

UNIVERSIDADE FEDERAL DE ITAJUBÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM
ENGENHARIA ELÉTRICA

Linearização de Amplificadores de Potência
de RF com Pré-Distorção Digital Ortogonal e
Realimentação Escalar

Henry Douglas Rodrigues

Itajubá, Outubro de 2017

UNIVERSIDADE FEDERAL DE ITAJUBÁ
PROGRAMA DE PÓS-GRADUAÇÃO EM
ENGENHARIA ELÉTRICA

Henry Douglas Rodrigues

Linearização de Amplificadores de Potência
de RF com Pré-Distorção Digital Ortogonal e
Realimentação Escalar

Tese submetida ao Programa de Pós-Graduação
em Engenharia Elétrica como parte dos requisitos
para obtenção do Título de Doutor em Ciências em
Engenharia Elétrica.

Área de Concentração: Microeletrônica

Orientador: Dr. Tales Cleber Pimenta
Coorientador: Dr. Rausley A. A. de Souza

Outubro de 2017
Itajubá

Agradecimentos

Agradeço a Deus pelas bênçãos na minha vida. Porque dele, e por meio dele, e para ele são todas as coisas. A ele, pois, a glória eternamente. Amém!

À minha amada esposa, Daniely, aos meus pais, Antonio e Simone e a todos os meus familiares, por seu incondicional incentivo e amor.

Aos orientadores, Prof. Tales e Prof. Rausley, pela ajuda na realização deste trabalho.

Ao Prof. Luciano Leonel Mendes pelo apoio na publicação dos artigos.

Ao colega Wheberth Damascena Dias pela contribuição no estudo de uma forma alternativa para obter-se o processo de ortogonalização.

Meus mais sinceros agradecimentos.

“A melhor maneira de prever o futuro é criá-lo”.
(Peter Drucker)

Resumo

Sistemas de comunicação sem fio empregam amplificadores de potência de RF, que são dispositivos intrinsecamente não-lineares. A não-linearidade distorce o sinal, gerando produtos de intermodulação que interferem tanto no próprio sinal a ser transmitido, quanto nas frequências adjacentes. Sistemas de linearização com pré-distorção digital são empregados para reduzir a intermodulação na saída de um transmissor. Tradicionalmente, os algoritmos de linearização são baseados em regressão linear e tem como critério de otimização minimizar o erro entre sinais no domínio do tempo. Essa solução apresenta alta complexidade e desempenho limitado. Este trabalho propõe um sistema de pré-distorção inovador em que a métrica a ser otimizada é uma medida custo escalar, proporcional à característica do sinal de saída do amplificador que deseja-se minimizar, como por exemplo, a intermodulação no domínio da frequência. A não-linearidade produzida pela pré-distorção é representada por meio de um modelo comportamental matemático contendo um conjunto de coeficientes. Assim, o algoritmo de linearização é feito ajustando-se cada coeficiente com o objetivo de minimizar a métrica escolhida. Um novo algoritmo de ortogonalização dos coeficientes remove a interação entre a intermodulação de diferentes ordens, resultando em uma convergência mais rápida e uma menor variação de potência na saída do amplificador ao longo das iterações. A identificação dos coeficientes a partir da medida custo é tratada como um problema de otimização numérica, genérico, em que qualquer método iterativo pode ser empregado. Devido à ortogonalização, o problema de otimização é simplificado, uma vez que o ajuste de um dado coeficiente não afeta os demais. A técnica proposta foi simulada computacionalmente, implementada em FPGA e testada com um amplificador de RF Doherty, comprovando seu funcionamento e eficácia. Além de demonstrar estes resultados, o trabalho também traz a fundamentação teórica da linearização através da pré-distorção digital, possibilitando ao leitor uma compreensão mais abrangente do assunto, incluindo sua implementação em tempo real. Esta técnica inovadora pode ser empregada em qualquer amplificador, cujo sinal a ser amplificado apresenta envoltória variante no tempo, o que engloba a grande maioria dos sistemas de comunicação sem fio.

Palavras-chaves: linearização, pré-distorção digital, sistema não-linear, amplificador de potência de RF, intermodulação, otimização numérica.

Abstract

Wireless communication systems employ RF power amplifiers, which are intrinsically non-linear devices. The non-linearity distorts the signal generating intermodulation products, which interfere not only on the transmitted signal, but also on adjacent frequencies. Digital pre-distortion linearization is employed to mitigate the intermodulation in the transmitter output. Traditionally, linearization algorithms are linear regression-based and its optimization criteria is to minimize the error between time-domain signals. This solution presents high complexity and limited performance. This thesis proposes an innovative digital pre-distortion system in which the optimization metric is a scalar cost measurement, proportional to the undesired amplifier's output signal characteristic, e.g., the intermodulation power in the frequency domain. The non-linearity produced by the pre-distortion is represented by a behavioral mathematical model containing a set of coefficients. Therefore, the linearization algorithm is performed by adjusting each coefficient until the chosen metric is minimized. A new coefficient orthogonalization algorithm removes the interaction between intermodulation of different orders, leading to a faster convergence and a small amplifier output power variation across the iterations. The coefficients identification becomes a generic optimization problem, in which any iterative method may be employed. Due to the orthogonalization, the optimization problem is simplified, considering the adjustment of a given coefficient will not affect the others. The proposed technique was computationally simulated, implemented on FPGA, and tested with a RF Doherty amplifier, proving its functionality and efficiency. The underlying theory behind digital pre-distortion linearization is also revised in this work, enabling the reader for a wider understanding of its principles, including its real-time implementation. This innovative technique can be employed with any signal presenting a time-varying envelope, which includes, the great majority of wireless communication systems.

Keywords: linearization, digital pre-distortion, non-linear system, RF power amplifier, intermodulation, numerical optimization.

Lista de ilustrações

Figura 1 – (a) Diagrama de blocos do sistema LINC. (b) Processo de decomposição.	20
Figura 2 – Diagrama de blocos do sistema de linearização com realimentação cartesiana.	21
Figura 3 – Diagrama de blocos do sistema com alimentação direta (ou <i>feed-forward</i>).	21
Figura 4 – (a) Sinal modulado em frequência (4-FSK) apresentando envoltória constante. (b) Sinal com múltiplas portadoras moduladas em quadratura apresentando envoltória variável.	27
Figura 5 – Potência de saída em função da potência de entrada do amplificador.	29
Figura 6 – Natureza dos sinais: (1) Complexo com tempo e amplitude discretos em banda base; (2) Complexo com tempo contínuo e amplitude discreta em banda base; (3) Complexo com tempo e amplitude contínuos em banda base e (4) Real com tempo e amplitude contínuos em banda passante com frequência de portadora igual a f_c . O valor absoluto da largura de faixa do sinal se mantem constante em todos os gráficos.	30
Figura 7 – Distorções não lineares no domínio do tempo, em que $x(n)$ e $y(n)$ representam a entrada e saída do sistema, respectivamente. a) Distorção provocada na magnitude do sinal. O símbolo $ \cdot $ representa o módulo de um sinal complexo. b) Distorção provocada na fase do sinal. O símbolo $\angle\cdot$ representa a fase de um sinal complexo.	31
Figura 8 – Funções de transferência de um sistema não-linear. A reta tracejada ilustra um caso ideal em que o sistema é totalmente linear. a) Função AM/AM mostrando a variação de ganho em função da magnitude do sinal de entrada. b) Função AM/PM mostrando a variação do deslocamento de fase da saída em relação à entrada em função da magnitude do sinal de entrada.	31
Figura 9 – DEP dos sinais de entrada e saída de um sistema não-linear, com uma máscara de emissão arbitrária.	32
Figura 10 – Espectro do sinal de saída do modelo série de potências para um sinal de dois tons como estímulo. Os números sobre cada raia indicam a ordem da intermodulação.	34
Figura 11 – Simulação mostrando o espectro do sinal de saída do modelo série de potências para um sinal de dois tons como estímulo.	34
Figura 12 – Sinal de saída do modelo banda base polinomial. Os produtos de intermodulação se apresentam somente na zona espectral da fundamental.	37
Figura 13 – Funções base do modelo polinomial em banda base para $K = 7$, ou seja, até 13ª ordem.	38

Figura 14 – Colunas da matriz de regressão linear no domínio da frequência, para um sinal de entrada com distribuição Gaussiana.	39
Figura 15 – Funções base do modelo Gaussiano ortogonal em banda base para $K = 7$	40
Figura 16 – Funções base do modelo uniforme ortogonal banda base para $K = 7$, ou seja, até 13ª ordem.	40
Figura 17 – Função de transferência instantânea. Em todos os sistemas utilizou-se o parâmetro $K = 7$. Nas funções com efeito memória acentuado e moderado, usou-se (2.34) com $M = 6$ e $M = 2$, respectivamente. No sistema sem memória empregou-se (2.20).	43
Figura 18 – Diagrama em blocos do sistema não-linear polinomial com memória conforme descrito por (2.34). Cada bloco $f_m(x, h_{i,m})$ corresponde a um sistema com polinômio em banda base sem memória.	44
Figura 19 – Diagrama de blocos do sistema não-linear polinomial com memória conforme descrito por (2.37). Cada bloco $f_m(x, h_{i,m})$ e $p_m(y, g_{i,m})$ corresponde a um sistema com polinômio em banda base sem memória.	45
Figura 20 – Modelos de a) Wiener e b) Hammerstein.	45
Figura 21 – Diagrama de blocos do processo de identificação de um sistema.	48
Figura 22 – Número de condicionamento em função da máxima ordem do polinômio para dois modelos distintos: polinomial banda base segundo (2.20) e ortogonal Gaussiano conforme (2.27).	52
Figura 23 – Convergência do algoritmo LMS para duas equações de modelo diferentes. a) Após $4, 1 \times 10^6$ de iterações para evidenciar a convergência lenta do modelo banda base polinomial de (2.20). b) Após 13000 iterações para evidenciar a convergência rápida do modelo ortogonal Gaussiano segundo (2.27).	54
Figura 24 – Comparação de convergência entre os algoritmos LMS e RPEM. a) Após 16384 iterações para evidenciar a convergência lenta do LMS. b) Após 400 iterações para evidenciar a convergência rápida do algoritmo RPEM.	56
Figura 25 – Diagrama de blocos de um sistema típico de linearização com pré-distorção digital.	57
Figura 26 – Arquiteturas de aprendizado. a) Aprendizado indireto e b) Aprendizado direto.	58
Figura 27 – Magnitude dos sinais em vários pontos do sistema de pré-distorção em função do número da amostra.	59
Figura 28 – Comparação entre as funções de transferência instantânea da magnitude: DPD, PA e do conjunto DPD em série com o PA.	59

Figura 29 – Funções AM/AM e AM/PM entre alguns pontos do sistema, sendo que DPD+PA corresponde à associação série do bloco do DPD com o amplificador.	60
Figura 30 – DEP dos sinais ao longo do sistema de pré-distorção.	61
Figura 31 – Diagrama de blocos do sistema proposto de linearização ortogonal com realimentação escalar (OSFL).	62
Figura 32 – Saída de um sistema não-linear com dois tons de entrada, em que apenas o coeficiente de 5 ^a ordem é não-nulo.	64
Figura 33 – Diagrama ilustrando o conceito de linearidade nos parâmetros.	64
Figura 34 – Cancelando o efeito do ajuste do coeficiente de ordem i nas ordens anteriores. Os coeficientes \mathbf{h}_R são limitados ao coeficiente $(i - 1)$	66
Figura 35 – Saída de um sistema não-linear com dois tons de entrada e coeficientes ortogonalizados. Apenas os produtos de 5 ^a ordem estão presentes nesta figura devido a um ajuste no coeficiente de mesma ordem.	67
Figura 36 – (a) Função custo suave para sinal de entrada periódico; (b) Contorno da função custo e trajetórias de medida custo para o algoritmo <i>Nelder-Mead Simplex</i> com e sem o procedimento de minimização alternada.	70
Figura 37 – Histórico de minimização do ACLR com ortogonalização dos coeficientes. Comparação entre os algoritmos de otimização numérica <i>Nelder-Mead Simplex</i> e <i>Quasi-Newton</i>	72
Figura 38 – Comparação entre o método proposto OSFL e o algoritmo de realimentação escalar sem ortogonalização . Foi medido ao longo das iterações a) ACLR para comparar tempo de convergência, e b) variação de potência de saída.	73
Figura 39 – Desempenho de intermodulação. a) sinal de entrada do DPD, i.e., o sinal não-distorcido, b) sinal de saída do PA sem o uso do DPD, c) sinal de saída do PA com DPD usando o método LS e d) sinal de saída do PA com DPD usando o método proposto OSFL com otimização numérica <i>Nelder-Mead Simplex</i>	74
Figura 40 – Funções de transferência instantâneas sem e com o DPD dadas por (a) AM/AM e (b) AM/PM.	75
Figura 41 – ACLR e número total de iterações para diferentes parâmetros M e K usando o algoritmo OSFL.	76
Figura 42 – NMSE entre os vetores $\mathbf{\Omega}_{iA}$ e $\mathbf{\Omega}_{iB}$ calculados para sequências com diferentes comprimentos N . Simulação realizada para os modelos ortogonal Gaussiano e polinomial em banda base.	77
Figura 43 – (a) Função custo não-suave. (b) Trajetória de otimização do coeficiente de 3 ^a ordem para o algoritmo <i>Pattern search</i>	78

Figura 44 – Histórico de minimização do ACLR no algoritmo <i>Pattern search</i> para seqüências de entrada aleatórias.	79
Figura 45 – Diagrama de blocos do sistema proposto de linearização ortogonal com realimentação de escalar (OSFL), com seleção de diferentes medidas custo.	80
Figura 46 – DEP de sinais para comparação de três métodos: sistema DPD baseado em LS e OSFL com métricas OOB e IB.	81
Figura 47 – Comparando processos de linearização com realimentação escalar usando Gram-Schmidt, quando há somente ortogonalização (equivalente ao OSFL proposto) e com ortonormalização. Comparação feita em termos de (a) ACLR e; (b) Potência de saída do PA ao longo das iterações.	84
Figura 48 – Rotação de um ângulo ϕ a partir de um número complexo c , gerando c'	87
Figura 49 – Região de convergência das rotações CORDIC e processo de transposição de quadrantes.	89
Figura 50 – Diagrama de blocos de um sistema DPD sem memória com ganho complexo implementado com LUT.	92
Figura 51 – Diagrama de blocos do DPD com LUTs para cada tomada de um sistema com memória.	93
Figura 52 – Diagrama de blocos do DPD com LUTs para cada ordem de um sistema com memória.	94
Figura 53 – LUTs com conteúdo puramente real para cada ordem de polinômio nos modelos a) banda base polinomial de (2.20) e b) Gaussiano ortogonal de (2.27).	95
Figura 54 – Erro de quantização das LUTs de cada ordem para os modelos a) banda base polinomial de (2.20) e b) Gaussiano ortogonal de (2.27).	96
Figura 55 – Diagrama de blocos simplificado do <i>hardware</i> implementado em FPGA SoC.	98
Figura 56 – Simulação do código VHDL do bloco DPD no Modelsim.	98
Figura 57 – Tela do Qsys onde é projetado o <i>hardware</i> do sistema embarcado SoC.	99
Figura 58 – Planta baixa da FPGA contendo o projeto compilado e bloco do DPD destacado.	100
Figura 59 – Resultado da conversão do código de <i>Mathscript</i> para C.	101
Figura 60 – Sinal de saída do DAC em FI e intermodulado pelo DPD para validar projeto em FPGA SoC.	102
Figura 61 – Demonstração dos efeitos de vazamento LO de frequência imagem após conversão para RF utilizando banda lateral inferior.	102
Figura 62 – Diagrama da bancada de testes em que o dispositivo BLF888E em configuração Doherty é linearizado com 100 W de potência média em sua saída.	103

Figura 63 – Sinal de saída do PA Doherty: (a) sem DPD; (b)-(f) com DPD; (b) 3ª ordem da 1ª tomada; (c) 5ª e 7ª ordem da 1ª tomada; (d) ordens 9ª a 15ª da 1ª tomada; (e) 3ª e 5ª ordens das 2ª e 3ª tomadas; (f) reajuste de todos os coeficientes para aproximar-se do mínimo global.	104
Figura 64 – Comparação entre métodos de regressão linear e o OSFL proposto: (a) sem DPD; (b) com DPD com o método OSFL; (c) DPD com o método LS	105
Figura 65 – Curvas CCDF para o sinal de saída do PA para níveis de potência distintos: 5 W, 20 W, 50 W, 100 W, 150 W e 200 W. A curva Gaussiana corresponde ao sinal ideal sem distorção.	106
Figura 66 – Saída do amplificador com DPD desligado para vários níveis de potência: 5 W, 20 W, 50 W, 100 W, 150 W e 200 W.	107
Figura 67 – CCDF em diferentes condições e pontos do sistema: (a) entrada do PA sem DPD (Gaussiana de referência); (b) saída do PA sem DPD; (c) entrada do PA com DPD; (d) saída do PA com DPD.	108
Figura 68 – Saída do amplificador com DPD ligado para vários níveis de potência: 5 W, 20 W, 50 W, 100 W, 150 W e 200 W.	109
Figura 69 – Constelação de um sinal OFDM na saída do PA com sistema de DPD desligado para 100 W de saída.	109
Figura 70 – Constelação de um sinal OFDM na saída do PA com sistema de DPD ligado para 100 W de saída.	110
Figura 71 – Diagrama de blocos do sistema proposto de linearização com realimentação de escalar (OSFL), tomando a medida custo usando um circuito analógico.	112
Figura 72 – Projeção do vetor \mathbf{b} no espaço das colunas de \mathbf{A}	115

Lista de tabelas

Tabela 1 – Comparação <i>qualitativa</i> das técnicas de linearização de amplificadores de potência de RF	22
Tabela 2 – Análise <i>quantitativa</i> das técnicas de linearização.	23
Tabela 3 – Termos de (2.27) correspondentes ao modelo polinomial ortogonal Gaussiano.	41
Tabela 4 – Média do número de condicionamento da matriz de regressão linear \mathbf{X} em diferentes modelos para diferentes distribuições do sinal de entrada \mathbf{x}	51
Tabela 5 – Medidas de desempenho do DPD.	61
Tabela 6 – Comparação de desempenho em termos de tempo de convergência e variação de potência de saída. Os algoritmos testados são o de realimentação escalar sem ortogonalização e o OSFL proposto.	73
Tabela 7 – Comparação de desempenho em termos de NMSE e ACLR.	74
Tabela 8 – Comparação de desempenho em termos de NMSE e métricas de intermodulação OOB e IB.	81
Tabela 9 – Tabela dos ângulos necessários para algoritmos CORDIC.	89
Tabela 10 – Classe de algoritmos CORDIC para algumas funções trigonométricas.	91
Tabela 11 – Recursos lógicos para duas abordagens de implementação das LUTs.	95
Tabela 12 – Valores da CCDF da relação de potência pico/média para o sistema de DPD desligado e ligado.	106
Tabela 13 – Comparativo da qualidade do sinal com o uso de DPD em termos de EVM e MER.	108

Lista de algoritmos

1	Lógica recursiva LMS	54
2	Lógica recursiva RPEM	56
3	O algoritmo proposto <i>Orthogonal Scalar Feedback Linearization</i>	71
4	Algoritmo de ortogonalização de Gram-Schmidt.	82
5	Algoritmo de ortogonalização de Gram-Schmidt modificado, baseado-se em matriz de projeção.	83
6	Conversão de um número complexo de coordenadas cartesianas $p + jq$ para polares $Ae^{j\theta}$ com CORDIC.	90

Lista de acrônimos

- ACLR *Adjacent Channel Leakage Rejection* - rejeição do vazamento do canal adjacente
- ADC *Analog to Digital Converter* - conversor analógico-digital
- AM/AM *Amplitude Modulation/Amplitude Modulation* - modulação em amplitude/modulação em amplitude
- AM/PM *Amplitude Modulation/Phase Modulation* - modulação em amplitude/modulação em fase
- ASIC *Application Specific Integrated Circuit* - circuito integrado de aplicação específica
- BER *Bit Error Rate* - taxa de erro de bit
- CCDF *Complementary Cumulative Distribution Function* - função distribuição cumulativa complementar
- CORDIC *COordinate Rotation Digital Computer* - computador digital de rotação de coordenadas
- DAC *Digital to Analog Converter* - conversor digital-analógico
- DC *Direct Current* - corrente contínua
- DDR3 *Double Data Rate 3* - taxa dupla de dados versão 3
- DEP Densidade Espectral de Potência
- DPD *Digital Predistortion* - pré-distorção digital
- DVB-T *Digital Video Broadcasting — terrestrial* - radiodifusão de video digital — terrestre
- EVM *Error Vector Magnitude* - magnitude do vetor de erro
- FI Frequência intermediária
- FIR *Finite Impulse Response* - resposta impulsiva finita
- FPGA *Field Programmable Gate Array* - matriz de portas lógicas programáveis em campo
- IB *In Band* - dentro da banda
- IIR *Infinite Impulse Response* - resposta impulsiva infinita

ISDB-T *Integrated Services Digital Broadcasting — Terrestrial* - radiodifusão de serviços digitais integrados — terrestre

LINC *Linear Amplification with Non-linear Components* - amplificação linear com componentes não-lineares

LMS *Least Mean Squares* - mínimo erro quadrático médio

LO *Local Oscillator* - oscilador local

LS *Least Squares* - mínimos quadrados

LUT *Look-Up Table* - tabela de consulta

MER *Modulation Error Ratio* - taxa de erro de modulação

NMSE *Normalized Mean Squared Error* - erro médio quadrático normalizado

OFDM *Orthogonal Frequency-Division Multiplexing* - multiplexação ortogonal por divisão em frequência

OOB *Out Of Band* - fora da banda

OSFL *Orthogonal Scalar Feedback Linearization* - linearização ortogonal com realimentação escalar

PA *Power Amplifier* - amplificador de potência

PAPR *Peak to Average Power Ratio* - relação entre a potência de pico e a potência média

PC *Personal Computer* - computador pessoal

RF *Radiofrequência*

RMS *Root Mean Square* - valor médio quadrático

ROM *Read Only Memory* - memória de leitura somente

RPEM *Recursive Prediction Error Method* - método de previsão de erro recursivo

SoC *System on Chip* - sistema em chip

SVD *Singular Value Decomposition* - decomposição em valores singulares

VHDL *Very high speed integrated circuits Hardware Description Language* - linguagem de descrição de *hardware* de circuitos integrados de alta velocidade

Lista de símbolos

\cdot	Argumento genérico
$ \cdot $	Módulo
$\angle\cdot$	Fase
$(\cdot)^T$	Transpor
$(\cdot)^\dagger$	Hermitiano: transpor e conjugar
$(\cdot)^{-1}$	Inversa
$\Re\{\cdot\}$	Parte real de uma grandeza complexa
$\Im\{\cdot\}$	Parte imaginária de uma grandeza complexa
$\mathbb{E}(\cdot)$	Valor esperado
$\arg \min_{(\cdot)} f(\cdot)$	Argumento que minimiza a função f
$\binom{n}{k}$	Coefficiente do binômio de Newton
$\frac{\partial}{\partial x} f(\cdot)$	Derivada parcial da função f em relação a x
\in	Pertence
\mathbb{R}	Conjunto dos números reais
\mathbb{C}	Conjunto dos números complexos
X	Matriz - símbolo maiúsculo e em negrito
x	Vetor coluna - símbolo minúsculo e em negrito
\vec{x}	Vetor linha - símbolo minúsculo, em negrito e com seta superior
\hat{h}	Parâmetro estimado - acento circunflexo
<i>n</i>	Escalar - símbolo maiúsculo ou minúsculo, sem negrito em itálico
$J(\cdot)$	Função custo
j	Unidade imaginária $\sqrt{-1}$

Sumário

1	INTRODUÇÃO	19
1.1	Motivação	19
1.2	Revisão bibliográfica e estado da arte	23
1.3	Objetivos e contribuições	25
1.4	Organização do trabalho	25
2	INTRODUÇÃO AOS SISTEMAS NÃO-LINEARES	27
2.1	Visão geral	27
2.2	Sistemas não-lineares	28
2.3	Modelos comportamentais de sistemas não-lineares	33
2.3.1	Série de potências	33
2.3.2	Polinomial banda base	35
2.3.3	Modelos ortogonais	38
2.3.4	Sistemas não-lineares com efeito memória	41
2.3.5	Modelos estruturados em blocos	45
2.4	Conclusão	46
3	IDENTIFICAÇÃO DE SISTEMAS NÃO-LINEARES E PRÉ-DISTORÇÃO	47
3.1	Introdução à identificação de sistemas	47
3.1.1	Métodos de iteração única	47
3.1.2	Métodos de múltiplas iterações	52
3.2	Pré-distorção	56
3.3	Conclusão	61
4	PROPOSTA DE UM SISTEMA DE LINEARIZAÇÃO ORTOGONAL COM REALIMENTAÇÃO ESCALAR	62
4.1	Sistema de DPD com realimentação escalar	62
4.2	Proposta de um sistema de DPD com realimentação escalar ortogonal	63
4.2.1	Ortogonalização dos coeficientes	65
4.2.2	Minimizando a medida custo para um dado coeficiente	68
4.2.3	Lógica do laço exterior	69
4.2.4	Resumo do algoritmo	69
4.3	Simulação do método proposto	71
4.4	Conformidade com sequências aleatórias	75
4.5	Otimização da intermodulação dentro da banda do canal	79
4.6	Processo de ortogonalização alternativo	81

4.7	Processo de ajuste manual	84
4.8	Conclusão	84
5	PRÉ-DISTORÇÃO EM TEMPO REAL	86
5.1	Trigonometria de ponto fixo	86
5.2	Ganho complexo	91
5.2.1	LUTs arranjadas por tomada	92
5.2.2	LUTs arranjadas por ordem	93
5.3	Conclusão	96
6	RESULTADOS EXPERIMENTAIS	97
6.1	Projeto em FPGA SoC	97
6.2	Teste com amplificador de alta potência Doherty	101
6.3	Conclusão	110
7	CONCLUSÕES	111
7.1	Trabalhos futuros	112
	APÊNDICE A – MÉTODO DOS MÍNIMOS QUADRADOS	114
	APÊNDICE B – PUBLICAÇÕES	117
B.1	Artigo aceito em periódico	117
B.2	Artigo submetido em periódico	117
	REFERÊNCIAS	118

1 Introdução

1.1 Motivação

Os sistemas de comunicação sem fio estão evoluindo a cada nova geração. Seja na área de radiodifusão, redes de comunicação móvel ou enlaces de comunicação ponto a ponto, tais sistemas estão exigindo cada vez mais desempenho e confiabilidade. Na maioria destas aplicações, o amplificador de potência (*Power Amplifier* - PA) de radiofrequência (RF) desempenha um papel crucial na camada física. A variação de envoltória do sinal a ser amplificado, combinado ao comportamento não-linear do PA, gera distorções no sinal que resultam em produtos de intermodulação, que interferem nos canais adjacentes e no próprio sinal a ser transmitido. Isso impõe limites no sistema [1], afetando diretamente:

- a máxima potência que pode-se extrair do PA, pois sempre haverá padrões regulatórios que impõe restrições nas emissões fora da faixa do canal como, por exemplo, uma máscara de emissão a ser atendida;
- o custo, uma vez que pode-se extrair menos Watts por unidade monetária investida no circuito de amplificação;
- o alcance do enlace, dado que a distorção dentro da banda aumenta o limiar de recepção;
- a eficiência energética reduz seu valor com a diminuição da razão entre potência de saída de RF e a potência quiescente;
- o tamanho e peso do circuito de amplificação, pois quanto menor a eficiência energética, maior a necessidade de se empregar soluções para dissipação de calor;
- a eficiência espectral, pois a intermodulação gerada fará com que o sinal ocupe uma maior largura de faixa.

Em sistemas que empregam modulação digital com múltiplas portadoras, a potência instantânea do sinal flutua significativamente ao longo do tempo. Isso resulta em um sinal a ser amplificado que apresenta uma alta relação entre a potência de pico e a potência média (*Peak to Average Power Ratio* - PAPR) [2], agravando ainda mais o problema da não-linearidade dos PAs, uma vez que, para uma dada potência de saída, haverá mais picos de sinal ultrapassando o ponto de saturação. Alternativamente, para não ultrapassar a saturação, o alto valor de PAPR implica em uma redução na potência média transmitida.

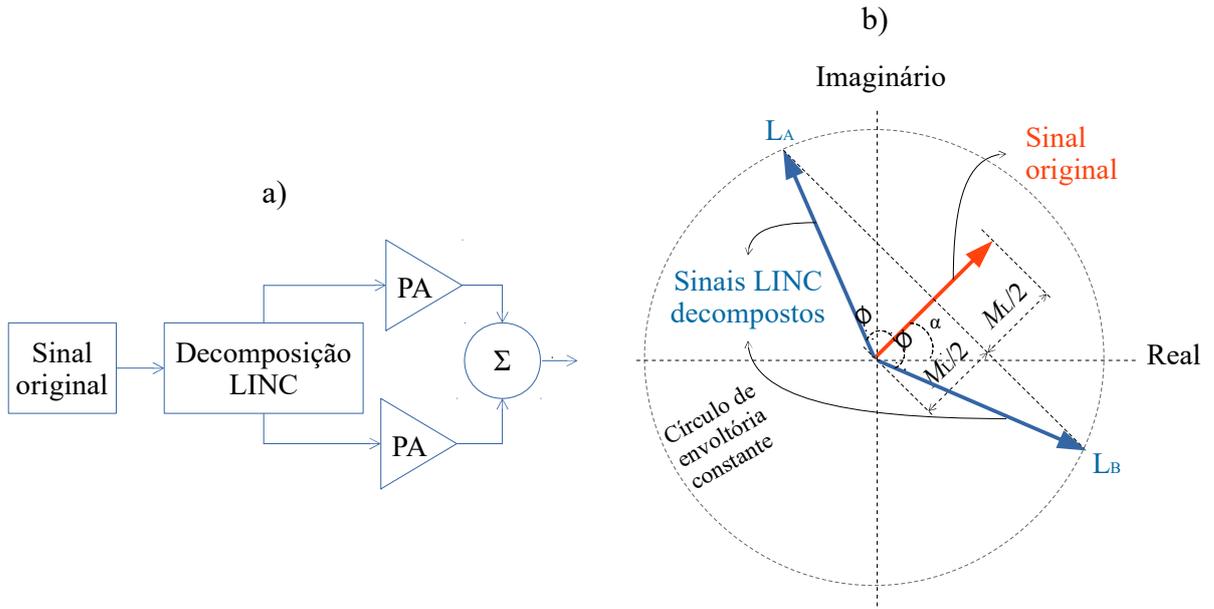


Figura 1 – (a) Diagrama de blocos do sistema LINC. (b) Processo de decomposição.

Existem muitos métodos para resolver o problema da distorção não-linear nos PAs, chamados de técnicas de linearização. A solução trivial é polarizar o amplificador em classe A, garantindo que toda a excursão do sinal se restrinja à região linear. A baixa eficiência energética desse método leva a problemas de consumo, aquecimento e tamanho que devem ser considerados. Outra solução é conhecida como amplificação linear com componentes não-lineares (*Linear Amplification with Non-linear Components* - LINC) [3]. Essa técnica se baseia na decomposição de um sinal de envoltória variável em dois sinais de envoltória constante e modulados em fase, para que depois de linearmente amplificados, sua combinação resulte no sinal original. O diagrama em blocos deste processo é ilustrado na Figura 1, em que L_A e L_B representam as duas componentes decompostas. A magnitude dos sinais decompostos é constante. A modulação de fase é dada por

$$\phi = \arccos\left(\frac{M_L}{2}\right), \quad (1.1)$$

em que M_L é a magnitude do sinal original. A fase final de cada uma das componentes são obtidas por meio de

$$\begin{aligned} \angle L_A &= \alpha + \phi \\ \angle L_B &= \alpha - \phi, \end{aligned} \quad (1.2)$$

em que α é a fase da amostra original. Além de duplicar todo o caminho de transmissão, cada sinal de envoltória constante apresenta uma largura de faixa aproximadamente 10 vezes maior do que a do sinal original [4].

Há também o método de realimentação cartesiana [5] com diagrama de blocos mostrado na Figura 2. Trata-se de um sistema de controle de laço fechado com realimentação negativa, em que \mathbf{x} , \mathbf{y} e \mathbf{z} são os sinais de entrada, compensado e realimentado,

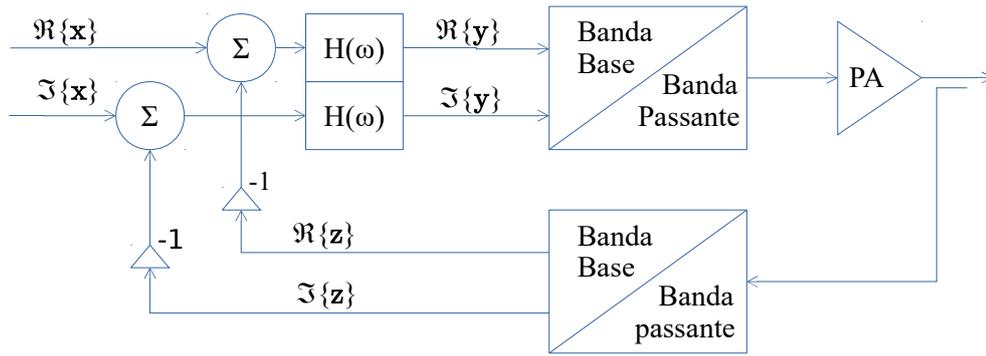


Figura 2 – Diagrama de blocos do sistema de linearização com realimentação cartesiana.

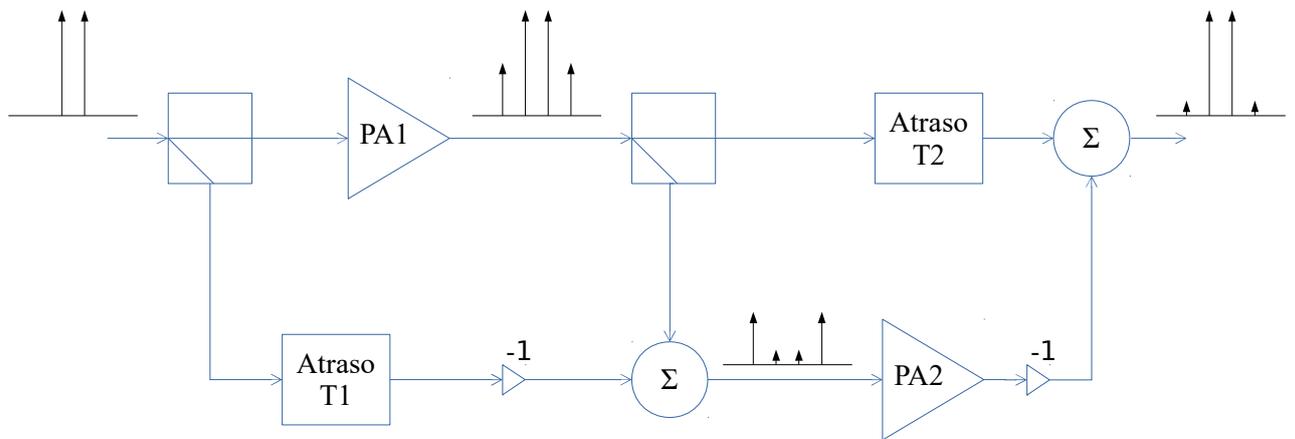


Figura 3 – Diagrama de blocos do sistema com alimentação direta (ou *feed-forward*).

respectivamente. A limitação, como todo sistema de realimentação negativa, é uma largura de faixa reduzida para garantir a estabilidade do laço, que tipicamente varia entre 100 kHz [6] e 5 MHz [7], incluindo os produtos de intermodulação. A principal diferença se dá no fato de que o sinal de erro é complexo (ou cartesiano).

Outra técnica conhecida é a amplificação com alimentação direta (ou *feed-forward*) [8], com diagrama de blocos ilustrado na Figura 3. Os produtos de intermodulação são detectados na saída do PA a partir de uma amostra em contra-fase de sua entrada. Esse sinal deve ser amplificado separadamente por um amplificador linear (PA2) e, em seguida, novamente combinado em contra-fase com o sinal na saída do PA não linear (PA1). Uma vantagem é a correção da intermodulação em largura de faixa da ordem de 125 MHz [9] a 500 MHz [10]. A principal desvantagem é a presença de um segundo amplificador que aumenta a complexidade e reduz a eficiência energética do sistema. Também, existe a pré-distorção de RF [11] que tem desempenho atrelado a circuitos analógicos, limitando o grau de precisão necessário para obter um resultado satisfatório.

O método mais promissor, alvo de inúmeras pesquisas na última década [12–14] e apresentando uma boa relação entre desempenho e complexidade é a linearização com

Tabela 1 – Comparação *qualitativa* das técnicas de linearização de amplificadores de potência de RF [23].

Técnica de linearização	Desvantagens (×) e vantagens (✓)
LINC	× Susceptibilidade à não-simetria dos caminhos de sinal. × Alto nível de complexidade. ✓ PAs podem operar em modo chaveado com alta eficiência.
<i>Feed-forward</i>	× Baixa eficiência energética. ✓ Maior largura de faixa quando comparado à outras técnicas.
Realimentação cartesiana	× Estabilidade do laço limita largura de faixa. ✓ Complexidade baixa.
Pré-distorção de RF	× Precisão limitada do modelo inverso que gera a pré-distorção. ✓ Implementação com baixa complexidade.
Pré-distorção digital	× Largura de faixa limitada pela frequência de amostragem. ✓ Redução da intermodulação com bom desempenho.

pré-distorção digital (*Digital Predistortion* - DPD). Esse se beneficia das vantagens do processamento digital, tais como flexibilidade, precisão, repetitividade e não-susceptível às condições do ambiente. Com o avanço da eletrônica, os dispositivos de processamento digital de sinais operam em taxas cada vez maiores, o que permitiu a implementação da técnica de DPD. Em lógica programável, por exemplo, uma matriz de portas lógicas programáveis em campo (*Field Programmable Gate Array* - FPGA) já é empregada para implementar o sistema de transmissão, que, agora, pode compartilhar recursos com o DPD. Alguns exemplos do processo de linearização com DPD podem ser encontrados em [15, 16] e em suas referências. A ideia consiste em introduzir uma distorção intencional no sinal de entrada do PA, de forma que compense as distorções introduzidas pelo mesmo. Por exemplo, o DPD expande a magnitude do sinal nas regiões em que o PA a comprime, e vice-versa. Ou seja, a associação em cascata do bloco do DPD com o PA tem como objetivo produzir um conjunto linear. As distorções do DPD são produzidas de acordo com um modelo comportamental, representado por uma equação que possui um conjunto de coeficientes. O formato da distorção é controlado pelo valor de tais coeficientes que são calculados de acordo com o algoritmo de linearização. O diagrama de blocos e mais detalhes sobre o sistema de DPD serão apresentados ao longo desse trabalho. Finalmente, a linearização com pré-distorção digital independe da forma de onda a ser amplificada. Como resultado, diversas pesquisas foram publicadas para padrões distintos como: LTE [17], DVB-T2 [18], ATSC [19], WLAN [20], WiMAX [21], WCDMA [22], entre outros.

As Tabelas 1 e 2 comparam as principais técnicas de linearização, de forma qualitativa e quantitativa. De acordo com o teorema da amostragem para sinais complexos, a largura de faixa de um sinal amostrado é igual a frequência de amostragem, f_s . Contudo, a largura de faixa efetiva da pré-distorção digital é uma fração de f_s , devido à banda de transição dos filtros de reconstrução do conversor digital-analógico (*Digital to*

Tabela 2 – Análise *quantitativa* das técnicas de linearização.

Técnica de linearização	Largura de Faixa [MHz]	Redução de intermodulação [dB]
Realimentação cartesiana [6, 7]	0,1 ~ 5	35
<i>Feed-forward</i> [9, 10]	125 ~ 500	20 ~ 40
Pré-distorção digital [24]	$= 0,6 \times f_s$ $= 0,6 \times 200 = 120$	20 ~ 30

Analog Converter - DAC) e anti-aliasing do conversor analógico-digital (*Analog to Digital Converter* - ADC), sendo 80%¹ um valor prático [25].

1.2 Revisão bibliográfica e estado da arte

Sistemas típicos de DPD utilizam algoritmos baseados em regressão linear [26, 27] que têm como objetivo minimizar o erro entre um sinal de referência e um sinal observado, ambos no domínio do tempo. Ao reduzir o erro entre o sinal de entrada do DPD e o de saída do PA no domínio do tempo, a intermodulação no domínio da frequência é indiretamente reduzida. Dado que há um atraso entre esses sinais, é necessário realizar um alinhamento temporal por intermédio de correlação cruzada. É também feita uma normalização de potência entre os dois sinais para garantir um ganho unitário no DPD. A partir das duas sequências de amostras alinhadas no tempo, a regressão linear é finalmente aplicada [28] para obter-se os coeficientes do DPD.

Técnicas de DPD com baixa complexidade que eficientemente reduzem a intermodulação são constantes focos de pesquisa na literatura [29–31]. Algoritmos baseados em regressão linear, tais como o método dos mínimos quadrados (*Least Squares* - LS) [32] ou predição de erro recursivo (*Recursive Prediction Error Method* - RPEM), consomem muitos recursos de processamento, uma vez que é necessário o alinhamento temporal entre os sinais de entrada e realimentação, assim como normalização de magnitude antes que a identificação de sistema seja feita. Além do mais, o critério de otimização de minimizar a energia do erro no domínio do tempo não é o mais adequado considerando que o objetivo final é a redução da intermodulação no domínio da frequência [20].

Para reduzir a complexidade e diretamente otimizar uma dada métrica no domínio da frequência, algoritmos de DPD com realimentação escalar foram propostos [33–36], se opondo às soluções típicas em dois sentidos:

i) o sinal realimentado é uma sequência (ou vetor) de amostras e não é diretamente usado no processo de estimação dos coeficientes do DPD, mas *indiretamente* por meio da tomada de uma medida custo considerando uma métrica no domínio da frequência. Por exemplo, a taxa de erro de modulação (*Modulation Error Ratio* - MER) para reduzir distorções

¹ Esta porcentagem é a relação entre a largura de faixa instantânea e a frequência de amostragem.

dentro da banda do canal, ou a rejeição do vazamento do canal adjacente (*Adjacent Channel Leakage Rejection - ACLR*) com o objetivo de reduzir as emissões fora da faixa do canal. O sinal realimentado se torna um escalar cujo valor é atualizado a cada iteração; ii) o processo de linearização se torna um problema de otimização numérica genérico, ao invés de regressão linear. O algoritmo baseia-se na ideia de iterativamente ajustar o valor de cada coeficiente do modelo com o objetivo de minimizar a medida custo. Além do mais, não há necessidade de alinhamento temporal e normalização de energia entre os sinais de entrada e realimentado.

Em [33], a energia das emissões fora da faixa foi tomada como informação escalar a ser minimizada no algoritmo. Contudo, a desvantagem desta abordagem é o elevado número de parâmetros, uma vez que cada ponto da curva da função de transferência é adaptado. Por exemplo, se a curva for quantizada com 10 bits de resolução, haverá 2048 parâmetros a serem ajustados, uma vez que cada ponto é um número complexo. Em [34] e [35], os autores usaram a potência do canal adjacente ponderada como realimentação escalar e trataram a estimação dos parâmetros como um problema de otimização numérica genérico. O DPD é tratado como um modelo matemático comportamental reduzindo o número de parâmetros a ser ajustado. Essa abordagem aplica as técnicas de otimização em um único coeficiente ou em um grupo de coeficientes para minimizar a medida escalar. Contudo, na saída do DPD, o ajuste de um dado coeficiente gera intermodulação não somente na sua ordem correspondente, mas também nas demais ordens mais baixas. Em outras palavras, o modelo é não-ortogonal em termos de produtos de intermodulação, de forma que o ajuste de um coeficiente de ordem alta afeta os demais de ordem baixa previamente ajustados. Portanto, esse algoritmo resulta em uma convergência lenta. No trabalho de [36] é introduzido o conceito de otimização com múltiplos objetivos, em que tanto a máscara espectral quanto a MER são simultaneamente otimizadas. Não obstante, a não ortogonalidade entre as ordens de intermodulação ainda se mostra presente, resultando em um elevado número de iterações.

Este trabalho apresenta um algoritmo de DPD com realimentação escalar, chamado de linearização ortogonal com realimentação escalar (*Orthogonal Scalar Feedback Linearization - OSFL*), em que a adaptação dos coeficientes torna o modelo ortogonal em termos de intermodulação. O processo de ortogonalização está relacionado à adaptação iterativa dos coeficientes de qualquer modelo de sistema não-linear *linear nos parâmetros*, de modo que um dado coeficiente apenas gere produtos de intermodulação na sua ordem correspondente, e não nas demais ordens mais baixas. Portanto, cada coeficiente pode ser adaptado para minimizar sua ordem de intermodulação, sem afetar o ajuste previamente feito nos demais coeficiente de ordem mais baixa. Com a ortogonalização proposta, o problema de otimização é simplificado ao ajuste de um coeficiente complexo repetido K vezes de forma independente, ao invés de ajustar K coeficientes complexos interdependentes. Simulações e resultados experimentais mostraram que o algoritmo proposto pode efetiva

e eficientemente compensar as distorções não-lineares, reduzindo a intermodulação e a MER.

1.3 Objetivos e contribuições

Dentro desse contexto, o presente trabalho tem como objetivo comprovar a tese de que é possível *obter um algoritmo para identificação dos coeficientes de um modelo de sistema não-linear, de forma ortogonal, a partir de uma medida custo escalar como realimentação, com o objetivo de linearizar um amplificador de potência de RF usando DPD.*

Com o método proposto, a linearização torna-se um problema de otimização numérica, cuja informação a ser minimizada é um escalar com valor atualizado a cada iteração. A principal contribuição deste trabalho é o novo algoritmo para ortogonalização dos coeficientes do modelo de sistema não-linear no domínio da intermodulação, que pode ser utilizado em conjunto com qualquer algoritmo de otimização numérica iterativo. A ortogonalização reduz a complexidade da otimização, fazendo com que cada coeficiente possa ser ajustado de forma independente sem interferir nos demais.

Comparando com técnicas de DPD baseadas em regressão linear, o algoritmo proposto apresenta um melhor desempenho na redução da intermodulação na saída do PA. Quando comparado com métodos de DPD com realimentação escalar, o OSFL apresenta uma convergência mais rápida e uma menor variação da potência de saída ao longo das iterações.

1.4 Organização do trabalho

O trabalho encontra-se organizado como a seguir.

Capítulo 2: é feita uma revisão dos aspectos teóricos básicos necessários para uma boa compreensão da tese proposta, contendo uma breve introdução aos sistemas não-lineares, assim como os principais modelos comportamentais matemáticos para representação de sistemas não-lineares. Os modelos são usados para obter-se um sinal de saída desejado a partir de um sinal de entrada e um conjunto de coeficientes.

Capítulo 3: são apresentados alguns métodos de identificação de sistemas não-lineares, que têm como objetivo obter o conjunto de coeficientes a partir da observação dos sinais de entrada e saída deste sistema. São abordadas técnicas de regressão linear baseadas em inversão matricial ou recursão, que correspondem a algoritmos de única ou múltiplas iterações, respectivamente.

Capítulo 4: apresenta o método inovador proposto para encontrar os coeficientes

do sistema não-linear que lineariza o PA. O algoritmo é explicado com detalhes, assim como sua fundamentação matemática. Finalmente, são apresentados os resultados de simulação relacionados à técnica proposta.

