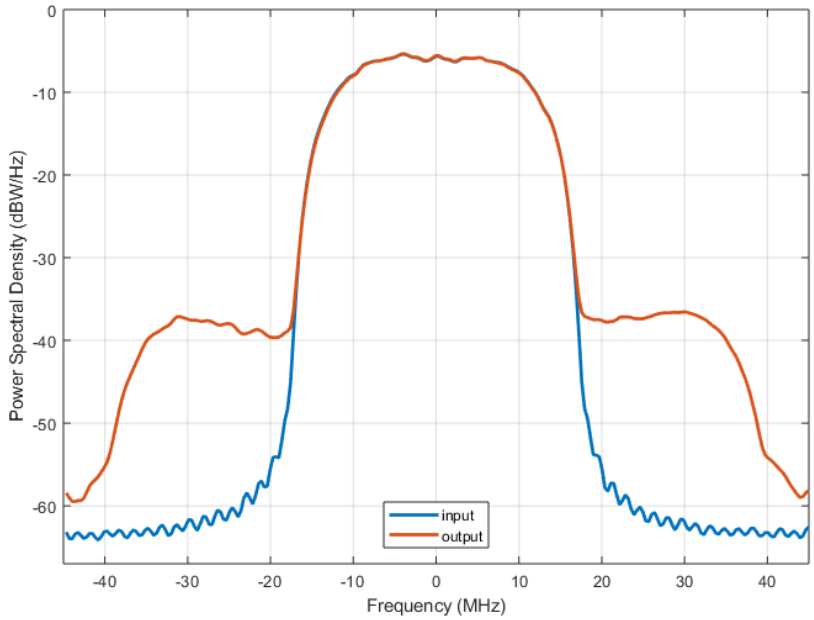
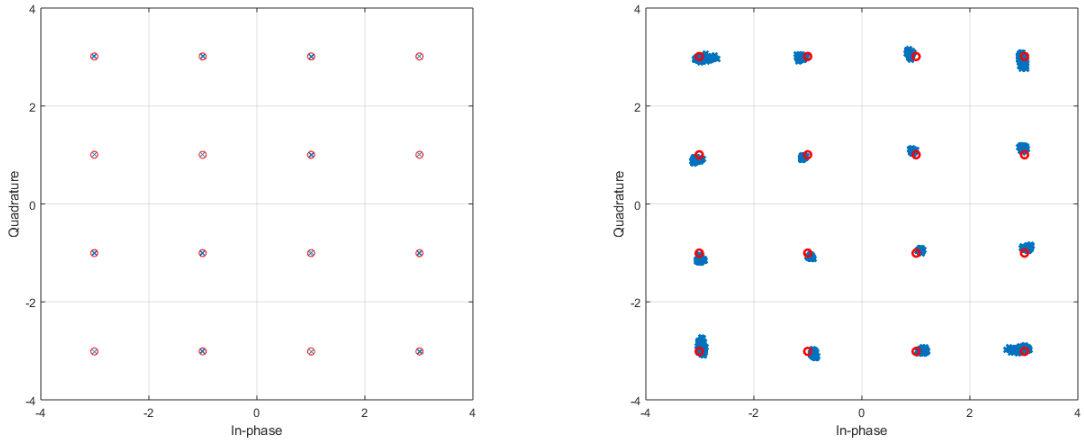



Figure 3.7 – Simulation of PA input and output signals.

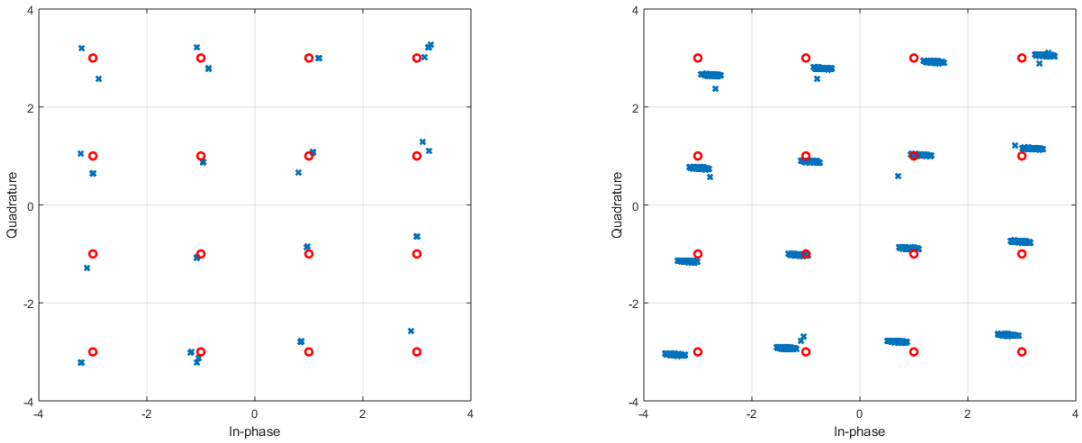



(a) Ideal TX (b) NL PA

Figure 3.8 – Simulation of constellation diagram at PA output.

Using the extended MP models above, Figures 3.10 and 3.11 show the impact, respectively, of a frequency-dependent IQM imbalance and of PA LMM. We observe from Figures 3.10 and 3.11 that, at low SNR, the system performance is dominated by the channel noise level. As the SNR increases, RF impairments contribute more significantly to degradation of BER, becoming the dominant factors at high SNR levels. Note that, due to the NL PA, the system BER improves more slowly with the signal power increase, because the PA becomes more NL.

Next, we apply the DPD technique to compensate for the RF impairments. First, the MP model is employed with parameters $K = 5$ and $M = 3$ and its coefficients are




(a) Frequency-independent imbalance (b) Frequency-dependent imbalance

Figure 3.9 – Simulation of constellation diagram at IQM output.

estimated using OLS during the DPD training phase, as presented in section 2.6. We observe that the MP model is able to adequately compensate for the NL PA, but its performance drastically deteriorates in the presence of PA LMM or IQM imbalance, as shown in Figures 3.10 and 3.11. Finally, a more sophisticated DPD model, the dual-input MP in eq. (A.2), to be discussed in the appendix A, with parameters $K = 5$ and $M = 3$, is able to jointly compensate for the NL PA and IQM imbalance, as shown in Figure 3.12. We conclude from these results that specialized models are required for the joint compensation of NL PA, IQM imbalance and/or PA LMM distortions, as will be discussed in the thesis.

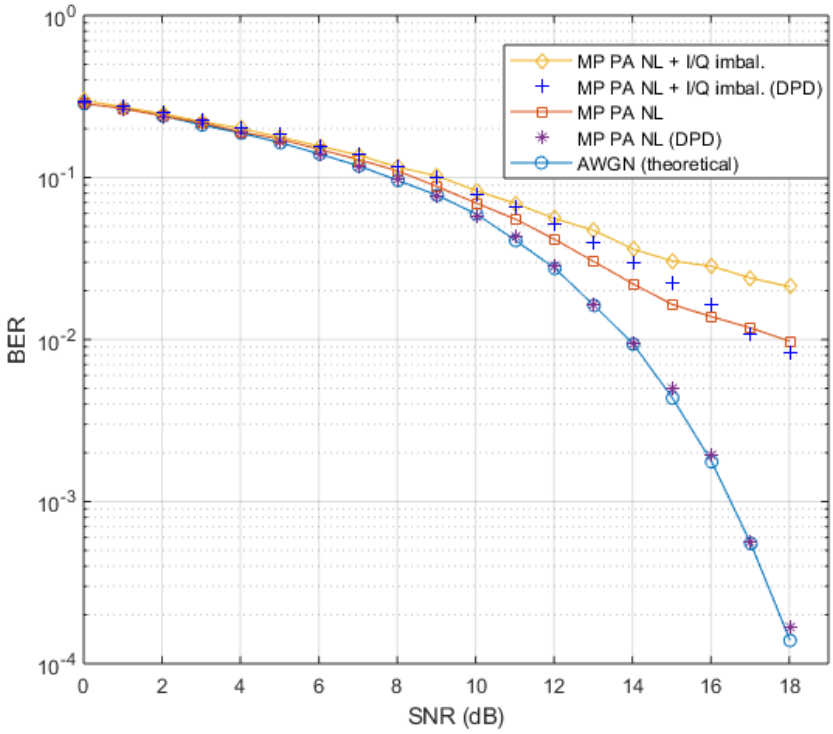



Figure 3.10 – Simulated BER vs. SNR with PA NL and IQM imbalance.

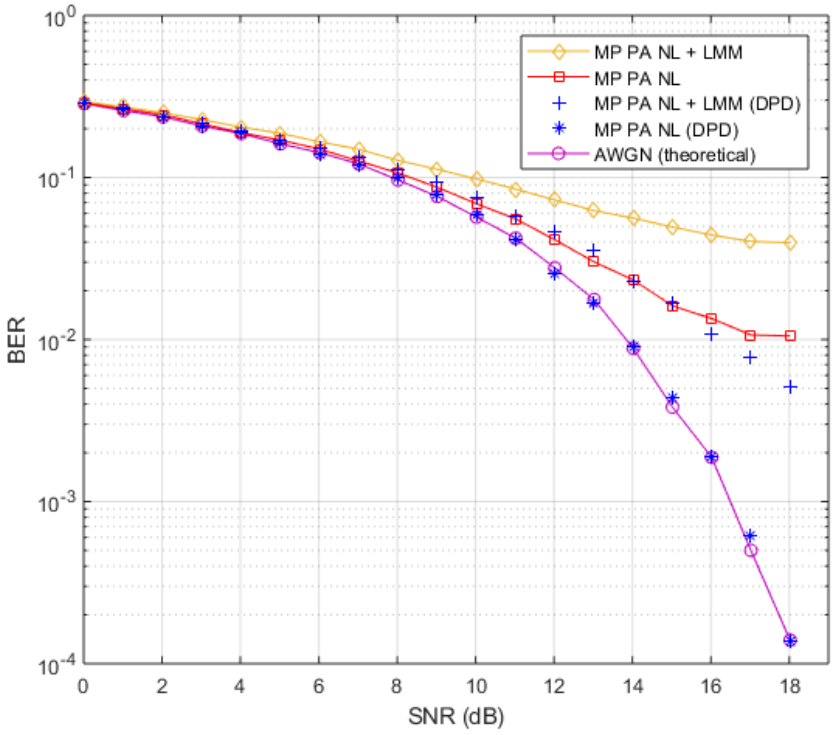



Figure 3.11 – Simulated BER vs. SNR with PA NL and LMM.

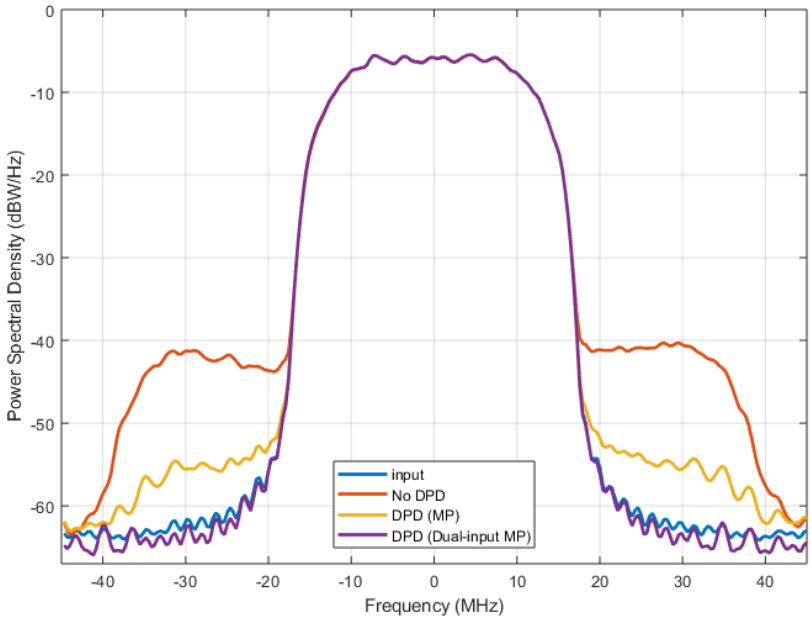



Figure 3.12 – Simulated residual spectral regrowth with NL PA and IQM imbalance.

## 3.4 CONCLUSIONS

This chapter presented the experimental test set-up used for the validation of the proposed DPD models and estimation techniques in this thesis. In addition, the figures of merit that used throughout this work were also defined. Finally, this chapter provided a preliminary simulation study to illustrate the detrimental impact of PA LMM and IQM imbalance on the performance of conventional DPD models, such as the MP, both in terms of simulated ACPR and BER.

# 4 VOLTERRA-PARAFAC AND LAGUERRE-VOLTERRA TECHNIQUES

The triangular FV model in eq. (2.4) requires a large number of coefficients to be estimated, being avoided in practice. In turn, as discussed in subsection 2.2.5, pruned-Volterra models have the disadvantage of not guaranteeing that a given structure is a good choice for an unknown PA or DPD, especially when physical knowledge is not available (ISAKSSON; WISELL; RONNOW, 2006). This is the motivation of this chapter, that is, to investigate alternative techniques proposed in the literature that avoid *a priori* prunning the Volterra model, but are able to reduce the complexity of the original model. The PARAFAC decomposition and the Laguerre expansion are the focus of this chapter, leading to the reduced complexity Volterra-PARAFAC (VP) and Laguerre-Volterra (LV) models, whose performances are then compared to those of the FV models. Also, in chapters 6 and 7, the VP and LV models are compared to the WHFB, sparse WHFB and the proposed pruned-Volterra models.

Sections 4.1 and 4.2 address, respectively, the doubly symmetric PARAFAC decomposition of Volterra kernels (BOUILLOC; FAVIER, 2012) and (CRESPO-CADENAS et al., 2014) and the expansion of Volterra kernels using Laguerre orthonormal basis functions (OLIVEIRA et al., 2011), (OLIVEIRA et al., 2012). The resulting VP and LV models are both capable of significantly reduce the complexity of the Volterra models, as analysed in section 4.3.

## 4.1 VOLTERRA-PARAFAC TECHNIQUE

This section investigates a new class of baseband Volterra models, developed in (BOUILLOC; FAVIER, 2012) and (CRESPO-CADENAS et al., 2014), called baseband Volterra–PARAFAC models, obtained using a doubly symmetric PARAFAC decomposition of high order Volterra kernels, treated as tensors. In order to explain how the VP technique works, first, we have to review, in the following subsection, a few concepts.

### 4.1.1 Tensor Definitions

A tensor is a multidimensional array. The number of ways or modes is called the order of a tensor. For example, a matrix is a tensor of order two. Tensors of order three or greater are called higher-order tensors. Let $n_1 \times n_2 \times \ldots \times n_m$ denote the size of an $m$-way tensor. The tensor is cubic if all the modes have the same size, i.e.,

$n = n_1 = n_2 = \cdots = n_m$ and $n$ is said to be the tensor dimension. Thus, the order of a tensor refers to the number of indices, while the dimension of a tensor, to the maximal value of each index. A tensor is symmetric if its elements do not change under any permutation of the indices. For example, a cubic three-way tensor $\mathcal{X} \in \mathbb{R}^{n \times n \times n}$ is symmetric if $x_{ijk} = x_{ikj} = x_{jik} = x_{jki} = x_{kij} = x_{kji}, \forall i, j, k = 1, \ldots, n$. The fibers of a tensor result from holding all but one index constant. In the mode-$i$ fiber, $i$ refers to the index that is not held constant. For a $3^{rd}$-order tensor, the three possible fibers are $\mathcal{X}(ij\cdot)$, $\mathcal{X}(i \cdot k)$ and $\mathcal{X}(\cdot jk)$. Tensors can be partially symmetric, for example, a cubic three-way tensor $\mathcal{X} \in \mathbb{R}^{n \times n \times n}$ is symmetric in modes one and two if $x_{ijk} = x_{jik}, \forall i, j, k = 1, \ldots, n$. The rank of a tensor is defined as the minimal number of rank-one tensors that approximately generates the tensor as their sum. These definitions are explored in the next subsection.

### 4.1.2 PARAFAC Decomposition

In the PARAFAC decomposition (KOLDA; BADER, 2009), (MARTÍ; BALDRICH, 2015), the general $k^{th}$-order tensor $\mathcal{H}$ is approximately decomposed into a sum of $R$ rank-one tensors, i.e., a sum of $R$ outer products of $k$ vectors:

$$\mathcal{H} = \sum_{r=1}^{R} \lambda_r \boldsymbol{a}_{(1),r} \circ \boldsymbol{a}_{(2),r} \circ \ldots \circ \boldsymbol{a}_{(k),r} + \boldsymbol{\epsilon}, \tag{4.1}$$

where $\boldsymbol{a}_{(i),r}$ is the $r^{th}$-column of the factor matrix $\boldsymbol{A}_{(i)}$, $\lambda_r$ are scalar weights, $\circ$ is the vector outer product and $\boldsymbol{\epsilon}$ are residuals. The minimal $R$ required in eq. (4.1) that allows $\boldsymbol{\epsilon}$ to be neglected is called the rank of $\mathcal{H}$.

Assuming the scalars $\lambda_r$ can be absorbed into the vectors and neglecting $\boldsymbol{\epsilon}$, a $3^{rd}$-order tensor $\mathcal{H}_3$ of rank $R_3$ is decomposed into a sum of $R_3$ rank-one tensors:

$$\mathcal{H}^{(3)} = \sum_{r=1}^{R_3} \boldsymbol{a}_r \circ \boldsymbol{b}_r \circ \boldsymbol{c}_r, \tag{4.2}$$

where $\boldsymbol{A}$ , $\boldsymbol{B}$ and $\boldsymbol{C}$ are the factor matrices. Note that in the element-wise notation, the PARAFAC decomposition can be given as:

$$h^{(3)}_{ijk} = \sum_{r=1}^{R_3} a_{ir} b_{jr} c_{kr} \tag{4.3}$$

and, equivalently, in the matricized format, as:

$$\boldsymbol{H}_{(1)} = \boldsymbol{A}(\boldsymbol{C} \odot \boldsymbol{B})^T, \tag{4.4a}$$

$$\boldsymbol{H}_{(2)} = \boldsymbol{B}(\boldsymbol{C} \odot \boldsymbol{A})^T, \tag{4.4b}$$

$$\boldsymbol{H}_{(3)} = \boldsymbol{C}(\boldsymbol{B} \odot \boldsymbol{A})^T, \tag{4.4c}$$

where $\boldsymbol{H}_{(i)}$, with $i = 1, 2, 3$, is the mode-$i$ matricization of the tensor $\mathcal{H}^{(3)}$ and $\odot$ is the Khatri-Rao product (KOLDA; BADER, 2009), which is defined below. Matricization is the process of reordering the elements of a tensor into a matrix. The mode-$i$ matricization of a tensor $\mathcal{H} \in \mathbb{R}^{I_1 \times \ldots \times I_N}$ is denoted by $\boldsymbol{H}_{(i)}$ and arranges the mode-$i$ fibers as the columns of the resulting matrix. The Khatri-Rao (column-wise Kronecker) product of matrices $\boldsymbol{A} \in \mathbb{R}^{I \times K}$ and $\boldsymbol{B} \in \mathbb{R}^{J \times K}$ is given by the matrix of size $IJ \times K$:

$$\boldsymbol{A} \odot \boldsymbol{B} = \begin{bmatrix} \boldsymbol{a}_1 \otimes \boldsymbol{b}_1 & \boldsymbol{a}_2 \otimes \boldsymbol{b}_2 & \ldots & \boldsymbol{a}_L \otimes \boldsymbol{b}_L \end{bmatrix}, \tag{4.5}$$

where $\boldsymbol{A}$ and $\boldsymbol{B}$ have the same number of columns, $\boldsymbol{a}_l$ and $\boldsymbol{b}_l$ $(1 \leq l \leq L)$ denote the $l^{th}$ columns of $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively. The Khatri-Rao product of $\boldsymbol{A}$ and $\boldsymbol{B}$ can also be expressed as:

$$\boldsymbol{A} \odot \boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}\, diag_1(\boldsymbol{A}) \\ \vdots \\ \boldsymbol{B}\, diag_L(\boldsymbol{A}) \end{bmatrix}, \tag{4.6}$$

where $diag_l(\boldsymbol{A})$ denotes the diagonal matrix formed from the $l^{th}$ row of $\boldsymbol{A}$.

### 4.1.3 Baseband Volterra-PARAFAC Model

The tensor decomposition of complex-valued baseband Volterra kernels was first investigated in (KIBANGOU; FAVIER, 2008). In the triangular baseband Volterra model, each $k^{th}$-order kernel is represented as a $k^{th}$-order tensor $\mathcal{H} \in \mathbb{C}^{(M+1)\times \ldots \times (M+1)}$, symmetric both with respect to the first $\frac{k+1}{2}$ indices and to the last $\frac{k-1}{2}$ ones as well, thus doubly-symmetric.

From (BOUILLOC; FAVIER, 2012), each doubly-symmetric tensor can be decomposed into two factor matrices of size $(M+1) \times R_k$, where $R_k$ denotes the rank of the $k^{th}$-order tensor. Using this property, we have that the baseband Volterra-PARAFAC DPD model for each NL order is given by a sum of rank-1 tensors, as follows:

$$\hat{x}_{VP}(n) = \sum_{\substack{k=1 \\ k\ odd}}^{K} \sum_{r=1}^{R_k} \left( \tilde{\boldsymbol{u}}^T(n) \boldsymbol{a}_r^{(k)} \right)^{\frac{k+1}{2}} \left( \tilde{\boldsymbol{u}}^H(n) \boldsymbol{b}_r^{(k)} \right)^{\frac{k-1}{2}}, \tag{4.7}$$

where $\tilde{\boldsymbol{u}}(n)$ is the input vector $\begin{bmatrix}\tilde{u}(n) & \dots & \tilde{u}(n-M)\end{bmatrix}^T$, $\boldsymbol{a}_r^{(k)}$ and $\boldsymbol{b}_r^{(k)}$ are the $r^{th}$ columns of the $k^{th}$-order factor matrices $\boldsymbol{A}^{(k)} \in \mathbb{C}^{(M+1)\times R_k}$ and $\boldsymbol{B}^{(k)} \in \mathbb{C}^{(M+1)\times R_k}$, respectively. Also note that $R_1$=1 and $\boldsymbol{a}^{(1)} = h_1(m_1)$.

In (CRESPO-CADENAS et al., 2014), additional symmetry properties of real PAs led to further simplifying the model. Thus, the factor matrices $\boldsymbol{A}^{(k)}$ and $\boldsymbol{B}^{(k)}$ in eq. (4.7) are related by (CRESPO-CADENAS et al., 2014):

$$\boldsymbol{b}_r^{(k)} = \boldsymbol{a}_r^{(k)^*}, \tag{4.8}$$

i.e., they are complex conjugate. Finally, the baseband Volterra-PARAFAC model is expressed as:

$$\hat{x}_{VP}(n) = \sum_{\substack{k=1 \\ k\ odd}}^{K} \sum_{r=1}^{R_k} |\tilde{\boldsymbol{u}}^T(n)\boldsymbol{a}_r^{(k)}|^{k-1}\tilde{\boldsymbol{u}}^T(n)\boldsymbol{a}_r^{(k)}, \tag{4.9}$$

which corresponds to Wiener models in parallel, as illustrated in Figure 4.1.

The number of model coefficients in this, reduced complexity, baseband Volterra-PARAFAC model is:

$$N_{VP} = \sum_{\substack{k=1 \\ k\ odd}}^{K} (M+1)R_k, \tag{4.10}$$

i.e., the sum of the numbers of elements of the factor matrices $\boldsymbol{A}^{(k)}$.

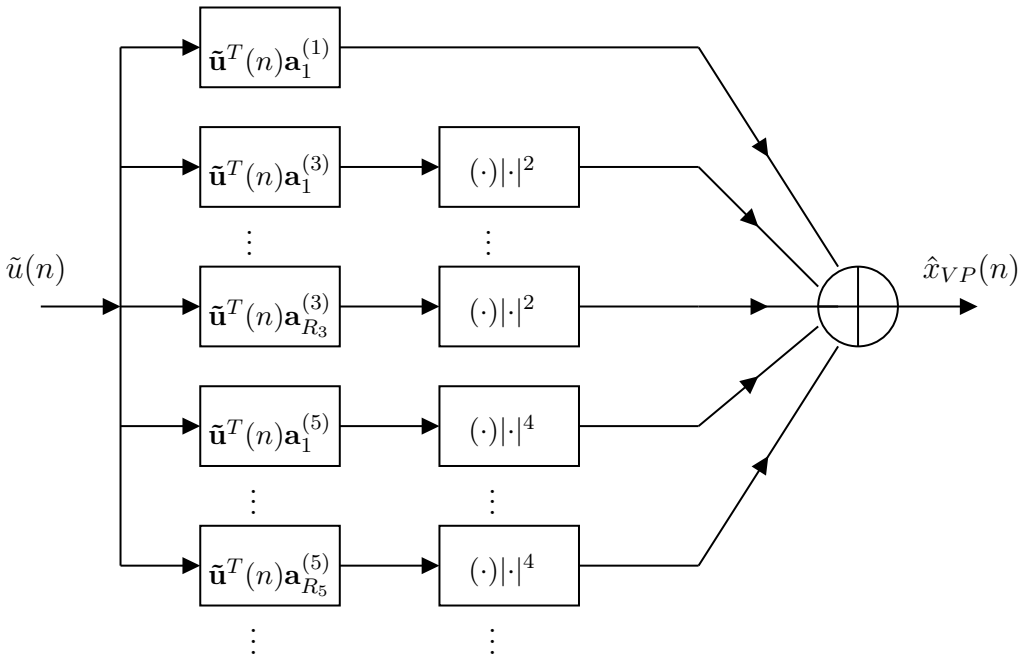



Figure 4.1 – Block diagram of the baseband Volterra-PARAFAC model.

### 4.1.4 Volterra-PARAFAC Estimation

This subsection discusses the estimation techniques proposed in the literature for the baseband Volterra-PARAFAC model. First, the alternating least squares (ALS) algorithm is considered. This algorithm has the disadvantage of requiring that the original full-Volterra kernels are first estimated, which is very expensive. Then, adaptive methods are presented that allow the Volterra-PARAFAC parameters to be estimated directly from input/output data. Two approaches, the adaptive complex least mean squares (CLMS) (BOUILLOC; FAVIER, 2012), (FAVIER; KIBANGOU; BOUILLOC, 2012) and non-linear least squares (NLS) algorithm using Levenberg-Marquardt (LM) (TOMASI; BRO, 2006), are capable of estimating Volterra-PARAFAC factor matrices directly from measured input and output datasets.

*Alternating Least Squares - ALS*

Fitting PARAFAC factor matrices to a given $3^{rd}$-order tensor, in the least squares sense, corresponds to finding the factor entries that minimise the following loss function:

$$J(a_{11}, a_{12}, \ldots, c_{KR}) = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(h_{ijk} - \sum_{r=1}^{R} a_{ir}b_{jr}c_{kr}\right)^2. \tag{4.11}$$

Using eq. (4.4a), the loss function is written as:

$$J(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}) = \|\boldsymbol{H}_{(1)} - \boldsymbol{A}(\boldsymbol{C} \odot \boldsymbol{B})^T\|_F^2, \tag{4.12}$$

where $\|\cdot\|_F$ is the Frobenius norm, defined as $\|\boldsymbol{X}\|_F = \|vec(\boldsymbol{X})\|_2$, in which $vec(\cdot)$ is the operator that stacks the matrix columns into a vector. Note that the Frobenius norm is also equal to $\langle X, X\rangle^{\frac{1}{2}}$ and $\langle\,,\rangle$ is the inner product (also called dot product). In the ALS algorithm, each of the factor matrices is evaluated separately, as follows:

$$\boldsymbol{A}^{(t)} = \boldsymbol{H}_{(1)}\left((\boldsymbol{C}^{(t-1)} \odot \boldsymbol{B}^{(t-1)})^\dagger\right)^T, \tag{4.13a}$$

$$\boldsymbol{B}^{(t)} = \boldsymbol{H}_{(2)}\left((\boldsymbol{C}^{(t-1)} \odot \boldsymbol{A}^{(t)})^\dagger\right)^T, \tag{4.13b}$$

$$\boldsymbol{C}^{(t)} = \boldsymbol{H}_{(3)}\left((\boldsymbol{B}^{(t)} \odot \boldsymbol{A}^{(t)})^\dagger\right)^T, \tag{4.13c}$$

where † denotes the Moore–Penrose pseudo-inverse. Firstly the matrices $\boldsymbol{B}$ and $\boldsymbol{C}$ are randomly initialized and the ALS sub-problems above are solved for each factor matrix in its turn, by solving a linear least squares problem. Then, $t$ is incremented and the steps are repeated until a predefined convergence criterion is met. Note that these expressions

could be generalized for higher tensor orders. Also note that the ALS algorithm does not guarantee that the global minimum is attained since it is an initialization-dependent iterative algorithm (BOQUÉ; FERRÉ, 2015).

In the case of a $k^{th}$-order doubly-symmetric tensor $\mathcal{H}$, the PARAFAC decomposition is given by:

$$\mathcal{H} = \sum_{r=1}^{R_k} (\boldsymbol{a}_r^{(k)})^{\frac{k+1}{2}} (\boldsymbol{a}_r^{(k)*})^{\frac{k-1}{2}}, \tag{4.14}$$

where the minimal value $R_k$ is said to be the symmetric rank of $\mathcal{H}$. The $k^{th}$-order tensor $\mathcal{H}$ is written as a summation of $R_k$ outer products of $\frac{k+1}{2}$ and $\frac{k-1}{2}$ identical vectors, respectively, $\boldsymbol{a}_r^{(k)}$ and $\boldsymbol{a}_r^{(k)*}$.

In this case, recalling from eq. (4.12), the ALS loss function of a $3^{rd}$-order doubly-symmetric tensor $\mathcal{H}_3$ can be expressed as:

$$J(\boldsymbol{A}) = \| \boldsymbol{H}_{(1)} - \boldsymbol{A}(\boldsymbol{A} \odot \boldsymbol{A}^*)^T \ \|_F^2. \tag{4.15}$$

Note that in this case the ALS sub-problems in eqs.(4.13a)-(4.13c) lead to repeated equations and become non-linear least squares (NLS) problems in the factor matrix $\boldsymbol{A}$. One approach discussed in the literature (KOLDA, 2014) is to use the standard ALS algorithm ignoring the symmetry, with the idea that it often converges to the symmetric solution (up to diagonal scaling). A simplification, proposed in (FAVIER; BOUILLOC, 2009) for single-symmetric Volterra kernels, is to randomly initialize $\boldsymbol{A}$ and iteratively compute $\boldsymbol{B}^{(t)} = \boldsymbol{A}^{(t-1)} \odot \boldsymbol{A}^{(t-1)*}$ and estimate $\boldsymbol{A}^{(t)} = \boldsymbol{H}_{(1)}((\boldsymbol{B}^{(t)})^{\dagger})^T$, until convergence.

To use the ALS algorithm, a set of Volterra kernels must first be estimated from the measured input/output data. Then, the ALS algorithm adjusts the PARAFAC model to the known Volterra kernels. In some cases, such as DPD, this procedure becomes very expensive, therefore, alternative adaptive approaches in the literature directly estimate approximate symmetric polyadic decompositions, instead of the symmetric Volterra kernels (FAVIER; KIBANGOU; BOUILLOC, 2012). Two popular approaches are the adaptive CLMS algorithm (BOUILLOC; FAVIER, 2012), (FAVIER; KIBANGOU; BOUILLOC, 2012), (CRESPO-CADENAS et al., 2014) and the NLS using the LM algorithm (TOMASI; BRO, 2006). Their main advantage is the ability to estimate the Volterra-PARAFAC factor matrices directly from the measured input/output data.

*Complex Least Mean Squares - CLMS*

The baseband Volterra-PARAFAC model is expressed as:

$$\hat{x}_{VP}(n) = \sum_{\substack{k=1 \\ k\ odd}}^{K} \sum_{r=1}^{R_k} |\tilde{\boldsymbol{u}}^T(n)\boldsymbol{a}_r^{(k)}|^{k-1}\tilde{\boldsymbol{u}}^T(n)\boldsymbol{a}_r^{(k)}, \tag{4.16}$$

where $R_1 = 1$ and $\boldsymbol{a}^{(1)} = \tilde{h}_1(m)$.

Defining the parameter vector:

$$\boldsymbol{\theta}^T = \begin{bmatrix} \boldsymbol{\theta}^{(1)^T} & \dots & \boldsymbol{\theta}^{(K)^T} \end{bmatrix}^T, \tag{4.17}$$

with $\boldsymbol{\theta}^{(1)} = \boldsymbol{a}^{(1)}$ and

$$\boldsymbol{\theta}^{(k)} = vec(\begin{bmatrix} \boldsymbol{a}_1^{(k)} & \dots & \boldsymbol{a}_{R_k}^{(k)} \end{bmatrix}), \tag{4.18}$$

where the vectorization operator stacks the column vectors of its argument.

The CLMS algorithm minimizes the cost function:

$$\begin{aligned} J(n, \boldsymbol{\theta}) &= \frac{1}{2}|\tilde{e}(n)|^2 \\ &= \frac{1}{2}\tilde{e}(n)\tilde{e}^*(n), \end{aligned} \tag{4.19}$$

where the error is the difference between the measured and modelled outputs $\tilde{e}(n) = \tilde{x}(n) - f(n, \boldsymbol{\theta})$ and $f(n, \boldsymbol{\theta}) = \hat{x}_{VP}(n)$.

Applying the steepest-descent algorithm, the update equation for the estimated parameters is (FAVIER; KIBANGOU; BOUILLOC, 2012):

$$\begin{aligned} \hat{\boldsymbol{\theta}}^{(k)}(n) = \hat{\boldsymbol{\theta}}^{(k)}(n-1) &+ \frac{1}{2}\mu_k\tilde{e}^*(n)\frac{\partial f(n, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{(k)^*}}|_{\hat{\boldsymbol{\theta}}^{(k)}(n-1)} + \\ &\frac{1}{2}\mu_k\tilde{e}(n)\frac{\partial f^*(n, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{(k)^*}}|_{\hat{\boldsymbol{\theta}}^{(k)}(n-1)}, \end{aligned} \tag{4.20}$$

where $\mu_k$ are the step-size parameters that trade-off the speed of adaptation and the noise in steady-state.

The gradients of the NL functions $f(n, \boldsymbol{\theta})$ and $f^*(n, \boldsymbol{\theta})$ with respect to the conjugated parameter vector $\boldsymbol{\theta}^{(k)^*}$ can be calculated by:

$$\frac{\partial f^*(n,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{(k)^*}}|_{\hat{\boldsymbol{\theta}}^{(k)}(n-1)} = \frac{(k+1)}{2}\hat{\boldsymbol{v}}^{(k)}(n) \otimes \tilde{\boldsymbol{u}}^*(n), \tag{4.21a}$$

$$\frac{\partial f(n,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^{(k)^*}}|_{\hat{\boldsymbol{\theta}}^{(k)}(n-1)} = \frac{(k-1)}{2}\hat{\boldsymbol{w}}^{(k)}(n) \otimes \tilde{\boldsymbol{u}}^*(n), \tag{4.21b}$$

where $\otimes$ is the Kronecker product, with:

$$\hat{\boldsymbol{v}}^{(k)}(n) = \left[\hat{\alpha}^{(k),1}(n) \quad \dots \quad \hat{\alpha}^{(k),R_k}(n)\right], \tag{4.22}$$

where $\hat{\alpha}^{(k),r}(n) = |\boldsymbol{u}^T(n)\boldsymbol{a}_r^{(k)}(n-1)|^{k-1}$ and

$$\hat{\boldsymbol{w}}^{(k)}(n) = \left[\hat{\beta}^{(k),1}(n) \quad \dots \quad \hat{\beta}^{(k),R_k}(n)\right], \tag{4.23}$$

where $\hat{\beta}^{(k),r}(n) = |\tilde{\boldsymbol{u}}^T(n)\boldsymbol{a}_r^{(k)}(n-1)|^{k-3}(\tilde{\boldsymbol{u}}^T(n)\boldsymbol{a}_r^{(k)}(n-1))^2$.

*Non-linear Least Squares using Levenberg-Marquardt - NLS-LM*

The LM is a robust algorithm for minimizing the NLS cost function in a numerically stable way and is proposed in (TOMASI; BRO, 2006) for fitting the Volterra-PARAFAC model. In this approach, the update equation of the estimated parameters becomes:

$$\hat{\boldsymbol{\theta}}^{(k)}(n) = \hat{\boldsymbol{\theta}}^{(k)}(n-1) + \left[\hat{\boldsymbol{J}}^{(k)^T}(n)\hat{\boldsymbol{J}}^{(k)}(n) + \lambda_k\, diag(\hat{\boldsymbol{J}}^{(k)^T}(n)\hat{\boldsymbol{J}}^{(k)}(n))\right]^{-1}\hat{\boldsymbol{J}}^{(k)^T}(n)\tilde{e}(n)\,, \tag{4.24}$$

where $\lambda_k$ is a non-negative damping parameter and $\boldsymbol{J}^{(k)}(n)$ is the Jacobian matrix of the NL function $f(n,\boldsymbol{\theta})$ with respect to the coefficients vector $\hat{\boldsymbol{\theta}}^{(k)}$ calculated at the point $\hat{\boldsymbol{\theta}}^{(k)}(n-1)$. Note that this algorithm converges to a local minimum of the cost function. Hence, good initial values of the parameters are very important to ensure the convergence of the estimates to the global minimum.

In section 4.3, simulation and experimental results of the baseband VP model are presented and discussed.

## 4.2 LAGUERRE-VOLTERRA TECHNIQUE

This section investigates the expansion of baseband Volterra kernels using orthonormal base functions (OBF), in particular Laguerre functions. The objective is to arrive at

the equivalent Laguerre-Volterra model, with reduced complexity. This technique, proposed for the first time by (WIENER, 1958), and is well covered, for example in (OLIVEIRA et al., 2011), (OLIVEIRA et al., 2012). The main idea is that, by choosing the OBF dynamics (i.e., poles) close to the real poles of the system to be modelled, the model's performance improve quickly with the number of coefficients. This leads to a model of reduced complexity, comparing to the Volterra model (KIBANGOU; FAVIER, 2005).

Several PA and DPD modelling papers expand the full-Volterra kernels using a truncated OBF series, in a technique known as the fixed pole expansion technique (FPET) (WILLIAMSON; HACIOGLU, 2001). The Kautz functions, i.e., a set of orthonormal functions with one complex-valued pole per NL order, have been applied in PA modelling (TEHRANI et al., 2008), (ISAKSSON; RONNOW, 2007) and DPD (ISAKSSON; WISELL; RONNOW, 2006). The Laguerre functions, with a single real pole per NL order, have been also applied in PA modelling (ZHU; BRAZIL, 2005). A generalized OBF, the Takenaka-Malmquist, having several complex poles per NL order, is applied in (SCHUMACHER; LIMA; OLIVEIRA, 2015). In (AMIN; HANDEL; RONNOW, 2017), fixed-pole MP and GMP models are developed using Laguerre functions. Also, OBF expansion has been applied to various other fields, such as control (ZHENG; ZAFIRIOU, 1995), (ROSA; CAMPELLO; AMARAL, 2007) and biological systems (MARMARELIS, 1993).

### 4.2.1 Laguerre Expansion

We focus on the Laguerre expansion, where the model is completely determined by a single real pole for each NL order. This makes the search for the unknown optimal poles faster than with the Kautz and generalized OBF functions. As illustrated in Figure 4.2, in the LV model, the input signal is first filtered by a linear filter bank composed of infinite impulse response (IIR) filters. The impulse responses of these filters follow a set of Laguerre functions, with pre-defined real poles. Next, the outputs of the filters are combined in a multiple-input static non-linearity. Assuming that the $k^{th}$-order Volterra kernel has fading memory with a given memory length $M_k$, i.e., $\tilde{h}_k(m_1,\ldots,m_k)=0$ for $m_i > M_k$ ($\forall i \in 1,\ldots,k$), the kernel is absolutely summable and can be approximated by a combination of terms from a truncated series of $F_k+1$ Laguerre functions (FAVIER; CAMPELLO; AMARAL, 2004):

$$\begin{aligned}&\tilde{h}_k(m_1,\ldots,m_k)\\&=\sum_{l_1=0}^{F_k}\sum_{l_2=l_1}^{F_k}\cdots\sum_{l_{\frac{k+1}{2}}=l_{\frac{k-1}{2}}}^{F_k}\cdots\sum_{l_{\frac{k+3}{2}}=0}^{F_k}\cdots\sum_{l_k=l_{k-1}}^{F_k}\tilde{c}_k(l_1,\ldots,l_k)\prod_{i=1}^{\frac{k+1}{2}}\phi_{k,l_i}(m_i)\prod_{i=\frac{k+3}{2}}^{k}\phi^*_{k,l_i}(m_i),\end{aligned}\tag{4.25}$$

where $\phi_{k,l}$ are the impulse responses of the Laguerre functions expanding the $k^{th}$-order kernel and $\tilde{c}_k(l_1,\ldots,l_k)$ are the triangular $k^{th}$-order expansion coefficients. The $l^{th}$ orthonormal Laguerre function is given by (ZHU; BRAZIL, 2005):

$$\phi_{k,l}(m) = \lambda_k^{\frac{m-l}{2}}(1-\lambda_k)^{\frac{1}{2}}\sum_{q=0}^{l}(-1)^q\binom{m}{q}\binom{l}{q}\lambda_k^{l-k}(1-\lambda_k)^k, \tag{4.26}$$

for $l = 0,\ldots,F_k$ and $m \geq 0$. The parameters $\lambda_k$, for $k = 1,3,\ldots$, are the real Laguerre poles, such that $|\lambda_k| < 1$. As mentioned, the parsimony of the resulting LV model depends on the choice of the Laguerre poles. Finally, note that the orthonormality property implies that:

$$\sum_{m=0}^{\infty}\phi_{k,l_i}(m)\phi_{k,l_j}(m) = \begin{cases}1, & \text{if } i=j.\\ 0, & \text{otherwise.}\end{cases} \tag{4.27}$$

As shown in Figure 4.2, in the $z$-domain, the Laguerre transfer functions correspond to:

$$\Phi_{k,l}(z,\lambda_k) = \frac{\sqrt{1-\lambda_k^2}}{1-z^{-1}\lambda_k}\left(\frac{-\lambda_k+z^{-1}}{1-z^{-1}\lambda_k}\right)^l, \tag{4.28}$$

where the term $\sqrt{1-\lambda_k^2}$ is a normalization factor, $\frac{1}{1-z^{-1}\lambda_k}$ is an IIR filter and each term

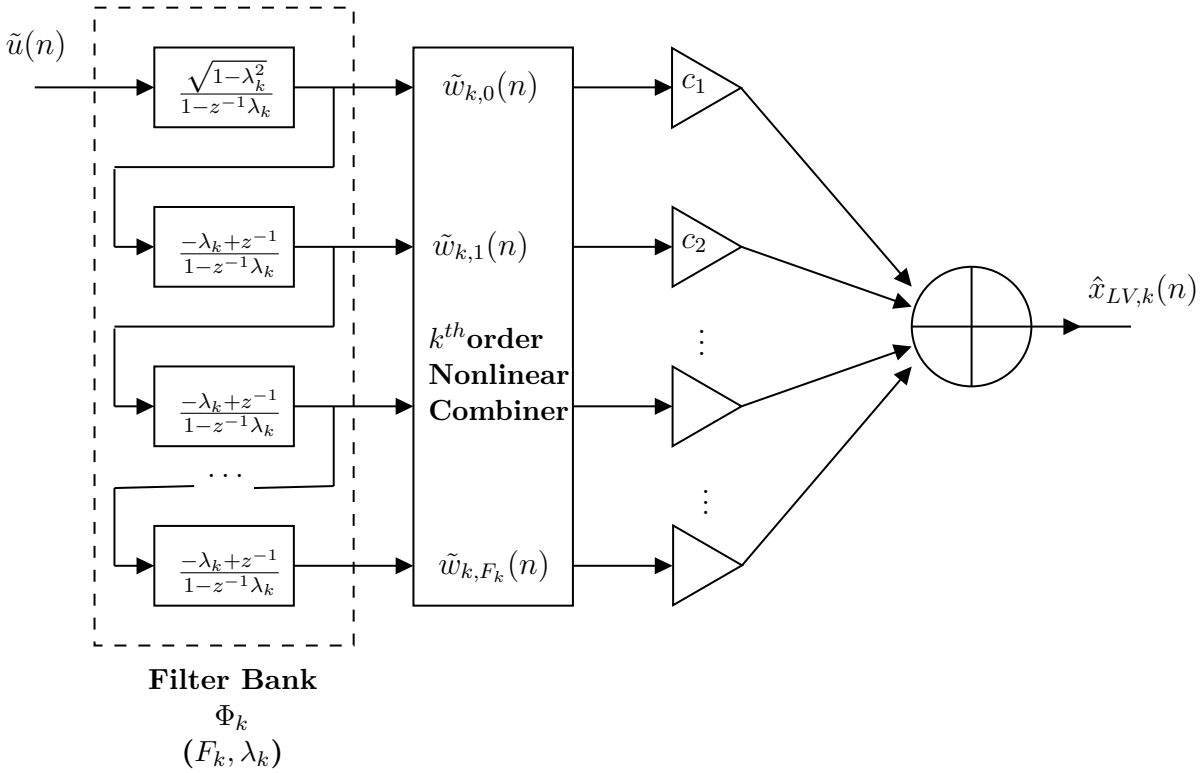



Figure 4.2 – Block diagram of the $k^{th}$-order Laguerre-Volterra model.