Capítulo 5: traz questões práticas de como implementar um sistema de pré-distorção digital em tempo real. O cálculo da equação do modelo em *hardware* digital traz desafios para os quais são apresentadas as soluções.

Capítulo 6: o método proposto foi implementado em FPGA e testado em um amplificador real. Os resultados experimentais são apresentados.

Capítulo 7: traz as conclusões mostrando as vantagens da técnica proposta, assim como sugestões para trabalhos futuros.

2 Introdução aos sistemas não-lineares

2.1 Visão geral

Sistemas de comunicação empregam PAs para amplificar o sinal transmitido. Esses dispositivos consomem uma porção considerável de energia elétrica utilizada pelo transmissor e apresentam um comportamento não-linear por natureza [37]. Por definição, um sistema não-linear é aquele que apresenta comportamentos diferentes para amplitudes de entrada distintas. Por exemplo, quando há uma variação no ganho do amplificador em função da magnitude do sinal de entrada. Quando a envoltória do sinal apresenta uma magnitude constante, como é o caso da modulação em frequência, não surgem produtos de intermodulação na saída do amplificador. Por outro lado, um sinal de *envoltória variável* que passa por um sistema não-linear sofre uma distorção gerando um efeito indesejado chamado de intermodulação. A intermodulação consiste em novas componentes de frequência que surgem a partir da interação das componentes de frequência do sinal original. Os produtos de intermodulação se apresentam em frequências fora da banda do sinal, interferindo nos canais adjacentes, assim como dentro da banda do sinal, interferindo no próprio sinal a ser transmitido. A Figura 4 ilustra exemplos da representação em banda base de sinais de entrada do amplificador com envoltória constante e variável.

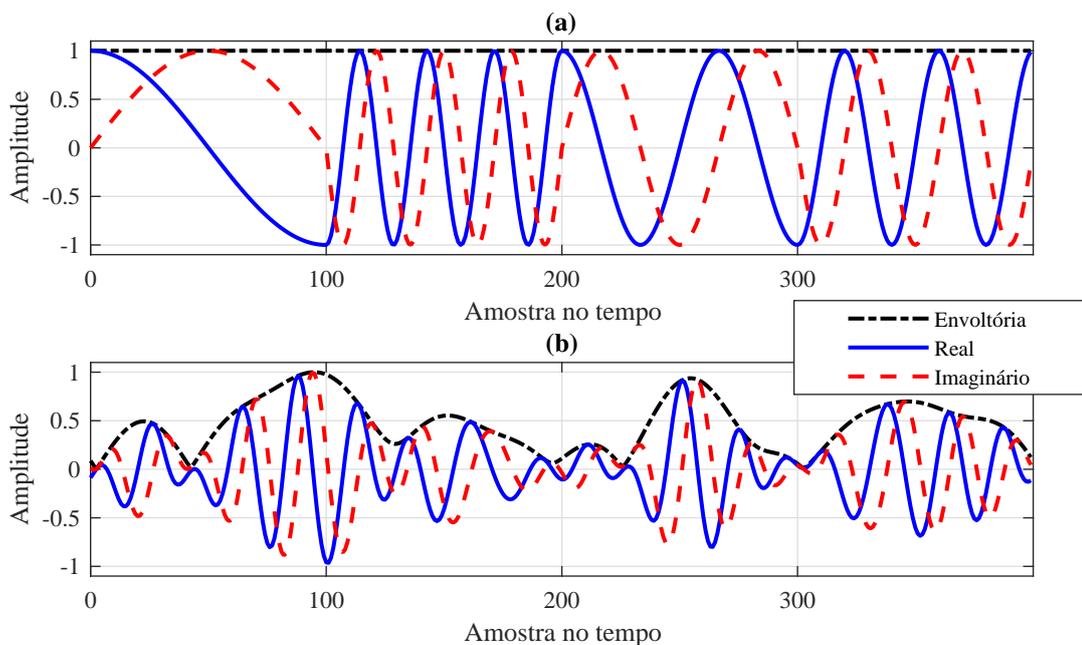


Figura 4 – (a) Sinal modulado em frequência (4-FSK) apresentando envoltória constante. (b) Sinal com múltiplas portadoras moduladas em quadratura apresentando envoltória variável.

A intermodulação gerada pelos PAs do transmissor irá degradar a qualidade do sinal transmitido, assim como gerar um crescimento espectral fora da banda do canal. Todos os sistemas de comunicação sem fio devem respeitar uma máscara de emissão espectral, para garantir a interoperabilidade das canalizações que compartilham o espectro de frequências. Devido à intermodulação, há uma limitação da quantidade de potência de RF que pode ser extraída de um PA para garantir que a máscara seja atendida.

Em sistemas de radiodifusão, como o de televisão digital por exemplo, a comunicação é unidirecional, significando que o processo de amplificação crítico acontece somente no transmissor. Os amplificadores presentes no receptor não são fontes de distorção não-linear pois operam recuados em relação ao ponto de saturação. O desafio no projeto de tais amplificadores é minimizar a figura de ruído. Em sistemas de comunicação móvel, a linearidade do amplificador de transmissão da unidade móvel afeta a duração de sua bateria [38]. Desta forma, independente da aplicação, surge a necessidade de reduzir a distorção não-linear gerada pelo PA, minimizando a intermodulação.

2.2 Sistemas não-lineares

Seja um sistema $\mathcal{F}(\cdot)$ que produz uma saída y para uma dada entrada x , em que

$$y = \mathcal{F}(x). \quad (2.1)$$

Sistemas lineares são aqueles que atendem à propriedade da aditividade

$$\mathcal{F}(x + z) = \mathcal{F}(x) + \mathcal{F}(z), \quad (2.2)$$

assim como a propriedade da homogeneidade, em que

$$y = \mathcal{F}(kx) = k \cdot \mathcal{F}(x), \quad (2.3)$$

para qualquer k . Sistemas não-lineares são aqueles que não atendem tais propriedades. Isso significa que função de transferência de um sistema não-linear é sensível à magnitude do sinal de entrada. Essa característica é claramente notada em amplificadores de potência de RF, no qual o ganho do mesmo varia em função da magnitude do sinal de entrada. O ganho sofre uma compressão à medida em que o nível do sinal de entrada aumenta [39]. A Figura 5 ilustra esse comportamento. Existe uma máxima potência de saída conhecida como ponto de saturação. Acima do ponto de saturação, o sinal é ceifado. Quanto menor for a potência de saída, maior é a distância entre o ponto de operação e a saturação e, portanto, o nível de intermodulação gerada é menor. Contudo, quanto maior o recuo de potência, menor será a eficiência energética do amplificador.

A Figura 6 mostra a natureza do sinal ao longo dos vários pontos de um sistema de comunicação. Na presença de um caminho de realimentação, a natureza do sinal é a mesma

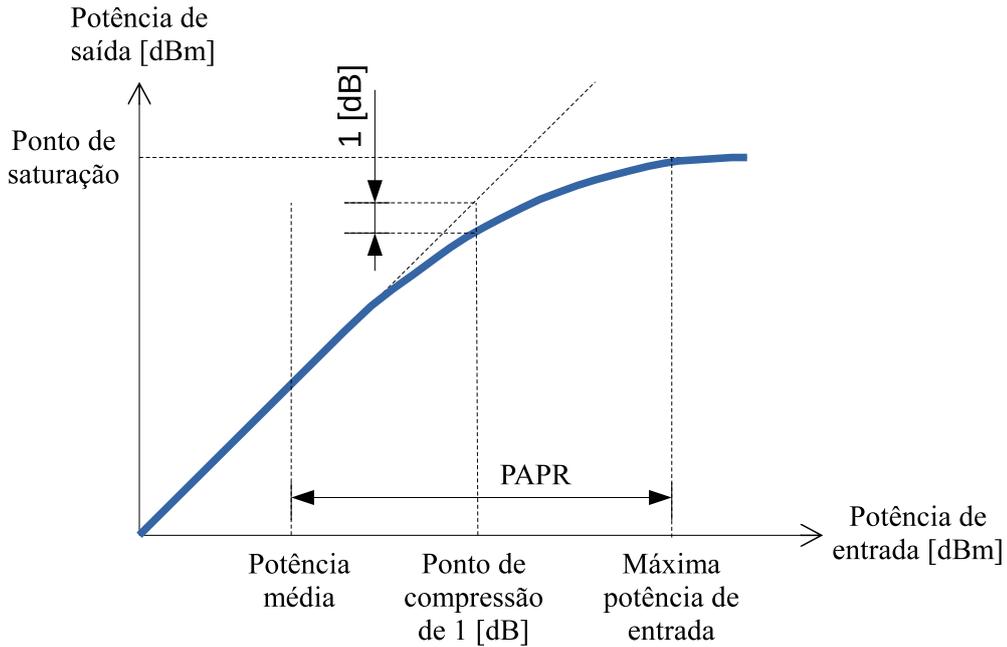


Figura 5 – Potência de saída em função da potência de entrada do amplificador.

em pontos correspondentes. O sinal de entrada do amplificador é sempre puramente real. Contudo, a geração do sinal é dada por sua versão equivalente em banda base, tratando-se de um sinal complexo.

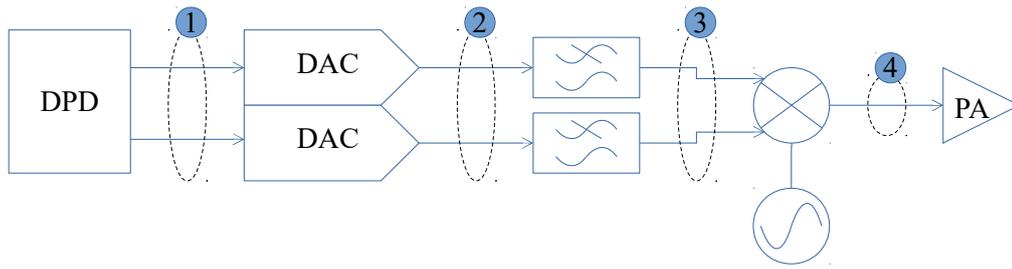
Sejam $x(n)$ e $y(n)$ sinais complexos de tempo discreto na entrada e saída do sistema não-linear, respectivamente, em que n denota a n -ésima amostra no tempo. A não-linearidade provoca distorções no sinal de saída desse sistema, podendo ser notadas tanto no domínio do tempo como da frequência. No domínio do tempo, a distorção pode ser notada quando a magnitude e fase dos sinais são alteradas em função da magnitude do sinal de entrada. A Figura 7 mostra sinais capturados a partir de um amplificador de potência. O sinal de saída é dado pelo sinal de entrada e uma função de transferência complexa [40]

$$y = \mathcal{G}(|x|)x, \quad (2.4)$$

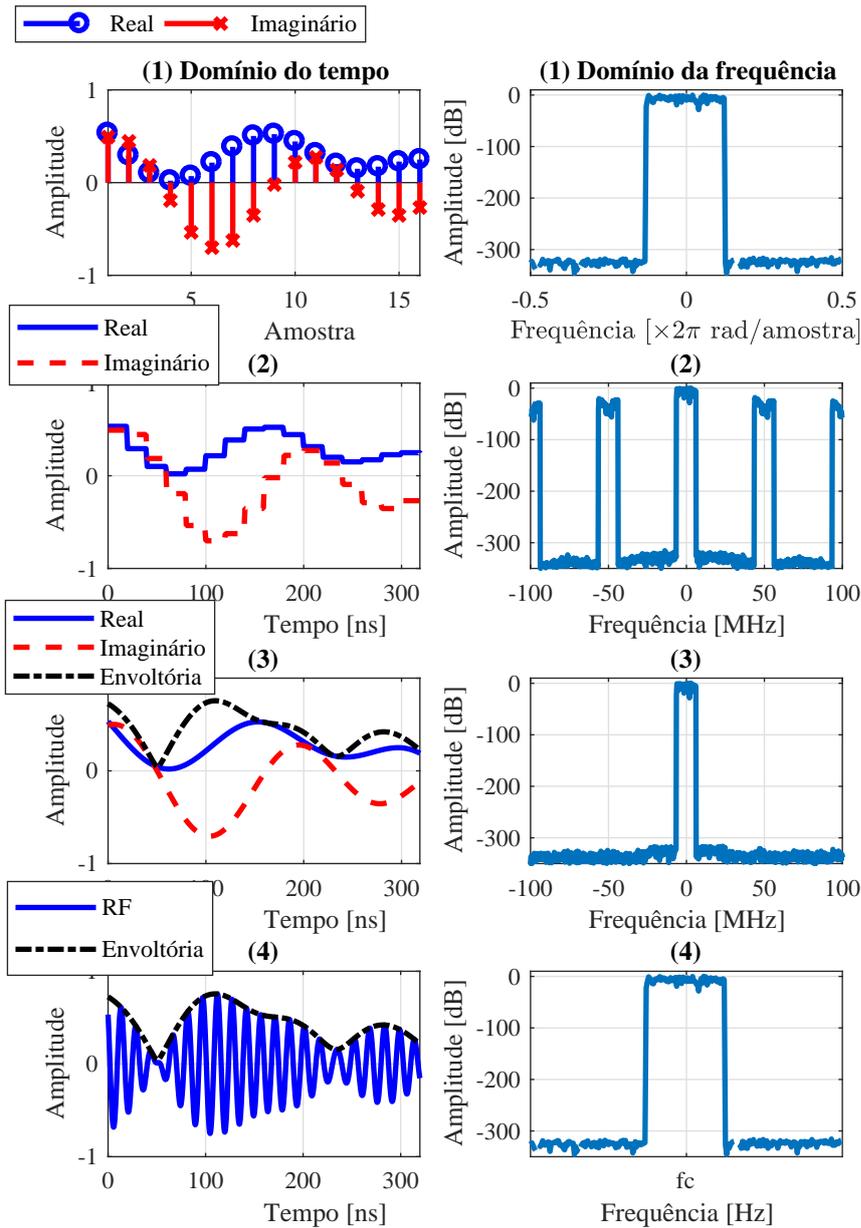
sendo que

$$\mathcal{G}(|x|) = \mathcal{G}_A(|x|)e^{j\mathcal{G}_P(|x|)}. \quad (2.5)$$

Os fatores $\mathcal{G}_A(|x|)$ e $\mathcal{G}_P(|x|)$ são conhecidos como as funções de transferência AM/AM (*Amplitude Modulation/Amplitude Modulation*) e AM/PM (*Amplitude Modulation/Phase Modulation*), respectivamente. Essas siglas denotam o significado de como a amplitude de saída é modulada em função da amplitude de entrada, assim como o deslocamento de fase da saída em relação a entrada depende da amplitude de entrada. A Figura 8 mostra estas funções de transferência, que dependem da magnitude do sinal de entrada, traçadas a partir de sinais capturados na saída de um PA. Note que o amplificador não somente comprime o ganho, mas também o expande para algumas amplitudes.



(a) Diagrama de blocos da cadeia de transmissão.



(b) Sinais no domínio do tempo e da frequência.

Figura 6 – Natureza dos sinais: (1) Complexo com tempo e amplitude discretos em banda base; (2) Complexo com tempo contínuo e amplitude discreta em banda base; (3) Complexo com tempo e amplitude contínuos em banda base e (4) Real com tempo e amplitude contínuos em banda passante com frequência de portadora igual a f_c . O valor absoluto da largura de faixa do sinal se mantém constante em todos os gráficos.

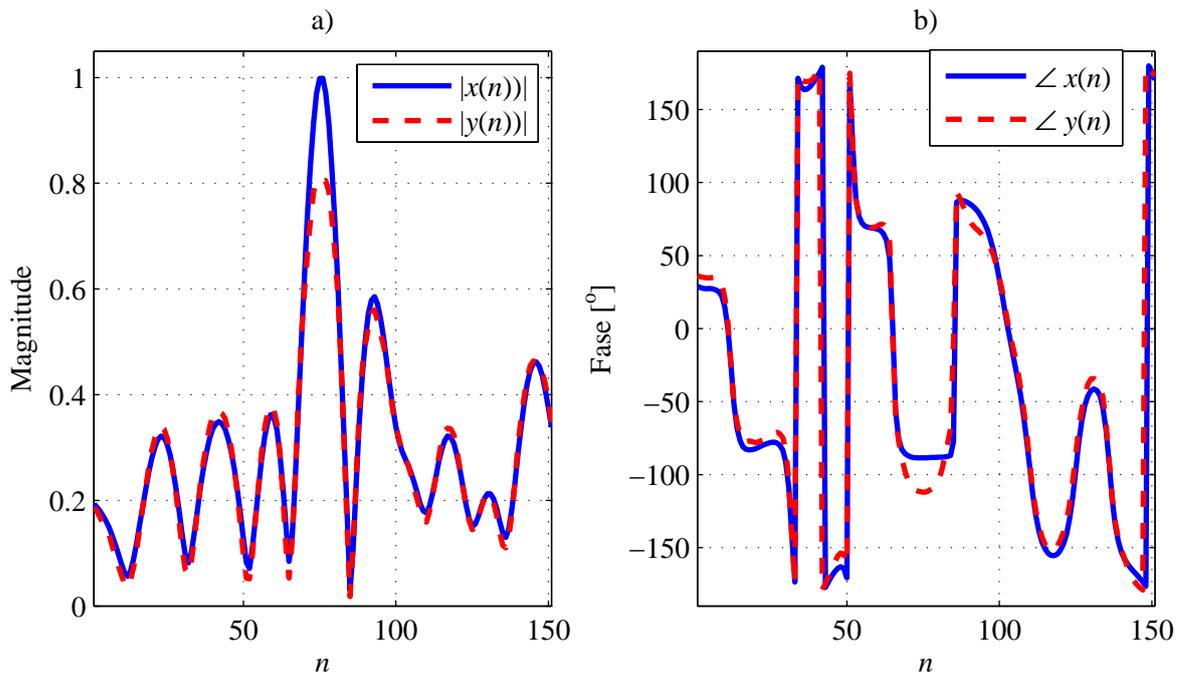


Figura 7 – Distorções não lineares no domínio do tempo, em que $x(n)$ e $y(n)$ representam a entrada e saída do sistema, respectivamente. a) Distorção provocada na magnitude do sinal. O símbolo $|\cdot|$ representa o módulo de um sinal complexo. b) Distorção provocada na fase do sinal. O símbolo $\angle \cdot$ representa a fase de um sinal complexo.

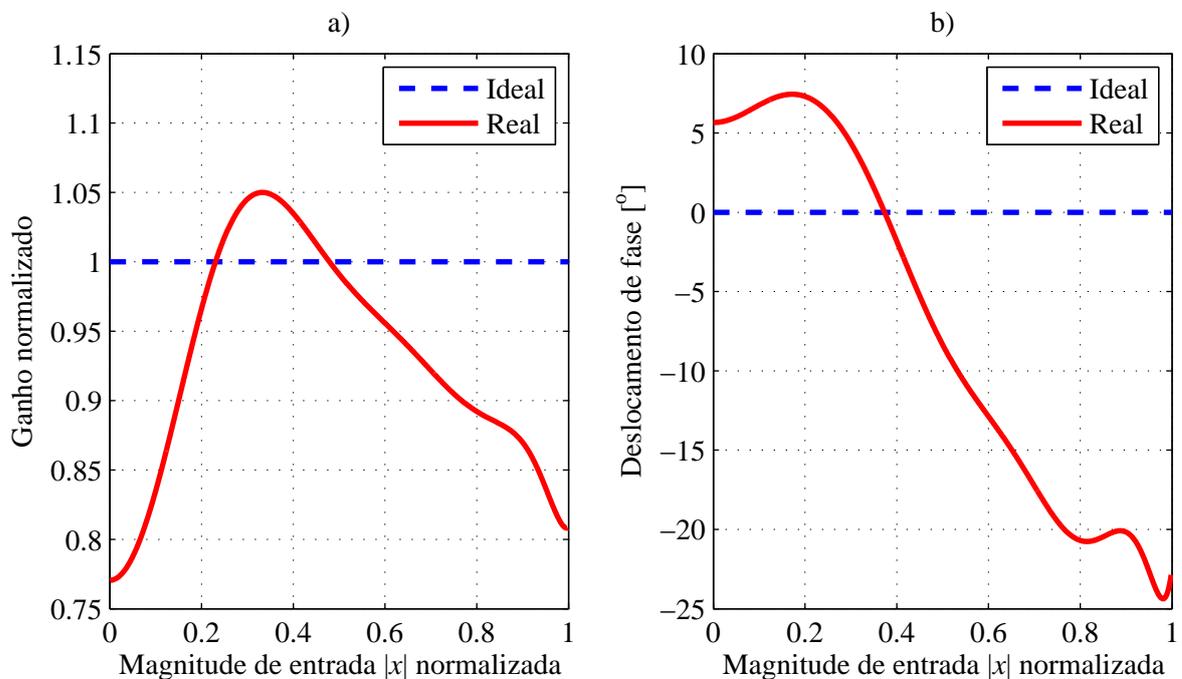


Figura 8 – Funções de transferência de um sistema não-linear. A reta tracejada ilustra um caso ideal em que o sistema é totalmente linear. a) Função AM/AM mostrando a variação de ganho em função da magnitude do sinal de entrada. b) Função AM/PM mostrando a variação do deslocamento de fase da saída em relação à entrada em função da magnitude do sinal de entrada.

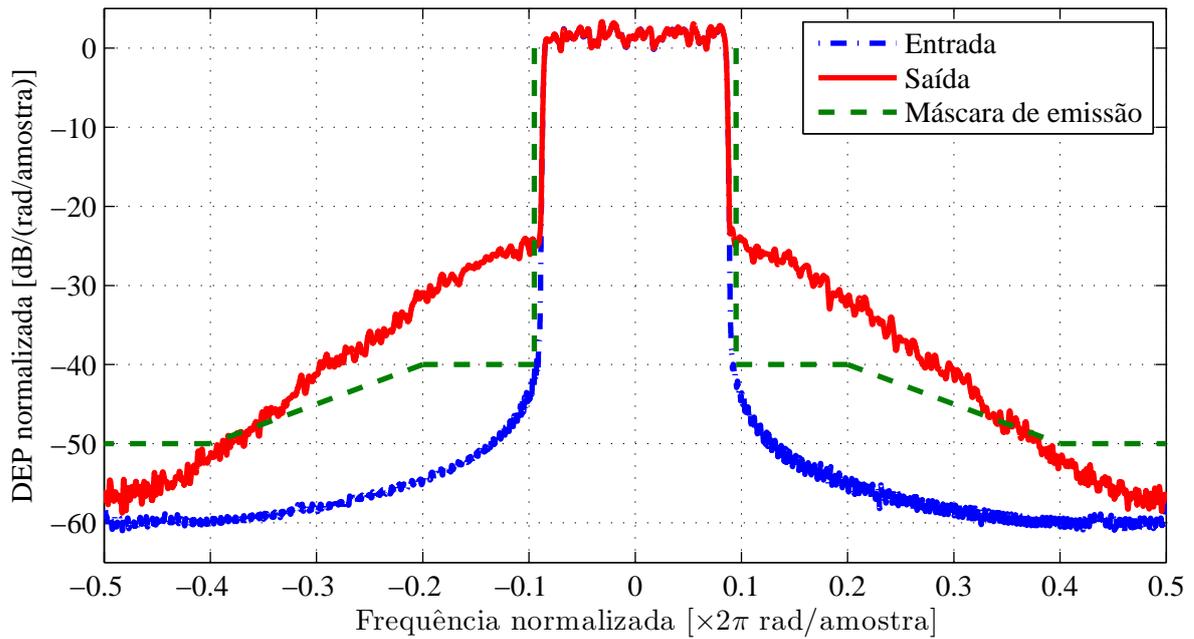


Figura 9 – DEP dos sinais de entrada e saída de um sistema não-linear, com uma máscara de emissão arbitrária.

A intermodulação dentro da banda do canal degrada a qualidade do sinal, afetando diretamente MER e indiretamente a taxa de erro de bit (*Bit Error Rate* - BER), devido ao aumento no limiar de recepção. A Figura 9 mostra a densidade espectral de potência (DEP) normalizada dos sinais na entrada e saída de um sistema não-linear. Está presente na mesma figura uma máscara de emissão arbitrária, mostrando a necessidade de reduzir os produtos de intermodulação para limitar a interferência nos canais adjacentes. Devido à mínima qualidade do sinal aceitável ou à máscara de emissão definida por órgãos regulatórios, a máxima potência de RF fornecida por um amplificador fica limitada pela intermodulação. A intermodulação fora da banda do canal poderia ser removida exclusivamente por meio do uso de um filtro passa faixa, mas dependendo da severidade desta máscara, pode ser necessário o uso de um filtro muito seletivo, de alta complexidade e elevada perda por inserção.

Uma outra solução seria diminuir a potência de RF de saída, afetando o ponto de operação do amplificador. Quanto menor a potência, menor a porção de tempo em que o sinal estará sofrendo ceifamento e, portanto, menor será a compressão. Esta solução funcionaria apenas para amplificadores classe A, cuja distorção se apresenta somente nas potências mais altas, próximas à saturação. Para amplificadores classe AB e Doherty, típicos nos estágios de amplificação de saída de sistemas de alta potência, esta solução é ineficiente, uma vez que o amplificador continua a apresentar distorções não-lineares mesmo para potências baixas. Além do mais, a eficiência energética do amplificador cai com a redução do ponto quiescente de operação. Sendo assim, a linearização do amplificador se faz necessária para reduzir a sua intermodulação gerada.

2.3 Modelos comportamentais de sistemas não-lineares

O modelo comportamental de um sistema é uma equação que produz um sinal de saída em função do sinal de entrada e de um conjunto de coeficientes [41]. Não é o objetivo do modelo descrever o sistema fisicamente. Ele será tratado como uma caixa preta com conteúdo desconhecido, em que apenas descreve-se seu comportamento. O modelo representa um sistema físico com sucesso quando o sinal de saída do modelo é tão próximo quanto possível do sinal de saída medido, para um mesmo sinal de entrada. Serão explicadas as características de cada modelo apresentado, iniciando-se com modelos simples sem memória até chegar em um modelo genérico com memória e suas simplificações. Esta teoria é importante, uma vez que o amplificador e o bloco de pré-distorção são representados por esses modelos. Todos os sinais apresentados nesse trabalho são de tempo discreto, visto que o processamento da pré-distorção ocorre no domínio digital.

2.3.1 Série de potências

A forma mais simples de representar um sistema não linear é por meio de polinômios, na forma de

$$y(n) = \sum_{i=1}^{\tilde{K}} \tilde{h}_i \tilde{x}^i(n), \quad (2.6)$$

em que $\tilde{x}(n)$ é o sinal de entrada em banda passante e \tilde{h}_i são os coeficientes dos polinômios. Esse é um modelo banda passante, cujo sinal de entrada é puramente real e pode ser descrito como

$$\tilde{x}(n) = A(n) \cos(\omega_0 n + \phi(n)). \quad (2.7)$$

Note que as ordens dos polinômios de (2.6) estão completas, incluindo pares e ímpares até a \tilde{K} -ésima ordem. Esse modelo apresenta uma característica indesejável: além dos produtos de intermodulação ao redor das frequências fundamentais, geram-se produtos de intermodulação extras localizados em outras zonas espectrais dos harmônicos e próximo de DC¹. A Figura 10 ilustra esse efeito para um sistema com série de potências de até 7ª ordem, na qual são discriminadas as intermodulações provenientes da soma e diferença das frequências dos dois tons de entrada, assim como os harmônicos. Como a implementação da pré-distorção em tempo real é feito no domínio digital, o aparecimento das demais zonas espectrais além da fundamental faz com que a frequência de amostragem necessária para acomodá-las seja muito alta. Em outras palavras, esse modelo facilmente provocaria o efeito de sobreposição de componentes de frequência (do inglês, *aliasing*). A Figura 11 mostra o resultado de uma simulação para um modelo série de potências de 7ª ordem, com zonas espectrais até o 7º harmônico. Na prática, o objetivo principal da pré-

¹ Define-se corrente contínua (*Direct Current* - DC) como sendo a componente espectral de frequência nula.

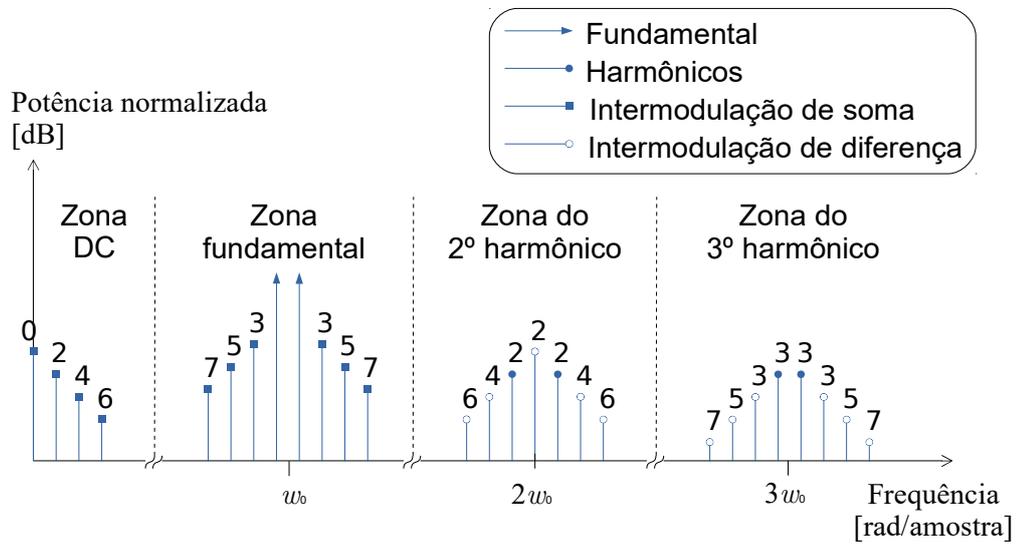


Figura 10 – Espectro do sinal de saída do modelo série de potências para um sinal de dois tons como estímulo. Os números sobre cada raia indicam a ordem da intermodulação.

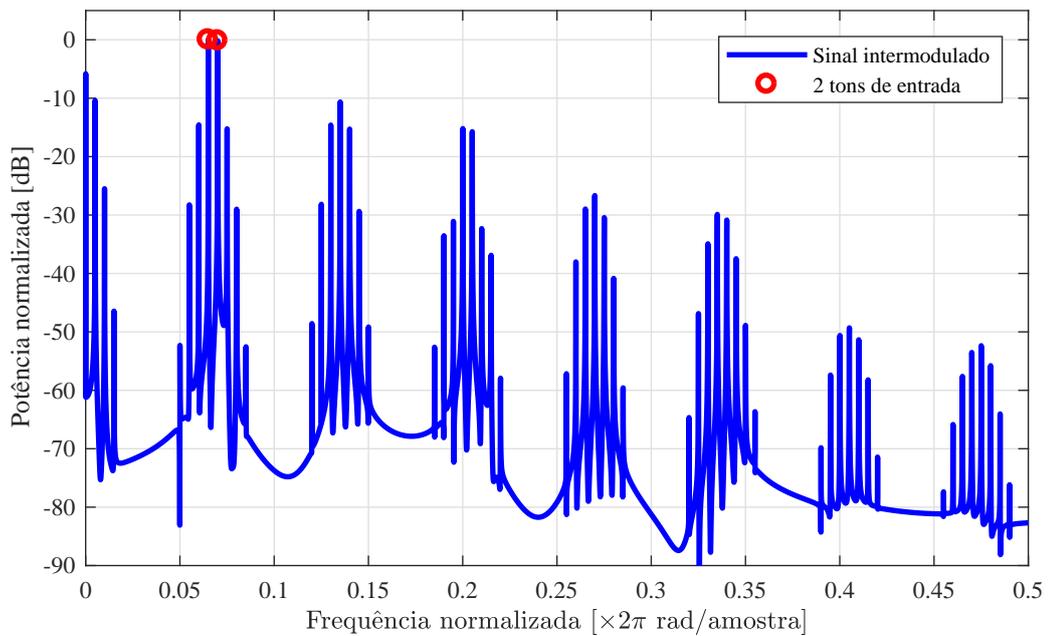


Figura 11 – Simulação mostrando o espectro do sinal de saída do modelo série de potências para um sinal de dois tons como estímulo.

distorção é reduzir os produtos de intermodulação apenas da zona fundamental. Para isto, um outro modelo comportamental deve ser considerado para representar o amplificador.

2.3.2 Polinomial banda base

O termo banda base para este modelo denota que a energia do sinal de entrada está concentrada ao redor da frequência fundamental. Adicionalmente, toda a intermodulação gerada se limita a esta região dentro da zona espectral da componente fundamental. A equação do modelo polinomial banda base é deduzida a partir de (2.6) em conformidade com [42]. Considere o sinal

$$x(n) = A(n)e^{j\phi(n)}, \quad (2.8)$$

como sendo a versão *banda base equivalente*, ou *envoltória complexa* ou *versão analítica* do sinal em banda passante $\tilde{x}(n)$. Em $x(n)$ está contida toda a informação que descreve o sinal original $\tilde{x}(n)$ e está localizada em banda base. A Equação (2.7) pode então ser reescrita como

$$\tilde{x}(n) = \Re \{ x(n)e^{j\omega_0 n} \}. \quad (2.9)$$

Aplicando a decomposição, tem-se

$$\tilde{x}(n) = \frac{1}{2} [x(n)e^{j\omega_0 n} + x^*(n)e^{-j\omega_0 n}], \quad (2.10)$$

em que $(\cdot)^*$ denota o conjugado do sinal.

Como descrito em [42], o sinal de entrada elevado à potência i pode ser expandido na expressão baseada em binômio de Newton

$$\begin{aligned} \tilde{x}^i(n) &= \left\{ \frac{1}{2} [x(n)e^{j\omega_0 n} + x^*(n)e^{-j\omega_0 n}] \right\}^i \\ &= \frac{1}{2^i} \sum_{k=0}^i \binom{i}{k} [x(n)]^k [x^*(n)]^{i-k} e^{j\omega_0(2k-i)n}. \end{aligned} \quad (2.11)$$

A condição para que as componentes do somatório de (2.11) contribuam somente para a primeira zona espectral é dada por $2k - i = \pm 1$, o que também implica em i assumir apenas valores ímpares, ou seja, $k = \frac{i+1}{2}$ ou $k = \frac{i-1}{2}$. Assim, a contribuição para a zona espectral da fundamental é

$$\tilde{x}_{1z}^{i \in \mathcal{I}}(n) = \frac{1}{2^i} \left\{ \binom{i}{\frac{i+1}{2}} [x(n)]^{\frac{i+1}{2}} [x^*(n)]^{i-\frac{i+1}{2}} e^{j\omega_0 n} + \binom{i}{\frac{i-1}{2}} [x(n)]^{\frac{i-1}{2}} [x^*(n)]^{i-\frac{i-1}{2}} e^{-j\omega_0 n} \right\}, \quad (2.12)$$

em que \mathcal{I} é o conjunto de números ímpares positivos. Considerando a propriedade do binômio de Newton

$$\binom{i}{\frac{i+1}{2}} = \binom{i}{\frac{i-1}{2}} \quad (2.13)$$

e também que

$$\begin{aligned} [x(n)]^{\frac{i+1}{2}} [x^*(n)]^{i-\frac{i+1}{2}} &= x(n) |x(n)|^{\frac{i-1}{2}} \\ [x(n)]^{\frac{i-1}{2}} [x^*(n)]^{i-\frac{i-1}{2}} &= x^*(n) |x(n)|^{\frac{i-1}{2}}, \end{aligned} \quad (2.14)$$

tem-se

$$\tilde{x}_{1z}^{i \in \mathcal{I}}(n) = \frac{1}{2^i} \binom{i}{\frac{i-1}{2}} \left\{ x(n) |x(n)|^{\frac{i-1}{2}} e^{j\omega_0 n} + x^*(n) |x(n)|^{\frac{i-1}{2}} e^{-j\omega_0 n} \right\}. \quad (2.15)$$

A versão banda base equivalente de (2.15) é dada por

$$x_{1z}^{i \in \mathcal{I}}(n) = \frac{1}{2^{i-1}} \binom{i}{\frac{i-1}{2}} x(n) |x(n)|^{\frac{i-1}{2}}. \quad (2.16)$$

A série de potências descrita em (2.6) pode ser reescrita para gerar somente produtos de intermodulação na primeira zona espectral

$$y(n) = \sum_{i \in \mathcal{I}} \tilde{h}_i \frac{1}{2^{i-1}} \binom{i}{\frac{i-1}{2}} x(n) |x(n)|^{\frac{i-1}{2}}, \quad (2.17)$$

em que i assume apenas valores ímpares e $K = \frac{\tilde{K} + 1}{2}$. Para que o somatório assumira todos os valores de i mantendo apenas polinômios de ordem ímpar, a equação se torna

$$y(n) = \sum_{i=1}^K \tilde{h}_i 2^{2(1-i)} \binom{2i-1}{i-1} x(n) |x(n)|^{2(i-1)}. \quad (2.18)$$

Fazendo

$$h_i = \tilde{h}_i 2^{2(1-i)} \binom{2i-1}{i-1}, \quad (2.19)$$

finalmente tem-se

$$y(n) = \sum_{i=1}^K h_i |x(n)|^{2(i-1)} x(n). \quad (2.20)$$

Como resultado, o sinal de saída de (2.20) apresentará produtos de intermodulação somente na primeira zona espectral, como ilustrado na Figura 12. As zonas espectrais da segunda harmônica em diante estão normalmente fora da faixa de operação do amplificador, ou seja, não são amplificadas e são normalmente eliminadas pelo uso de filtros de RF. Logo, os únicos produtos de intermodulação realmente necessários para representar o comportamento do PA são aqueles presentes na zona fundamental. Portanto, o modelo polinomial banda base resolve o problema da série de potências, gerando apenas os produtos de intermodulação necessários para o processo de linearização.

Analisando (2.20), observa-se que é possível reescrevê-la por meio de uma multiplicação matricial. Para isso, definem-se os seguintes vetores coluna

$$\mathbf{x} = [x(0), x(1) \dots, x(N-1)]^T \quad (2.21)$$

$$\mathbf{y} = [y(0), y(1) \dots, y(N-1)]^T \quad (2.22)$$

$$\mathbf{h} = [h(1), h(2) \dots, h(K)]^T \quad (2.23)$$

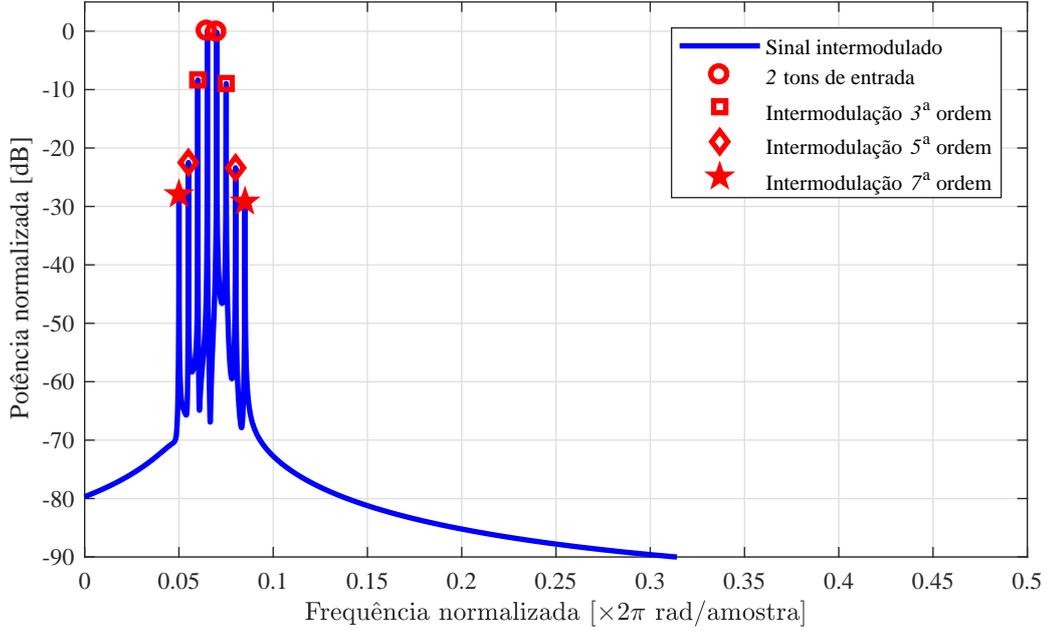


Figura 12 – Sinal de saída do modelo banda base polinomial. Os produtos de intermodulação se apresentam somente na zona espectral da fundamental.

com $\mathbf{x} \in \mathbb{C}^{N \times 1}$, $\mathbf{y} \in \mathbb{C}^{N \times 1}$ e $\mathbf{h} \in \mathbb{C}^{K \times 1}$. A matriz de regressão linear é dada por

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K] \quad (2.24)$$

sendo que $\mathbf{X} \in \mathbb{C}^{N \times K}$ e $\mathbf{x}_i \in \mathbb{C}^{N \times 1}$ se refere a i -ésima coluna de \mathbf{X} dada por

$$\mathbf{x}_i = \begin{bmatrix} |x(0)|^{2(i-1)}x(0) \\ |x(1)|^{2(i-1)}x(1) \\ \vdots \\ |x(N-1)|^{2(i-1)}x(N-1) \end{bmatrix}. \quad (2.25)$$

Finalmente, (2.20) pode ser reescrita como

$$\mathbf{y} = \mathbf{X}\mathbf{h}. \quad (2.26)$$

Analisando o significado das colunas \mathbf{x}_i da matriz de regressão linear \mathbf{X} , percebe-se que cada coluna corresponde a um dado coeficiente h_i e, portanto, à uma dada ordem do polinômio, ou seja, o sinal de saída \mathbf{y} é obtido a partir da combinação linear das colunas de \mathbf{X} ponderadas por \mathbf{h} . Se o vetor de entrada \mathbf{x} for uma sequência de N amostras linearmente espaçadas no intervalo $[0, 1]$, então as colunas \mathbf{x}_i são chamadas de funções base do modelo. Estas funções estão presentes na Figura 13, para $N = 1024$ e $K = 7$. A máxima ordem será $2K - 1$, ou seja, 13ª ordem. Nota-se que a função base de primeira ordem é uma reta e não pode provocar distorção não-linear no sinal de entrada. Esta apenas controla o ganho linear do sistema. Para as demais curvas, cada uma provocará distorções não-lineares de uma dada ordem ímpar. Outra forma de analisar as colunas da matriz de regressão linear

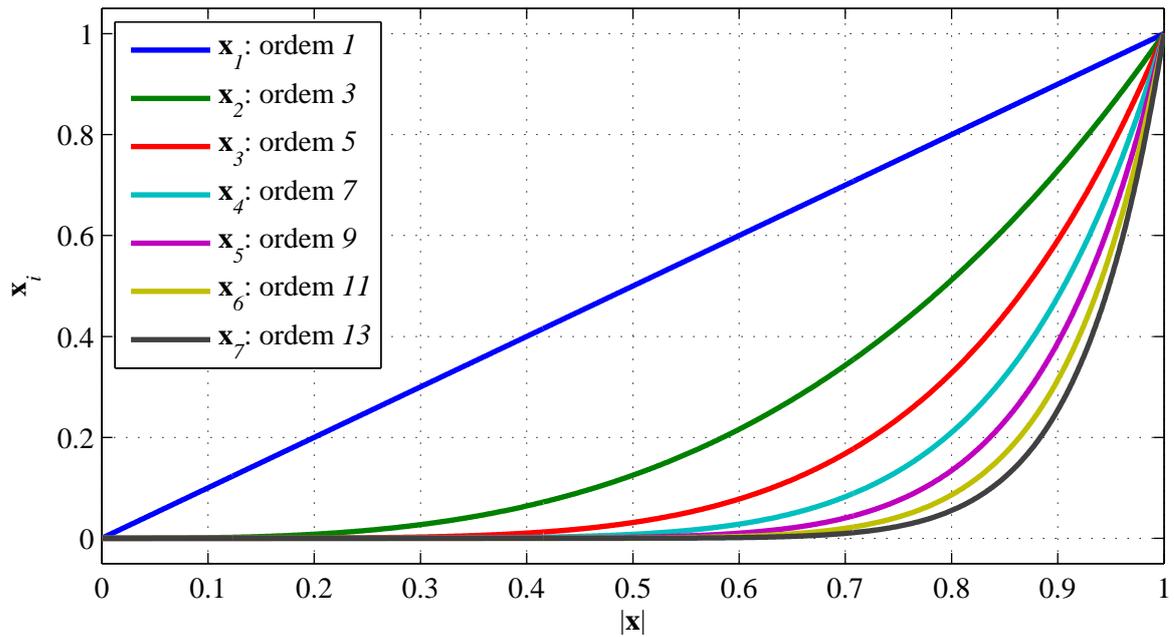


Figura 13 – Funções base do modelo polinomial em banda base para $K = 7$, ou seja, até 13ª ordem.

é considerar que \mathbf{x} corresponde a um sinal com uma dada distribuição. Por exemplo, arbitrariamente faz-se as partes real e imaginária de cada elemento de \mathbf{x} ser uma variável aleatória com distribuição Gaussiana, com média nula e desvio padrão unitário. Tal sinal foi posteriormente limitado em frequência. A Figura 14, ilustra cada uma das colunas \mathbf{x}_i no domínio da frequência a partir da transformada discreta de Fourier.

Ao analisar (2.26) nota-se que o sinal de saída é basicamente o escalonamento complexo das colunas \mathbf{x}_i pelos coeficientes \mathbf{h}_i . Assim, o sinal de saída é a combinação linear de cada coluna \mathbf{x}_i com o seu respectivo coeficiente complexo h_i promovendo um ajuste de magnitude e fase. Nota-se também que a largura de faixa do sinal intermodulado é $2i - 1$ vezes maior do que a largura de faixa do sinal de entrada, para cada ordem de índice i . Em outras palavras, o coeficiente de 3ª ordem gera um sinal com largura de faixa três vezes maior. O coeficiente de 5ª ordem gera um sinal com largura de faixa cinco vezes maior, e assim por diante.