$\frac{-\lambda_k+z^{-1}}{1-z^{-1}\lambda_k}$ is an all-pass section.

### 4.2.2 Baseband Laguerre-Volterra Model

The LV DPD model is given by:

$$\begin{aligned}\hat{x}_{LV}(n) &= \sum_{\substack{k=1\\ k\ odd}}^{K} \hat{x}_{LV,k}(n) = \\ &\sum_{\substack{k=1\\ k\ odd}}^{K} \sum_{l_1=0}^{F_k} \sum_{l_2=l_1}^{F_k} \cdots \sum_{l_{\frac{k+1}{2}}=l_{\frac{k-1}{2}}}^{F_k} \cdots \sum_{l_{\frac{k+3}{2}}=0}^{F_k} \cdots \sum_{l_k=l_{k-1}}^{F_k} \tilde{c}_k(l_1,\ldots,l_k) \\ &\prod_{i=1}^{\frac{k+1}{2}} \tilde{w}_{k,l_i}(n) \prod_{i=\frac{k+3}{2}}^{k} \tilde{w}^*_{k,l_i}(n),\end{aligned} \tag{4.29}$$

where $\tilde{w}_{k,l}(n)$ corresponds to the input signal $\begin{bmatrix}\tilde{u}(n) & \ldots & \tilde{u}(n-M_k)\end{bmatrix}^T$ filtered by the $l^{th}$ Laguerre function in the memory span $[0, M_k]$, i.e.:

$$\begin{aligned}\tilde{w}_{k,l}(n) &= \sum_{m=0}^{M_k} \phi_{k,l}(m)\tilde{u}(n-m) \\ &= \Phi_{k,l}(z,\lambda_k)\tilde{u}(n).\end{aligned} \tag{4.30}$$

The terms of the LV model output in eq. (4.29) follow the topology in Figure 4.2, which is known as the Wiener-Bose topology (MARMARELIS, 1993). Also note that, if $\lambda_k$=0, $\Phi_{k,l}(z) = z^{-l}$, $l = 0,\ldots,F_k$, i.e., Volterra basis functions are obtained.

Figure 4.3 shows the first four Laguerre impulse response functions, given by eq. (4.26), with parameter $\lambda$=0.7, and also the Volterra impulse train. We observe that the Laguerre basis functions spread over time, thus the LV model is expected to require a reduced number of coefficients. As per eq. (4.26), the Laguerre poles $\lambda_k$ determine the Laguerre dynamics, i.e., the rate of exponential asymptotic decline. The poles have to be decided *a priori*, based on any physical knowledge, or the pole selection is obtained by a NL optimization algorithm. Also, an analytical approach is developed in (FAVIER; CAMPELLO; AMARAL, 2004). Once the poles $\boldsymbol{\lambda}_k$, for $k = 1,3,\ldots,$ are selected, the model is linear in the parameters and the coefficients vectors $c_k(l_1,\ldots,l_k)$ are estimated by OLS, as follows:

$$\hat{\boldsymbol{\theta}}(\boldsymbol{\lambda}_k) = \arg\min_{\boldsymbol{\theta}} \sum_{n=0}^{N} |\tilde{x}(n) - \boldsymbol{\psi}(n,\boldsymbol{\lambda}_k)^T \hat{\boldsymbol{\theta}}(\boldsymbol{\lambda}_k)|^2, \tag{4.31}$$

where $\boldsymbol{\theta}$ is a parameter vector containing all coefficients $c_k(l_1, \ldots, l_k)$ and $\boldsymbol{\psi}(n)$ contains all the products $\tilde{w}_{k,l_1}(n)\tilde{w}_{k,l_2}(n)\ldots\tilde{w}^*_{k,l_k}(n)$ from eq. (4.29).

The number of coefficients in the LV model is given by:

$$N_{LV} = \sum_{\substack{k=1 \\ k\ odd}}^{K} \binom{F_k + \frac{k+1}{2}}{\frac{k+1}{2}} \binom{F_k + \frac{k-1}{2}}{\frac{k-1}{2}}. \tag{4.32}$$

i.e., it is independent of the system memory, but depends on the number of Laguerre basis functions $F_k$, for $k = 1, 3, \ldots$.

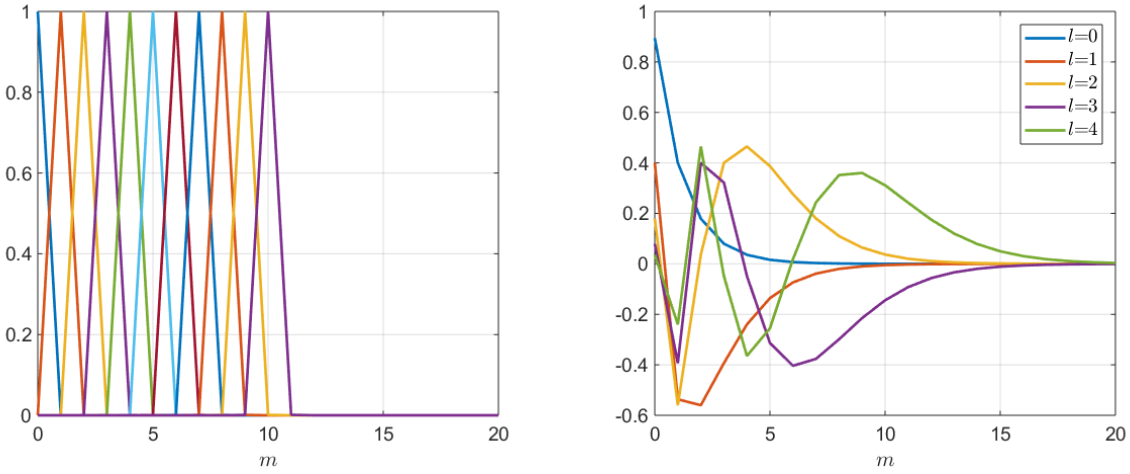
### 4.2.3 Laguerre-Volterra Estimation

The Laguerre poles in this study were obtained by a constrained NL least squares optimization, in which each pole position is optimized within the range $-1 \leq \lambda \leq 1$ based on minimizing the squared norm of the residual signal:

$$\min_{\boldsymbol{\lambda}} \sum_{n=0}^{N} |\tilde{x}(n) - \boldsymbol{\psi}(n, \boldsymbol{\lambda}_k)^T \hat{\boldsymbol{\theta}}(\boldsymbol{\lambda}_k)|^2, \tag{4.33}$$

where $\boldsymbol{\theta}$ is a parameter vector containing all coefficients $c_k(l_1, \ldots, l_k)$ and $\boldsymbol{\psi}(n)$ contains all the products $\tilde{w}_{k,l_1}(n)\tilde{w}_{k,l_2}(n)\ldots\tilde{w}^*_{k,l_k}(n)$ from eq. (4.29).

In the next section, simulation and experimental results of the baseband LV model are presented and discussed.




(a) Volterra ($M$=10) (b) Laguerre ($\lambda$=0.7)

Figure 4.3 – Volterra and Laguerre impulse response functions.

## 4.3 SIMULATION AND EXPERIMENTAL VALIDATION

The experimental validation presented in this section is based on measured input and output data from the LMM PA described in section 3.1 with VSWR 4:1 and the experimental test set-up described therein. The baseband VP and LV DPD models studied in this chapter are compared to the triangular FV and MV models. Afterwards, in chapters 6 and 7, the VP and LV models are also compared to the WHFB and to the proposed pruned-Volterra models.

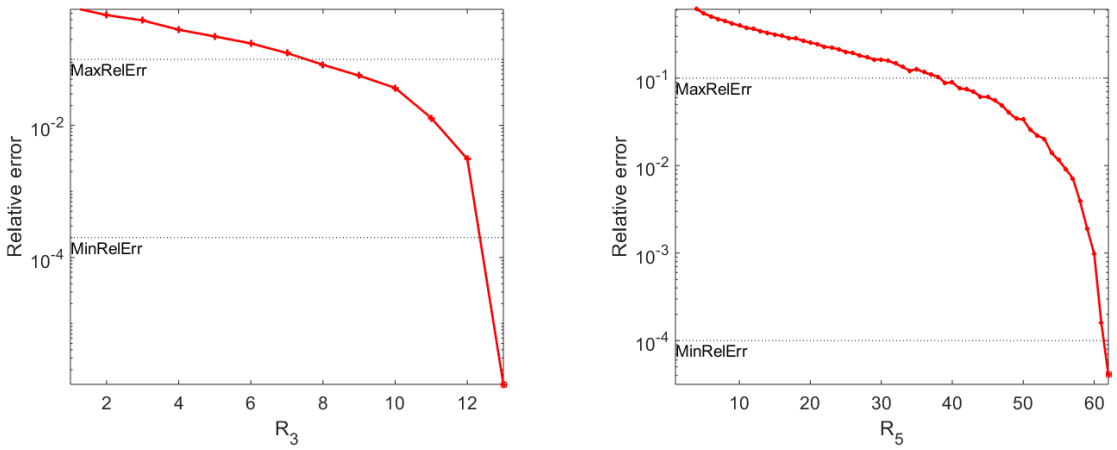
### 4.3.1 Baseband Volterra-PARAFAC Results

The PARAFAC decomposition of the tensors is performed using the Tensorlab 3.0 package (VERVLIET et al., 2016). First, the coefficients of the baseband triangular FV DPD model are estimated by LS regression using input and output samples. Then, the coefficients are expressed as doubly-symmetric kernels and, for each NL order, converted to tensors. Note that the PARAFAC decomposition requires an estimate of the tensor rank, which is often a difficult problem, involving trial and error (VERVLIET et al., 2016); the *rankest* tool from the Tensorlab package provides an estimate of the tensor rank. Finally, the Volterra-PARAFAC model is fitted in the LS sense using the ALS technique.

Figure 4.4 shows the $3^{rd}$- and $5^{th}$-order tensors rank estimations obtained from the *rankest* tool, for memory depth $M$, respectively, equal to 5 and 4. This tool evaluates the relative error between the tensor and its PARAFAC decomposition approximations (VERVLIET et al., 2016), $\|\boldsymbol{\epsilon}\|_F^2/\|\mathcal{H}\|_F^2$, where $\boldsymbol{\epsilon}$ is the residual in eq. (4.1), for increasing rank values, until $\boldsymbol{\epsilon}$ is negligible. In our work, the L-curve and further analysis of results are used to find the optimal trade-off between model accuracy and complexity reduction, thus arriving to the tensor rank estimation. Figure 4.5 compares the residual spectral regrowth after DPD obtained by the triangular full-Volterra ($FV_{\triangle}$) and the VP model (ALS estimation), with NL orders $K$=3 and $K$=5, memory length $M$=5, $R_3$=11, $R_5$=45 and $N$ =50,000 samples. As shown in the figure, the VP (ALS) model achieved ACPR and NMSE very close to the $FV_{\triangle}$, with much less coefficients, as detailed in the following tables.

Tables 4.1 and 4.2 summarize the numbers of coefficients, ACPR and NMSE obtained, respectively, with $K$=3 and $K$=5, including the VP ALS, CLMS and NLS estimation approaches, discussed in section 4.1. The CLMS and NLS approaches do not require that the $FV_{\triangle}$ kernels are first estimated, thus being advantageous in terms of computational cost and scalability. Note that the NLS LM algorithm uses the MATLAB optimization tool *lsqnonlin* (MATHWORKS, 2017) and that a local minimum may be reached, depending on the initialization. As shown on the tables, in this study, both the

CLMS and NLS-LM algorithms did not achieve the performance level of the ALS algorithm. Note that the tables also present results for the LV model, to be discussed in the next subsection. Finally, Figure 4.6 shows the convergence of the NLS LM algorithm, in terms of the norm of the residuals for the first 30 iterations.




(a) $3^{rd}$-order tensor rank (b) $5^{th}$-order tensor rank

Figure 4.4 – Tensor rank estimation, $k$=3 and $k$=5.

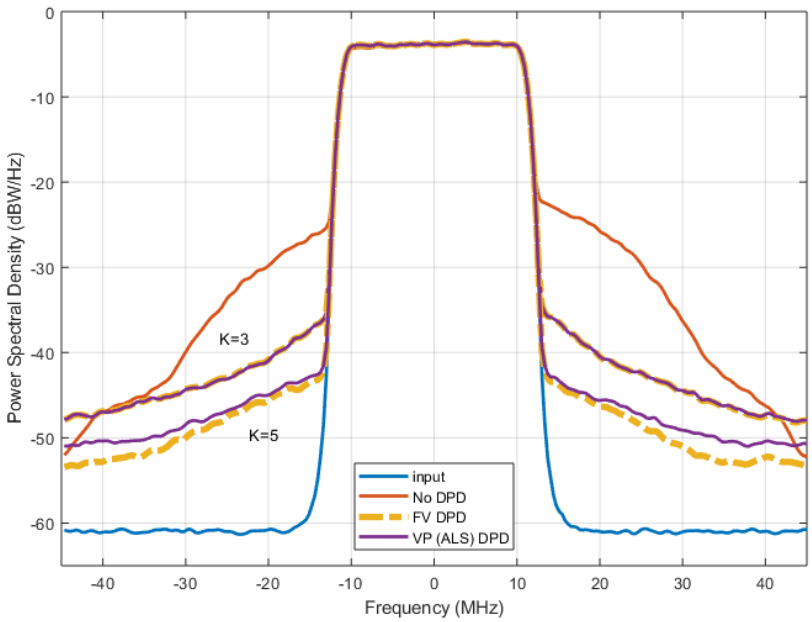



Figure 4.5 – Residual spectral regrowth for LMM PA with FV and VP DPD models ($K$=3 and $K$=5).

Now, we extend the VP ALS estimation applied above to higher NL orders $K$=7 and $K$=9. Due to the high computational cost of high order PARAFAC decompositions, however, instead of starting from the $FV_{\triangle}$ model, we apply the PARAFAC decomposition to the triangular modified Volterra ($MV_{\triangle}$) model from eq. (2.7), in which shorter memory

Table 4.1 – DPD performance for LMM PA with FV, VP and LV models ($K$=3, $M$=5 and $N$=50,000).

| | **Num. coeffs.** | **Max. ACPR (dBc)** | **NMSE (dB)** |
|---|---|---|---|
| No DPD | N/A | -23.2 | 3.3 |
| $FV_{\triangle}$ | 132 | -37.5 | -26.3 |
| VP (ALS) ($R_k = [1, 11]$) | 66 | -37.5 | -25.3 |
| VP (CLMS alg.) ($R_k = [1, 11]$, $\mu_1$=0.1, $\mu_3$=0.07) | 66 | -37.1 | -25.1 |
| VP (NLS LM alg.) ($R_k = [1, 11]$) | 66 | -38.1 | -24.8 |
| LV ($F_k = [5, 4]$) | 81 | -37.7 | -26.0 |

Table 4.2 – DPD performance for LMM PA with FV, VP and LV models ($K$=5, $M$=5, $N$=50,000).

| | **Num. coeffs.** | **Max. ACPR (dBc)** | **NMSE (dB)** |
|---|---|---|---|
| No DPD | N/A | -23.2 | 3.3 |
| $FV_{\triangle}$ | 1,308 | -42.5 | -32.5 |
| VP (ALS) ($R_k = [1, 11, 45]$) | 336 | -41.7 | -31.0 |
| VP (CLMS alg.) ($R_k = [1, 11, 45]$, $\mu_1$=0.1, $\mu_3$=0.07, $\mu_5$=0.07) | 336 | -39.5 | -29.3 |
| VP (NLS LM alg.) ($R_k = [1, 11, 45]$) | 336 | -39.7 | -29.3 |
| LV ($F_k = [5, 4, 3]$) | 281 | -41.6 | -31.5 |

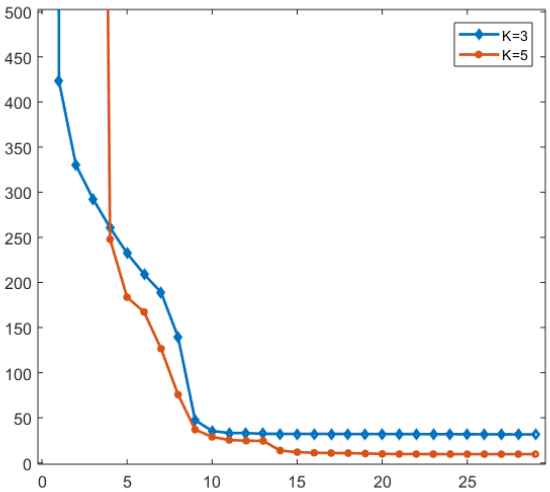
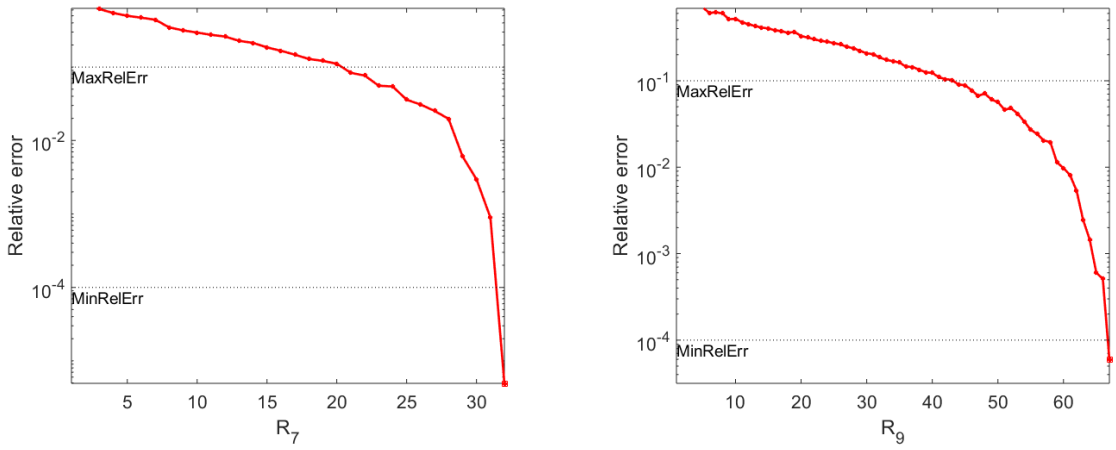



Figure 4.6 – NLS LM algorithm norm of residuals versus iterations.

lengths are configured for higher NL orders. Figure 4.7 shows the $7^{th}$- and $9^{th}$-order tensors rank estimation for $M$=2 obtained from the *rankest* tool. Tables 4.3 and 4.4 summarize the numbers of coefficients, ACPR and NMSE obtained, respectively, with $K$=7 and $K$=9, with

the VP ALS model. Figure 4.8 compares the residual spectral regrowth after DPD obtained by the $MV_{\triangle}$ and the VP model (ALS), with NL orders $K$=7 and $K$=9, memory length $M_k = [5, 5, 4, 3, 2]$, $R_k = [1, 11, 45, 42, 32]$ and $N$ =50,000 samples. Also results for the LV model are presented, to be discussed in the next subsection. As can be seen from the tables, the PARAFAC decomposition is able to significantly reduce the number of coefficients in the FV and MV DPD models, being a general approach for running complexity reduction. The ALS algorithm shows the disadvantage of fitting the PARAFAC model to *a priori* known Volterra kernels, thus requiring that the FV model is first estimated, which may be very costly. On the other hand, the adaptive CLMS and NLS LM algorithms approximate the VP model directly from input/output data, but, in this study, had to be run several times to avoid local minima and did not achieve the ALS performance. The results with the CLMS and NLS-LM algorithms were only presented for $K$=3 and $K$=5 (Tables 4.1 and 4.2).




(a) $7^{th}$-order tensor rank (b) $9^{th}$-order tensor rank

Figure 4.7 – Tensor rank estimation, $k$=7 and $k$=9.

Table 4.3 – DPD performance for LMM PA with MV, VP and LV models ($K$=7, $M$=5 and $N$=50,000).

| | **Num. coeffs.** | **Max. ACPR (dBc)** | **NMSE (dB)** |
|---|---|---|---|
| No DPD | N/A | -23.2 | 3.3 |
| $FV_{\triangle}$ | 8,364 | -46.3 | -35.0 |
| $MV_{\triangle}$ $(M_k = [5, 5, 4, 2])$ | 807 | -43.4 | -33.6 |
| VP (ALS) $(M_k = [5, 5, 4, 2])$ $(R_k = [1, 13, 45, 32])$ | 405 | -43.3 | -33.5 |
| LV $(F_k = [5, 4, 3, 2])$ | 431 | -43.1 | -33.5 |
| $MV_{\triangle}$ $(M_k = [5, 5, 3, 2])$ | 482 | -42.4 | -32.9 |
| VP (ALS) $(M_k = [5, 5, 3, 2])$ $(R_k = [1, 13, 30, 32])$ | 300 | -42.4 | -32.9 |
| LV $(F_k = [5, 4, 3, 1])$ | 301 | -42.8 | -33.0 |

Table 4.4 – DPD performance for LMM PA with FV, VP and LV models ($K$=9, $M$=5 and $N$=50,000).

| | **Num. coeffs.** | **Max. ACPR (dBc)** | **NMSE (dB)** |
|---|---|---|---|
| No DPD | N/A | -23.2 | 3.3 |
| $FV_{\triangle}$ | 11,875 | -46.5 | -35.3 |
| $MV_{\triangle}$ $(M_k = [5,5,4,3,2])$ | 1672 | -44.0 | -34.5 |
| VP (ALS) $(M_k = [5,5,4,3,2])$ $(R_k = [1,13,45,42,30])$ | 567 | -43.8 | -34.0 |
| LV $(F_k = [5,4,3,2,2])$ | 746 | -43.7 | -34.2 |
| $MV_{\triangle}$ $(M_k = [5,5,4,2,2])$ | 1122 | -43.5 | -33.9 |
| VP (ALS) $(M_k = [5,5,4,2,2])$ $(R_k = [1,13,45,32,30])$ | 501 | -43.5 | -33.6 |
| LV $(F_k = [5,4,3,2,1])$ | 461 | -43.3 | -33.8 |

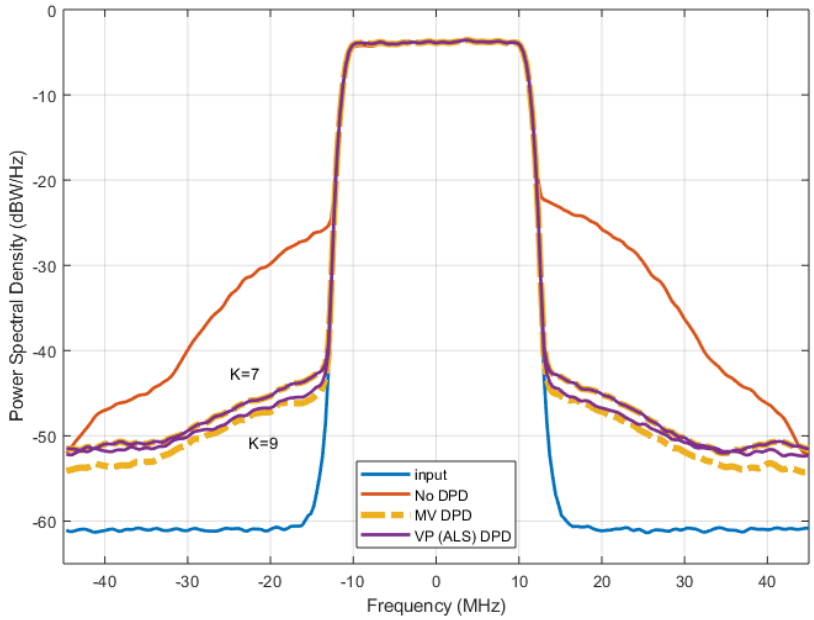


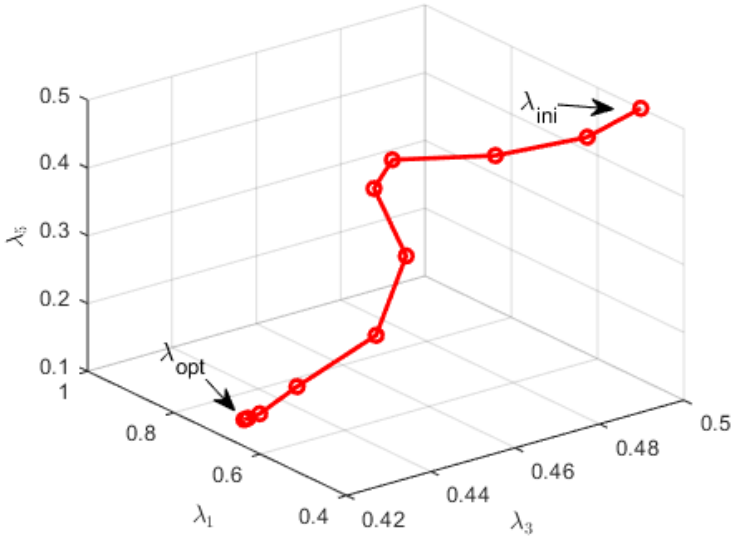

Figure 4.8 – Residual spectral regrowth for LMM PA with MV and VP DPD models ($K$=7 and $K$=9).

### 4.3.2 Baseband Laguerre-Volterra Results

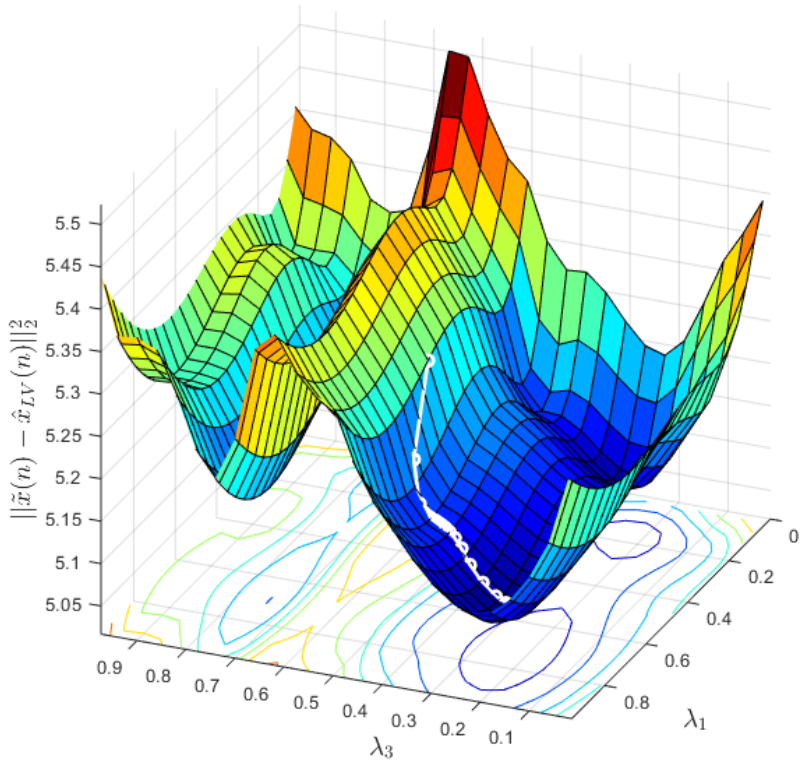
First, the optimal poles of the Laguerre functions, one real-valued pole per NL order, have to be obtained. The search for the optimal Laguerre poles is costly, but independent of the memory length. In this study, we used the MATLAB optimization tool *lsqnonlin* (MATHWORKS, 2017), running the algorithm for several initializations in order to avoid local minima. Figure 4.9 shows the pole optimization path of the LV model for $K$=5, with $F_k = [4,3,2]$ and initialization at $\boldsymbol{\lambda}_{ini} = [0.5, 0.5, 0.5]$. The figure also shows the objective function surface, i.e., the squared norm of the residual signal, illustrated for a grid of $\lambda_1$ and $\lambda_3$ values, while keeping $\lambda_5$ at its optimal value, previously obtained. Note that the

objective function shows several points of local minimum, thus the optimization solution depends on the initialization. The same optimization was also run for $K$=3. As shown in the figure, the Laguerre pole selection is a non-convex NL optimization problem, highly dependent on the initialization and subject to local minima. In this study, the optimization had to be run several times. Returning to Tables 4.1 and 4.2, they present the results for the LV model with the optimized poles, respectively, for $K$=3 and $K$=5, comparing them to the $FV_{\triangle}$ and VP models in terms of numbers of coefficients, ACPR and NMSE. As shown on the tables, the LV model also achieved ACPR and NMSE very close to the $FV_{\triangle}$, with much less coefficients. Note from the block diagram in Figure 4.2 that the number of coefficients in a LV model is independent of the system memory, thus being advantageous over the FV model for systems with low frequency characteristics, e.g., an RF PA with long-term memory effects.

The LV model results for NL orders $K$=7 and $K$=9 are presented, respectively, in Tables 4.3 and 4.4, comparing them to the $FV_{\triangle}$ and VP models. Moreover, in Figures 4.10 and 4.11, the residual spectral regrowth after DPD obtained with the $MV_{\triangle}$ and the LV models are compared for $K$=7 and $K$=9, respectively. As observed from the tables, the LV model is, similarly to the VP model, a general approach for significantly reducing the number of $FV_{\triangle}$ and $MV_{\triangle}$ DPD coefficients and, therefore, their running complexity. Finally, the LV model does not require the $FV_{\triangle}$ model coefficients to be first estimated; this aspect increases its scalability comparing to the VP (ALS) model.




(a) Optimization path.




(b) Objective function surface.

Figure 4.9 – Optimization of the LV poles for $K$=5.

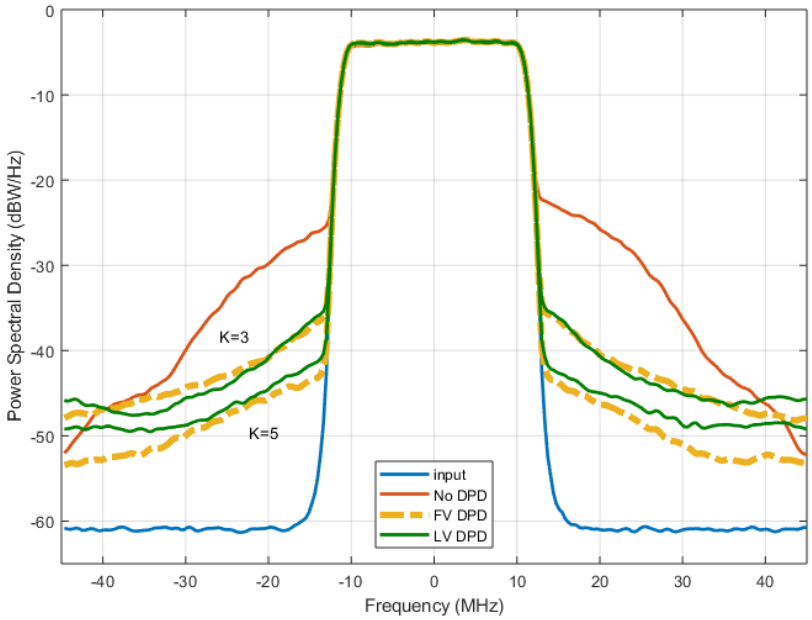



Figure 4.10 – Residual spectral regrowth for LMM PA with FV and LV DPD models ($K$=3 and $K$=5).

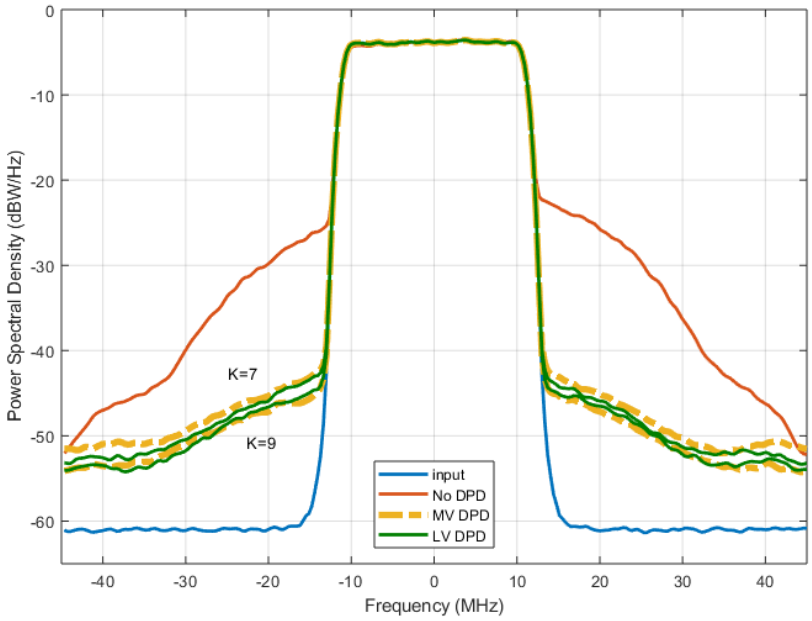



Figure 4.11 – Residual spectral regrowth for LMM PA with MV and LV DPD models ($K$=7 and $K$=9).

## 4.4 CONCLUSIONS

In this chapter, the PARAFAC decomposition and Laguerre expansion were applied to the triangular FV model, leading, respectively, to the baseband Volterra-PARAFAC and Laguerre-Volterra models. Without *a priori* enforcing a Volterra pruning strategy, these techniques were able to significantly reduce the numbers of coefficients and the running complexity of the original models. On the other hand, both techniques rely on NL

optimizations and tend to become very costly as the model dimensions increase. In the next chapter, a simplified DPD behavioural model, linear in the coefficients, is derived based on the FV model and prior physical knowledge of PA LMM.

# 5 PROPOSED WHFB MODEL FOR THE COMPENSATION OF LMM PA

This chapter is motivated by several practical cases in which PAs are expected to suffer from LMM, as discussed in subsection 1.2.2, and also by the fact that conventional DPD polynomial models, such as the GMP model, are satisfactory when the PA is perfectly matched (CRIPPS, 2006), but have their performance degraded under LMM conditions, even after re-estimating the model coefficients, as studied in (SANCHEZ-PEREZ et al., 2011), (ZENTENO; ISAKSSON; HANDEL, 2015), (LANDIN; BENGTSSON; ISAKSSON, 2009). The contribution of this chapter is to propose the application of the physically-inspired WHFB model (MKADEM et al., 2014) as a simplified DPD behavioural model for PAs under LMM. To the authors' knowledge, such approach has not been applied in the LMM scenario. Besides, the implementation of this approach is less complex than those of the DPD models presented in section 2.3, as discussed therein and in section 5.2. The choice of the WHFB model is justified in this chapter using the Volterra series analysis, from which the DPD closed-form expression is derived. The ability of the proposed model to compensate for PA LMM is experimentally verified in chapter 6. This chapter follows the ILA DPD topology from Figure 2.12, assuming that the components in the feedback path are ideal, including the directional coupler, IQ down-converter and a pair of analog-to-digital converters (ADC).

The main results from chapters 5 and 6 were published in (HEMSI; PANAZIO, 2021).

## 5.1 JUSTIFICATION OF THE WHFB DPD MODEL FOR LMM PA

The Volterra framework (SCHETZEN, 2006) plays an important role in DPD of high-power and wideband PAs subject to NL behaviour and memory effects. In practice, the FV model is avoided, due to the exponential growth on its number of coefficients, as the model parameters increase; thus, pruned-Volterra polynomial models are preferred. DPD techniques for amplifiers linearisation are widely covered in the literature, although the major part of the works, e.g. for general applications (KIM; KONSTANTINOU, 2001), (MORGAN et al., 2006), (DING et al., 2004), (LEE; FIGUEIREDO, 2006) and for mobile terminals (PRESTI; KIMBALL; ASBECK, 2012), (NORRIS et al., 2007), assumes that the PA is perfectly terminated, i.e., the load absorbs all the incoming power, and zero reflections are returned to the amplifier. In this context, the MP and GMP models are

commonly used as DPD inverse models. As shown in (ZENTENO; ISAKSSON; HANDEL, 2015), (LANDIN; BENGTSSON; ISAKSSON, 2009), (VUOLEVI; RAHKONEN, 2003), the MP and GMP DPD models have their performance degraded when the PA suffers from LMM, due to frequency dependencies at the baseband, fundamental and harmonic frequency bands (PEDRO; MAAS, 2005).

This section justifies the use of the physically-inspired, Volterra-based WHFB model (MKADEM et al., 2014), as a simplified DPD behavioural model for PA LMM, seeking for a trade-off between model accuracy and complexity.

Figure 5.1 shows a simplified equivalent circuit of a typical Class B or Class AB PA circuit with a field-effect transistor (FET). In the figure, the impact of PA LMM is considered by applying the Volterra series analysis NL currents method (MAAS, 2003), between $x_{RF}(t)$ to $y_{RF}(t)$, the RF PA input and output voltage signals, respectively. In order to obtain simplified expressions for the input-to-output NL transfer functions (TF), the following assumptions are made:

- The single active device is a FET in single-ended common-source mode,
- The device internal reactances are negligible at the operating band,
- Mildly NL behaviour in absence of hard clipping and cut-off,
- The (input) gate current is linear,
- The PA is stable, i.e., low internal feedback,
- The active device self-heating is not considered.

These assumptions lead to the simplified PA equivalent circuit in Figure 5.1, as in (ZHU; PEDRO; CUNHA, 2007) and (PEDRO; CARVALHO; LAVRADOR, 2003), where the elements in dashed line are neglected. As in (PEDRO; CARVALHO; LAVRADOR, 2003) and (FAGER et al., 2004), this analysis is based on recognizing the (output) drain current $i_{ds}(t)$ as the major contributor to the FET's non-linear behaviour. The current $i_{ds}(t)$ is assumed a memoryless mildly NL function of both the gate and drain voltages, $v_{gs}(t)$ and $v_{ds}(t)$, approximated by a bi-dimensional Taylor series expansion around the quiescent point:

$$\begin{aligned} i_{ds}(t) \simeq\ & I_{ds} + g_m v_{gs}(t) + K_{2gm} v_{gs}^2(t) + K_{3gm} v_{gs}^3(t) \\ & + g_o v_{ds}(t) + K_{2go} v_{ds}^2(t) + K_{3go} v_{ds}^3(t) + K_{2gmgo} v_{gs} v_{ds}(t) \\ & + K_{3gm2go} v_{gs}^2 v_{ds}(t) + K_{3gmgo_2} v_{gs} v_{ds}^2(t) + \ldots, \end{aligned} \tag{5.1}$$

where $I_{ds}$ is the output DC current, the coefficients $g_m$, $K_{2gm}$, $K_{3gm}$, $\cdots$ are non-linear transconductance and $g_o$, $K_{2go}$, $K_{3go}$, $\cdots$, the non-linear output conductances. The

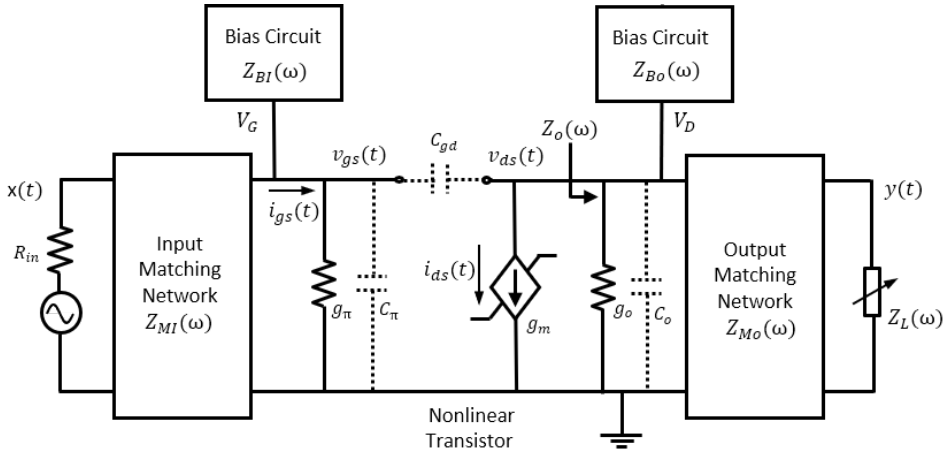



Figure 5.1 – Simplified equivalent circuit of the FET-based PA.

coefficients are calculated by partial derivatives, such as $g_m = \partial i_{ds}(t)/\partial v_{gs}(t)$, $g_o = \partial i_{ds}(t)/\partial v_{ds}(t)$, etc.

The output impedance $Z_o(\omega)$, in the frequency domain, at the transistor (output) drain node, neglecting package parasitics, is the association of the intrinsic output conductance and extrinsic output MN, BN and load impedances:

$$Z_o(\omega) \simeq g_o \parallel Z_{Bo}(\omega) \parallel (Z_{Mo}(\omega) + Z_L(\omega)), \tag{5.2}$$

where $\parallel$ represents the parallel association in this expression. The impact of the output impedance to the PA NL behaviour involves the following feedback (ZHU; PEDRO; CUNHA, 2007):

- The drain (output) voltage $v_{ds}(t)$ is the negative of the drain current $i_{ds}(t)$ times the output impedance, $Z_o(\omega)$,
- In turn, the drain current $i_{ds}(t)$ is a NL function of both the gate and drain voltages, $v_{gs}(t)$ and $v_{ds}(t)$, as per eq. (5.1).

The Volterra series analysis allows the extrinsic PA voltage TFs to be obtained, by accounting for the embedding circuitry in Figure 5.1. In order to obtain simplified expressions for the input-to-output NL TF, the NL current source $i_{ds}(t)$ is linearised and associated in parallel with NL current sources, one for each NL order under analysis (MAAS, 2003). The resulting first-order linear input-to-output TF is given by:

$$H_1(\omega_1) = -M_i(\omega_1) g_m Z_o(\omega_1) M_o(\omega_1), \tag{5.3}$$

where $\omega_1$ is a fundamental band frequency and $M_i(\omega)$ and $M_o(\omega)$ are the voltage TFs of the input and output MNs, respectively. Note that the PA linear voltage TF $H_1(\omega)$ at the

fundamental band depends on the output impedance function $Z_o(\omega)$ that is expected to be flat under nominal conditions. Load mismatch conditions may cause frequency-dependencies in eq. (5.2) and linear distortions on the TF in eq. (5.3).

Recursively, $H_2(\omega_1, \omega_2)$ and $H_3(\omega_1, \omega_2, \omega_3)$, i.e., the second- and third-order nonlinear TFs, respectively, are given by:

$$\begin{aligned} H_2(\omega_1, \omega_2) = -M_o(\omega_1 + \omega_2) Z_o(\omega_1 + \omega_2) \Big\{ & K_{2gm} M_i(\omega_1) M_i(\omega_2) \\ & + \frac{1}{2} K_{2gmgo} [H_1(\omega_1) + H_1(\omega_2)] + K_{2go} H_1(\omega_1) H_1(\omega_2) \Big\} \end{aligned} \tag{5.4}$$

and

$$\begin{aligned} H_3(\omega_1, \omega_2, \omega_3) = -M_o(\omega_1 + \omega_2 + \omega_3) Z_o(\omega_1 + \omega_2 + \omega_3) \Big\{ & K_{3gm} M_i(\omega_1) M_i(\omega_2) M_i(\omega_3) \\ & + \frac{1}{3} K_{3gm2go} [H_1(\omega_1) + H_1(\omega_2) + H_1(\omega_3)] + \frac{1}{3} K_{3gmgo2} [H_1(\omega_1) H_1(\omega_2) \\ & + H_1(\omega_1) H_1(\omega_3) + H_1(\omega_2) H_1(\omega_3)] + K_{3go} H_1(\omega_1) H_1(\omega_2) H_1(\omega_3) \\ & + \frac{1}{3} K_{2gmgo} [H_2(\omega_1, \omega_2) + H_2(\omega_1, \omega_3) + H_2(\omega_2, \omega_3)] \\ & + \frac{2}{3} K_{2go} [H_1(\omega_1) H_2(\omega_2, \omega_3) + H_1(\omega_2) H_2(\omega_1, \omega_3) + H_1(\omega_3) H_2(\omega_1, \omega_2)] \Big\}, \end{aligned} \tag{5.5}$$

where $\omega_1$, $\omega_2$ and $\omega_3$ are fundamental band frequencies.

Now, the NL PA RF output signal $y_{RF}(t)$ is approximated by the truncated Volterra series:

$$y_{RF}(t) = \sum_{k=1}^{K} y_{RF_k}(t), \tag{5.6}$$

assuming that the DC term is zero and $K$ is the truncation NL order. Each $k^{th}$-order output term, $y_{RF_k}(t)$, is obtained by the $k$-dimensional convolution integral:

$$y_{RF_k}(t) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h_k(\tau_{1:k}) \prod_{i=1}^{k} x_{RF}(t - \tau_i) \; d\tau_1 \dots d\tau_k, \tag{5.7}$$

where $x_{RF}(t)$ is the RF input signal and $h_k(\tau_{1:k})$, denoting $(\tau_{1:k}) = (\tau_1, \dots, \tau_k)$, is the $k$-dimensional Volterra kernel. In the frequency domain, the $k^{th}$-order output term $Y_{RF_k}(\omega)$ is given by:

$$Y_{RF_k}(\omega) = H_k(\omega_{1:k}) \prod_{i=1}^{k} X_{RF}(\omega_i), \tag{5.8}$$

with $(\omega_{1:k}) = (\omega_1, \dots, \omega_k)$. The input $X_{RF}(\omega)$ is the Fourier transform of $x_{RF}(t)$, i.e., $X_{RF}(\omega) = \int_{-\infty}^{\infty} x_{RF}(\tau) e^{-j\omega\tau} d\tau$, and $H_k(\omega_{1:k})$ is the $k$-dimensional Fourier transform of

$h_k(t_{1:k})$, defined as follows:

$$H_k(\omega_{1:k}) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} h_k(\tau_{1:k}) e^{-j(\omega_1\tau_1+\dots+\omega_k\tau_k)} \; d\tau_1 \dots d\tau_k. \tag{5.9}$$

The NL functions $Y_{RF_k}(\omega)$ in eq. (5.8) contain several frequency components at baseband, fundamental and harmonic bands, but only the output signal components at the fundamental zone (PEDRO; MAAS, 2005) are of interest for behavioural models and DPD, thus are considered hereafter. Moreover, these models are typically expressed in the LPE notation, by using the complex envelopes of the RF signals. For example, the complex envelope $\tilde{x}(t)$ of the RF input signal is extracted from $x_{RF}(t) = \Re\{\tilde{x}(t)e^{j\omega_c t}\}$, where $\omega_c$ is the carrier frequency. Retaining only the components at the fundamental zone, the LPE of the output signal in eq. (5.6), $\tilde{y}(t)$, is composed of:

- The linear term $\tilde{Y}_1(\omega)$:

$$\tilde{Y}_1(\omega) = \tilde{H}_1(\omega)\tilde{X}(\omega), \tag{5.10}$$

  where the LPE kernel $\tilde{H}_1(\omega)$ is a function of the non-ideal frequency responses of the MN and the load impedance $Z_L(\omega)$, as per eqs. (5.2) and (5.3). Note that $\tilde{Y}_1(\omega)$ represents the fundamental zone linear parcel, shifted in frequency around DC. In the time-domain, this corresponds to the linear convolution:

$$\tilde{y}_1(t) = \int_{-\infty}^{\infty} \tilde{h}_1(\tau_1)\tilde{x}(t-\tau_1)d\tau_1, \tag{5.11}$$

  where $\tilde{y}_1(t)$ and $\tilde{h}_1(t)$ are the LPE of $y_{RF_1}(t)$ and $h_1(t)$. Next, the $\tilde{y}_1(t)$ expression is converted to discrete time, assuming a finite memory length $M$, resulting in linear memory effects, as follows:

$$\tilde{y}_1(n) = \sum_{m=0}^{M} \alpha_{1,m}\tilde{x}\left(n-m\right), \tag{5.12}$$

  where $\alpha_{1,m}$ are the linear coefficients.

- Odd-order intermodulation (IM) distortions produced by $\tilde{Y}_k(\omega)$, with $k$ odd, given by:

$$\tilde{Y}_k(\omega) = \tilde{H}_k(\omega_{1:k}) \prod_{i=1}^{\frac{(k+1)}{2}} \tilde{X}(\omega_i) \prod_{i=\frac{(k+3)}{2}}^{k} \tilde{X}^*(\omega_i), \tag{5.13}$$

  where $\tilde{H}_k(\omega_{1:k})$ is proportional to the frequency responses of the MN and the load impedance $Z_L(\omega)$ at the IM frequencies, as per eqs. (5.2) and (5.5). In the time-domain, this corresponds to the $k$-dimensional convolution:

$$\tilde{y}_k(t) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \tilde{h}_k(\tau_{1:k}) \prod_{i=1}^{\frac{(k+1)}{2}} \tilde{x}(t-\tau_i) \prod_{i=\frac{(k+3)}{2}}^{k} \tilde{x}^*(t-\tau_i) \; d\tau_1 \dots d\tau_k. \tag{5.14}$$

Next, converting to discrete time, with a finite memory length $M$, and retaining only the diagonal terms (KIM; KONSTANTINOU, 2001), i.e., sample delays $m_1 = \cdots = m_k = m$, kernels $\tilde{h}_k(\tau_{1:k})$ reduce to $\alpha_{m,k}$ and the odd-order distortion leads to:

$$\tilde{y}_k(n) = \sum_{m=0}^{M} \alpha_{m,k} \tilde{x}\left(n-m\right) \left|\tilde{x}\left(n-m\right)\right|^{k-1}, \tag{5.15}$$

where $\alpha_{m,k}$ are the model coefficients. Note that eq. (5.15) produces the MP model, in which all kernels are reduced to dimensionality one. By allowing a sample delay between the signal and its powered envelope, the GMP model is obtained: $\sum_{m=0}^{M}\sum_{l=-L}^{L} \alpha_{m,l,k} \tilde{x}\left(n-m\right) \left|\tilde{x}\left(n-m-l\right)\right|^{k-1}$, in which the kernels have one extra dimension and $\alpha_{m,l,k}$ are the corresponding GMP model coefficients.

- Baseband IM distortions produced by $\tilde{Y}_k(\omega)$, with $k$ even, converted to the fundamental zone by mixing with odd-order terms, as per eq. (5.5). The even-order distortion components are given by:

$$\tilde{Y}_k(\omega) = \tilde{H}_k(\omega_{1:k}) \prod_{i=1}^{\frac{k}{2}} \tilde{X}(\omega_i) \prod_{i=\frac{k}{2}+1}^{k} \tilde{X}^*(\omega_i), \tag{5.16}$$

where $\tilde{H}_k(\omega_{1:k})$ is proportional to the baseband output impedance, such as $Z_o(\omega_1 - \omega_2)$, as in eq. (5.4). In the time-domain, eq. (5.16) corresponds to:

$$\tilde{y}_k(t) = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \tilde{h}_k(\tau_{1:k}) \prod_{i=1}^{\frac{k}{2}} \tilde{x}(t-\tau_i) \prod_{i=\frac{k}{2}+1}^{k} \tilde{x}^*(t-\tau_i) d\tau_1 \ldots d\tau_k. \tag{5.17}$$

Next, converting to discrete time, with a finite memory length $L$, and retaining only the diagonal terms, i.e., $l_1 = \cdots = l_k = l$, the baseband distortion leads to:

$$\tilde{y}_k(n) = \sum_{l=0}^{L} \alpha_{l,k} \left|\tilde{x}\left(n-l\right)\right|^{k}, \tag{5.18}$$

where $\alpha_{l,k}$ are the model coefficients. Finally, the remix of baseband terms with the odd-order terms is modelled by enlarging the previous structure in eq. (5.15) with an additional dimension with different sample delays, as follows:

$$\tilde{y}_{k+q}(n) = \sum_{m=0}^{M} \sum_{\substack{q=0 \\ q\ even}}^{Q} \sum_{l=0}^{L} \alpha_{m,q,l,k} \tilde{x}\left(n-m\right) \left|\tilde{x}\left(n-m\right)\right|^{k-1} \left|\tilde{x}\left(n-m-l\right)\right|^{q}, \tag{5.19}$$

where $\alpha_{m,q,l,k}$ are the corresponding model coefficients, with $k$ odd.

The resulting polynomial justifies the choice of the WHFB model from eq. (2.43) for PA LMM. Comparing to the MP and GMP models, the added terms enhance the PA modelling capabilities, by capturing NL interactions among the output signal reflections and the non-linearity, caused by in-band frequency dependencies.

### 5.1.1 Measurement-based Model Extraction

A measurement-based model extraction approach for the WHFB model is briefly discussed here. The objective is to extract the frequency response of the feedback LTI filter, which models the impact of LMM on the PA model. The simplified PA equivalent circuit shown in Figure 5.1 is converted into the physically-inspired bandpass feedback topology in Figure 5.2, proposed in (PEDRO; CARVALHO; LAVRADOR, 2003), (ZHU; PEDRO; CUNHA, 2007). In this block diagram, the PA transistor is represented by a (memoryless) SNL block and the MN and BN in-band frequency responses, by three LTI filters. The input and output linear filters $M_i(\omega)$ and $M_o(\omega)$, in series with the transistor, correspond to linear memory effects that model non-constant frequency responses of the MN and BN at the fundamental band. The NL memory effects, mainly caused by load mismatch (LANDIN; BENGTSSON; ISAKSSON, 2009), (ZHU; PEDRO; CUNHA, 2007), i.e., the reflection and re-ingress of delayed output samples into the SNL block input (PEDRO; MAAS, 2005), (VUOLEVI; RAHKONEN, 2003), (CARVALHO; PEDRO, 2002), are modelled by the LTI filter in the feedback loop, $M_f(\omega)$. In order to simplify the model, the next step is to replace the recursive WHFB model in Figure 5.2 by the truncated feed-forward structure in Figure 5.3, (MAZIERE et al., 2005), (BOLSTAD et al., 2011), (CUNHA et al., 2007), (CUNHA; PEDRO; CABRAL, 2007) . This simplification assumes that each signal sample at the output of the NL block only enters the feedback loop once, as explained in (ZHU; PEDRO; CUNHA, 2007).

In the simplified feed-forward model, $M_i(\omega)$ and $M_o(\omega)$ are bandpass filters at the fundamental band, whereas the frequency response of the $M_f(\omega)$ filter accounts for LMM effects, such as signal reflections, at baseband, fundamental and harmonic frequency bands. The approximate output signal from the feed-forward model, in the frequency domain, is given by:

$$\begin{aligned} Y(\omega) &= M_o(\omega)\ f\Big\{M_i(\omega)X(\omega) + M_f(\omega)Y(\omega)M_o^{-1}(\omega)\Big\} \\ &\simeq M_o(\omega)\, f\{M_i(\omega)X(\omega)\} + M_o(\omega)\, f\Big\{M_i(\omega)X(\omega) + M_f(\omega)\, f\{M_i(\omega)X(\omega)\}\Big\}, \end{aligned} \tag{5.20}$$

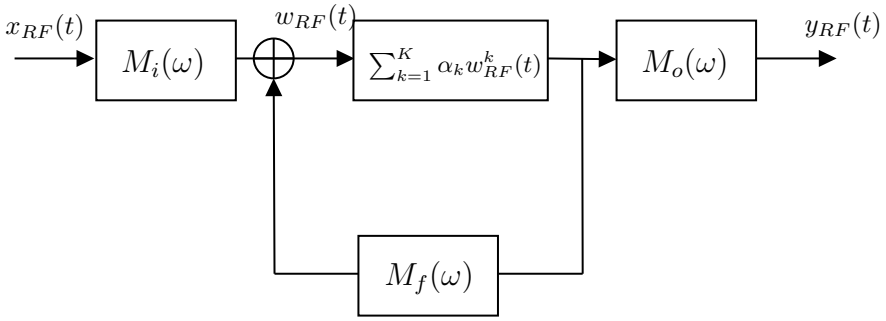



Figure 5.2 – Block diagram of the equivalent WHFB model.

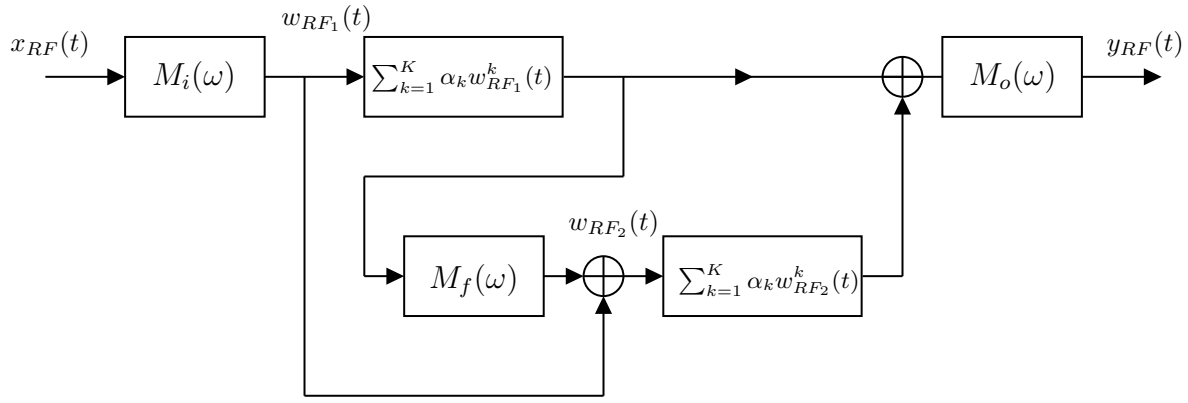



Figure 5.3 – Block diagram of the approximate feed-forward model.

where $f\{\dot{}\}$ is the transistor (memoryless) SNL function. $X(\omega)$ and $Y(\omega)$ are the PA input and output signals, respectively, in the frequency-domain.

Next, the objective is to extract a model for the PA under a given LMM condition, i.e., to extract the frequency response of the feedback LTI filter. This way, it is assumed that the transistor NL function is known, including the model coefficients $\boldsymbol{\alpha}$, and that the frequency responses of the filters $M_i(\omega)$ and $M_o(\omega)$, with source and load terminations, are also known, either by measurement or simulation. In this case, the measurement-based procedure for finding the frequency response of the feedback LTI filter, $M_f(\omega)$, and, thus, the impact of LMM on the PA, is to apply a two-tone probing (excitation) signal at the PA input and measure the output distortion levels at some specific frequency points in the fundamental band, such as at third-order inter-modulation (IM3) points. The measured distortion levels are, then, mapped into the parcels (or contributions) that are expected from the remix of distortion terms originated at baseband and harmonics with the input signal, as illustrated in Table 5.1. The table presents some of these distortion terms that remix with the input signal, due to LMM, and sum up at IM3 frequencies. The two-tone input signal has center frequency $\omega_c$, with tone spacing $\Delta\omega$ and tone amplitude $A$. By running an automated program that sweeps $\omega_c$ and $\Delta\omega$ and maps the measured distortion levels to the expected IM3 levels from the table, the frequency response of $M_f(\omega)$ can be approximately extracted. Note that this frequency response corresponds to the reflection responses of the BN and MN at baseband and $2^{nd}$-harmonic bands.

Table 5.1 – IM3 caused by NL remix of distortion terms.

| $\omega_1$ | $\omega_2$ | **Non-linear order** | **IM3 distortion** ($\omega_c \pm 3\Delta\omega/2$) | **Level** |
|---|---|---|---|---|
| $\omega_c \pm \Delta\omega/2$ | $\omega_c \mp \Delta\omega/2$ | $3^{\text{rd}}$ | $2\omega_1 - \omega_2$ | $\frac{3}{4}A^3\alpha_3$ |
| $\pm\Delta\omega$ | $\omega_c \pm \Delta\omega/2$ | $2^{\text{nd}}$ | $\omega_1 + \omega_2$ | $A^3 M_f(\Delta\omega)\alpha_2^2$ |
| $\pm\Delta\omega$ | $\omega_c \mp \Delta\omega/2$ | $3^{\text{rd}}$ | $2\omega_1 + \omega_2$ | $\frac{3}{4}A^5 M_f(\Delta\omega)\alpha_2\alpha_3$ |
| $2\omega_c \pm \Delta\omega$ | $\omega_c \mp \Delta\omega/2$ | $2^{\text{nd}}$ | $\omega_1 - \omega_2$ | $\frac{1}{2}A^3 M_f(2\omega)\alpha_2^2$ |

## 5.2 PROPOSED WHFB DPD MODEL FOR LMM PA

In the previous section, the WHFB polynomial model is proposed as a simplified behavioural model for PA under LMM. As shown in (SCHETZEN, 2006), (LJUNG; FORSSELL, 1999), the DPD inverse model can also be implemented using the same structure. Therefore, the proposed DPD model is expressed as follows:

$$\begin{aligned}\hat{x}(n) = &\sum_{m=0}^{M_1}\sum_{\substack{k=1\\ k\ odd}}^{K_1}\beta_{m,k}^{(1)}\tilde{u}(n-m)\left|\tilde{u}(n-m)\right|^{k-1} + \sum_{m=0}^{M_2}\sum_{l=1}^{L_2}\sum_{\substack{k=1\\ k\ odd}}^{K_2}\sum_{\substack{q=2\\ q\ even}}^{Q_2}\beta_{m,l,k,q}^{(2)}\tilde{u}(n-m)\\ &\left|\tilde{u}(n-m)\right|^{k-1}\left|\tilde{u}(n-m-l)\right|^{q} + \sum_{m=0}^{M_3}\sum_{l=1}^{L_3}\sum_{\substack{k=0\\ k\ even}}^{K_3}\sum_{\substack{q=1\\ q\ odd}}^{Q_3}\beta_{m,l,k,q}^{(3)}\tilde{u}(n-m-l)\\ &\left|\tilde{u}(n-m-l)\right|^{q-1}\left|\tilde{u}(n-m)\right|^{k},\end{aligned} \tag{5.21}$$

where $\tilde{u}(n)$ and $\hat{x}(n)$ are, respectively, the inverse model normalized input and the output signals. The parameters $M_1$ and $K_1$ are, respectively, the memory depth and power order on the first summation. Likewise, $M_2$, $L_2$, $K_2$ and $Q_2$ are, respectively, the memory depth, envelope time-shift, aligned envelope power order and time-shifted envelope power order on the second summation. Finally, $M_3$, $L_3$, $K_3$ and $Q_3$ are, respectively, the memory depth, envelope time-shift, time-shifted envelope power order and aligned envelope power order on the third summation. This way, the model has NL order given by $K + Q = max\{K_1, K_2 + Q_2, K_3 + Q_3\}$ and memory length $M + L = max\{M_1, M_2 + L_2, M_3 + L_3\}$. Finally, in the appendix A, the WHFB model is extended to an unified approach that simultaneously compensates for PA LMM and IQM imbalance.

Advantages of the proposed WHFB DPD model, comparing to other Volterra-based models:

- The proposed structure includes cross-terms not accounted for with the GMP model, but is less complex than the triangular FV models, thus becoming computationally more affordable.
- All model coefficients in eq. (5.21) can be estimated by OLS, since the expression is linear in the coefficients.
- The proposed model is only based on four parameters, $M$, $L$, $K$ and $Q$, assuming that all summations in eq. (5.21) have the same parameters.
- The proposed DPD structure reduces to the GMP model when the additional coefficients become negligible.

Now, we consider the PA LMM DPD models, from the literature, presented in section 2.3. Firstly, the WHFB model is compared to load-dependent behavioural models parametrized by the load reflection coefficient in (CAI et al., 2015), (DHAR et al., 2018), (DUNN et al., 2017). The proposed DPD approach shows the following advantages:

- No special training procedures are required, i.e., model estimation is performed during the TX transmission, using the actual modulated signals, without any interruptions.
- The reflection coefficient is implicitly considered in the DPD coefficients, thus its measurement is not required.
- The proposed model is able to compensate for LMM at the envelope and harmonic frequency bands as well.
- An extended version of the proposed model can be applied to the joint compensation of IQM imbalance and PA LMM.

Next, comparing the WHFB model to the DIDO models from the literature (CAI; GONÇALVES; PEDRO, 2017), (ZARGAR; BANAI; PEDRO, 2015b):

- The reflected signals are implicitly considered in the DPD structure, not requiring to be individually measured.
- The proposed model shows lower complexity, both in terms of estimation and running costs.

Given that operational/ environmental conditions, such as LMM, might be intermittent, Figure 5.4 proposes a scheme in which the actual DPD model can be switched between a baseline and an advanced mode, respectively the MP/GMP and the WHFB models. While the baseline mode is running in the predistorter, its NMSE is being computed and compared against a target. Whenever the baseline mode is unable to reach the pre-defined NMSE level, due to operational conditions, such as LMM, supply voltage drift, temperature variations, etc., firstly the DPD training block tries to re-estimate the model coefficients, keeping the same parameters. Secondly, the training block sets new model parameters, re-estimate and re-compute the NMSE, until no significant improvements are obtained using the baseline model (LANDIN; ISAKSSON; HANDEL, 2008). Finally, when required, the algorithm switches to the advanced mode and adjusts the WHFB model parameters. Once the WHFB model is running, the MP/GMP model performance is periodically re-evaluated, allowing the DPD scheme to switch back to its less computationally intensive mode.

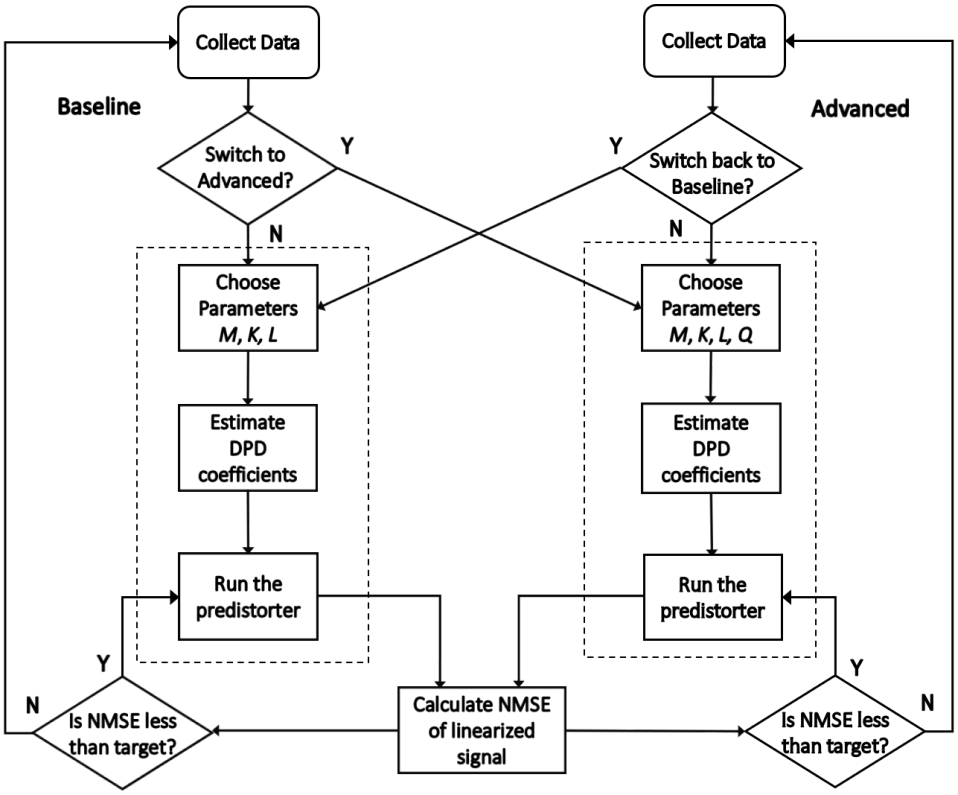



Figure 5.4 – Block diagram of the switching DPD model technique.

### 5.2.1 WHFB Least Squares Estimation

As described in section 2.6, the WHFB model coefficients are estimated by OLS regression from measured input/output data. The WHFB PA model in eq. (5.21) consists of three polynomial functions, thus the coefficient vector to be estimated is built by concatenation, as follows:

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}^{(1)} \\ \boldsymbol{\beta}^{(2)} \\ \boldsymbol{\beta}^{(3)} \end{bmatrix}, \tag{5.22}$$

where:

$$\boldsymbol{\beta}^{(1)} = \begin{bmatrix} \beta_{01}^{(1)} & \dots & \beta_{M_1K_1}^{(1)} \end{bmatrix}^T, \tag{5.23}$$

$$\boldsymbol{\beta}^{(2)} = \begin{bmatrix} \beta_{0110}^{(2)} & \dots & \beta_{M_2L_2K_2Q_2}^{(2)} \end{bmatrix}^T \tag{5.24}$$

and

$$\boldsymbol{\beta}^{(3)} = \begin{bmatrix} \beta_{0101}^{(3)} & \dots & \beta_{M_3L_3K_3Q_3}^{(3)} \end{bmatrix}^T. \tag{5.25}$$

In order to estimate the DPD coefficients, we collect $N$ samples of the PA input and normalized output signals. Accordingly, the inverse model regression matrix $\boldsymbol{\Phi}(\tilde{\boldsymbol{u}})$, computed from the PA normalized output samples, is given as:

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\Phi}^{(1)} & \boldsymbol{\Phi}^{(2)} & \boldsymbol{\Phi}^{(3)} \end{bmatrix}, \tag{5.26}$$

where $\boldsymbol{\Phi}^{(i)}$ is the sub-matrix of regressors produced using the $i^{th}$ polynomial term in eq. (5.21), with $i = 1, 2, 3$, and given by:

$$\boldsymbol{\Phi}^{(i)} = \begin{bmatrix} \boldsymbol{\Phi}^{(i)}(0) \\ \vdots \\ \boldsymbol{\Phi}^{(i)}(N-1) \end{bmatrix}, \tag{5.27}$$

where:

$$\boldsymbol{\Phi}^{(1)}(n) = \begin{bmatrix} \tilde{u}(n) \\ \vdots \\ \tilde{u}(n)\,|\tilde{u}(n)|^{K_1-1} \\ \vdots \\ \tilde{u}(n-M_1) \\ \vdots \\ \tilde{u}(n-M_1)\,|\tilde{u}(n-M_1)|^{K_1-1} \end{bmatrix}^T, \tag{5.28}$$

$$\boldsymbol{\Phi}^{(2)}(n) = \begin{bmatrix} \tilde{u}(n)\,|\tilde{u}(n-1)|^{2} \\ \vdots \\ \tilde{u}(n)\,|\tilde{u}(n)|^{K_2-1}\,|\tilde{u}(n-L_2)|^{Q_2} \\ \vdots \\ \tilde{u}(n-M_2)\,|\tilde{u}(n-M_2-1)|^{2} \\ \vdots \\ \tilde{u}(n-M_2)\,|\tilde{u}(n-M_2)|^{K_2-1}\,|\tilde{u}(n-M_2-L_2)|^{Q_2} \end{bmatrix}^T \tag{5.29}$$

and

$$\boldsymbol{\Phi}^{(3)}(n) = \begin{bmatrix} \tilde{u}(n-1)\,|\tilde{u}(n)| \\ \vdots \\ \tilde{u}(n-L_3)\,|\tilde{u}(n-L_3)|^{K_3-1}\,|\tilde{u}(n)|^{Q_3} \\ \vdots \\ \tilde{u}(n-M_3-1)\,|\tilde{u}(n-M_3)| \\ \vdots \\ \tilde{u}(n-M_3-L_3)\,|\tilde{u}(n-M_3-L_3)|^{K_3-1}\,|\tilde{u}(n-M_3)|^{Q_3} \end{bmatrix}^T . \tag{5.30}$$

Finally, the DPD training reference output signal is given by:

$$\tilde{\boldsymbol{x}} = \begin{bmatrix} \tilde{x}(0) \\ \vdots \\ \tilde{x}(N-1) \end{bmatrix}, \tag{5.31}$$

whose elements are delayed samples of $\tilde{\boldsymbol{x}}$. Given that the model is linear in the parameters, OLS minimizes the corresponding error between the modelled and the measured output signals, leading to the following estimated coefficients vector:

$$\hat{\boldsymbol{\beta}}_{OLS} = (\boldsymbol{\Phi}^H\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^H\tilde{\boldsymbol{x}}, \tag{5.32}$$

where $\boldsymbol{\Phi}^H\boldsymbol{\Phi}$ is assumed as non-singular.

## 5.3 CONCLUSIONS

This chapter combined both the behavioural modelling framework and prior physical knowledge about the PA LMM condition to derive a simplified DPD model to compensate for PA LMM. Besides, the model is linear in the coefficients, which is an important aspect for DPD. In the next chapter, the large number of coefficients of the WHFB polynomial model is reduced by applying parsimonious, sparse estimation techniques (DAVENPORT et al., 2012). Additionally, it is experimentally verified the ability of the proposed model to compensate for PA LMM.

# 6 SPARSE ESTIMATION TECHNIQUES

The inherently large number of degrees of freedom of the WHFB model motivates, in this chapter, the parsimonious, sparse estimation of the model, using least absolute shrinkage and selection operator (LASSO) and group-wise LASSO, also applied in compressed-sensing. With these techniques, the number of coefficients of the original model is significantly reduced, thus reducing the DPD running complexity. Additionally, this chapter proposes the use of block-oriented LASSO extensions, such as group-LASSO and sparse-group LASSO, for model dimensioning, i.e., for setting parameters values. The LASSO is chosen due to its sparsity-inducing properties, convexity, robustness and strong theoretical guarantees compared to other CS techniques, such as the OMP (ELAD, 2010), (DAVIES; RILLING; BLUMENSATH, 2012). Finally, this chapter also proposes a simplified, approximate technique, in which the most relevant blocks in the model are selected prior to running LASSO, resulting in lower estimation cost. Experimental results demonstrate the ability of the models to adequately linearize PAs subject to LMM.

## 6.1 SPARSE LEAST SQUARES ESTIMATION

From eq. (2.67), the modelled output signal $\hat{\boldsymbol{x}}$ can be expressed as a commonly noisy linear combination of the model regressors, i.e., $\hat{\boldsymbol{x}}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\Phi}\hat{\boldsymbol{\beta}} + \boldsymbol{\epsilon}$, where $\boldsymbol{\Phi} \in \mathbb{C}^{N\times R}$ is the regression matrix, whose columns are polynomial regressors $\{\boldsymbol{\phi}_r\}_{r=1}^{R}$, $N$ is the number of samples and $R$, the number of regressors. In section 2.6, the coefficients vector was found by OLS. In practice, however, as the model dimensions increase, the regression matrix tends to become ill-conditioned, due to the high correlation among regressors in the polynomial model, e.g., $\tilde{u}\,(n-m)$, $\tilde{u}\,(n-m)\,|\tilde{u}\,(n-m)|$, etc., and among data samples, e.g., $\tilde{u}\,(n-m)$, $\tilde{u}\,(n-m-l)$, etc. Thus, the accuracy and stability of the OLS estimation are compromised, as indicated by an exponential growth on the correlation (Gram) matrix condition number, $cond(\boldsymbol{\Phi}^H\boldsymbol{\Phi})$ (RAICH; QIAN; ZHOU, 2004). This issue can be alleviated by reducing the model dimensionality. Sparse estimation seeks an approximation of the original model, by eliminating the less relevant or redundant model regressors, thus reducing the initial model dimensionality (DAVENPORT et al., 2012).

The coefficients vector $\boldsymbol{\beta} \in \mathbb{C}^{R\times 1}$ is said to be $k$-sparse if $\|\boldsymbol{\beta}\|_0 \leq k$, with $k \ll R$, i.e., having at most $k$ nonzero entries. In this definition, $\|\boldsymbol{\beta}\|_0$ is the $\ell_0$ pseudo-norm, given by $card\,(\,supp(\boldsymbol{\beta}))$.

A sparse approximate model assumes that the measured signal $\tilde{\boldsymbol{x}}$ can be well approximated by $\hat{\boldsymbol{x}}$, whose values are generated as linear combination of, at most, $k \ll R$

columns of $\mathbf{\Phi}$, i.e.:

$$\tilde{\boldsymbol{x}} \approx \mathbf{\Phi}\boldsymbol{\beta}^{(k)} + \boldsymbol{\epsilon}_k, \tag{6.1}$$

where $\boldsymbol{\beta}^{(k)}$ belongs to the set of $k$-sparse vectors from $\mathbb{C}^{R\times 1}$, and $\boldsymbol{\epsilon}_k$ is the small error of the $k$-sparse approximation, such that $\|\boldsymbol{\epsilon}_k\|_2^2 \ll \|\tilde{\boldsymbol{x}}\|_2^2$. The support set $supp(\boldsymbol{\beta}^{(k)}) = \{i : |\beta_i| \neq 0, i = 1, \ldots, R\}$ is also referred to as the active set of the sparse model.

In practice, subset selection and the resulting sparse models are important due to the following advantages:

- Find a solution for under-determined systems (section 2.6).
- Avoid the less accurate OLS estimation when the matrix $\mathbf{\Phi}$ becomes ill-conditioned.
- Reduce the required DSP memory and running computational cost.
- Avoid model over-fitting and improve prediction accuracy.
- Avoid multicollinearity, by eliminating regressors that correspond to linear combinations.
- Improve model interpretation, by only retaining a small subset of regressors that are more explanatory.

As introduced in section 2.6, high dimensionality is a major challenge for the OLS estimation of Volterra-based models, leading to the ridge regression estimation. Meanwhile, several cases admit parsimonious, sparse approximations, in which only a few of the original coefficients are non-zero, because either the physical system is sparse or the initial model was over-parametrized. Indeed, usually PA and DPD NL systems admit sparse approximations (KEKATOS; GIANNAKIS, 2011), since the Volterra kernels do not play an equally important role in the models. However, it is very difficult to estimate *a priori* which terms to be kept, since their contribution to the model is inherently associated to physical knowledge of the system, such as dynamic characteristics of the PA to be modelled. Therefore, we resort to the subset selection and sparse techniques described in this section to derive sparse models for DPD systems. As presented in section 2.6, the ridge regression, which is based on the minimization of the $\ell_2$ norm of $\boldsymbol{\beta}$, is not able to induce sparsity. A first approach to explore sparsity among polynomial regressors is to solve, instead, the $\ell_0$ pseudo-norm constrained optimization problem, as follows:

$$\underset{\boldsymbol{\beta}\in\mathbb{C}^{R\times 1}}{\text{minimize}} \quad \frac{1}{2N}\|\tilde{\boldsymbol{x}} - \mathbf{\Phi}\boldsymbol{\beta}\|_2^2 \quad \text{subject to} \quad \|\boldsymbol{\beta}\|_0 \leq t. \tag{6.2}$$

In this case, however, the optimal recovery of the sparse support of $\boldsymbol{\beta}$ is a non-convex and combinatorial problem, that usually involves searching over all the $\binom{R}{k}$ possible support sets, leading to non-deterministic polynomial time (NP-hard) complexity.

For $R$ large, eq. (6.2) is computationally infeasible and most of the attention has been focused on feasible methods that approximate the optimal solution, such as the matching pursuit and LASSO regularized regression approaches. Matching pursuit and orthogonal matching pursuit (OMP) are greedy subset selection methods that iteratively update one entry of $\hat{\boldsymbol{\beta}}$ at a time, until a sparsity constraint is reached, arriving to an approximate solution to the $\ell_0$ norm minimization problem (ELAD, 2010), (DAVIES; RILLING; BLUMENSATH, 2012).

In turn, LASSO penalizes the high dimensionality of the model and induces sparsity, by efficiently and accurately solving the convex relaxation problem that trades-off between a quadratic error measure and the $\ell_1$ norm of coefficients. In this chapter, we pursue parsimonious, sparse approximations of the WHFB model, by applying the LASSO technique to the original regression matrix. LASSO is chosen due to its sparsity-inducing properties, convexity, robustness and strong theoretical guarantees compared to other CS techniques. As discussed in (ELAD, 2010), (DAVIES; RILLING; BLUMENSATH, 2012), the greedy pursuit performance guarantees are typically weaker than LASSO, both in terms of variable selection consistency and accuracy. Also note that CS-based techniques has the additional advantage of reducing the DPD running complexity, whereas techniques that carry out a regression basis change, such as the principal component analysis (PCA) in (LOPEZ–BUENO et al., 2018), (GILABERT et al., 2013) imply an increase in DPD running cost (SCHUARTZ et al., 2019). Examples in the literature that have applied LASSO to reduce the dimensionality of PA and DPD models are (ZENTENO et al., 2015), (KEKATOS; GIANNAKIS, 2011), (WISELL; JALDEN; HANDEL, 2008).

### 6.1.1 LASSO

In the LASSO approach (TIBSHIRANI, 1996), (HASTIE; TIBSHIRANI; WAINWRIGHT, 2015), a convex relaxation of the problem in eq. (6.2) replaces the intractable $\ell_0$ norm constraint by the $\ell_1$ norm. The LASSO batch estimator minimizes the sum of squared errors subject to the sum of the absolute value of the coefficients being less than a constant:

$$\underset{\boldsymbol{\beta}\in\mathbb{C}^{R\times 1}}{\text{minimize}} \quad \frac{1}{2N}\|\tilde{\boldsymbol{x}}-\boldsymbol{\Phi}\boldsymbol{\beta}\|_2^2 \quad \text{subject to} \quad \|\boldsymbol{\beta}\|_1 \leq t \tag{6.3}$$

which can be equivalently expressed in Lagrangian form as:

$$\hat{\boldsymbol{\beta}}^{(\lambda)} = \underset{\boldsymbol{\beta}\in\mathbb{C}^{R\times 1}}{\arg\min} \quad (\frac{1}{2N}\|\tilde{\boldsymbol{x}}-\boldsymbol{\Phi}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1), \tag{6.4}$$

i.e., as an augmented objective function with a weighted sum of constraints. In this equation, $\ell_1$ norm is $\|\boldsymbol{\beta}\|_1 = \sum_{i=1}^{R}|\beta_i| = \sum_{i=1}^{R}\sqrt{\Re\{\beta_i\}^2 + \Im\{\beta_i\}^2}$. Without loss of generality, assume that the matrix $\boldsymbol{\Phi}$ is standardized, such that each of its columns $\boldsymbol{\phi}_j$ is centered, i.e., $\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{\phi}_j)_i = 0$ and normalized, i.e., $\frac{1}{N}\boldsymbol{\phi}_j^H\boldsymbol{\phi}_j = 1$, for all $j = 1, \ldots, R$. Also $\tilde{\boldsymbol{x}}$ is centered, i.e., $\frac{1}{N}\sum_{i=1}^{N}\tilde{x}_i = 0$. These centering conditions allow to omit the intercept term $\beta_0$. The LASSO regression induces sparsity through a $\ell_1$ norm penalty function, in which $\lambda \geq 0$ is the sparsity parameter that trades-off between sparsity and prediction accuracy. The optimization problem in eqs. (6.3) and (6.4) is also known as basis pursuit denoising and combines convexity and sparsity. Solving for $\hat{\boldsymbol{\beta}}^{(\lambda)}$ is a quadratic programming problem with linear inequality constraint and efficient and stable convex optimization techniques can be used.

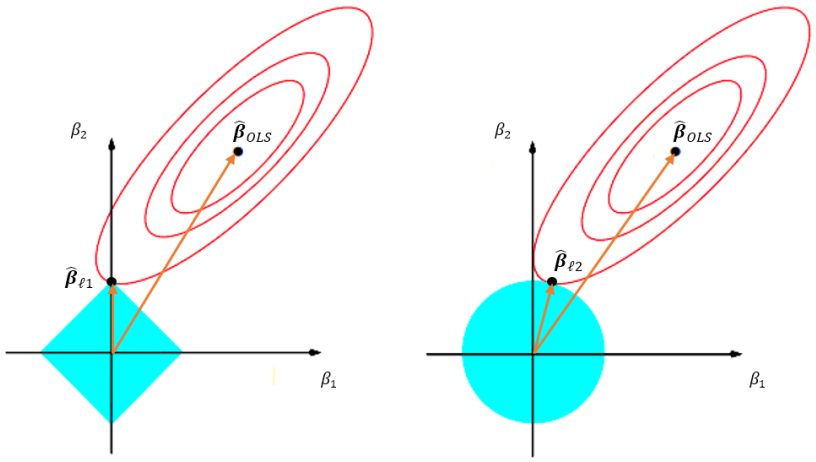



Figure 6.1 – Least squares regularization using $\ell_1$- (left) and $\ell_2$-norm (right).