2.3.3 Modelos ortogonais

Durante o processo de identificação de sistemas não-lineares, a ser explicado na Seção 3.1.1, existe um parâmetro denominado número de condicionamento da matriz de regressão linear. Esse parâmetro pode ser minimizado fazendo com que $\mathbb{E}[\mathbf{X}^\dagger \mathbf{X}] = \mathbf{I}$, em que os operadores $\mathbb{E}[\cdot]$ e $(\cdot)^\dagger$ denotam o valor esperado e o Hermitiano do argumento, respectivamente. Ou seja, o modelo pode tornar a matriz \mathbf{X} estocasticamente ortogonal, dependendo da distribuição do sinal de entrada \mathbf{x} . Isso significa que, na média, o produto

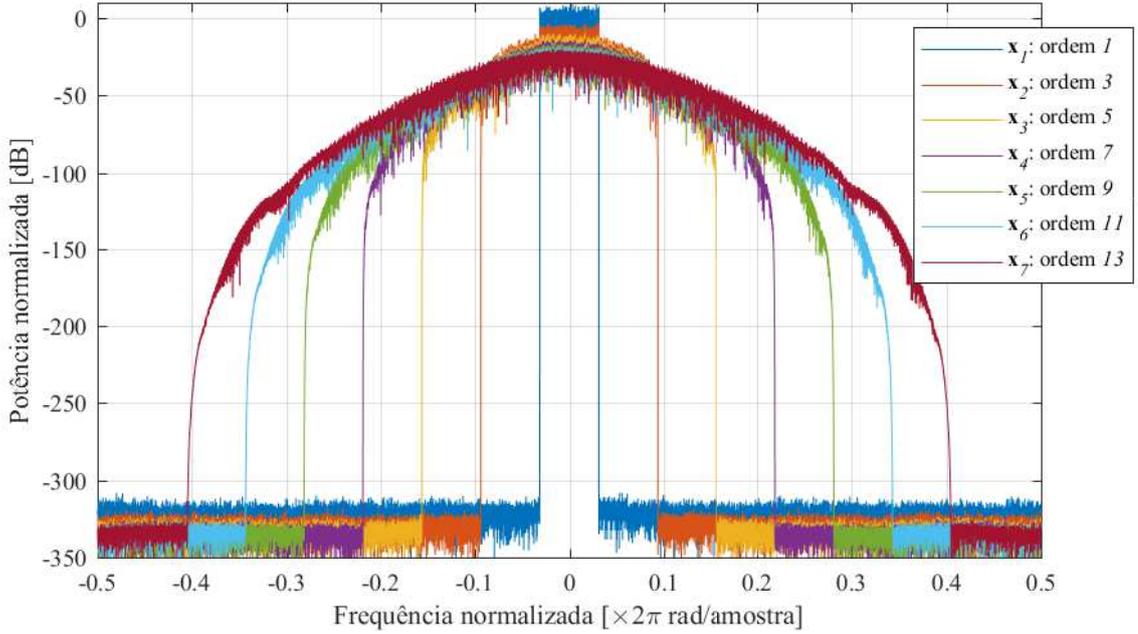


Figura 14 – Colunas da matriz de regressão linear no domínio da frequência, para um sinal de entrada com distribuição Gaussiana.

interno entre duas colunas distintas de \mathbf{X} é nulo. Há alguns modelos ortogonais propostos na literatura que são especializados para determinados tipos de distribuição. Em [43] foi proposto uma equação de modelo que minimiza o número de condicionamento para sinais com distribuição Gaussiana. Nesse modelo, o sinal de saída é dado por

$$y(n) = \sum_{i=1}^K h_i \sum_{l=1}^i \frac{(-1)^{i-l} \sqrt{i}}{l!} \binom{i-1}{l-1} |x(n)|^{2(l-1)} x(n). \quad (2.27)$$

Ao obter a matriz de regressão linear a partir desse modelo, tem-se

$$\mathbf{x}_i = \begin{bmatrix} \sum_{l=1}^i \frac{(-1)^{i-l} \sqrt{i}}{l!} \binom{i-1}{l-1} |x(0)|^{2(l-1)} x(0) \\ \sum_{l=1}^i \frac{(-1)^{i-l} \sqrt{i}}{l!} \binom{i-1}{l-1} |x(1)|^{2(l-1)} x(1) \\ \vdots \\ \sum_{l=1}^i \frac{(-1)^{i-l} \sqrt{i}}{l!} \binom{i-1}{l-1} |x(N-1)|^{2(l-1)} x(N-1) \end{bmatrix}, \quad (2.28)$$

que gera as funções base de acordo com a Figura 15. A Tabela 3 mostra os termos de (2.27) para $K = 5$ (9ª ordem).

Em [44] também foi definido um conjunto de polinômios ortogonais para sinais com distribuição uniforme, embora não seja comum encontrar tais sinais em aplicações práticas. A sua equação é dada por

$$y(n) = \sum_{i=1}^K h_i \sum_{l=1}^i (-1)^{l+i} \frac{(i+l)!}{(l-1)!(l+1)!(i-l)!} |x(n)|^{l-1} x(n). \quad (2.29)$$

A Figura 16 ilustra as funções base para esse caso.

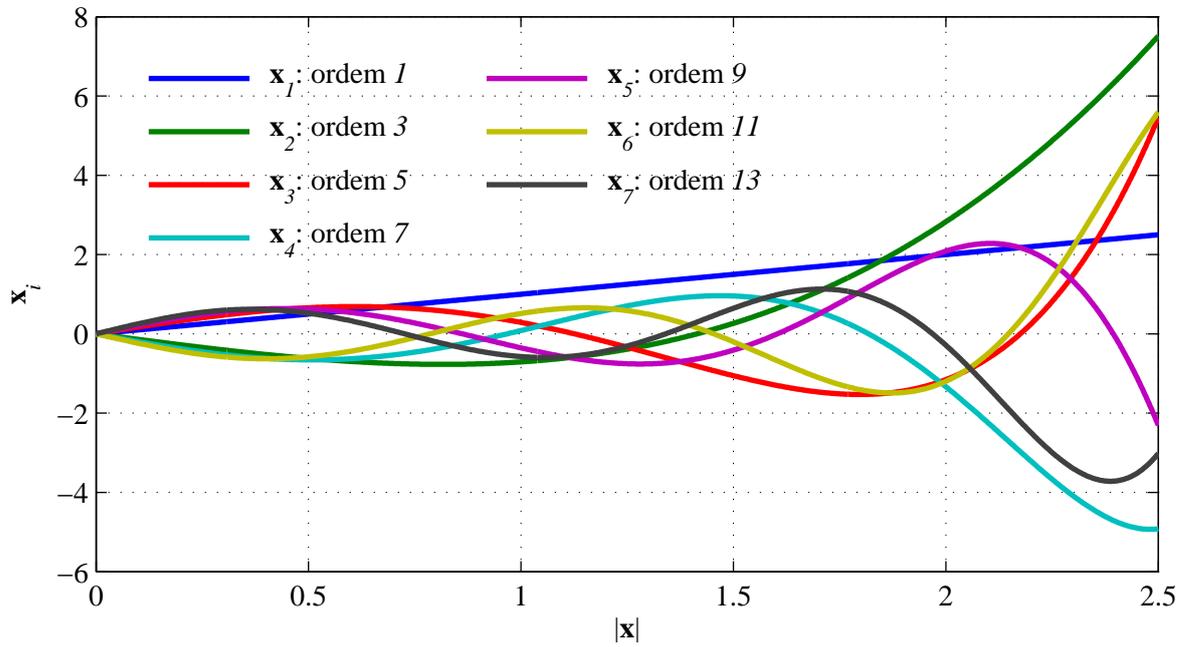


Figura 15 – Funções base do modelo Gaussiano ortogonal em banda base para $K = 7$.

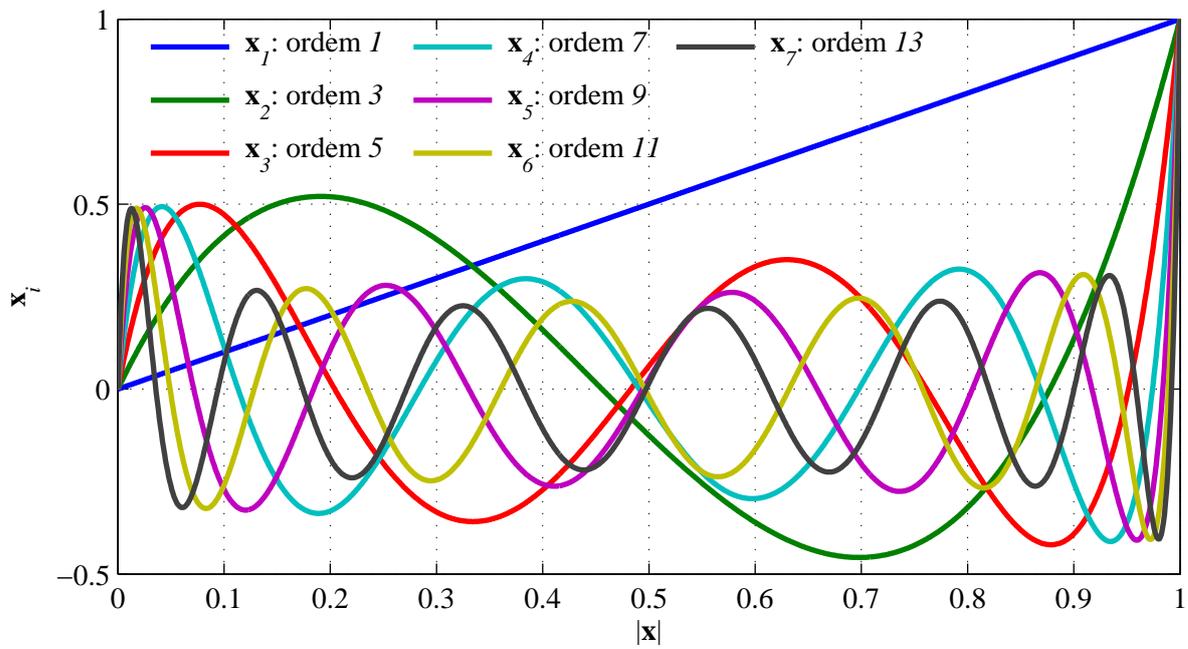


Figura 16 – Funções base do modelo uniforme ortogonal banda base para $K = 7$, ou seja, até 13ª ordem.

Tabela 3 – Termos de (2.27) correspondentes ao modelo polinomial ortogonal Gaussiano.

Índice da ordem	Polinômio base correspondente a coluna \mathbf{x}_i
$i = 1$	$x(n)$
$i = 2$	$\left[-\sqrt{2} + \frac{\sqrt{2}}{2} x(n) ^2\right] x(n)$
$i = 3$	$\left[\sqrt{3} - \sqrt{3} x(n) ^2 + \frac{\sqrt{3}}{6} x(n) ^4\right] x(n)$
$i = 4$	$\left[-2 + 3 x(n) ^2 - x(n) ^4 + \frac{1}{12} x(n) ^6\right] x(n)$
$i = 5$	$\left[\sqrt{5} - 2\sqrt{5} x(n) ^2 + \sqrt{5} x(n) ^4 - \frac{\sqrt{5}}{6} x(n) ^6 + \frac{\sqrt{5}}{120} x(n) ^8\right] x(n)$

2.3.4 Sistemas não-lineares com efeito memória

Por definição, um sistema sem memória é aquele que cuja saída no instante atual depende da entrada somente no mesmo instante. Todos os modelos apresentados anteriormente não possuem memória. Em sistemas com memória (ou dinâmicos), a saída no instante atual depende da entrada não somente no instante atual, mas também dos instantes anteriores [45]. Um filtro de resposta impulsiva finita (*Finite Impulse Response - FIR*) é um exemplo de sistema com memória, uma vez que a linha de atraso faz com que a saída dependa de mais de uma amostra da entrada. O fato é que PAs apresentam distorções não-lineares com memória [46]. Esta característica deve-se a diferentes fatores como: realimentação do sinal de saída para a entrada, desacoplamento de dreno insuficiente (em um transistor de efeito de campo), variações instantâneas na temperatura da junção, entre outros. Logo, um sistema de DPD sem memória não poderá compensar os efeitos de um PA com memória.

O modelo não-linear com memória mais genérico presente na literatura é o de Volterra [47], dado por

$$y(n) = \sum_{i=1}^K \sum_{m_1=0}^{M_1-1} \cdots \sum_{m_i=0}^{M_i-1} h_i(m_1, m_2, \dots, m_i) \prod_{j=1}^i x(n - m_j). \quad (2.30)$$

em que $h_i(m_1, m_2, \dots, m_i)$ são os coeficientes do modelo, correspondendo a uma função i -dimensional. A saída y no instante n depende da entrada não somente do instante atual, correspondendo a $m_j = 0$, mas nos instantes passados, ou seja, para $m_j > 0$. Assim, o comprimento de memória do i -ésimo termo é dado por M_i . Abrindo o primeiro somatório,

tem-se

$$\begin{aligned}
y(n) = & \sum_{m_1=0}^{M_1-1} h_1(m_1)x(n - m_1) + \\
& \sum_{m_1=0}^{M_1-1} \sum_{m_2=0}^{M_2-1} h_2(m_1, m_2)x(n - m_1)x(n - m_2) + \\
& \vdots \\
& \sum_{m_1=0}^{M_1-1} \cdots \sum_{m_K=0}^{M_K-1} h_K(m_1, \dots, m_K) \prod_{j=1}^K x(n - m_j).
\end{aligned} \tag{2.31}$$

Percebe-se que o primeiro termo corresponde à equação de convolução, podendo produzir somente distorções lineares. Do segundo termo em diante, as expressões se apresentam de forma parecida com a convolução, mas o produtório entre as versões do sinal de entrada geram as distorções não-lineares. Basicamente, um sistema não-linear com memória é aquele que gera tanto distorções lineares quanto não-lineares. Percebe-se também que o número de coeficientes aumenta exponencialmente com a ordem. Por exemplo, os coeficientes $h_2(m_1, m_2)$ do segundo termo correspondem a uma função tridimensional. Genericamente, os coeficientes $h_K(m_1, m_2, \dots, m_k)$ correspondem a uma função K -dimensional. Por esta razão, é interessante aplicar uma simplificação para reduzir o número de coeficientes. Ao tornar

$$h_i(m_1, m_2, \dots, m_i) = \begin{cases} h_{i,m}, & \text{para } m = m_1 = m_2 = \dots = m_i \\ 0, & \text{caso contrário} \end{cases} \tag{2.32}$$

a equação (2.30) é simplificada para

$$y(n) = \sum_{i=1}^K \sum_{m=0}^{M-1} h_{i,m} [x(n - m)]^i \tag{2.33}$$

em que M é a profundidade de memória. Mais uma vez, o termo do sinal de entrada elevado a potência i será alterado para gerar apenas produtos de intermodulação na região fundamental, como a seguir

$$y(n) = \sum_{i=1}^K \sum_{m=0}^{M-1} h_{i,m} |x(n - m)|^{2(i-1)} x(n - m). \tag{2.34}$$

Esse modelo é conhecido como polinômio com memória [48].

A principal diferença entre os modelos sem e com memória pode ser notada na função de transferência instantânea traçada na Figura 17. Para geração dessa figura, um sinal com distribuição Gaussiana correlacionada passou por três modelos distintos. Nesse gráfico, a cada instante de tempo correspondente a um valor de amostra n , produz-se um ponto cujas coordenadas cartesianas são os módulos do sinal de entrada versus a saída. O efeito memória provoca um espalhamento nos pontos quando comparado com o sistema sem memória. Este espalhamento é provocado pelo passado histórico das amostras de

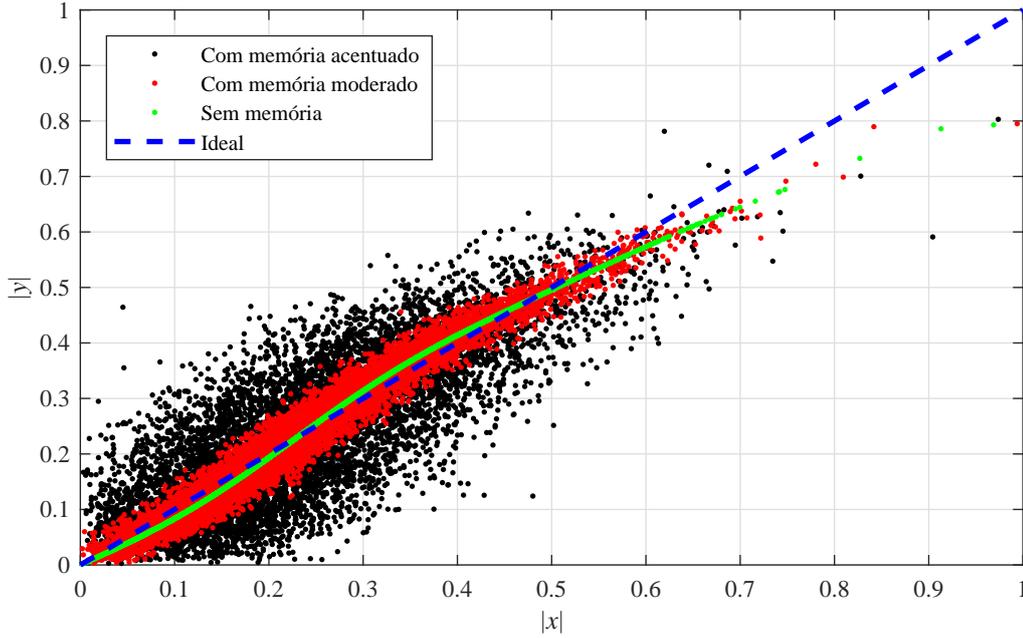


Figura 17 – Função de transferência instantânea. Em todos os sistemas utilizou-se o parâmetro $K = 7$. Nas funções com efeito memória acentuado e moderado, usou-se (2.34) com $M = 6$ e $M = 2$, respectivamente. No sistema sem memória empregou-se (2.20).

entrada, desde o instante $n - (M - 1)$ até o instante n . Quanto maior a profundidade de memória, maior o espalhamento.

Ao analisar (2.34), abrindo o somatório em m , tem-se

$$\begin{aligned}
 y(n) = & \sum_{i=1}^K h_{i,0} |x(n)|^{2(i-1)} x(n) + \\
 & \sum_{i=1}^K h_{i,1} |x(n-1)|^{2(i-1)} x(n-1) + \\
 & \vdots \\
 & \sum_{i=1}^K h_{i,M-1} |x(n-(M-1))|^{2(i-1)} x(n-(M-1)).
 \end{aligned} \tag{2.35}$$

Note que cada termo corresponde a (2.20), porém com uma versão do sinal de entrada atrasada de m amostras. Assim, cada um dos m termos corresponde a um sistema polinomial sem memória com seu próprio conjunto de coeficientes $h_{i,m}$. Contudo, o somatório de todos os termos resulta em um sistema polinomial com memória. A Figura 18 ilustra o diagrama de blocos equivalente a (2.35). Observe que esse diagrama de blocos assemelha-se ao de um filtro FIR. Seria exatamente igual se os polinômios representados pelos blocos $f_m(x, h_{i,m})$ fossem multiplicadores por constantes, ou, as tomadas (do inglês, *taps*) do filtro. Sendo assim, é possível concluir que o sistema polinomial com memória é equivalente a um filtro de resposta impulsiva finita com tomadas não-lineares.

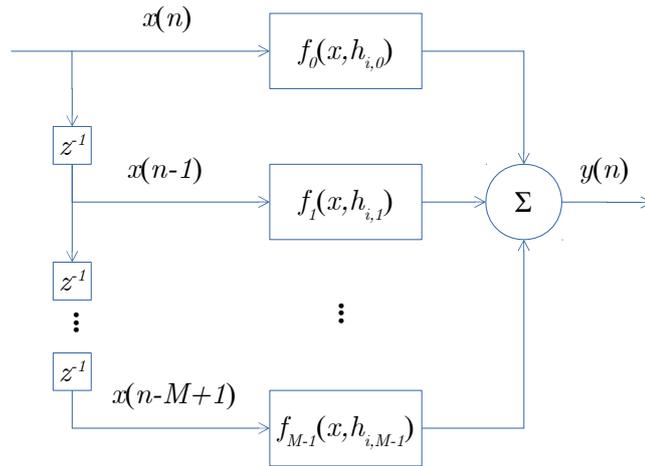


Figura 18 – Diagrama em blocos do sistema não-linear polinomial com memória conforme descrito por (2.34). Cada bloco $f_m(x, h_{i,m})$ corresponde a um sistema com polinômio em banda base sem memória.

A partir desta conclusão, pode-se estender o conceito do polinômio com memória. A forma mais genérica de descrever o comportamento de sistemas lineares de tempo discreto é dada pela equação de diferenças

$$y(n) = \sum_{m=0}^{M-1} b_m x(n-m) + \sum_{m=1}^{R-1} a_m y(n-m), \quad (2.36)$$

em que b_m e a_m são os coeficientes constantes da parte não-recursiva e recursiva do sistema, respectivamente. Logo, R corresponde à profundidade de memória da parte recursiva. A equação (2.36) pode ser modificada para contemplar a geração de distorções não-lineares. Para isto, os coeficientes de ganho linear b_m e a_m devem ser substituídos por polinômios descritos em (2.20). Desta forma, um novo modelo polinomial com memória recursivo surge como

$$y(n) = \sum_{m=0}^{M-1} \sum_{i=1}^K h_{i,m} |x(n-m)|^{2(i-1)} x(n-m) + \sum_{m=1}^{R-1} \sum_{i=1}^K g_{i,m} |y(n-m)|^{2(i-1)} y(n-m). \quad (2.37)$$

O diagrama de blocos do referido sistema é ilustrado na Figura 19. A principal vantagem desta estrutura é que devido à realimentação da saída, ou recursão dada pelos coeficientes $g_{i,m}$, a profundidade de memória se estende além do número de coeficientes. Contudo, a recursividade pode provocar instabilidade quando não implementada de maneira adequada. Este sistema corresponde a um filtro de resposta impulsiva infinita (*Infinite Impulse Response - IIR*) com tomadas não-lineares.

Finalmente, alguns modelos sem memória podem ser alterados para contemplar o efeito memória. Por exemplo, o Gaussiano ortogonal de (2.27) pode ser modificado para

$$y(n) = \sum_{m=0}^{M-1} \sum_{i=1}^{K_m} h_{i,m} \sum_{l=1}^i \frac{(-1)^{i-l} \sqrt{i}}{l!} \binom{i-1}{l-1} |x(n-m)|^{2(l-1)} x(n-m). \quad (2.38)$$

Contudo, ao acrescentar o efeito memória, a propriedade de ortogonalidade é eliminada.

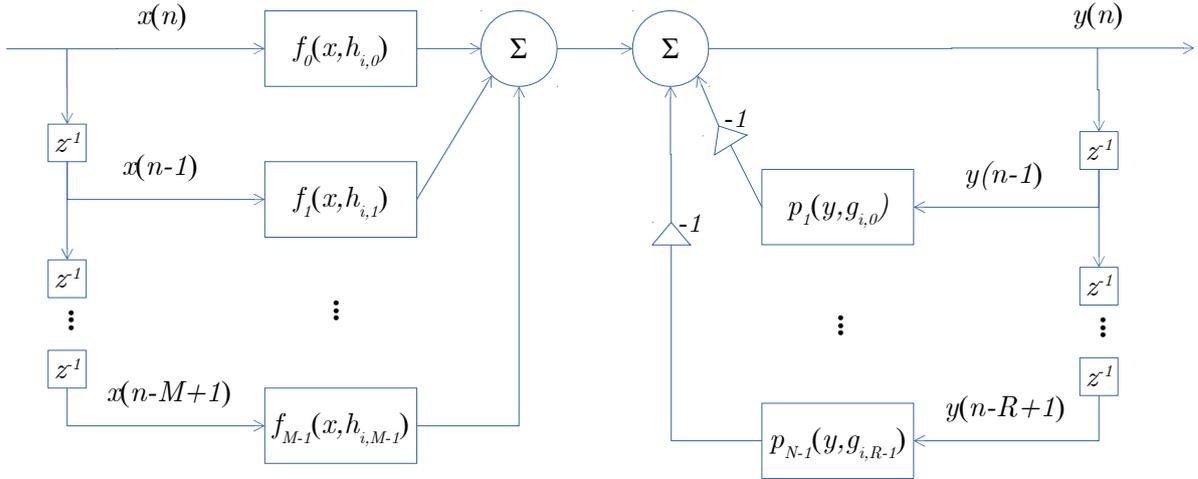


Figura 19 – Diagrama de blocos do sistema não-linear polinomial com memória conforme descrito por (2.37). Cada bloco $f_m(x, h_{i,m})$ e $p_m(y, g_{i,m})$ corresponde a um sistema com polinômio em banda base sem memória.

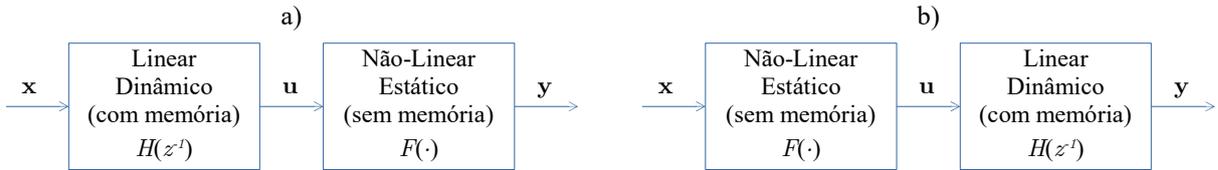


Figura 20 – Modelos de a) Wiener e b) Hammerstein.

2.3.5 Modelos estruturados em blocos

Existem modelos na literatura formados por meio da associação de blocos de sistemas não-lineares estáticos (sem memória) com lineares dinâmicos (com memória). O conjunto desses dois blocos formam um sistema não-linear com memória. A Figura 20 mostra os mais conhecidos sistemas com blocos associados em cascata, em que $F(\cdot)$ e $H(z^{-1})$ denotam a não-linearidade estática e linearidade dinâmica, respectivamente. O bloco linear dinâmico pode ser representado pela equação de diferenças como em (2.36). O bloco não-linear estático pode ser representado por qualquer modelo não-linear sem memória apresentado anteriormente como, por exemplo, o polinomial banda base de (2.20).

A equação que descreve o sistema de Wiener [49] é dada por

$$y(n) = \sum_{k=1}^K c_k \left(\sum_{m=0}^{M-1} h_m x(n-m) \right)^k, \quad (2.39)$$

enquanto que o modelo de Hammerstein [26] é representado por

$$y(n) = \sum_{m=0}^{M-1} h_m \left(\sum_{k=1}^K c_k x^k(n-m) \right). \quad (2.40)$$

Note que apesar de se adotar um polinômio simples para representar o bloco não-linear estático, o sistema de Wiener é *não-linear nos parâmetros*, uma vez que os

coeficientes h_m estão elevados a potência k . Esta é uma desvantagem a ser explicada no Capítulo 3.

2.4 Conclusão

Esse capítulo apresentou noções sobre um sistema não-linear, necessárias para o entendimento de pré-distorção digital. Discutiu-se sobre o comportamento típico de um PA e a natureza do sinal ao longo do diagrama de blocos do sistema, além de apresentar uma seção sobre os modelos comportamentais de sistemas não-lineares. O modelo é necessário para representar o comportamento de um sistema não-linear a partir de uma equação matemática, obtendo o sinal de saída a partir do sinal de entrada e um conjunto de parâmetros. Tal modelo pode ser utilizado para representar tanto o amplificador quanto o bloco de pré-distorção.

3 Identificação de sistemas não-lineares e pré-distorção

O projeto de um sistema de linearização de amplificadores é subdividido em duas grandes partes: identificação de sistemas e pré-distorção em tempo real. Neste capítulo, são apresentados os conceitos básicos sobre estas partes. Na parte de identificação são apresentados diferentes métodos divididos em duas categorias: de única ou múltiplas iterações. Os métodos apresentados neste capítulo são baseados em regressão linear, tendo como objetivo minimizar um sinal de erro no domínio do tempo. Na parte de pré-distorção é feita uma explanação conceitual, detalhando os sinais e a relação entre os mesmos nos vários pontos do sistema.

3.1 Introdução à identificação de sistemas

Nos modelos apresentados no capítulo anterior, o sinal de saída é obtido em função do sinal de entrada e um conjunto de coeficientes. O processo de identificação de um modelo consiste em obter os coeficientes a partir da observação dos sinais de entrada e saída de um sistema. O objetivo é calcular os coeficientes \mathbf{h} do modelo, a partir da mesma entrada \mathbf{x} , para que o modelo produza uma saída $\hat{\mathbf{y}}$ semelhante o suficiente com a saída do sistema a ser modelado \mathbf{y} . O processo de identificação tem como objetivo encontrar os coeficientes do modelo que minimizam

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}. \quad (3.1)$$

A Figura 21 representa esse processo.

Uma figura de mérito pode ser usada para medir quão preciso foi o processo de identificação, isto é, medir a semelhança entre os vetores \mathbf{y} e $\hat{\mathbf{y}}$. Esta medida é o erro quadrático médio normalizado (*Normalized Mean Squared Error* - NMSE), definido como

$$\text{NMSE [dB]} = 10 \log \left[\frac{(\mathbf{y} - \hat{\mathbf{y}})^\dagger (\mathbf{y} - \hat{\mathbf{y}})}{\mathbf{y}^\dagger \mathbf{y}} \right]. \quad (3.2)$$

em que \mathbf{y} e $\hat{\mathbf{y}}$ são vetores coluna contendo a sequência de amostras complexas dos sinais.

3.1.1 Métodos de iteração única

Todos os modelos de sistemas não-lineares apresentados na Seção 2.3, com exceção de Wiener, são lineares nos parâmetros. Isso significa que a equação do modelo pode ser

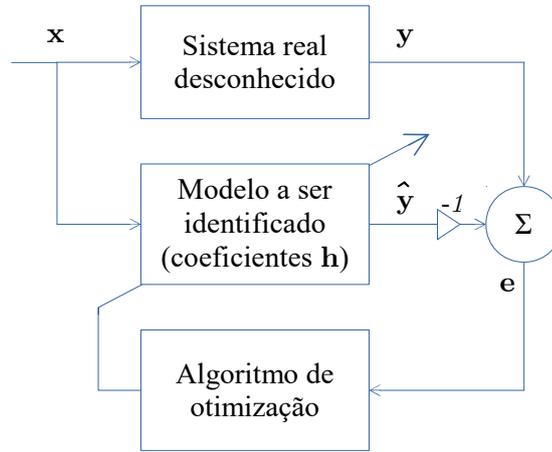


Figura 21 – Diagrama de blocos do processo de identificação de um sistema.

representada por uma multiplicação de matrizes como em (2.26). Detalhando esta equação tem-se

$$\begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(N-1) \end{bmatrix} = \begin{bmatrix} X_{0,1} & X_{0,2} & \cdots & X_{0,K} \\ X_{1,1} & X_{1,2} & \cdots & X_{1,K} \\ \vdots & \vdots & \ddots & \vdots \\ X_{N-1,1} & X_{N-1,2} & \cdots & X_{N-1,K} \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_K \end{bmatrix} \quad (3.3)$$

sendo que os elementos $X_{n,i}$ são obtidos a partir da equação do modelo em função do sinal de entrada \mathbf{x} , tal como apresentado em (2.24). As variáveis \mathbf{y} e \mathbf{X} são conhecidas e deseja-se encontrar os valores dos coeficientes \mathbf{h} . Em (3.3) tem-se um sistema de equações lineares do tipo: i) subdeterminado se $N < K$, ii) determinado se $N = K$ e iii) super determinado se $N > K$. Normalmente, o número de amostras observadas é muito maior que o número de coeficientes a ser identificado, ou seja, $N \gg K$ resultando em um sistema de equações lineares super determinado. Não há necessariamente solução exata para esse sistema de equações. Logo, a estratégia do processo de identificação é encontrar valores para \mathbf{h} que minimizem o NMSE. Na abordagem padrão para solucionar esse problema, multiplicam-se os dois membros de (2.26) pela Hermitiana de \mathbf{X} , resultando em

$$\mathbf{X}^\dagger \mathbf{X} \mathbf{h} = \mathbf{X}^\dagger \mathbf{y}, \quad (3.4)$$

que é a equação normal cuja solução¹ é dada pelo método LS,

$$\hat{\mathbf{h}} = (\mathbf{X}^\dagger \mathbf{X})^{-1} \mathbf{X}^\dagger \mathbf{y}, \quad (3.5)$$

em que $(\cdot)^{-1}$ indica a inversão da matriz. A matriz $(\mathbf{X}^\dagger \mathbf{X})^{-1} \mathbf{X}^\dagger$ é chamada de pseudo-inversa de \mathbf{X} . A dedução de (3.5) a partir da minimização do erro quadrático é demonstrada no Apêndice A. A saída do modelo identificado é dada por $\hat{\mathbf{y}} = \mathbf{X} \hat{\mathbf{h}}$.

¹ Na plataforma MATLABTM, a equação (3.5) é implementada pelo operador *contra barra* por meio do comando: $\mathbf{h} = \mathbf{X} \backslash \mathbf{y}$.

Há um outro algoritmo para solução do LS conhecido como fatoração QR [50], que simplifica o cálculo de $\hat{\mathbf{h}}$ em (3.5) ao aplicar a decomposição ortogonal-triangular. Esta fatoração, faz com que uma dada matriz seja representada por uma multiplicação de duas outras tal como $\mathbf{A} = \mathbf{QR}$, em que $\mathbf{Q} \in \mathbb{C}^{N \times N}$ é uma matriz ortogonal e $\mathbf{R} \in \mathbb{C}^{N \times K}$ é uma matriz triangular superior. Uma matriz unitária \mathbf{Q} garante que $\mathbf{Q}^\dagger \mathbf{Q} = \mathbf{I}$. Enquanto que na matriz triangular superior do tipo

$$\mathbf{R} = \begin{bmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,K} \\ 0 & r_{2,2} & \cdots & r_{2,K} \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & r_{K,K} \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}. \quad (3.6)$$

Aplicando esta decomposição na matriz de regressão linear \mathbf{X} , (3.5) tem-se

$$\begin{aligned} \hat{\mathbf{h}} &= ((\mathbf{QR})^\dagger (\mathbf{QR}))^{-1} (\mathbf{QR})^\dagger \mathbf{y} \\ &= (\mathbf{R}^\dagger \mathbf{Q}^\dagger \mathbf{QR})^{-1} \mathbf{R}^\dagger \mathbf{Q}^\dagger \mathbf{y} \\ &= (\mathbf{R}^\dagger \mathbf{R})^{-1} \mathbf{R}^\dagger \mathbf{Q}^\dagger \mathbf{y} \\ &= \mathbf{R}^{-1} (\mathbf{R}^\dagger)^{-1} \mathbf{R}^\dagger \mathbf{Q}^\dagger \mathbf{y} \end{aligned} \quad (3.7)$$

e, finalmente,

$$\hat{\mathbf{h}} = \mathbf{R}^{-1} \mathbf{Q}^\dagger \mathbf{y}. \quad (3.8)$$

Considerando que (2.26) é um sistema super determinado, é possível simplificar (3.8) lançando mão de uma decomposição QR econômica [51]. Desta forma, $\mathbf{A} = \mathbf{Q}'\mathbf{R}'$ em que $\mathbf{Q}' \in \mathbb{C}^{N \times K}$ e $\mathbf{R}' \in \mathbb{C}^{K \times K}$ com as mesmas propriedades anteriores. Ao reescrever (3.8) como

$$\mathbf{R}'\hat{\mathbf{h}} = \mathbf{Q}'^\dagger \mathbf{y}. \quad (3.9)$$

e considerando $\mathbf{Q}'^\dagger \mathbf{y} = \mathbf{b}$ tem-se

$$\begin{bmatrix} r'_{1,1} & r'_{1,2} & \cdots & r'_{1,K} \\ 0 & r'_{2,2} & \cdots & r'_{2,K} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r'_{K,K} \end{bmatrix} \begin{bmatrix} \hat{h}_1 \\ \hat{h}_2 \\ \vdots \\ \hat{h}_K \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_K \end{bmatrix}. \quad (3.10)$$

Nota-se, então, que não é necessário fazer a inversão de \mathbf{R}' . Os coeficientes podem

ser encontrados pelo método de substituição retroativa dado por

$$\left\{ \begin{array}{l} \hat{h}_K = \frac{b_K}{r'_{K,K}} \\ \hat{h}_{K-1} = \frac{b_{K-1} - r'_{K-1,K} \hat{h}_K}{r'_{K-1,K-1}} \\ \vdots \\ \hat{h}_i = \frac{b_i - \sum_{j=i+1}^K r'_{i,j} \hat{h}_j}{r'_{i,i}} \end{array} \right. \quad (3.11)$$

Há também um terceiro método para solução do LS, chamado de decomposição em valores singulares (*Singular Value Decomposition* - SVD). Considerando o sistema de equações a ser resolvido como super determinado, para uma dada matriz $\mathbf{A} \in \mathbb{C}^{N \times K}$, a fatoração SVD econômica é dada por $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\dagger$, em que $\mathbf{U} \in \mathbb{C}^{N \times K}$, $\mathbf{\Sigma} \in \mathbb{C}^{K \times K}$ e $\mathbf{V}^\dagger \in \mathbb{C}^{K \times K}$. As matrizes \mathbf{U} e \mathbf{V}^\dagger são unitárias e $\mathbf{\Sigma}$ é uma diagonal contendo os valores singulares de \mathbf{A} . Ao decompor \mathbf{X} , (3.4) pode ser reescrita como

$$\begin{aligned} \mathbf{V}\mathbf{\Sigma}^\dagger\mathbf{U}^\dagger\mathbf{U}\mathbf{\Sigma}\mathbf{V}^\dagger\hat{\mathbf{h}} &= \mathbf{V}\mathbf{\Sigma}^\dagger\mathbf{U}^\dagger\mathbf{y} \\ \mathbf{V}\mathbf{\Sigma}^\dagger\mathbf{\Sigma}\mathbf{V}^\dagger\hat{\mathbf{h}} &= \mathbf{V}\mathbf{\Sigma}^\dagger\mathbf{U}^\dagger\mathbf{y} \\ \mathbf{\Sigma}\mathbf{V}^\dagger\hat{\mathbf{h}} &= \mathbf{U}^\dagger\mathbf{y}. \end{aligned} \quad (3.12)$$

Substituindo $\mathbf{V}^\dagger\hat{\mathbf{h}}$ por \mathbf{a} e $\mathbf{U}^\dagger\mathbf{y}$ por \mathbf{b} , tem-se um sistema do tipo

$$\begin{bmatrix} \lambda_{1,1} & 0 & \cdots & 0 \\ 0 & \lambda_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_{K,K} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_K \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_K \end{bmatrix}. \quad (3.13)$$

Sendo $\mathbf{\Sigma}$ uma matriz diagonal, os valores de \mathbf{a} são obtidos diretamente sem ser necessário o processo de substituição retroativa

$$\hat{\mathbf{h}} = \mathbf{V}\mathbf{a}. \quad (3.14)$$

De maneira alternativa, é possível trabalhar sobre (3.12), obtendo

$$\begin{aligned} \mathbf{\Sigma}\mathbf{V}^\dagger\hat{\mathbf{h}} &= \mathbf{U}^\dagger\mathbf{y} \\ \hat{\mathbf{h}} &= \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^\dagger\mathbf{y}, \end{aligned} \quad (3.15)$$

que é uma forma menos complexa de se obter a solução, pois $\mathbf{\Sigma}$ é uma matriz diagonal.

Há um parâmetro ligado à estabilidade numérica do processo de inversão matricial, chamado número de condicionamento. Em um sistema de equação lineares, como o de

(2.26), o número de condicionamento² da matriz \mathbf{X} é definido como

$$\kappa(\mathbf{X}) = \frac{\lambda_{max}}{\lambda_{min}}, \quad (3.16)$$

em que λ_{max} e λ_{min} correspondem ao máximo e mínimo valor singular de \mathbf{X} .

Os valores singulares, por sua vez, são definidos como a raiz quadrada dos *autovalores* de $\mathbf{X}^\dagger \mathbf{X}$. Deve-se interpretar $\kappa(\mathbf{X})$ como uma constante que dá a proporção de quão sensível a solução \mathbf{h} será em função de uma variação em \mathbf{y} e do erro de arredondamento do processo de cálculo. Desta forma, um melhor resultado será obtido quanto menor for o valor de $\kappa(\mathbf{X})$. Em outras palavras, matriz $\mathbf{X}^\dagger \mathbf{X}$ é pior condicionada quanto maior o número de condicionamento $\kappa(\mathbf{X})$ multiplicado pelo erro de arredondamento relativo da máquina de cálculo. A solução LS com decomposição SVD é particularmente útil para esses casos.

Os modelos de sistemas não-lineares especializados apresentados na Seção 2.3.3 têm como objetivo diminuir o número de condicionamento da matriz \mathbf{X} . A ideia baseia-se em gerar uma matriz de regressão linear \mathbf{X} ortogonal, dada a distribuição do sinal de entrada \mathbf{x} . O máximo e mínimo autovalores de uma matriz identidade \mathbf{I} , por sua vez, são iguais a 1. Isso gera o mínimo número de condicionamento possível igual a 1.

A Tabela 4 mostra a diferença no número de condicionamento para três tipos de modelos, para diferentes distribuições. Para obtenção desta tabela, foram gerados sinais complexos com distribuição Gaussiana e uniforme com 65536 amostras. Para cada tipo de sinal, a matriz de regressão linear foi montada conforme as equações de três modelos distintos usando $K = 8$ (15ª ordem). O número de condicionamento foi calculado para cada caso. Finalmente, todo o processo foi repetido 1000 vezes e calculado uma média para cada caso.

Tabela 4 – Média do número de condicionamento da matriz de regressão linear \mathbf{X} em diferentes modelos para diferentes distribuições do sinal de entrada \mathbf{x} .

Modelo	Distribuição de \mathbf{x}	
	Gaussiana	Uniforme
Polinomial banda base conforme (2.20)	$1,03 \times 10^9$	186
Ortogonal Gaussiano conforme (2.27)	259	$4,7 \times 10^{10}$
Ortogonal uniforme conforme (2.29)	$4,3 \times 10^{15}$	115,7

Fica claro que o número de condicionamento cai consideravelmente para o modelo especializado de acordo com sua distribuição correspondente. É importante notar que o

² Na plataforma MATLABTM, o número de condicionamento de uma dada matriz \mathbf{X} é dado pela função `cond(X)`.

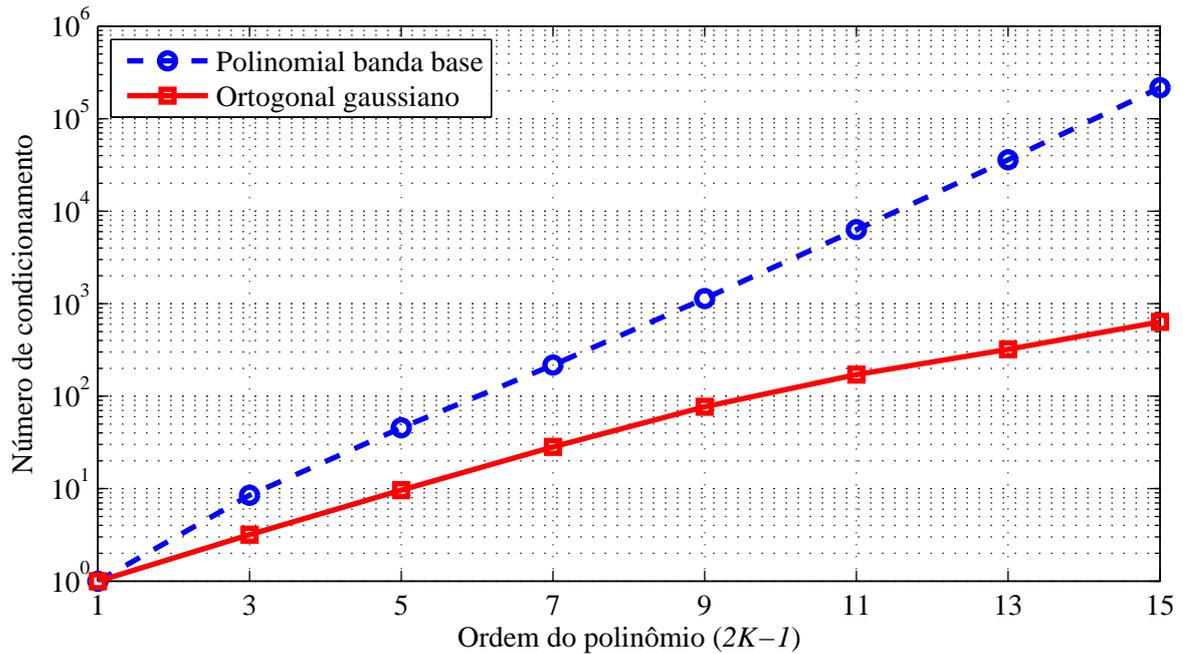


Figura 22 – Número de condicionamento em função da máxima ordem do polinômio para dois modelos distintos: polinomial banda base segundo (2.20) e ortogonal Gaussiano conforme (2.27).

número de condicionamento é dependente da máxima ordem do polinômio. A Figura 22 mostra o número de condicionamento em função da máxima ordem do polinômio para dois modelos distintos.

3.1.2 Métodos de múltiplas iterações

Os algoritmos de otimização apresentados na Seção 3.1.1 empregam uma única iteração, utilizando todas as amostras dos sinais observados para encontrar os coeficientes. Há casos em que é mais interessante obter os coeficientes de maneira recursiva (adaptativa), ou seja, empregando múltiplas iterações. Desta forma, não é necessário uma inversão matricial. Existem inúmeros algoritmos adaptativos na literatura. Neste trabalho serão abordados os principais, destacando suas características, vantagens e desvantagens.

O mais conhecido e simples é o algoritmo de mínimo erro quadrático médio (*Least Mean Squares* - LMS) [52]. Para ilustrar seu uso na identificação de um sistema não-linear, reconsidere (2.26) para um dado instante de tempo discreto n . Logo,

$$y(n) = \vec{\mathbf{x}}(n)\mathbf{h}(n). \quad (3.17)$$

em que $\vec{\mathbf{x}}(n)$ corresponde a uma linha da matriz de regressão linear \mathbf{X} , nesse trabalho

chamado de *vetor de regressão linear*. Para o modelo de (2.20) tem-se

$$\vec{\mathbf{x}}(n) = \begin{bmatrix} x_{1,n} \\ x_{2,n} \\ x_{3,n} \\ \vdots \\ x_{K,n} \end{bmatrix}^T = \begin{bmatrix} x(n) \\ |x(n)|^2 x(n) \\ |x(n)|^4 x(n) \\ \vdots \\ |x(n)|^{2(K-1)} x(n) \end{bmatrix}^T. \quad (3.18)$$

O LMS é uma simplificação do método de descida mais inclinada (*Steepest Descent*) [53] no qual deseja-se minimizar a função custo

$$J(n) = \mathbb{E} [e(n)e^*(n)]. \quad (3.19)$$

O erro entre a saída do sistema observado $y(n)$ e a saída do modelo sendo identificado $\hat{y}(n)$ é dado por

$$e(n) = y(n) - \hat{y}(n) \quad (3.20)$$

sendo

$$\hat{y}(n) = \vec{\mathbf{x}}(n)\hat{\mathbf{h}}^*(n). \quad (3.21)$$

O algoritmo baseia-se em encontrar o gradiente da função custo dado por

$$\begin{aligned} \nabla_{\hat{\mathbf{h}}^*} J(n) &= \nabla_{\hat{\mathbf{h}}^*} \mathbb{E} [e(n)e^*(n)] \\ &= 2\mathbb{E} [\nabla_{\hat{\mathbf{h}}^*} [e(n)] e^*(n)], \end{aligned} \quad (3.22)$$

sendo

$$\nabla_{\hat{\mathbf{h}}^*} = \begin{bmatrix} \frac{\partial}{\partial \hat{h}_1} & \frac{\partial}{\partial \hat{h}_2} & \cdots & \frac{\partial}{\partial \hat{h}_K} \end{bmatrix}. \quad (3.23)$$

Como

$$\begin{aligned} \nabla_{\hat{\mathbf{h}}^*} [e(n)] &= \nabla_{\hat{\mathbf{h}}^*} [y(n) - \vec{\mathbf{x}}(n)\hat{\mathbf{h}}^*(n)] \\ &= -\vec{\mathbf{x}}(n), \end{aligned} \quad (3.24)$$

o gradiente da função custo é

$$\nabla_{\hat{\mathbf{h}}^*} J(n) = -2\mathbb{E} [\vec{\mathbf{x}}(n)e^*(n)]. \quad (3.25)$$

Como o gradiente $\nabla_{\hat{\mathbf{h}}^*} J(n)$ é um vetor que aponta na direção de maior variação da função custo, os coeficientes devem ser atualizados na direção contrária

$$\begin{aligned} \hat{\mathbf{h}}(n+1) &= \hat{\mathbf{h}}(n) - \frac{\mu}{2} \nabla_{\hat{\mathbf{h}}^*} J(n) \\ &= \hat{\mathbf{h}}(n) + \mu \mathbb{E} [\vec{\mathbf{x}}(n)e^*(n)] \end{aligned} \quad (3.26)$$

em que μ é uma contante de tamanho do passo. O LMS é uma implementação estocástica do método *Steepest Descent*. Sendo assim, o valor esperado $\mathbb{E} [\vec{\mathbf{x}}(n)e^*(n)]$ é grosseiramente substituído pelo seu valor instantâneo resultando em

$$\hat{\mathbf{h}}(n+1) = \hat{\mathbf{h}}(n) + \mu \vec{\mathbf{x}}(n)e^*(n), \quad (3.27)$$

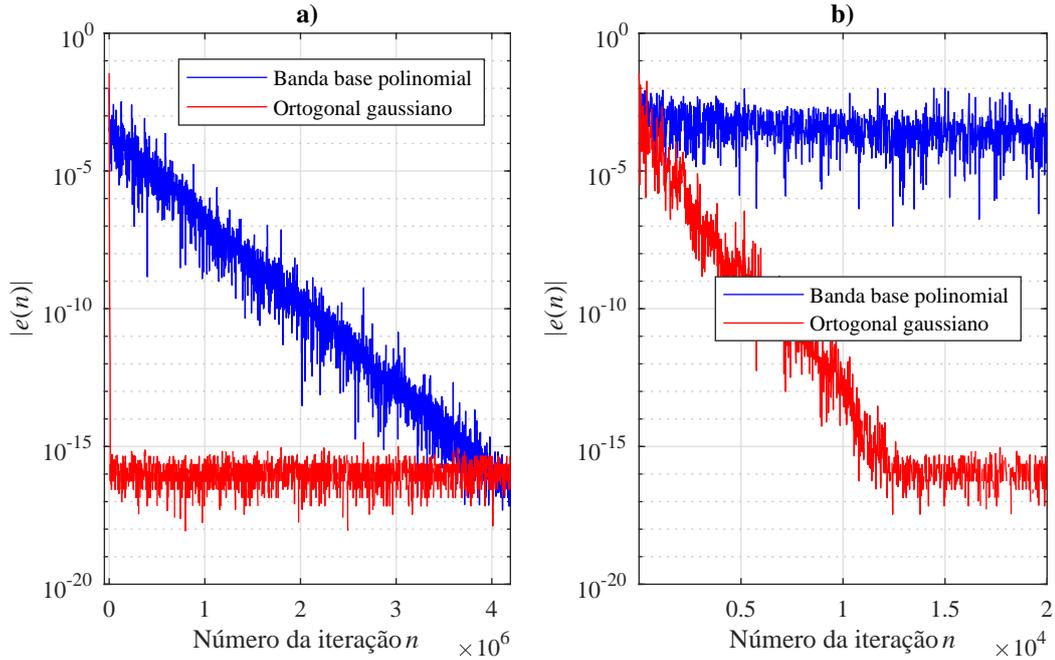


Figura 23 – Convergência do algoritmo LMS para duas equações de modelo diferentes. a) Após $4,1 \times 10^6$ de iterações para evidenciar a convergência lenta do modelo banda base polinomial de (2.20). b) Após 13000 iterações para evidenciar a convergência rápida do modelo ortogonal Gaussiano segundo (2.27).

em que os coeficientes a serem aplicados são $\hat{\mathbf{h}}^*$. Em [54], é mostrado que pode-se obter uma taxa de aprendizagem ótima fazendo

$$\mu = \frac{1}{\bar{\mathbf{x}}(n)\bar{\mathbf{x}}^\dagger(n)}. \quad (3.28)$$

O resumo desse algoritmo é descrito no **Algoritmo 1**.

Algoritmo 1 Lógica recursiva LMS

Inicialize $\hat{\mathbf{h}}(0) = [1, 0, \dots, 0]^T$

para $n \leftarrow 0$ **até** $N - 1$ **faça**

 Obtenha o vetor de regressão linear $\bar{\mathbf{x}}(n)$ de acordo com a equação do modelo

 Obtenha a saída do sistema identificado por meio de (3.21)

 Calcule o erro por meio de (3.20)

 Adapte os coeficientes $\hat{\mathbf{h}}(n + 1)$ de acordo com (3.28)

fim para

Os coeficientes do sistema identificado são dados por $\hat{\mathbf{h}}^*(N)$

O número de condicionamento para os métodos de múltiplas iterações, tem uma influência direta na velocidade de convergência. Para um sinal com distribuição Gaussiana, foram comparados dois modelos distintos: banda base polinomial e ortogonal Gaussiano. Ambos foram identificados usando o algoritmo LMS. A Figura 23 mostra a magnitude do sinal de erro $e(n)$ em função do número da iteração n para esses dois casos. O modelo

banda base polinomial atingiu uma magnitude de erro da ordem de 10^{-15} com aproximadamente $4,1 \times 10^6$ de iterações, enquanto que o modelo ortogonal Gaussiano atingiu a mesma ordem de magnitude com apenas 13000 iterações. Fica claro a influência do número de condicionamento na velocidade de convergência.

Métodos baseados em gradiente possuem uma taxa de convergência linear e lenta. Uma convergência mais rápida pode ser atingida modificando o critério de otimização para determinar a direção de busca. O método conhecido como RPEM [55] permite reduzir o tempo de convergência empregando uma função custo definida como

$$J(\hat{\mathbf{h}}) = \frac{1}{2N} \sum_{n=0}^{N-1} [(y(n) - \hat{y}(n|n-1))^* (y(n) - \hat{y}(n|n-1))]. \quad (3.29)$$

A saída do sistema a ser identificado é dada por

$$\hat{y}(n) = \vec{\mathbf{x}}(n)\hat{\mathbf{h}}(n). \quad (3.30)$$

O erro entre a saída do sistema observado $y(n)$ e a saída do modelo sendo identificado $\hat{y}(n)$ é dado por (3.20). O RPEM requer o gradiente negativo de $e(n)$ em relação aos parâmetros $\hat{\mathbf{h}}(n)$ dado por

$$\begin{aligned} \psi(n) &= -\nabla_{\hat{\mathbf{h}}} e(n) \\ &= \nabla_{\hat{\mathbf{h}}} \hat{y}(n) \\ &= \vec{\mathbf{x}}(n). \end{aligned} \quad (3.31)$$

A simplificação desse gradiente se dá pelo fato do modelo ser linear nos parâmetros. A obtenção dos coeficientes usando o método RPEM é atingido através de

$$\begin{aligned} e(n) &= y(n) - \hat{y}(n) \\ \lambda(n) &= \lambda_0 \lambda(n-1) + 1 - \lambda_0 \\ S(n) &= \vec{\mathbf{x}}(n)\mathbf{P}(n-1)\vec{\mathbf{x}}^T(n) + \lambda(n) \\ \mathbf{P}(n) &= \frac{\mathbf{P}(n-1) - \mathbf{P}(n-1)\vec{\mathbf{x}}^T(n)S(n)^{-1}\vec{\mathbf{x}}(n)\mathbf{P}(n-1)}{\lambda(n)} \\ \hat{\mathbf{h}}(n) &= \hat{\mathbf{h}}(n-1) + \mathbf{P}(n-1)\vec{\mathbf{x}}^T(n)e(n). \end{aligned} \quad (3.32)$$

O fator de esquecimento $\lambda(n)$ cresce exponencialmente para 1, em que a taxa λ_0 e o valor $\lambda(0)$ são ajustáveis. O valor inicial dado para a matriz $\mathbf{P}(n)$ é normalmente $\mathbf{P}(0) = \rho\mathbf{I}$, em que ρ é uma constante que reflete a confiança no valor inicial dos coeficientes $\hat{\mathbf{h}}(0)$.

A Figura 24 compara os dois algoritmos apresentados para modelo ortogonal Gaussiano. Os parâmetros iniciais foram $\lambda_0 = 0,9999$, $\lambda(0) = 0,9$ e $\rho = 10^4$. Assim como no LMS, não foi necessária uma inversão de matriz, dado que o termo $S^{-1}(n)$ de (3.32) é um escalar. Contudo, o RPEM é significativamente mais rápido, uma vez que levou apenas 200 iterações para convergir. O resumo desse algoritmo é mostrado no **Algoritmo 2**.

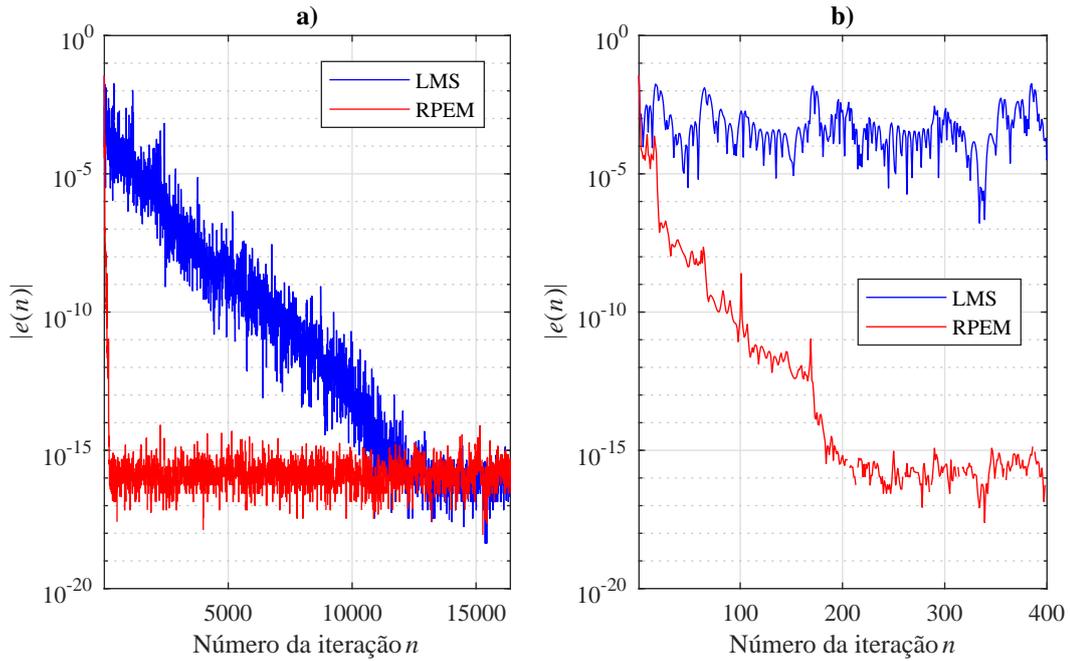


Figura 24 – Comparação de convergência entre os algoritmos LMS e RPEM. a) Após 16384 iterações para evidenciar a convergência lenta do LMS. b) Após 400 iterações para evidenciar a convergência rápida do algoritmo RPEM.

Algoritmo 2 Lógica recursiva RPEM

Inicialize $\hat{\mathbf{h}}(0) = [1, 0, \dots, 0]^T$

Inicialize λ_0

Inicialize $\lambda(0)$

Inicialize ρ

$\mathbf{P}(0) = \rho \mathbf{I}$

para $n \leftarrow 1$ **até** $N - 1$ **faça**

Adapte os coeficientes conforme (3.32)

fim para

Os coeficientes do sistema identificado são dados por $\hat{\mathbf{h}}(N - 1)$

Concluindo, a combinação do RPEM com um modelo de polinômios ortogonais, correspondendo a uma dada distribuição, resulta em uma convergência mais rápida.

3.2 Pré-distorção

A pré-distorção é usada para compensar as distorções não lineares provocadas pelo amplificador de RF. A ideia é aplicar uma distorção no sinal de entrada, de forma que ao passar pelo amplificador, as distorções não-lineares introduzidas por este dispositivo façam com que o sinal de saída seja uma versão linearmente amplificada do sinal de entrada, ou seja, o DPD expande a magnitude do sinal enquanto que o PA comprimir e vice-versa. Semelhantemente, o DPD produz um deslocamento de fase na direção oposta do deslocamento de fase promovido pelo PA. A Figura 25 ilustra o diagrama de blocos de

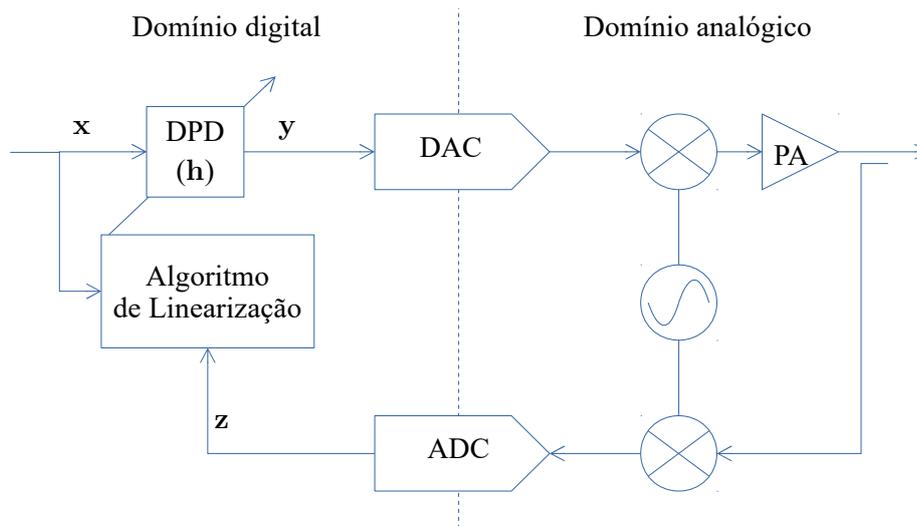


Figura 25 – Diagrama de blocos de um sistema típico de linearização com pré-distorção digital.

um sistema típico com DPD. Nesse diagrama, \mathbf{x} representa o sinal entregue pelo sistema de modulação, independente do padrão que gerou a forma de onda. O sinal modulado se apresenta na forma complexa e pode estar em banda base ou em frequência intermediária (FI). Este sinal passa pelo bloco de DPD gerando a pré-distorção de acordo com os coeficientes \mathbf{h} . O sinal pré-distorcido \mathbf{y} é convertido para o domínio analógico usando um DAC. Haverá então uma conversão para a frequência de operação em RF. O sinal de RF é amplificado por um conjunto de amplificadores associados em série, chamados de estágios de excitação. Nesse ponto, pode-se ter um ou mais amplificadores em paralelo, chamado de estágio de saída. Todos esses amplificadores estão representados por um único bloco PA. É tomada uma amostra do sinal distorcido na saída do PA, normalmente usando um acoplador direcional. A partir daí é feito o processo contrário da transmissão. Usando o mesmo oscilador, o sinal de RF é convertido para banda base ou FI. Este sinal é convertido para o domínio digital usando um ADC. Utilizando o sinal realimentado \mathbf{z} e o sinal \mathbf{x} (sem distorção), o bloco Algoritmo de Linearização utiliza algum método para determinar os coeficientes \mathbf{h} do DPD.

Existem alguns conceitos distintos na forma com que é realizado o processo de aprendizagem que determina os coeficientes do DPD [56]. A Figura 26 ilustra duas formas distintas. O processo de aprendizagem indireta identifica um sistema pós inverso, que tem como sinal de entrada a saída do PA e como sinal de saída a entrada do PA. Assim, quando o sinal distorcido na saída do PA passa pelo sistema pós-inverso, esse produz um sinal sem distorção. Não é possível usar o modelo pós-inverso como um bloco de pós-distorção por questões práticas. Felizmente, o modelo pós-inverso pode ser colocado antes do PA e o resultado será o mesmo, gerando assim uma pré-distorção [12, 57]. Logo, os coeficientes calculados no modelo pós-inverso são copiados para o bloco de DPD.

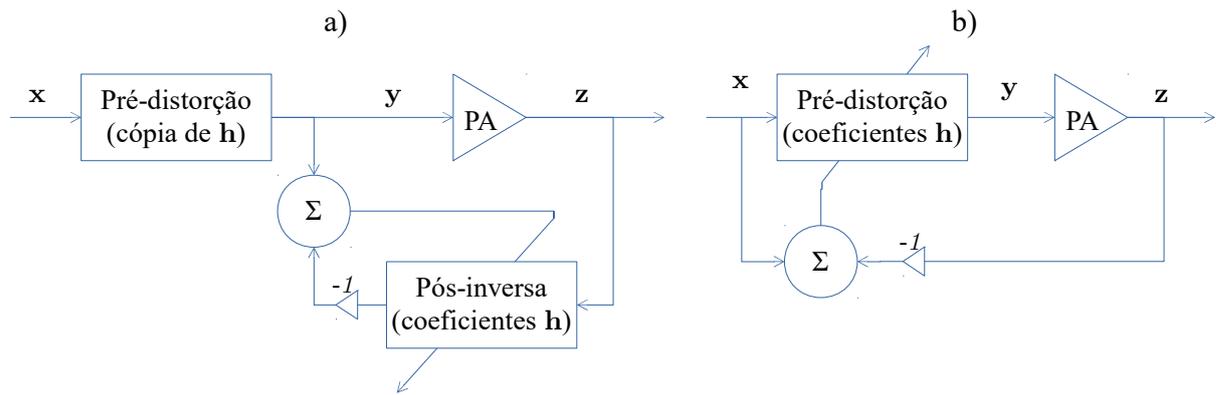


Figura 26 – Arquiteturas de aprendizado. a) Aprendizado indireto e b) Aprendizado direto.

O processo de aprendizado direto é baseado em uma arquitetura de laço fechado que opera em tempo real. Quando o algoritmo de linearização for adaptativo, novos valores de coeficientes são aplicados diretamente no DPD a cada iteração. Assim, não existe um pré-cálculo seguido de cópia de coeficientes, como no aprendizado indireto.

Seja uma simulação de um sistema de pré-distorção que usa a arquitetura de aprendizado indireto. O sistema pós-inverso foi identificado usando o método de mínimos quadrados de acordo com (3.5) e modelo polinomial com memória conforme (2.34). As Figuras 27-30 ilustram os resultados obtidos para melhor explicar o funcionamento da pré-distorção. A Figura 27 mostra a magnitude dos sinais em vários pontos do sistema. Pode-se observar uma sequência de amostras ao redor de um pico de potência, onde acontece a maior distorção provocada pelo amplificador. Os sinais de entrada, saída do DPD e saída do PA correspondem aos vetores \mathbf{x} , \mathbf{y} e \mathbf{z} da Figura 25, respectivamente. Note que a saída do PA sem DPD sofreu compressão nas amplitudes mais altas e expansão nas magnitudes mais baixas, comparado ao sinal de entrada. Naturalmente, também houve distorções de fase, mas que não foram ilustradas. Para compensar estas distorções, a saída do DPD expandiu nas amplitudes mais altas e comprimiu nas amplitudes mais baixas. O sinal na saída do PA com DPD foi corrigido na maioria das amplitudes. Com exceção da região do pico de magnitude, todas as demais partes do sinal foram compensadas. O motivo pelo qual a região de pico não ser corrigida é um problema conhecido. Muitos sistemas podem ser invertidos somente para um intervalo de amplitude de entrada restrito [49]. Este problema será explicado com mais detalhes na Seção 7.1 e é tratado como um trabalho futuro. Mesmo com esta restrição, a correção resulta em redução na potência dos produtos de intermodulação no domínio da frequência, que é o objetivo final de todo este trabalho.

Na Figura 28, a função de transferência instantânea é mostrada entre vários blocos do sistema. Devido ao espalhamento dos pontos, conclui-se que há efeito memória no amplificador. A única forma de combater tal efeito é com uma pré-distorção também

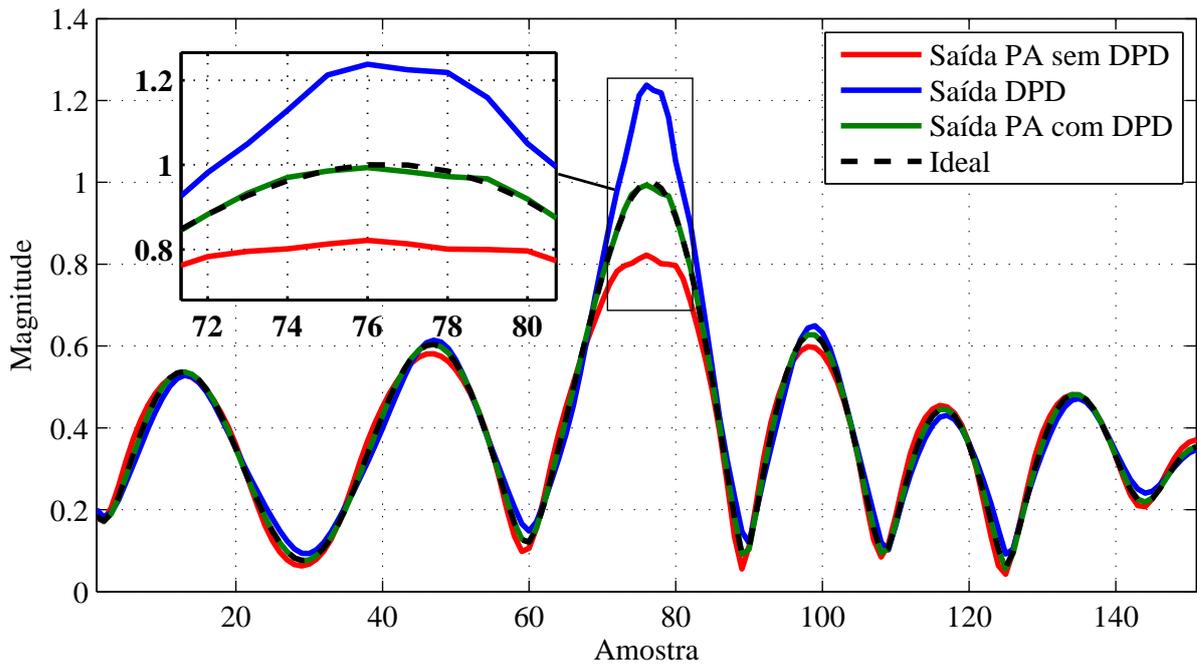


Figura 27 – Magnitude dos sinais em vários pontos do sistema de pré-distorção em função do número da amostra.

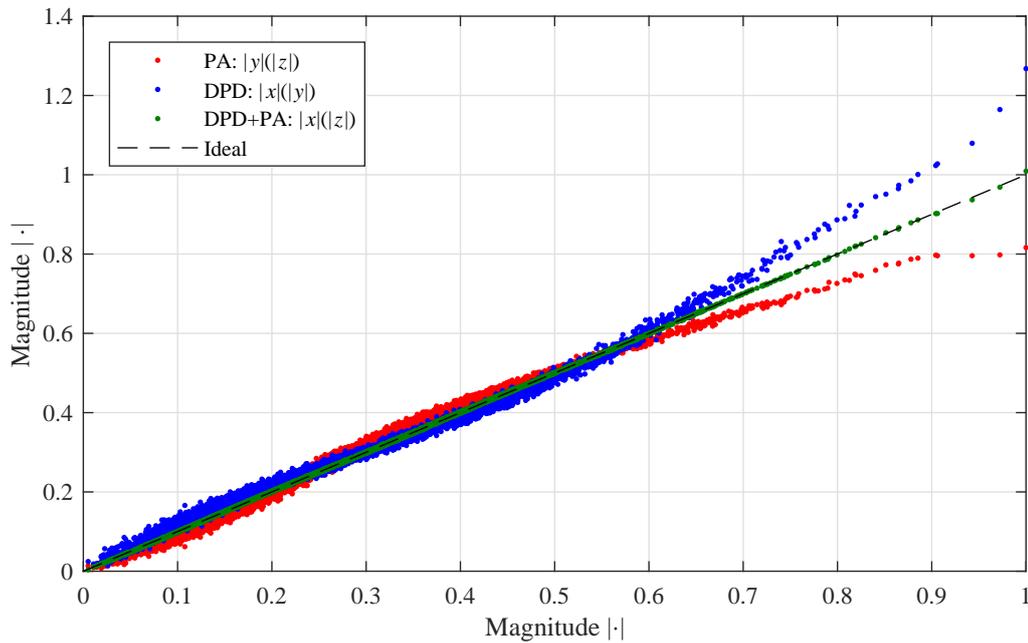


Figura 28 – Comparação entre as funções de transferência instantânea da magnitude: DPD, PA e do conjunto DPD em série com o PA.

com efeito memória. A função de transferência da combinação de DPD e PA diminuiu o espalhamento, compensando o efeito memória, e também se aproximou da linha reta representando um sistema linear. A reduzida quantidade de pontos nas magnitudes mais altas indicam uma menor probabilidade do sinal de entrada apresentar picos de potência.

As funções de transferência AM/AM e AM/PM são apresentadas na Figura 29.

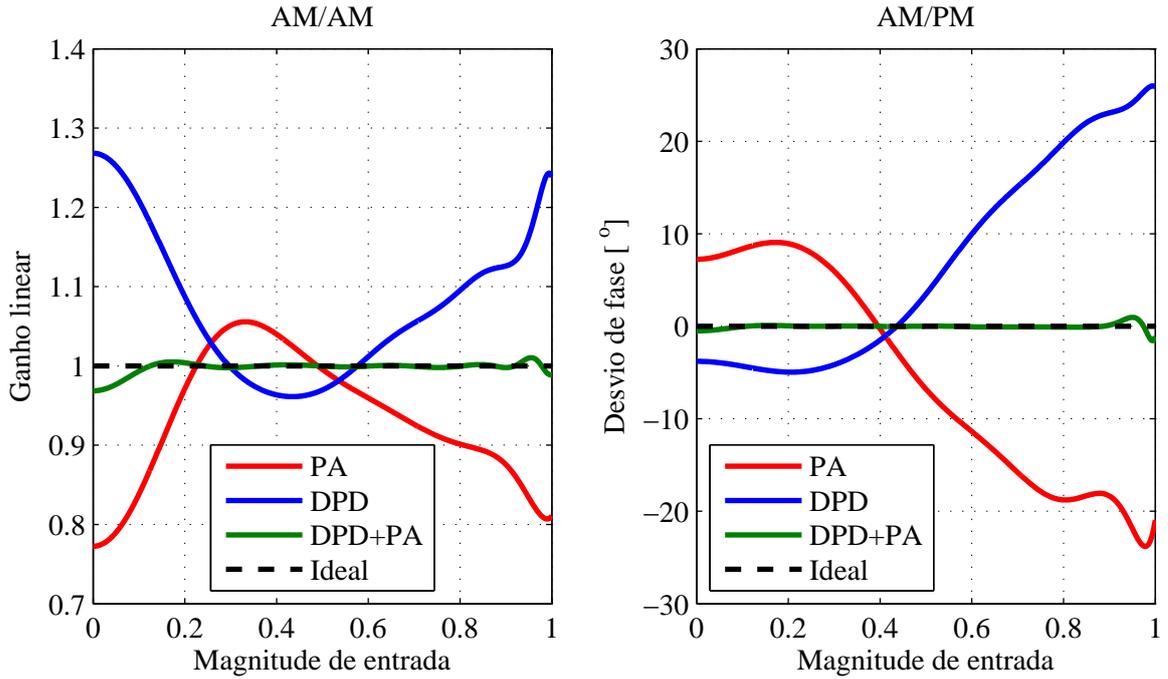


Figura 29 – Funções AM/AM e AM/PM entre alguns pontos do sistema, sendo que DPD+PA corresponde à associação série do bloco do DPD com o amplificador.

Pode-se notar o comportamento de compensação das características não-lineares do PA. Apesar do nome de pós-inversa, a curva AM/AM do DPD não é literalmente a inversão da curva correspondente do PA, ou seja, não se trata de um espelhamento em torno da reta ideal.

Finalmente, os sinais no domínio da frequência são mostrados na Figura 30. Note que a intermodulação do sinal na saída do DPD se assemelha muito à intermodulação observada na saída do PA, quando não havia pré-distorção. Com a presença do DPD, o sinal na saída do PA tem sua intermodulação consideravelmente reduzida. Para demonstrar a redução da intermodulação, serão usados duas figuras de mérito: o NMSE, que mede o erro entre dois sinais no domínio do tempo, e o ACLR que mede a intermodulação no domínio da frequência, definido como

$$\text{ACLR} = C(\mathbf{h}) = \frac{\int_{f \in \text{OOB}} P_{\mathbf{z}}(f) df}{\int_{f \in \text{IB}} P_{\mathbf{z}}(f) df}, \quad (3.33)$$

em que $P_{\mathbf{z}}(f)$ representa a DEP de \mathbf{z} na frequência f . A DEP é medida nos conjuntos de frequências fora da banda (*Out Of Band* - OOB) e dentro da banda (*In Band* - IB) do canal. A Tabela 5 mostra os resultados. Houve uma redução de 28,42 dB no NMSE, mostrando que a forma de onda na saída do PA se tornou mais parecida com a da entrada. A redução de 26,68 dB no ACLR corresponde à queda da potência da intermodulação no domínio da frequência.

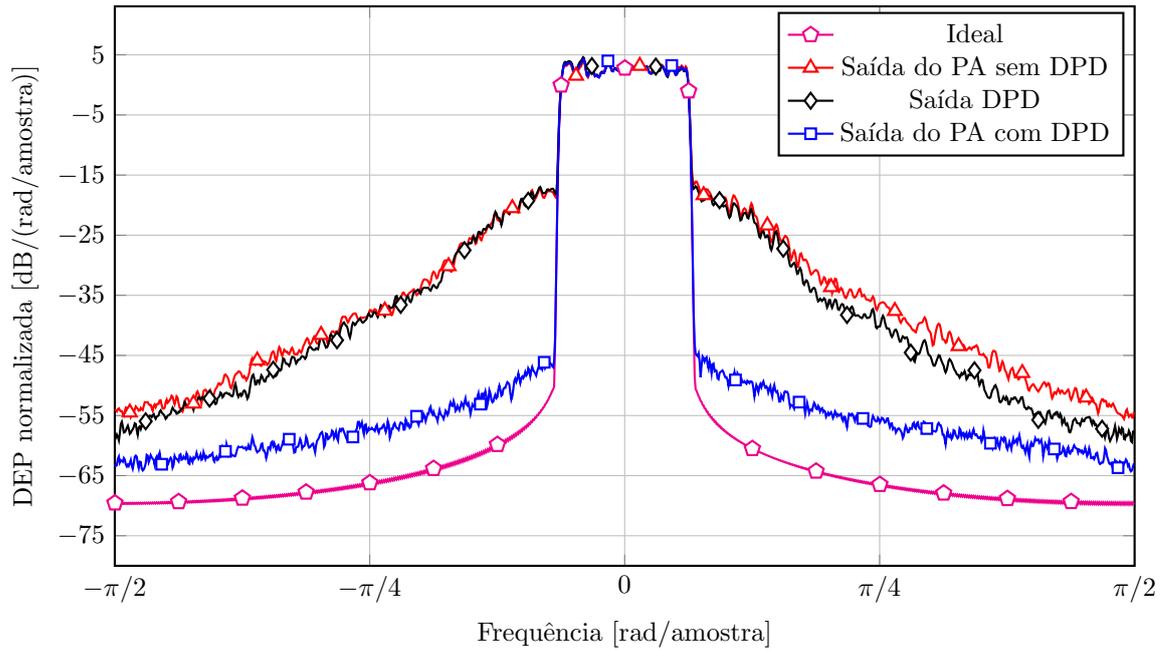


Figura 30 – DEP dos sinais ao longo do sistema de pré-distorção.

Tabela 5 – Medidas de desempenho do DPD.

Figuras de mérito	Sem DPD	Com DPD
NMSE [dB]	-16,07	-44,49
ACLR [dB]	-24,01	-50,69

3.3 Conclusão

Este capítulo apresentou o método de cálculo do modelo do DPD através de aprendizado indireto. Essa técnica identifica o modelo pós-inverso do amplificador, realizado através de métodos de identificação de sistemas usando regressão linear. O capítulo concluiu revisando conceitos básicos de pré-distorção. A identificação usando regressão linear, tem como objetivo a minimização do erro no domínio do tempo, o que indiretamente reduz a intermodulação no domínio da frequência. O algoritmo proposto nesse trabalho, é uma alternativa concorrente aos métodos de DPD usando regressão linear, onde a métrica é tomada diretamente no domínio da frequência como critério de otimização, resultando em um melhor desempenho na redução da intermodulação. De qualquer maneira, todos os conceitos apresentados no presente capítulo têm importância para a compreensão do método proposto no Capítulo 4.

4 Proposta de um sistema de linearização ortogonal com realimentação escalar

O algoritmo proposto neste capítulo, é uma alternativa concorrente aos métodos de DPD usando regressão linear, que minimizam o erro no domínio do tempo, indiretamente reduzindo a intermodulação no domínio da frequência. No algoritmo de realimentação escalar, o critério de otimização é a minimização direta de uma dada métrica no domínio da frequência. Esse conceito será explicado em detalhes a seguir, com foco na proposta do algoritmo de ortogonalização para os sistemas de DPD com realimentação escalar, que é a principal contribuição deste trabalho.

4.1 Sistema de DPD com realimentação escalar

O diagrama de blocos dos sistemas de DPD com realimentação escalar [33–36] é dado pela Figura 31. A ideia básica desse sistema é tomar uma medida custo $C(\mathbf{h})$, em que $C(\mathbf{h}) \in \mathbb{R}$, a partir do sinal de saída do PA, \mathbf{z} , e ajustar os coeficientes do modelo do DPD, \mathbf{h} , para minimizar $C(\mathbf{h})$. O sinal na saída do PA não será diretamente usado pelo algoritmo de adaptação, mas indiretamente por meio da tomada de uma dada medida custo. A cada nova iteração, a função de medida custo gera um novo escalar $C(\mathbf{h})$ proporcional à característica do sinal que deseja-se minimizar.

A medida custo pode ser selecionada de acordo com um dado objetivo:

- atender uma máscara de emissão espectral – medir a intermodulação em frequências

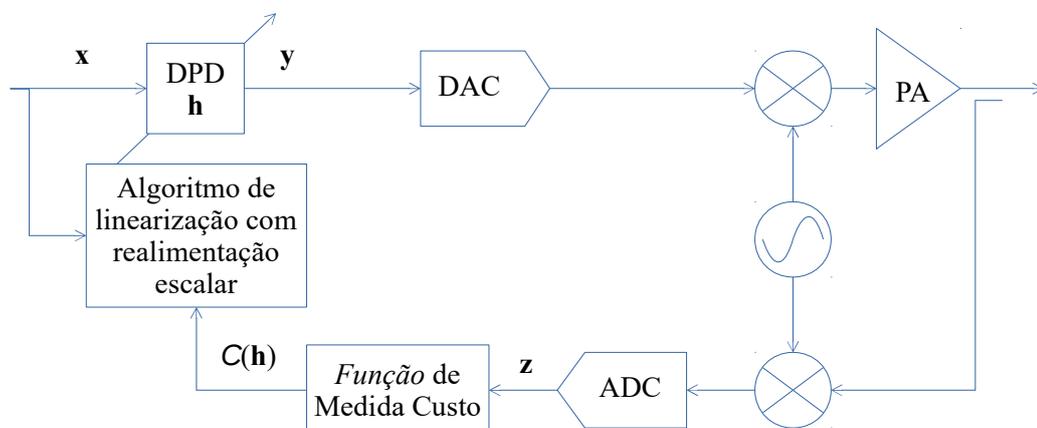


Figura 31 – Diagrama de blocos do sistema proposto de linearização ortogonal com realimentação escalar (OSFL).

específicas e selecionar a pior delas em relação à máscara;

- minimizar a interferência fora da banda nos canais adjacentes – medir o ACLR; ou
- minimizar a interferência dentro da banda – medir a MER.

Na descrição do método e simulações a seguir, a medida custo empregada foi o ACLR.

O processo de linearização é definido como um problema de otimização numérica da função custo $C(\mathbf{h})$, em que deve-se encontrar o valor ótimo das variáveis dependentes \mathbf{h} , denotado por \mathbf{h}_{opt} , de forma que $C(\mathbf{h}_{\text{opt}}) = \text{ACLR}_{\text{min}}$. Otimização numérica é um campo amplo da matemática e há muitos algoritmos disponíveis na literatura [58, 59]. Neste trabalho, os algoritmos podem ser selecionados¹ considerando que a função custo a ser minimizada é convexa, suave e com objetivo único, devido à realimentação escalar. Os algoritmos precisam ser multivariáveis, baseados em gradiente ou livres de gradiente. Qualquer algoritmo de otimização pode ser empregado, desde que seja de natureza iterativa [60].

4.2 Proposta de um sistema de DPD com realimentação escalar ortogonal

O problema de ajustar os coeficientes independentemente baseia-se no fato de que o coeficiente de 1ª ordem (ganho linear) é afetado pelo coeficiente de 3ª ordem. Similarmente os coeficientes de 1ª e 3ª ordem são afetados pelo ajuste do coeficiente de 5ª ordem. Genericamente, um ajuste feito no coeficiente h_i afeta os ajustes previamente feitos nos coeficientes h_1, h_2, \dots, h_{i-1} , i.e., os coeficientes não são ortogonais em termos de produtos de intermodulação. Para demonstrar esse problema, considere um sinal de dois tons passando por um sistema polinomial banda base de (2.20) de 5ª ordem. Arbitrariamente, os coeficientes desse modelo exemplo são: $h_1 = 0$ (1ª ordem), $h_2 = 0$ (3ª ordem) e $h_3 = 0, 1 + j0, 1$ (5ª ordem). A Figura 32 mostra o sinal de saída desse sistema. Note que o coeficiente de 5ª ordem é o único a apresentar valor não nulo, mesmo assim resultou não somente em um produto de intermodulação de 5ª ordem, mas também nos de 3ª e 1ª. Portanto, para manter os benefícios atingidos pelo ajuste dos coeficientes anteriores, para cada novo valor de h_i , os coeficientes h_1, h_2, \dots, h_{i-1} precisam ser recalculados e adaptados para anular o efeito de h_i nas ordens inferiores. A esse processo de adaptação dá-se o nome *ortogonalização dos coeficientes do modelo*.

O algoritmo de ortogonalização proposto necessita que o modelo matemático a ser identificado seja um sistema não-linear de tempo discreto *linear nos parâmetros*. Na Seção 2.3, foram apresentados alguns modelos que atendem essa exigência. Ser linear

¹ O *Optimization Toolbox* e o *Global Optimization Toolbox* do MATLAB fornecem muitos algoritmos.

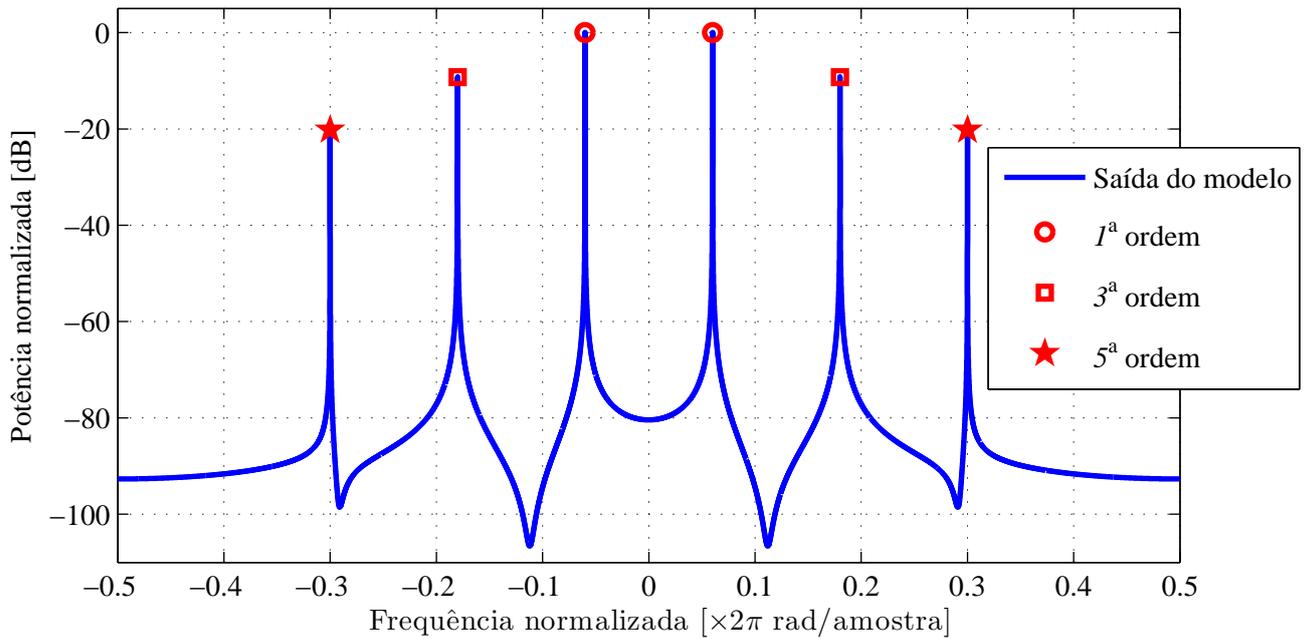


Figura 32 – Saída de um sistema não-linear com dois tons de entrada, em que apenas o coeficiente de 5ª ordem é não-nulo.

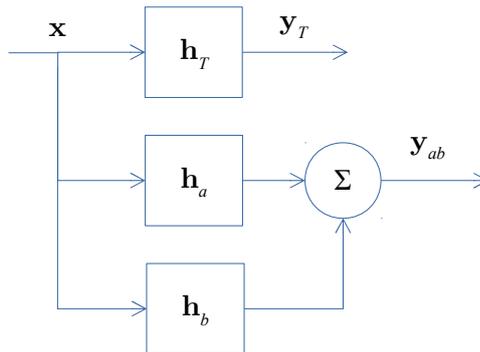


Figura 33 – Diagrama ilustrando o conceito de linearidade nos parâmetros.

nos parâmetros significa poder representar a equação do modelo por meio de uma multiplicação entre uma matriz e um vetor de coeficientes. Tais sistemas não apresentam o princípio da superposição para o sinal de entrada, por serem sistemas não-lineares, mas apresentam o princípio da *superposição nos coeficientes*. Matematicamente, os sistemas da Figura 33 são considerados lineares nos parâmetros quando $\mathbf{y}_T = \mathbf{y}_{ab}$ se $\mathbf{h}_T = \mathbf{h}_a + \mathbf{h}_b$. Os modelos polinomial banda base de (2.20), Gaussiano ortogonal de (2.27), uniforme ortogonal de (2.29), Volterra de (2.30) e polinômio com memória de (2.34) atendem esse requisito. Porém, nem todos sistemas não-lineares são lineares nos parâmetros. O sistema de Wiener [49] de (2.39) é um exemplo.

A ortogonalização promove uma simplificação no problema de otimização de $C(\mathbf{h})$, uma vez que cada coeficiente pode ser ajustado independentemente para otimizar uma

função de duas variáveis dada por

$$\text{ACLR} = C(\Re\{h_i\}, \Im\{h_i\}) \quad (4.1)$$

em que $\Re\{h_i\}$ e $\Im\{h_i\}$ denotam a parte real e imaginária de h_i . O processo de ajuste para minimizar a função é repetida K vezes, em que K é o número de coeficientes do modelo. O algoritmo completo foi subdividido em três partes distintas, como apresentado nas subseções a seguir.

4.2.1 Ortogonalização dos coeficientes

Sejam três conjuntos de coeficientes definidos como

$$\mathbf{h}_L = [h_{L_1}, h_{L_2}, \dots, h_{L_i}, \dots, h_{L_K}]^T \quad (4.2)$$

$$\mathbf{h}_C = [h_{C_1}, h_{C_2}, \dots, h_{C_i}, \dots, h_{C_K}]^T \quad (4.3)$$

$$\mathbf{h}_D = [h_{D_1}, h_{D_2}, \dots, h_{D_i}, \dots, h_{D_K}]^T \quad (4.4)$$

em que $(\cdot)^T$ denota a operação de transposição. Os coeficientes $\mathbf{h}_L \in \mathbb{C}^{K \times 1}$ correspondem ao último ajuste válido, $\mathbf{h}_C \in \mathbb{C}^{K \times 1}$ ao ajuste da ordem atual e $\mathbf{h}_D \in \mathbb{C}^{K \times 1}$ remete aos coeficientes aplicados no DPD. O valor inicial dos coeficientes são atribuídos como $\mathbf{h}_L = \mathbf{h}_C = \mathbf{h}_D = [1, 0, \dots, 0]^T$. Essa inicialização equivale a nenhuma pré-distorção aplicada em que o primeiro coeficiente corresponde ao ganho linear unitário. O ajuste dos coeficientes h_{C_i} será feito na ordem $i = 2, \dots, K$. O coeficiente h_{C_1} nunca será ajustado uma vez que é desejável manter o ganho de potência unitário no sistema de DPD. Considerando que $\mathbf{x} = [x_0, x_1, \dots, x_{N-1}]^T$, $\mathbf{x} \in \mathbb{C}^{N \times 1}$ seja o sinal de entrada, sendo N o número de amostras, extenso o suficiente para reproduzir as estatísticas do sinal. Para cada iteração na qual o valor do coeficiente h_{C_i} está sendo ajustado, os coeficientes $h_{D_1}, h_{D_2}, \dots, h_{D_{i-1}}$ precisam ser atualizados. O vetor diferença (ou residual) entre a saída dos sistemas com coeficientes \mathbf{h}_L e \mathbf{h}_C , ilustrado na Figura 34, é definido como

$$\mathbf{r} = \mathbf{X}\mathbf{h}_L - \mathbf{X}\mathbf{h}_C \quad (4.5)$$

com $\mathbf{r} \in \mathbb{C}^{N \times 1}$, em que a matriz de regressão linear \mathbf{X} é definida como

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_i, \dots, \mathbf{x}_K] \quad (4.6)$$

em que $\mathbf{x}_i \in \mathbb{C}^{N \times 1}$ se refere a i -ésima coluna de $\mathbf{X} \in \mathbb{C}^{N \times K}$ definida de acordo com o modelo de sistema não-linear de tempo discreto adotado. Por uma questão de simplicidade, o polinômio banda base dado por (2.20) foi escolhido. Logo, \mathbf{x}_i será dado por (2.25).