Figure 6.1, from (FRIEDMAN; HASTIE; TIBSHIRANI, 2010), illustrates the sparsity inducing property of LASSO, comparing it to the ridge regression solution from eq. (2.70). For a two-dimensional solution space, the figure represents the elliptical least squares error contours, centered at the optimal least squares solution $\hat{\boldsymbol{\beta}}_{OLS}$, approaching the $\ell_1$ regularization diamond and $\ell_2$ regularization ball (ridge regression), respectively, given by $|\beta_1|+|\beta_2| \leq t$ and $\beta_1^2 + \beta_2^2 \leq t$. The error contours often touch the $\ell_1$ diamond at a corner, where one (or more) parameters are zero, whereas the $\ell_2$ constraint region is spherical.

One important result is that $\hat{\boldsymbol{\beta}}^{(\lambda)}$ is the problem solution if and only if it satisfies the Karush-Kuhn-Tucker (KKT) conditions (TIBSHIRANI, 1996). This implies that the correlation between each regressor $\boldsymbol{\phi}_i$ to be included in the model and the residual must exceed the threshold value $\lambda$, otherwise $\hat{\beta}_i^{(\lambda)}$ is set to zero, i.e.:

$$\frac{1}{N}|\boldsymbol{\phi}_i^H(\tilde{\boldsymbol{x}} - \boldsymbol{\Phi}\hat{\boldsymbol{\beta}}^{(\lambda)})| > \lambda, \tag{6.5}$$

with $i = 1, \ldots, R$. The idea is that if $\hat{\beta}_i^{(\lambda)}$ cannot move in any direction away from zero without increasing the penalty more than improving the fit, then $\hat{\beta}_i^{(\lambda)} = 0$ is its global minimum. Note that the solution of one $\hat{\beta}_i^{(\lambda)}$ is dependent upon all the other components of $\hat{\boldsymbol{\beta}}^{(\lambda)}$, therefore there is no closed form solution.

### 6.1.2 Group-wise LASSO

The LASSO subset selection is based on the importance of each regressor, individually, to the model. Group sparsity can also be introduced in the model, taking advantage of prior knowledge of the physical system. In this work, block-wise LASSO extensions, such as group-LASSO (KRONVALL et al., 2016), (MEIER; GEER; BUHLMANN, 2008), (YUAN; LIN, 2006) and sparse-group LASSO (SIMON et al., 2013), are applied for selecting the dominant blocks of regressors from the initial model. Consider a linear regression model with $R$ regressors divided into $G$ non-overlapping groups, such that the model is written as, (YUAN; LIN, 2006):

$$\tilde{\boldsymbol{x}} = \sum_{g=1}^{G} \boldsymbol{\Phi}_g \boldsymbol{\beta}_g + \boldsymbol{\epsilon}, \tag{6.6}$$

where $\boldsymbol{\Phi}_g$ contains the $g^{th}$ group of regressors from $\boldsymbol{\Phi}$ and $\boldsymbol{\beta}_g$ is the corresponding coefficients vector. Without loss of generality, assume that the sub-matrices $\boldsymbol{\Phi}_g$ are standardized and $\tilde{\boldsymbol{x}}$ is centered. In this case, a single-level selection decides which groups to be chosen, whereas a bi-level selection allows to decide which variables are more relevant within each of the groups.

The following group-wise LASSO techniques are considered:

- To explore the model block-wise sparsity, the group-LASSO criterion divides the $R$ regressors of the model into $G$ disjoint groups (YUAN; LIN, 2006). Group sparsity is achieved by minimizing the following cost function:

$$\frac{1}{2N}\|\tilde{\boldsymbol{x}} - \sum_{g=1}^{G} \boldsymbol{\Phi}_g \boldsymbol{\beta}_g\|_2^2 + \lambda \sum_{g=1}^{G} \sqrt{d_g}\|\boldsymbol{\beta}_g\|_2, \tag{6.7}$$

  where $d_g$ is the length of the corresponding coefficients vector, $\boldsymbol{\beta}_g$, and $\lambda$ is the group-LASSO sparsity parameter.

From KKT, a condition similar to eq. (6.5) is valid for group-LASSO:

$$\frac{1}{N}\|\boldsymbol{\Phi}_g^H(\tilde{\boldsymbol{x}} - \sum_{j=1}^{G}\boldsymbol{\Phi}_j\hat{\boldsymbol{\beta}}_j^{(\lambda)})\|_2 > \lambda\sqrt{d_g}. \tag{6.8}$$

The penalty function is comprised of the $\ell_2$ norm of the groups. The grouping of variables is useful, e.g., in cases when variables are highly correlated with each other. Group LASSO focuses on groups of variables instead of on individual variables. This way, group-LASSO is unable to perform variable selection at the individual level:

  - Group LASSO sets all coefficients within a group either to zero or non-zero values, i.e., is not able to achieve sparsity within the group.
  - The groups need to be predefined, i.e., the regression model does not provide a direct mechanism to obtain the grouping.

- To explore both the model block-wise sparsity and the regressors sparsity within each block, the sparse-group LASSO (SGL) criterion minimizes the following cost function (SIMON et al., 2013):

$$\begin{aligned}\frac{1}{2N}\|\tilde{\boldsymbol{x}} - \sum_{g=1}^{G}\boldsymbol{\Phi}_g\boldsymbol{\beta}_g\|_2^2 + (1-\alpha)\lambda\sum_{g=1}^{G}\sqrt{d_g}\|\boldsymbol{\beta}_g\|_2 \\ + \alpha\lambda\sum_{g=1}^{G}\|\boldsymbol{\beta}_g\|_1,\end{aligned} \tag{6.9}$$

where the sparse group LASSO uses a weight parameter $\alpha \in [0, 1]$ for combining the LASSO and group LASSO penalties. Equivalently, the parameters $\lambda_1 = (1-\alpha)\lambda$ and $\lambda_2 = \alpha\lambda$ could be defined. A similar method is the elastic net regularization, proposed in (ZOU; HASTIE, 2005). In both cases, the $\ell_1/\ell_2$ regularization is desirable in many applications of regression and classification.

### 6.1.3 Coordinate-descent Algorithm

The LASSO objective function in eq. (6.3) can be split into a differentiable squared error function $\frac{1}{2N}\|\tilde{\boldsymbol{x}} - \boldsymbol{\Phi}\boldsymbol{\beta}\|_2^2$ and a non-differentiable regularization term, $\lambda\|\boldsymbol{\beta}\|_1$, which is strictly convex in each coordinate. Thus, the LASSO objective function can be minimized coordinate-wise, converging to the global minimum, under relatively mild conditions. The coordinate-descent (CD) algorithm, represented in Figure 6.2, iteratively cycle through the regressors of $\boldsymbol{\Phi}$, while minimizing the objective function with respect to the coefficient $\beta_i$, all other coefficients are holding fixed at their most recent values. Let $\boldsymbol{\phi}_i$ denote the $i^{th}$ column of $\boldsymbol{\Phi}$ and $\boldsymbol{\Phi}_{-i}$ denote all the remaining columns, except $\boldsymbol{\phi}_i$. Then the problem is expressed as:

$$\hat{\beta}_i^{(\lambda)} = \arg\min_{\beta_i \in \mathbb{C}} \quad (\frac{1}{2N}\|\boldsymbol{r}_i - \boldsymbol{\phi}_i\beta_i\|_2^2 + \lambda|\beta_i| + \lambda\|\boldsymbol{\beta}_{-i}\|_1), \tag{6.10}$$

where $\boldsymbol{r}_i = \tilde{\boldsymbol{x}} - \boldsymbol{\Phi}_{-i}\boldsymbol{\beta}_{-i}$ is the residual between the response $\tilde{\boldsymbol{x}}$ and the current fit using fixed $\boldsymbol{\beta}_{-i}$ and $\boldsymbol{\Phi}_{-i}$. The solution is found by computing the sub-differential and equating to zero, leading to the soft-threshold function, as follows:

$$\hat{\beta}_i^{(\lambda)} = sgn(\hat{\beta}_i^{OLS})\Big(|\hat{\beta}_i^{OLS}| - \frac{\lambda}{2}\Big)_+, \tag{6.11}$$

where $\hat{\beta}_i^{OLS} = \frac{1}{2N}\boldsymbol{r}_i^H\boldsymbol{\phi}_i$, $sgn$ is the complex signum function and $\Big(|\hat{\beta}_i^{OLS}| - \frac{\lambda}{2}\Big)_+$ corresponds to $max(|\hat{\beta}_i^{OLS}| - \frac{\lambda}{2}, 0)$.

The algorithm applies the soft-threshold update to each element of $\boldsymbol{\beta}$ at a time, then updating the residual and the OLS estimates, until converge is achieved.

Coordinate-descent algorithm

**Input** $\boldsymbol{\Phi}$, $\tilde{\boldsymbol{x}}$, $\lambda$
**Output** $\hat{\boldsymbol{\beta}}$

Initialize $\boldsymbol{\beta} = \mathbf{0}$
**repeat**
  **for** $i = 1, \ldots, R$ **do**
    Compute the residual: $\boldsymbol{r}_i = \tilde{\boldsymbol{x}} - \boldsymbol{\Phi}_{-i}\boldsymbol{\beta}_{-i}$
    Compute the OLS coefficient: $\beta_i^{OLS} = \frac{1}{2N}\boldsymbol{r}_i^H\boldsymbol{\phi}_i$
    Update $\beta_i^{(\lambda)} = sgn(\beta_i^{OLS})\Big(|\beta_i^{OLS}| - \frac{\lambda}{2}\Big)_+$
  **end for**
**until** convergence
$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$
**return** $\hat{\boldsymbol{\beta}}$

Figure 6.2 – Coordinate-descent algorithm.

Similarly to the LASSO coordinate-descent algorithm, the group coordinate-descent algorithm optimizes the objective function with respect to a single group at a time, and cycles through the groups until convergence:

$$\hat{\boldsymbol{\beta}}_g^{(\lambda)} = \arg\min_{\boldsymbol{\beta}_g \in \mathbb{C}^{d_g \times 1}} \quad (\frac{1}{2N}\|\boldsymbol{r}_g - \boldsymbol{\Phi}_g\boldsymbol{\beta}_g\|_2^2 + \lambda\sqrt{d_g}\|\boldsymbol{\beta}_g\|_2 + \lambda\sum_{\substack{k=1 \\ k\neq g}}^{G}\sqrt{d_k}\|\boldsymbol{\beta}_k\|_2), \tag{6.12}$$

where $\boldsymbol{r}_g = \tilde{\boldsymbol{x}} - \sum_{k=1,k\neq g}^{G}\boldsymbol{\Phi}_k\boldsymbol{\beta}_k$ is the residual between the response $\tilde{\boldsymbol{x}}$ and the current fit using $\sum_{k=1,k\neq g}^{G}\boldsymbol{\Phi}_k\boldsymbol{\beta}_k$.

The solution is given by the multivariate soft-threshold function, as follows:

$$\hat{\boldsymbol{\beta}}_g^{(\lambda)} = \frac{\hat{\boldsymbol{\beta}}_g^{OLS}}{\|\hat{\boldsymbol{\beta}}_g^{OLS}\|_2}\Big(\|\hat{\boldsymbol{\beta}}_g^{OLS}\|_2 - \frac{\lambda\sqrt{d_g}}{2}\Big)_+, \tag{6.13}$$

where $\Big(\|\hat{\boldsymbol{\beta}}_g^{OLS}\|_2 - \frac{\lambda\sqrt{d_g}}{2}\Big)_+$ corresponds to $max(\|\hat{\boldsymbol{\beta}}_g^{OLS}\|_2 - \frac{\lambda\sqrt{d_g}}{2}, 0)$ and $\hat{\boldsymbol{\beta}}_g^{OLS} = \frac{1}{2N}\boldsymbol{r}_g^H\boldsymbol{\Phi}_g$. The algorithm applies the multivariate soft-threshold update repeatedly, updating the residual and the OLS estimates, until converge is achieved.

### 6.1.4 Choice of the Regularization Parameter

Choosing the parameter $\lambda$ is fundamental to the performance of LASSO, since it controls the shrinkage and variable selection. There are two major approaches for choosing the regularization parameter: the minimization of some information criterion, e.g., AIC, and cross validation, as described next.

- Cross validation: the regularization parameter is chosen by fitting the model to a subset of the data (known as training set) and validating the choice through the remaining of the data (test set). The most common approach uses the K-fold cross validation, in which the data set is partitioned into K sub-sets, one of them is retained as test set and the remaining as training sets (ARLOT; CELISSE, 2010), (HASTIE; TIBSHIRANI; WAINWRIGHT, 2015).

- Information criteria: using the AIC score (AKAIKE, 1974), each candidate solution is evaluated using eq. (3.4), such as, for example, in (TOSINA et al., 2015) and (BANKS; JOYNER, 2017). The AIC is a widely used selection criterion for choosing the best approximating model from several competing models fitted to a dataset. Note that the AIC trades-off between the residual sum-of-squares and the number of estimated coefficients (BANKS; JOYNER, 2017). As the model accuracy increases, the logarithmic term becomes more negative. The best $\lambda$ value, denoted by $\lambda_o$, corresponds to the smallest AIC score. Observe that, for sample size $N$ small compared to the number of regressors, it is recommended instead to use the corrected AIC (HURVICH; TSAI, 1989) score.

Finally, the alternative L-curve method traces a parametric curve of the residual error versus the solution norm on a log-log plot. The residual error norm, $\log\Big(\frac{1}{2N}\|\tilde{\boldsymbol{x}} - \boldsymbol{\Phi}\boldsymbol{\beta}\|_2^2\Big)$, is placed on the horizontal axis, whereas the solution norm, $\log\|\boldsymbol{\beta}\|_1$, is on the vertical axis. Their values are computed for predefined values of the regularization parameter $\lambda$. For lower $\lambda$ values, the solution norm significantly decreases as $\lambda$ is increased, while

the residual error is kept almost the same. For higher $\lambda$ values, the solution norm is not significantly reduced as $\lambda$ is increased, but the residual error rapidly increases. At the corner of the curve, the regularization value minimizes the norm of the solution, subject to the smallest possible residual error.

## 6.2 PA/DPD BEHAVIOURAL MODEL SIZING

Besides determining the best pruned-Volterra model structure that accurately describes an unknown NL PA or DPD system, another challenging task involving a Volterra-based behavioural model is its model sizing, or dimensioning, i.e., setting suitable values for the parameters of a given model structure.

Typically, model sizing is a very costly multi-dimensional combinatorial problem, performed by exhaustive search (try-and-error) (GOTTHANS; BAUDOIN; MBAYE, 2013), (HANZALA; SHARAWI; HAMMI, 2015). This section briefly presents some of the major state-of-the-art model sizing approaches, classified into heuristic search, orthogonalization and approximate techniques. The aim is to contextualize the CS-based model sizing technique that is proposed in the next sections of this chapter for the WHFB model. The proposed CS-based (sparse) sizing approach takes advantage of the block-wise modular structure proposed for the model to select the most relevant model blocks, using Group-LASSO (KRONVALL et al., 2016), (MEIER; GEER; BUHLMANN, 2008), (YUAN; LIN, 2006) and Sparse-group LASSO (SIMON et al., 2013). This sparse sizing approach is presented and evaluated in details in sections 6.3 and 6.4.

### 6.2.1 Heuristic Search

A variety of heuristic search techniques have been proposed in the literature, in the context of Volterra series model sizing, such as hill-climbing (WANG et al., 2018), particle swarm optimization (PSO)(MALHOTRA; SAPPAL, 2019), (ABDELHAFIZ et al., 2013) and evolutionary algorithms (LI et al., 2016). Although these techniques reduce the computational cost that is required by the exhaustive search approach, the main disadvantage in the context of DPD of heuristic techniques is still their high computational cost. In fact, due to the re-estimation of the residual sum of squares (RSS) of the model for each point in the search grid, these approaches become very costly and suffer from scalability issues (WANG et al., 2016).

### 6.2.2 Orthogonalization

Another approach for model sizing is based on an orthogonal regression matrix, which implies that the model estimation is separable (and also more robust). Using an orthogonal regression matrix, it is possible to separate the effect (or contribution) of

each regressor/ block of regressors to the performance of the model, thus, gradually add regressors/ blocks and incrementally run the least squares estimation of the model coefficients until arriving to the most desirable model sizing. In some particular cases, (QUINDROIT et al., 2013), (YANG et al., 2013), (MKADEM; WU; BOUMAIZA, 2012), (SCHUMACHER; LIMA; OLIVEIRA, 2015), the models are developed in terms of an orthogonal set of polynomials, but this applies only to certain probability distributions of the input signal and is not assumed in the thesis. A more general approach is to have a closed-form model expression that is not based on orthogonal polynomials and use the QR or SVD decompositions to orthogonalize the original regression matrix, eliminating the correlation among its regressors. Firstly, let us consider the QR decomposition of the regression matrix in eq. (2.68). The modified Gram-Schmidt procedure (BJORCK, 1967) decomposes the original $N \times R$ matrix $\mathbf{\Phi}$, as follows:

$$\mathbf{\Phi} = \mathbf{\Theta}\boldsymbol{R}, \tag{6.14}$$

where the $N \times R$ matrix $\mathbf{\Theta}$ has orthonormal columns and the $R \times R$ transition matrix $\boldsymbol{R}$ is upper triangular. The transformed model $R \times 1$ coefficients vector, $\hat{\boldsymbol{\beta}}'_{OLS}$, is related to the original one from eq. (2.69) by: $\hat{\boldsymbol{\beta}}_{OLS} = \boldsymbol{R}\,\hat{\boldsymbol{\beta}}'_{OLS}$. The least squares estimation of the transformed model is separable, since the correlation matrix $\mathbf{\Theta}^H\mathbf{\Theta}$ equals the $R \times R$ identity matrix.

Also note that the QR decomposition allows a more robust LS estimation of the model coefficients, because $\mathbf{\Theta}$ is well-conditioned. In the context of PA modelling and DPD, the QR decomposition is applied for model sizing in (TARVER et al., 2017), (DEVI; KURUP, 2017), (FAIG et al., 2019), (YU; JIANG, 2013). (QIAN et al., 2014), (PIAZZA; RAO; OTTERSTEN, 2013). The following steps illustrate the procedure:

**Initialization:** Initial values for NL order $K$ and memory depth $M$ are chosen. The $N \times R_1$ regression matrix block $\mathbf{\Phi}_{(1)}$ is assembled and ortho-normalized, producing $\mathbf{\Theta}_{(1)}$.

**Step 1)** For initial model sizing $K$ and $M$

$$\hat{\boldsymbol{\beta}}'_{OLS(1)} = \mathbf{\Theta}^H_{(1)}\tilde{\boldsymbol{x}}, \tag{6.15}$$

$$\hat{\boldsymbol{x}}_{(1)} = \mathbf{\Theta}_{(1)}\hat{\boldsymbol{\beta}}'_{OLS(1)}, \tag{6.16}$$

$$\hat{\boldsymbol{e}}_{(1)} = \tilde{\boldsymbol{x}} - \hat{\boldsymbol{x}}_{(1)}, \tag{6.17}$$

where $\boldsymbol{\Theta}_{(1)}$ is the orthonormal matrix with the initial model sizing, $\hat{\boldsymbol{\beta}}'_{OLS(1)}$ is the transformed model coefficients vector at step 1 and $\hat{\boldsymbol{e}}_{(1)}$ is the residual error signal. Regressors can be removed or added from/to $\boldsymbol{\Theta}_{(1)}$, since the residual errors are orthogonal to each other. Note that in eq. (6.15) the coefficients are decoupled, i.e., can be obtained individually by $\hat{\beta}'_{OLS(i)} = \boldsymbol{\theta}_i^H \tilde{\boldsymbol{x}}$.

**Step 2)** Until the residual error is above a given target, assemble a new $N \times R_2$ regression matrix block $\boldsymbol{\Phi}_{(2)}$, ortho-normalize it with respect to the previous ones and estimate the corresponding incremental coefficients, as follows:

$$\hat{\boldsymbol{\beta}}'_{OLS(2)} = \boldsymbol{\Theta}_{(2)}^H \hat{\boldsymbol{e}}_{(1)}, \tag{6.18}$$

$$\hat{\boldsymbol{x}}_{(2)} = \begin{bmatrix} \boldsymbol{\Theta}_{(1)} & \boldsymbol{\Theta}_{(2)} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}}'_{OLS(1)} \\ \hat{\boldsymbol{\beta}}'_{OLS(2)} \end{bmatrix}, \tag{6.19}$$

$$\hat{\boldsymbol{e}}_{(2)} = \tilde{\boldsymbol{x}} - \hat{\boldsymbol{x}}_{(2)}. \tag{6.20}$$

Next, let us discuss the SVD decomposition of the regression matrix $\boldsymbol{\Phi}$. The SVD procedure decomposes the original $N \times R$ matrix $\boldsymbol{\Phi}$, as follows:

$$\boldsymbol{\Phi} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^H, \tag{6.21}$$

where $\boldsymbol{U}$ and $\boldsymbol{V}$ are square and unitary (complex-valued orthogonal) matrices of size, respectively, $N \times N$ and $R \times R$, and $\boldsymbol{\Sigma}$ is a $N \times R$ diagonal matrix. Similarly, the $R \times R$ correlation matrix $\boldsymbol{\Phi}^H\boldsymbol{\Phi}$ can be written as $\boldsymbol{\Phi}^H\boldsymbol{\Phi} = \boldsymbol{V}\boldsymbol{\Sigma}^2\boldsymbol{V}^H$, where the columns of $\boldsymbol{V}$ contain the eigenvectors of $\boldsymbol{\Phi}^H\boldsymbol{\Phi}$ and the $R \times R$ matrix $\boldsymbol{\Sigma}^2 = diag(\lambda_1, \ldots, \lambda_R)$, the corresponding eigenvalues.

This way, the most significant regressors correspond to the largest singular values. Using the matrices $\boldsymbol{V}$ and $\boldsymbol{\Sigma}^2$, sort the regressors by their importance to the model and finally arrive to an appropriate model sizing, by only keeping the most relevant regressors (dimension grading) (LOPEZ–BUENO et al., 2018), (GILABERT et al., 2013).

Although orthogonalizing the regression matrix $\boldsymbol{\Phi}$ simplifies the estimation and model sizing tasks associated to the model in eq. (6.1), in the context of DPD the transformation of the original, observable regression basis, $\boldsymbol{\Phi}$, into an orthogonal basis $\boldsymbol{\Theta}$ severely impacts the running cost of the model, since a matrix-to-matrix multiplication is required before the DPD filtering of the input signal (SCHUARTZ et al., 2019).

### 6.2.3 Approximate Techniques

The techniques presented so far are computationally very intensive. There are also some alternative, approximate methods proposed in the literature that attempt to arrive to a model sizing estimate in less costly ways. Firstly, the NL order $K$ of the model can be approximated by assuming that the static non-linearity parcel dominates the NL behaviour (FEHRI; BOUMAIZA, 2013). Thus, setting the memory indexes to zero, the Volterra-based models are reduced to the static power series:

$$\hat{x}_S(n) = \sum_{\substack{k=1 \\ k\ odd}}^{K} \beta_k \tilde{u}(n) \left|\tilde{u}(n)\right|^{k-1}, \tag{6.22}$$

where $\boldsymbol{\beta}_k$ corresponds to the Volterra coefficients with all $m = 0$. The power series model output $\hat{x}_S(n)$ is used to estimate the NL order, by computing the RSS with increasing orders, until only marginal improvements are achieved or a given threshold is reached.

Likewise, the memory depth of the model can be approximated directly from the input/output dataset, therefore avoiding try-and-error or several model estimations. The major approximate approaches in this case are the Lipschitz quotients (HE; ASADA, 1993), (O'BRIEN et al., 2006), (DOOLEY et al., 2015), z-score technique (LI; ZHU, 2020) and other empirical methods (CHOUCHANE; BOUZRARA; MESSAOUD, 2018), (MURILLO; YAGUE-JIMENEZ, 2017),(LIU et al., 2007). Next, we describe the estimation of the required input sample delays using the Lipschitz quotients (HE; ASADA, 1993). Let us consider that the Lipschitz quotients are computed for a range of values of the parameter $l$, from zero to a maximum predefined value. For each value of $l$, the Lipschitz quotients $q_{ij}^{(l)}$ are computed using samples $i$ and $j$, distant to each other up to $l$ samples,

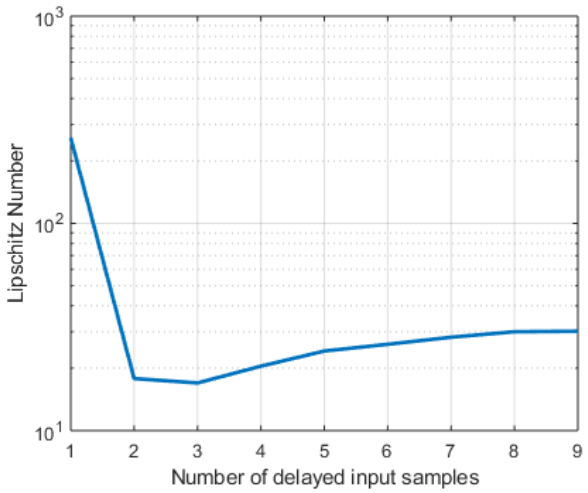



Figure 6.3 – Lipschitz quotients versus input delay.

for $i, j = 1, \ldots, N$, by the expression:

$$q_{ij}^{(l)} = \frac{|x(i) - x(j)|}{|u(i) - u(j)|}, i \neq j; i, j = 1, \ldots, N \tag{6.23}$$

where $|u(i) - u(j)|$ is the distance between two samples $u(i)$ and $u(j)$ of the input signal and $|x(i) - x(j)|$, the distance between two output samples. Next, keep only the $p$ largest quotients among all $q_{ij}^{(l)}$ for each $l$, with $p$ a small fraction of $N$, and calculate for $l$ the average Lipschitz quotient, by:

$$\bar{q}^{(l)} = \Big( \prod_{k=1}^{p} \sqrt{l} q_k^{(l)} \Big)^{1/p} \tag{6.24}$$

where $q_k^{(l)} \in \{p \text{ largest quotients } q_{ij}^{(l)}\}$. Repeating this process for the range of values of $l$ and plotting $\bar{q}^{(l)}$ versus $l$ provides a *knee* on the curve that corresponds to the maximum lag of the input variable that should be chosen. In short, if $l$ is the optimal lag then $\bar{q}^{(l+1)}$ is very close to $\bar{q}^{(l)}$, and $\bar{q}^{(l-1)}$ is much larger than $\bar{q}^{(l)}$.

As a numerical example, Figure 6.3 shows the Lipschitz quotients versus the memory depth of the input signal, computed using a measured PA input/output dataset. The higher the number of relevant input delays that are missing, the larger the Lipschitz quotient, while the addition of redundant input delays does not change significantly the Lipschitz quotient value.

## 6.3 PROPOSED WHFB BLOCK STRUCTURE

In addition to LASSO, block-wise LASSO extensions (such as group-LASSO and SGL) are also considered for selecting the dominant blocks of regressors from the initial WHFB model and discarding unnecessary blocks. This procedure gives a set of the most relevant model blocks, leading to a systematic model sizing approach.

To this end, the model in eq. (5.21) has to be re-arranged into a block-wise modular structure. The WHFB regression matrix is firstly divided into the MP, GMP and WHFB-type disjoint sub-matrices. To simplify the notation, the same parameters $M$, $K$, $L$, $Q$ are assumed in the sub-matrices. Next, each of the sub-matrices is further divided into non-overlapping blocks of regressors. The procedure is described in more details as follows:

**Step 1)** The regression matrix $\mathbf{\Phi}$ corresponding to eq. (6.1) is partitioned into sub-matrices that contain, respectively, the MP, GMP and WHFB-type regressors nested into $\mathbf{\Phi}$, avoiding any repetition of regressors:

$$\mathbf{\Phi} = \begin{bmatrix} \mathbf{\Phi}_{mp} & \mathbf{\Phi}_{gmp} & \mathbf{\Phi}_{whfb} \end{bmatrix}. \tag{6.25}$$

The sub-matrix $\mathbf{\Phi}_{mp}$ is given by eq. (5.28), whereas $\mathbf{\Phi}_{gmp}$ corresponds to eq. (5.29), with $K_2 = 1$ and eq. (5.30), with $Q_3 = 0$. $\mathbf{\Phi}_{whfb}$ corresponds to eqs. (5.29)-(5.30), otherwise. Note that this approach reduces to the MP or GMP models, when the additional WHFB terms become negligible.

**Step 2)** Each of the sub-matrices is further partitioned into disjoint blocks. Each block is defined by a fixed value of one or more parameters in the polynomial model. In this work, the parameter that is chosen to be fixed is the sample delay between the input and output signals. This choice is based on the reasonable assumption that the correlation between regressors, such as $\tilde{u}(n-m)|\tilde{u}(n-m)|^{(k-1)}$ and the response signal $\tilde{x}(n)$ is less significant as the sample delay between them increases, which is based on the physical characteristics of the PA (STAUDINGER; NANAN; WOOD, 2010) and is also verified experimentally. Therefore, the sub-matrices are divided in blocks, as follows:

- The MP sub-matrix is given by:

$$\mathbf{\Phi}_{mp} = \{\mathbf{\Phi}_{mp}^{(i)}\}_{i=1}^{M+1}, \tag{6.26}$$

where $i$ is the block index for $m = 0, \ldots, M$ and $\mathbf{\Phi}_{mp}^{(i)}$ contains the $i^{th}$ group of regressors of $\mathbf{\Phi}_{mp}$. Inside each MP block, regressors are $\tilde{u}\,(n-m)\,|\tilde{u}\,(n-m)|^{(k-1)}$, where $k = 1, \ldots, K$, with $k$ odd, $n = 1, \ldots, N$ and block size is $N \times (K+1)/2$.

- The GMP sub-matrix is given by:

$$\mathbf{\Phi}_{gmp} = \{\mathbf{\Phi}_{gmp}^{(i)}\}_{i=1}^{M+1}, \tag{6.27}$$

where $i$ is the block index for $m = 0, \ldots, M$ and $\mathbf{\Phi}_{gmp}^{(i)}$ contains the $i^{th}$ group of regressors of $\mathbf{\Phi}_{gmp}$. Inside each GMP block, regressors are $\tilde{u}\,(n-m)\,|\tilde{u}\,(n-m-l)|^{q}$, where $q = 2, \ldots, Q$, with $q$ even, $l = -L, \ldots L$, $l \neq 0$, $n = 1, \ldots, N$ and block size is $N \times QL$. Note that if $k = 1$ or $l = 0$ the corresponding terms would reduce to MP terms, so these values are neglected.

- In the WHFB sub-matrix $\mathbf{\Phi}_{whfb}$, the sample delays $m$ and $l$ in eq. (5.21) are replaced by $v = m + l$, in order to avoid the repetition of terms. The WHFB blocks are defined as:

$$\mathbf{\Phi}_{whfb} = \{\mathbf{\Phi}_{whfb}^{(i)}\}_{i=1}^{M+2L+1}, \tag{6.28}$$

where $i$ is the block index for $v = -L, \ldots, M + L$ and $\mathbf{\Phi}_{whfb}^{(i)}$ contains the $i^{th}$ group of regressors of $\mathbf{\Phi}_{whfb}$. Inside each WHFB block, the regressors are given by $\tilde{u}(n-v)\left|\tilde{u}(n-w)\right|^{k}\left|\tilde{u}(n-v)\right|^{(q-1)}$, where $w = \max(0, v-L), \ldots, \min(M, v+L)$, $k = 2, \ldots, K$, with $k$ even, $q = 1, \ldots, Q$, with $q$ odd, $n = 1, \ldots, N$ and block size is $N \times (\min(M, v+L) - \max(0, v-L))(K-1)(Q+1)/4$. Note that if $k = 0$ the corresponding terms would reduce to MP terms.

The procedure proposed in eqs. (6.26)-(6.28) leads to a WHFB modular structure that is suitable for the Group-LASSO and SGL sparse estimation. These techniques take advantage of the modular structure of the WHFB model, but become computationally expensive for large data sets. In the next section, a simplified, approximate technique is developed, in which the most relevant WHFB blocks are selected prior to running the LASSO estimation, resulting in lower estimation cost than other techniques.

## 6.4 PROPOSED SPARSE BLOCK SELECTION ALGORITHM

The proposed technique follows a divide-and-conquer (CHEN; XIE, 2014) approximate approach, firstly selecting the most relevant blocks of the WHFB regression matrix $\mathbf{\Phi}$. In order to select the most relevant blocks, the proposed technique computes, for each WHFB block, the norm of the correlation vector, which contains the correlations of the regressors in the block to the response signal. Then, LASSO is applied only to the blocks that are mostly correlated with the response signal, as detailed in the algorithm described next. By performing the model dimensioning prior to LASSO, a more affordable regression matrix is achieved and the estimation cost is reduced.

The following inputs are required by the algorithm:

- The N-length complex sample training datasets $\{\tilde{u}(n), \tilde{x}(n)\}_{n=0}^{N-1}$,
- The WHFB initial model size $M$, $K$, $L$ and $Q$,
- The initial model block structure, for example: $M+1$ MP blocks, $M+1$ GMP blocks and $M + 2L + 1$ WHFB blocks,
- The block correlation thresholds $T_{mp}$, $T_{gmp}$ and $T_{whfb}$ for the MP, GMP and WHFB sub-matrices,
- The LASSO sparsity parameter vector $\boldsymbol{\lambda}$,

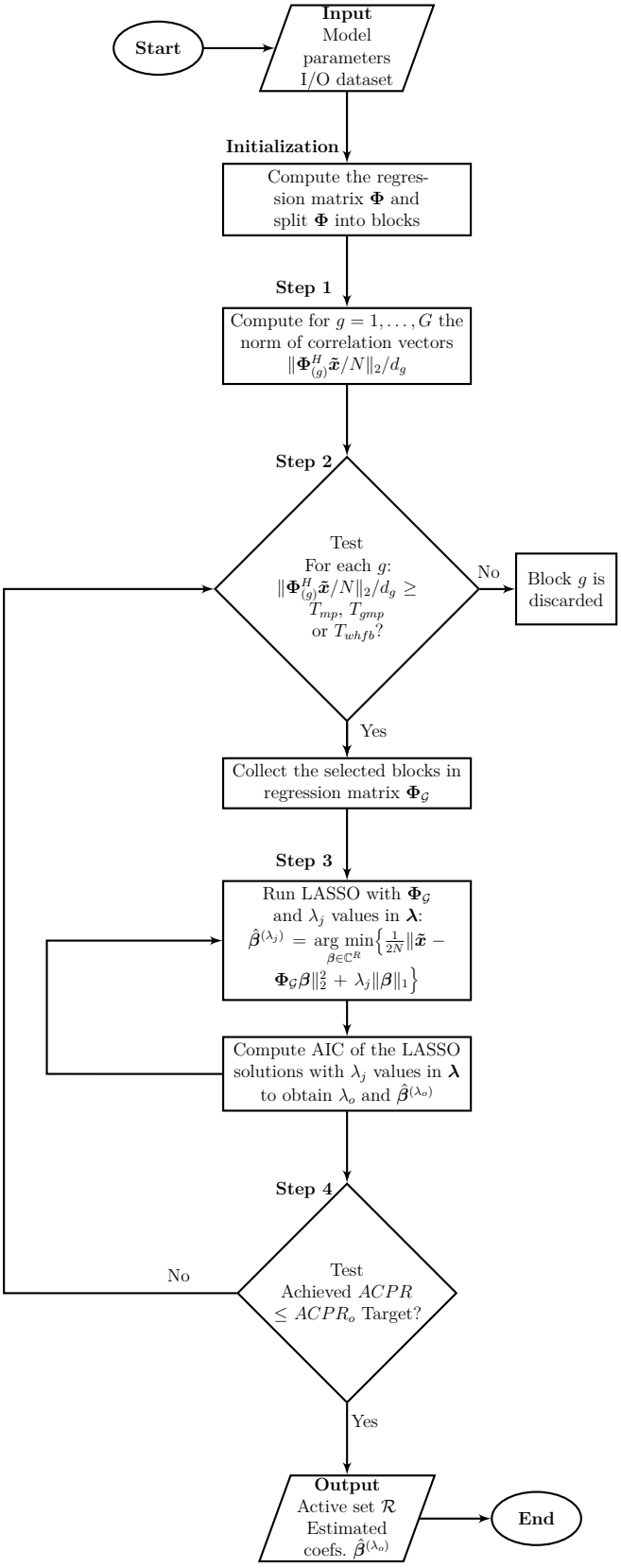



Figure 6.4 – Proposed sparse block selection algorithm.

- The target ACPR, $ACPR_o$.

As shown in Figure 6.4, the algorithm comprises the following steps:

**Initialization:** The proposed algorithm computes the WHFB regression matrix $\mathbf{\Phi}$ from eqs. (6.26)-(6.28), where each sub-matrix is partitioned into blocks, as described in the previous section.

**Step 1)** The norm of the correlation vector of each block $\mathbf{\Phi}_{(g)}$ of each sub-matrix and the response signal $\tilde{\boldsymbol{x}}$ is computed as $\|\mathbf{\Phi}_{(g)}^H \tilde{\boldsymbol{x}}/N\|_2/d_g$, where $d_g$ is the number of regressors in block $g$;

**Step 2)** A threshold function selects the block $\mathbf{\Phi}_{(g)}$ of each sub-matrix in the model only if its correlation vector norm is larger than the corresponding threshold value, i.e., $T_{mp}$, $T_{gmp}$ or $T_{whfb}$, otherwise the block is discarded. For each sub-matrix, the indexes of the selected blocks are stored in the support set $\mathcal{G}$;

**Step 3)** The LASSO technique is applied to $\mathbf{\Phi}_{\mathcal{G}}$ containing the groups in the solution set in $\mathcal{G}$ and the best value of the sparse parameter, $\lambda_o$, is obtained using the AIC score in eq. (3.4);

**Step 4)** The support set $\mathcal{R}$ of active regressors is obtained from the LASSO selection, i.e., for which $\hat{\beta}^{(\lambda_o)} \neq 0$. The selected coefficients are then re-estimated using OLS, as recommended in (MEINSHAUSEN, 2007). After the algorithm is complete, $\mathcal{R}$, and $\hat{\boldsymbol{\beta}}^{(\lambda_o)}$ are copied to the predistorter function. The algorithm compares the resulting ACPR with its target and decides whether to adjust initial parameters and repeat, or terminate. The out-of-band ACPR metric is defined in eq. (3.2). Additionally, the DPD performance is also evaluated in terms of the in-band NMSE metric, defined in eq. (3.5).

As described in this section, LASSO and its extensions achieve a sparse, parsimonious WHFB model estimation that reduces the DPD running cost and alleviates the regression matrix ill-conditioning.

The next section presents the experimental results.

## 6.5 EXPERIMENTAL VALIDATION

Measured data for PA LMM, obtained with the experimental test set-up described in section 3.1, is used to validate the DPD techniques proposed in chapters 5 and 6. The WHFB model from chapter 5, as well as sparse WHFB models from chapter 6, are compared in terms of numbers of coefficients, ACPR and NMSE, to well-known DPD models, such as the MP and GMP, and to the Volterra-PARAFAC and Laguerre-Volterra models from chapter 4. Also, the block selection sparse algorithm proposed in section 6.4 is validated in this section.

### 6.5.1 DPD Model Comparison

This subsection compares the ability of the MP, GMP and WHFB polynomial models, based in eqs. (2.9), (2.17) and (5.21), to linearize the PA under load mismatch. The results are based in experimental data measured as described in section 3.1 , with input power $P_{in}$= −8 dBm, signal bandwidth BW= 20 MHz and load mismatch VSWR of 4:1. The original dense MP, GMP and WHFB models are considered in this item, i.e., sparsity is not explored, and the model coefficients are estimated by OLS using the Gram-Schmidt factorization, to avoid the ill-conditioned inversion of the matrix $\mathbf{\Phi}^H\mathbf{\Phi}$. Figure 6.5 presents the residual ACPR values after linearization versus the numbers of regressors, comparing PA perfectly matched and VSWR=4:1 load mismatched conditions. The model parameters are determined as follows. The best combinations of MP parameters $M$ and $K$, in terms of the residual out-of-band distortion, is chosen for each number of regressors $(M+1)\times(K+1)/2$. Both parameters are swept from 1 to 10. The ACPR is gradually reduced, until a region of diminishing improvements, i.e., a MP *plateau* region is reached. At this point, no further significant enhancements are provided by the MP model, thus, the GMP model is introduced, by keeping $M$=7, $K$=9 and varying the parameter $L$ from 1 to 6. A GMP *plateau* region is reached for $L$=5 and the WHFB model is introduced, keeping $L$=5 and sweeping the parameter $Q$ from 1 to 6. As illustrated in the figure, lower values of $L$ lead to a compromised performance of the WHFB model and, finally, $Q$=5 is chosen.

As shown in Figure 6.5, in order to comply with an ACPR level ranging from −45 to −50 dBc, imposed by wireless standards (SESIA; TOUFIK; BAKER, 2009), the MP and/or GMP models suffice in the perfectly matched PA case, whereas, only the WHFB model is enough under load mismatch conditions. In Figure 6.6, the residual spectral regrowth obtained with the MP, GMP and WHFB DPD models under load mismatch are compared, assuming model parameters $M$= 7, $K$= 9, $L$= 5 and $Q$= 5. The ACPR with the MP and GMP models are as high as −42.0 and −44.7 dBc, respectively, while

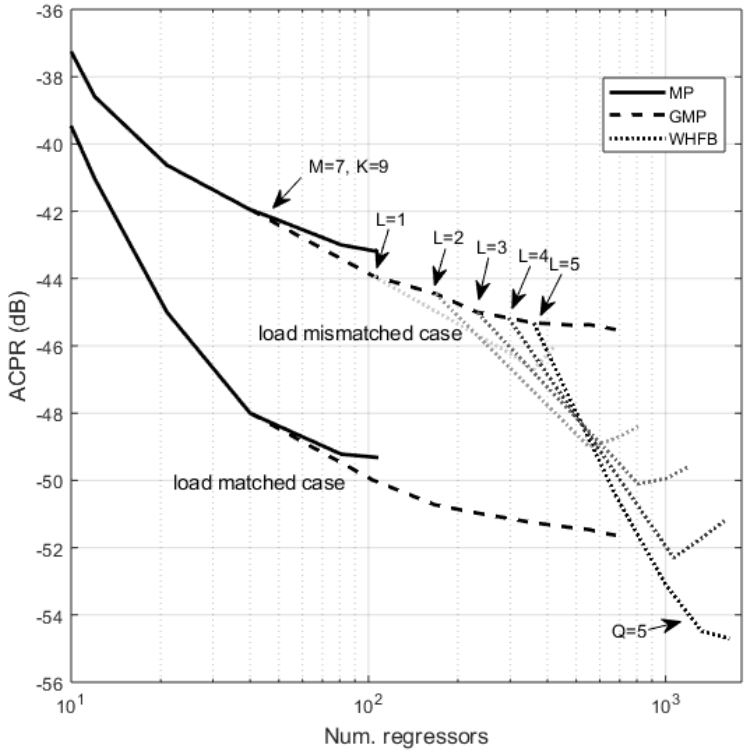



Figure 6.5 – Comparison of DPD model performance for load matched and LMM PA.

the WHFB achieves −53.3 dBc. The NMSE with the MP, GMP and WHFB models are, respectively, −29.5, −32.7 and −33.9 dB and the numbers of coefficients are 40, 360 and 1320, respectively. Figure 6.7 shows the constellation diagrams, after demodulating the OFDM symbols, of the MP, GMP and WHFB DPD output signals. The corresponding EVMs of the output signals without DPD and with the MP, GMP and WHFB DPD models are, respectively, 16.8%, 3.5%, 2.4% and 1.7%. Comparing the ACPR, NMSE and EVM values of the models, we can see that both in-band and out-of-band distortions are reduced with the WHFB model. Out-of-band distortions are reduced even more significantly, which is important, since the system must obey spectral emission masks. The following subsections analyse the linearization performance versus the numbers of coefficients for sparse estimation techniques, with initial model parameters $M$= 7, $K$= 9, $L$= 5, $Q$= 5 and $N$= 30,000 training samples.