A equação (4.5) pode ser detalhada como

$$\mathbf{r} = \begin{bmatrix} x_{1,0}(h_{L_1} - h_{C_1}) + \dots + x_{i,0}(h_{L_i} - h_{C_i}) + \dots + x_{K,0}(h_{L_K} - h_{C_K}) \\ x_{1,1}(h_{L_1} - h_{C_1}) + \dots + x_{i,1}(h_{L_i} - h_{C_i}) + \dots + x_{K,1}(h_{L_K} - h_{C_K}) \\ \vdots \\ x_{1,N-1}(h_{L_1} - h_{C_1}) + \dots + x_{i,N-1}(h_{L_i} - h_{C_i}) + \dots + x_{K,N-1}(h_{L_K} - h_{C_K}) \end{bmatrix} \quad (4.7)$$

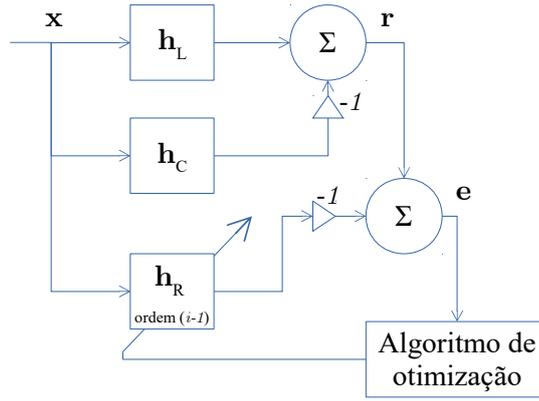


Figura 34 – Cancelando o efeito do ajuste do coeficiente de ordem i nas ordens anteriores. Os coeficientes \mathbf{h}_R são limitados ao coeficiente $(i - 1)$.

e por definição do algoritmo OSFL, $h_{L_k} = h_{C_k}$ para $k \neq i$, ou seja, quando o i -ésimo coeficiente está sendo ajustado, todos os outros coeficientes \mathbf{h}_L e \mathbf{h}_C são iguais. Consequentemente, (4.7) pode ser simplificado para

$$\mathbf{r} = \mathbf{x}_i (h_{L_i} - h_{C_i}). \quad (4.8)$$

Um sistema com entrada \mathbf{x} e coeficientes \mathbf{h}_R , $\mathbf{h}_R \in \mathbb{C}^{(i-1) \times 1}$, necessita ser identificado para produzir uma saída tão similar quanto possível do vetor residual \mathbf{r} dado por (4.8). Como apenas os coeficientes de ordem inferior a do que está sendo ajustado necessitam ser atualizados, o sistema residual identificado é limitado ao coeficiente de índice $(i - 1)$. Este processo é mostrado no diagrama de blocos da Figura 34. O sistema residual pode ser modelado usando o método dos mínimos quadrados como a seguir

$$J(\mathbf{h}_R) = (\mathbf{r} - \mathbf{R}\mathbf{h}_R)^\dagger (\mathbf{r} - \mathbf{R}\mathbf{h}_R) \quad (4.9)$$

$$\hat{\mathbf{h}}_R = \arg \min_{\mathbf{h}_R} J(\mathbf{h}_R), \quad (4.10)$$

em que

$$\mathbf{R} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{i-1}] \quad (4.11)$$

é a matriz de regressão com ordem polinomial limitado a $i - 1$. A função custo $J(\mathbf{h}_R)$ retorna o erro quadrático que é um escalar a ser minimizado no sentido dos mínimos quadrados. A solução de (4.10) é dada por

$$\hat{\mathbf{h}}_R = (\mathbf{R}^\dagger \mathbf{R})^{-1} \mathbf{R}^\dagger \mathbf{x}_i (h_{L_i} - h_{C_i}). \quad (4.12)$$

Finalmente, os coeficientes \mathbf{h}_D a serem aplicados no sistema DPD são obtidos pela combinação linear

$$h_{D_k} = \begin{cases} h_{L_k} + \hat{h}_{R_k}, & \text{para } k = 1, \dots, i - 1 \\ h_{C_k}, & \text{para } k = i \\ h_{L_k}, & \text{para } k = i + 1, \dots, K \end{cases}. \quad (4.13)$$

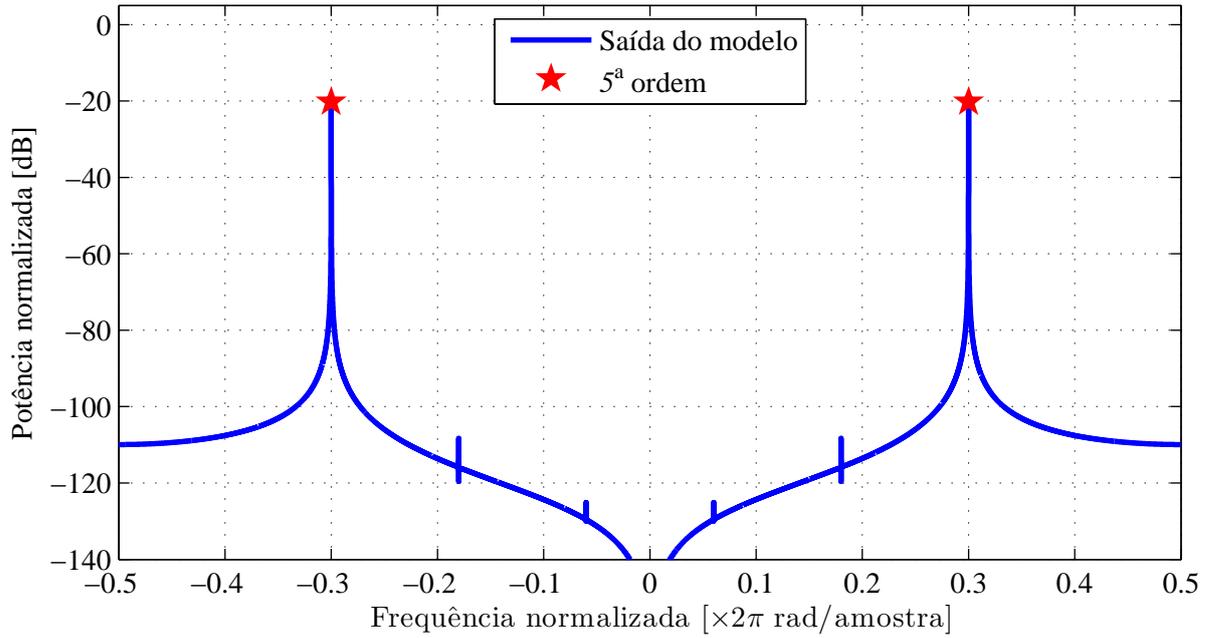


Figura 35 – Saída de um sistema não-linear com dois tons de entrada e coeficientes ortogonalizados. Apenas os produtos de 5ª ordem estão presentes nesta figura devido a um ajuste no coeficiente de mesma ordem.

Devido à limitação da ordem da matriz de regressão \mathbf{R} , quando o coeficiente h_{C_i} está sendo ajustado, apenas os coeficientes h_{D_k} com $k = 1, 2, \dots, (i - 1)$ são atualizados pelo processo de ortogonalização. O conjunto de coeficientes \mathbf{h}_D é aplicado no DPD e fornece a independência necessária entre as ordens polinomiais, ou seja, cada coeficiente gera produtos de intermodulação somente na sua ordem.

Dando continuidade no exemplo da Figura 32, a lógica de ortogonalização dos coeficientes será aplicada como a seguir. Considerando $\mathbf{h}_L = [0, 0, 0]^T$ e $\mathbf{h}_C = [0, 0, (0, 1 + j0, 1)]^T$, os coeficientes h_{D_1} e h_{D_2} serão atualizados devido ao ajuste em h_{C_3} . Os valores que promovem a ortogonalização entre as ordens são $\mathbf{h}_D = [(0, 0313 + j0, 0313), (-0, 1250 - j0, 1250), (0, 1 + j0, 1)]^T$. Os valores atualizados dependem da distribuição do sinal de entrada e do modelo escolhido. Quando o mesmo sinal de dois tons de entrada passa pelo sistema com coeficientes \mathbf{h}_D , um sinal de saída é produzido conforme ilustrado na Figura 35. Note que agora existem apenas os produtos de intermodulação de 5ª ordem, como consequência do ajuste no coeficiente de 5ª ordem. As intermodulações de 1ª e 3ª ordem foram canceladas devido ao processo de ortogonalização.

Este processo pode ser mais simplificado reescrevendo (4.12) como

$$\hat{\mathbf{h}}_R = \mathbf{\Omega}_i (h_{L_i} - h_{C_i}) \quad (4.14)$$

em que

$$\mathbf{\Omega}_i = (\mathbf{R}^\dagger \mathbf{R})^{-1} \mathbf{R}^\dagger \mathbf{x}_i. \quad (4.15)$$

Cada vetor Ω_i possui $i - 1$ elementos que dependem apenas do sinal de entrada e das equações do modelo adotado. Portanto, não depende dos valores dos coeficientes que estão sendo ajustados. Isso significa que se a distribuição do sinal de entrada \mathbf{x} é invariante no tempo, o conjunto de vetores Ω_i são constantes e precisam ser calculados uma única vez durante a inicialização. O valor de $\hat{\mathbf{h}}_R$ será resultado de uma proporção linear entre o vetor Ω_i e o escalar $(h_{L_i} - h_{C_i})$. Assim, a complexidade de implementação desse algoritmo é reduzida significativamente uma vez que o processo de inversão matricial para calcular Ω_i precisa ser feito apenas K vezes (uma para cada ordem), independente do número de iterações.

4.2.2 Minimizando a medida custo para um dado coeficiente

Nesta etapa, o objetivo é encontrar o valor ótimo do coeficiente sendo ajustado para minimizar a medida custo. Como descrito na Subseção 4.2.1, para cada valor de h_{C_i} , o conjunto de coeficientes \mathbf{h}_D é atualizado e aplicado no bloco do DPD. Então, uma nova medida custo é tomada na saída do PA. Um novo valor para h_{C_i} pode ser escolhido para a próxima iteração e o processo se repete. Note que o valor de \mathbf{h}_L permanece constante e somente será atualizado quando a medida custo for minimizada com sucesso para o coeficiente atual h_{C_i} . Para que o DPD seja capaz de produzir distorções AM/PM, os coeficientes do modelo são números complexos em que a parte real e imaginária devem ser ajustadas independentemente. Isso significa que a minimização da função custo para um dado coeficiente é um problema de otimização de uma função de duas variáveis. Neste ponto, uma variação de sub-algoritmos pode ser aplicada.

Com relação aos algoritmos de otimização numérica, os algoritmos *Nelder-Mead Simplex*² [61] e *Quasi-Newton*³ [62] foram investigados. O *Nelder-Mead Simplex* é um método multidimensional livre de gradiente. É considerado um algoritmo de busca direta devido ao seu comportamento eurístico. Uma geometria simplex inicial é calculada ao avaliar a função para valores arbitrários das variáveis independentes. Para cada iteração, a geometria simplex é atualizada baseada em contração, expansão, reflexão ou encolhimento dependendo dos valores dos vértices. O método *Quasi-Newton* baseado em gradiente usa um procedimento de busca linear, lançando mão a fórmula *Broyden-Fletcher-Goldfarb-Shanno* para atualizar uma aproximação da matriz Hessiana. Este se opõe ao método de Newton no sentido de que o valor exato da matriz Hessiana ou Jacobiana não está disponível para dar um passo no mínimo local. Uma comparação dos dois algoritmos será apresentada na Seção 4.3.

Outra alternativa para o problema da otimização é aplicar a minimização alternada [63]. Seja $h_{C_i} = p + jq = (p, q)$, em que p e q são as partes real e imaginária de h_{C_i}

² *Nelder-Mead Simplex* está disponível no MATLAB como a função `fminsearch`.

³ O método *Quasi-Newton* está disponível no MATLAB na função `fminunc`.

com $p \in P$ e $q \in Q$. Seja $C : P \times Q \rightarrow \mathbb{R}$ a função custo. O objetivo é minimizar C sobre $P \times Q$, portanto

$$\min_{(p,q) \in P \times Q} C(p, q). \quad (4.16)$$

Ajustar as variáveis p e q simultaneamente não é trivial, mas uma pode ser ajustada enquanto a outra é mantida fixa. Nesse caso, o algoritmo de minimização alternada começa com a iteração $\ell = 1$ e um valor arbitrário q_0 , sendo que

$$\begin{aligned} p_\ell &\in \arg \min_{p \in P} C(p, q_{\ell-1}) \\ q_\ell &\in \arg \min_{q \in Q} C(p_\ell, q). \end{aligned} \quad (4.17)$$

Desta maneira apenas uma variável é ajustada por vez, caso o algoritmo de otimização numérica não seja multivariável. Se o algoritmo convergir,

$$\lim_{\ell \rightarrow \infty} C(p_\ell, q_\ell) = \min_{(p,q) \in P \times Q} C(p, q). \quad (4.18)$$

A Figura 36(a) mostra a função custo do ACLR, com aparência suave, uma vez que a sequência de entrada é periódica. A Figura 36(b) contém a trajetória da minimização da função custo e sua projeção em um gráfico de contorno para o coeficiente de 3ª ordem, usando ou não a minimização alternada para o algoritmo de otimização *Nelder-Mead Simplex*. Verifica-se que a trajetória da minimização alternada apresenta apenas conversões de ângulo reto, uma vez que uma das partes real ou imaginária é fixada enquanto a outra é minimizada.

4.2.3 Lógica do laço exterior

O último ajuste de h_{C_i} gerou um conjunto de coeficientes \mathbf{h}_D . Uma vez que a medida custo foi minimizada para um dado coeficiente com índice i , \mathbf{h}_L e \mathbf{h}_C necessitam ser atualizados fazendo $\mathbf{h}_L = \mathbf{h}_D$ e $\mathbf{h}_C = \mathbf{h}_D$. O processo é então repetido para o próximo coeficiente h_{C_i} , em que $i \leftarrow i + 1$. Então se repete para todos os demais coeficientes até que $i = K$. Para uma melhor precisão, todo o processo pode ser repetido fazendo $i = 2$, mas no próximo ciclo, os coeficientes \mathbf{h}_L não são inicializados. O coeficiente h_{C_1} de ordem 1 não é ajustado para manter o ganho linear unitário. Contudo, o coeficiente de 1ª ordem é adaptado pelo algoritmo de ortogonalização. Em outras palavras, o algoritmo de otimização numérica não ajusta h_{C_1} , mas o algoritmo de ortogonalização adapta h_{D_1} para remover o efeito dos coeficientes de mais alta ordem no de 1ª ordem.

4.2.4 Resumo do algoritmo

O algoritmo OSFL pode ser resumido conforme descrito em **Algoritmo 3**. A declaração ‘Gerar novo h_{C_i} para minimizar $C(\mathbf{h}_D)$ ’ é feita de acordo com o algoritmo de

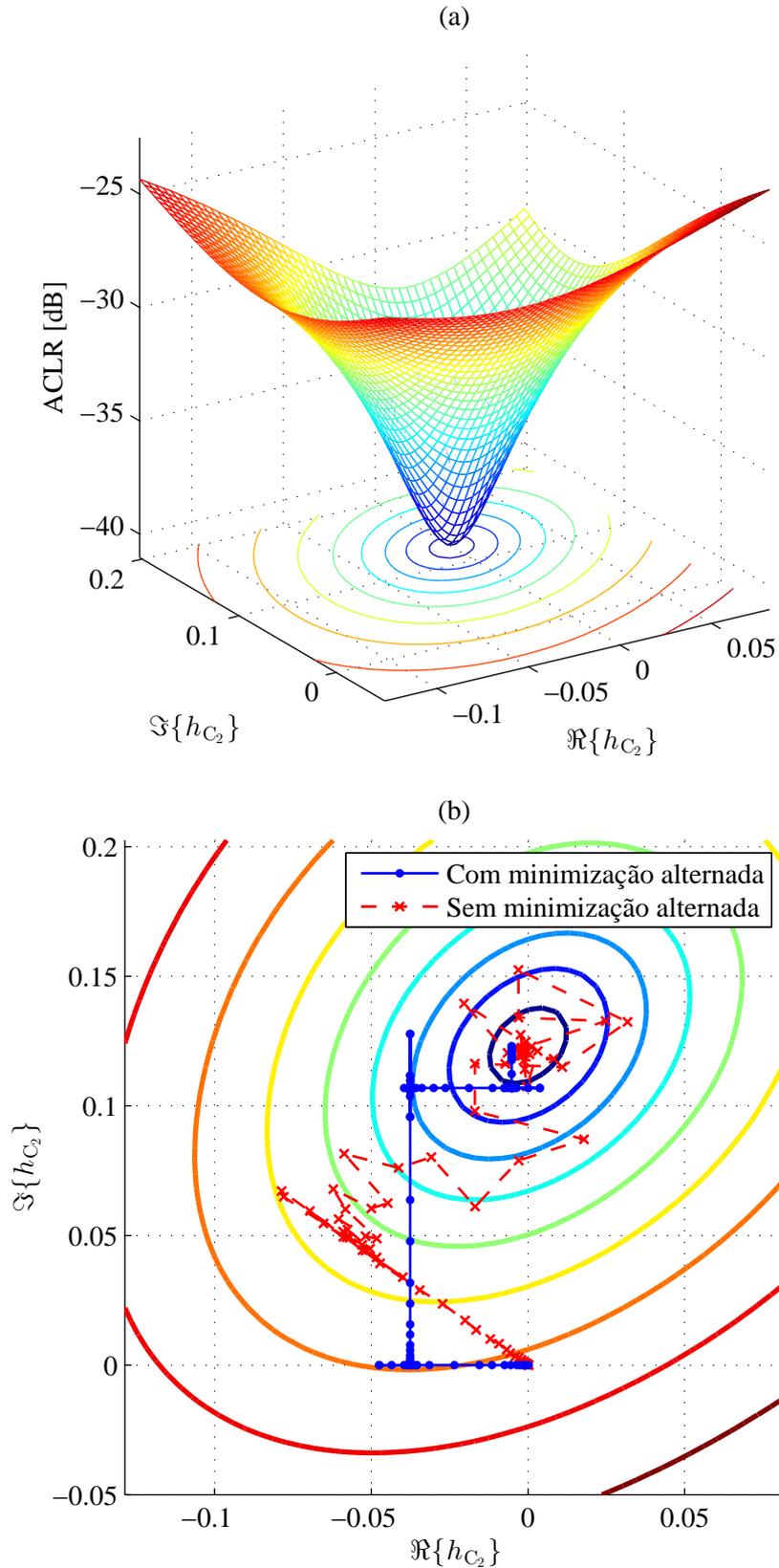


Figura 36 – (a) Função custo suave para sinal de entrada periódico; (b) Contorno da função custo e trajetórias de medida custo para o algoritmo *Nelder-Mead Simplex* com e sem o procedimento de minimização alternada.

Algoritmo 3 O algoritmo proposto *Orthogonal Scalar Feedback Linearization*

```

para  $i \leftarrow 2$  até  $K$  faça
  Calcule o vetor  $\Omega_i$  usando (4.15)
fim para
Inicialize  $\mathbf{h}_L = \mathbf{h}_C = \mathbf{h}_D = [1, 0, \dots, 0]^T$ 
para  $g \leftarrow 1$  até  $N_{\text{GlobalLoop}}$  faça
  para  $i \leftarrow 2$  até  $K$  faça
    enquanto Critério de parada não atendido faça
       $\triangleright$  Iterar de acordo com o algoritmo de otimização
      Gerar novo  $h_{C_i}$  para minimizar  $C(\mathbf{h}_D)$ 
      Calcule  $\hat{\mathbf{h}}_R$  de acordo com (4.14)
      Adapte  $\mathbf{h}_D$  de acordo com (4.13) e aplique no DPD
      Tome uma nova medida custo  $C(\mathbf{h}_D)$ , da saída do PA
    fim enquanto
     $\mathbf{h}_L = \mathbf{h}_D$ 
     $\mathbf{h}_C = \mathbf{h}_D$ 
  fim para
fim para

```

otimização numérica escolhido, assim como o critério de parada. Note que a minimização alternada não foi usada, dado que as partes real e imaginária de h_{C_i} estão sendo ajustadas simultaneamente.

O laço global (repetido $N_{\text{GlobalLoop}}$ vezes) leva ao reajuste de todos os coeficientes, o que pode efetivamente continuar a reduzir a intermodulação. Testes empíricos mostraram que fazer $N_{\text{GlobalLoop}} = 2$ é o suficiente para alcançar um resultado satisfatório.

4.3 Simulação do método proposto

O método OSFL proposto foi simulado na plataforma MATLAB. A distribuição Gaussiana correlacionada foi adotada para o sinal de entrada, uma vez que modela precisamente um grande número de sinais modulados, tal como a multiplexação ortogonal por divisão em frequência (*Orthogonal Frequency-Division Multiplexing* - OFDM) [64] com um grande número de subportadoras. Esta é a técnica de modulação empregada na maioria das tecnologias de comunicação sem fio. Adicionalmente, é bastante conhecida a característica de sinais OFDM possuírem uma alta PAPR [2], resultando em compressão e ceifamento dos picos de magnitude no processo de amplificação. O sinal de entrada não segue nenhum padrão de modulação, corresponde apenas a um sinal complexo com distribuição Gaussiana limitado em largura de faixa. Foi gerado um conjunto de 2048 sequências com 16384 amostras cada. A sequência com a maior PAPR foi selecionada desse conjunto para ser usada repetidamente ao longo de todas as iterações. Por apresentar os maiores picos de magnitude, esta sequência sofre uma maior compressão do amplificador, e tende a ser o pior caso, i.e., qualquer outra sequência (dentro do conjunto de 2048) ou tomada

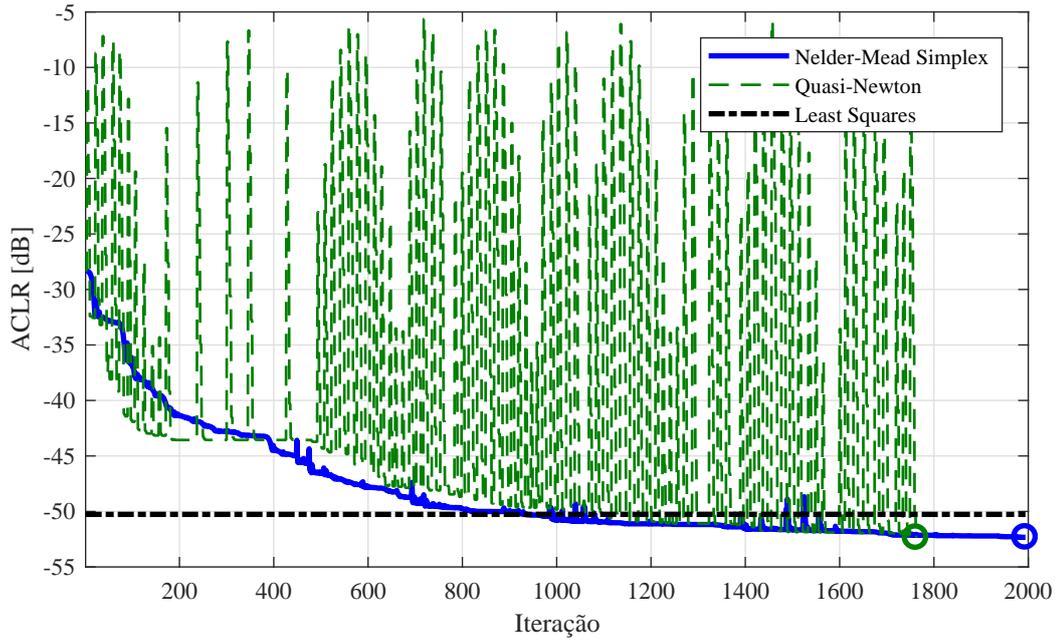


Figura 37 – Histórico de minimização do ACLR com ortogonalização dos coeficientes. Comparação entre os algoritmos de otimização numérica *Nelder-Mead Simplex* e *Quasi-Newton*.

de média levaria a um resultado melhor. O caso em que uma nova sequência é gerada a cada iteração será tratado na Seção 4.4. O ACLR foi escolhido como sendo a medida custo a ser minimizada. Os coeficientes do modelo do PA foram identificados a partir de sinais capturados, significando que esse comportamento corresponde ao amplificador empregado na implementação prática, apresentada na Capítulo 6.

Com relação ao algoritmo de otimização numérica, uma comparação foi feita entre o *Nelder-Mead Simplex* e *Quasi-Newton* cujos resultados são apresentados na Figura 37. Apesar do *Quasi-Newton* apresentar uma convergência mais rápida, nota-se picos do ACLR no início do ajuste de cada coeficiente. Como esses picos ultrapassam o valor inicial do ACLR, esta é uma característica indesejável. O *Nelder-Mead Simplex* foi adotado como melhor solução devido à característica de minimização monotonicamente decrescente, ao custo de algumas iterações a mais.

O método OSFL é comparado com outras técnicas existentes em termos de desempenho na redução de intermodulação, velocidade de convergência e variação de potência de saída do PA ao longo das iterações. Para todos os métodos, o modelo Gaussiano de (2.38) foi usado para representar o DPD com os parâmetros $M = 3$, $K_0 = 10$, $K_1 = 6$ e $K_2 = 3$, correspondendo às ordens 19^a, 11^a e 5^a para as três tomadas, respectivamente. Isso significa que o modelo tem $\sum_{m=0}^{M-1} K_m = 19$ coeficientes complexos ou 38 variáveis a serem ajustadas para minimizar a medida custo. Na comparação de velocidade de convergência e variação de potência, o método de realimentação escalar sem ortogonalização foi usado conforme em [34, 35]. Ao comparar a redução de intermodulação medindo o ACLR,

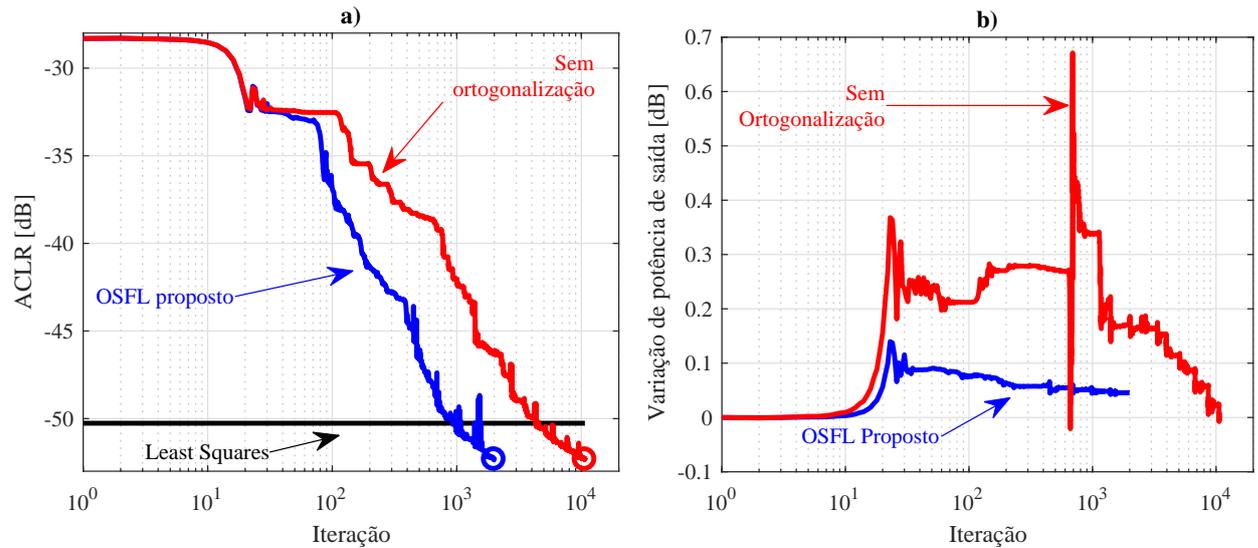


Figura 38 – Comparação entre o método proposto OSFL e o algoritmo de realimentação escalar sem ortogonalização [34, 35]. Foi medido ao longo das iterações a) ACLR para comparar tempo de convergência, e b) variação de potência de saída.

Tabela 6 – Comparação de desempenho em termos de tempo de convergência e variação de potência de saída. Os algoritmos testados são o de realimentação escalar sem ortogonalização [34, 35] e o OSFL proposto.

Algoritmo	Número total de iterações	Variação pico-a-pico na potência de saída [dB]
Sem ortogonalização	10679	0,69
OSFL proposto	1994	0,14

o método LS baseado em regressão linear foi usado conforme descrito em [32].

O histórico do ACLR e variação de potência de saída ao longo das iterações são comparadas na Figura 38. A comparação é feita entre dois algoritmos: o método proposto OSFL (com a ortogonalização dos coeficientes) e o algoritmo de realimentação escalar sem ortogonalização conforme descrito em [34, 35], todos empregando o *Nelder-Mead Simplex* como algoritmo de otimização numérica. O método LS é mostrado como referência na forma de uma reta, devido à sua natureza de iteração única. Nota-se que ambos algoritmos são capazes de atingir o mesmo valor de ACLR final. Contudo, há uma diferença no número total de iterações até a convergência e na variação de potência mostrados na Tabela 6. O OSFL diminui o número de iterações em 5,3 vezes e restringe a variação de potência. Estas vantagens se devem ao processo de ortogonalização, uma vez que o ajuste de um coeficiente não afetará os anteriores. O número de iterações corresponde ao número de medidas custo tomadas, que é igual ao número de iterações do laço interno do algoritmo OSFL.

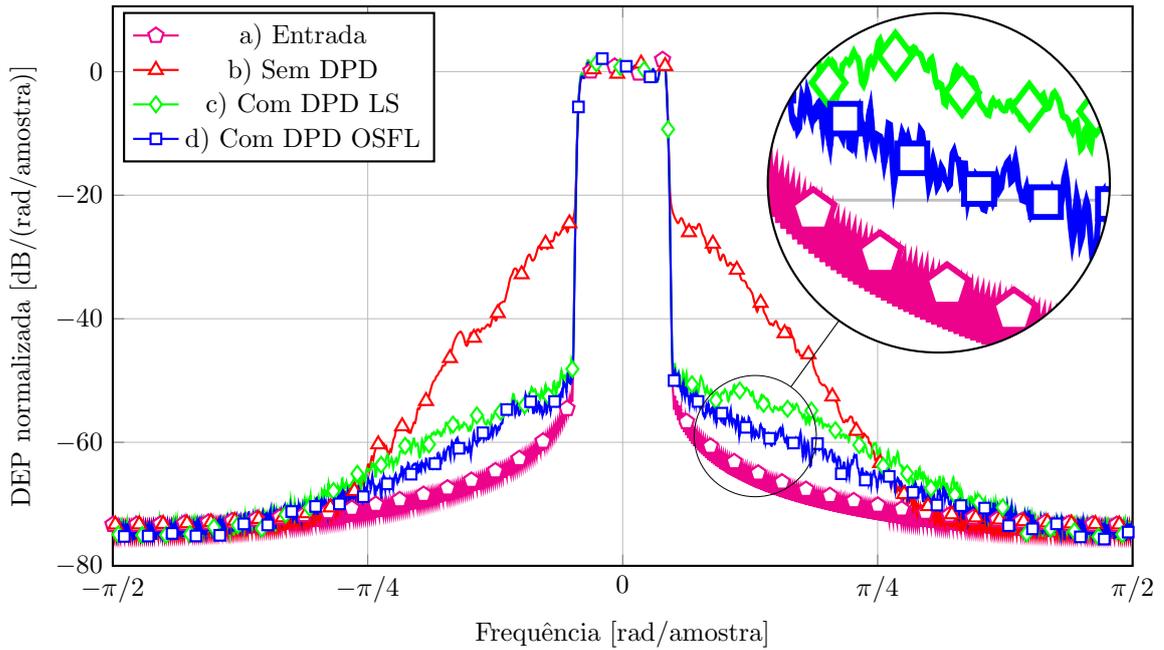


Figura 39 – Desempenho de intermodulação. a) sinal de entrada do DPD, i.e., o sinal não-distorcido, b) sinal de saída do PA sem o uso do DPD, c) sinal de saída do PA com DPD usando o método LS e d) sinal de saída do PA com DPD usando o método proposto OSFL com otimização numérica *Nelder-Mead Simplex*.

Tabela 7 – Comparação de desempenho em termos de NMSE e ACLR.

Comparação	NMSE [dB]	ACLR [dB]
Sem DPD	-19,74	-28,31
Com DPD LS	-44,73	-50,26
Com DPD OSFL	-27,81	-53,67

A Figura 39 compara a DEP⁴ dos sinais de entrada do sistema e saída do PA, assumindo diferentes esquemas de DPD. O OSFL apresentou um resultado melhor do que o método LS em termos de intermodulação. A comparação é quantificada na Tabela 7, usando duas figuras de mérito distintas: o ACLR e o NMSE entre o sinal de entrada do DPD e a saída linearizada do PA, sendo que os valores em negrito destacam os melhores resultados. Conforme esperado, o LS apresentou o melhor NMSE, uma vez que o erro quadrático no domínio do tempo é a métrica do critério de otimização desse algoritmo. Ao compensar o erro temporal, o DPD acaba por corrigir não somente as distorções não-lineares, mas também as lineares desnecessariamente. O OSFL proposto apresentou o melhor ACLR, que, por sua vez, é o critério de otimização.

Os gráficos AM/AM e AM/PM sem e com a ação do DPD são fornecidos na Figura 40.

⁴ A estimação da densidade espectral de potência foi feita usando a função do MATLAB `pwelch` com uma janela Gaussiana. Obviamente, diferentes técnicas de estimação, tipos de janela ou ambos, podem levar a diferentes estimativas, bem como diferentes valores finais do ACLR.

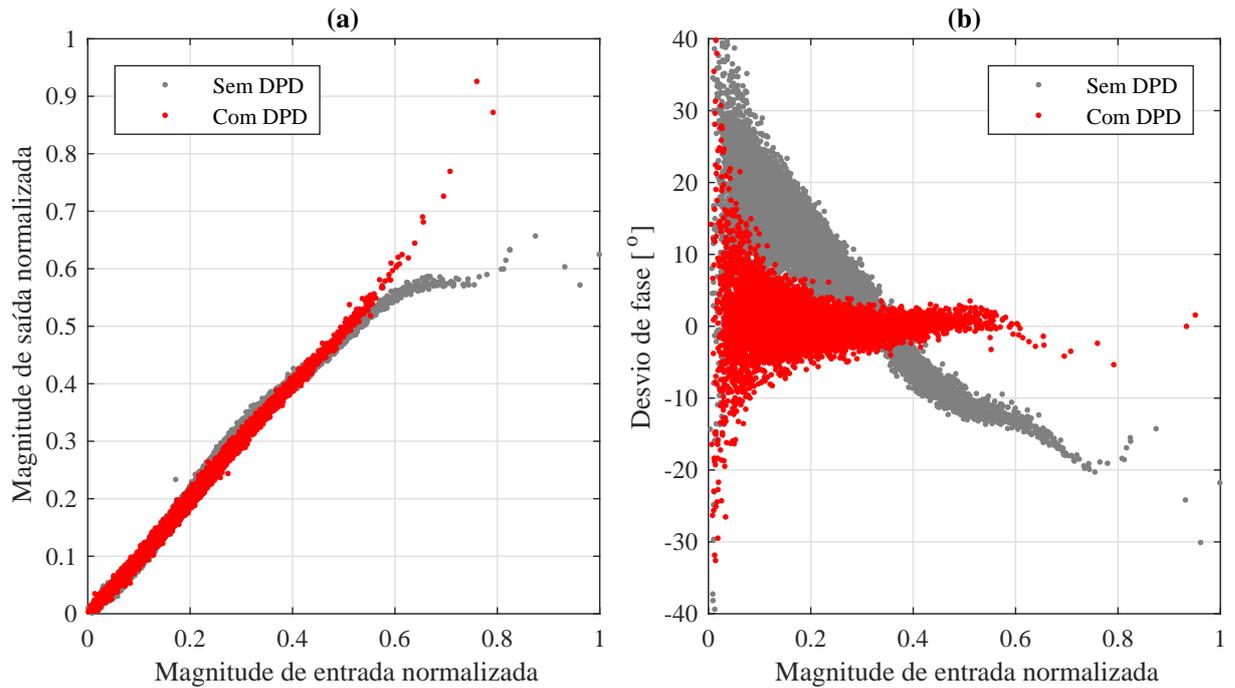


Figura 40 – Funções de transferência instantâneas sem e com o DPD dadas por (a) AM/AM e (b) AM/PM.

No algoritmo OSFL, o número total de iterações e o desempenho de intermodulação variam de acordo com o número de coeficientes, o algoritmo de otimização numérica escolhido e o critério de parada para considerar a medida custo minimizada para um dado coeficiente. O critério de parada para quebrar o laço **while** do **Algoritmo 3** é uma variável de ajuste para cada algoritmo de otimização numérica. A Figura 41 mostra o ACLR final e o número total de iterações para diferentes parâmetros M e K .

4.4 Conformidade com sequências aleatórias

Na Seção 4.3 desse capítulo, uma sequência foi repetidamente empregada como sinal de entrada. Nesta condição, a medida custo é uma constante para um dado conjunto de coeficientes. Em aplicações práticas, o sinal de entrada corresponde a sequências aleatórias. Linearizar um PA em uma aplicação prática com uma sequência repetida significaria conduzir a estimação dos parâmetros com a transmissão interrompida. Nesse caso, uma sequência de treinamento é transmitida e não o sinal contendo informações de usuários durante a linearização, o que é indesejável. Para sequências de entrada aleatórias, o sinal está em constante mudança causando variação do ACLR para um dado conjunto de coeficientes. Isso traz outro desafio, dado que a função custo a ser minimizada torna-se *não-suave* (ou de aparência ruidosa) e também variante no tempo. Os algoritmos de otimização numérica podem não apresentar o mesmo comportamento quando comparados com o resultado alcançado com uma função custo *suave*, podendo não convergir, por exemplo. A função custo para sequências aleatórias pode se tornar mais suave ao tomar a medida

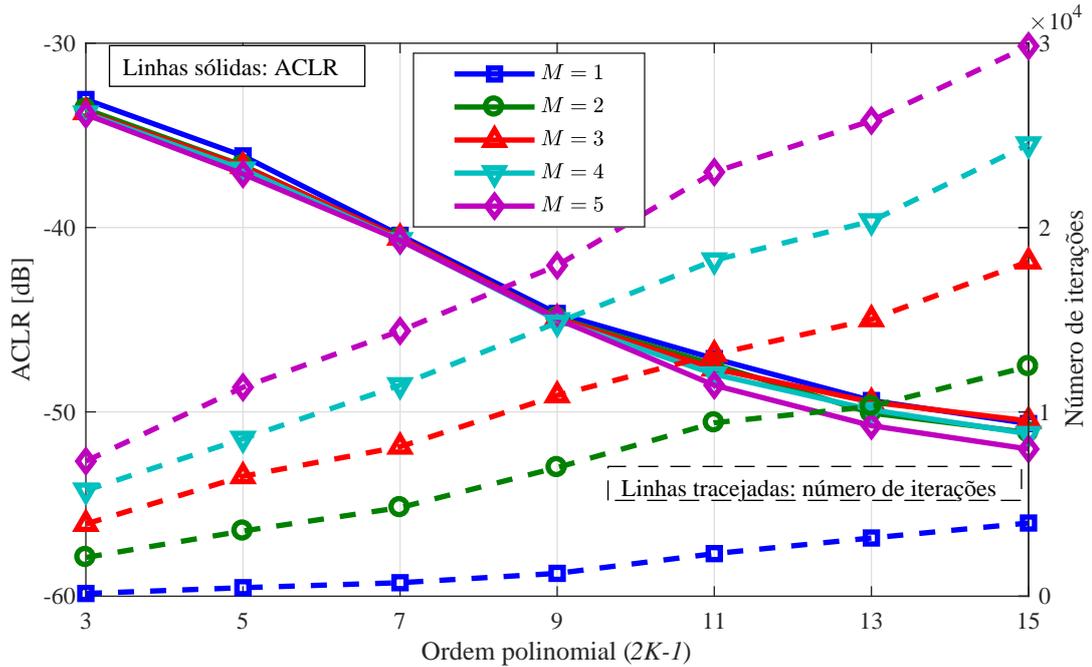


Figura 41 – ACLR e número total de iterações para diferentes parâmetros M e K usando o algoritmo OSFL.

custo sobre uma sequência mais longa do sinal, promovendo uma detecção mais precisa devido às mudanças causadas pela variação dos coeficientes. Isso resulta em melhorias, mas não resolve o problema. Algumas considerações são feitas com relação ao algoritmo OSFL proposto, quando usado com sequências aleatórias.

Na Seção 4.2.1 foi mostrado que a adaptação dos coeficientes dentro do processo de ortogonalização depende do conjunto de vetores Ω_i , que por sua vez, são dependentes da sequência de entrada. Isso sugere que um conjunto de vetores Ω_i , calculados para uma dada sequência não seriam válidos para as sequências subsequentes. Um teste de simulação mostra que esse erro é inversamente proporcional ao comprimento da sequência N . Portanto, esse problema pode ser superado fazendo N longo o suficiente, desde que o sinal de entrada seja estacionário no sentido amplo. O teste gera uma sequência de entrada, separando-a em duas metades, A e B, de forma que Ω_{iA} e Ω_{iB} são calculados para cada metade. Então, o NMSE entre os vetores Ω_{iA} e Ω_{iB} é calculado para cada índice de coeficiente i , como mostrado na Figura 42. Nota-se que o cálculo foi feito para os modelos polinomial banda base de (2.20) e Gaussiano ortogonal de (2.27), em que $K = 5$. A diminuição do NMSE com o aumento de N mostra que Ω_i tende para um valor constante. O modelo ortogonal Gaussiano promove uma maior precisão no algoritmo de ortogonalização, devido ao menor NMSE para um dado N . Desta forma, mesmo com sequências de entrada aleatórias, os vetores Ω_i podem ser calculados uma única vez e serem considerados constantes, desde que o comprimento da sequência seja longo o suficiente e os sinais de entrada sejam estacionários no sentido amplo.

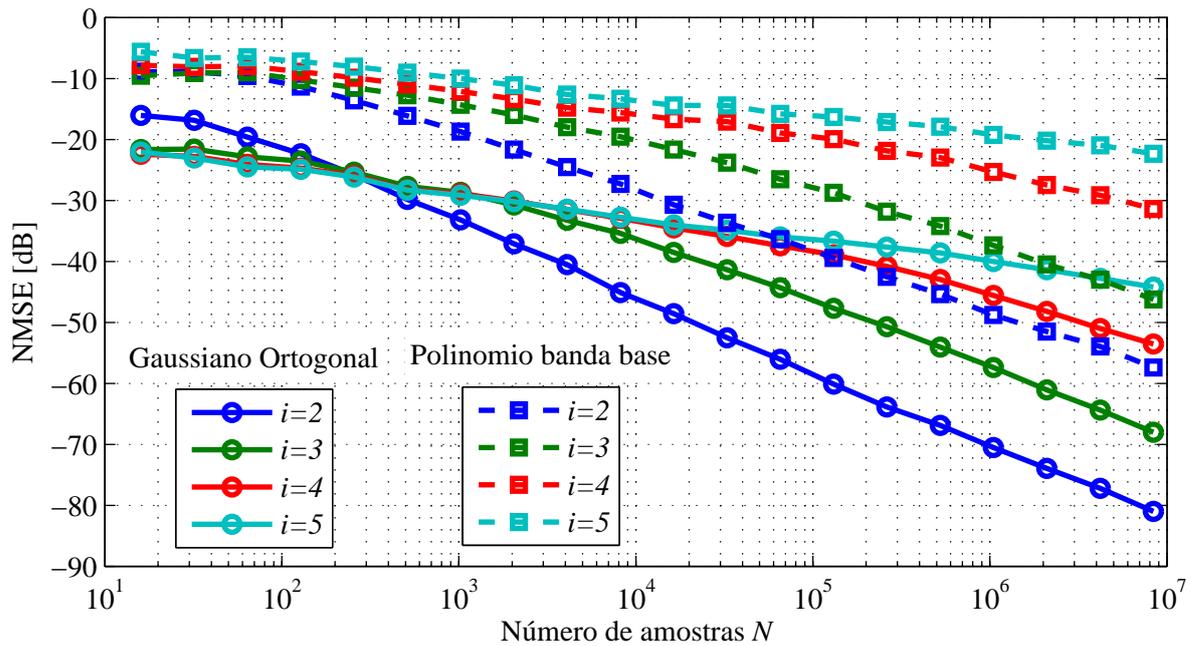


Figura 42 – NMSE entre os vetores Ω_{iA} e Ω_{iB} calculados para seqüências com diferentes comprimentos N . Simulação realizada para os modelos ortogonal Gaussiano e polinomial em banda base.

Considerando que a função custo para cada coeficiente é não-suave e não-estática, foi observado em simulações que os algoritmos de otimização *Nelder-Mead Simplex* e *Quasi-Newton* podem não convergir. Para tais casos, um algoritmo estocástico é mais apropriado, como o de busca de padrão (*Pattern search*) [65] que é um otimizador global. Observa-se que o processo de ortogonalização dos coeficientes, conforme descrito no **Algoritmo 3**, permanece o mesmo. Neste caso, apenas o método de otimização numérica para um dado coeficiente é modificado. A função custo e a trajetória de otimização para o coeficiente de 3ª ordem são mostradas na Figura 43. Nessa simulação, as seqüências de entrada aleatórias são geradas com 16384 amostras cada. Para cada iteração, o ACLR é calculado a partir da média de dez medidas custo. Os vetores Ω_i são calculados uma única vez a partir de uma seqüência de 1048576 amostras. O teste de desempenho e convergência é mostrado na Figura 44. O resultado do LS é apresentado como referência não mais como uma linha reta, uma vez que o ACLR varia com o tempo para um conjunto de coeficientes de valores fixos. Apesar de levar mais iterações do que os algoritmos apresentados anteriormente, esse método foi capaz de convergir com um resultado de ACLR final superior ao do LS. Uma característica notável do *Pattern search* é o aumento do ACLR nas iterações iniciais de cada coeficiente. Porém, ao contrário do *Quasi-Newton*, esse é um método de otimização restrita que permite impor limites no valor máximo dos coeficientes, restringindo os picos de ACLR.

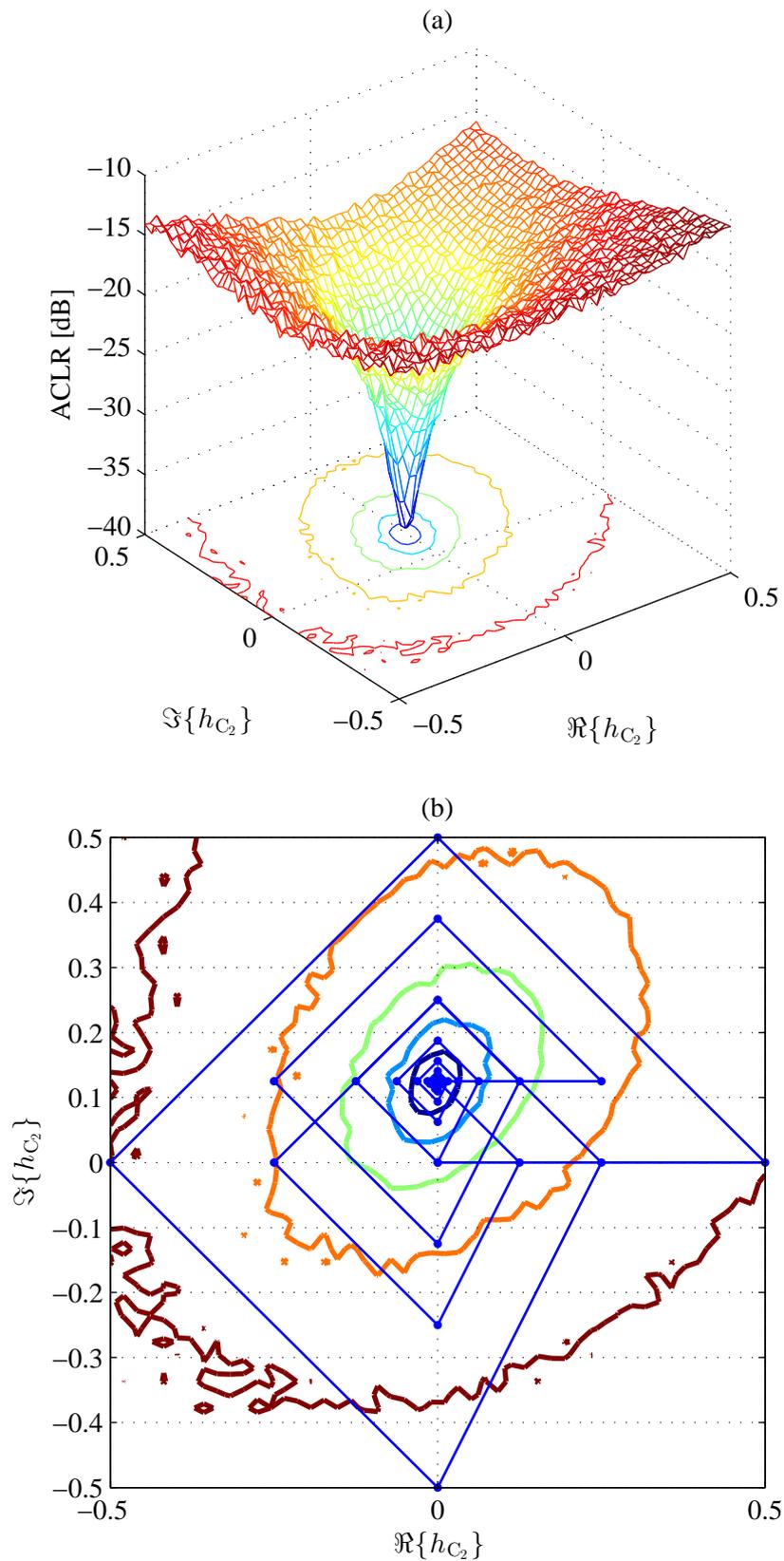


Figura 43 – (a) Função custo não-suave. (b) Trajetória de otimização do coeficiente de 3ª ordem para o algoritmo *Pattern search*.

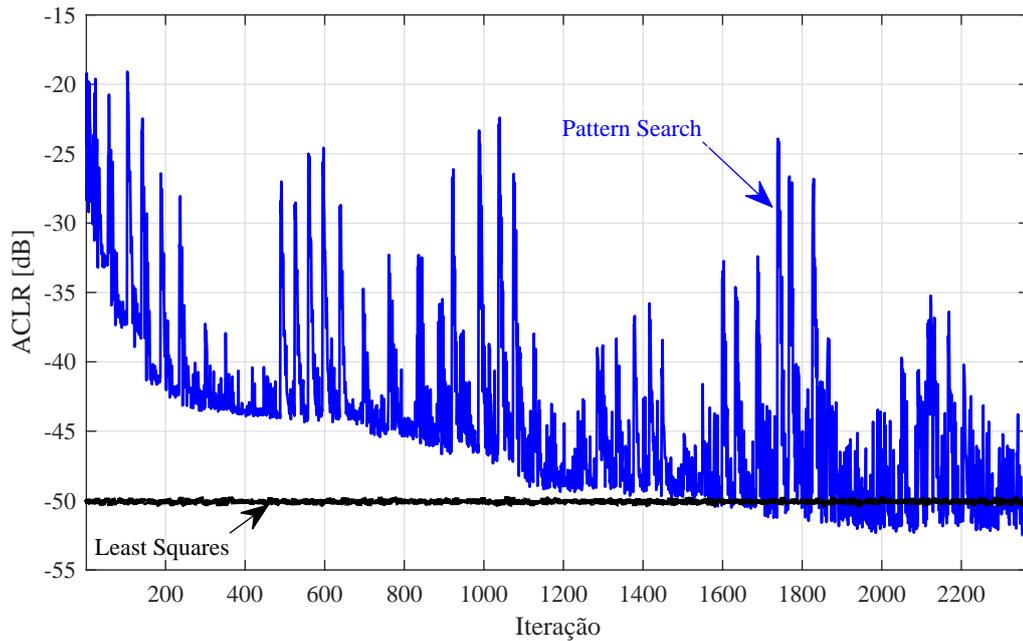


Figura 44 – Histórico de minimização do ACLR no algoritmo *Pattern search* para sequências de entrada aleatórias.

4.5 Otimização da intermodulação dentro da banda do canal

Os produtos de intermodulação gerados pelo PA estão localizados tanto OOB quanto IB. A qualidade do sinal, e conseqüentemente o limiar de recepção, serão afetados pela intermodulação IB, enquanto as emissões OOB e a margem em relação à máscara espectral serão afetados pela intermodulação OOB. A primeira impacta no desempenho do sistema e a segunda causa interferência nos sistemas ocupando as bandas adjacentes. Sendo assim, a máxima potência do PA estará sempre limitada tanto pela máscara espectral, que é afetada pela intermodulação OOB, quanto pela qualidade do sinal, que é afetada pela intermodulação IB.

Para padrões em que a máscara não é rigorosa e a máxima potência do PA é limitada pela qualidade do sinal, a melhor métrica é a distorção IB, tomando a MER como medida custo. Por exemplo, padrões de televisão digital como DVB-T (*Digital Video Broadcasting — terrestrial - DVB-T*)/T2 e ISDB-T (*Integrated Services Digital Broadcasting — Terrestrial - ISDB-T*) apresentam diferentes requisitos em relação à máscara de emissão [66] e MER para qualidade de sinal. Para DVB-T, à medida em que a potência de saída do PA é aumentada, a especificação de MER falha primeiro, enquanto a máscara ainda é atendida. De forma semelhante, para ISDB-T quando a potência de saída do PA é aumentada, as emissões OOB falharão primeiro, enquanto a MER ainda atende as especificações. Isso significa que é preferível adotar o ACLR como medida custo para o ISDB-T, assim como uma métrica IB para o DVB-T.

Propõe-se então uma metodologia em que um sistema de DPD pode ser direta-

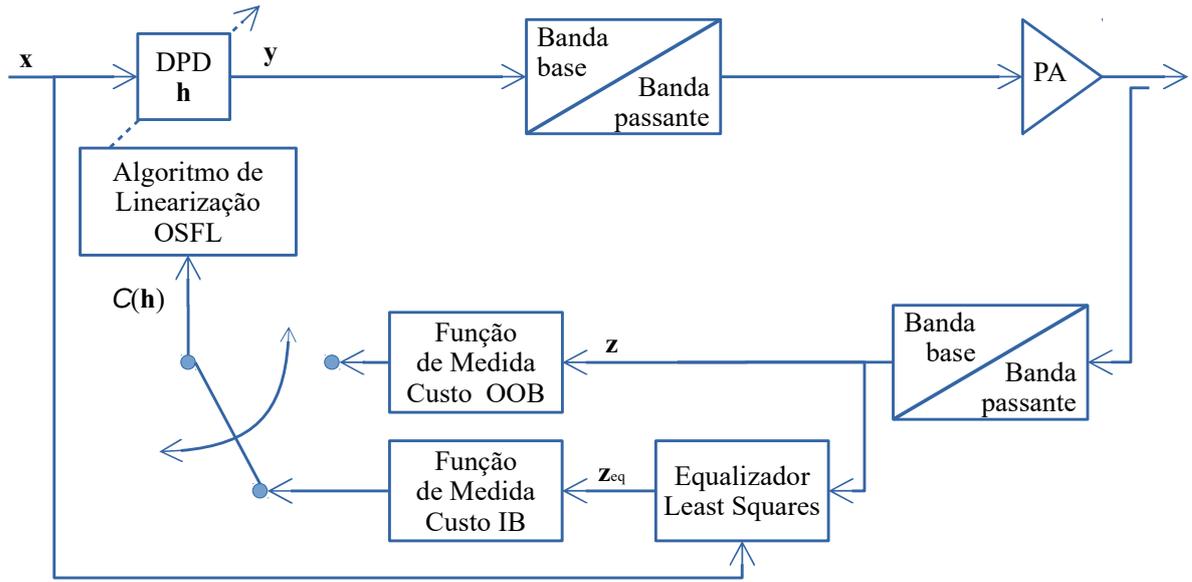


Figura 45 – Diagrama de blocos do sistema proposto de linearização ortogonal com realimentação de escalar (OSFL), com seleção de diferentes medidas custo.

mente otimizado para uma dada métrica, escolhida de acordo com a característica limitante, com diagrama de blocos ilustrado na Figura 45. A medida custo IB é definida como a razão entre a potência do sinal erro IB e a potência do sinal IB, ambos no domínio da frequência, dado por

$$C_{IB}(\mathbf{h}) = \frac{\int_{f \in IB} P_e(f) df}{\int_{f \in IB} P_z(f) df}, \quad (4.19)$$

em que $P_e(f)$ representa a DEP do sinal erro na saída do equalizador LS, conforme mostrado na Figura 45. Medir a potência do erro não é uma tarefa direta. A primeira opção seria demodular o sinal. A segunda opção é dada pela extração do sinal erro no domínio do tempo gerado a partir da versão equalizada de \mathbf{z} que remove suas distorções lineares, restando apenas as não-lineares. A segunda opção foi escolhida e é dada por $\mathbf{e} = \mathbf{x} - \mathbf{z}_{eq}$, em que as seqüências \mathbf{x} e \mathbf{z}_{eq} necessitam estar alinhadas no tempo antes de realizar a subtração.

Uma comparação foi feita para medir o desempenho da linearização OSFL em função de diferentes medidas custo. Foram comparados três casos: o algoritmo baseado em regressão linear LS e o OSFL com as medidas custo OOB e IB. A Figura 46 mostra a DEP para os seguintes grupos de sinal: i) o sinal de entrada (sistema linear) e a saída do PA sem DPD como referências; ii) a saída do PA com DPD para o LS [32], que minimiza o erro quadrático médio no domínio do tempo e para o DPD OSFL para as métricas OOB de (3.33) e IB de (4.19) e; iii) o sinal erro para todos os algoritmos plotado apenas para $f \in IB$. A redução da intermodulação é visualmente percebida na Figura 46. Para quantificar a melhoria de maneira objetiva, a Tabela 8 compara os três métodos usando

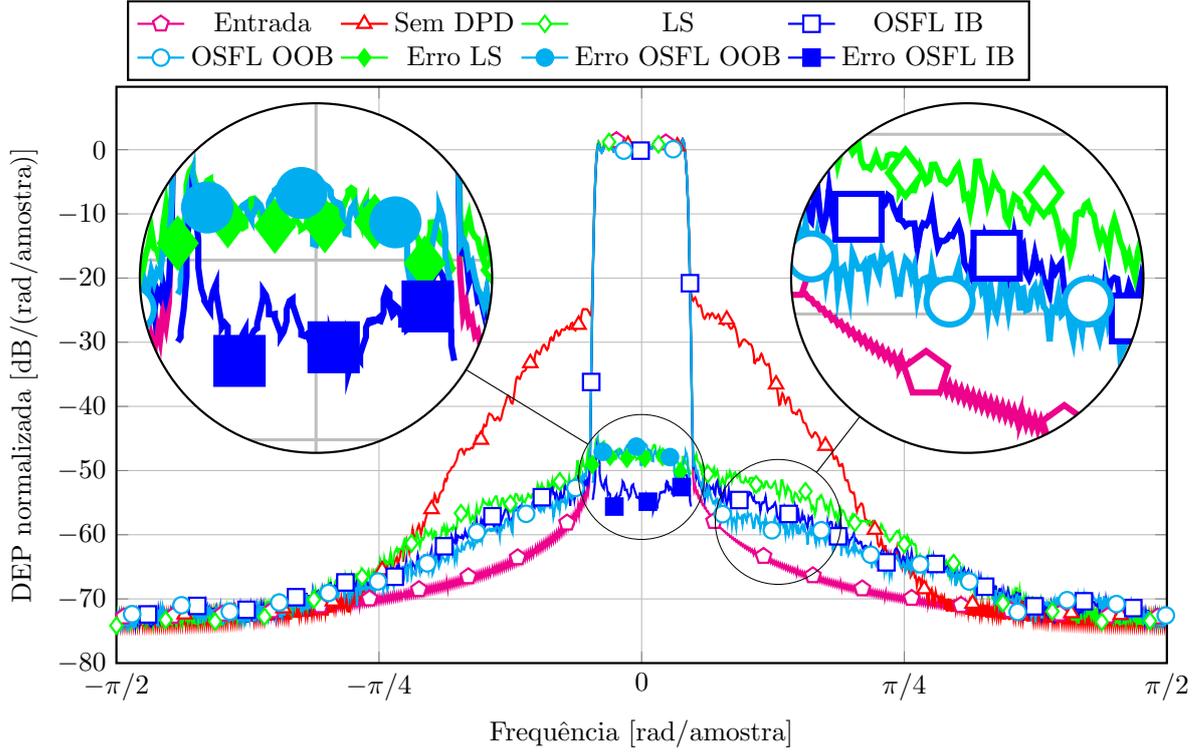


Figura 46 – DEP de sinais para comparação de três métodos: sistema DPD baseado em LS e OSFL com métricas OOB e IB.

Tabela 8 – Comparação de desempenho em termos de NMSE e métricas de intermodulação OOB e IB.

Método	NMSE	Intermodulação fora da banda [dB]	Intermodulação dentro da banda [dB]
Sem DPD	-19,93	-28,55	-22,28
Com DPD LS	-44,56	-50,06	-47,56
Com DPD OSFL OOB	-31,25	-54,24	-47,60
Com DPD OSFL IB	-31,44	-52,73	-52,89

as três diferentes métricas: o NMSE entre o sinal de entrada do DPD e o de saída do PA no domínio do tempo e o nível relativo das intermodulações OOB e IB. Os números em negrito destacam os melhores resultados dentro de cada métrica. O algoritmo OSFL proposto supera todos os demais considerando a métrica escolhida. Em outras palavras a emissão OOB é a menor para o OSFL OOB, assim como a distorção IB é a menor para o OSFL IB. Quando comparado com o DPD LS convencional, houve uma melhoria de 5,33 dB na intermodulação IB.

4.6 Processo de ortogonalização alternativo

Considerando a forma matricial da equação do modelo comportamental $\mathbf{y} = \mathbf{X}\mathbf{h}$ de (2.26), o processo ortogonalização descrito na Seção 4.2 adapta os coeficientes \mathbf{h} para obter

o efeito desejado. De forma alternativa, pode-se atuar na matriz \mathbf{X} para obter o mesmo efeito de acordo com o algoritmo de ortogonalização de Gram-Schmidt [67], conforme descrito no **Algoritmo 4**, em que $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_K]$ é a versão ortogonalizada de \mathbf{X} .

Algoritmo 4 Algoritmo de ortogonalização de Gram-Schmidt.

```

 $\mathbf{u}_1 = \frac{\mathbf{x}_1}{\sqrt{\mathbf{x}_1^\dagger \mathbf{x}_1}}$ 
para  $i \leftarrow 2$  até  $K$  faça
   $\mathbf{u}_i = \mathbf{x}_i$ 
  para  $j \leftarrow 1$  até  $i - 1$  faça
     $\mathbf{u}_i = \mathbf{u}_i - \frac{\mathbf{u}_i^\dagger \mathbf{u}_j}{\mathbf{u}_j^\dagger \mathbf{u}_j} \mathbf{u}_j$             $\triangleright$  Remove de  $\mathbf{u}_i$  a projeção de  $\mathbf{u}_i$  em  $\mathbf{u}_j$ 
  fim para
   $\mathbf{u}_i = \frac{\mathbf{u}_i}{\sqrt{\mathbf{u}_i^\dagger \mathbf{u}_i}}$             $\triangleright$  Normaliza para que  $\mathbf{u}_i^\dagger \mathbf{u}_i = 1$ 
fim para

```

Com a ortogonalização, o produto interno entre colunas distintas de \mathbf{U} é nulo, i.e., $\mathbf{U}^\dagger \mathbf{U}$ é uma matriz diagonal. Além disso, devido à etapa de normalização, $\mathbf{U}^\dagger \mathbf{U} = \mathbf{I}$, ou seja, \mathbf{U} é *ortonormal*. Dessa forma, a saída do modelo ortonormalizado é obtida por meio de $\mathbf{y} = \mathbf{U}\mathbf{h}$. Quando a normalização é omitida do **Algoritmo 4**, obtêm-se exatamente o mesmo resultado do processo de ortogonalização de coeficientes conforme descrito na Seção 4.2.

A seguir são comparadas semelhanças e diferenças entre um modelo ortogonalizado com o método de Gram-Schmidt e o modelo Gaussiano ortogonal de (2.27). Primeiramente, nota-se que o Gaussiano resulta em $\mathbb{E}[\mathbf{X}^\dagger \mathbf{X}] = \mathbf{I}$, apenas quando \mathbf{x} apresenta distribuição Gaussiana. Enquanto com Gram-Schmidt pode-se obter a ortogonalização com qualquer distribuição. Também, para o Gaussiano apenas a *média* do produto interno entre colunas distintas de \mathbf{X} é nula, enquanto para o Gram-Schmidt o produto interno entre colunas distintas de \mathbf{U} é nulo para a sequência \mathbf{x} usada para montar a matriz \mathbf{X} .

Quando o efeito memória é adicionado ao modelo Gaussiano ortogonal de (2.27), formando o Gaussiano com memória de (2.38), o último perde a propriedade da ortogonalidade. Em tese, o algoritmo de Gram-Schmidt poderia ser aplicado no modelo com memória descrito em (2.38) para se recuperar a ortogonalidade para uma dada sequência de entrada. Contudo, observou-se que, para esse caso, o **Algoritmo 4** não produziu o efeito esperado, ou seja, $\mathbf{U}^\dagger \mathbf{U} \neq \mathbf{I}$. Isso aconteceu por duas razões: i) a não-ortogonalidade de uma matriz apresenta grau, medido pelo número de condicionamento da matriz \mathbf{X} . Em um modelo com memória, \mathbf{X} apresenta um alto número de condicionamento devido à forte correlação entre as colunas de \mathbf{X} , uma vez que o conteúdo das colunas se repetem deslocadas de $m = 0, 1, \dots, M - 1$ amostras para as várias tomadas, e; ii) porque o método de

duplo laço **for** conforme descrito em **Algoritmo 4** acumula erro ao longo das iterações que remove as projeções. O método de ortogonalização dos coeficientes descrito na Seção 4.2 não sofre essa limitação.

Para resolver o problema para os modelos com memória, o algoritmo de Gram-Schmidt foi alterado conforme descrito no **Algoritmo 5**.

Algoritmo 5 Algoritmo de ortogonalização de Gram-Schmidt modificado, baseado-se em matriz de projeção.

```


$$\mathbf{u}_1 = \frac{\mathbf{x}_1}{\sqrt{\mathbf{x}_1^\dagger \mathbf{x}_1}}$$

para  $i \leftarrow 2$  até  $K$  faça
   $\mathbf{u}_i = \mathbf{x}_i$ 
   $\mathbf{R}_i = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{i-1}]$ 
   $\mathbf{u}_i = \mathbf{u}_i - \mathbf{R}_i (\mathbf{R}_i^\dagger \mathbf{R}_i)^{-1} \mathbf{R}_i^\dagger \mathbf{u}_i$  ▷ Remove de  $\mathbf{u}_i$  a projeção de  $\mathbf{u}_i$  em  $\mathbf{u}_j$ 
  simultaneamente para  $j = 1, 2, \dots, i - 1$ 
   $\mathbf{u}_i = \frac{\mathbf{u}_i}{\sqrt{\mathbf{u}_i^\dagger \mathbf{u}_i}}$  ▷ Normaliza para que  $\mathbf{u}_i^\dagger \mathbf{u}_i = 1$ 
fim para

```

A versão modificada apresenta agora um único laço **for**, dado que a projeção do vetor \mathbf{u}_i nas colunas anteriores é removida em uma única etapa por meio do conceito de matriz de projeção. Com essa modificação não há erro acumulativo, sendo possível a ortogonalização de modelos com memória. Por definição [68],

$$\mathbf{P}_i \triangleq \mathbf{R}_i (\mathbf{R}_i^\dagger \mathbf{R}_i)^{-1} \mathbf{R}_i^\dagger \quad (4.20)$$

é a matriz de projeção do subespaço \mathbf{R}_i . Na multiplicação de \mathbf{P}_i por um vetor \mathbf{u}_i , o resultado é a projeção de \mathbf{u}_i no subespaço vetorial definido pelas colunas de \mathbf{R}_i , ou seja, $\mathbf{P}_i \mathbf{u}_i = \hat{\mathbf{u}}_i$, em que $\hat{\mathbf{u}}_i$ é a projeção ortogonal de \mathbf{u}_i em \mathbf{R}_i .

Ao usar o algoritmo de ortogonalização de Gram-Schmidt, na sua versão modificada quando necessário, teremos um modelo ortogonalizado para qualquer combinação de coeficientes. Isso significa que o processo de linearização pode ser obtido diretamente ajustando-se coeficiente a coeficiente de acordo com um algoritmo de otimização numérica, sem a necessidade de se adaptar os coeficientes a cada iteração. Quando comparado ao OSFL descrito no **Algoritmo 3**, a diferença se dá por causa da ortonormalização. Uma simulação de desempenho foi feita comparando o OSFL proposto (equivalendo ao Gram-Schmidt sem normalização) e o Gram-Schmidt ortonormal, em que os resultados são ilustrados na Figura 47. Nota-se um aumento na velocidade de convergência para o caso ortonormal. Outra característica da ortonormalização é uma menor dispersão na magnitude dos coeficientes. Por exemplo, o desvio padrão do conjunto de coeficientes da Figura 47 para o método ortogonal e ortonormal foi de 380,41 e 0,21, respectivamente. Isso é uma vantagem quando emprega-se um método de otimização numérica restrita,

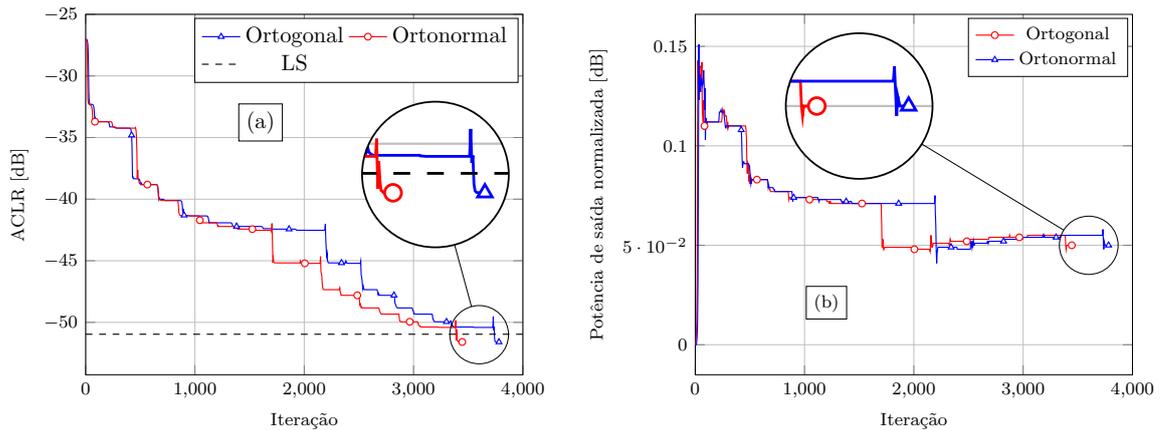


Figura 47 – Comparando processos de linearização com realimentação escalar usando Gram-Schmidt, quando há somente ortogonalização (equivalente ao OSFL proposto) e com ortonormalização. Comparação feita em termos de (a) ACLR e; (b) Potência de saída do PA ao longo das iterações.

em que impõe-se um limite máximo na magnitude dos coeficientes para restringir o valor máximo de ACLR ao longo das iterações.

4.7 Processo de ajuste manual

É possível remover completamente o caminho de realimentação fazendo o ajuste manual dos coeficientes. Desta forma, a medida custo é tomada visualmente por um operador humano, ao checar o sinal de saída do PA em um analisador de espectro, em uma operação de tentativa e erro. Apesar de não ser um processo automático, é adequado para aplicações de baixo custo. Em tais sistemas sem realimentação, o DPD é tipicamente realizado fazendo desenhos manuais das curvas AM/AM e AM/PM. Este procedimento apresenta um tempo de convergência longo. Ao aplicar o algoritmo OSFL manualmente, os valores dos coeficientes são ajustados ao invés de desenhar curvas. O tempo de convergência do ajuste manual dos coeficientes se torna mais rápido do que manualmente desenhar curvas AM/AM e AM/PM, devido ao número de parâmetros reduzidos e ao processo de ortogonalização.

Aplicações como transmissores de TV Digital de baixa potência são extremamente sensíveis ao custo. A remoção completa de todo o circuito de realimentação, incluindo o ADC, seria economicamente interessante neste cenário.

4.8 Conclusão

O presente capítulo apresentou o método de linearização OSFL proposto, que é a principal contribuição deste trabalho. Destaca-se o desenvolvimento teórico para o

processo de adaptação dos coeficientes que resulta na ortogonalização dos produtos de intermodulação no domínio da frequência. Isso permite a criação de um sistema de pré-distorção com realimentação escalar, em que os coeficientes são ajustados de forma independente de acordo com algum algoritmo de otimização numérica, com o objetivo de reduzir a medida custo. O método proposto foi simulado e os resultados obtidos foram satisfatórios.

5 Pré-distorção em tempo real

O projeto de um sistema de pré-distorção digital pode ser dividido em duas grandes partes: o bloco do DPD responsável por pré-distorcer o sinal de entrada em tempo real e a identificação do modelo inverso do PA cujas características são carregadas no bloco do DPD. Neste capítulo são apresentados os conceitos necessários para gerar a pré-distorção no sinal de entrada, amostra por amostra, na velocidade da frequência de amostragem. Os sistemas de comunicação sem fio modernos apresentam sinais com grande largura de faixa, exigindo uma frequência de amostragem também alta para representá-los. Além do mais, a não-linearidade introduzida pelo DPD gera um crescimento espectral na forma de intermodulação, exigindo uma largura de faixa ainda maior. Gerar a pré-distorção em tempo real significa implementar a equação do modelo escolhido em função do sinal de entrada e um conjunto de coeficientes. Em sistemas típicos, o bloco do DPD opera com taxas de amostragem entre 50 MS/s e 200 MS/s, ou seja, o intervalo de tempo entre as amostras variam de 5 ns a 20 ns. Devido a estas especificações, a pré-distorção deve ser executada não por *software*, mas por algum *hardware* digital, que pode ser uma FPGA ou um circuito integrado de aplicação específica (*Application Specific Integrated Circuit* - ASIC). Em outras palavras, pode ser implementada usando portas lógicas e registradores. Além do mais os sinais são quantizados em número finito de bits correspondendo a números inteiros, isto é, comumente usa-se a representação com ponto fixo. A seção 5.1 apresenta algumas operações trigonométricas complexas feitas a partir de tais representações. A seção 5.2 aborda diferentes interpretações de uma dada equação de modelo, resultando em diferentes formas de implementação, cada qual com suas características e vantagens.

5.1 Trigonometria de ponto fixo

Praticamente todas equações de modelo não-linear da Seção 2.3 utilizam o conceito de polinômio em banda base, em que são gerados apenas produtos de intermodulação próximos da frequência fundamental. Para tais modelos, é necessário calcular o módulo do sinal de entrada $|x(n)|$. Seja $x = p + jq$, da trigonometria

$$|x| = \sqrt{p^2 + q^2}. \quad (5.1)$$

O cálculo da magnitude do sinal complexo é trivial por *software*. O cálculo de $p^2 + q^2$ que corresponde a $x^*(n)x(n)$ é fácil de ser obtido por *hardware* por meio de multiplicadores. Contudo, considerando que p e q são números inteiros com representação de ponto fixo, a operação de raiz quadrada não é trivial. Para resolver tal problema, faz-se uso dos algoritmos de trigonometria de ponto fixo chamados de computador digital

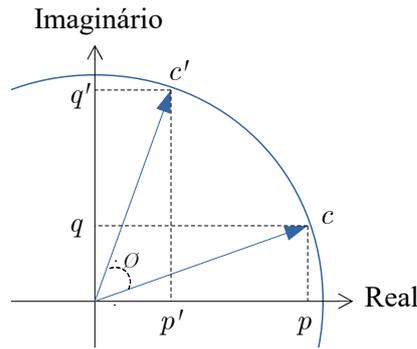


Figura 48 – Rotação de um ângulo ϕ a partir de um número complexo c , gerando c' .

de rotação de coordenadas (*COordinate Rotation DIgital Computer* - CORDIC) [69]. Trata-se de uma classe de algoritmos com uso eficiente do *hardware* empregando somente operações básicas como deslocamento de bits e somas.

O CORDIC baseia-se na rotação de um vetor no plano complexo, como apresentado na Figura 48. A operação matemática que gera o número complexo c' a partir de c é

$$\begin{aligned} c' &= c e^{j\phi} \\ &= (p + jq) [\cos(\phi) + j \operatorname{sen}(\phi)] \\ p' + jq' &= p \cos(\phi) - q \operatorname{sen}(\phi) + j [p \operatorname{sen}(\phi) + q \cos(\phi)], \end{aligned} \quad (5.2)$$

podendo ser representada pela multiplicação matricial

$$\begin{bmatrix} p' \\ q' \end{bmatrix} = \begin{bmatrix} \cos(\phi) & -\operatorname{sen}(\phi) \\ \operatorname{sen}(\phi) & \cos(\phi) \end{bmatrix} \begin{bmatrix} p \\ q \end{bmatrix}. \quad (5.3)$$

Multiplicando o segundo membro da equação por $\frac{\cos(\phi)}{\cos(\phi)}$, tem-se

$$c' = \cos(\phi) \begin{bmatrix} 1 & -\tan(\phi) \\ \tan(\phi) & 1 \end{bmatrix} c. \quad (5.4)$$

Discretizando o número de rotações para um conjunto $\{\theta_i\}$, $i = 0, 1, \dots, R - 1$,

$$\begin{aligned} \tan(\phi_i) &= \pm 2^{-i} \\ \phi_i &= \arctan(\pm 2^{-i}). \end{aligned} \quad (5.5)$$

Assim, (5.4) torna-se

$$c_{i+1} = \cos(\arctan(\pm 2^{-i})) \begin{bmatrix} 1 & \mp 2^{-i} \\ \pm 2^{-i} & 1 \end{bmatrix} c_i, \quad (5.6)$$

em que c_i é o valor complexo assumindo a i -ésima rotação dentro do conjunto de R possíveis rotações.

Como a função arco-tangente apresenta simetria ímpar, os ângulos podem ser ainda mais restritos a apenas valores positivos, como

$$c_{i+1} = \cos(\arctan(2^{-i})) \begin{bmatrix} 1 & -d_i 2^{-i} \\ d_i 2^{-i} & 1 \end{bmatrix} c_i. \quad (5.7)$$

em que $d_i \in \{-1, 1\}$ e faz-se $d_i = -1$ para obter uma rotação no sentido horário. Após R rotações sucessivas, o valor do cosseno é uma constante dada por

$$K_R = \prod_{i=0}^{R-1} \cos(\arctan(2^{-i})) = \prod_{i=0}^{R-1} (1 + 2^{-2i})^{-\frac{1}{2}}, \quad (5.8)$$

independente do sentido de rotação uma vez que $\cos(-\phi) = \cos(\phi)$. Quando o número de rotações tender para o infinito a constante torna-se

$$K_\infty = \prod_{i=0}^{\infty} (1 + 2^{-2i})^{-\frac{1}{2}} = 0,607252935008881\dots \quad (5.9)$$

Isso significa que K_R pode ser pré-calculado de acordo com o número de rotações R e ser multiplicado apenas após a última rotação como um fator de escalonamento. Assim sendo, a equação da rotação aplicada em cada iteração do algoritmo é dada por

$$c_{i+1} = \begin{bmatrix} 1 & -d_i 2^{-i} \\ d_i 2^{-i} & 1 \end{bmatrix} c_i. \quad (5.10)$$

com partes real e imaginárias dadas por

$$\begin{cases} p_{i+1} = p_i - d_i 2^{-i} q_i \\ q_{i+1} = q_i + d_i 2^{-i} p_i. \end{cases} \quad (5.11)$$

Na Equação (5.11) há apenas operações simples de soma (ou subtração) e divisões por alguma potência de 2, que é obtida deslocando-se bits. A Tabela 9 mostra valores dos ângulos para 10 rotações. Como os ângulos de rotação são restritos à tabela de arco-tangente, a rotação por um ângulo θ qualquer é obtida a partir de sucessivas rotações selecionando-se a direção de rotação d_i de forma que

$$\theta \simeq \sum_{i=0}^R d_i \phi_i \quad (5.12)$$

em que (5.12) torna-se uma igualdade quando $R \rightarrow \infty$. Porém, ainda que infinitas rotações sejam feitas no mesmo sentido, há um ângulo máximo de convergência dado por

$$\theta_{\max} = \sum_{i=0}^{\infty} \arctan(2^{-i}) = 99,882965834689614\dots^\circ. \quad (5.13)$$

Desta forma, o algoritmo de rotação converge para os 1° e 4° quadrantes e uma pequena parcela nos 2° e 3° quadrantes, ou seja, de $-\theta_{\max}$ a θ_{\max} . Para funcionar em toda a circunferência é necessário contornar o problema fazendo

$$c_0 = \begin{cases} c, & \text{se } -90^\circ < \angle c < 90^\circ \\ -c, & \text{caso contrário,} \end{cases} \quad (5.14)$$

Tabela 9 – Tabela dos ângulos necessários para algoritmos CORDIC.

i	$\tan(\phi) = 2^{-i}$	$\phi = \arctan(2^{-i}) [^\circ]$
0	1	45
1	1/2	26,5651
2	1/4	14,0362
3	1/8	7,1250
4	1/16	3,5763
5	1/32	1,7899
6	1/64	0,8952
7	1/128	0,4476
8	1/256	0,2238
9	1/512	0,1119

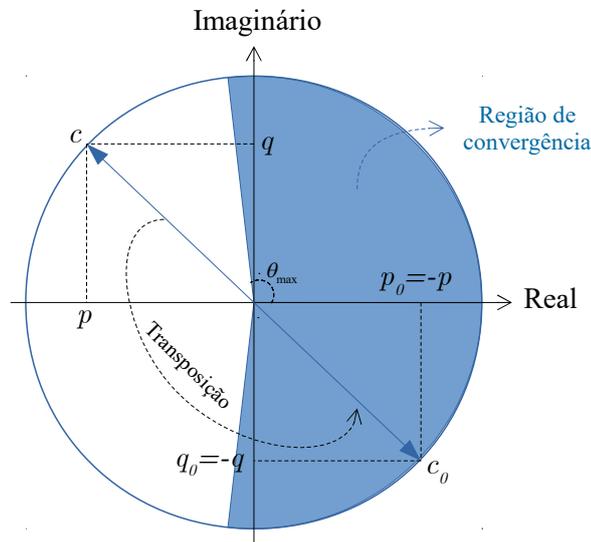


Figura 49 – Região de convergência das rotações CORDIC e processo de transposição de quadrantes.

que troca o sinal das partes real e imaginária de c . Isso corresponde a uma transposição dos 2º ou 3º quadrantes não-convergentes para os 4º ou 1º quadrantes convergentes, respectivamente. A partir de c_0 , fazem-se as rotações sucessivas. O processo inverso da transposição de quadrantes deve ser feito, quando o resultado da rotação final que gera c_R for obtido. A Figura 49 ilustra a região de convergência e o processo de transposição de quadrantes.

Considerando o objetivo inicial de calcular $|x| = \sqrt{p^2 + q^2}$, aplica-se o algoritmo CORDIC que resulta na conversão de coordenadas cartesianas para polares, de acordo com o **Algoritmo 6**. Este algoritmo tem como objetivo provocar rotações sucessivas de acordo com o sinal da parte imaginária, até que a mesma seja zerada projetando todo o ponto sobre o eixo real. Quando isto acontece, a magnitude do sinal complexo corresponde à parte real e sua fase ao somatório de ângulos rotacionados. Para a aplicação do DPD, a informação de fase é desnecessária e será descartada.

Algoritmo 6 Conversão de um número complexo de coordenadas cartesianas $p + jq$ para polares $Ae^{j\theta}$ com CORDIC.

Monte a tabela arco-tangente $\phi_i = \arctan(2^{-i})$ com $i = 0, 1, \dots, R - 1$

Calcule a constante K_R de acordo com (5.8)

Coordenadas cartesianas são dadas por

$$p = \Re\{x\}$$

$$q = \Im\{x\}$$

Transposição dos quadrantes de acordo com o sinal da parte imaginária

se $q < 0$ então

$$P_0 = -p$$

$$Q_0 = -q$$

senão

$$P_0 = p$$

$$Q_0 = q$$

fim se

$$Z = 0$$

para $i \leftarrow 0$ até $R - 1$ faça

▷ Rotações sucessivas

Defina a direção de rotação

se $Q_i < 0$ então

$$d_i = 1$$

senão

$$d_i = -1$$

fim se

Obter P_D a partir de P_i deslocado i bits para a direita

Obter Q_D a partir de Q_i deslocado i bits para a direita

$$P_{i+1} = P_i - d_i Q_D$$

$$Q_{i+1} = Q_i + d_i P_D$$

$$Z_{i+1} = Z_i - \phi_i$$

fim para

Escalonamento pela constante. $A = K_R P_R$

se $q < 0$ então

▷ Transpõe para o quadrante original

$$\theta = Z_R + \pi$$

senão

$$\theta = Z_R$$

fim se

Tabela 10 – Classe de algoritmos CORDIC para algumas funções trigonométricas.

Aplicação	Entrada	Inicialização	Critério de convergência	Resultado	Saída
Conversão Cartesiana para Polar	$x = p + jq$ $p = \Re\{x\}$ $q = \Im\{x\}$	$P_0 = p$ $Q_0 = q$ $Z_0 = 0$	Escolher d_i para $Q_{i+1} = 0$	$P_R = A$ $Q_R = 0$ $Z = \theta$	$x = Ae^{j\theta}$ $A = \sqrt{p^2 + q^2}$ $\theta = \arctan\left(\frac{q}{p}\right)$
Conversão Polar para Cartesiana	$x = Ae^{j\theta}$ $A = x $ $\theta = \angle x$	$P_0 = A$ $Q_0 = 0$ $Z_0 = 0$	Escolher d_i para $Z_{i+1} = \theta$	$P_R = p$ $Q_R = q$ $Z = \theta$	$x = p + jq$ $p = A \cos(\theta)$ $q = A \sin(\theta)$
Cosseno e Seno	Ângulo θ	$P_0 = 1$ $Q_0 = 0$ $Z_0 = 0$	Escolher d_i para $Z_{i+1} = \theta$	$P_R = p$ $Q_R = q$ $Z = \theta$	$p = \cos(\theta)$ $q = \sin(\theta)$

Para fins informativos, há, na verdade, uma classe de algoritmos CORDIC cujas lógicas são mostradas na Tabela 10. O algoritmo CORDIC é eficiente para ser executado em *hardware* pois são necessários apenas somas, deslocamento binário, uma tabela de arco-tangente e um escalonamento por uma constante.

5.2 Ganho complexo

O objetivo é implementar a equação de um dado modelo de sistema não-linear em tempo real e, portanto, o bloco do DPD. Este deve produzir um sinal de saída $y(n)$, a partir do sinal de entrada $x(n)$ e um conjunto de coeficientes \mathbf{h} por meio da equação do modelo, para cada instante de tempo discreto n . Todas as equações de modelos linear nos parâmetros, podem ser reescritas da seguinte forma

$$y(n) = f(|x(n)|, \mathbf{h})x(n) \quad (5.15)$$

sendo que $f(|x(n)|, \mathbf{h})$ fornece um número complexo em função da magnitude do sinal de entrada e do conjunto de coeficientes. Este número complexo por sua vez é multiplicado pelo próprio sinal de entrada x para cada instante de tempo n . Sendo assim, a função $f(|x(n)|, \mathbf{h})$ promove um novo valor de ganho de magnitude e deslocamento de fase para cada novo valor de magnitude do sinal de entrada, cujo comportamento é definido pelos valores dos coeficientes. Considere como exemplo, o modelo polinomial banda base de (2.20). Assim, tem-se

$$y(n) = \sum_{i=1}^K h_i |x(n)|^{2(i-1)} x(n) \quad (5.16)$$

$$y(n) = f(|x(n)|, \mathbf{h})x(n)$$

em que

$$f(|x(n)|, \mathbf{h}) = \sum_{i=1}^K h_i |x(n)|^{2(i-1)}. \quad (5.17)$$

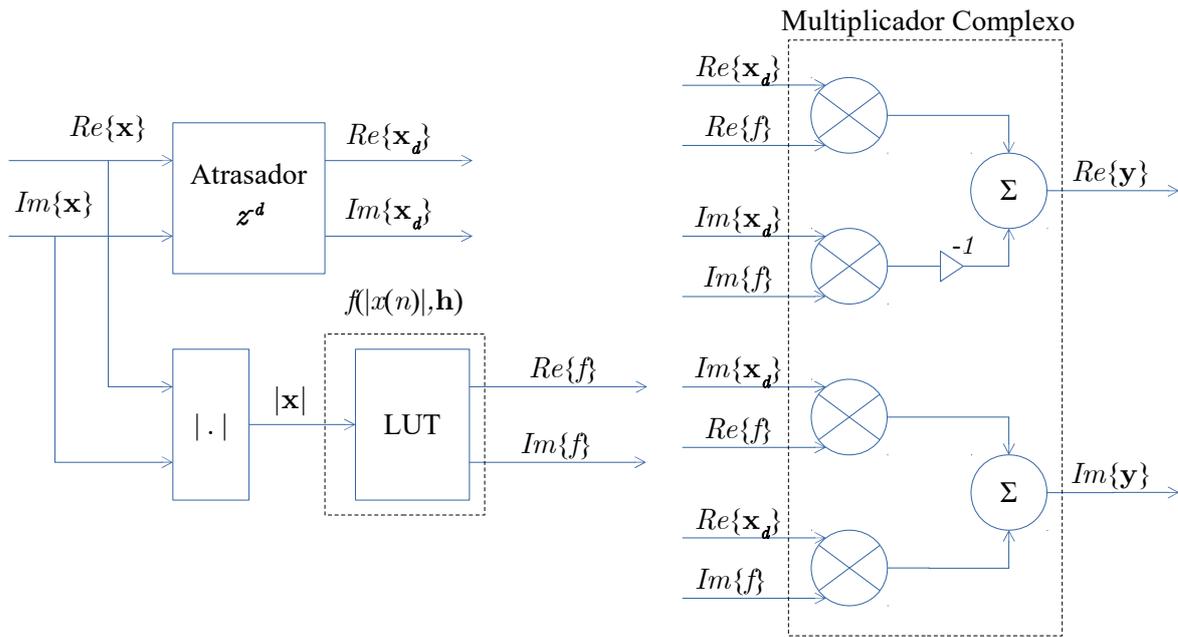


Figura 50 – Diagrama de blocos de um sistema DPD sem memória com ganho complexo implementado com LUT.

Sob a ótica do ganho complexo, $f(|x(n)|, \mathbf{h})$ pode ser implementada separadamente. De forma simplificada, a Figura 50 mostra um diagrama de blocos do sistema de ganho complexo, sendo que o bloco atrasador é necessário para compensar a latência dos blocos situados abaixo do mesmo. Há uma vantagem no fato do argumento $|x(n)|$ da função $f(|x(n)|, \mathbf{h})$ ser puramente real, resultando em uma função bidimensional. Esta função, por sua vez, é pré-calculada como resultado do processo de identificação e armazenada em uma tabela de consulta (*Look-Up Table* - LUT), sendo que o próprio sinal $|x(n)|$ é o endereço de leitura desta memória. Há uma grande vantagem nesta implementação, uma vez que o tamanho da memória é determinado pelo número de bits de quantização do sinal $|x(n)|$ que a endereça. O número de bits de endereço da LUT, N_A , é independente do número de bits do sinal a ser pré-distorcido, N_S . Quando $N_A < N_S$, haverá um mesmo endereço de LUT, e, portanto, um mesmo ganho complexo a ser provocado, para cada $2^{N_S - N_A}$ magnitudes do sinal de entrada. Isso determina a resolução com que as características do amplificador poderão ser modeladas. Nas próximas subseções são apresentados dois diferentes modos de se implementar um sistema baseado em LUTs a partir de $f(|x(n)|, \mathbf{h})$.

5.2.1 LUTs arranjadas por tomada

Para generalizar o conceito, utiliza-se, por exemplo, o polinômio com memória de (2.34). A equação do modelo usando a função de ganho complexo inicialmente dada por

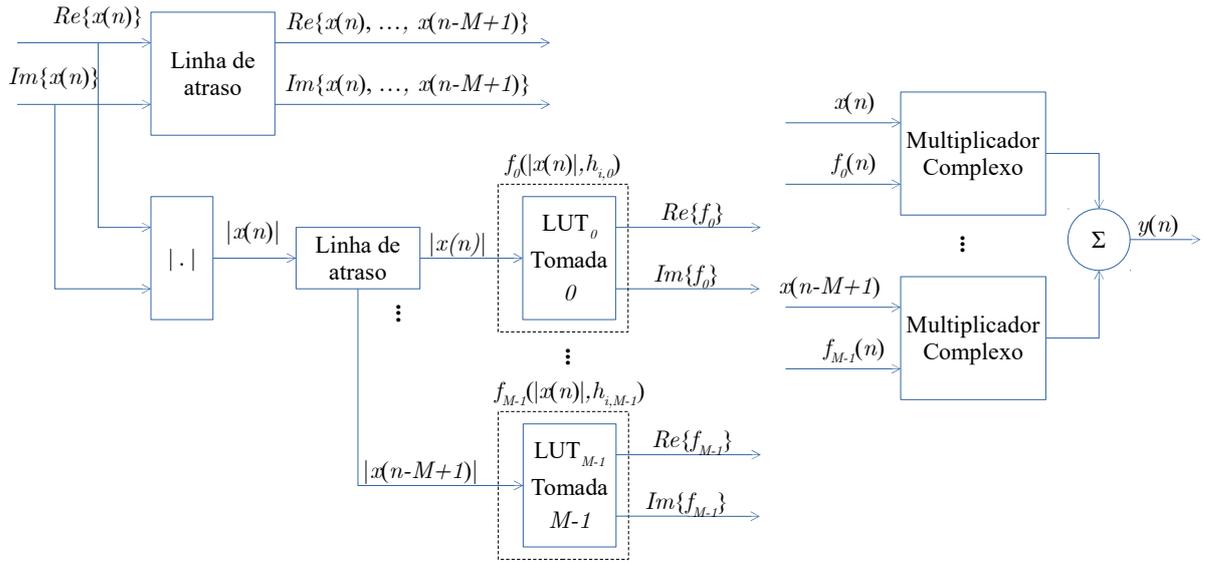


Figura 51 – Diagrama de blocos do DPD com LUTs para cada tomada de um sistema com memória.

(5.16) é redefinida como

$$y(n) = \sum_{m=0}^{M-1} f_m(|x(n-m)|, h_{i,m})x(n-m). \quad (5.18)$$

Nesta abordagem, os coeficientes $h_{i,m}$ serão incorporados dentro das LUTs que são calculadas de acordo com a equação

$$\text{LUT}_m = f_m(|x(n-m)|, h_{i,m}) = \sum_{i=1}^K h_{i,m} |x(n-m)|^{2(i-1)} \quad (5.19)$$

sendo que $|x(n)|$ assume 2^{N_A} valores linearmente distribuídos no intervalo $(0, \max(|x(n)|))$, sendo $\max(|x(n)|)$ o máximo valor que $|x(n)|$ pode assumir. Assim, haverá M LUTs complexas, uma para cada tomada. Logo, cada uma das M LUTs são iguais às suas respectivas funções de ganho complexo $f_m(\cdot)$. Como os coeficientes foram incorporados dentro da função que compõe a LUT, haverá apenas uma LUT para cada tomada, independente do número de polinômios K , com máxima ordem $2K - 1$. A Figura 51 ilustra o diagrama de blocos desse tipo de abordagem. O bloco linha de atraso da Figura 51 se difere do atrasador da Figura 50 pois existem M saídas, cada uma correspondendo a uma versão atrasada do sinal de entrada.