## 6.5.2 LASSO Results

The large size of the WHFB model implies increased DPD running cost and the risk of ill-conditioning of the regression matrix during the estimation, thus motivating sparse selection techniques. The LASSO optimization problem in eq. (6.3) is applied to $\mathbf{\Phi}$ in eq. (5.21), with the model parameter values defined above. First, $\mathbf{\Phi}$ is standardized and the output signal is centered, then the LASSO algorithm is run for sparse parameter

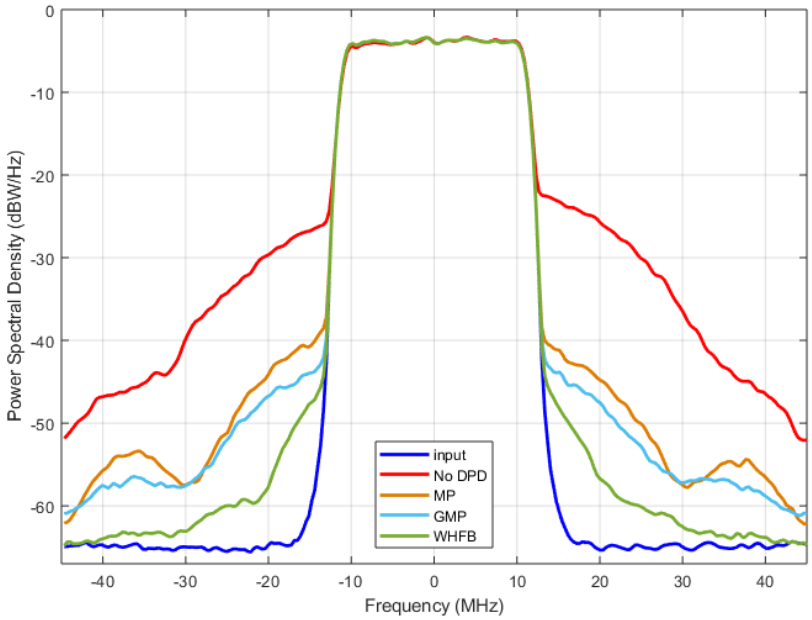



Figure 6.6 – Residual spectral regrowth for LMM PA with WHFB model.

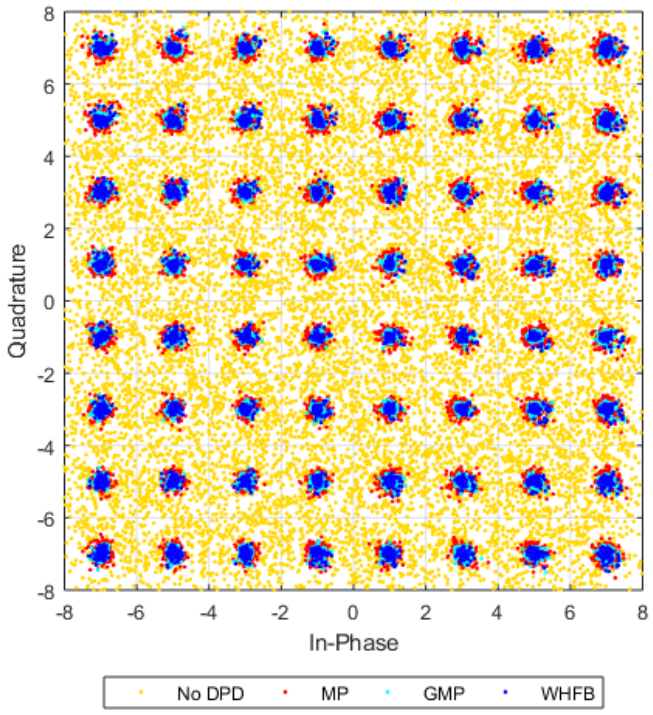



Figure 6.7 – Constellation diagram for LMM PA with WHFB model.

values defined in the vector $\boldsymbol{\lambda}$. The DPD models were implemented using MATLAB and the LASSO, using the SLEP 4.1 tool (LIU; JI; YE, 2009), although other methods could be used (EFRON et al., 2004), (FRIEDMAN; HASTIE; TIBSHIRANI, 2010). In

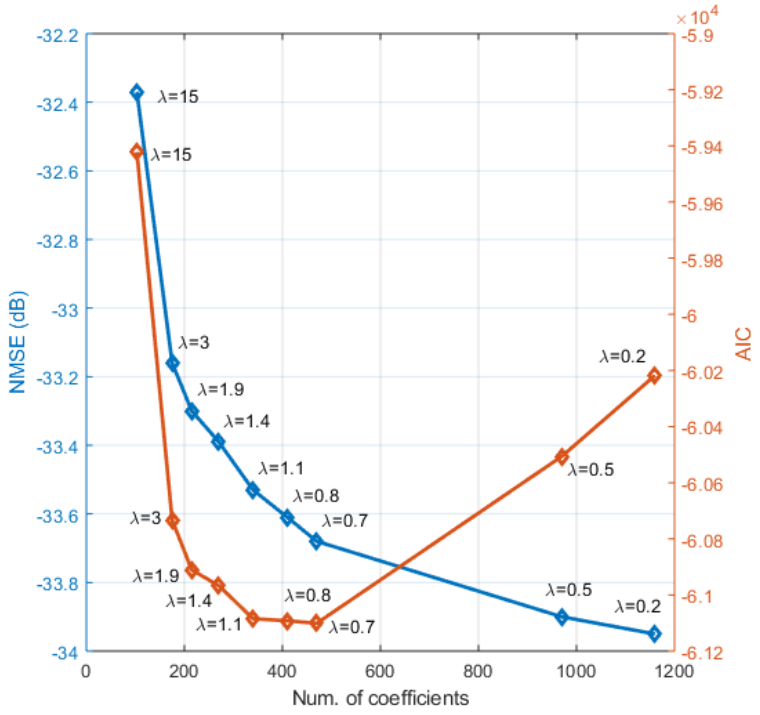



Figure 6.8 – NMSE (dB) and AIC versus number of coefficients (sweep $\lambda$).

Figure 6.8, the *knee* in the NMSE curve indicates sparse solutions for which very small degradations of the model performance are observed. This is confirmed using the AIC curve in the same figure, i.e., a good trade-off between accuracy and complexity is achieved close to the minimum AIC value. Note that a value of $\lambda$ that is too large can prevent the LASSO from capturing the signal behaviour, while too small, can lead to over-fitting. The best sparsity parameter $\lambda_o$ is determined by the AIC score from eq. (3.4). Figure 6.9 illustrates the residual spectral regrowth achieved with the GMP, WHFB LASSO, WHFB group-wise LASSO and the proposed algorithm, all for approximately the same number of coefficients. Firstly considering LASSO, for the selected value $\lambda_o$=1, LASSO is able to reduce the number of coefficients in the WHFB model from 1320 to 356, thus reducing the DPD running cost, while yielding a ACPR performance of −49.4 dBc and NMSE of −33.5 dB. Note that $\lambda_o$=1 for LASSO is chosen by the AIC criterion and resulted in approximately the same number of coefficients as the GMP model. Figure 6.10 shows the constellation diagrams of the WHFB and WHFB LASSO DPD output signals. As can be seen, their constellation diagrams are very similar; thus, the in-band performance of the sparse WHFB model is very close to the WHFB model. The EVM of the WHFB LASSO output signal is 1.9%. The LASSO sparsity effect, i.e., the reduction in the number of coefficients, is approximately 4:1 in this case. Note that the model coefficients selected by LASSO are post-estimated using OLS, as recommended in the literature. The GMP is known as providing a good trade-off between model complexity and performance in the load matched case (TEHRANI et al., 2010), but showed performance degradation under

load mismatch.

### 6.5.3 Group-LASSO and Sparse group-LASSO Results

In the next two subsections, group-wise LASSO extensions are considered. For this, the modular WHFB structure in eqs. (6.26)-(6.28) is obtained by partitioning $\Phi$ into MP, GMP and WHFB sub-matrices and then into $M+1$, $M+1$ and $M+2L+1$ blocks, respectively. In this subsection, group-LASSO and SGL techniques from eqs. (6.7) and (6.9) are applied to the WHFB model, the corresponding sparsity parameters are optimized using the AIC score and finally adjusted to result in approximately the same number of coefficients as with LASSO. Returning to Figure 6.9, the residual spectral regrowth obtained with group-LASSO and SGL can be also visualized. Group-LASSO and SGL are able to reduce the number of coefficients in the WHFB model from 1320 to 414 and 380, respectively, while yielding an ACPR performance of $-45.5$ and $-49.2$ dBc and NMSE of $-33.2$ and $-33.6$ dB, respectively. Note that using group-LASSO only the sparsity among blocks is explored, i.e., once a block is selected, all regressors within the block are selected, thus the benefit is expected to be less attractive than with SGL. The results show that SGL provides a better solution than group-LASSO, both in terms of accuracy and sparsity, but this is at the expense of a higher estimation cost, needed for optimizing both the group and regressor sparsity parameters $\lambda$ and $\alpha$ in a two-dimensional grid. In the next subsection, the proposed algorithm is considered.

### 6.5.4 Proposed Block Selection Sparse Algorithm Results

This item illustrates the algorithm proposed in section 6.4. As described therein, the norms of the block correlation vectors are compared to the thresholds $T_{mp}$, $T_{gmp}$ or $T_{whfb}$, in order to find the interval of blocks in each sub-matrix that is highly correlated to the output signal and neglect the tails in the initial model. The selected blocks in each sub-matrix are shown in gray in Figure 6.11, corresponding to a simple model dimensioning approach. The initial assumption that the correlations with the output signal become less important as the block lagging/leading increases (STAUDINGER; NANAN; WOOD, 2010) is verified, although the blocks in $\mathbf{\Phi}$ are correlated to each other. Note that the algorithm can refine the threshold values, after the correlation norms are calculated, in order to adjust the selection to the interval of dominant blocks. Also note that the algorithm can further sub-divide each selected block, allowing the dimensioning of other model parameters, such as the model power order. For simplicity, this is not explored here. In Figure 6.11, the threshold values are set to 0.15 and the initial model size is shrunk from 1320 to 715 regressors, thus LASSO is run over a more affordable regression matrix. In the LASSO step, several values of the sparse parameter are considered and, in this subsection, $\lambda_o$=0.7 is chosen based on AIC. Finally, back to Figure 6.9, the residual spectral regrowth obtained

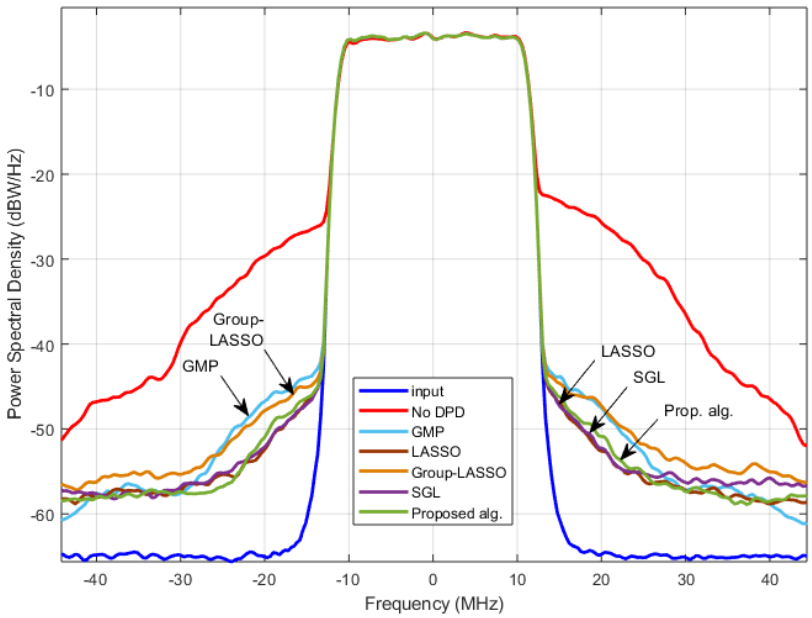



Figure 6.9 – Residual spectral regrowth for LMM PA with sparse WHFB.

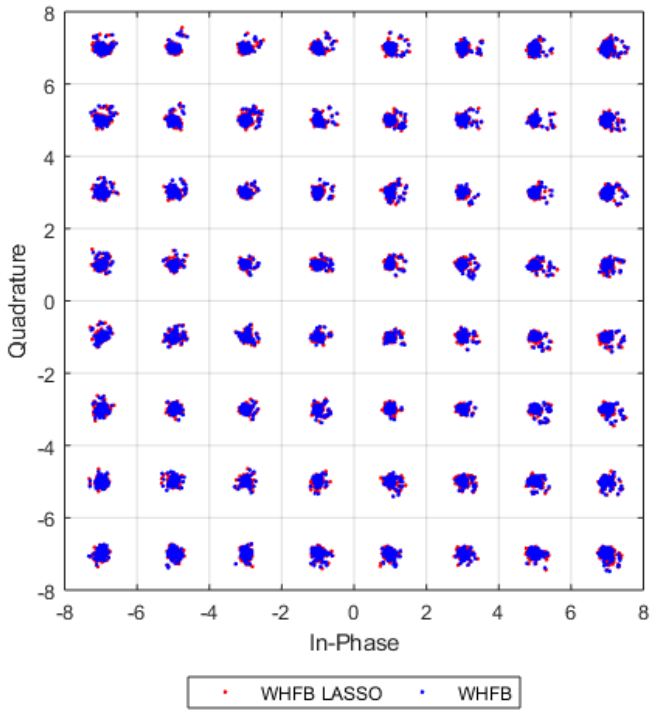



Figure 6.10 – Constellation diagram for LMM PA with WHFB LASSO.

with the proposed algorithm is illustrated. After completing the proposed algorithm, the number of coefficients in the WHFB model is reduced from 1320 to 350, while achieving an ACPR performance of −48.6 dBc and NMSE of −33.5 dB. The summary of the ACPR

and NMSE results achieved with the techniques from subsections 6.5.1 to 6.5.4 is provided in Table 6.1. Note that in the table, LASSO with $\lambda$=2 is also provided for comparison. This value of *lambda* reduces the number of coefficients, but also the model performance, not being optimal in terms of AIC. The table also includes the VP and LV models from chapter 4, showing that the sparse WHFB model adequately compensate the LMM PA under study, resulting in lower numbers of coefficients than with VP and LV models. The computational cost of the LASSO and group-wise LASSO solutions is considered in the next subsection.

Table 6.1 – DPD performance for LMM PA with WHFB and sparse estimation ($M$= 7, $K$= 9, $L$= 5, $Q$= 5, $N$= 30,000).

| | **Num. coeffs.** | **Max. ACPR (dBc)** | **NMSE (dB)** |
|---|---|---|---|
| No DPD | - | -23.2 | 3.3 |
| MP OLS | 40 | -42.0 | -29.5 |
| GMP OLS | 360 | -44.7 | -32.7 |
| WHFB OLS | 1320 | -53.3 | -33.9 |
| WHFB LASSO ($\lambda_o$=1) | 356 | -49.4 | -33.5 |
| WHFB LASSO ($\lambda$=2) | 202 | -47.4 | -33.3 |
| WHFB Group-LASSO | 414 | -45.5 | -33.2 |
| WHFB SGL | 380 | -49.2 | -33.6 |
| WHFB Prop. alg. ($\lambda_o$=0.7) | 350 | -48.6 | -33.5 |
| VP (ALS) ($M_k = [5, 5, 4, 2, 2]$) ($R_k = [1, 13, 45, 32, 30]$) | 501 | -43.5 | -33.6 |
| LV ($M_k = [5, 5, 4, 3, 2]$) ($F_k = [5, 4, 3, 1, 1]$) | 331 | -43.2 | -33.6 |

### 6.5.5 Computational Cost

The DPD computational complexity can be classified into estimation, adaptation and running complexity (TEHRANI et al., 2010). The estimation complexity refers to the DPD training, while the adaptation part refers to additional training required to adapt to new operating and/or environmental conditions. Finally, the running (or execution) complexity mainly involves real-time pre-distortion filtering. From (TEHRANI et al., 2010), the running cost of Volterra-based models in floating point operations (FLOPs) per sample can be approximated by $8R - 2$, where $R$ is the number of regressors in the given model.

The estimation cost for OLS, including the Gram-Schmidt factorization, is assessed by $\mathcal{O}(2NR^2)$ FLOPs, where $N$ is the sample size and $R$, the number of regressors. For LASSO and extensions, it is assumed that the convex minimization is solved by the coordinate descent algorithm (FRIEDMAN; HASTIE; TIBSHIRANI, 2010) over the range of sparsity parameters defined for each problem. Note that one cycle of coordinate descent

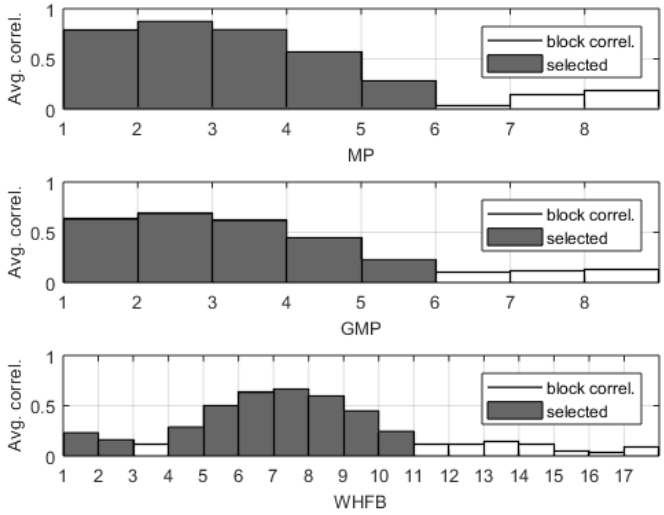



Figure 6.11 – Selected MP, GMP and WHFB blocks using the proposed algorithm.

costs $\mathcal{O}(NR)$ operations. The resulting estimation cost is approximated by $\mathcal{O}(spNR)$, where $s$ is the number of $\lambda$ values and $p$ is the number of iterations until convergence, both depending on the cost function (TIBSHIRANI; HASTIE; FRIEDMAN, 2010) under consideration. Typically, the DPD training is computed off-line and holds valid for a large amount of running data. In scenarios where the PA is subject to impedance mismatch, due to operational and/or environmental changes more frequent adaptations are expected, making the proposed algorithm even more advantageous. The resulting values for running and estimation complexities are presented in Table 6.2.

Table 6.2 – DPD complexity with WHFB and sparse estimation ($M$= 7, $K$= 9, $L$= 5, $Q$= 5, $N$= 30,000).

| | **Num. coeffs.** | **Estimation cost (FLOPs)** | **Running cost (FLOPs/sample)** |
|---|---|---|---|
| No DPD | - | - | - |
| MP OLS | 40 | $3\times10^3$ | 318 |
| GMP OLS | 360 | $2\times10^5$ | 2880 |
| WHFB OLS | 1320 | $3\times10^6$ | 10560 |
| WHFB LASSO ($\lambda_o$=1) | 356 | $3\times10^8$ | 2846 |
| WHFB LASSO ($\lambda$=2) | 202 | $3\times10^8$ | 1614 |
| WHFB Group-LASSO | 414 | $10^8$ | 3310 |
| WHFB SGL | 380 | $5\times10^8$ | 3038 |
| WHFB Prop. alg. ($\lambda_o$=0.7) | 350 | $3\times10^7$ | 2798 |

## 6.6 CONCLUSIONS

This chapter tackled the high dimensionality of the WHFB model employing sparse techniques, such as LASSO and block-wise LASSO. Then, this chapter demonstrated the

in-band and out-of-band performance improvements provided by the WHFB and sparse WHFB models, comparing to the more traditional MP and GMP DPD models. Finally, the chapter proposed an algorithm that partially alleviates the estimation cost of LASSO, while maintaining the DPD model performance. This algorithm is a more affordable sparse selection alternative for high-dimensional polynomial models.

# 7 FLEXIBLE BASEBAND VOLTERRA MODEL PRUNING TECHNIQUE

The Volterra framework (SCHETZEN, 2006) plays an important role in DPD modelling for high-power and wideband PAs, but, in practice, the $FV_{\triangle}$ model in eq. (2.4) is avoided, due to the curse of dimensionality, i.e., the large number of coefficients to be estimated, the corresponding length of the DPD filter and its running cost. As presented in section 2.2, this has motivated several Volterra pruning techniques in the literature, the most prominent of them are the MP, GMP, DDR and block-oriented models. However, choosing the pruned-Volterra structure that, for a given model complexity, best describes an unknown PA or DPD is not an easy problem. This chapter is motivated by the challenging task of *a priori* choosing a pruned-Volterra model, among a wide range of candidate solutions, that is able to accurately describe an unknown PA or DPD, specially when no physical knowledge about the system is available.

The objective of this chapter is to propose a flexible yet economical pruned-Volterra structure from the $FV_{\triangle}$ model that generates a combination of MP/GMP, WHFB and DDR regressors, thus being able to describe a wide range of NL systems. The parameters of the flexible model, such as the polynomial dimension (i.e., the number of input cross-terms in the regressors), the exponents of the input terms and the memory depths, are adjusted separately for each NL order. Model sizing, i.e., finding appropriate values for the model parameters, is performed through block-oriented LASSO techniques, such as group-LASSO and sparse-group LASSO, also applied in the previous chapter. The contribution in this chapter is that the proposed models present a significantly reduced DPD running complexity, compared to the triangular FV models, while are more flexible than pruning approaches in the literature, such as MP/GMP and DDR models.

## 7.1 CONTEXT OF PRUNED-VOLTERRA MODELS

The triangular $FV_{\triangle}$ and $MV_{\triangle}$ baseband models are given, respectively, in eqs. (2.4) and (2.7). The major advantage of the $MV_{\triangle}$ model is that its dimensionality can be controlled by setting shorter memory spans for the higher NL orders. However, the number of $MV_{\triangle}$ coefficients in eq. (2.8) still is given by a doubly-combinatorial expression, which makes the model computationally costly as increase the NL order $K$ and a set of memory depths $\{M_k\}$. Therefore, additional Volterra pruning techniques have been developed in

the literature, among them, the most prominent are described as follows:

- The MP model (KIM; KONSTANTINOU, 2001), expressed in eq. (2.9), only considers one-dimensional terms, i.e., regressors that only depend on a single sample of the input signal (SCHUARTZ et al., 2019), such as $\tilde{u}(n-m_1)\,|\tilde{u}(n-m_1)|^2$. By avoiding input cross-terms, the MP model largely reduces its size, so $k^{th}$-dimensional FV kernels with length $M^k$ are replaced by a one-dimensional, length $M$, coefficients vector.

- Likewise, the GMP model (MORGAN et al., 2006), in eq. (2.17), also limits the number of input cross-terms, but allowing up to two-dimensional terms. Thus, GMP regressors are given by up to two distinct input samples, apart from each other by up to $l \leq L$ samples, as for example $\tilde{u}(n-m_1)\,|\tilde{u}(n-m_1 \pm l)|^2$.

- The DDR model (ZHU; DOOLEY; BRAZIL, 2006), in eq. (2.18), whose strategy first splits the FV model into its static and dynamic parts, then limits the power orders of the delayed input samples (on the dynamic part) to the $1^{st}$, $2^{nd}$ or higher order. For example, $\tilde{u}(n-m_1)|\tilde{u}(n)|^2$ belongs to the $1^{st}$-order DDR model, whereas $\tilde{u}(n-m_1)\tilde{u}(n-m_2)\tilde{u}^*(n)|\tilde{u}(n)|^2$, to the $2^{nd}$-order DDR model.

Comparing the MP/GMP and DDR pruning strategies above, the former limits the polynomial dimensions in the model, while the last restricts the power orders of the delayed input samples; the polynomial dimension is not directly controlled. Also note that the baseband DDR model requires the input signal and its complex conjugate, which increases rapidly the number of required coefficients and the implementation cost with the DDR order (GUAN; ZHU, 2011).

Starting from eq. (2.7), a more flexible Volterra pruning strategy is proposed in section 7.3. Firstly, the multi-index notation is introduced.

## 7.2 MULTI-INDEX STRUCTURE

A mathematical concept known as multiset (mset or bag) (GIRISH; JOHN, 2012), (ALBERT, 1991) is introduced to represent the collections of memory multi-indexes in the triangular Volterra models in eqs. (2.4) and (2.7). The concept of msets have been used in multiple applications (SINGH et al., 2007), including in some recent signal processing papers, e.g., (DAVIS et al., 2018), (ZAKARIA; JOHN; GIRISH, 2019). An mset is an unordered collection of objects where repetitions are admitted. An mset differs from a set in the sense that each object has a multiplicity, i.e., a natural number, not necessarily one,

that indicates how many times it is an element of the mset.

**Definition 1. Multiset (mset or bag).** An mset $\mathsf{A}$ drawn from the underlying set $\mathsf{S} = \{s_1, s_2, \ldots, s_n\}$ is a collection of elements from $\mathsf{S}$ where duplicates are allowed. The multiplicity (number of occurrences) of each object $s_i$ from the set $\mathsf{S}$ in the multiset $\mathsf{A}$ is given by the counting function $C(\mathsf{S}) : \mathsf{S} \to \mathbb{N}_0$, where $\mathbb{N}_0$ is the set of the natural numbers including zero. For example, let $\mathsf{S} = \{0, 1, 2, 3\}$ and $C(\mathsf{S}) = \{3, 0, 0, 1\}$, then the mset $\mathsf{A} = [0, 0, 0, 3]$.

**Definition 2. Cardinality.** The cardinality of an mset $\mathsf{A}$ drawn from $\mathsf{S}$ is given by $card(\mathsf{A}) = \sum_{\mathsf{S}} C(\mathsf{S})$. From the previous example, $card(\mathsf{A}) = 4$.

This way, for each NL order $k$ in eq. (2.7), let the collection of memory delay multi-indexes drawn from the finite set of integers $\mathsf{S} = \{0, 1, \ldots, M_k\}$ be represented by the following set:

$$\mathsf{M}^{(k)} = \{[m_1], \underbrace{[m_2, m_3, \ldots, m_{\frac{k+1}{2}}]}_{\mathsf{M}_1^{(k)}}, \underbrace{[m_{\frac{k+3}{2}}, m_{\frac{k+5}{2}}, \ldots, m_k]}_{\mathsf{M}_{1*}^{(k)}}\}, \tag{7.1}$$

in which the first element is the index $m_1$ for one non-conjugated input and, next, msets $\mathsf{M}_1^{(k)} = [m_2, m_3, \ldots, m_{\frac{k+1}{2}}]$ and $\mathsf{M}_{1*}^{(k)} = [m_{\frac{k+3}{2}}, m_{\frac{k+5}{2}}, \ldots, m_k]$, each with cardinality $\frac{k-1}{2}$, contain, respectively, the memory indexes for the remaining non-conjugated and the conjugated inputs. Additionally, the $MV_{\triangle}$ model in eq. (2.7) imposes that $m_{\frac{k+1}{2}} \geq m_{\frac{k-1}{2}} \geq \ldots \geq m_2$ and $m_{m_k} \geq m_{k-1} \geq \ldots \geq m_{\frac{k+3}{2}}$. Let $\mathcal{M}^{(k)}$ be a class containing all distinct sets $\mathsf{M}^{(k)}$. Table 7.1 illustrates a few of the sets in $\mathcal{M}^{(5)}$ for $k$=5 and $M_k$=3. Eq. (2.8) gives the total number of 200 sets in this case.

**Definition 3. Mset membership.** Let $\mathsf{A}$ be an mset drawn from the set $\mathsf{S}$. The support of $\mathsf{A}$ is $supp(\mathsf{A}) := \{s_i \in \mathsf{S} \mid C(s_i) > 0\}$. Thus, mset membership $s_i \in \mathsf{A}$ is defined as $s_i \in supp(\mathsf{A})$, otherwise $s_i$ is not a member of $\mathsf{A}$. In the previous example, $supp(\mathsf{A}) = \{0, 3\}$.

**Definition 4. Equal msets.** Two msets $\mathsf{A}$ and $\mathsf{B}$ drawn from the set $\mathsf{S}$ are equal, written as $\mathsf{A} = \mathsf{B}$, if and only if: $supp(\mathsf{A}) = supp(\mathsf{B})$ and for any object $s_i \in supp(\mathsf{A})$, $C_{\mathsf{A}}(s_i) = C_{\mathsf{B}}(s_i)$ (every element in $\mathsf{A}$ is also in $\mathsf{B}$ and conversely).

Table 7.1 – Doubly-combinatorial multisets $\mathcal{M}^{(5)}$ ($k$=5 and $M_k$=3).

| count | $\mathsf{M}^{(5)}$ |
|---|---|
| 1 | 0 \| 0 0 \| 0 0 |
| 2 | 0 \| 0 0 \| 0 1 |
| 3 | 0 \| 0 0 \| 0 2 |
| 4 | 0 \| 0 0 \| 0 3 |
| 5 | 0 \| 0 0 \| 1 1 |
| 6 | 0 \| 0 0 \| 1 2 |
| 7 | 0 \| 0 0 \| 1 3 |
| 8 | 0 \| 0 0 \| 2 2 |
| 9 | 0 \| 0 0 \| 2 3 |
| 10 | 0 \| 0 0 \| 3 3 |
| 11 | 0 \| 0 1 \| 0 0 |
| ⋮ | ⋮ \| ⋮ ⋮ \| ⋮ ⋮ |
| 100 | 0 \| 3 3 \| 3 3 |
| 101 | 1 \| 1 1 \| 0 0 |
| 102 | 1 \| 1 1 \| 0 1 |
| ⋮ | ⋮ \| ⋮ ⋮ \| ⋮ ⋮ |
| 198 | 3 \| 3 3 \| 2 2 |
| 199 | 3 \| 3 3 \| 2 3 |
| 200 | 3 \| 3 3 \| 3 3 |

## 7.3 PROPOSED MODEL STRUCTURE

This section proposes a generalized pruned-Volterra model structure, able to include regressors from the MP/GMP, WHFB and DDR models. Besides, the proposed pruned-Volterra model avoids complex-conjugate input terms and is linear in the coefficients. It also provides a good trade-off between model size and accuracy.

The proposed model has a block-wise modular structure that allows the sizing of its initial dimensions through group sparse LASSO (SIMON et al., 2013). This sparse estimation technique selects the dominant blocks of regressors in the model and discards unnecessary ones. To this end, the model's regression matrix is generated as the concatenation of non-overlapping matrix blocks, as detailed in this section.

The pruning strategy is described next in two steps. Step 1 replaces the doubly-combinatorial structure in eq. (2.7) by a generalized MP/GMP expression, whose structure is single combinatorial, by only retaining the polynomials that have non-conjugated and conjugated sample pairs with the same memory delays. This corresponds to impose in eq. (7.1) that, for each $k$:

$$\mathsf{M}_{1*}^{(k)} = \mathsf{M}_{1}^{(k)}, \tag{7.2}$$

i.e., the two msets are equal. In this case, the set $\mathsf{M}^{(k)}$ at step 1 is reduced to the following

reduced memory delay multi-indexes:

$$\mathsf{M}^{(k)} = \{m_1, \underbrace{[m_2, m_3, \ldots, m_{\frac{k+1}{2}}]}_{\mathsf{M}_1^{(k)}}\}, \tag{7.3}$$

where, now, the mset $\mathsf{M}_1^{(k)} = [m_2, m_3, \ldots, m_{\frac{k+1}{2}}]$ contains the memory indexes of squared amplitude terms.

The pruned-Volterra model is given by:

$$\begin{aligned}
\hat{x}_{RV_1}(n) \; &= \sum_{\substack{k=1 \\ k\ odd}}^{K} \sum_{m_1=0}^{M_k} \sum_{m_2=0}^{M_k} \cdots \sum_{m_{\frac{k+1}{2}}=m_{\frac{k-1}{2}}}^{M_k} \tilde{h}_k(m_{1:\frac{k+1}{2}})\tilde{u}(n-m_1) \\
&\prod_{i=2}^{\frac{k+1}{2}} \tilde{u}(n-m_i)\ \tilde{u}^*(n-m_i) \\
&= \sum_{\substack{k=1 \\ k\ odd}}^{K} \sum_{m_1=0}^{M_k} \sum_{m_2=0}^{M_k} \cdots \sum_{m_{\frac{k+1}{2}}=m_{\frac{k-1}{2}}}^{M_k} \tilde{h}_k(m_{1:\frac{k+1}{2}})\tilde{u}(n-m_1) \\
&\prod_{i=2}^{\frac{k+1}{2}} |\tilde{u}(n-m_i)|^2 ,
\end{aligned} \tag{7.4}$$

where $RV_1$ stands for the reduced-Volterra model in step 1.

Eq. (7.4) corresponds to a generalized memory polynomial model and requires, for each NL order, only $\frac{k+1}{2}$ summations. In fact, it is inspired by the MP and GMP models. Note that its triangular structure imposes that $m_{\frac{k+1}{2}} \geq m_{\frac{k-1}{2}} \geq \ldots \geq m_2$. The total number of coefficients is:

$$N_{RV_1} = \sum_{\substack{k=1 \\ k\ odd}}^{K} (M_k + 1) \left(\!\!\binom{M_k+1}{\frac{k-1}{2}}\!\!\right), \tag{7.5}$$

where $\left(\!\!\binom{M_k+1}{\frac{k-1}{2}}\!\!\right)$ is the multiset binomial coefficient, given by $\binom{M_k+\frac{k-1}{2}}{\frac{k-1}{2}}$.

Let $\mathcal{M}^{(k)}$ be a class containing all distinct sets $\mathsf{M}^{(k)}$ using eq. (7.3). Table 7.2 illustrates a few of the $\mathcal{M}^{(5)}$ sets for $k$=5 and $M_k$=3. Using eq. (7.5), the total number of 40 sets is obtained.

Next, step 2 keeps only the terms in eq. (7.4) that have polynomial dimensions up to a parameter $L_k$ plus one, where $L_k$ is chosen for each $k$, in the range $[1, \frac{k-1}{2}]$, and is later dimensioned through model sizing. The definition of polynomial dimension is given

Table 7.2 – Multisets $\mathcal{M}^{(5)}$ ($k$=5 and $M_k$=3).

| count | $\mathsf{M}^{(5)}$ |
|---|---|
| 1 | 0 \| 0 0 |
| 2 | 0 \| 0 1 |
| 3 | 0 \| 0 2 |
| 4 | 0 \| 0 3 |
| 5 | 0 \| 1 1 |
| 6 | 0 \| 1 2 |
| 7 | 0 \| 1 3 |
| 8 | 0 \| 2 2 |
| 9 | 0 \| 2 3 |
| 10 | 0 \| 3 3 |
| 11 | 1 \| 0 0 |
| ⋮ | ⋮ \| ⋮ ⋮ |
| 20 | 1 \| 3 3 |
| 21 | 2 \| 0 0 |
| ⋮ | ⋮ \| ⋮ ⋮ |
| 30 | 2 \| 3 3 |
| 31 | 3 \| 0 0 |
| ⋮ | ⋮ \| ⋮ ⋮ |
| 40 | 3 \| 3 3 |

in (SCHUARTZ et al., 2019) as the maximum number of distinct cross-terms allowed in the memory polynomial. This way, in step 2, the maximum number of distinct squared amplitude cross-terms is limited, similarly to the MP/GMP and WHFB models. This corresponds to selecting only the multi-indexes from $\mathsf{M}^{(k)}$ in which the same delays appear repeatedly. Let $\mathsf{M}^{(k,l)}$, for $l = 1, \ldots, L_k$, represent the multi-indexes in which $(\frac{k+1}{2} - l)$ indexes assume the same value, as follows:

$$\mathsf{M}^{(k,l)} = \{m_1, [\underbrace{m_2, \ldots, m_{\frac{k+1}{2}-l+1}}_{\text{same value}}, m_{\frac{k+1}{2}-l+2}, \ldots, m_{\frac{k+1}{2}}]\}. \tag{7.6}$$

Equivalently, $\mathsf{M}^{(k,l)}$ can be more economically expressed assigning multiplicity $(\frac{k+1}{2} - l)$ to the delay value in $m_2$:

$$\mathsf{M}^{(k,l)} = \{m_1, \underbrace{[m_2, m_3, \ldots, m_{l+1}]}_{\mathsf{M}_1^{(k,l)}}\}, \tag{7.7}$$

where $m_{l+1} > m_l > \cdots > m_3 > m_2$. The multiplicity in eq. (7.7) and other multiplicities assigned to indexes are considered by the polynomial exponents, as presented next. The pruned-Volterra model is given by:

$$\hat{x}_{RV_2}(n) = \sum_{\substack{k=1 \\ k\ odd}}^{K} \sum_{m_1=0}^{M_k} \sum_{l=1}^{L_k} \sum_{m_2=0}^{M_k} \sum_{m_3>m_2}^{M_k} \cdots \sum_{m_{l+1}>m_l}^{M_k} \sum_{j=1}^{\eta_k} \tilde{h}_k(m_{1:l+1}, j) \tilde{u}(n-m_1) \prod_{i=2}^{l+1} |\tilde{u}(n-m_i)|^{v_j(i-1)}, \tag{7.8}$$

where $RV_2$ stands for the reduced-Volterra model in step 2, $L_k+1$ is the maximum $k^{th}$-order polynomial dimension in the model and $\boldsymbol{v}_j \in \mathsf{V}^{(k,l)}$. The set $\mathsf{V}^{(k,l)}$ contains the exponent vectors $\boldsymbol{v}_j$ for the polynomials $(k,l)$, i.e., of order $k$ with $l$ distinct squared amplitude terms. Additionally, the parameter $\eta_k$ limits the maximum number of exponent vectors from each $\mathsf{V}^{(k,l)}$ allowed for the $k^{th}$-order terms in the model. The following expressions are proposed for calculating the vectors $\boldsymbol{v}_j$:

- For $l = 1$:

$$\mathsf{V}^{(k,1)} = \{[k+1-2l]\}, \tag{7.9}$$

- For $l = 2$:

$$\mathsf{V}^{(k,2)} = \{[k+1-2l,\ 2], \tag{7.10a}$$
$$[k-1-2l,\ 4], \tag{7.10b}$$
$$[k-3-2l,\ 6]\}, \tag{7.10c}$$

- For $l = 3$:

$$\mathsf{V}^{(k,3)} = \{[k+1-2l,\ 2,\ 2], \tag{7.11a}$$
$$[k-1-2l,\ 4,\ 2], \tag{7.11b}$$
$$[k-3-2l,\ 6,\ 2], \tag{7.11c}$$
$$[k-3-2l,\ 4,\ 4], \tag{7.11d}$$
$$[k-3-2l,\ 6,\ 2]\} \tag{7.11e}$$

and so on.

The sets of exponent vectors $\boldsymbol{v}_j$ for various $(k,l)$, with $1 \leq l \leq \frac{k-1}{2}$, are given in Table 7.3. For $l$=3, eqs. (7.11a)-(7.11b) result in very unequal exponents, e.g. with $k$=15, result in [10,2,2] and [8,4,2]. Additionally, eqs. (7.11c)-(7.11e) can also be used, for better distributed exponent values, as long as they are positive. In the example, the latter equations result in exponents [6,6,2] and [6,4,4]. However, we adopt a simplification inspired by the DDR model strategy that only assign low exponents to the more delayed

input samples, therefore just the first equations are used and the latter are disregarded. In the example above, only using eq. (7.11a), just exponents 2 are assigned to the more delayed input samples, resulting in a more economical model. This way, for each $\mathsf{V}^{(k,l)}$, only the first $\eta_k$ vectors are included in the model, i.e., $\eta_k$ is the number of exponent vectors allowed for $k^{th}$-order terms. As presented so far, the set of parameters $\{L_k\}$ controls the polynomial dimensions in each order of the model, similarly to the MP/GMP and WHFB strategy, whereas the sets of memory depths $\{M_k\}$ and exponent vector lengths $\{\eta_k\}$ limit the exponents assigned to delayed input samples, similarly to the DDR strategy.

Table 7.3 – Exponent vectors in the proposed pruning method.

| $k$ | $L_k$ | $l$ | $\mathsf{V}^{(k,l)}$ |
|---|---|---|---|
| 1 | 0 | 0 | $\emptyset$ |
| 3 | 1 | 1 | {[2]} |
| 5 | 2 | 1 | {[4]} |
| | | 2 | {[2,2]} |
| 7 | 3 | 1 | {[6]} |
| | | 2 | {[4,2]} |
| | | 3 | {[2,2,2]} |
| 9 | 4 | 1 | {[8]} |
| | | 2 | {[6,2], [4,4]} |
| | | 3 | {[4,2,2]} |
| | | 4 | {[2,2,2,2]} |
| 11 | 5 | 1 | {[10]} |
| | | 2 | {[8,2], [6,4]} |
| | | 3 | {[6,2,2], [4,4,2]} |
| | | 4 | {[4,2,2,2]} |
| | | 5 | {[2,2,2,2,2]} |
| 13 | 6 | 1 | {[12]} |
| | | 2 | {[10,2], [8,4], [6,6]} |
| | | 3 | {[8,2,2], [6,4,2], [4,4,4]} |
| | | 4 | {[6,2,2,2], [4,4,2,2]} |
| | | 5 | {[4,2,2,2,2]} |
| | | 6 | {[2,2,2,2,2,2]} |

Table 7.3 – Exponent vectors in the proposed pruning method (cont.).

| $k$ | $L_k$ | $l$ | $\mathsf{V}^{(k,l)}$ |
|---|---|---|---|
| 15 | 7 | 1 | {[14]} |
| | | 2 | {[12,2], [10,4], [8,6]} |
| | | 3 | {[10,2,2], [8,4,2], [6,6,2], [6,4,4]} |
| | | 4 | {[8,2,2,2], [6,4,2,2], [4,4,4,2]} |
| | | 5 | {[6,2,2,2,2], [4,4,2,2,2]} |
| | | 6 | {[4,2,2,2,2,2]} |
| | | 7 | {[2,2,2,2,2,2,2]} |

The total number of coefficients assuming $\eta^{(k)} = 1$ is given by:

$$N_{RV_2} = \sum_{\substack{k=1 \\ k\ odd}}^{K} (M_k + 1) \left(\!\!\binom{M_k + 1}{L_k}\!\!\right), \tag{7.12}$$

where $\left(\!\!\binom{M_k+1}{L_k}\!\!\right) = \binom{M_k+L_k}{L_k}$, further reducing the model size when $L_k < \frac{k-1}{2}$.