5.2.2 LUTs arranjadas por ordem

Outra abordagem de formação das LUTs é remover os coeficientes $h_{i,m}$ de sua equação. Desta forma, haverá K LUTs reais, uma para cada ordem, independente do número de tomadas, sendo que a saída de cada LUT é escalonada pelos coeficientes complexos

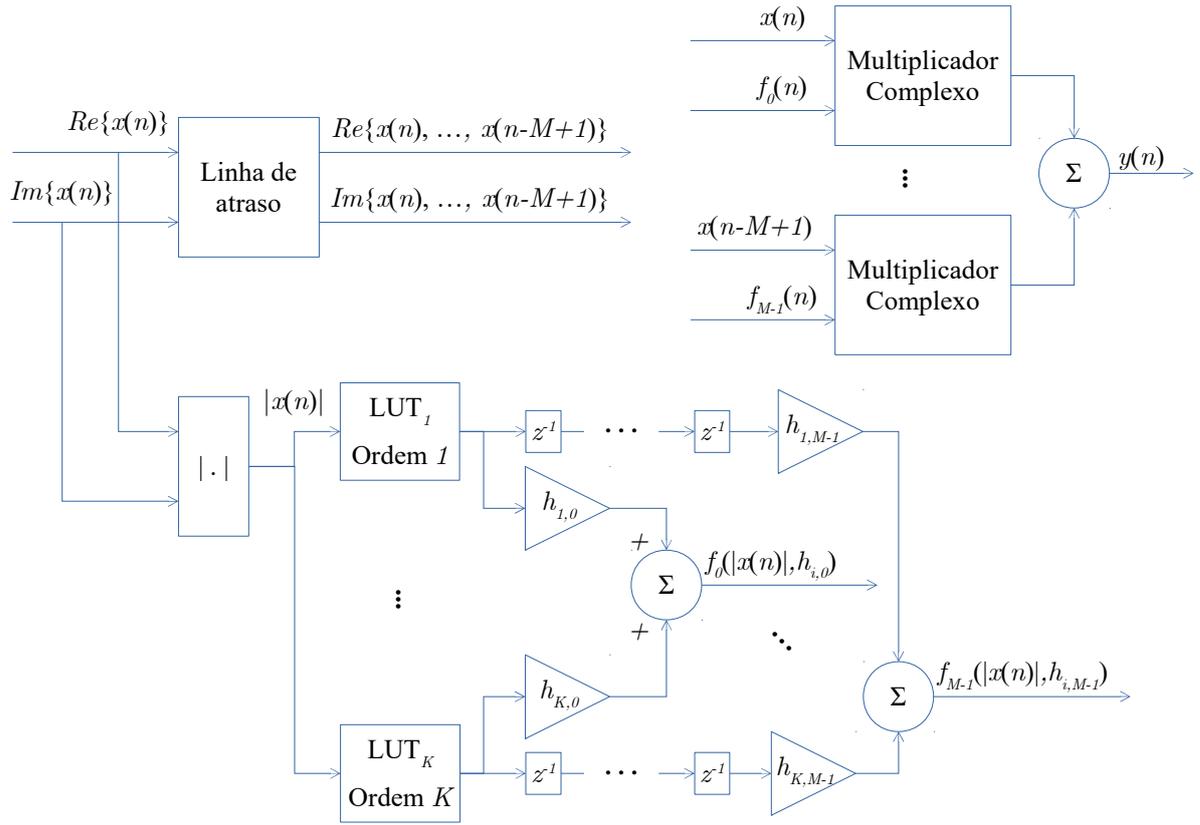


Figura 52 – Diagrama de blocos do DPD com LUTs para cada ordem de um sistema com memória.

$h_{i,m}$. Cada função de ganho complexo é dada por

$$f_m(|x(n)|, h_{i,m}) = \sum_{i=1}^K h_{i,m} \text{LUT}_i(n-m), \quad (5.20)$$

em que cada LUT referente a uma dada ordem polinomial é calculada de acordo com

$$\text{LUT}_i = f_i(|x(n)|) = |x(n)|^{2(i-1)}. \quad (5.21)$$

Observa-se uma diferença importante em relação ao caso anterior. Ao retirar os coeficientes $h_{i,m}$, os dados armazenados nas LUTs tornam-se reais. A Figura 52 ilustra o diagrama de blocos desse tipo de abordagem. Como as saídas das LUTs são sinais reais, a multiplicação pelos coeficientes corresponde à metade de um multiplicador complexo. Além do mais, como os coeficientes não estão incorporados nas LUTs, o conteúdo das mesmas é fixo, podendo ser implementadas com memórias de somente leitura (*Read Only Memory* - ROM). A aplicação da pré-distorção se resume em alterar o valor dos coeficientes nos multiplicadores. A faixa dinâmica do valor dos coeficientes traz um outro desafio considerando sua representação em ponto fixo e uma solução para esse problema está fora do escopo deste trabalho.

Nesse método de implementação de LUTs por ordem há uma vantagem em utilizar modelos como o Gaussiano de (2.38). As Figuras 53 e 54 ilustram esta vantagem. No

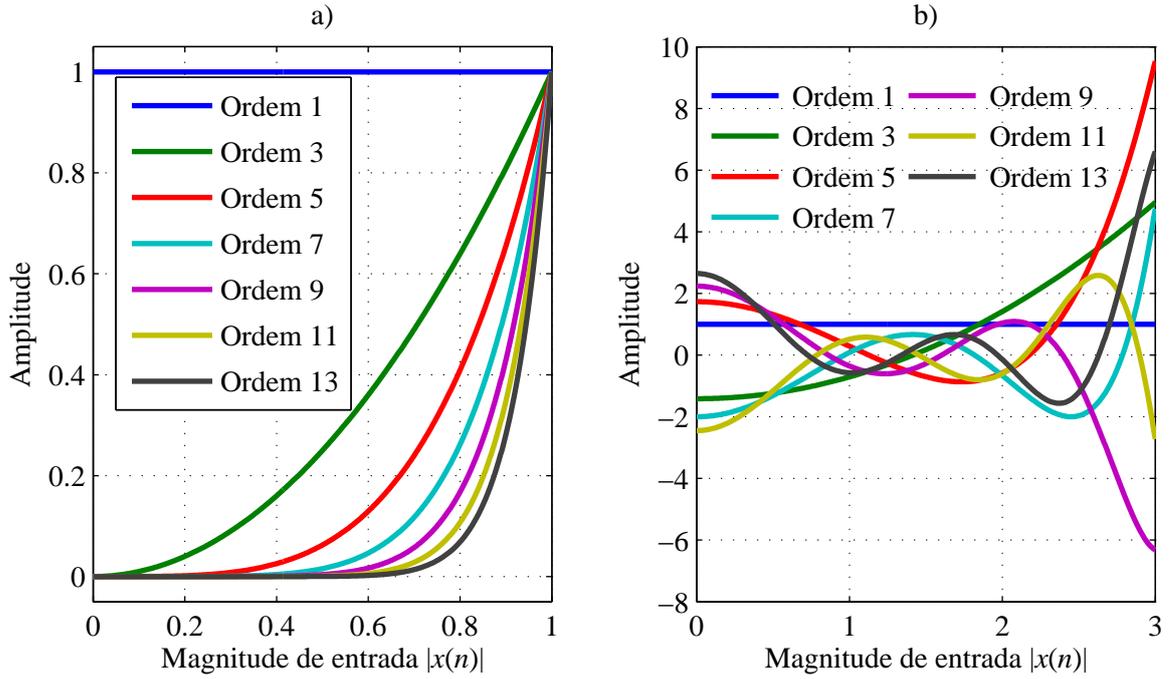


Figura 53 – LUTs com conteúdo puramente real para cada ordem de polinômio nos modelos a) banda base polinomial de (2.20) e b) Gaussiano ortogonal de (2.27).

Tabela 11 – Recursos lógicos para duas abordagens de implementação das LUTs.

Aplicação	LUTs por tomada	LUTs por ordens
Bits de Memória	$2^{N_A+1} N_D M$	$2^{N_A} N_D (K - 1)$
Multiplicadores	$4M$	$4M + 2K$
Somadores	$4M - 2$	$2K + 4M - 4$

modelo banda base polinomial convencional de (2.20), as LUTs apresentam valores muito próximos de zero, principalmente com o aumento da ordem. Ao quantizar as LUTs, ou seja, adotar um número de bits N_D para os dados em cada endereço da memória, obtém-se um erro de quantização que pode afetar o desempenho do DPD [70]. Por outro lado, em modelos ortogonais como o de (2.27), as LUTs para cada ordem de polinômio apresentam um formato que reduz o erro de quantização. O erro de quantização presente no modelo não-ortogonal poderia ser resolvido aplicando um quantizador não-linear. Por exemplo, usando uma função $\log(|x(n)|)$ no sinal de endereçamento das LUTs, o que aumenta a resolução para as baixas magnitudes. A vantagem mostrada anteriormente se refere a um sistema com quantização linear.

Fazendo uma comparação, a primeira abordagem incorpora os coeficientes nas LUTs, havendo uma LUT para cada uma das M tomadas, independente do número de ordens K . A segunda, deixa os coeficientes $h_{i,m}$ fora das LUTs e haverá uma LUT para cada das K ordens, independente do número de tomadas M . Para quantificar o uso de recursos lógicos, a Tabela 11 compara os dois métodos. Concluindo, para otimizar a utilização de recursos lógicos, a abordagem LUTs por tomada é interessante para sistemas

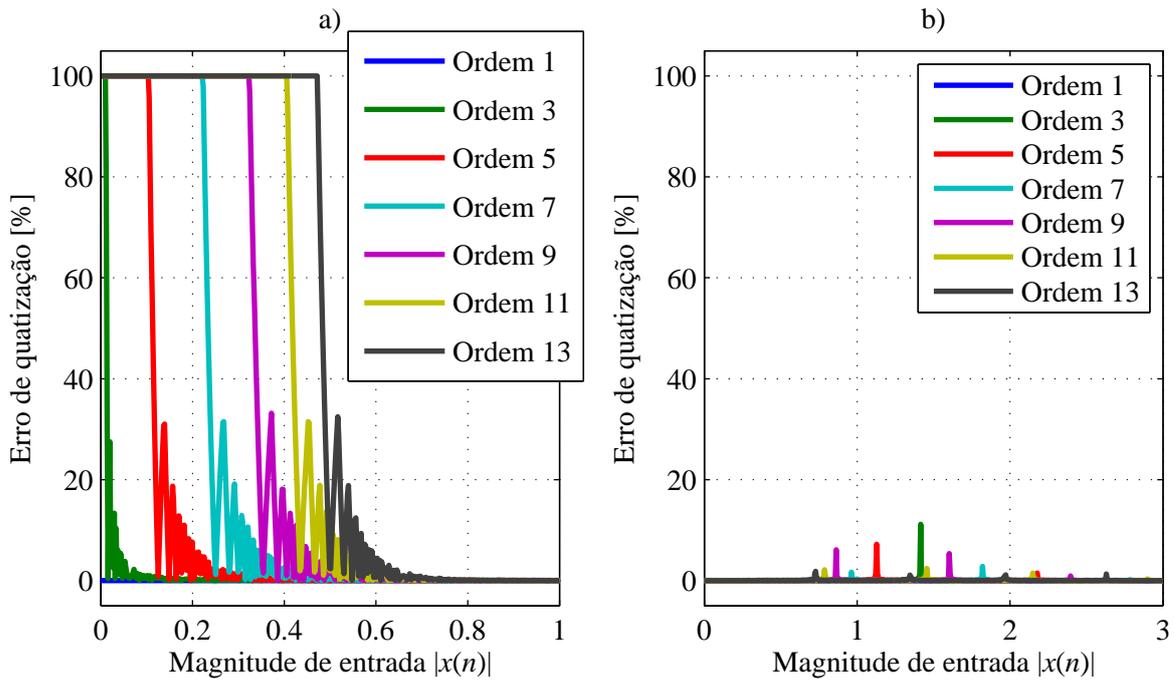


Figura 54 – Erro de quantização das LUTs de cada ordem para os modelos a) banda base polinomial de (2.20) e b) Gaussiano ortogonal de (2.27).

com poucas tomadas de ordem alta. Enquanto que a abordagem de LUTs por ordem é indicada para sistemas com muitas tomadas de ordem baixa.

5.3 Conclusão

No capítulo anterior, os coeficientes da pré-distorção foram calculados de acordo com o método com realimentação escalar proposto. Contudo, independente do algoritmo de DPD, é preciso aplicar a pré-distorção no sinal de entrada *em tempo real* a partir da equação do modelo adotada e do conjunto de coeficientes calculados. Este capítulo sintetiza as informações necessárias para obter o bloco de pré-distorção em tempo real. Esta é uma etapa importante na implementação do sistema como um todo.

6 Resultados experimentais

As simulações realizadas anteriormente correspondem a um primeiro passo no processo de validação do algoritmo OSFL proposto. A validação se dá por completa ao implementar tal sistema e testá-lo com um amplificador não-linear. Ao medir a redução da intermodulação na saída do amplificador devido ao uso da linearização, garante-se não somente que o modelo adotado é capaz representar o comportamento do amplificador, mas também que o algoritmo de ortogonalização funciona em casos práticos. Neste capítulo, serão apresentados os resultados obtidos a partir da implementação do método proposto. Foram empregados esforços nas áreas de *software*, *hardware* digital, eletrônica analógica e RF. Na parte digital usou-se uma FPGA híbrida com sistema em pastilha (*System on Chip* - SoC), contendo tanto lógica programável quanto um processador executando o sistema operacional Linux. No decorrer do capítulo, mais detalhes são expostos com ênfase na melhoria das características do sinal de RF, devido ao uso da linearização.

6.1 Projeto em FPGA SoC

Nesta subseção, será descrita a metodologia de desenvolvimento utilizada no projeto em FPGA SoC, ou seja, será apresentado o processo para desenvolver tanto a parte de lógica programável utilizando-se linguagem de descrição de *hardware*, quanto a parte de *software* na forma de uma aplicação Linux programada em linguagem C. A Figura 55 ilustra uma versão simplificada do diagrama de blocos do *hardware* implementado em FPGA, destacando as principais partes do sistema.

O bloco do DPD foi implementado de acordo com o diagrama de blocos da Figura 51, em um sistema com memória de três tomadas. Sua funcionalidade foi primeiramente simulada em MATLAB, em que foi gerado um arquivo contendo vetores de teste de entrada para fins de verificação e validação. O bloco inclui o conversor de coordenadas cartesianas para polares usando CORDIC (conforme mostrado no Capítulo 5), LUTs, atrasadores, multiplicadores complexos, entre outros.

Códigos de linguagem de descrição de *hardware* de circuitos integrados de alta velocidade (*Very high speed integrated circuits Hardware Description Language* - VHDL) foram escritos, compilados e simulados no *software* Modelsim, conforme ilustrado na Figura 56. Esta simulação do VHDL gera um arquivo com vetores de teste de saída, que são novamente analisados em MATLAB para serem comparados com os valores esperados para fins de validação.

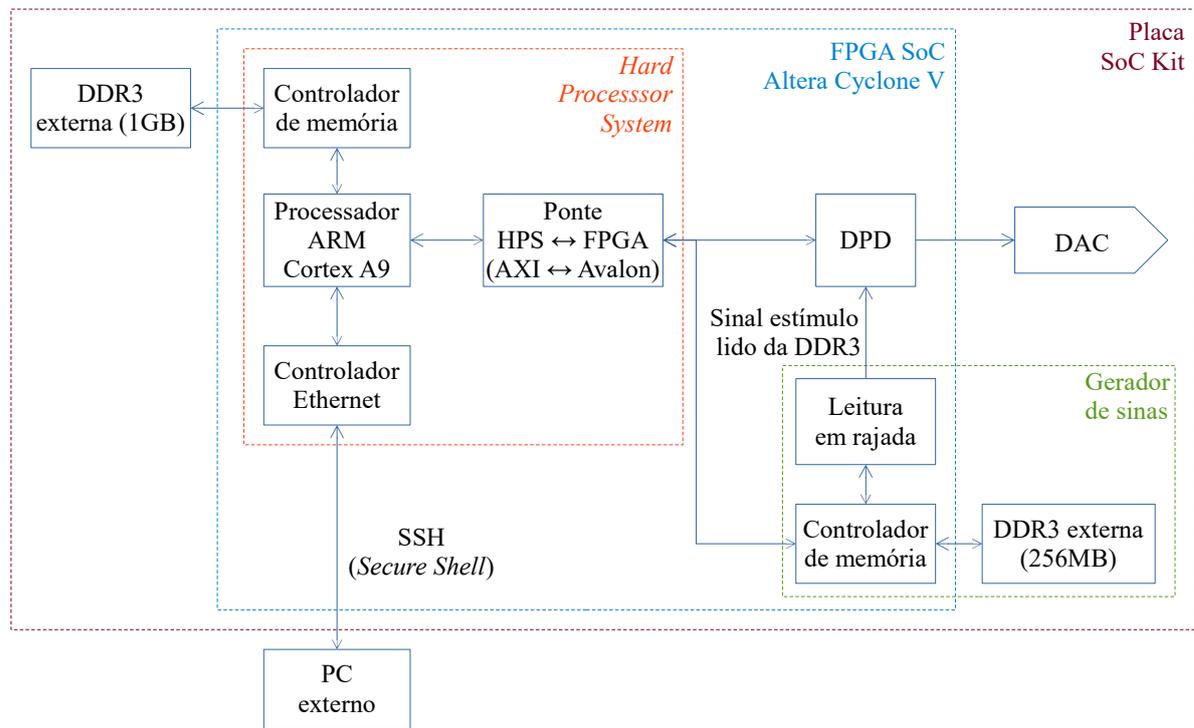


Figura 55 – Diagrama de blocos simplificado do *hardware* implementado em FPGA SoC.

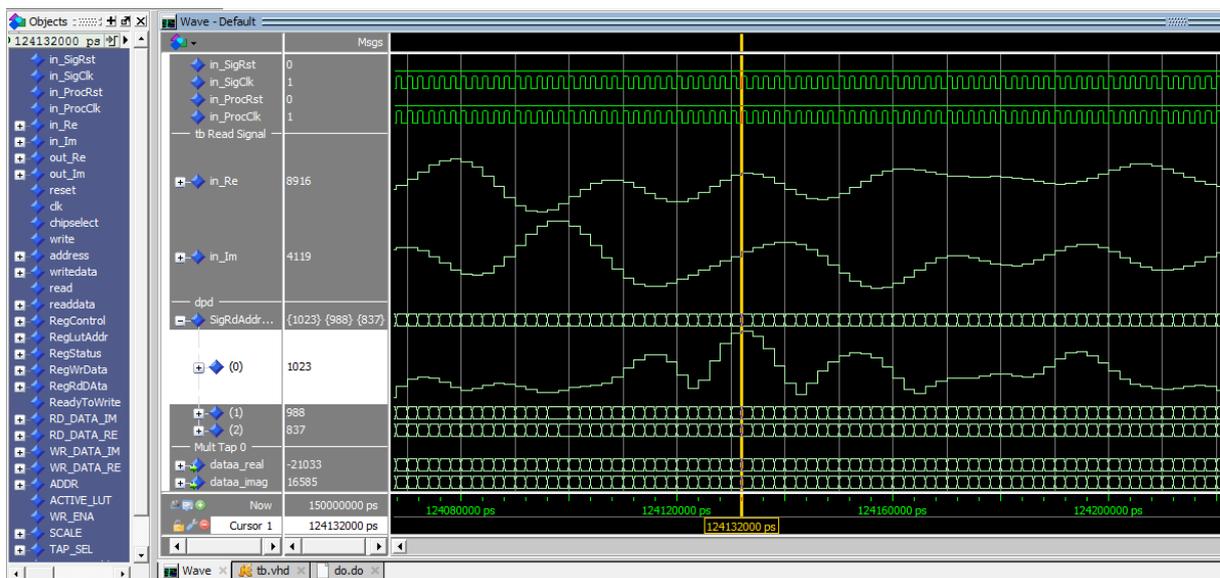


Figura 56 – Simulação do código VHDL do bloco DPD no Modelsim.

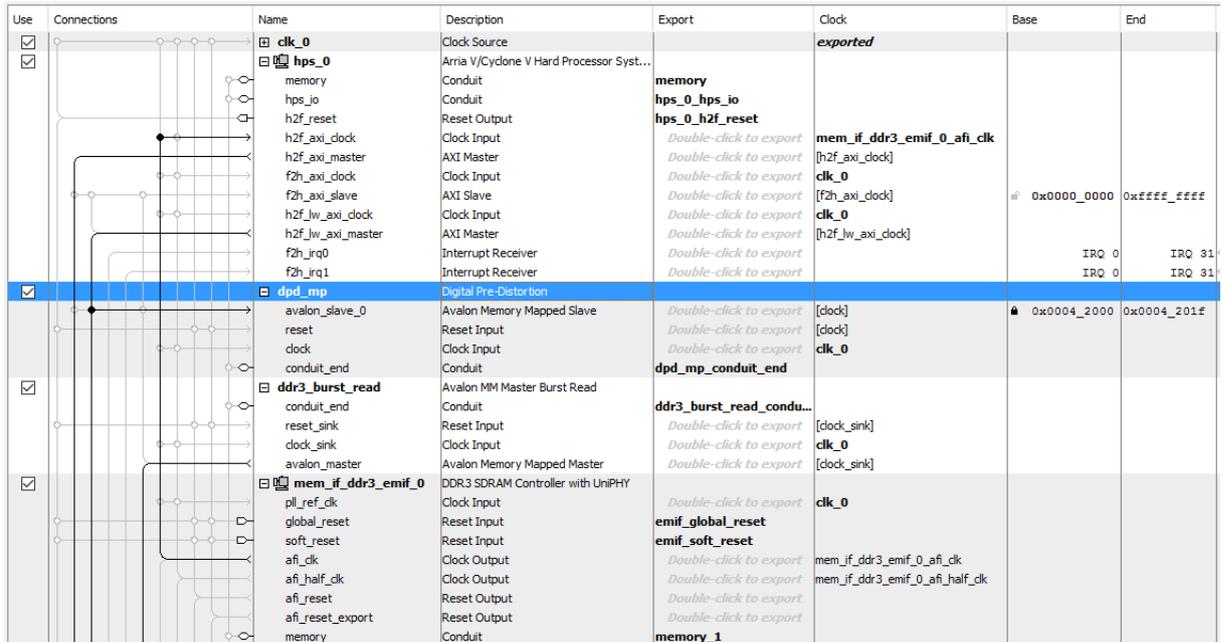


Figura 57 – Tela do Qsys onde é projetado o *hardware* do sistema embarcado SoC.

No módulo Qsys do *software* Quartus II, foi criado um projeto do *hardware* do sistema embarcado, formado por um processador ARM Cortex-A9 e um conjunto de periféricos já presentes na pastilha da FPGA. Há uma grande flexibilidade nesta arquitetura SoC, uma vez que se pode acrescentar periféricos ao processador desenvolvidos pelo próprio usuário. A Figura 57 mostra o diagrama de conexões entre os elementos do sistema com destaque para: i) *hps_0*: o processador e seus periféricos; ii) *dpd_mp*: o bloco do DPD como um periférico escravo do processador para permitir que uma aplicação do sistema operacional Linux possa escrever nas LUTs com base no valor dos coeficientes; iii) *ddr3_burst_read*: bloco também desenvolvido nesse trabalho para ler o conteúdo da memória de taxa dupla de dados versão 3 (*Double Data Rate 3* - DDR3) em rajada através de seu controlador (*mem_if_ddr_emif_0*), onde está armazenado as amostras do sinal de entrada que será reproduzido repetidamente. A Figura 58 mostra a planta baixa da pastilha da FPGA com o projeto compilado.

O algoritmo de linearização proposto com realimentação escalar e ortogonalização dos coeficientes foi desenvolvido em MATLAB, e necessita ser executado no ambiente de Linux embarcado. Como a linguagem *Mathscript* do MATLAB não pode ser interpretada no sistema embarcado, o algoritmo precisa ser convertido para linguagem C. Foi utilizada uma ferramenta chamada *MATLAB Coder*, usada para converter códigos *.m* em *.c* automaticamente. A Figura 59 ilustra esse processo de conversão.

O código C é compilado utilizando a ferramenta *Eclipse for DS-5*, sendo executado como uma aplicação do Linux. Neste ponto, pode-se verificar o desempenho de execução do *software* e se os resultados obtidos a partir do C compilado foram iguais aos do *Mathscript* interpretados no simulador.

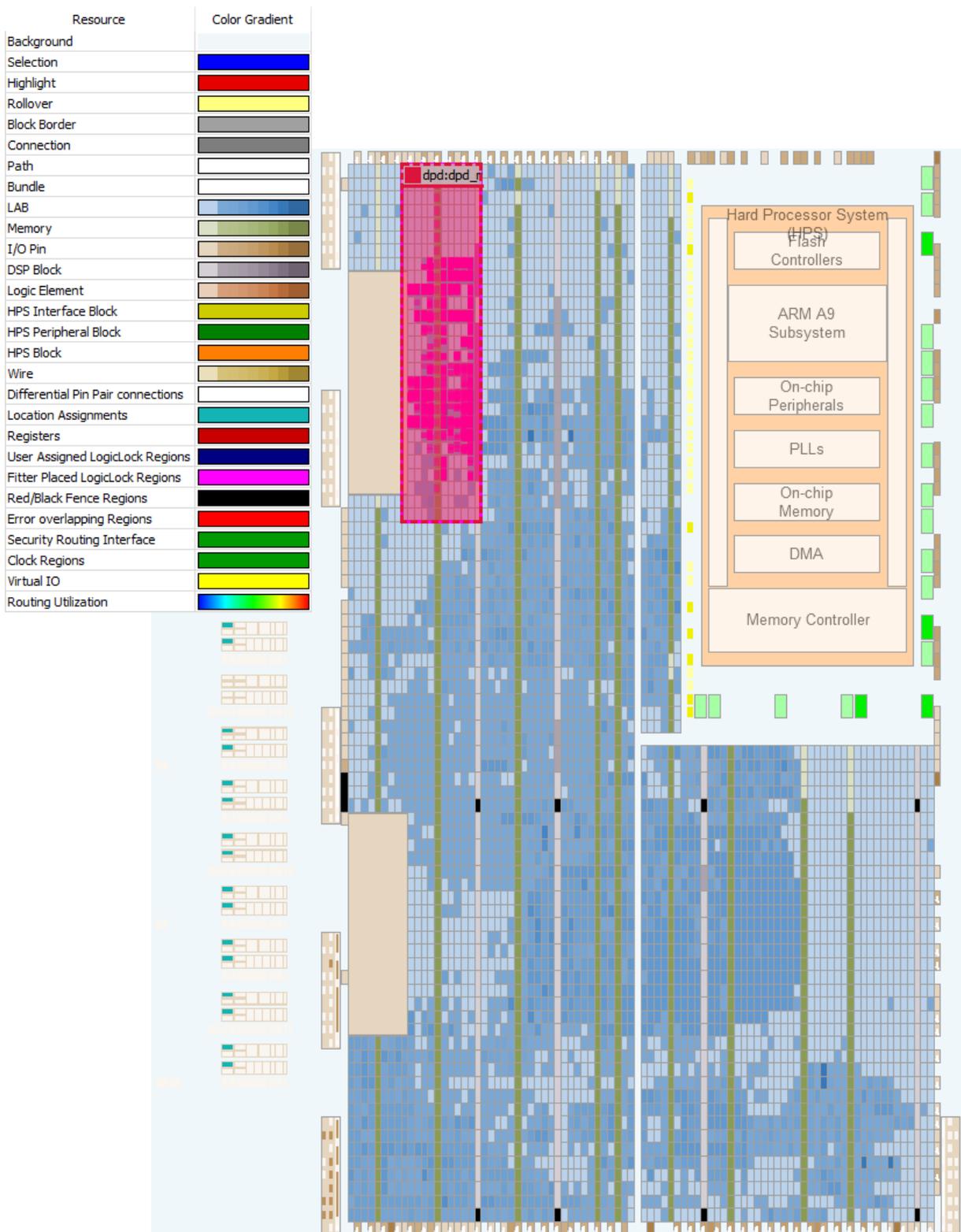
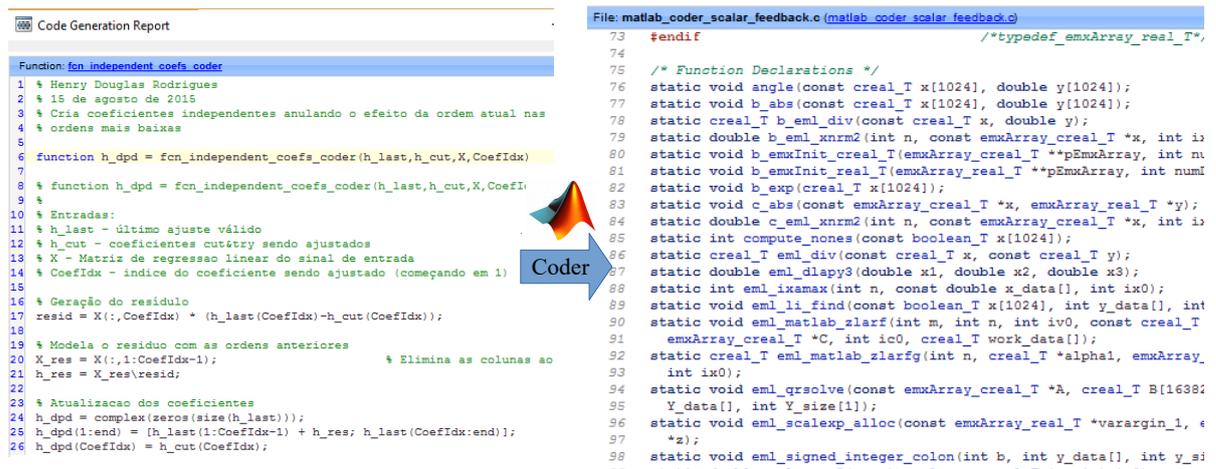


Figura 58 – Planta baixa da FPGA contendo o projeto compilado e bloco do DPD destacado.



```

Code Generation Report

Function: fcn_independent_coefs_coder
1 % Henry Douglas Rodrigues
2 % 15 de agosto de 2015
3 % Cria coeficientes independentes anulando o efeito da ordem atual nas
4 % ordens mais baixas
5
6 function h_dpd = fcn_independent_coefs_coder(h_last,h_cut,X,CoeffIdx)
7
8 % function h_dpd = fcn_independent_coefs_coder(h_last,h_cut,X,CoeffI
9 %
10 % Entradas:
11 % h_last - último ajuste válido
12 % h_cut - coeficientes cutítry sendo ajustados
13 % X - Matriz de regressão linear do sinal de entrada
14 % CoefIdx - índice do coeficiente sendo ajustado (começando em 1)
15
16 % Geração do resíduo
17 resid = X(:,CoefIdx) * (h_last(CoefIdx)-h_cut(CoefIdx));
18
19 % Modelar o resíduo com as ordens anteriores
20 X_res = X(:,1:CoefIdx-1); % Elimina as colunas ao
21 h_res = X_res/resid;
22
23 % Atualização dos coeficientes
24 h_dpd = complex(zeros(size(h_last)));
25 h_dpd(1:end) = [h_last(1:CoefIdx-1) + h_res; h_last(CoefIdx:end)];
26 h_dpd(CoefIdx) = h_cut(CoefIdx);

File: matlab_coder_scalar_feedback.c (matlab_coder_scalar_feedback.c)
73 #endif /*typedef_emxArray_real_T*/
74
75 /* Function Declarations */
76 static void angle(const creal_T x[1024], double y[1024]);
77 static void b_abs(const creal_T x[1024], double y[1024]);
78 static creal_T b_eml_div(const creal_T x, double y);
79 static double b_eml_xnrm2(int n, const emxArray_creal_T *x, int ix);
80 static void b_emxInit_creal_T(emxArray_creal_T **pEmxArray, int nu);
81 static void b_emxInit_real_T(emxArray_real_T **pEmxArray, int num);
82 static void b_exp(creal_T x[1024]);
83 static void c_abs(const emxArray_creal_T *x, emxArray_real_T *y);
84 static double c_eml_xnrm2(int n, const emxArray_creal_T *x, int ix);
85 static int compute_nones(const boolean_T x[1024]);
86 static creal_T eml_div(const creal_T x, const creal_T y);
87 static double eml_dlap3(double x1, double x2, double x3);
88 static int eml_ixamax(int n, const double x_data[], int ix0);
89 static void eml_li_find(const boolean_T x[1024], int y_data[], int
90 static void eml_matlab_zlarf(int m, int n, int i0, const creal_T
91 emxArray_creal_T *C, int ic0, creal_T work_data[]);
92 static creal_T eml_matlab_zlarfg(int n, creal_T *alpha, emxArray
93 int ix0);
94 static void eml_qrsolve(const emxArray_creal_T *A, creal_T B[16384
95 Y_data[], int Y_size[1]);
96 static void eml_scalexp_alloc(const emxArray_real_T *varargin_1, e
97 *z);
98 static void eml_signed_integer_colon(int b, int y_data[], int y_si

```

Figura 59 – Resultado da conversão do código de *Mathscript* para C.

Uma vez que o projeto no Quartus II foi compilado e carregado, o mesmo estará repetidamente lendo a memória DRR3, contendo um sinal Gaussiano complexo com amostras correlacionadas, quantizadas com 16 bits em cada uma das partes real e imaginária. Este sinal passa pelo bloco do DPD, cujas LUTs são escritas pela aplicação Linux de acordo com os coeficientes calculados. E, finalmente, o sinal passa pelo DAC que gera uma FI complexa. Para validar o funcionamento, são arbitrariamente experimentados valores não-nulos para os coeficientes do DPD. Observou-se, então, o crescimento espectral produzido no sinal de FI conforme mostrado na Figura 60.

6.2 Teste com amplificador de alta potência Doherty

Para melhor avaliar o desempenho do método proposto, um arranjo experimental foi implementado. A placa de processamento em tempo real foi a SoC *Kit* com uma FPGA *Altera Cyclone V System on Chip* [71]. O sinal de entrada foi gerado previamente em MATLAB e então carregado em uma memória de alta vazão DDR3, presente nesta placa. O sinal que corresponde a uma sequência de 65.536 amostras com distribuição Gaussiana correlacionada foi gerado uma única vez e reproduzido repetidamente em tempo real. O sinal complexo tem uma largura de faixa de 6 MHz centrado em uma FI de 12,5 MHz com uma taxa de amostragem de 50 MS/s, apresentando uma PAPR de 12,2 dB. O sinal complexo na saída dos DACs é colocado a nível de FI e não em banda base para que o vazamento de oscilador local (*Local Oscillator* - LO) e frequência imagem fiquem localizados fora da banda do canal. A Figura 61 ilustra os referidos vazamentos provenientes da etapa de conversão para RF. A FI foi convertida para 479 MHz na faixa de UHF.

O amplificador empregado na análise do DPD possui topologia *Doherty* de alta eficiência [72], e foi construído com o dispositivo NXP BLF888E operando em uma potên-

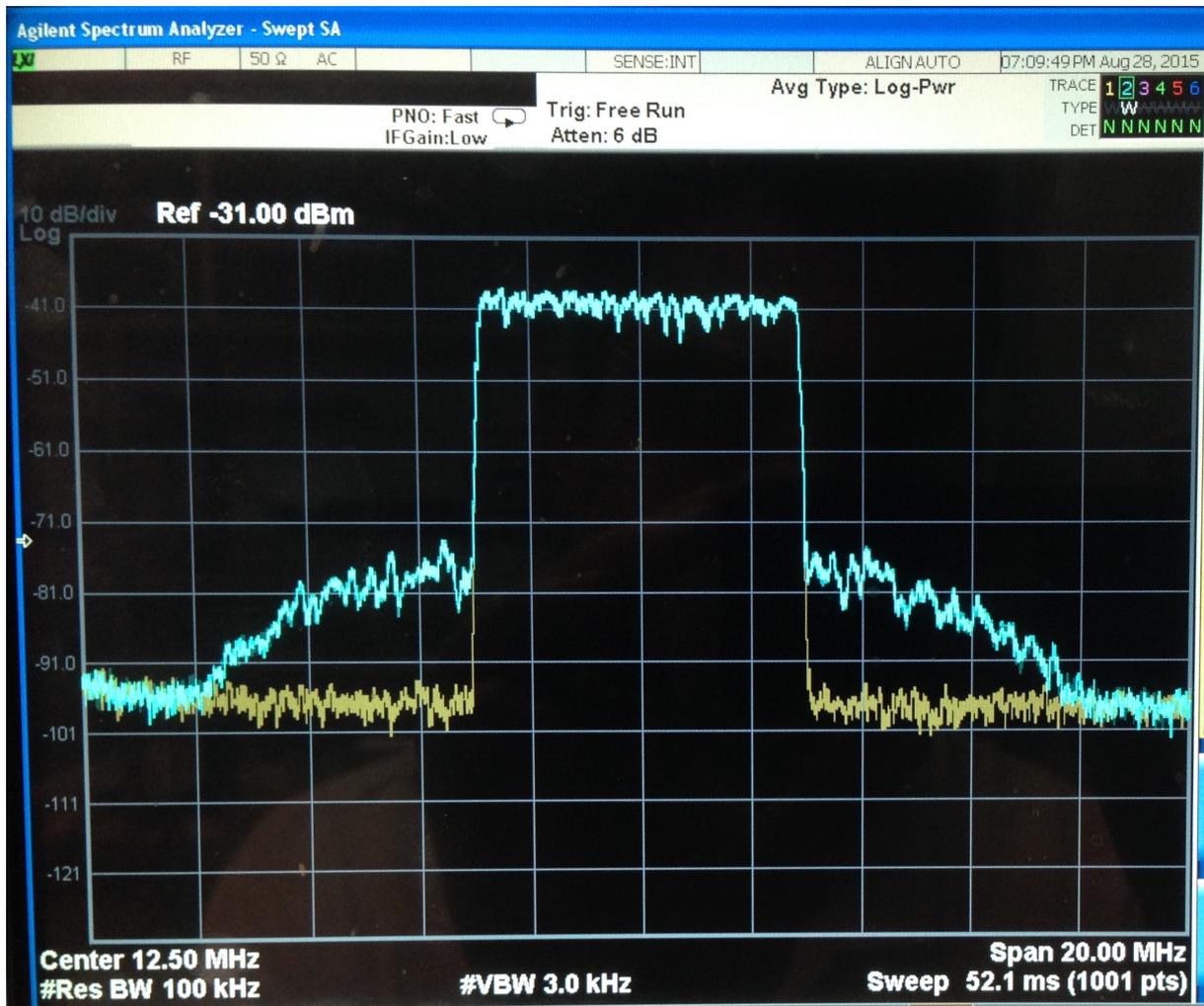


Figura 60 – Sinal de saída do DAC em FI e intermodulado pelo DPD para validar projeto em FPGA SoC.

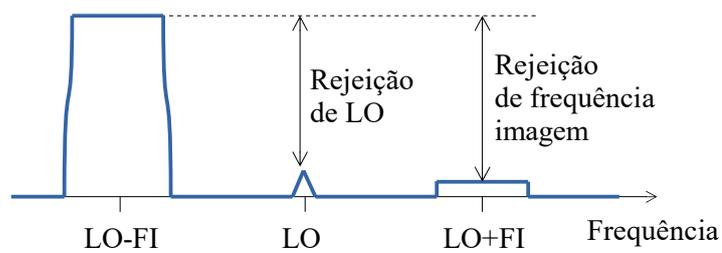


Figura 61 – Demonstração dos efeitos de vazamento LO de frequência imagem após conversão para RF utilizando banda lateral inferior.

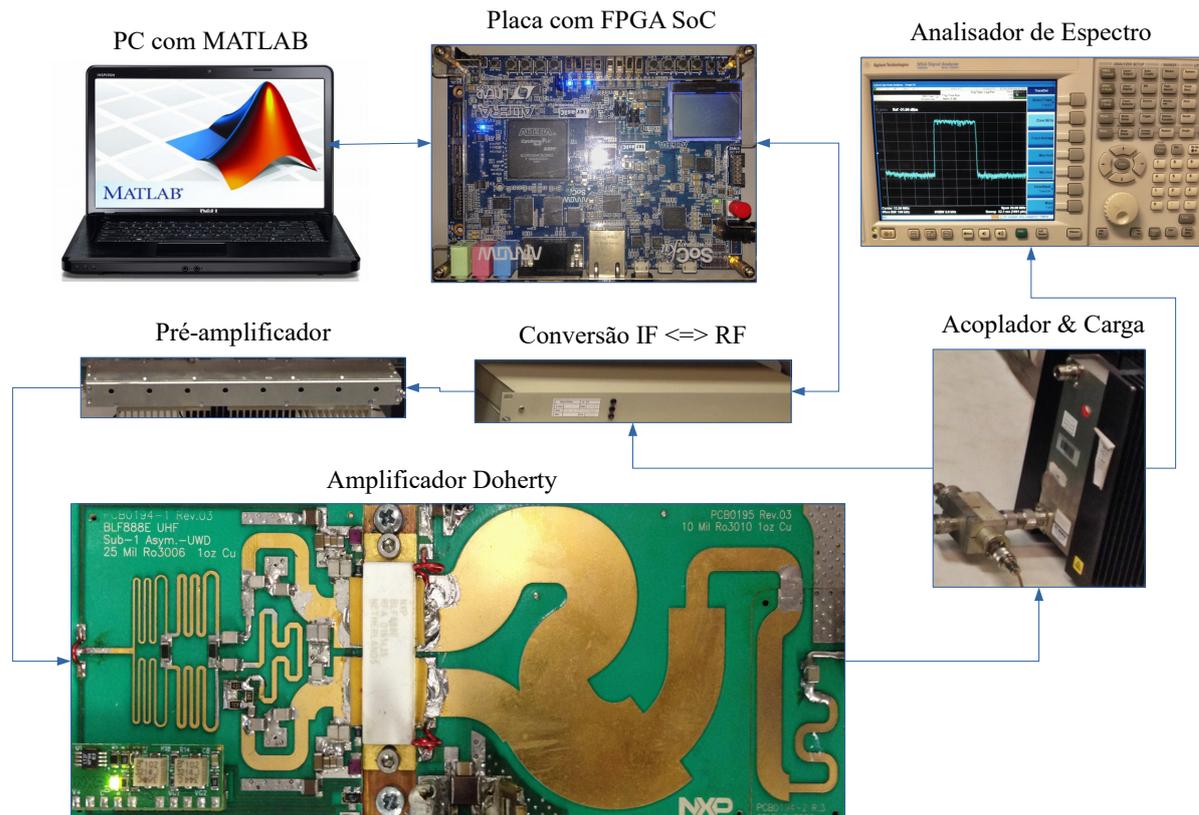


Figura 62 – Diagrama da bancada de testes em que o dispositivo BLF888E em configuração Doherty é linearizado com 100 W de potência média em sua saída.

cia de RF de 100 W de valor médio quadrático (*Root Mean Square* - RMS) na sua saída. Sua frequência de operação cobre a faixa de 470 MHz a 806 MHz. A teoria de operação do amplificador Doherty é baseada na modulação de impedâncias promovida pela combinação de dois amplificadores, chamados de principal e auxiliar que são polarizados em classe AB (ou B) e classe C, respectivamente. O amplificador auxiliar permanece desligado a maior parte do tempo, sendo ligado apenas durante os picos de potência instantâneos. Por esse motivo, o amplificador Doherty apresenta um nível significativo de intermodulação e é uma das topologias de PAs mais desafiadoras para se linearizar. Isso significa que os resultados obtidos em um amplificador Doherty podem ser repetidos em qualquer outra topologia, inclusive com expectativa de melhor desempenho.

Um diagrama com fotos da bancada de testes é mostrado na Figura 62. O sinal de FI complexo é conectado a um módulo de conversão para RF. Considerando o nível de saída do conversor FI/RF, é preciso usar um pré-amplificador intermediário, dado que o amplificador Doherty requer um nível de excitação superior ao fornecido pelo conversor FI/RF. Um acoplador direcional foi utilizado para tomar uma amostra do sinal antes de ser terminado em uma carga resistiva de 50 Ω . O analisador de espectro tomou a medida custo, medindo o nível de intermodulação por meio do ACLR.

A equação do modelo para o DPD é o polinômio com memória conforme (2.34).

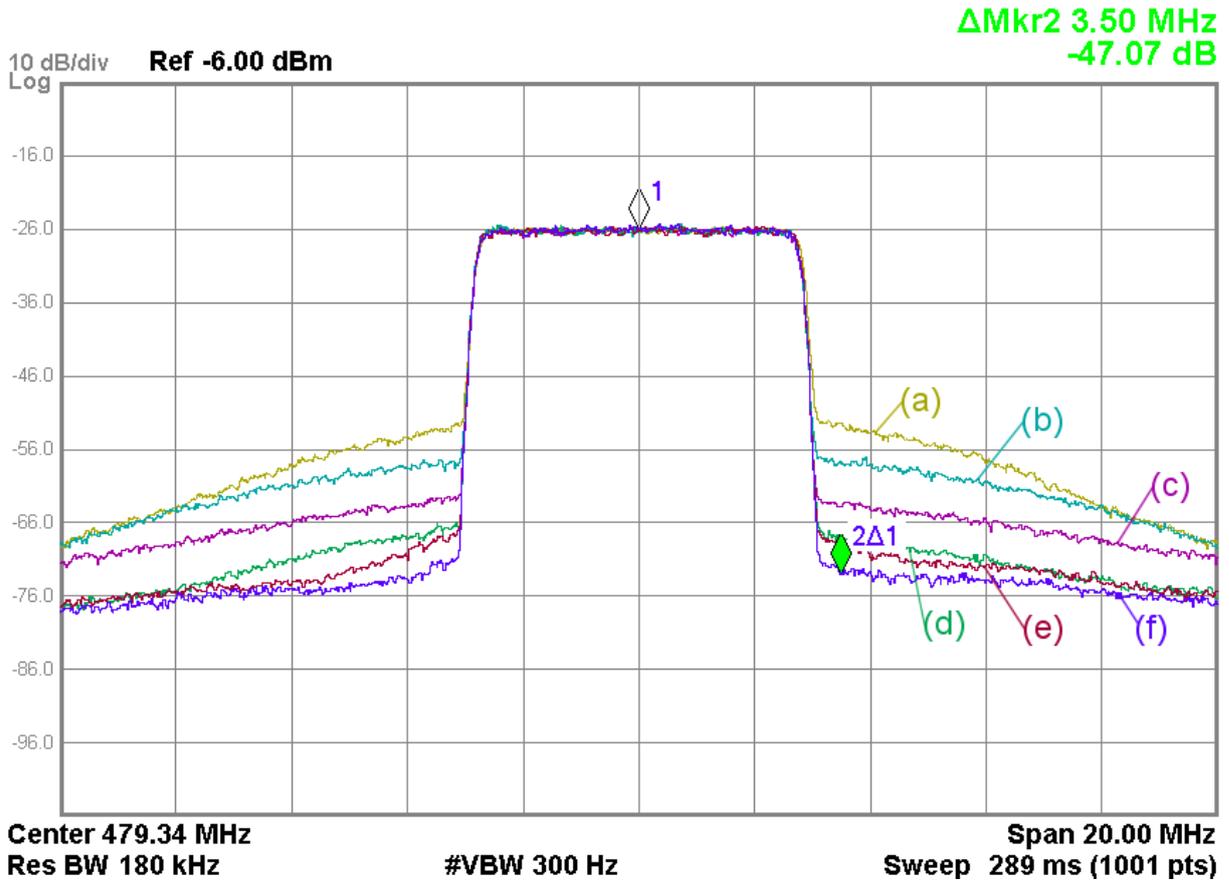


Figura 63 – Sinal de saída do PA Doherty: (a) sem DPD; (b)-(f) com DPD; (b) 3ª ordem da 1ª tomada; (c) 5ª e 7ª ordem da 1ª tomada; (d) ordens 9ª a 15ª da 1ª tomada; (e) 3ª e 5ª ordens das 2ª e 3ª tomadas; (f) reajuste de todos os coeficientes para aproximar-se do mínimo global.