Let $\mathbf{\Phi}$ be the regression matrix of the model $RV_2$ in eq. (7.8). Let us express $\mathbf{\Phi}$ as the concatenation of the $k^{th}$-order sub-matrices $\mathbf{\Phi}^{(k)}$, with $k = 1, \ldots, K$:

$$\mathbf{\Phi} = [\mathbf{\Phi}^{(1)} \quad \mathbf{\Phi}^{(3)} \ldots \quad \mathbf{\Phi}^{(k)} \ldots \quad \mathbf{\Phi}^{(K)}]. \tag{7.13}$$

Moreover, each of the sub-matrices is composed of disjoint blocks $\mathbf{\Phi}_j^{(k,l)}$, each of which has a fixed value of polynomial dimension $l$ ($1 \leq l \leq L_k$) and of exponent index $j$ in $\mathsf{V}^{(k,l)}$. This choice of partitioning each $\mathbf{\Phi}^{(k)}$ into blocks $\mathbf{\Phi}_j^{(k,l)}$ is based on the reasonable assumption that higher polynomial dimensions $l$ are only included to the model if required, by either manual selection of the initial parameters or a block-oriented sparse regression technique. Thus, the initial model parameter choices $K$, $M_k$, $L_k$ and $\eta_k$ are refined by sparse techniques, such as SGL.

The regression matrix $\mathbf{\Phi}$ and its blocks $\mathbf{\Phi}_j^{(k,l)}$ are implemented as follows:

**Initialize:** Given an initial choice of $K$, $\{M_k\}$, $\{L_k\}$ and $\{\eta_k\}$, for $k = 1, \ldots, K$ ($k$ odd):

**Step 1)** For each $k$:

**Step 2)** For each $l$, from 1 to $L_k$:

**a)** Create the class $\mathcal{M}^{(k,l)}$ containing all distinct sets $\mathsf{M}^{(k,l)}$ of memory indexes $\{m_1, [m_2, m_3, \ldots, m_{l+1}]\}$ from eq. (7.7) , with indexes ranging from zero to $M_k$ and also $m_{l+1} > m_l > \cdots > m_3 > m_2$.

**Step 3)** For each $j$, from 1 to $min(card(\mathsf{V}^{(k,l)}), \eta_k)$:

**a)** Use exponent vector $\boldsymbol{v}_j \in \mathsf{V}^{(k,l)}$ from Table 7.3.

**Step 4)** For $n = 0, \ldots, N-1$:

**a)** The input vector is given by:

$$\tilde{\boldsymbol{u}}(n) = \begin{bmatrix} \tilde{u}(n) & \tilde{u}(n-1) & \ldots & \tilde{u}(n-M_k) \end{bmatrix}. \tag{7.14}$$

**b)** Compute the elements $\tilde{u}(n-m_1)\prod_{i=2}^{l+1}|\tilde{u}(n-m_i)|^{v_j(i-1)}$ of the row vector $\boldsymbol{\phi}_j^{(k,l)}(n)$, using $\tilde{\boldsymbol{u}}(n)$ and the memory indexes $\mathsf{M}^{(k,l)} \in \mathcal{M}^{(k,l)}$. The block $\boldsymbol{\Phi}_j^{(k,l)}$ of the regression matrix is given by:

$$\boldsymbol{\Phi}_j^{(k,l)} = \begin{bmatrix} \boldsymbol{\phi}_j^{(k,l)}(0) \\ \vdots \\ \boldsymbol{\phi}_j^{(k,l)}(N-1) \end{bmatrix}. \tag{7.15}$$

**Step 5)** The complete regression matrix $\boldsymbol{\Phi}$ is the concatenation of the $k^{th}$-order matrices obtained by cycling the previous steps.

**Step 6)** Finally, SGL is applied to the initial regression matrix $\boldsymbol{\Phi}$, retaining only the most relevant sub-blocks in each block $\boldsymbol{\Phi}_j^{(k,l)}$, as detailed in section 7.4.

Table 7.4 shows examples of the proposed model terms with $m_1 = 0$, $m_2 = 1$ and $1 \leq l \leq L_k$, for $L_k = \frac{k-1}{2}$ for each order $k$. Also note that all $\boldsymbol{v}_j \in \mathsf{V}^{(k,l)}$ are included in the example.

Table 7.4 – Examples of polynomials in the proposed pruning method.

| | | Examples | |
|---|---|---|---|
| $k$ | $l$ | **Memory** $(m_1 \vert m_2 \ldots m_{l+1})$ | Polynomials |
| 1 | 0 | 0 | $u(n)$ |
| 3 | 1 | 0\|1 | $u(n)\vert u(n-1)\vert^2$ |
| 5 | 1 | 0\|1 1 | $u(n)\vert u(n-1)\vert^4$ |
| | 2 | 0\|1 2 | $u(n)\vert u(n-1)\vert^2\vert u(n-2)\vert^2$ |
| 7 | 1 | 0\|1 1 1 | $u(n)\vert u(n-1)\vert^6$ |
| | 2 | 0\|1 1 2 | $u(n)\vert u(n-1)\vert^4\vert u(n-2)\vert^2$ |
| | 3 | 0\|1 2 3 | $u(n)\vert u(n-1)\vert^2\ldots\vert u(n-3)\vert^2$ |

Table 7.4 – Examples of polynomials in the proposed pruning method (cont.).

| | | Examples | |
|---|---|---|---|
| $k$ | $l$ | **Memory** $(m_1 \vert m_2 \dots m_{l+1})$ | Polynomials |
| 9 | 1 | $0 \vert 1\ 1\ 1\ 1$ | $u(n)\,\vert u(n-1)\vert^8$ |
| | 2 | $0 \vert 1\ 1\ 1\ 2$ | $u(n)\,\vert u(n-1)\vert^6\,\vert u(n-2)\vert^2$ |
| | 2 | $0 \vert 1\ 1\ 2\ 2$ | $u(n)\,\vert u(n-1)\vert^4\,\vert u(n-2)\vert^4$ |
| | 3 | $0 \vert 1\ 1\ 2\ 3$ | $u(n)\,\vert u(n-1)\vert^4\,\vert u(n-2)\vert^2\,\vert u(n-3)\vert^2$ |
| | 4 | $0 \vert 1\ 2\ 3\ 4$ | $u(n)\,\vert u(n-1)\vert^2 \dots \vert u(n-4)\vert^2$ |
| 11 | 1 | $0 \vert 1\ 1\ 1\ 1\ 1$ | $u(n)\,\vert u(n-1)\vert^{10}$ |
| | 2 | $0 \vert 1\ 1\ 1\ 1\ 2$ | $u(n)\,\vert u(n-1)\vert^8\,\vert u(n-2)\vert^2$ |
| | 2 | $0 \vert 1\ 1\ 1\ 2\ 2$ | $u(n)\,\vert u(n-1)\vert^6\,\vert u(n-2)\vert^4$ |
| | 3 | $0 \vert 1\ 1\ 1\ 2\ 3$ | $u(n)\,\vert u(n-1)\vert^6\,\vert u(n-2)\vert^2\,\vert u(n-3)\vert^2$ |
| | 3 | $0 \vert 1\ 1\ 2\ 2\ 3$ | $u(n)\,\vert u(n-1)\vert^4\,\vert u(n-2)\vert^4\,\vert u(n-3)\vert^2$ |
| | 4 | $0 \vert 1\ 1\ 2\ 3\ 4$ | $u(n)\,\vert u(n-1)\vert^4\,\vert u(n-2)\vert^2 \dots \vert u(n-4)\vert^2$ |
| | 5 | $0 \vert 1\ 2\ 3\ 4\ 5$ | $u(n)\,\vert u(n-1)\vert^2 \dots \vert u(n-5)\vert^2$ |
| 13 | 1 | $0 \vert 1\ 1\ 1\ 1\ 1\ 1$ | $u(n)\,\vert u(n-1)\vert^{12}$ |
| | 2 | $0 \vert 1\ 1\ 1\ 1\ 1\ 2$ | $u(n)\,\vert u(n-1)\vert^{10}\,\vert u(n-2)\vert^2$ |
| | 2 | $0 \vert 1\ 1\ 1\ 1\ 2\ 2$ | $u(n)\,\vert u(n-1)\vert^8\,\vert u(n-2)\vert^4$ |
| | 2 | $0 \vert 1\ 1\ 1\ 2\ 2\ 2$ | $u(n)\,\vert u(n-1)\vert^6\,\vert u(n-2)\vert^6$ |
| | 3 | $0 \vert 1\ 1\ 1\ 1\ 2\ 3$ | $u(n)\,\vert u(n-1)\vert^8\,\vert u(n-2)\vert^2\,\vert u(n-3)\vert^2$ |
| | 3 | $0 \vert 1\ 1\ 1\ 2\ 2\ 3$ | $u(n)\,\vert u(n-1)\vert^6\,\vert u(n-2)\vert^4\,\vert u(n-3)\vert^2$ |
| | 3 | $0 \vert 1\ 1\ 2\ 2\ 3\ 3$ | $u(n)\,\vert u(n-1)\vert^4\,\vert u(n-2)\vert^4\,\vert u(n-3)\vert^4$ |
| | 4 | $0 \vert 1\ 1\ 1\ 2\ 3\ 4$ | $u(n)\,\vert u(n-1)\vert^6\,\vert u(n-2)\vert^2 \dots \vert u(n-4)\vert^2$ |
| | 4 | $0 \vert 1\ 1\ 2\ 2\ 3\ 4$ | $u(n)\,\vert u(n-1)\vert^4\,\vert u(n-2)\vert^4\,\vert u(n-3)\vert^2\,\vert u(n-4)\vert^2$ |
| | 5 | $0 \vert 1\ 1\ 2\ 3\ 4\ 5$ | $u(n)\,\vert u(n-1)\vert^4\,\vert u(n-2)\vert^2 \dots \vert u(n-5)\vert^2$ |
| | 6 | $0 \vert 1\ 2\ 3\ 4\ 5\ 6$ | $u(n)\,\vert u(n-1)\vert^2 \dots \vert u(n-6)\vert^2$ |

Table 7.4 – Examples of polynomials in the proposed pruning method (cont.).

| | | Examples | |
|---|---|---|---|
| $k$ | $l$ | **Memory** $(m_1 \vert m_2 \dots m_{l+1})$ | Polynomials |
| | 1 | $0 \vert 1\ 1\ 1\ 1\ 1\ 1\ 1$ | $u(n)\,\vert u(n-1)\vert^{14}$ |
| | 2 | $0 \vert 1\ 1\ 1\ 1\ 1\ 1\ 2$ | $u(n)\,\vert u(n-1)\vert^{12}\,\vert u(n-2)\vert^{2}$ |
| | 2 | $0 \vert 1\ 1\ 1\ 1\ 1\ 2\ 2$ | $u(n)\,\vert u(n-1)\vert^{10}\,\vert u(n-2)\vert^{4}$ |
| | 2 | $0 \vert 1\ 1\ 1\ 1\ 2\ 2\ 2$ | $u(n)\,\vert u(n-1)\vert^{8}\,\vert u(n-2)\vert^{6}$ |
| | 3 | $0 \vert 1\ 1\ 1\ 1\ 1\ 2\ 3$ | $u(n)\,\vert u(n-1)\vert^{10}\,\vert u(n-2)\vert^{2}\,\vert u(n-3)\vert^{2}$ |
| | 3 | $0 \vert 1\ 1\ 1\ 1\ 2\ 2\ 3$ | $u(n)\,\vert u(n-1)\vert^{8}\,\vert u(n-2)\vert^{4}\,\vert u(n-3)\vert^{2}$ |
| | 3 | $0 \vert 1\ 1\ 1\ 2\ 2\ 2\ 3$ | $u(n)\,\vert u(n-1)\vert^{6}\,\vert u(n-2)\vert^{6}\,\vert u(n-3)\vert^{2}$ |
| | 3 | $0 \vert 1\ 1\ 1\ 2\ 2\ 3\ 3$ | $u(n)\,\vert u(n-1)\vert^{6}\,\vert u(n-2)\vert^{4}\,\vert u(n-3)\vert^{4}$ |
| 15 | 4 | $0 \vert 1\ 1\ 1\ 1\ 2\ 3\ 4$ | $u(n)\,\vert u(n-1)\vert^{8}\,\vert u(n-2)\vert^{2}\dots\vert u(n-4)\vert^{2}$ |
| | 4 | $0 \vert 1\ 1\ 1\ 2\ 2\ 3\ 4$ | $u(n)\,\vert u(n-1)\vert^{6}\,\vert u(n-2)\vert^{4}\,\vert u(n-3)\vert^{2}\,\vert u(n-4)\vert^{2}$ |
| | 4 | $0 \vert 1\ 1\ 2\ 2\ 3\ 3\ 4$ | $u(n)\,\vert u(n-1)\vert^{4}\,\vert u(n-2)\vert^{4}\,\vert u(n-3)\vert^{4}\,\vert u(n-4)\vert^{2}$ |
| | 5 | $0 \vert 1\ 1\ 1\ 2\ 3\ 4\ 5$ | $u(n)\,\vert u(n-1)\vert^{6}\,\vert u(n-2)\vert^{2}\dots\vert u(n-5)\vert^{2}$ |
| | 5 | $0 \vert 1\ 1\ 2\ 2\ 3\ 4\ 5$ | $u(n)\,\vert u(n-1)\vert^{4}\,\vert u(n-2)\vert^{4}\,\vert u(n-3)\vert^{2}\dots\vert u(n-5)\vert^{2}$ |
| | 6 | $0 \vert 1\ 1\ 2\ 3\ 4\ 5\ 6$ | $u(n)\,\vert u(n-1)\vert^{4}\,\vert u(n-2)\vert^{2}\dots\vert u(n-6)\vert^{2}$ |
| | 7 | $0 \vert 1\ 2\ 3\ 4\ 5\ 6\ 7$ | $u(n)\,\vert u(n-1)\vert^{2}\dots\vert u(n-7)\vert^{2}$ |

## 7.4 PROPOSED MODEL SIZING

The regression matrix $\mathbf{\Phi}$ of the pruned-Volterra model is composed of the $k^{th}$-order sub-matrices $\mathbf{\Phi}^{(k)}$, $k = 1, \dots, K$, with $k$ odd, each of which composed of disjoint blocks $\mathbf{\Phi}_j^{(k,l)}$, as illustrated in Figure 7.1. Note that in some cases the indexes $j$ and/or $l$ can be omitted.

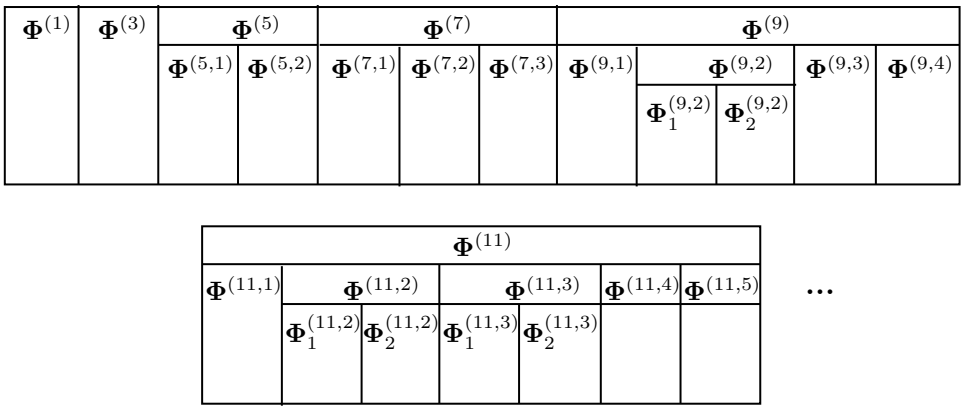



Figure 7.1 – Proposed structure in blocks $\mathbf{\Phi}_j^{(k,l)}$

For model sizing using group-wise LASSO selection, each block $\mathbf{\Phi}_j^{(k,l)}$ is further divided into sub-blocks. Thus, Group-LASSO and SGL techniques select the most relevant sub-blocks within each block of the model. The strategy adopted here is to divide each block of order $k$ (except for $k$=1) into sub-blocks of length $M_k + 1$. Given the parameters $k$, $M_k$ and $L_k$, the number of sub-blocks of length $M_k + 1$ in a block $(k, l)$ is:

$$N_{sb}^{(k,l)} = \binom{L_k - 1}{l - 1}\binom{M_k + 1}{l}. \tag{7.16}$$

Thus, in eq. (7.16), the number of sub-blocks depends on $M_k$, $L_k$ and $l$; for example, with parameters $M_k$=8 and $L_k$=3, the numbers of sub-blocks per block $(k, l)$ are $N_{sb}^{(k,l)} =$ 9 $(l = 1)$; 72 $(l = 2)$ and 84 $(l = 3)$ sub-blocks, i.e., in total, 165 sub-blocks for each $\mathbf{\Phi}_j^{(k,l)}$. Now, for parameters $M_k$=8 and $L_k$=2, the number of sub-blocks per block $(k, l)$ is $N_{sb}^{(k,l)} =$ 9 $(l = 1)$ and 36 $(l = 2)$, i.e., 45 sub-blocks in total.

Next, the SGL technique, described in section 6.1, is applied to the regression matrix $\mathbf{\Phi}$, retaining only the most relevant sub-blocks in each block. After the model sizing and sparse estimation task is completed, the selected sub-blocks indexes, regressor indexes and corresponding model coefficients are copied to the predistorter.

### 7.4.1 Kronecker Product Implementation

In this subsection, we illustrate an efficient way to compute separately each regression block $\mathbf{\Phi}_j^{(k,l)}$ in eq. (7.15). Firstly, we propose a modified truncated Kronecker operator, $\oslash_{-}^{q}$, that allows to implement polynomials $(k, l)$ and is also presented in appendix B.

For each $k$, the input vector $\tilde{\boldsymbol{u}}(n)$ at instant $n$ is adjusted for length $M_k + 1$, as in eq. (7.14). Then we define the squared amplitude input vector, as follows:

$$\tilde{\boldsymbol{u}}^{(2)}(n) = \begin{bmatrix} |\tilde{u}(n)|^2 & |\tilde{u}(n-1)|^2 & \dots & |\tilde{u}(n-M_k)|^2 \end{bmatrix}^T. \tag{7.17}$$

Next, the modified truncated Kronecker product of $\tilde{\boldsymbol{u}}^{(2)}(n)$ by itself is given by:

$$\oslash_{-}^{2}\tilde{\boldsymbol{u}}^{(2)}(n) \;=\; \begin{bmatrix} |\tilde{u}(n)|^2\tilde{\boldsymbol{u}}^{(2)}(n-1) \\ |\tilde{u}(n-1)|^2\tilde{\boldsymbol{u}}^{(2)}(n-2) \\ \vdots \\ |\tilde{u}(n-M_k+1)|^2\tilde{\boldsymbol{u}}^{(2)}(n-M_k) \end{bmatrix}, \tag{7.18}$$

where:

$$\tilde{\boldsymbol{u}}^{(2)}(n-i) = \begin{bmatrix} |\tilde{u}(n-i)|^2 & |\tilde{u}(n-i-1)|^2 & \dots & |\tilde{u}(n-M_k)|^2 \end{bmatrix}^T. \tag{7.19}$$

The high order modified truncated Kronecker product of $\tilde{\boldsymbol{u}}^{(2)}(n)$ is given, recursively, by:

$$\oslash_{-}^{q}\tilde{\boldsymbol{u}}^{(2)}(n) = \begin{bmatrix} |\tilde{u}(n)|^2 \oslash_{-}^{q-1}\tilde{\boldsymbol{u}}^{(2)}(n-1) \\ |\tilde{u}(n-1)|^2 \oslash_{-}^{q-1}\tilde{\boldsymbol{u}}^{(2)}(n-2) \\ \vdots \\ |\tilde{u}(n-M_k+q-1)|^2 \oslash_{-}^{q-1}\tilde{\boldsymbol{u}}^{(2)}(n-M_k+q-2) \end{bmatrix}. \tag{7.20}$$

Finally, the proposed expressions for the rows of $\boldsymbol{\Phi}_j^{(k,l)}$, for $k \geq 3$, are given by:

- For $j = 1$:

$$\boldsymbol{\Phi}_1^{(k,l)}(n) = \tilde{\boldsymbol{u}}(n) \otimes \left[\tilde{\boldsymbol{u}}^{(k+1-2l)}(n) \oslash_{-} (\oslash_{-}^{l-1}\tilde{\boldsymbol{u}}^{(2)}(n))\right], \tag{7.21}$$

where $\boldsymbol{\Phi}_j^{(1)}(n) = \tilde{\boldsymbol{u}}(n)$, $\oslash_{-}^{0}\tilde{\boldsymbol{u}}^{(2)}(n) = \mathbf{1}$ and $\oslash_{-}^{1}\tilde{\boldsymbol{u}}^{(2)}(n) = \tilde{\boldsymbol{u}}^{(2)}(n)$.

Note that $(k+1-2l)$ is even and $\tilde{\boldsymbol{u}}^{(k+1-2l)}(n)$ is an element-wise powered version of $\tilde{\boldsymbol{u}}^{(2)}(n)$, thus:

$$\tilde{\boldsymbol{u}}^{(k+1-2l)}(n) \oslash_{-} (\oslash_{-}^{l-1}\tilde{\boldsymbol{u}}^{(2)}(n)) = \begin{bmatrix} |\tilde{u}(n)|^{k+1-2l} \oslash_{-}^{l-1}\tilde{\boldsymbol{u}}^{(2)}(n) \\ |\tilde{u}(n-1)|^{k+1-2l} \oslash_{-}^{l-1}\tilde{\boldsymbol{u}}^{(2)}(n-1) \\ \vdots \\ |\tilde{u}(n-M_k)|^{k+1-2l} \oslash_{-}^{l-1}\tilde{\boldsymbol{u}}^{(2)}(n-M_k) \end{bmatrix}. \tag{7.22}$$

- For $j = 2$:

$$\boldsymbol{\Phi}_2^{(k,l)}(n) = \tilde{\boldsymbol{u}}(n) \otimes \left[\tilde{\boldsymbol{u}}^{(k-1-2l)}(n) \oslash_{-} \tilde{\boldsymbol{u}}^{(4)}(n) \oslash_{-} (\oslash_{-}^{l-2}\tilde{\boldsymbol{u}}^{(2)}(n))\right]. \tag{7.23}$$

Appendix B also provides a few examples. The next section presents the experimental results.

## 7.5 EXPERIMENTAL VALIDATION

Measured data for PA LMM, obtained with the experimental test set-up described in section 3.1, is used to validate the DPD models proposed in this chapter. The proposed pruned-Volterra models, $RV_1$ in eq. (7.4) and $RV_2$ in eq. (7.8), are compared in this section

to the triangular FV and MV models from eqs. (2.4) and (2.7). The proposed models are also compared to the VP, LV and $MV_{\triangle}$ LASSO models, as well as to the MP, GMP, WHFB and sparse WHFB models. The $MV_{\triangle}$ LASSO model corresponds to the sparse estimation of the $MV_{\triangle}$ coefficients, similarly to the sparse full-Volterra estimation in (KEKATOS; GIANNAKIS, 2011), (ZENTENO et al., 2015), (ABDELHAFIZ et al., 2014). Note that the estimation cost of running LASSO over the $FV_{\triangle}$ and $MV_{\triangle}$ models is very high, thus this approach becomes less scalable than running LASSO over pruned-Volterra models. Also note that the best values of the sparse parameter, $\lambda_o$, in the following sparse models, are also obtained using the AIC score in eq. (3.4).

### 7.5.1 $RV_1$ and $RV_2$ Results

The original dense $RV_1$ and $RV_2$ models are considered in this item, i.e. sparsity is not yet explored, and the model coefficients are estimated by OLS. Note that the estimation cost of the $FV_{\triangle}$ and $MV_{\triangle}$ models grow fast with the NL order $K$, so this study is limited to $K$=9. Table 7.5 shows the DPD performance of the models with parameters $K = 7$, $M = 5$ and sets $\{M_k\}$ as shown in the table. Additional parameters for the $RV_2$ model are $L_k = [0, 1, 2, 2]$ and $\eta_k = 1$, for $1 \leq k \leq K$. As observed on the tables, both the $FV_{\triangle}$ and $MV_{\triangle}$ model sizes grow rapidly with the parameters, due to the doubly-combinatorial in eq. (2.7). Even reducing the memory depths of the higher order terms, the $MV_{\triangle}$ model still suffers from the curse of dimensionality. In turn, both the $RV_1$ and $RV_2$ models replace the doubly-combinatorial by eq. (7.12), substantially reducing the model size. Finally, $RV_2$ model is able to further reduce model redundancies, by allowing the user to set the polynomial dimensions and exponent vectors for each NL order.

Note that, in this subsection, the selection of polynomial dimensions and exponent vectors is performed manually, through $\{L_k\}$ and $\eta_k$, whereas, in the next subsection, it is performed by the LASSO and group-wise LASSO algorithms. The table also includes the VP and LV models, from chapter 4, and the $MV_{\triangle}$ LASSO model, i.e., the sparse estimation of $MV_{\triangle}$ coefficients. Next, Table 7.6 shows the DPD performance with $K = 9$, $M = 4$, $M_k \leq 4$ and for the $RV_2$ model, $L_k = [0, 1, 2, 2, 2]$ and $\eta_k = 1$. From the results so far, the proposed $RV_2$ model approximate the DPD performance of the $MV_{\triangle}$ model, while being much more affordable for practical implementation than the $FV_{\triangle}$ and $MV_{\triangle}$ structures. The tables also show that the proposed $RV_2$ model results in lower numbers of coefficients than the VP, LV and $MV_{\triangle}$ LASSO, for the same performance. Additionally, Figure 7.2 shows the constellation diagrams for the $MV_{\triangle}$ and $RV_2$ DPD output signals, for the model parameters $K = 9$ and $M_k = [4, 4, 4, 3, 3]$. As can be seen, their constellation diagrams are very similar, thus, their in-band performance are very close. The EVM of the $MV_{\triangle}$ and $RV_2$ output signals are, respectively, 1.8% and 2.1%. These results validate

the pruned-Volterra $RV_2$ model.

Table 7.5 – DPD performance for LMM PA with $FV_\triangle$, $MV_\triangle$, $RV_1$ and $RV_2$ ($K$=7, $N$= 50,000).

| | **Num. coeffs.** | **Max. ACPR (dBc)** | **NMSE (dB)** |
|---|---|---|---|
| No DPD | - | -23.2 | 3.3 |
| $FV_\triangle$ $(M = 5)$ | 8,364 | -46.3 | -35.0 |
| $MV_\triangle$ $(M_k = [5, 5, 5, 4])$ | 3,758 | -44.8 | -34.2 |
| $MV_\triangle$ $(M_k = [5, 5, 5, 3])$ | 2,008 | -44.4 | -33.5 |
| $RV_1$ $(M_k = [5, 5, 5, 5])$ | 504 | -43.9 | -32.9 |
| $RV_1$ $(M_k = [5, 5, 5, 4])$ | 343 | -43.6 | -32.5 |
| $RV_2$ $(M_k = [5, 5, 5, 5])$ | 294 | -43.8 | -32.5 |
| $RV_2$ $(M_k = [5, 5, 5, 4])$ | 243 | -43.6 | -32.4 |
| VP (ALS) $(M_k = [5, 5, 4, 2])$ $(R_k = [1, 13, 45, 32])$ | 405 | -43.0 | -33.2 |
| LV $(M_k = [5, 5, 4, 2])$ $(F_k = [5, 5, 3, 1])$ | 352 | -42.9 | -33.3 |
| $MV_\triangle$ LASSO $(M_k = [5, 5, 5, 4])$ $(\lambda = 24)$ | 281 | -42.7 | -32.2 |

Table 7.6 – DPD performance for LMM PA with $FV_\triangle$, $MV_\triangle$, $RV_1$ and $RV_2$ ($K$=9, $N$= 50,000).

| | **Num. coeffs.** | **Max. ACPR (dBc)** | **NMSE (dB)** |
|---|---|---|---|
| No DPD | - | -23.2 | 3.3 |
| $FV_\triangle$ $(M = 4)$ | 11,875 | -45.4 | -34.9 |
| $MV_\triangle$ $(M_k = [4, 4, 4, 3, 3])$ | 3,265 | -44.5 | -34.1 |
| $MV_\triangle$ $(M_k = [4, 4, 4, 3, 2])$ | 1,620 | -44.0 | -33.6 |
| $RV_1$ $(M_k = [4, 4, 4, 4, 4])$ | 630 | -44.2 | -33.7 |
| $RV_1$ $(M_k = [4, 4, 4, 3, 3])$ | 325 | -43.3 | -32.8 |
| $RV_2$ $(M_k = [4, 4, 4, 4, 4])$ | 255 | -44.1 | -33.1 |
| $RV_2$ $(M_k = [4, 4, 4, 3, 3])$ | 185 | -43.1 | -32.4 |
| VP (ALS) $(M_k = [4, 4, 4, 3, 2])$ $(R_k = [1, 9, 40, 32, 32])$ | 474 | -43.3 | -33.3 |
| LV $(M_k = [4, 4, 4, 3, 2])$ $(F_k = [4, 4, 3, 2, 1])$ | 461 | -43.3 | -33.8 |
| $MV_\triangle$ LASSO $(M_k = [4, 4, 4, 3, 3])$ $(\lambda_o = 20)$ | 354 | -43.5 | -32.5 |

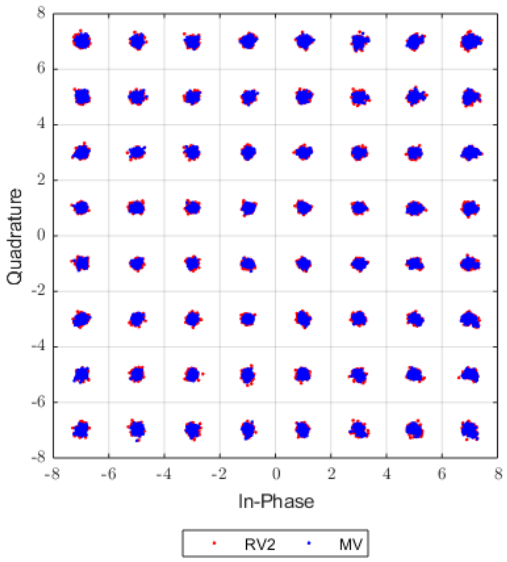



Figure 7.2 – Constellation diagram for LMM PA with $RV_2$ model ($K$=9).

### 7.5.2 Sparse $RV_2$ Model Results

Now, the DPD performance of the $RV_1$ and sparse $RV_2$ models are compared using parameters $K$=9, $M_k = [11, 11, 8, 6, 5]$ and, for the $RV_2$ model, $L_k = [0, 1, 2, 2, 2]$ and $\eta_k = 1$ in Table 7.7. Group-wise LASSO optimization is applied to the $RV_2$ model, to show the model size reduction that can be achieved. In order to run the parsimonious selection of the most relevant $RV_2$ regressors blocks, firstly, $\mathbf{\Phi}$ is standardized and the output signal is centered. Then, Group-LASSO and SGL algorithms are run for a pre-defined set of the corresponding sparse parameter values and the best values are determined using the AIC score in eq. (3.4). Figure 7.3 illustrates the residual spectral regrowth achieved with the $RV_1$, $RV_2$, $RV_2$ LASSO, $RV_2$ Group-LASSO and $RV_2$ SGL models. It is concluded from the results that the proposed sparse models are able to approximate the DPD performance of the $FV_\triangle$ and $MV_\triangle$ models being much more affordable in terms of running cost. Figure 7.4 shows the regressors selected by LASSO, Group-LASSO and SGL, illustrating the difference between the LASSO and group-wise LASSO selection criteria, i.e., that sub-blocks are first selected in the group-wise LASSO strategies. The initial size of 883 coefficients of the $RV_2$ model is reduced to approximately 330 coefficients using sparse techniques. The Group-LASSO and SGL algorithms selected, respectively 41 and 50 sub-blocks, out of 107 sub-blocks of the $RV_2$ model.

Table 7.7 – DPD performance for LMM PA with $RV_1$ and $RV_2$ ($K$=9, $N$= 50,000).

| | **Num. coeffs.** | **Max. ACPR (dBc)** | **NMSE (dB)** |
|---|---|---|---|
| No DPD | - | -23.2 | 3.3 |
| $RV_1$ ($M_k = [11, 11, 8, 6, 5]$) | 1,905 | -48.4 | -34.9 |
| $RV_2$ ($M_k = [11, 11, 8, 6, 5]$) | 883 | -48.2 | -34.2 |
| $RV_2$ LASSO ($\lambda_o = 7$) | 332 | -46.9 | -33.9 |
| $RV_2$ Group-LASSO ($\lambda_o$ =40) | 337 | -47.0 | -33.4 |
| $RV_2$ SGL ($\lambda_o$ =7, $\alpha_o$ =8) | 324 | -47.3 | -33.5 |

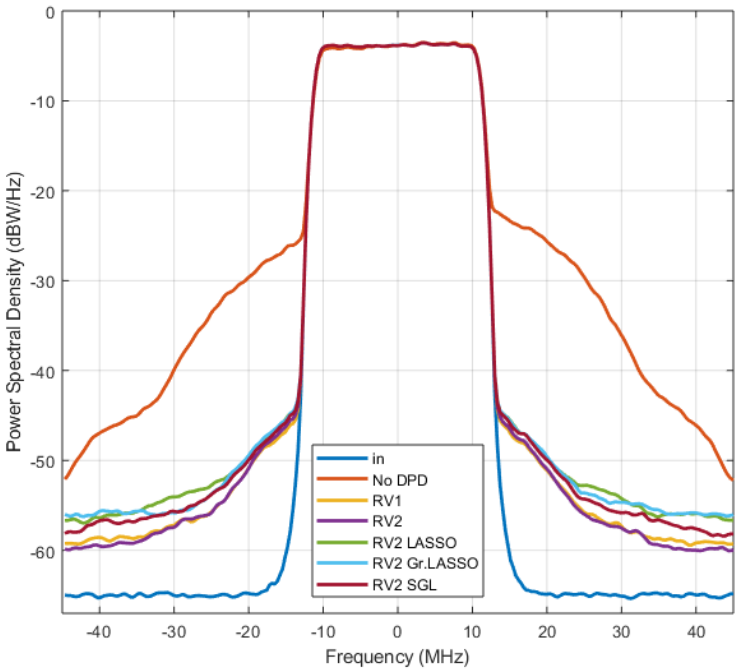



Figure 7.3 – Residual spectral regrowth for LMM PA with sparse $RV_2$ model ($K$=9).

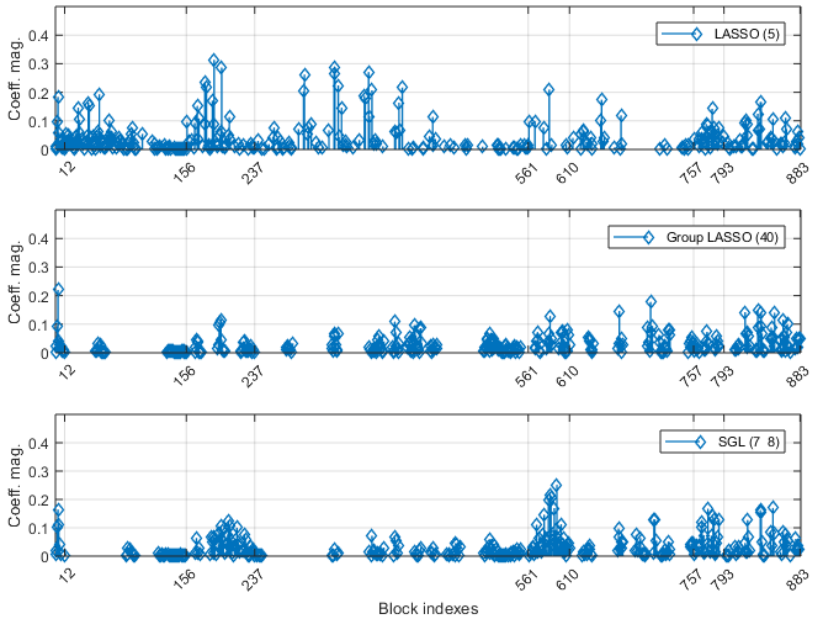



Figure 7.4 – Sparse selection of the $RV_2$ model ($K$=9).

### 7.5.3 Comparison with the WHFB Model

Finally, the $RV_2$ model is compared to the MP, GMP and WHFB models using experimental data measured for PA LMM. Table 7.8 compares the DPD performance of the $RV_2$ LASSO, $RV_2$ Group-LASSO and $RV_2$ SGL, with $K$=15, $M_k = [13, 13, 10, 10, 9, 7, 7, 7]$, $L_k = [0, 1, 2, 2, 2, 1, 1, 1]$ and $\eta_k = 1$, the MP, GMP and WHFB LASSO, WHFB Group-LASSO and WHFB SGL from chapter 6, all for approximately the same number of coefficients. The corresponding residual spectral regrowth is illustrated in Figure 7.5. The results show that the proposed model improves the ACPR performance of the GMP model by 3 to 4 dB, with the same number of coefficients. We can also conclude that the WHFB and sparse WHFB models are more cost-effective for LMM PA than the $RV_2$ models. However, the $RV_2$ models are expected to cover a wider range of operational/environmental conditions, due to their flexibility. Figure 7.6 shows the regressors, selected by LASSO, and/or sub-blocks selected by Group-LASSO and SGL. The initial size of 2404 coefficients of the $RV_2$ model is reduced to approximately 335 coefficients using sparse techniques. The Group-LASSO and SGL algorithms selected, respectively, 33 and 50 sub-blocks, out of 226 sub-blocks of the initial $RV_2$ model.

Table 7.8 – DPD performance for LMM PA with WHFB, sparse WHFB and $RV_2$, ($K$=15 and $N$= 50,000).

| | **Num. coeffs.** | **Max. ACPR (dBc)** | **NMSE (dB)** |
|---|---|---|---|
| No DPD | - | -23.2 | 3.3 |
| MP OLS | 40 | -42.0 | -29.5 |
| GMP OLS | 360 | -44.7 | -32.7 |
| WHFB OLS | 1320 | -53.3 | -33.9 |
| WHFB LASSO ($\lambda_o$=1) | 356 | -49.4 | -33.5 |
| WHFB Group-LASSO ($\lambda_o$ =37) | 414 | -45.5 | -33.2 |
| WHFB SGL ($\lambda_o$=12, $\alpha_o$=7) | 380 | -49.2 | -33.6 |
| WHFB Proposed alg. ($\lambda_o$=0.7) | 350 | -48.6 | -33.5 |
| $RV_2$ OLS | 2404 | -52.5 | -35.1 |
| $RV_2$ LASSO ($\lambda_o$=7) | 338 | -48.8 | -33.6 |
| $RV_2$ Group-LASSO ($\lambda_o$ =25) | 338 | -48.7 | -33.1 |
| $RV_2$ SGL ($\lambda_o$=9.5, $\alpha_o$=9) | 335 | -47.5 | -33.2 |

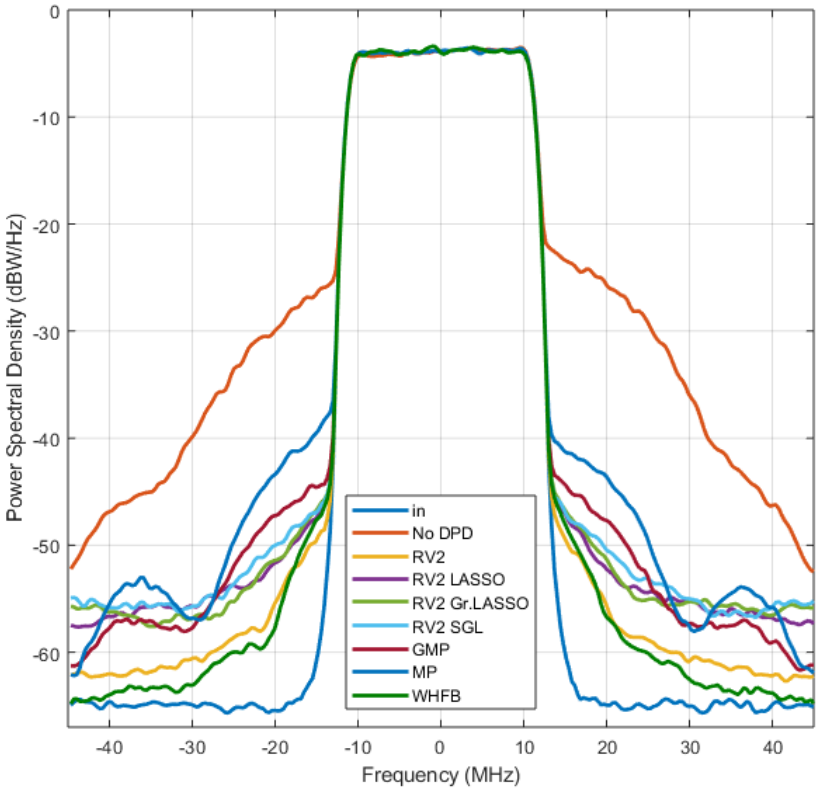



Figure 7.5 – Residual spectral regrowth with sparse $RV_2$ model ($K$=15).

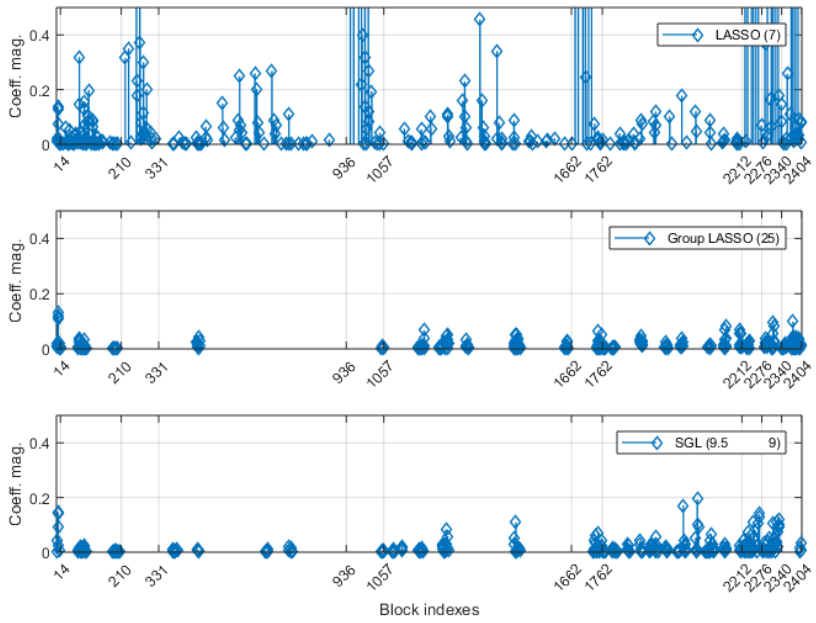



Figure 7.6 – Sparse selection of $RV_2$ model ($K$=15).

## 7.6 CONCLUSIONS

This chapter proposed a flexible and economical pruned-Volterra model that shows significantly reduced running cost comparing to the triangular FV model, while being more flexible than the MP/GMP and DDR models. From the experimental results of section 7.5.3, we observe that the $RV_2$ model was able to achieve a DPD performance for the PA LMM condition very close to that obtained with the WHFB model, with only the latter being designed based on prior information of PA LMM. The advantages of the proposed sparse $RV_2$ model are summarized as follows:

- The model is able to include a mix of MP, GMP, DDR, WHFB and higher dimension (up to $L_k$) polynomials that are selected depending on the operational/environmental conditions of the PA.
- Its initial parameters $\{M_k\}$, $\{L_k\}$ and $\eta_k$ can be defined separately for each NL order, avoiding the curse of dimensionality.
- Its model structure is modular, thus enables model sizing by the selection of the most relevant sub-blocks. After model sizing is complete, only the coefficients and indexes of the selected sub-blocks/ regressors are passed to the DPD predistorter.

The complete experimental validation of the $RV_2$ model through other PAs and operational conditions remains as a future work.

# 8 CONCLUSIONS AND PERSPECTIVES

This thesis investigated DSP-based compensation techniques for RF analog impairments in modern wideband wireless transmitters, including the non-idealities caused by NL PAs and LMM that degrade the spectral purity and modulation quality. Several practical scenarios of PAs subject to LMM motivated the research for more advanced DPD behavioural models, able to overcome the limitations of polynomial models such as the MP and GMP, while simultaneously being less complex than existing PA LMM DPD approaches. In chapter 5, the WHFB polynomial was proposed as a simplified DPD behavioural model for LMM PAs, being justified by the Volterra series analysis and experimentally validated in chapter 6. This model was also extended to the joint compensation of IQM imbalance.

The high dimensionality of the proposed DPD model was addressed in chapter 6, using parsimonious, sparse estimation techniques, such as LASSO and block-wise LASSO extensions. It was shown that using LASSO and block-wise LASSO, the number of required coefficients is significantly reduced, thus reducing the DPD filter length and, proportionally, the DPD running cost. Additionally, block-oriented selection techniques, such as Group-LASSO and Sparse-group LASSO, can also be applied for the model dimensioning, i.e., setting the values of the model parameters. Finally, a simplified, approximate technique was proposed in chapter 6, in which the most relevant WHFB blocks are selected prior to running the LASSO, resulting in lower estimation cost than with LASSO. The WHFB and sparse WHFB models were experimentally validated in section 6.5, by comparing their performance to other models, including to the Volterra-PARAFAC and Laguerre-Volterra models from chapter 4. The extension of the WHFB model for the joint compensation of IQM imbalance is proposed in appendix A, but was not experimentally validated and remains as a future work.

In chapter 7, a Volterra pruning strategy is proposed that leads to a flexible and modular memory polynomial, in which the model parameters are chosen independently for each NL order. Group LASSO was used to select the relevant blocks of regressors in the model. This flexible approach is expected to accurately describe a wide range of PA operational/ environmental conditions.

## 8.1 PERSPECTIVES AND FUTURE WORK

This thesis dealt with techniques, such as behavioural modelling, Volterra model pruning, linear and non-linear least squares regression, CS-based sparse selection and tensor decomposition, that can also be applied in a variety of analog RF impairments, other PA operational/ environmental conditions and wireless transceiver configurations (e.g., MIMO). Moreover, the techniques above, together with other ones, such as SVD/ PCA, ANN, low-rank approximations, etc., are at the intersection of data-driven computational methods, optimization and applied mathematics and have been widely employed to model, predict and control non-linear, dynamic and high-dimensional complex systems in numerous modern applications.

## 8.2 COMPARISON TABLES

This section provides in Tables 8.1, 8.2 and 8.3 an overview of the Pros and Cons of the model structures and identification techniques discussed in the thesis.

Table 8.1 – Comparison of behavioural models.

| Model | Pros | Cons |
|---|---|---|
| Full-Volterra | - Universal approximator for mild NL mappings<br>- Model is linear in the parameters<br>- Inherit stability from FIR filters<br>- The model complexity can be reduced by employing the PARAFAC decomposition and Laguerre expansion | - Curse of dimensionality, i.e., impractical amount of parameters to be estimated<br>- Computational cost<br>- Data over-fitting<br>- Regression matrix becomes ill-conditioned |

Table 8.1 – Comparison of behavioural models (cont.).

| Model | Pros | Cons |
|---|---|---|
| Memory-polynomials | - Lower computational cost<br>- Scalability<br>- Reduced amount of parameters to be estimated | - Loss of generality<br>- Cannot guarantee that the most relevant NL interactions of the system are captured by the model<br>- Choosing the memory polynomial that, with a given complexity, best describes an unknown system is challenging<br>- Model sizing is based on try-and-error or heuristic methods |

Table 8.2 – Comparison of memory polynomial models.

| Model | Pros | Cons |
|---|---|---|
| MP | - Only two parameters are required to define the model<br>- Small number of coefficients | - Inability to suppress off-diagonal terms may reduce the model performance when these terms are non negligible |
| GMP | - Better performance than MP, by adding off-diagonal terms to the model<br>- In general, satisfactory trade-off between complexity and accuracy | - Up to eight parameters are required, including leading and lagging NL orders and memory lengths<br>- Model performance degrades under specific operational/ environmental conditions, such as LMM PA and IQM imbalance |
| WHFB | - Better performance than MP and GMP models, useful under specific operational/ environmental conditions, such as PA LMM<br>- Only four parameters are required to define the model<br>- The model complexity can be reduced by CS-based sparse estimation | - Computationally intensive<br>- Regression matrix may become ill-conditioned |

Table 8.3 – Comparison of estimation techniques.

| Estimation Technique | Pros | Cons |
|---|---|---|
| OLS | - Linear technique<br>- Convex cost function | - Unable to search for parsimonious, sparse solutions for over-parametrized models<br>- Large regression matrices may result in numerical instability and ill-conditioning |
| LASSO | - Convex and linear optimization<br>- Able to explore sparsity of the original regression matrix<br>- Lower computational cost than modelling techniques that require NL optimization<br>- Unstructured model pruning, not requiring *a priori* knowledge of the system or sensitivity analysis<br>- Smoothly trades-off between accuracy and model size, by varying the sparsity parameter<br>- Allows higher NL orders and memory depths to be explored, in an economical way | - Tuning the sparsity parameter requires cross validation or information criteria<br>- Convex relaxation of the $\ell-0$ norm optimization problem may lead to some performance degradation<br>- LASSO estimation is more costly than OLS, but less costly than NL optimization |

Table 8.3 – Comparison of estimation techniques (cont.).

| **Estimation Technique** | **Pros** | **Cons** |
|---|---|---|
| Volterra-PARAFAC | - Able to reduce the number of parameters, while approximating the original FV kernels | - The estimation algorithms are nonlinear and the ALS algorithm requires that the FV kernels are firstly estimated |
| | - PARAFAC estimation can be performed directly from input-output data by iterative algorithms, such as CLMS, NLS LM | - PARAFAC decomposition is a non-convex NL optimization problem, highly dependent on the initialization, subject to find convergence issues and local minima |
| | - Tensorlab and other tools are capable of estimating the tensor rank | - Tensor rank estimation and PARAFAC decomposition are time consuming |
| | - *A priori* physical knowledge of the system to be modelled is not required | |
| Laguerre-Volterra | - Able to reduce the number of parameters, while approximating the FV performance | - OBF pole selection is a non-convex NL optimization problem, highly dependent on the initialization, subject to local minima and convergence not guaranteed |
| | - In this thesis, presented better scalability than the PARAFAC approach when the Volterra parameters increase | |
| | - Allow to explore higher NL orders and memory lengths, in a more economical way than the FV model | |

# BIBLIOGRAPHY

ABDELHAFIZ, A. et al. Digital predistortion of LTE-A power amplifiers using compressed sampling based unstructured pruning of Volterra series. *IEEE Transactions on Microwave Theory and Techniques*, v. 62, Sep. 2014.

ABDELHAFIZ, A. H. et al. A PSO based memory polynomial predistorter with embedded dimension estimation. *IEEE Transactions on Broadcasting*, v. 59, n. 4, p. 665–673, Dec. 2013. ISSN 0018-9316.

ABIDI, A. Direct-conversion radio transceivers for digital communications. In: *Proceedings ISSCC '95 - International Solid-State Circuits Conference*. [S.l.: s.n.], 1995. p. 186–187.

AKAIKE, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, v. 19, n. 6, p. 716–723, Dec. 1974. ISSN 0018-9286.

AKMAL, M. et al. The effect of baseband impedance termination on the linearity of GaN HEMTs. In: *The 40th European Microwave Conference*. [S.l.: s.n.], 2010. p. 1046–1049.

ALBERT, J. Algebraic properties of bag data types. In: *Proceedings of the 17th International Conference on Very Large Data Bases*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1991. (VLDB '91), p. 211–219. ISBN 1558601503.

ALSHARIF, M. H.; NORDIN, R. Evolution towards fifth generation (5G) wireless networks: Current trends and challenges in the deployment of millimetre wave, massive MIMO, and small cells. *Telecommunication Systems*, v. 64, n. 4, p. 617–637, Apr. 2017. ISSN 1572-9451. Disponível em: <https://doi.org/10.1007/s11235-016-0195-x>.

AMIN, S.; HANDEL, P.; RONNOW, D. Digital predistortion of single and concurrent dual-band radio frequency GaN amplifiers with strong nonlinear memory effects. *IEEE Transactions on Microwave Theory and Techniques*, v. 65, p. 2453–2464, Jul. 2017.

ANTTILA, L.; HANDEL, P.; VALKAMA, M. Joint mitigation of power amplifier and I/Q modulator impairments in broadband direct-conversion transmitters. *IEEE Transactions on Microwave Theory and Techniques*, v. 58, n. 4, p. 730–739, Apr. 2010.

ANTTILA, L.; VALKAMA, M.; RENFORS, M. Circularity-based I/Q imbalance compensation in wideband direct-conversion receivers. *IEEE Transactions on Vehicular Technology*, v. 57, n. 4, p. 2099–2113, Jul. 2008.

ANTTILA, L.; VALKAMA, M.; RENFORS, M. Frequency-selective I/Q mismatch calibration of wideband direct-conversion transmitters. *IEEE Transactions on Circuits and Systems II: Express Briefs*, v. 55, n. 4, p. 359–363, Apr. 2008.

ARLOT, S.; CELISSE, A. A survey of cross-validation procedures for model selection. *Statist. Surv.*, The American Statistical Association, the Bernoulli Society, the Institute of Mathematical Statistics, and the Statistical Society of Canada, v. 4, p. 40–79, Jul. 2010. Disponível em: <https://doi.org/10.1214/09-SS054>.

ARMADA, A. G. Understanding the effects of phase noise in orthogonal frequency division multiplexing (OFDM). *IEEE Transactions on Broadcasting*, v. 47, n. 2, p. 153–159, Jun. 2001.

BANKS, H.; JOYNER, M. L. AIC under the framework of least squares estimation. *Applied Mathematics Letters*, v. 74, p. 33 – 45, Dec. 2017. ISSN 0893-9659. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0893965917301623>.

BARRADAS, F. M. et al. Polynomials and LUTs in PA behavioral modeling: A fair theoretical comparison. *IEEE Transactions on Microwave Theory and Techniques*, v. 62, n. 12, p. 3274–3285, Dec. 2014.

BARRADAS, F. M. et al. Power, linearity, and efficiency prediction for MIMO arrays with antenna coupling. *IEEE Transactions on Microwave Theory and Techniques*, v. 65, n. 12, p. 5284–5297, Dec. 2017. ISSN 0018-9480.

BERG, M.; SONKKI, M.; SALONEN, E. Experimental study of hand and head effects to mobile phone antenna radiation properties. In: *3rd European Conference on Antennas and Propagation*. [S.l.: s.n.], 2009. p. 437–440.

BEZOOIJEN, A. van; MAHMOUDI, R.; ROERMUND, A. van. Adaptive methods to preserve power amplifier linearity under antenna mismatch conditions. *IEEE Transactions on Circuits and Systems I: Regular Papers*, v. 52, n. 10, p. 2101–2108, Oct. 2005.

BEZOOIJEN, A. van; MAHMOUDI, R.; ROERMUND, A. van. *Adaptive RF Front-Ends for Hand-held Applications*. Springer Netherlands, 2011. (Analog Circuits and Signal Processing). ISBN 9789048199358. Disponível em: <https://books.google.com.br/books?id=kLWlHurURGQC>.

BISHOP, E. A generalization of the Stone-Weierstrass theorem. *Pacific Journal of Mathematics*, Pacific Journal of Mathematics, A Non-profit Corporation, v. 11, n. 3, p. 777–783, 1961. Disponível em: <https://doi.org/>.

BJORCK, A. Solving least squares problems by Gram–Schmidt orthogonalization. *BIT*, v. 7, p. 1–21, Jan. 1967.

BOLSTAD, A. et al. Identification and compensation of Wiener-Hammerstein systems with feedback. In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. [S.l.: s.n.], 2011. p. 4056–4059.

BONFIM, E.; LIMA, E. A modified two dimensional Volterra-based series for the low-pass equivalent behavioral modeling of RF power amplifiers. *Progress In Electromagnetics Research M*, v. 47, p. 27–35, Jan. 2016.

BOQUÉ, R.; FERRÉ, J. Chapter 1 - Fundamentals of Parafac. In: *Fundamentals and Analytical Applications of Multiway Calibration*. Elsevier, 2015, (Data Handling in Science and Technology, v. 29). p. 7–35. Disponível em: <https://www.sciencedirect.com/science/article/pii/B9780444635273000011>.

BOSCH, W.; GATTI, G. Measurement and simulation of memory effects in predistortion linearizers. *IEEE Transactions on Microwave Theory and Techniques*, v. 37, n. 12, p. 1885–1890, Dec. 1989.

BOUILLOC, T.; FAVIER, G. Nonlinear channel modeling and identification using baseband Volterra–Parafac models. *Signal Processing*, v. 92, n. 6, p. 1492 – 1498, 2012. ISSN 0165-1684. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0165168411004397>.

BOYLE, K. R.; YUAN, Y.; LIGTHART, L. P. Analysis of mobile phone antenna impedance variations with user proximity. *IEEE Transactions on Antennas and Propagation*, v. 55, n. 2, p. 364–372, Feb. 2007.

BRAITHWAITE, R. N. Digital predistortion of an RF power amplifier using a reduced Volterra series model with a memory polynomial estimator. *IEEE Transactions on Microwave Theory and Techniques*, v. 65, n. 10, p. 3613–3623, Oct. 2017. ISSN 0018-9480.

BRINKHOFF, J.; PARKER, A. E. Effect of baseband impedance on FET intermodulation. *IEEE Transactions on Microwave Theory and Techniques*, v. 51, n. 3, p. 1045–1051, Mar. 2003. ISSN 0018-9480.

BRINKHOFF, J.; PARKER, A. E.; LEUNG, M. Baseband impedance and linearization of FET circuits. *IEEE Transactions on Microwave Theory and Techniques*, v. 51, n. 12, p. 2523–2530, Dec. 2003. ISSN 0018-9480.

CABRAL, P. M.; PEDRO, J. C.; CARVALHO, N. B. Bias networks impact on the dynamic AM/AM contours in microwave power amplifiers. In: *International Workshop on Integrated Nonlinear Microwave and Millimeter-Wave Circuits*. [S.l.: s.n.], 2006. p. 38–41.

CAI, J.; GONÇALVES, R.; PEDRO, J. C. A new complex envelope behavioral model for load mismatched power amplifiers. *International Journal of RF and Microwave Computer-Aided Engineering*, v. 27, n. 6, Feb. 2017. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/mmce.21097>.

CAI, J. et al. Nonlinear behavioral modeling dependent on load reflection coefficient magnitude. *IEEE Transactions on Microwave Theory and Techniques*, v. 63, n. 5, p. 1518–1529, May 2015. ISSN 0018-9480.

CANTONI, A.; TUTHILL, J. Digital compensation of frequency dependent imperfections in direct conversion I/Q modulators. In: *IEEE International Symposium on Circuits and Systems*. [S.l.: s.n.], 2007. p. 269–272.

CAO, H. et al. Compensation of transmitter distortion using a nonlinear modeling approach. In: *Workshop on Integrated Nonlinear Microwave and Millimetre-Wave Circuits*. [S.l.: s.n.], 2008. p. 131–134.

CAO, H. et al. I/Q imbalance compensation using a nonlinear modeling approach. *IEEE Transactions on Microwave Theory and Techniques*, v. 57, n. 3, p. 513–518, Mar. 2009.

CARVALHO, N.; PEDRO, J. Two-tone IMD asymmetry in microwave power amplifiers. In: *IEEE MTT-S International Microwave Symposium Digest (Cat. No.00CH37017)*. [S.l.: s.n.], 2000. p. 445–448 vol.1.

CARVALHO, N. B. de; PEDRO, J. A comprehensive explanation of distortion sideband asymmetries. *IEEE Transactions on Microwave Theory and Techniques*, v. 50, n. 9, p. 2090–2101, Sep. 2002.

CAVERS, J. Amplifier linearization using a digital predistorter with fast adaptation and low memory requirements. *IEEE Transactions on Vehicular Technology*, v. 39, n. 4, p. 374–382, Nov. 1990.

CAVERS, J. The effect of quadrature modulator and demodulator errors on adaptive digital predistorters for amplifier linearization. *IEEE Transactions on Vehicular Technology*, v. 46, n. 2, p. 456–466, May 1997.

CHEANG, C.; MAK, P.; MARTINS, R. P. A hardware-efficient feedback polynomial topology for DPD linearization of power amplifiers: Theory and FPGA validation. *IEEE Transactions on Circuits and Systems I: Regular Papers*, v. 65, n. 9, p. 2889–2902, Sep. 2018. ISSN 1549-8328.

CHEN, X.; XIE, M. ge. A split-and-conquer approach for analysis of extraordinarily large data. *Statistica Sinica*, Institute of Statistical Science, Academia Sinica, v. 24, n. 4, p. 1655–1684, Oct. 2014. ISSN 10170405, 19968507. Disponível em: <http://www.jstor.org/stable/24310963>.

CHOUCHANE, S.; BOUZRARA, K.; MESSAOUD, H. Estimation of the order and the memory of Volterra model from input/output observations. *International Journal of Modelling, Identification and Control*, v. 29, p. 244, Jan. 2018.

COLANTONIO, P.; GIANNINI, F.; LIMITI, E. Nonlinear analysis for power amplifiers. In: *High Efficiency RF and Microwave Solid State Power Amplifiers*. John Wiley  Sons, Ltd, 2009. cap. 3, p. 85–129. ISBN 9780470746547. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470746547.ch3>.

CRESPO-CADENAS, C. et al. On nonlinear amplifier modeling and identification using baseband Volterra-Parafac models. *Signal Processing*, v. 96, p. 401 – 405, 2014. ISSN 0165-1684. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0165168413003939>.

CRESPO-CADENAS, C. et al. A new approach to pruning Volterra models for power amplifiers. *IEEE Transactions on Signal Processing*, v. 58, n. 4, p. 2113–2120, 2010.

CRIPPS, S. C. *RF Power Amplifiers for Wireless Communications, Second Edition (Artech House Microwave Library (Hardcover))*. USA: Artech House, Inc., 2006. ISBN 1596930187.

CUNHA, T. R.; PEDRO, J. C.; CABRAL, P. M. Design of a power-amplifier feed-forward RF model with physical knowledge considerations. *IEEE Transactions on Microwave Theory and Techniques*, v. 55, n. 12, p. 2747–2756, Dec. 2007.

CUNHA, T. R. et al. General nonlinear feed-forward RF model for power amplifiers. In: *IEEE/MTT-S International Microwave Symposium*. [S.l.: s.n.], 2007. p. 2027–2030.

DAVENPORT, M. A. et al. Introduction to compressed sensing. In: ELDAR, Y. C.; KUTYNIOK, G. (Ed.). *Compressed Sensing: Theory and Applications*. [S.l.]: Cambridge University Press, 2012. p. 1–64.

DAVIES, M.; RILLING, G.; BLUMENSATH, T. Greedy algorithms for compressed sensing. In: ELDAR, Y.; KUTYNIOK, G. (Ed.). *Compressed Sensing*. [S.l.]: Cambridge University Press, 2012. p. 348–393. ISBN 978-1-107-00558-7.

DAVIS, C. G. et al. Alias-free products of signals near Nyquist rate. *IEEE Transactions on Signal Processing*, v. 66, n. 16, p. 4151–4159, 2018.

DEVI, R. V. S.; KURUP, D. G. Sparse identification of memory effects and nonlinear dynamics for developing parsimonious behavioral models of RF power amplifiers. In: *2017 IEEE MTT-S International Microwave and RF Conference (IMaRC)*. [S.l.: s.n.], 2017. p. 1–4.

DHAR, S. K. et al. A reflection-aware unified modeling and linearization approach for power amplifier under mismatch and mutual coupling. *IEEE Transactions on Microwave Theory and Techniques*, v. 66, n. 9, p. 4147–4157, Sep. 2018. ISSN 0018-9480.

DING, L. et al. Frequency-dependent modulator imbalance in predistortion linearization systems: modeling and compensation. In: *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*. [S.l.: s.n.], 2003. p. 688–692 Vol.1.

DING, L. et al. Compensation of frequency-dependent gain/phase imbalance in predistortion linearization systems. *IEEE Transactions on Circuits and Systems I: Regular Papers*, v. 55, n. 1, p. 390–397, Feb. 2008. ISSN 1549-8328.

DING, L. et al. Memory polynomial predistorter based on the indirect learning architecture. In: *Global Telecommunications Conference, 2002. GLOBECOM '02. IEEE.* [S.l.: s.n.], 2002. p. 967–971 vol.1.

DING, L. et al. A robust digital baseband predistorter constructed using memory polynomials. *IEEE Transactions on Communications*, v. 52, n. 1, p. 159–165, Jan. 2004. ISSN 0090-6778.

DOOLEY, J. et al. Estimation of sparse memory taps for RF power amplifier behavioral models. *IEEE Microwave and Wireless Components Letters*, v. 25, p. 64–66, Jan. 2015.

DU, T. et al. An accurate physical knowledge-based model for behavioral modeling and digital predistortion of power amplifiers. *International Journal of RF and Microwave Computer-Aided Engineering*, v. 23, n. 6, p. 682–689, Jun. 2013. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/mmce.20704>.

DUNN, Z. et al. Impedance-dependent wideband digital predistortion of solid-state radar amplifiers. *IEEE Transactions on Aerospace and Electronic Systems*, v. 53, n. 5, p. 2290–2303, Oct. 2017. ISSN 0018-9251.

EBADI, Z. S.; SALEH, R. Adaptive compensation of RF front-end nonidealities in direct conversion receivers. *IEEE Transactions on Circuits and Systems II: Express Briefs*, v. 55, n. 4, p. 354–358, Apr. 2008.

EFRON, B. et al. Least angle regression. *Ann. Statist.*, The Institute of Mathematical Statistics, v. 32, n. 2, p. 407–499, Apr. 2004. Disponível em: <https://doi.org/10.1214/009053604000000067>.

ELAD, M. *Sparse and Redundant Representations - From Theory to Applications in Signal and Image Processing*. [S.l.]: Springer, 2010. ISBN 978-1-4419-7010-7.

ENZINGER, H. *Behavioral Modeling and Digital Predistortion of Radio Frequency Power Amplifiers*. Tese (Doutorado) — Graz University of Technology, Austria, Mar. 2018.

EUN, C.; POWERS, E. J. A new Volterra predistorter based on the indirect learning architecture. *IEEE Transactions on Signal Processing*, v. 45, n. 1, p. 223–227, Jan. 1997. ISSN 1941-0476.

FAGER, C. et al. Prediction of smart antenna transmitter characteristics using a new behavioral modeling approach. In: *2014 IEEE MTT-S International Microwave Symposium (IMS2014)*. [S.l.: s.n.], 2014. p. 1–4.

FAGER, C. et al. A comprehensive analysis of IMD behavior in RF CMOS power amplifiers. *IEEE Journal of Solid-State Circuits*, v. 39, n. 1, p. 24–34, Jan. 2004. ISSN 0018-9200.

FAIG, H. et al. Dimensions-reduced Volterra digital pre-distortion based on orthogonal basis for band-limited nonlinear opto-electronic components. *IEEE Photonics Journal*, v. 11, n. 1, p. 1–13, Feb. 2019.

FAVIER, G. An overview of system modeling and identification. In: *11th International conference on Sciences and Techniques of Automatic control computer engineering (STA'2010)*. Monastir, Tunisia: [s.n.], 2010. Disponível em: <https://hal.archives-ouvertes.fr/hal-00718864>.

FAVIER, G.; BOUILLOC, T. Parametric complexity reduction of Volterra models using tensor decompositions. In: *2009 17th European Signal Processing Conference*. [S.l.: s.n.], 2009. p. 2288–2292.

FAVIER, G.; CAMPELLO, R.; AMARAL, W. C. D. Optimal expansions of discrete-time Volterra models using Laguerre functions. *Automatica*, Elsevier, v. 40, p. 815–822, 2004. Disponível em: <https://hal.archives-ouvertes.fr/hal-00719336>.

FAVIER, G.; KIBANGOU, A.; CAMPELLO, R. Nonlinear system modelling by means of orthonormal basis functions (Invited paper). In: . [S.l.: s.n.], 2003.

FAVIER, G.; KIBANGOU, A. Y.; BOUILLOC, T. Nonlinear system modeling and identification using Volterra-Parafac models. *International Journal of Adaptive Control and Signal Processing*, v. 26, p. 30–53, Jan. 2012. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/acs.1272>.

FEHRI, B.; BOUMAIZA, S. Application of embedding dimension estimation to Volterra series-based behavioral modeling and predistortion of wideband RF power amplifier. *International Journal of Microwave and Wireless Technologies*, Cambridge University Press, v. 5, n. 2, p. 115–122, 2013.

FEHRI, B.; BOUMAIZA, S. Baseband equivalent volterra series for behavioral modeling and digital predistortion of power amplifiers driven with wideband carrier aggregated signals. *IEEE Transactions on Microwave Theory and Techniques*, v. 62, n. 11, p. 2594–2603, Nov. 2014.

FERNANDES, C. A. R.; MOTA, J. C. M.; FAVIER, G. MIMO Volterra modeling for nonlinear communication channels. *Learning  Nonlinear Models*, SBRN, n. 2, p. 71–92, Ago. 2010.

FETTWEIS, G. et al. Dirty RF: a new paradigm. *Int. J. Wireless Inf. Networks*, v. 14, p. 133–148, 06 2007.

FRANKLIN, J. N. On Tikhonov's method for ill-posed problems. *Mathematics of Computation*, American Mathematical Society, v. 28, n. 128, p. 889–907, 1974. ISSN 00255718, 10886842. Disponível em: <http://www.jstor.org/stable/2005354>.

FRIEDMAN, J.; HASTIE, T.; TIBSHIRANI, R. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, Articles*, v. 33, n. 1, p. 1–22, Jan. 2010. ISSN 1548-7660. Disponível em: <https://www.jstatsoft.org/v033/i01>.

FU, J.-S.; MORTAZAWI, A. Improving power amplifier efficiency and linearity using a dynamically controlled tunable matching network. *IEEE Transactions on Microwave Theory and Techniques*, v. 56, n. 12, p. 3239–3244, Dec. 2008.

GHANNOUCHI, F. M.; HAMMI, O. Behavioral modeling and predistortion. *IEEE Microwave Magazine*, v. 10, n. 7, p. 52–64, 2009.

GIBIINO, G. P. et al. A two-port nonlinear dynamic behavioral model of RF PAs subject to wideband load modulation. *IEEE Transactions on Microwave Theory and Techniques*, v. 66, n. 2, p. 831–844, Feb. 2018. ISSN 1557-9670.

GILABERT, P. L. et al. Order reduction of wideband digital predistorters using principal component analysis. In: *2013 IEEE MTT-S International Microwave Symposium Digest (MTT)*. [S.l.: s.n.], 2013. p. 1–7. ISSN 0149-645X.

GIRISH, K.; JOHN, S. J. Multiset topologies induced by multiset relations. *Information Sciences*, v. 188, p. 298 – 313, 2012. ISSN 0020-0255. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0020025511006086>.

GOTTHANS, T.; BAUDOIN, G.; MBAYE, A. Optimal order estimation for modeling and predistortion of power amplifiers. In: *2013 IEEE International Conference on Microwaves, Communications, Antennas and Electronic Systems (COMCAS 2013)*. [S.l.: s.n.], 2013. p. 1–4.

GUAN, L.; ZHU, A. Simplified dynamic deviation reduction-based Volterra model for Doherty power amplifiers. In: *2011 Workshop on Integrated Nonlinear Microwave and Millimetre-Wave Circuits*. [S.l.: s.n.], 2011. p. 1–4.

GUAN, L.; ZHU, A. Optimized low-complexity implementation of least squares based model extraction for digital predistortion of RF power amplifiers. *IEEE Transactions on Microwave Theory and Techniques*, v. 60, n. 3, p. 594–603, Mar. 2012. ISSN 1557-9670.

HANZALA, M.; SHARAWI, M. S.; HAMMI, O. Accurate estimation of memory polynomial model dimension for dynamic nonlinear RF power amplifiers. In: *2015 IEEE 10th Conference on Industrial Electronics and Applications (ICIEA)*. [S.l.: s.n.], 2015. p. 145–150.

HASTIE, T.; TIBSHIRANI, R.; WAINWRIGHT, M. *Statistical Learning with Sparsity: The Lasso and Generalizations*. [S.l.]: Chapman and Hall/CRC, 2015. ISBN 1498712169.

HAUSMAIR, K. et al. Prediction of nonlinear distortion in wideband active antenna arrays. *IEEE Transactions on Microwave Theory and Techniques*, v. 65, n. 11, p. 4550–4563, Nov. 2017. ISSN 0018-9480.

HE, X.; ASADA, H. A new method for identifying orders of input-output models for nonlinear dynamic systems. In: *1993 American Control Conference*. [S.l.: s.n.], 1993. p. 2520–2523.

HEMSI, C.; PANAZIO, C. Sparse estimation technique for digital pre-distortion of impedance-mismatched power amplifiers. *Circuits, Systems, and Signal Processing*, p. 1–29, Feb. 2021.

HORLIN, F.; BOURDOUX, A. *Digital Compensation for Analog Front-Ends: A New Approach to Wireless Transceiver Design*. [S.l.: s.n.], 2008. 1-255 p.

HUANG, X.; CARON, M. Gain/phase imbalance and DC offset compensation in quadrature modulators. In: *IEEE International Symposium on Circuits and Systems. Proceedings (Cat. No.02CH37353)*. [S.l.: s.n.], 2002. p. IV–IV.

HUANG, X.; CARON, M. A novel type-based vector modulator self-calibration technique. In: *IEEE International Symposium on Circuits and Systems*. [S.l.: s.n.], 2009. p. 924–927.

HURVICH, C. M.; TSAI, C.-L. Regression and time series model selection in small samples. *Biometrika*, v. 76, n. 2, p. 297–307, Jun. 1989. ISSN 0006-3444. Disponível em: <https://doi.org/10.1093/biomet/76.2.297>.

ILVONEN, J. et al. Mobile terminal antenna performance with the user's hand: Effect of antenna dimensioning and location. *IEEE Antennas and Wireless Propagation Letters*, v. 10, p. 772–775, 2011.

ISAKSSON, M.; RONNOW, D. A parameter-reduced Volterra model for dynamic RF power amplifier modeling based on orthonormal basis functions. *International Journal of RF and Microwave Computer-Aided Engineering*, v. 17, p. 542–551, Jun. 2007. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/mmce.20253>.

ISAKSSON, M.; WISELL, D.; RONNOW, D. Wide-band dynamic modeling of power amplifiers using radial-basis function neural networks. *IEEE Transactions on Microwave Theory and Techniques*, v. 53, n. 11, p. 3422–3428, 2005.

ISAKSSON, M.; WISELL, D.; RONNOW, D. A comparative analysis of behavioral models for RF power amplifiers. *IEEE Transactions on Microwave Theory and Techniques*, v. 54, n. 1, p. 348–359, Jan. 2006.

JI, D.; JEON, J.; KIM, J. A novel load mismatch detection and correction technique for 3G/4G load insensitive power amplifier application. *IEEE Transactions on Microwave Theory and Techniques*, v. 63, n. 5, p. 1530–1543, May 2015. ISSN 0018-9480.

KEERTI, A.; PHAM, A.-V. H. RF characterization of SiGe HBT power amplifiers under load mismatch. *IEEE Transactions on Microwave Theory and Techniques*, v. 55, n. 2, p. 207–214, Feb. 2007.

KEKATOS, V.; GIANNAKIS, G. B. Sparse Volterra and polynomial regression models: Recoverability and estimation. *Trans. Sig. Proc.*, IEEE Press, Piscataway, NJ, USA, v. 59, n. 12, p. 5907–5920, Dec. 2011. ISSN 1053-587X. Disponível em: <https://doi.org/10.1109/TSP.2011.2165952>.

KESHMIRI, F.; CRAEYE, C. A Green's function approach for analysis of body-area-network antennas. In: *Loughborough Antennas & Propagation Conference*. [S.l.: s.n.], 2009. p. 769–772.

KIBANGOU, A.; FAVIER, G. Selection of generalized orthonormal bases for second-order Volterra filters. *Signal Processing*, v. 85, p. 2371–2385, May 2005.

KIBANGOU, A. Y.; FAVIER, G. Matrix and tensor decompositions for identification of block-structured nonlinear channels in digital transmission systems. In: *2008 IEEE 9th Workshop on Signal Processing Advances in Wireless Communications*. [S.l.: s.n.], 2008. p. 281–285.

KIM, J.; KONSTANTINOU, K. Digital predistortion of wideband signals based on power amplifier model with memory. *Electronics Letters*, v. 37, n. 23, p. 1417–1418, Nov. 2001.

KIM, Y.-D.; JEONG, E.-R.; LEE, Y. H. Adaptive compensation for power amplifier nonlinearity in the presence of quadrature modulation/demodulation errors. *IEEE Transactions on Signal Processing*, v. 55, n. 9, p. 4717–4721, Sep. 2007.

KOLDA, T. Numerical optimization for symmetric tensor decomposition. *Mathematical Programming*, v. 151, Oct. 2014.

KOLDA, T. G.; BADER, B. W. Tensor decompositions and applications. *SIAM Review*, v. 51, n. 3, p. 455–500, 2009. Disponível em: <https://doi.org/10.1137/07070111X>.

KORPI, D.; ANTTILA, L.; VALKAMA, M. Nonlinear self-interference cancellation in MIMO full-duplex transceivers under crosstalk. *EURASIP Journal on Wireless Communications and Networking*, v. 2017, n. 1, p. 24, Feb. 2017. ISSN 1687-1499. Disponível em: <https://doi.org/10.1186/s13638-017-0808-4>.

KRONVALL, T. et al. Sparse modeling of chroma features. *Signal Processing*, v. 130, Jun. 2016.

KU, H.; MCKINLEY, M.; KENNEY, J. Quantifying memory effects in RF power amplifiers. *IEEE Transactions on Microwave Theory and Techniques*, v. 50, n. 12, p. 2843–2849, Dec. 2002.

LANDIN, P.; BENGTSSON, O.; ISAKSSON, M. Power amplifier behavioural model mismatch sensitivity and the impact on digital predistortion performance. In: *2009 European Microwave Conference (EuMC)*. [S.l.: s.n.], 2009. p. 338–341.

LANDIN, P.; ISAKSSON, M.; HANDEL, P. Comparison of evaluation criteria for power amplifier behavioral modeling. In: *IEEE MTT-S International Microwave Symposium Digest*. [S.l.: s.n.], 2008. p. 1441–1444.

LANDIN, P. N. et al. Two novel memory polynomial models for modeling of RF power amplifiers. *International Journal of Microwave and Wireless Technologies*, Cambridge University Press, v. 7, n. 1, p. 19–29, Feb. 2015.

LEE, B. M.; FIGUEIREDO, R. J. P. D. Adaptive predistorters for linearization of high-power amplifiers in OFDM wireless communications. *Circuits, Systems, and Signal Processing*, v. 25, p. 59–80, Feb. 2006.

LEUNG, H.; ZHU, Z. *Signal Processing for RF Impairment Mitigation in Wireless Communications*. [S.l.: s.n.], 2014.

LI, L. et al. Evolutionary multiobjective optimization for digital predistortion architectures. In: NOGUET, D.; MOESSNER, K.; PALICOT, J. (Ed.). *Cognitive Radio Oriented Wireless Networks*. Cham: Springer International Publishing, 2016. p. 498–510. ISBN 978-3-319-40352-6.

LI, Y.; ZHU, A. On-demand real-time optimizable dynamic model sizing for digital predistortion of broadband RF power amplifiers. *IEEE Transactions on Microwave Theory and Techniques*, v. 68, p. 2891–2901, Jul. 2020.

LIMA, E. G.; CUNHA, T. R.; PEDRO, J. C. A physically meaningful neural network behavioral model for wireless transmitters exhibiting PM–AM/PM–PM distortions. *IEEE Transactions on Microwave Theory and Techniques*, v. 59, n. 12, p. 3512–3521, Dec. 2011.

LIU, J.; JI, S.; YE, J. *SLEP: Sparse Learning with Efficient Projections*. [S.l.], 2009. Disponível em: <http://www.public.asu.edu/~jye02/Software/SLEP>.

LIU, T.; BOUMAIZA, S.; GHANNOUCHI, F. M. Dynamic behavioral modeling of 3G power amplifiers using real-valued time-delay neural networks. *IEEE Transactions on Microwave Theory and Techniques*, v. 52, n. 3, p. 1025–1033, 2004.

LIU, T. et al. Quantitative measurements of memory effects in wideband RF power amplifiers driven by modulated signals. *IEEE Microwave and Wireless Components Letters*, v. 17, p. 79–81, Jan. 2007.

LJUNG, L.; FORSSELL, U. An alternative motivation for the indirect approach to closed-loop identification. *IEEE Transactions on Automatic Control*, v. 44, n. 11, p. 2206–2209, Nov. 1999.

LOPEZ–BUENO, D. et al. Independent digital predistortion parameters estimation using adaptive principal component analysis. *IEEE Transactions on Microwave Theory and Techniques*, v. 66, n. 12, p. 5771–5779, Dec. 2018.

MAAS, S. *Nonlinear Microwave and RF Circuits*. Artech House, 2003. ISBN 9781580536110. Disponível em: <https://books.google.com.br/books?id=SSw6gWLG-d4C>.

MAK, P.; U, S.; MARTINS, R. P. Transceiver architecture selection: Review, state-of-the-art survey and case study. *IEEE Circuits and Systems Magazine*, v. 7, n. 2, p. 6–25, Feb. 2007. ISSN 1531-636X.

MALHOTRA, M.; SAPPAL, A. S. SSA optimized digital pre-distorter for compensating non-linear distortion in high power amplifier. *Telecommunication Systems*, v. 72, n. 2, p. 179–188, Oct. 2019. ISSN 1572-9451. Disponível em: <https://doi.org/10.1007/s11235-019-00565-9>.

MARMARELIS, V. Identification of nonlinear biological systems using Laguerre expansions of kernels. *Annals of biomedical engineering*, v. 21, p. 573–89, Nov. 1993.

MARTÍ, R. B.; BALDRICH, J. F. Chapter 1 - Fundamentals of Parafac. *Data Handling in Science and Technology*, v. 29, p. 7–35, 2015.

MATHWORKS The. *MATLAB Optimization Toolbox*. 2017. Natick, MA, USA.

MAZIERE, C. et al. A system level model of solid state amplifiers with memory based on a nonlinear feedback loop principle. In: *European Microwave Conference*. [S.l.: s.n.], 2005.

MEIER, L.; GEER, S. van de; BUHLMANN, P. The group LASSO for logistic regression. *Journal of the Royal Statistical Society Series B*, v. 70, p. 53–71, Feb. 2008.

MEINSHAUSEN, N. Relaxed Lasso. *Computational Statistics and Data Analysis*, v. 52, n. 1, p. 374 – 393, Sep. 2007. ISSN 0167-9473. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0167947306004956>.

MESSAOUDI, N. et al. A comparative study of power amplifiers' sensitivity to load mismatch: Single branch vs. Doherty architectures. In: *2007 Canadian Conference on Electrical and Computer Engineering*. [S.l.: s.n.], 2007. p. 1543–1546. ISSN 0840-7789.

MINGO, J. de; CRESPO, A.; VALDOVINOS, A. Input impedance antenna automatic matching system. In: *The 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*. [S.l.: s.n.], 2002. p. 1872–1876 vol.4.

MINGO, J. de et al. An RF electronically controlled impedance tuning network design and its application to an antenna input impedance automatic matching system. *IEEE Transactions on Microwave Theory and Techniques*, v. 52, n. 2, p. 489–497, Feb. 2004. ISSN 0018-9480.

MITOLA, J.; MAGUIRE, G. Q. Cognitive radio: making software radios more personal. *IEEE Personal Communications*, v. 6, n. 4, p. 13–18, 1999.

MKADEM, F. et al. Complexity-reduced Volterra series model for power amplifier digital predistortion. *Analog Integrated Circuits and Signal Processing*, v. 79, n. 2, p. 331–343, May 2014. ISSN 1573-1979. Disponível em: <https://doi.org/10.1007/s10470-014-0266-4>.

MKADEM, F.; WU, D. Y.; BOUMAIZA, S. Wiener G-functionals for nonlinear power amplifier digital predistortion. In: *2012 IEEE/MTT-S International Microwave Symposium Digest*. [S.l.: s.n.], 2012. p. 1–3.

MOON, J.; KIM, B. Enhanced Hammerstein behavioral model for broadband wireless transmitters. *IEEE Transactions on Microwave Theory and Techniques*, v. 59, n. 4, p. 924–933, 2011.

MORGAN, D. R. et al. A generalized memory polynomial model for digital predistortion of RF power amplifiers. *IEEE Transactions on Signal Processing*, v. 54, n. 10, p. 3852–3860, Oct. 2006. ISSN 1053-587X.

MORISHITA, H.; KIM, Y.; FUJIMOTO, K. Analysis of handset antennas in the vicinity of the human body by the electromagnetic simulator. *IEICE Transactions on Electronics*, E84C, p. 937–947, Jul. 2001.

MUHONEN, K.; KAVEHRAD, M.; KRISHNAMOORTHY, R. Look-up table techniques for adaptive digital predistortion: a development and comparison. *IEEE Transactions on Vehicular Technology*, v. 49, n. 5, p. 1995–2002, Sep. 2000.

MURILLO, J. L. B.; YAGUE-JIMENEZ, V. A method for informed selection of memory-length and nonlinearity-order parameters in Volterra-Wiener systems from exponential sweep excitations. *Multidimensional Systems and Signal Processing*, v. 29, p. PP, Nov. 2017.

MYOUNG, S.-S.; KIM, Y.-H.; YOOK, J.-G. Impact of group delay in RF BPF on impulse radio systems. In: *IEEE MTT-S International Microwave Symposium Digest, 2005.* [S.l.: s.n.], 2005.

NAGATA, Y. Linear amplification technique for digital mobile communications. In: *IEEE 39th Vehicular Technology Conference.* [S.l.: s.n.], 1989. p. 159–164 vol.1.

NASKAS, N.; PAPANANOS, Y. Neural-network-based adaptive baseband predistortion method for RF power amplifiers. *IEEE Transactions on Circuits and Systems II: Express Briefs*, v. 51, n. 11, p. 619–623, 2004.

NGOYA, E.; MONS, S. Progress for behavioral challenges: A summary of time-domain behavioral modeling of RF and microwave subsystems. *IEEE Microwave Magazine*, v. 15, n. 6, p. 91–105, Sep. 2014. ISSN 1527-3342.

NIELSEN, T. et al. Modeling power amplifiers with antenna mismatch. In: *IEEE International Symposium on Circuits and Systems.* [S.l.: s.n.], 2005. p. 5083–5086 Vol. 5.

NORRIS, G. et al. Application of digital adaptive pre-distortion to mobile wireless devices. In: *2007 IEEE Radio Frequency Integrated Circuits (RFIC) Symposium.* [S.l.: s.n.], 2007. p. 247–250. ISSN 1529-2517.

O'BRIEN, B. et al. Estimation of memory length for RF power amplifier behavioral models. In: *2006 European Microwave Conference.* [S.l.: s.n.], 2006. p. 680–682.

OLIVEIRA, G. H. C. et al. An introduction to models based on Laguerre, Kautz and other related orthonormal functions - Part I: linear and uncertain models. *International Journal of Modelling, Identification and Control - IJMIC*, Inderscience Publishers, v. 14, p. 121–132, 2011.

OLIVEIRA, G. H. C. et al. An introduction to models based on Laguerre, Kautz and other related orthonormal functions - Part II: non-linear models. *International Journal of Modelling, Identification and Control - IJMIC*, Inderscience Publishers, 2012.

PAVAN, S.; SHIVAPPA, S. Nonidealities in traveling wave and transversal FIR filters operating at microwave frequencies. *IEEE Transactions on Circuits and Systems I: Regular Papers*, v. 53, n. 1, p. 177–192, Jan. 2006.

PEDRO, J.; PEREZ, J. Accurate simulation of GaAs MESFET's intermodulation distortion using a new drain-source current model. *IEEE Transactions on Microwave Theory and Techniques*, v. 42, n. 1, p. 25–33, Jan. 1994.

PEDRO, J. C.; CARVALHO, N. B.; LAVRADOR, P. M. Modeling nonlinear behavior of band-pass memoryless and dynamic systems. In: *IEEE MTT-S International Microwave Symposium Digest, 2003.* [S.l.: s.n.], 2003. v. 3, p. 2133–2136. ISSN 0149-645X.

PEDRO, J. C.; MAAS, S. A. A comparative overview of microwave and wireless power-amplifier behavioral modeling approaches. *IEEE Transactions on Microwave Theory and Techniques*, v. 53, n. 4, p. 1150–1163, Apr. 2005. ISSN 0018-9480.

PEDRO, J. C.; NUNES, L. C.; CABRAL, P. M. A simple method to estimate the output power and efficiency load–pull contours of Class-B power amplifiers. *IEEE Transactions on Microwave Theory and Techniques*, v. 63, n. 4, p. 1239–1249, Apr. 2015.

PIAZZA, R.; RAO, B. S. M. R.; OTTERSTEN, B. Data predistortion for multicarrier satellite channels using orthogonal memory polynomials. In: *2013 IEEE 14th Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. [S.l.: s.n.], 2013. p. 689–693.

PRESTI, C. D.; KIMBALL, D. F.; ASBECK, P. M. Closed-loop digital predistortion system with fast real-time adaptation applied to a handset WCDMA PA module. *IEEE Transactions on Microwave Theory and Techniques*, v. 60, n. 3, p. 604–618, Mar. 2012. ISSN 0018-9480.

QI, H.; BENEDIKT, J.; TASKER, P. J. Nonlinear data utilization: From direct data lookup to behavioral modeling. *IEEE Transactions on Microwave Theory and Techniques*, v. 57, n. 6, p. 1425–1432, Jun. 2009. ISSN 1557-9670.

QIAN, H. et al. A low-complexity digital predistortion algorithm for power amplifier linearization. *IEEE Transactions on Broadcasting*, v. 60, n. 4, p. 670–678, Dec. 2014.

QUINDROIT, C. et al. FPGA implementation of orthogonal 2D digital predistortion system for concurrent dual-band power amplifiers based on time-division multiplexing. *IEEE Transactions on Microwave Theory and Techniques*, v. 61, n. 12, p. 4591–4599, Dec. 2013.

RAAB, F. et al. Power amplifiers and transmitters for RF and microwave. *IEEE Transactions on Microwave Theory and Techniques*, v. 50, n. 3, p. 814–826, Mar. 2002.

RAHATI, A.; MOTAMEDI, S.; SHARIFIAN, S. A novel generalized parallel two-box structure for behavior modeling and digital predistortion of RF power amplifiers at lte applications. *Circuits, Systems, and Signal Processing*, v. 37, Nov. 2017.

RAICH, R.; QIAN, H.; ZHOU, G. T. Orthogonal polynomials for power amplifier modeling and predistorter design. *IEEE Transactions on Vehicular Technology*, v. 53, n. 5, p. 1468–1479, Sep. 2004.

RAICH, R.; ZHOU, G. On the modeling of memory nonlinear effects of power amplifiers for communication applications. In: *Proceedings of 2002 IEEE 10th Digital Signal Processing Workshop, 2002 and the 2nd Signal Processing Education Workshop.* [S.l.: s.n.], 2002. p. 7–10.

RAPP, C. Effects of HPA-nonlinearity on 4-DPSK/OFDM-signal for a digital sound broadcasting system. In: . [S.l.: s.n.], 1991. p. 179–184.

RAWAT, K.; HASHMI, M. S.; GHANNOUCHI, F. M. Dual-band RF circuits and components for multi-standard software defined radios. *IEEE Circuits and Systems Magazine*, v. 12, n. 1, p. 12–32, First quarter 2012. ISSN 1531-636X.

RAZAVI, B. Design considerations for direct-conversion receivers. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, v. 44, n. 6, p. 428–435, Jun. 1997.

ROSA, A.; CAMPELLO, R.; AMARAL, W. Choice of free parameters in expansions of discrete-time Volterra models using Kautz functions. *Automatica*, v. 43, p. 1084–1091, Jun. 2007.

SAFFAR, D. et al. A compound structure and a single-step identification procedure for I/Q and DC offset impairments and nonlinear distortion modeling and compensation in wireless transmitters. *International Journal of RF and Microwave Computer-Aided Engineering*, v. 23, p. 367–377, Mai 2013.

SALEH, A. Frequency-independent and frequency-dependent nonlinear models of TWT amplifiers. *IEEE Transactions on Communications*, v. 29, n. 11, p. 1715–1720, 1981.

SANCHEZ-PEREZ, C. et al. Improving digital predistortion mismatch sensitivity using tunable matching networks. In: *IEEE 73rd Vehicular Technology Conference (VTC Spring)*. [S.l.: s.n.], 2011. p. 1–5.

SCHETZEN, M. *The Volterra and Wiener Theories of Nonlinear Systems*. Melbourne, FL, USA: Krieger Publishing Co., Inc., 2006. ISBN 1575242834.

SCHREURS, D. et al. *RF Power Amplifier Behavioral Modeling*. 1. ed. United Kingdom: Cambridge University Press, 2008. (The Cambridge RF and Microwave Engineering Series). ISBN 978-0-521-88173-9.

SCHUARTZ, L. et al. Reduced-complexity polynomials with memory applied to the linearization of power amplifiers with real-time discrete gain control. *Circuits, Systems, and Signal Processing*, v. 38, Feb. 2019.

SCHUMACHER, R.; LIMA, E.; OLIVEIRA, G. RF power amplifier behavioral modeling based on Takenaka–Malmquist–Volterra series. *Circuits, Systems, and Signal Processing*, v. 35, Aug. 2015.

SESIA, S.; TOUFIK, I.; BAKER, M. *LTE, The UMTS Long Term Evolution: From Theory to Practice*. [S.l.]: Wiley Publishing, 2009. ISBN 0470697164.

SEVIC, J.; BURGER, K.; STEER, M. A novel envelope-termination load-pull method for ACPR optimization of RF/microwave power amplifiers. In: *IEEE MTT-S International Microwave Symposium Digest (Cat. No.98CH36192)*. [S.l.: s.n.], 1998. p. 723–726 vol.2.

SIMON, N. et al. A sparse-group Lasso. *Journal of Computational and Graphical Statistics*, [American Statistical Association, Taylor  Francis, Ltd., Institute of Mathematical Statistics, Interface Foundation of America], v. 22, n. 2, p. 231–245, Jun. 2013. ISSN 10618600. Disponível em: <http://www.jstor.org/stable/43304828>.

SIMPSON, G. et al. Load-pull + NVNA = enhanced X-parameters for pa designs with high mismatch and technology-independent large-signal device models. In: *2008 72nd ARFTG Microwave Measurement Symposium*. [S.l.: s.n.], 2008. p. 88–91.

SINGH, D. et al. An overview of the applications of multisets. *Novi Sad J. Math*, v. 37, p. 73–92, Jan. 2007.

SMAINI, L. *RF Analog Impairments Modeling for Communication Systems Simulation: Application to OFDM-based Transceivers*. Wiley, 2012. ISBN 9781118438879. Disponível em: <https://books.google.com.br/books?id=McBwgm8_W1UC>.

STANKOVIC, L. et al. A tutorial on sparse signal reconstruction and its applications in signal processing. *Circuits, Systems, and Signal Processing*, v. 38, p. 1206–1263, Aug. 2018.

STAUDINGER, J.; NANAN, J.; WOOD, J. Memory fading Volterra series model for high power infrastructure amplifiers. In: *2010 IEEE Radio and Wireless Symposium (RWS)*. [S.l.: s.n.], 2010. p. 184–187.

TARVER, C. et al. Low-complexity, multi sub-band digital predistortion: Novel algorithms and SDR verification. *Journal of Signal Processing Systems*, Nov. 2017.

TEHRANI, A. S. et al. A comparative analysis of the complexity/accuracy tradeoff in power amplifier behavioral models. *IEEE Transactions on Microwave Theory and Techniques*, v. 58, n. 6, p. 1510–1520, Jun. 2010. ISSN 1557-9670.

TEHRANI, A. S. et al. Orthonormal-basis power amplifier model reduction. In: *2008 Workshop on Integrated Nonlinear Microwave and Millimetre-Wave Circuits*. [S.l.: s.n.], 2008. p. 39–42.

TIBSHIRANI, R. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, [Royal Statistical Society, Wiley], v. 58, n. 1, p. 267–288, Jan. 1996. ISSN 00359246. Disponível em: <http://www.jstor.org/stable/2346178>.

TIBSHIRANI, R.; HASTIE, T.; FRIEDMAN, J. Regularized paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, v. 33, Feb. 2010.

TOMASI, G.; BRO, R. A comparison of algorithms for fitting the Parafac model. *Computational Statistics  Data Analysis*, v. 50, n. 7, p. 1700 – 1734, 2006. ISSN 0167-9473. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0167947304003895>.

TOSINA, J. R. et al. Behavioral modeling and predistortion of power amplifiers under sparsity hypothesis. *IEEE Transactions on Microwave Theory and Techniques*, v. 63, n. 2, p. 745–753, Feb. 2015. ISSN 1557-9670.

TUTHILL, J.; CANTONI, A. Efficient compensation for frequency-dependent errors in analog reconstruction filters used in IQ modulators. *IEEE Transactions on Communications*, v. 53, n. 3, p. 489–496, Mar. 2005.

VALKAMA, M. RF impairment compensation for future radio systems. In: *Multi-Mode/Multi-Band RF Transceivers for Wireless Communications*. [S.l.]: John Wiley  Sons, Ltd, 2010. cap. 15, p. 451 – 496. ISBN 9780470634455.

VALKAMA, M.; RENFORS, M.; KOIVUNEN, V. Advanced methods for I/Q imbalance compensation in communication receivers. *IEEE Transactions on Signal Processing*, v. 49, n. 10, p. 2335–2344, Oct. 2001.

VALKAMA, M.; RENFORS, M.; KOIVUNEN, V. Compensation of frequency-selective I/Q imbalances in wideband receivers: models and algorithms. In: *IEEE Third Workshop on Signal Processing Advances in Wireless Communications (SPAWC'01). Workshop Proceedings (Cat. No.01EX471)*. [S.l.: s.n.], 2001. p. 42–45.

VALKAMA, M.; RENFORS, M.; KOIVUNEN, V. Blind I/Q imbalance compensation in OFDM receivers based on adaptive I/Q signal decorrelation. In: *IEEE International Symposium on Circuits and Systems*. [S.l.: s.n.], 2005. p. 2611–2614 Vol. 3.

VALKAMA, M.; RENFORS, M.; KOIVUNEN, V. Blind I/Q signal separation-based solutions for receiver signal processing. *EURASIP Journal on Applied Signal Processing*, Hindawi Publishing Corporation, n. 2005:16, p. 2708–2718, 2005. ISSN 1110-8657.

VERBEYST, F.; BOSSCHE, M. Measurement-based behavioral model under mismatched conditions: a new and easy approach for an accurate model. In: *European Microwave Conference*. [S.l.: s.n.], 2005.

VERSPECHT, J. et al. Broad-band, multi-harmonic frequency domain behavioral models from automated large-signal vectorial network measurements. In: *IEEE MTT-S International Microwave Symposium Digest, 2005.* [S.l.: s.n.], 2005.

VERVLIET, N. et al. *Tensorlab 3.0*. 2016. Available online. Disponível em: <https://www.tensorlab.net>.

VUOLEVI, J.; RAHKONEN, T. Analysis of third-order intermodulation distortion in common-emitter BJT and HBT amplifiers. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, v. 50, n. 12, p. 994–1001, Dec. 2003.

WANG, S. et al. Comparison of hill-climbing and genetic algorithms for digital predistortion models sizing. In: *2016 IEEE International Conference on Electronics, Circuits and Systems (ICECS)*. [S.l.: s.n.], 2016. p. 289–292.

WANG, S. et al. A novel algorithm for determining the structure of digital predistortion models. *IEEE Transactions on Vehicular Technology*, v. 67, n. 8, p. 7326–7340, Aug. 2018. ISSN 0018-9545.

WIENER, N. *Nonlinear problems in random theory*. USA: Wiley, 1958. ISBN 1596930187.

WILLIAMS, D.; LECKEY, J.; TASKER, P. A study of the effect of envelope impedance on intermodulation asymmetry using a two-tone time domain measurement system. In: *IEEE MTT-S International Microwave Symposium Digest (Cat. No.02CH37278)*. [S.l.: s.n.], 2002. p. 1841–1844 vol.3.

WILLIAMSON, G.; HACIOGLU, R. Reduced complexity Volterra models for nonlinear system identification. *EURASIP Journal on Advances in Signal Processing*, v. 2001, Dec. 2001.

WISELL, D.; JALDEN, J.; HANDEL, P. Behavioral power amplifier modeling using the LASSO. In: *2008 IEEE Instrumentation and Measurement Technology Conference*. [S.l.: s.n.], 2008. p. 1864–1867. ISSN 1091-5281.

WOODINGTON, S. et al. A novel measurement based method enabling rapid extraction of a RF waveform look-up table based behavioral model. In: *IEEE MTT-S International Microwave Symposium Digest*. [S.l.: s.n.], 2008. p. 1453–1456.

XU, J. et al. Neural-based dynamic modeling of nonlinear microwave circuits. *IEEE Transactions on Microwave Theory and Techniques*, v. 50, n. 12, p. 2769–2780, 2002.

YANG, G. et al. 2D orthogonal polynomials for concurrent dual-band digital predistortion. In: *2013 IEEE MTT-S International Microwave Symposium Digest (MTT)*. [S.l.: s.n.], 2013. p. 1–3.

YAO, S. et al. Orthonormal basis functions for memory predistorter in oversampled systems. *International Journal of RF and Microwave Computer-Aided Engineering*, v. 25, n. 3, p. 202–212, Sep. 2014. Disponível em: <https://onlinelibrary.wiley.com/doi/abs/10.1002/mmce.20850>.

YE, Y. et al. Analysis and decomposition of the nonlinearities in RF power amplifiers. In: *6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM)*. [S.l.: s.n.], 2010. p. 1–4.

YU, X.; JIANG, H. Digital predistortion using adaptive basis functions. *IEEE Transactions on Circuits and Systems I: Regular Papers*, v. 60, n. 12, p. 3317–3327, Dec. 2013.

YUAN, M.; LIN, Y. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society Series B*, v. 68, p. 49–67, Feb. 2006.

ZAKARIA, A.; JOHN, S. J.; GIRISH, K. P. Multiset filters. *Journal of the Egyptian Mathematical Society*, v. 27, n. 51, Dec. 2019.

ZARGAR, H.; BANAI, A.; PEDRO, J. DIDO behavioral model extraction setup using uncorrelated envelope signals. In: *2015 European Microwave Conference (EuMC)*. [S.l.: s.n.], 2015. p. 646–649.

ZARGAR, H.; BANAI, A.; PEDRO, J. C. A new double input-double output complex envelope amplifier behavioral model taking into account source and load mismatch effects. *IEEE Transactions on Microwave Theory and Techniques*, v. 63, n. 2, p. 766–774, Feb. 2015. ISSN 0018-9480.

ZENTENO, E.; ISAKSSON, M.; HANDEL, P. Output impedance mismatch effects on the linearity performance of digitally predistorted power amplifiers. *IEEE Transactions on Microwave Theory and Techniques*, v. 63, n. 2, p. 754–765, Feb. 2015.

ZENTENO, E. et al. Low complexity predistortion and equalization in nonlinear multicarrier satellite communications. *Journal on Advances in Signal Processing*, v. 2015, Mar. 2015.

ZHENG, Q.; ZAFIRIOU, E. Nonlinear system identification for control using Volterra-Laguerre expansion. In: *Proceedings of 1995 American Control Conference - ACC'95*. [S.l.: s.n.], 1995. v. 3, p. 2195–2199 vol.3.

ZHU, A.; BRAZIL, T. J. RF power amplifier behavioral modeling using Volterra expansion with Laguerre functions. In: *IEEE MTT-S International Microwave Symposium Digest, 2005.* [S.l.: s.n.], 2005. p. 963–966.

ZHU, A.; DOOLEY, J.; BRAZIL, T. J. Simplified Volterra series based behavioral modeling of RF power amplifiers using deviation-reduction. In: *2006 IEEE MTT-S International Microwave Symposium Digest*. [S.l.: s.n.], 2006. p. 1113–1116. ISSN 0149-645X.

ZHU, A.; PEDRO, J. C.; BRAZIL, T. J. Dynamic deviation reduction-based Volterra behavioral modeling of RF power amplifiers. *IEEE Transactions on Microwave Theory and Techniques*, v. 54, n. 12, p. 4323–4332, 2006.

ZHU, A.; PEDRO, J. C.; CUNHA, T. R. Pruning the Volterra series for behavioral modeling of power amplifiers using physical knowledge. *IEEE Transactions on Microwave Theory and Techniques*, v. 55, n. 5, p. 813–821, May 2007. ISSN 0018-9480.

ZHU, Z.; LEUNG, H.; HUANG, X. Challenges in reconfigurable radio transceivers and application of nonlinear signal processing for RF impairment mitigation. *IEEE Circuits and Systems Magazine*, v. 13, n. 1, p. 44–65, First quarter 2013. ISSN 1531-636X.

ZOU, H.; HASTIE, T. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, v. 67, n. 2, p. 301–320, 2005. Disponível em: <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2005.00503.x>.

# APPENDIX A – JOINT COMPENSATION OF PA LMM AND IQM IMBALANCE

This appendix proposes an extension of the WHFB model targeting the joint compensation of PA LMM and IQM imbalances, discussed in section 2.4. The experimental validation of this model is not presented in this thesis and remains as a future work.

IQM imbalance compensation is described in (VALKAMA; RENFORS; KOIVUNEN, 2001a), (ANTTILA; VALKAMA; RENFORS, 2008b), (ANTTILA; VALKAMA; RENFORS, 2008a) and is discussed here in the context of the joint compensation of IQM and PA LMM. Therefore, this section extends the WHFB DPD model to the joint compensation of IQM imbalance.

In the literature, IQM imbalance and PA NL can be mitigated either separately or jointly. The individual compensation is a two-step process in which the IQM imbalance is first compensated (CANTONI; TUTHILL, 2007), (EBADI; SALEH, 2008). In this case, the compensation structure is composed of the cascaded PA DPD and IQM compensation blocks, in this order. The solution requires extra hardware, since both the IQM and the PA output signals are fed back to DSP. By its turn, the joint mitigation approach relies on a parallel structure that performs the one-step compensation of PA NL and IQM imbalance (ANTTILA; HANDEL; VALKAMA, 2010), allowing all coefficients to be simultaneously estimated by the OLS technique. Besides, no additional hardware is required.

Two state-of-the-art models for the one-step, joint compensation of PA NL and IQM frequency-dependent imbalance are considered in this section: the dual-input polynomial model (CAO et al., 2008) and (CAO et al., 2009), and the dual-branch parallel Hammerstein (PH) model (ANTTILA; HANDEL; VALKAMA, 2010). In the first approach, we also consider that (SAFFAR et al., 2013) replaced the Volterra model in the original dual-input DPD structure by the MP model, thus reducing its complexity and increasing the model scalability. Note that some other approaches have been also proposed in the literature, such as non-data-aided IQM compensation schemes, where statistical properties of the input and output complex baseband signals are used, e.g., the cumulative distribution functions (CDF) (HUANG; CARON, 2009), the circularity property of circular modulation formats (VALKAMA; RENFORS; KOIVUNEN, 2001a), (ANTTILA; VALKAMA; RENFORS, 2008a), and the I and Q correlation of the output signal (since IQM imbalance creates correlation between them) (VALKAMA; RENFORS; KOIVUNEN, 2005a). Finally, higher-order statistics can also be used in a blind signal separation (BSS) technique applied to the I and Q components, as proposed in (VALKAMA; RENFORS; KOIVUNEN, 2005b).

First, let us consider the dual-branch PH joint-DPD model, proposed in (ANTTILA; HANDEL; VALKAMA, 2010), as an augmented version of the IQM model in Figure 2.9, using the PH model in each branch. The dual-branch PH model processes separately the input signal and its complex conjugate, as shown in Figure A.1. The output signal is modelled by memoryless polynomials with orders $K_a$ and $K_b$, respectively, for the non-conjugate and conjugate input signals, respectively, and by FIR filters with depths $\{M_k\}$, associated to each PH branch.

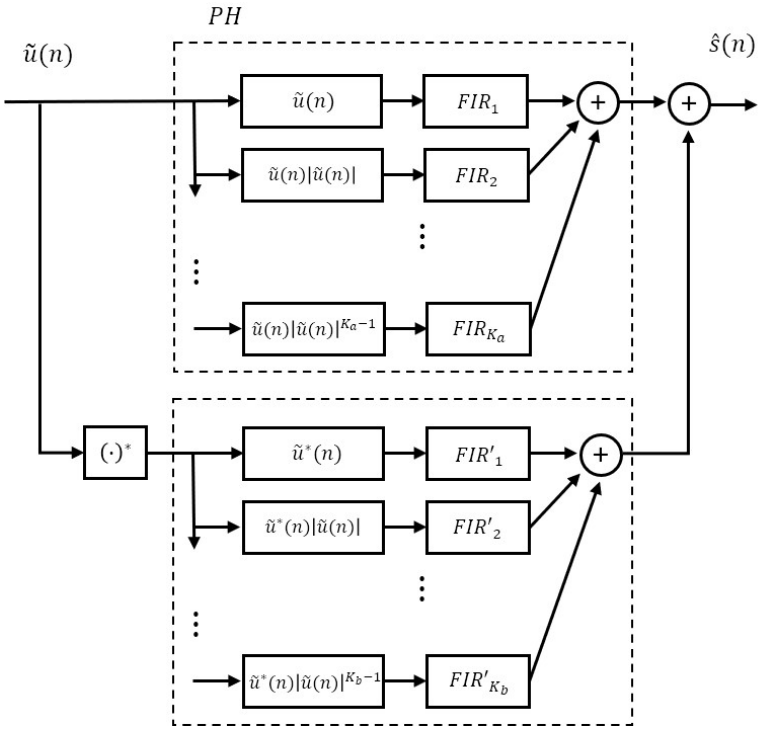



Figure A.1 – Dual-branch PH joint DPD model.

The dual-branch PH DPD model is expressed as:

$$\hat{s}\,(n) = \sum_{\substack{k=1 \\ k\ odd}}^{K_1} \sum_{m=0}^{M_k} \beta_{m,k}^{(1)} \tilde{u}\,(n-m)\,|\tilde{u}\,(n-m)|^{k-1} + \sum_{\substack{k=1 \\ k\ odd}}^{K_2} \sum_{m=0}^{M_k} \beta_{m,k}^{(2)} \tilde{u}^*\,(n-m)\,|\tilde{u}\,(n-m)|^{k-1}\,, \tag{A.1}$$

where $M_k$ is the $k^{th}$-order branch FIR filter depth, $K_1$ and $K_2$ are the pre-distorter non-linear orders of the non-conjugate and conjugate MP blocks, $\tilde{u}(n)$, $\tilde{u}^*(n)$ and $\hat{s}(n)$ are, respectively, the ILA DPD input, conjugate input and modelled output signals.

As shown in (VALKAMA; RENFORS; KOIVUNEN, 2001a), the IQ imbalance implies a widely linear transformation that is also present in the compensation structure.

The second model is the dual-input MP model that processes separately the real and imaginary parts of the complex-valued baseband input signal (GHANNOUCHI; HAMMI, 2009). In Figure A.2, the dual-input MP block input signals are $v_I(n)$ and $v_Q(n)$, the real

and imaginary parts of the normalized LPE PA output signal, $u_I(n)$ and $u_Q(n)$, after passing through an IQM imbalance model, represented by the real filters $\boldsymbol{g}_{II}$, $\boldsymbol{g}_{IQ}$, $\boldsymbol{g}_{QI}$ and $\boldsymbol{g}_{QQ}$.

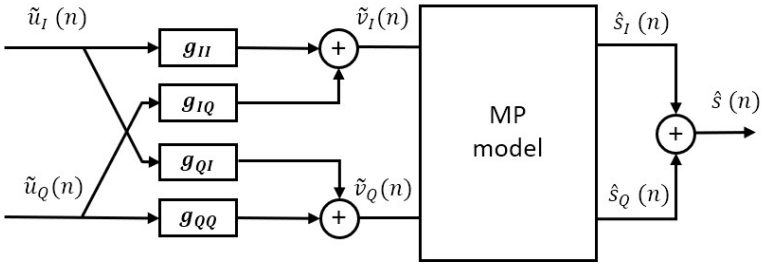



Figure A.2 – Dual-input MP joint DPD model.

The dual-input MP model output can be written as:

$$\begin{aligned}\hat{s}(n) &= \sum_{m=0}^{M}\sum_{\substack{k=1\\ k\ odd}}^{K}\beta_{m,k}\tilde{v}(n-m)\left|\tilde{v}(n-m)\right|^{k-1}\\ &= \sum_{m=0}^{M}\sum_{\substack{k=1\\ k\ odd}}^{K}\beta_{m,k}\Big(\tilde{v}_I(n-m)+j\tilde{v}_Q(n-m)\Big)\\ &\quad\Big[\tilde{v}_I^2(n-m)+\tilde{v}_Q^2(n-m)\Big]^{\frac{k-1}{2}},\end{aligned} \tag{A.2}$$

where $v_I(n)$ and $v_Q(n)$ are the IQM imbalance model output signals, given by:

$$\tilde{v}_I(n) = \boldsymbol{g}_{II}\tilde{\boldsymbol{u}}_I(n) + \boldsymbol{g}_{IQ}\tilde{\boldsymbol{u}}_Q(n) \tag{A.3}$$

and

$$\tilde{v}_Q(n) = \boldsymbol{g}_{QI}\tilde{\boldsymbol{u}}_I(n) + \boldsymbol{g}_{QQ}\tilde{\boldsymbol{u}}_Q(n). \tag{A.4}$$

Replacing eq. (A.3) and (A.4) into eq. (A.2), the joint DPD model output is expressed as:

$$\begin{aligned}\hat{s}(n) = \sum_{m=0}^{M}\sum_{\substack{k=1\\ k\ odd}}^{K}\beta_{m,k}&\Big[g_{II}(m)\tilde{u}_I(n-m)+g_{IQ}(m)\tilde{u}_Q(n-m)\\ &+j\Big(g_{QI}(m)\tilde{u}_I(n-m)+g_{QQ}(m)\tilde{u}_Q(n-m)\Big)\Big]\ \Big[(g_{II}(m)\tilde{u}_I(n-m)\\ &+g_{IQ}\tilde{u}_Q(n-m))^2+(g_{QI}\tilde{u}_I(n-m)+g_{QQ}\tilde{u}_Q(n-m))^2\Big]^{\frac{k-1}{2}}.\end{aligned} \tag{A.5}$$

Then, the expression (A.5) is developed for $k = 1$ and $k = 3$, as in (GHANNOUCHI; HAMMI, 2009):

For $k = 1$:

$$\hat{s}_I(n)_{k=1} = \sum_{m=0}^{M} \beta_{m,1} \Big( g_{II}(m)\tilde{u}_I(n-m) + g_{IQ}(m)\tilde{u}_Q(n-m) \Big) \tag{A.6}$$

and

$$\hat{s}_Q(n)_{k=1} = \sum_{m=0}^{M} \beta_{m,1} \Big( g_{QI}(m)\tilde{u}_I(n-m) + g_{QQ}(m)\tilde{u}_Q(n-m) \Big). \tag{A.7}$$

For $k = 3$:

$$\begin{aligned} \hat{s}_I(n)_{k=3} = \sum_{m=0}^{M} \beta_{m,3} \Big[ & (g_{II}^3 + g_{II}g_{QI}^2)\tilde{u}_I^3(n-m) + (3g_{II}^2 g_{IQ} + g_{IQ}g_{QI}^2 \\ & + 2g_{II}g_{QI}g_{QQ})\tilde{u}_I^2(n-m)\tilde{u}_Q(n-m) + (3g_{II}g_{IQ}^2 + g_{II}g_{QQ}^2 \\ & + 2g_{IQ}g_{QI}g_{QQ})\tilde{u}_I(n-m)\tilde{u}_Q^2(n-m) + (g_{IQ}^3 + g_{IQ}g_{QQ}^2)\tilde{u}_Q^3(n-m) \Big] \end{aligned} \tag{A.8}$$

and

$$\begin{aligned} \hat{s}_Q(n)_{k=3} = \sum_{m=0}^{M} \beta_{m,3} \Big[ & (g_{QI}^3 + g_{II}^2 g_{QI})\tilde{u}_I^3(n-m) + (3g_{QI}^2 g_{QQ} + g_{QQ}g_{II}^2 \\ & + 2g_{II}g_{QI}g_{IQ})\tilde{u}_I^2(n-m)\tilde{u}_Q(n-m) + (3g_{QI}g_{QQ}^2 + g_{QI}g_{IQ}^2 \\ & + 2g_{II}g_{IQI}g_{QQ})\tilde{u}_I(n-m)\tilde{u}_Q^2(n-m) + (g_{QQ}^3 + g_{IQ}g_{QQ}^2)\tilde{u}_Q^3(n-m) \Big]. \end{aligned} \tag{A.9}$$

By developing for the next odd-order NL terms, the dual-input MP model is expressed by:

$$\hat{s}_I(n) = \sum_{m=0}^{M} \sum_{\substack{k=1 \\ k\ odd}}^{K} \sum_{r=0}^{k} \beta_{m,k,r}^{(I)} \tilde{u}_I^{k-r}(n-m)\tilde{u}_Q^r(n-m) \tag{A.10}$$

and

$$\hat{s}_Q(n) = \sum_{m=0}^{M} \sum_{\substack{k=1 \\ k\ odd}}^{K} \sum_{r=0}^{k} \beta_{m,k,r}^{(Q)} \tilde{u}_I^{k-r}(n-m)\tilde{u}_Q^r(n-m), \tag{A.11}$$

where $M$ and $K$ are the pre-distorter memory depth and NL order, $z_I(n)$ and $z_Q(n)$ are the real and imaginary parts of the modelled pre-distorted IQM input signal, whereas $u_I(n)$ and $u_Q(n)$ are the real and imaginary parts of the normalized PA output signal. $\boldsymbol{\beta}^{(I)}$ and $\boldsymbol{\beta}^{(Q)}$ are the I- and Q-component coefficients vectors. The following expression is equivalent to eqs. (A.10) and (A.11):

$$\hat{s}\,(n) = \sum_{m=0}^{M} \sum_{\substack{k=1 \\ k\ odd}}^{K} \sum_{r=0}^{k} \beta_{m,k,r} \tilde{u}_I^{k-r}(n-m) \tilde{u}_Q^{r}(n-m), \tag{A.12}$$

where $\boldsymbol{\beta}$ is the complex-valued coefficients vector, such that $\boldsymbol{\beta} = \boldsymbol{\beta}^{(I)} + j\boldsymbol{\beta}^{(Q)}$.

Likewise, using the dual-input joint DPD model approach, the WHFB DPD model can be extended to the joint compensation of PA LMM and IQM imbalance, as follows:

$$\begin{aligned}
\hat{s}\,(n) = & \sum_{m=0}^{M_1} \sum_{\substack{k=1 \\ odd}}^{K_1} \sum_{r=0}^{k} \beta_{m,k,r}^{(1)} \tilde{u}_I^{k-r}(n-m) \tilde{u}_Q^{r}(n-m) \\
& + \sum_{m=0}^{M_2} \sum_{l=1}^{L_2} \sum_{\substack{k=1 \\ odd}}^{K_2} \sum_{r=0}^{k} \beta_{m,k,r}^{(2,1)} \tilde{u}_I^{k-r}(n-m) \tilde{u}_Q^{r}(n-m) \\
& \sum_{\substack{q=0 \\ even}}^{Q_2} \sum_{s=0}^{q} \beta_{m,q,s}^{(2,2)} \tilde{u}_I^{q-s}(n-m-l) \tilde{u}_Q^{s}(n-m-l) \\
& + \sum_{m=0}^{M_3} \sum_{l=1}^{L_3} \sum_{\substack{k=1 \\ odd}}^{K_3} \sum_{r=0}^{k} \beta_{m,k,r}^{(3,1)} \tilde{u}_I^{k-r}(n-m-l) \tilde{u}_Q^{r}(n-m-l) \\
& \sum_{\substack{q=0 \\ even}}^{Q_3} \sum_{s=0}^{q} \beta_{m,q,s}^{(3,2)} \tilde{u}_I^{q-s}(n-m) \tilde{u}_Q^{s}(n-m),
\end{aligned} \tag{A.13}$$

where $\tilde{u}_I(n)$, $\tilde{u}_Q(n)$ and $\hat{s}(n)$ are, respectively, the I and Q components of the inverse model normalized input and the output signals and the parameters $M_1$ and $K_1$ are, respectively, the memory depth and power order on the first summation. Likewise, $M_2$, $L_2$, $K_2$ and $Q_2$ are, respectively, the memory depth, envelope time-shift, aligned envelope power order and time-shifted envelope power order on the second summation. Finally, $M_3$, $L_3$, $K_3$ and $Q_3$ are, respectively, the memory depth, envelope time-shift, time-shifted envelope power order and aligned envelope power order on the third summation. This way, the model has NL order given by $K + Q = max\{K_1, K_2 + Q_2, K_3 + Q_3\}$ and memory length $M + L = max\{M_1, M_2 + L_2, M_3 + L_3\}$.

# APPENDIX B – KRONECKER PRODUCTS

The Kronecker product of the matrices $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ and $\boldsymbol{B} \in \mathbb{C}^{p\times q}$, with size $mp \times nq$, is defined as:

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} a_{1,1}\boldsymbol{B} & a_{1,2}\boldsymbol{B} & \dots & a_{1,n}\boldsymbol{B} \\ a_{2,1}\boldsymbol{B} & a_{2,2}\boldsymbol{B} & \dots & a_{2,n}\boldsymbol{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1}\boldsymbol{B} & a_{m,2}\boldsymbol{B} & \dots & a_{m,n}\boldsymbol{B} \end{bmatrix}. \tag{B.1}$$

In the case of column vectors, $\boldsymbol{a} \in \mathbb{C}^{m\times 1}$ and $\boldsymbol{b} \in \mathbb{C}^{p\times 1}$, the Kronecker and the outer products are related by $\boldsymbol{a} \otimes \boldsymbol{b} = vec(\boldsymbol{b} \circ \boldsymbol{a})$, where $\circ$ is the outer product and $vec$, the vectorization operator. Thus,

$$\boldsymbol{a} \otimes \boldsymbol{b} = vec(\boldsymbol{b}\boldsymbol{a}^T) = \begin{bmatrix} a_1b_1 & \dots & a_1b_p & a_2b_1 & \dots & a_2b_p & \dots & a_mb_1 & \dots & a_mb_p \end{bmatrix}^T, \tag{B.2}$$

with size $mp \times 1$.

The Kronecker product of a vector by itself, represented by $\otimes^2\boldsymbol{a}$, produces redundant cross-terms, as from the previous equation; for example, $a_1a_2$ and $a_2a_1$. In the simple case of $\otimes^2\boldsymbol{a}$, the redundancies can be eliminated multiplying the result by a vectorized lower triangular matrix, as follows:

$$\begin{aligned} vec\left(\begin{bmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & 0 & 0 \\ \vdots & \vdots & \ddots & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}\right)(*) \underbrace{\begin{bmatrix} a_1\boldsymbol{a}^T & a_2\boldsymbol{a}^T & \dots & a_m\boldsymbol{a}^T \end{bmatrix}^T}_{\otimes^2\boldsymbol{a}} \\ = \begin{bmatrix} a_1^2 & a_1a_2 \dots a_1a_m & 0 & a_2^2 \dots a_2a_m & 0 \dots 0 & a_m^2 \end{bmatrix}^T, \end{aligned} \tag{B.3}$$

where $(*)$ is the Hadamard (element-wise) product.

To avoid any redundancies in the general case, the truncated Kronecker product of a vector by itself is defined (FERNANDES; MOTA; FAVIER, 2010) and (FAVIER; KIBANGOU; CAMPELLO, 2003), as follows:

$$\oslash^2 \boldsymbol{a} \;= \boldsymbol{a} \oslash \boldsymbol{a} = \begin{bmatrix} a_1\boldsymbol{a}_1 \\ a_2\boldsymbol{a}_2 \\ \vdots \\ a_i\boldsymbol{a}_i \\ \vdots \\ a_m\boldsymbol{a}_m \end{bmatrix}, \tag{B.4}$$

where $\boldsymbol{a}_i = \begin{bmatrix} a_i & a_{i+1} & \dots & a_m \end{bmatrix}^T$, for $i = 1, \dots, m$. As can be seen, the truncated Kronecker product $\oslash^2 \boldsymbol{a}$ suppresses the symmetric terms in $\otimes^2 \boldsymbol{a}$. Also higher-order truncated Kronecker operators $\oslash^q \boldsymbol{a} \equiv \underbrace{\boldsymbol{a} \oslash \dots \oslash \boldsymbol{a}}_{\text{q}}$ can be defined, recursively, as:

$$\oslash^q \boldsymbol{a} \;= \begin{bmatrix} a_1 \oslash^{q-1} \boldsymbol{a}_1 \\ a_2 \oslash^{q-1} \boldsymbol{a}_2 \\ \vdots \\ a_m \oslash^{q-1} \boldsymbol{a}_m \end{bmatrix}, \tag{B.5}$$

with $\oslash^1 \boldsymbol{a} = \boldsymbol{a}$. These operators allow the representation of triangular Volterra models, such as in eqs. (2.4) and eq. (2.7) and the computation of their regression matrices, as shown in section 2.6.

In section 7.4, the proposed implementation of the model in eq. (7.8) is expressed using a modified truncated Kronecker product, defined as follows:

$$\oslash_-^2 \boldsymbol{a} \;= \begin{bmatrix} a_1\boldsymbol{a}_2 \\ a_2\boldsymbol{a}_3 \\ \vdots \\ a_i\boldsymbol{a}_{i+1} \\ \vdots \\ a_{m-1}\boldsymbol{a}_m \end{bmatrix} \tag{B.6}$$

and the corresponding high-order modified truncated Kronecker operator, $\oslash_-^q$:

$$\oslash_-^q \boldsymbol{a} \;= \begin{bmatrix} a_1 \oslash_-^{q-1} \boldsymbol{a}_2 \\ a_2 \oslash_-^{q-1} \boldsymbol{a}_3 \\ \vdots \\ a_{m-q+1} \oslash_-^{q-1} \boldsymbol{a}_{m-q+2} \end{bmatrix}. \tag{B.7}$$

Next, we provide a few examples:

From eq. (7.8), a pruned-Volterra model similar to the WHFB can be obtained setting $\{L_k\} = 2$:

- For $l = 1$

$$\boldsymbol{\Phi}^{(k,1)}(n) = \tilde{\boldsymbol{u}}(n) \ \otimes \tilde{\boldsymbol{u}}^{(k-1)}(n); \quad \text{for all } k. \tag{B.8}$$

- For $l = 2$

$$\boldsymbol{\Phi}^{(k,2)}_{j=1}(n) = \tilde{\boldsymbol{u}}(n) \ \otimes \left[\tilde{\boldsymbol{u}}^{(k-3)}(n) \oslash_{-} \tilde{\boldsymbol{u}}^{(2)}(n)\right]; \quad \text{for } k \geq 3, \tag{B.9}$$

$$\boldsymbol{\Phi}^{(k,2)}_{j=2}(n) = \tilde{\boldsymbol{u}}(n) \ \otimes \left[\tilde{\boldsymbol{u}}^{(k-5)}(n) \oslash_{-} \tilde{\boldsymbol{u}}^{(4)}(n)\right]; \quad \text{for } k \geq 5. \tag{B.10}$$

In the next example, let us consider the model with the following parameters: $K$=11, $L_7$=3, $L_9$=2 and $L_{11}$=1. The model terms are:

- For $l = 1$

$$\boldsymbol{\Phi}^{(k,1)}(n) = \tilde{\boldsymbol{u}}(n) \ \otimes \tilde{\boldsymbol{u}}^{(k-1)}(n); \quad \text{for } 1 \leq k \leq 11. \tag{B.11}$$

- For $l = 2$

$$\boldsymbol{\Phi}^{(k,2)}_{j=1}(n) = \tilde{\boldsymbol{u}}(n) \ \otimes \left[\tilde{\boldsymbol{u}}^{(k-3)}(n) \oslash_{-} \tilde{\boldsymbol{u}}^{(2)}(n)\right]; \quad \text{for } 3 \leq k \leq 9, \tag{B.12}$$

$$\boldsymbol{\Phi}^{(k,2)}_{j=2}(n) = \tilde{\boldsymbol{u}}(n) \ \otimes \left[\tilde{\boldsymbol{u}}^{(k-5)}(n) \oslash_{-} \tilde{\boldsymbol{u}}^{(4)}(n)\right]; \quad \text{for } 5 \leq k \leq 9. \tag{B.13}$$

- For $l = 3$

$$\boldsymbol{\Phi}^{(k,3)}_{j=1}(n) = \tilde{\boldsymbol{u}}(n) \ \otimes \left[\tilde{\boldsymbol{u}}^{(k-5)}(n) \oslash_{-} (\oslash^{2}_{-} \tilde{\boldsymbol{u}}^{(2)}(n))\right]; \quad \text{for } k = 7. \tag{B.14}$$

This latter example shows that, by choosing $L_k$ values independently, high orders can be included to the model in a modular and more economical way.