Os parâmetros são $M = 3$, $K_1 = 8$, $K_2 = 3$ e $K_3 = 2$, correspondendo à 15ª, 5ª e 3ª ordem para as três tomadas, respectivamente. A Figura 63 mostra o resultado da linearização à medida em que os coeficientes foram independentemente ajustados de forma ortogonal. Pode ser notado a contribuição de cada coeficiente, ou grupo de coeficientes na redução da intermodulação. Observa-se que a melhoria da Figura 63(e) para a Figura 63(f) foi obtida ao repetir todo o processo de ajuste, recomeçando a partir do coeficiente de 3ª ordem da 1ª tomada.

Uma comparação de desempenho foi feita com o método de referência LS [32] e a Figura 64 mostra os resultados. Como pode ser visto, o desempenho é o mesmo em frequências próximas dos limites do canal. Para frequências mais distantes do centro do canal, o algoritmo proposto foi levemente superado. Como o resultado do desempenho é praticamente o mesmo para esse nível de intermodulação, o método proposto provou ser simples e tão eficiente quanto a abordagem convencional.

A função distribuição cumulativa complementar (*Complementary Cumulative Distribution Function* - CCDF) da relação entre a potência instantânea e média é uma medida

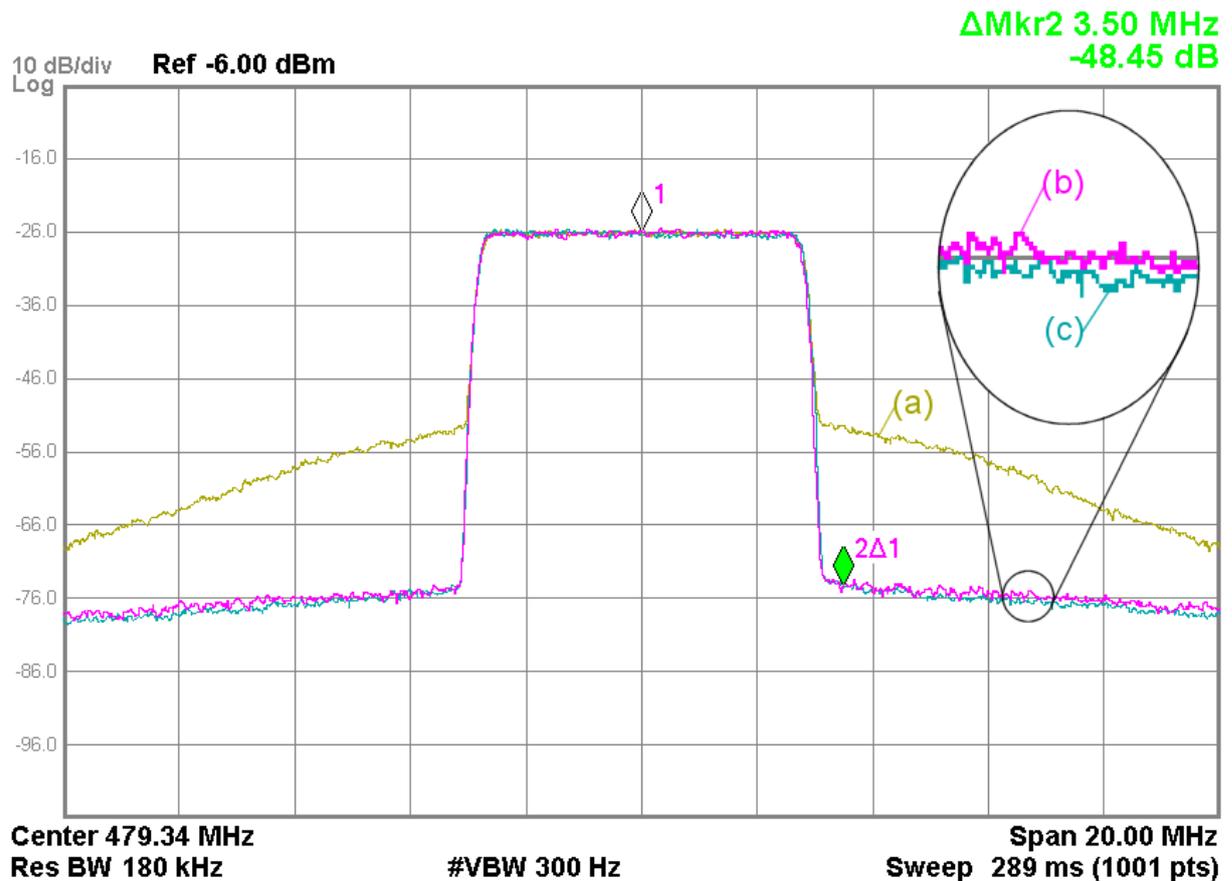


Figura 64 – Comparação entre métodos de regressão linear e o OSFL proposto: (a) sem DPD; (b) com DPD com o método OSFL; (c) DPD com o método LS [32].

que mostra a probabilidade da potência instantânea apresentar um valor acima da abscissa. A Figura 65 mostra várias curvas de CCDF do sinal na saída do PA sem a influência do DPD para níveis de potência distintos. Nota-se que a medida em que a potência sobe, o amplificador comprime uma parcela maior do sinal, fazendo com que a probabilidade de haver picos de potência acima da potência média seja reduzida. Tomando a CCDF de uma distribuição Gaussiana (que corresponde ao sinal ideal) como referência, os níveis de potência mais baixos, de 5 W e 20 W, sofreram expansão de magnitude dos picos. Tal comportamento é comum em amplificadores classe AB e Doherty. Isso significa que mesmo fazendo um grande recuo de potência, o uso de algum tipo de técnica de linearização ainda é necessário para combater a intermodulação [73]. Do ponto de vista do espectro do sinal, a Figura 66 mostra o efeito do aumento da potência na intermodulação do sinal. Neste gráfico, também é possível verificar que, para baixos valores de potência, o sinal já é distorcido pelo PA, em função do efeito da expansão.

A Figura 67 mostra a CCDF da relação entre a potência instantânea e média para os sinais de entrada e saída do amplificador em duas condições distintas: com e sem o DPD. Pode-se notar a influência do amplificador ao longo dos vários níveis de amplitude do sinal. A entrada do PA sem o uso de DPD corresponde ao sinal de en-

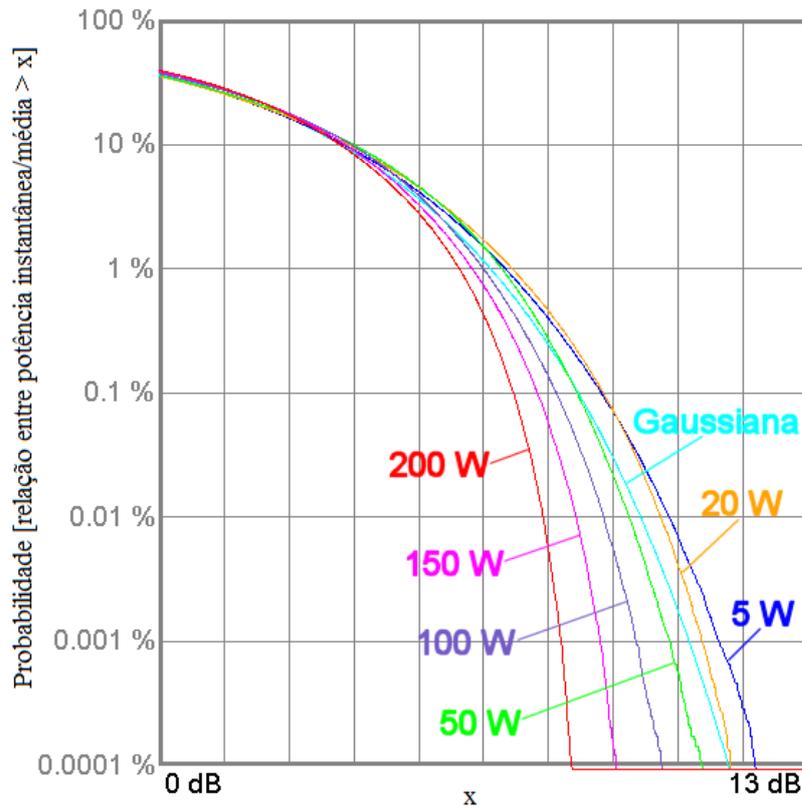


Figura 65 – Curvas CCDF para o sinal de saída do PA para níveis de potência distintos: 5 W, 20 W, 50 W, 100 W, 150 W e 200 W. A curva Gaussiana corresponde ao sinal ideal sem distorção.

Tabela 12 – Valores da CCDF da relação de potência pico/média para o sistema de DPD desligado e ligado.

Probabilidade (Relação de potência instantânea/média) > x	x [dB]	
	Sem DPD	Com DPD
10 [%]	3,63	3,4
1 [%]	6,63	6,4
0,1 [%]	8,37	8,47
0,01 [%]	9,65	10
0,001 [%]	10,57	11,08
0,0001 [%]	11,13	11,69

trada com distribuição Gaussiana. Ao analisar a saída do PA, ainda sem o uso de DPD, nota-se claramente o efeito provocado pelo PA de expandir as amplitudes mais baixas e comprimir as altas. Observando o sinal na entrada do PA com o uso de DPD, nota-se a ação contrária às distorções provocadas pelo PA: compressão nas baixas amplitudes e expansão nas altas. Finalmente, na saída do PA com DPD, pode-se observar que o CCDF novamente se aproxima da Gaussiana. A Tabela 12 apresenta os valores da curva CCDF para diferentes probabilidades (ou porcentagem do tempo) medindo a influência do DPD. A Figura 68 mostra o sinal de saída do PA ao aplicar o DPD para os diferentes níveis

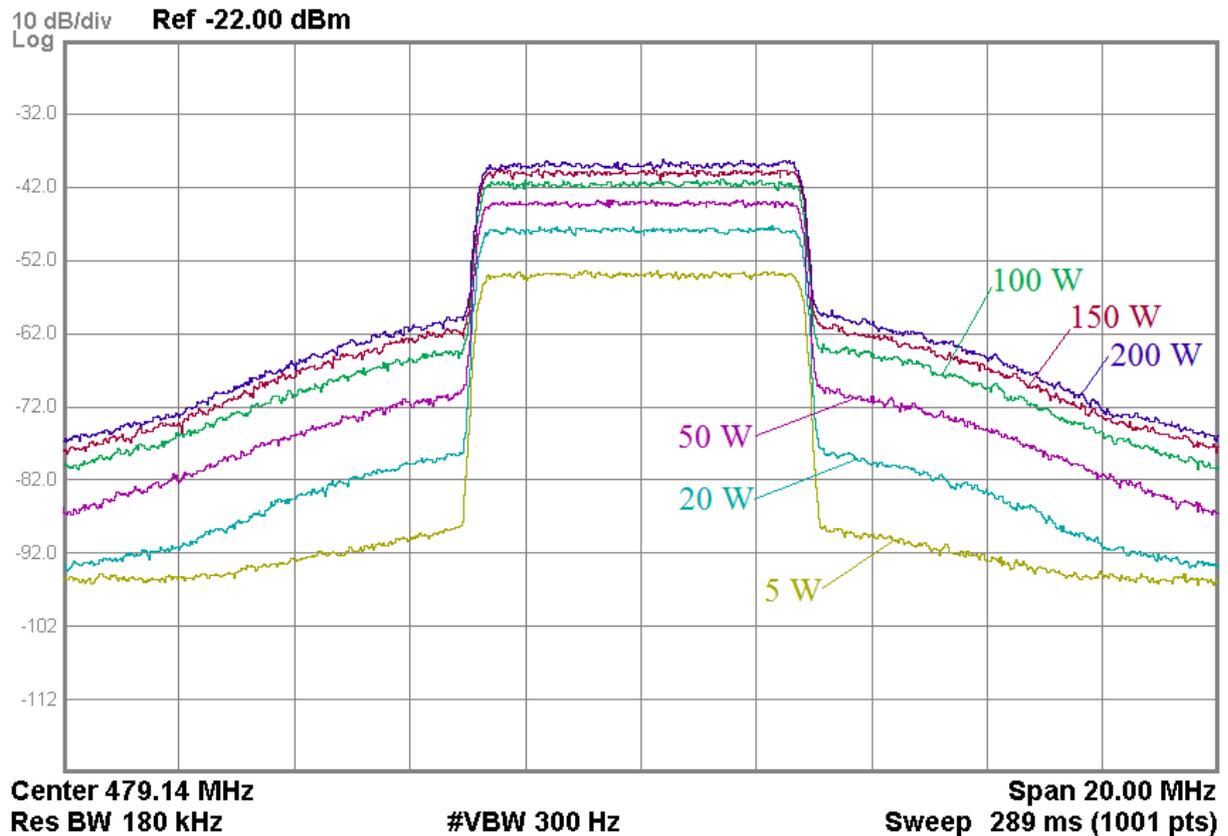


Figura 66 – Saída do amplificador com DPD desligado para vários níveis de potência: 5 W, 20 W, 50 W, 100 W, 150 W e 200 W.

de potência. Observa-se que a redução na intermodulação é menor à medida em que a potência é aumentada. Isso ocorre porque existe um limite físico de potência máxima que o PA consegue fornecer. Quando a potência instantânea do sinal ultrapassar esse limite, ocorrerá uma saturação do amplificador, resultando no ceifamento do sinal. Tal efeito não é corrigível com o uso de linearização, o que explica a perda de desempenho com o aumento da potência. Neste caso, o desempenho só pode ser melhorado com técnicas de redução de PAPR [74].

A linearização reduz a intermodulação tanto fora quanto dentro do canal. Uma forma de analisar a degradação causada pela intermodulação dentro do canal é medir a qualidade da constelação demodulada. As Figuras 69 e 70 mostram a constelação do sinal na saída do PA com o DPD desligado e ligado, respectivamente. A Tabela 13 mostra os parâmetros de qualidade do sinal para o sistema de DPD desligado e ligado. Dentre tais parâmetros estão a MER e a magnitude do vetor de erro (*Error Vector Magnitude* - EVM). Nota-se uma melhoria de desempenho, com a redução do EVM e aumento da MER, devido ao uso da linearização com DPD.

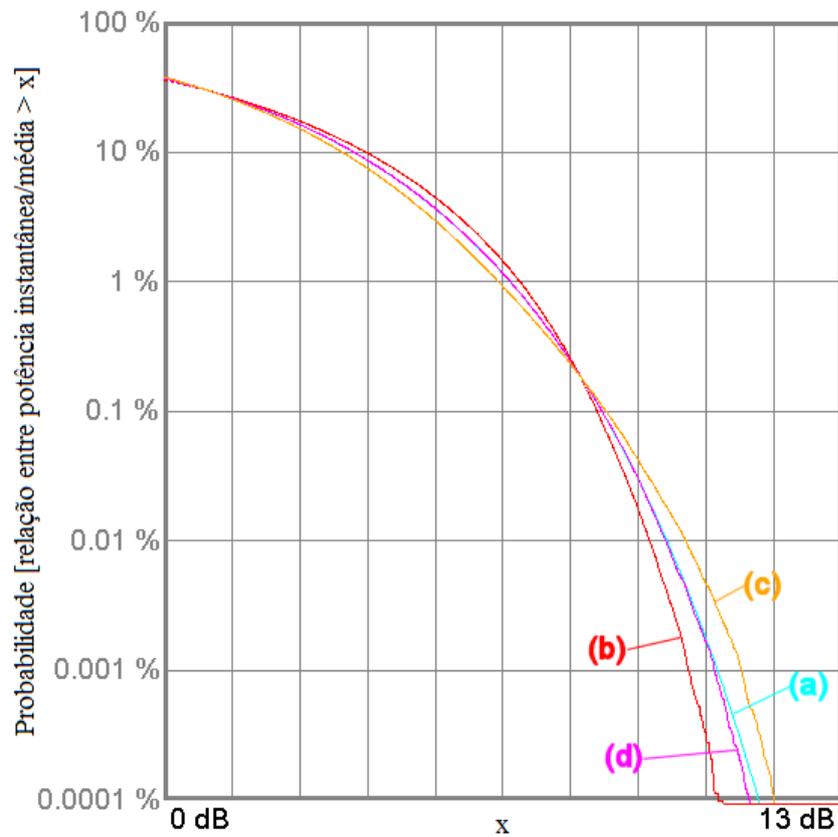


Figura 67 – CCDF em diferentes condições e pontos do sistema: (a) entrada do PA sem DPD (Gaussiana de referência); (b) saída do PA sem DPD; (c) entrada do PA com DPD; (d) saída do PA com DPD.

Tabela 13 – Comparativo da qualidade do sinal com o uso de DPD em termos de EVM e MER.

Parâmetro	Sem DPD	Com DPD
EVM [%]	9,24	0,95
EVM pico [%]	47,38	5,1
MER [dB]	20,68	40,44
MER pico [dB]	6,49	25,85
Erro de magnitude [%]	6,92	0,7
Erro de magnitude pico [%]	47,93	3,97

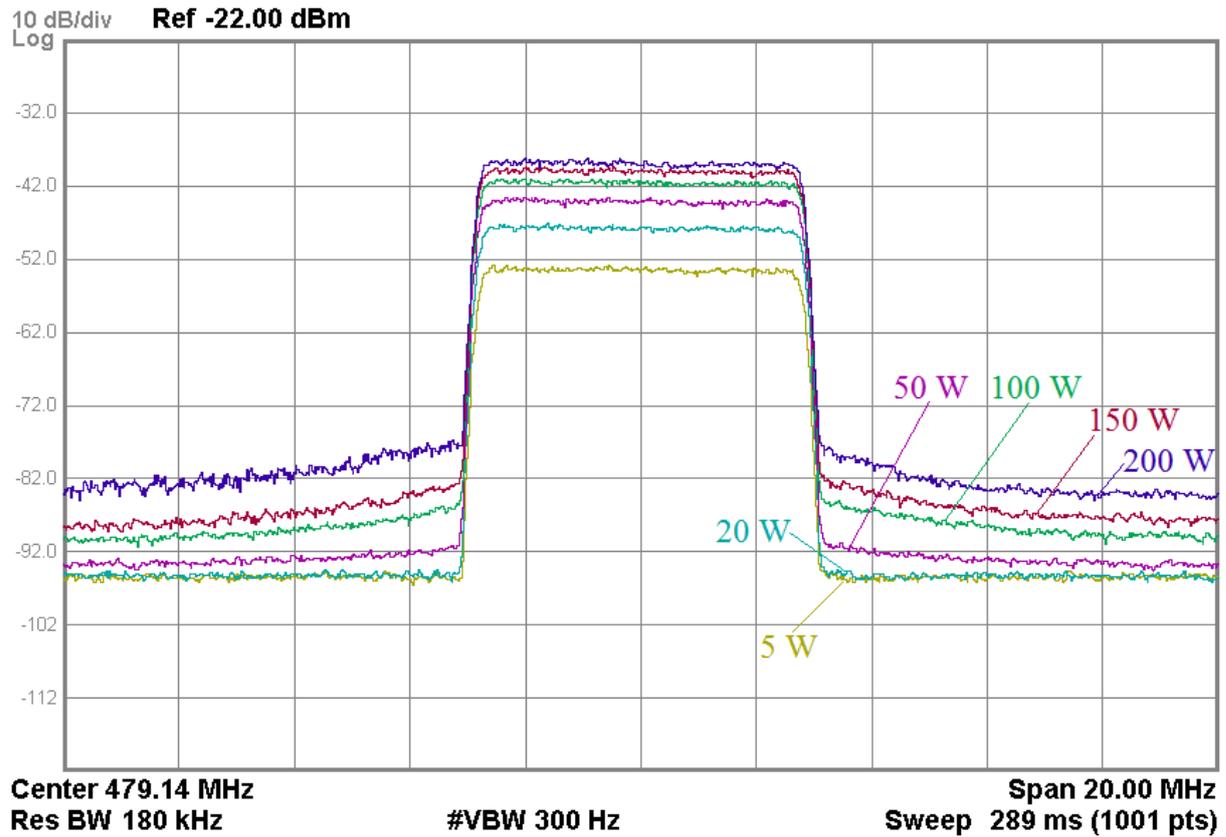


Figura 68 – Saída do amplificador com DPD ligado para vários níveis de potência: 5 W, 20 W, 50 W, 100 W, 150 W e 200 W.

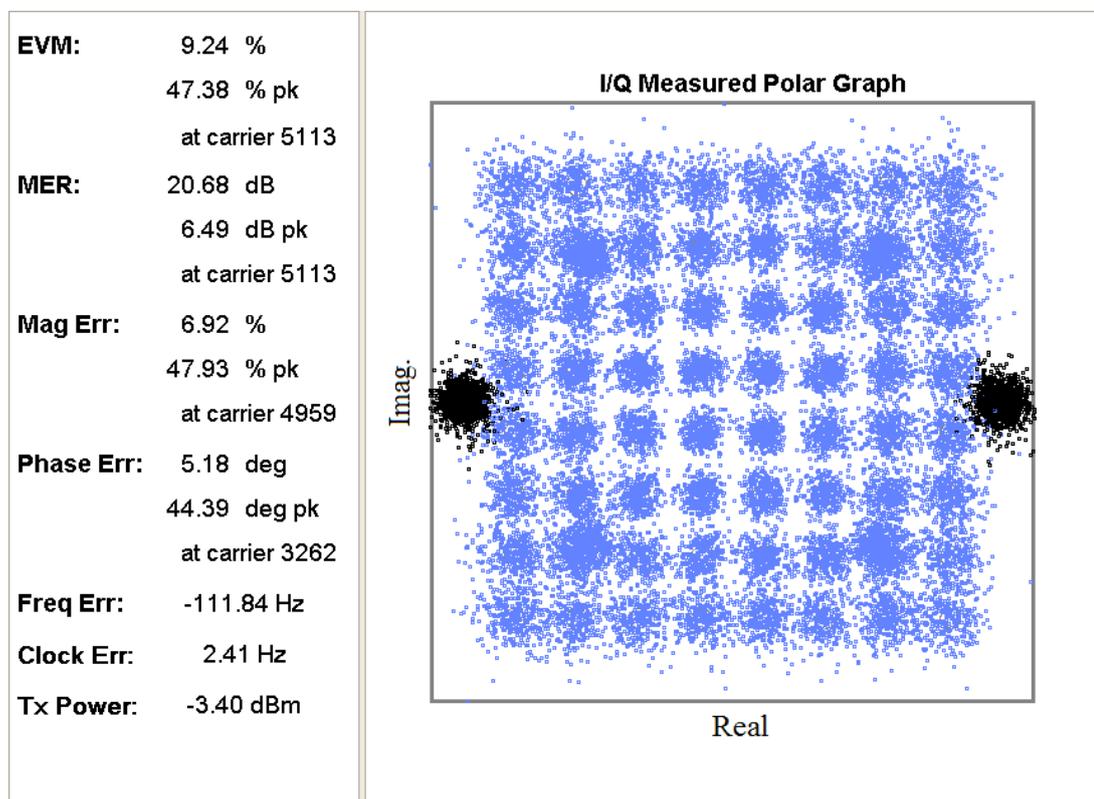


Figura 69 – Constelação de um sinal OFDM na saída do PA com sistema de DPD desligado para 100 W de saída.

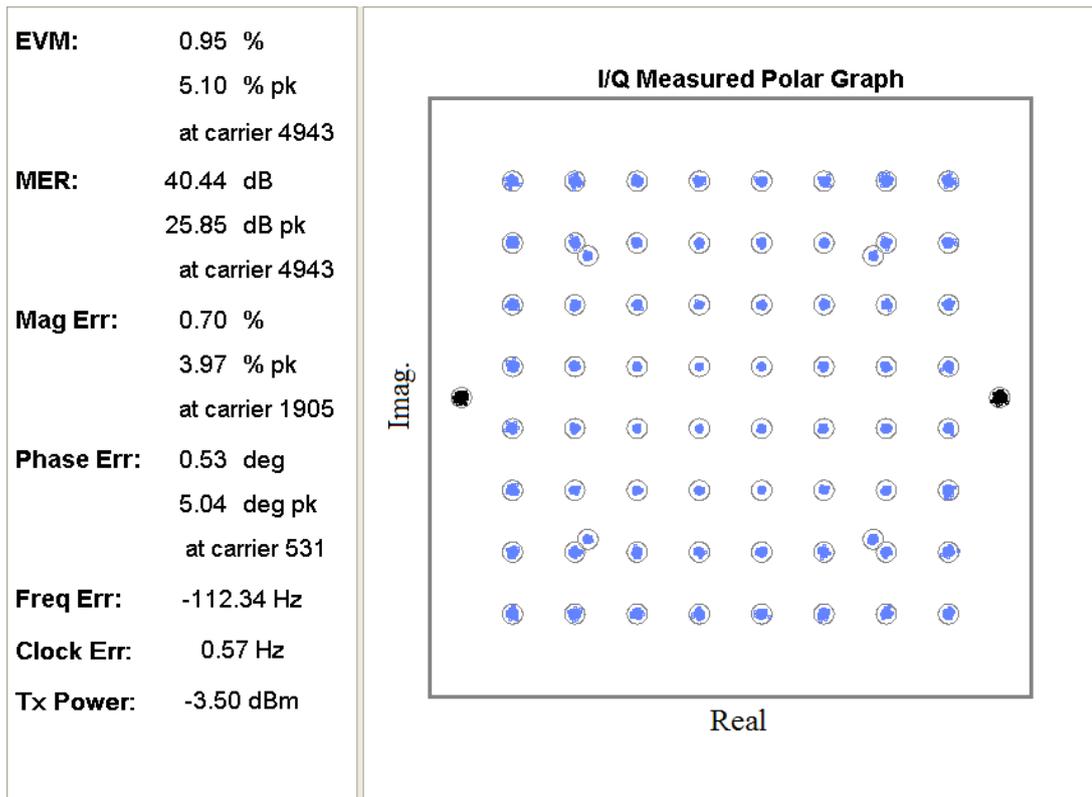


Figura 70 – Constelação de um sinal OFDM na saída do PA com sistema de DPD ligado para 100 W de saída.

6.3 Conclusão

Esse capítulo discorreu sobre a metodologia de desenvolvimento e apresentou os resultados obtidos na implementação do método proposto. As medidas de laboratório comprovam a eficácia do sistema de DPD proposto. O desempenho da linearização foi avaliado através da redução da intermodulação tanto fora da banda, medindo as emissões fora da faixa, quanto dentro banda, observando a melhoria da MER. Ao comparar com o método LS baseado em regressão linear, o desempenho foi equivalente e satisfatório.

7 Conclusões

A linearização usando DPD desempenha um papel crucial na redução da intermodulação na saída de um PA, para que especificações sejam atendidas tais como, atender uma máscara de emissão e garantir uma mínima qualidade de sinal. Contudo, a alta complexidade do DPD pode dificultar sua aplicação ou limitar seu desempenho. O método proposto de linearização ortogonal com realimentação escalar aborda esse problema e oferece as seguintes vantagens, quando comparado com métodos de regressão linear:

- versatilidade – devido a possibilidade de escolher diferentes métricas a serem otimizadas;
- implementação simples – a minimização da medida custo se torna um problema de otimização numérica genérico;
- nenhuma necessidade de alinhamento temporal e normalização de ganho entre os sinais de entrada e saída do PA antes da identificação – para tomar a métrica fora da banda, por exemplo;
- melhor desempenho na redução da intermodulação – a métrica escolhida no domínio da frequência é diretamente otimizada e não indiretamente por meio de uma métrica no domínio do tempo.

Quando comparado com outros métodos de realimentação escalar, o processo de ortogonalização promove as seguintes vantagens:

- convergência rápida – da ordem de 5 vezes mais rápido;
- baixa variação na potência de saída ao longo das iterações – reduzida de 0,69 [dB] para 0,14 [dB] de pico-a-pico.

Os PAs variam seu comportamento em função de fatores como: potência de saída, frequência de operação, temperatura, envelhecimento, entre outros [75]. Isso significa que se o PA já estiver linearizado, a alteração de um desses fatores resulta em um novo comportamento que degrada o desempenho de intermodulação. Para os casos em que esta degradação não é aceitável, deve-se executar o processo de linearização mais uma vez, ou executá-lo continuamente.

As especificações impostas pelos padrões de comunicação sem fio podem ser diretamente otimizadas. O método proposto neste trabalho é flexível e pode ser adaptado para outras formas de onda, medidas custo, equações de modelo comportamental não-linear e algoritmos de otimização numérica.

7.1 Trabalhos futuros

No diagrama de blocos do sistema proposto na Figura 31, a medida custo é tomada digitalmente em um bloco localizado após o ADC. Considerando que a informação realimentada é um escalar que muda seu valor a cada iteração, surge uma oportunidade para simplificar o circuito de realimentação como um todo. Se o processo de tomar a medida custo for implementado em um circuito analógico, não haverá a necessidade de digitalizar diretamente o sinal realimentado, apenas a saída do circuito de medida custo. No sistema convencional, a frequência de amostragem do ADC precisa ser suficientemente alta para acomodar, não somente a largura de faixa do sinal, mas também a largura de faixa da intermodulação, que é tipicamente cinco vezes maior do que a largura de faixa do sinal [16]. Com um circuito de medida custo, o sinal realimentado torna-se uma tensão cujo valor é proporcional à medida custo em regime permanente. Ou seja, o sinal realimentado somente muda seu valor a cada iteração. Isso reduz a taxa de amostragem do ADC significativamente, tornando-a independente da largura de faixa do sinal, dependendo somente da taxa temporal das iterações. O sinal de realimentação de um sistema típico de DPD apresenta uma largura de faixa, incluindo os produtos de intermodulação, entre 30 MHz e 100 MHz. Assim, seria necessário um ADC com taxa de amostragem entre 60 MS/s e 200 MS/s. Se houver 100 iterações por segundo, a frequência de amostragem será 100 Hz. Elimina-se também a necessidade de um circuito de captura, em que é necessário uma memória com alta taxa de escrita, uma vez que o sinal realimentado deve ser armazenado em tempo real. Esta solução torna-se muito atrativa para sistemas com alta largura de faixa. A Figura 71 ilustra o diagrama de blocos dessa ideia.

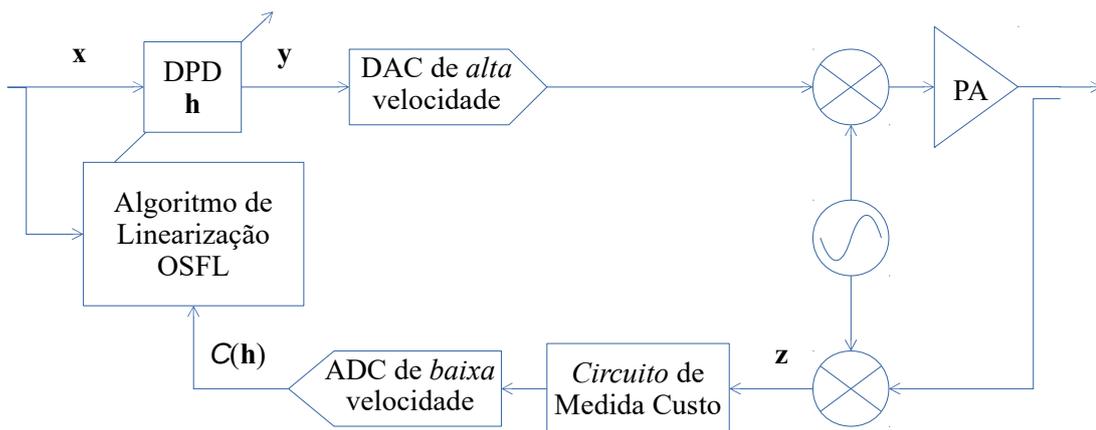


Figura 71 – Diagrama de blocos do sistema proposto de linearização com realimentação de escalar (OSFL), tomando a medida custo usando um circuito analógico.

A continuação natural desta linha de pesquisa é projetar um circuito analógico para extração da medida custo, a partir do sinal realimentado da saída do PA. O projeto desse circuito traz desafios, uma vez que é necessário uma sensibilidade para detectar a

intermodulação que pode apresentar níveis de potência até 50 dB abaixo do sinal. Dependendo da solução adotada, toda a complexidade eliminada no ADC pode ser reintroduzida. Logo, tal circuito teria uma importância para garantir a simplificação do sistema como um todo. Caso seja possível implementar um circuito de medida custo com complexidade e desempenho satisfatórios, haverá uma vantagem.

Em sistemas de radiodifusão, o transmissor opera em uma frequência específica ao longo de sua vida útil. Assim, é normalmente utilizado um filtro passa-faixa sintonizado no canal de operação, com o objetivo de atenuar as emissões fora da faixa. Logo, para atender à máscara de emissão, é empregado a linearização com DPD em conjunto com o filtro passa-faixas. Devido à alta potência de saída de um transmissor, o filtro apresenta uma estrutura mecânica construída a partir de cavidades ressonantes, por exemplo. As imperfeições de construção ou ajuste desse filtro podem provocar assimetria na sua resposta em frequência. Isso faz com que as emissões fora da faixa inferiores e superiores ao canal apresente valores diferentes, provocando um desbalanceamento nos produtos de intermodulação na saída do filtro. Compensar esse desbalanceamento com equalização linear não é possível, pois o desbalanceamento na intermodulação ocorre fora da faixa do canal, onde não há energia do sinal de entrada. A medida custo do algoritmo proposto pode ser selecionada para compensar esse balanceamento. Ou seja, fazendo que o DPD atue mais nos produtos de intermodulação de uma banda lateral em detrimento da outra, compensando o efeito de desbalanceamento do filtro. Além do mais, mesmo na ausência de um filtro passa-faixas, o efeito memória do amplificador pode provocar uma assimetria nos produtos de intermodulação.

O tempo de convergência do OSFL se mostrou mais rápido quando comparado em termos de número de iterações com outros métodos de realimentação escalar. Caso a comparação fosse feita com o LS, que é um algoritmo de iteração única, deveria ser feita em tempo absoluto, ou seja, em segundos. Nesse caso, o LS é o mais rápido uma vez que é executado na ordem de segundos, enquanto o OSFL na ordem de minutos. Embora o OSFL tenha sido proposto como uma alternativa para os métodos de regressão linear como o LS, por exemplo, os dois algoritmos não precisam ser concorrentes, mas podem operar de forma conjunta. Se o LS fosse executado e os coeficientes resultantes fossem utilizados como valores iniciais para o OSFL, o tempo de execução seria reduzido e a redução na intermodulação seria maior, pois aproximadamente 90% das iterações do OSFL são gastas para atingir o desempenho do LS e as 10% finais para superá-lo. Essa é uma possibilidade para um trabalho futuro.

Finalmente, também é válido investigar outros algoritmos de otimização numérica global adequados para sequências aleatórias de entrada, como o algoritmo genético [76].

APÊNDICE A – Método dos mínimos quadrados

Seja um sistema de equações lineares do tipo

$$\mathbf{A}\mathbf{x} = \mathbf{b} \quad (\text{A.1})$$

sendo que $\mathbf{A} \in \mathbb{C}^{N \times K}$, $\mathbf{x} \in \mathbb{C}^{K \times 1}$ e $\mathbf{b} \in \mathbb{C}^{K \times 1}$. O sistema é super-determinado se $N > K$, isto é, o número de equações é maior que o número de incógnitas. Reescrevendo (A.1) tem-se

$$\begin{bmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,K} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,K} \\ \vdots & \vdots & \ddots & \vdots \\ A_{N,1} & A_{N,2} & \cdots & A_{N,K} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_K \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_K \end{bmatrix}. \quad (\text{A.2})$$

em que \mathbf{A} e \mathbf{b} são conhecidos, porém \mathbf{x} é o conjunto de variáveis que se deseja encontrar. Suponha que não há solução exata para (A.1), ou seja, não há uma combinação linear das colunas de \mathbf{A} que resulta em \mathbf{b} . Desta forma os valores estimados de \mathbf{x} , dados por $\hat{\mathbf{x}}$ produzirão um sistema

$$\mathbf{A}\hat{\mathbf{x}} = \hat{\mathbf{b}}. \quad (\text{A.3})$$

Por não haver solução exata, então $\mathbf{b} \neq \hat{\mathbf{b}}$. Matricialmente isto é equivalente a dizer que \mathbf{b} não pertence ao espaço definido pelas colunas de \mathbf{A} . Pode-se definir uma função custo como sendo o erro quadrático em função de $\hat{\mathbf{x}}$ como

$$\begin{aligned} J(\hat{\mathbf{x}}) &= (\mathbf{A}\hat{\mathbf{x}} - \mathbf{b})^\dagger (\mathbf{A}\hat{\mathbf{x}} - \mathbf{b}) \\ &= \|\mathbf{A}\hat{\mathbf{x}} - \mathbf{b}\|_2^2 \end{aligned} \quad (\text{A.4})$$

sendo que $\|\cdot\|_p$ denota a norma-p de um vetor e $J(\hat{\mathbf{x}}) \in \mathbb{R}^{1 \times 1}$ podendo ser reescrito como

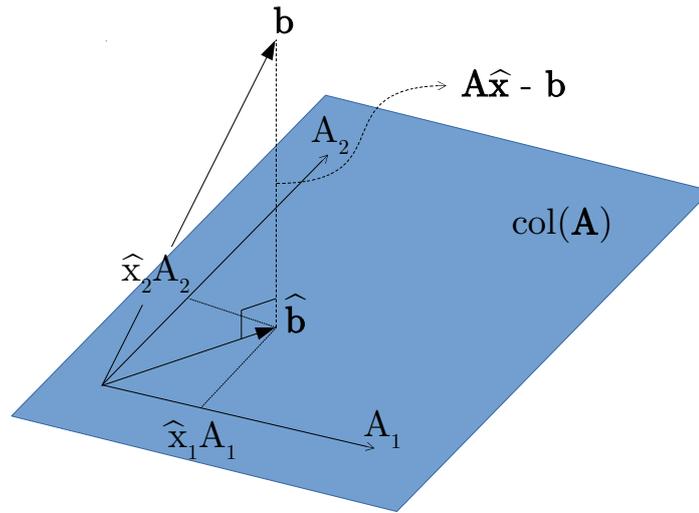
$$\begin{aligned} J(\hat{\mathbf{x}}) &= \sum_{i=1}^N (\hat{b}_i - b_i)^* (\hat{b}_i - b_i) \\ &= \sum_{i=1}^N |\hat{b}_i - b_i|^2 \end{aligned} \quad (\text{A.5})$$

em que $(\cdot)^*$ denota a operação de conjugar um número complexo. Abrindo (A.4) tem-se

$$J(\hat{\mathbf{x}}) = \mathbf{b}^\dagger \mathbf{b} - \mathbf{b}^\dagger \mathbf{A}\hat{\mathbf{x}} - \hat{\mathbf{x}}^\dagger \mathbf{A}^\dagger \mathbf{b} + \hat{\mathbf{x}}^\dagger \mathbf{A}^\dagger \mathbf{A}\hat{\mathbf{x}} \quad (\text{A.6})$$

e como $\mathbf{b}^\dagger \mathbf{A}\hat{\mathbf{x}} = \hat{\mathbf{x}}^\dagger \mathbf{A}^\dagger \mathbf{b}$ resulta em

$$J(\hat{\mathbf{x}}) = \mathbf{b}^\dagger \mathbf{b} - 2\mathbf{b}^\dagger \mathbf{A}\hat{\mathbf{x}} + \hat{\mathbf{x}}^\dagger \mathbf{A}^\dagger \mathbf{A}\hat{\mathbf{x}}. \quad (\text{A.7})$$


 Figura 72 – Projeção do vetor \mathbf{b} no espaço das colunas de \mathbf{A} .

O método dos mínimos quadrados se baseia em encontrar os valores de $\hat{\mathbf{x}}$ que minimizam o erro quadrático $J(\hat{\mathbf{x}})$. Para isto diferencia-se a função custo em relação ao vetor $\hat{\mathbf{x}}$ e iguala-se o resultado a zero, como mostrado a seguir

$$\frac{\partial J(\hat{\mathbf{x}})}{\partial \hat{\mathbf{x}}} = 0 \quad (\text{A.8})$$

$$\frac{\partial}{\partial \hat{\mathbf{x}}} \mathbf{b}^\dagger \mathbf{b} - 2 \frac{\partial}{\partial \hat{\mathbf{x}}} \mathbf{b}^\dagger \mathbf{A} \hat{\mathbf{x}} + \frac{\partial}{\partial \hat{\mathbf{x}}} \hat{\mathbf{x}}^\dagger \mathbf{A}^\dagger \mathbf{A} \hat{\mathbf{x}} = 0.$$

A primeira derivada será zerada pois o termo $\mathbf{b}^\dagger \mathbf{b}$ não depende de $\hat{\mathbf{x}}$. Lançando mão das propriedades

$$\frac{\partial}{\partial \theta} \mathbf{v}^\dagger \theta = \mathbf{v} \quad (\text{A.9})$$

$$\frac{\partial}{\partial \theta} \theta^\dagger \mathbf{v} \theta = 2\mathbf{v} \theta,$$

(A.8) fica

$$-2\mathbf{A}^\dagger \mathbf{b} + 2\mathbf{A}^\dagger \mathbf{A} \hat{\mathbf{x}} = 0, \quad (\text{A.10})$$

o que resulta na equação normal dada por

$$\mathbf{A}^\dagger \mathbf{A} \hat{\mathbf{x}} = \mathbf{A}^\dagger \mathbf{b} \quad (\text{A.11})$$

cuja solução é

$$\hat{\mathbf{x}} = (\mathbf{A}^\dagger \mathbf{A})^{-1} \mathbf{A}^\dagger \mathbf{b}. \quad (\text{A.12})$$

Fica provado então que a solução da equação normal corresponde à minimização do erro quadrático e, portanto, à solução do método dos mínimos quadrados.

Há uma interpretação geométrica para o método LS dada pela Figura 72. Como não há solução exata para (A.1), o vetor \mathbf{b} não pertence ao espaço das colunas de \mathbf{A} . Isto é, \mathbf{b} não pode ser formado por uma combinação linear das colunas de \mathbf{A} e, portanto,

está fora do plano imagem de \mathbf{A} . Encontrar a melhor estimativa de $\hat{\mathbf{x}}$ que minimiza o erro quadrático entre \mathbf{b} e $\hat{\mathbf{b}}$ significa fazer uma projeção ortogonal de \mathbf{b} no espaço coberto pelas colunas de \mathbf{A} . Desta forma, a projeção faz com que o erro $J(\hat{\mathbf{x}})$ fique perpendicular ao plano das colunas de \mathbf{A} que corresponde à menor distância entre \mathbf{b} e $\hat{\mathbf{b}}$, minimizando o erro.

APÊNDICE B – Publicações

Até o presente momento, esse trabalho gerou um artigo aceito para publicação em uma revista internacional.

B.1 Artigo aceito em periódico

Revista *IEEE Transactions on Broadcasting*, de Qualis A-1 de um artigo relacionado ao método de linearização proposto.

- Henry D. Rodrigues, Tales C. Pimenta, Rausley A. A. de Souza, Luciano L. Mendes. Orthogonal Scalar Feedback Digital Pre-Distortion Linearization. *IEEE Transactions on Broadcasting*.

B.2 Artigo submetido em periódico

Submissão na revista *IEEE Wireless Communications Letters*, de Qualis A-1 de um artigo relacionado ao método de linearização proposto.

- Henry D. Rodrigues, Wheberth D. Dias, Tales C. Pimenta, Rausley A. A. de Souza, Luciano L. Mendes. Orthonormal Scalar-Feedback Digital Pre-Distortion Linearization for In/Out-of-band Optimization. *IEEE Wireless Communications Letters*.

Referências

- [1] S. O’Leary, *Understanding Digital Terrestrial Broadcasting*. Artech House digital audio and video library, Artech House, 2000.
- [2] H. Ochiai and H. Imai, “On the distribution of the peak-to-average power ratio in OFDM signals,” *IEEE Transactions on Communications*, vol. 49, pp. 282–289, Feb 2001.
- [3] H. Chireix, “High power outphasing modulation,” *Proceedings of the Institute of Radio Engineers*, vol. 23, pp. 1370–1392, Nov 1935.
- [4] H. D. Rodrigues, “High efficiency transmission - green amplification,” in *NAB Broadcasting Engineering Conference*, 2012.
- [5] J. L. Dawson and T. H. Lee, “Cartesian feedback for RF power amplifier linearization,” in *American Control Conference, 2004. Proceedings of the 2004*, vol. 1, pp. 361–366 vol.1, June 2004.
- [6] H. Cho, M. S. Kim, J. Jeong, J. h. Van, S. c. Jung, H. c. Park, K. h. Lim, S. w. Kwon, J. Y. Kim, S. C. Song, C. S. Park, and Y. Yang, “A 30 W cartesian feedback transmitter with 40 % efficiency incorporating an inverted Doherty amplifier,” in *2007 Asia-Pacific Microwave Conference*, pp. 1–4, Dec 2007.
- [7] N. Delaunay, N. Deltimple, E. Kerhervé, and D. Belot, “A RF transmitter linearized using cartesian feedback in CMOS 65nm for UMTS standard,” in *Power Amplifiers for Wireless and Radio Applications (PAWR), 2011 IEEE Topical Conference on*, pp. 49–52, Jan 2011.
- [8] T. J. Bennett and R. F. Clements, “Feedforward an alternative approach to amplifier linearization,” *Radio and Electronic Engineer*, vol. 44, pp. 257–262, May 1974.
- [9] M. K. Ibrahim, “Feedforward linearization of a power amplifier for wireless communication systems,” *Journal of Babylon University/Engineering Sciences*, vol. 21, no. 4, pp. 1183–1193, 2013.
- [10] H. Ma and Q. Feng, “An improved design of feed-forward power amplifier,” *PIERS Online*, vol. 3, no. 4, pp. 363–367, 2007.
- [11] “RF predistortion (RFPD) theory of operation.” <https://www.maximintegrated.com/en/products/comms/wireless-rf/rf-predistorters/predistortion-linearization/rfpal-technology/rfpd-theory-operation.html>, 2016. [*Online*; acessado em 18 de Agosto de 2016].

- [12] D. R. Morgan, Z. Ma, J. Kim, M. G. Zierdt, and J. Pastalan, “A generalized memory polynomial model for digital predistortion of RF power amplifiers,” *IEEE Transactions on Signal Processing*, vol. 54, pp. 3852–3860, Oct 2006.
- [13] C. Yu, L. Guan, E. Zhu, and A. Zhu, “Band-limited Volterra series-based digital predistortion for wideband RF power amplifiers,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, pp. 4198–4208, Dec. 2012.
- [14] B. Ai, Z. x. Yang, C. y. Pan, S. g. Tang, and T. t. Zhang, “Analysis on LUT based predistortion method for HPA with memory,” *IEEE Transactions on Broadcasting*, vol. 53, pp. 127–131, Mar. 2007.
- [15] J. Kim, Y. Y. Woo, J. Moon, and B. Kim, “A new wideband adaptive digital predistortion technique employing feedback linearization,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 56, no. 2, pp. 385–392, 2008.
- [16] Y. Liu, W. Pan, S. Shao, and Y. Tang, “A general digital predistortion architecture using constrained feedback bandwidth for wideband power amplifiers,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 63, no. 5, pp. 1544–1555, 2015.
- [17] M. V. D. Nair, R. Giofrè, L. Piazzon, and P. Colantonio, “A comparative study on digital predistortion techniques for Doherty amplifier for LTE applications,” in *Integrated Nonlinear Microwave and Millimetre-wave Circuits (INMMiC), 2014 International Workshop on*, pp. 1–3, April 2014.
- [18] M. Allegue-Martinez, M. J. Madero-Ayora, J. G. Doblado, C. Crespo-Cadenas, J. Reina-Tosina, and V. Baena, “Digital predistortion technique with in-band interference optimisation applied to DVB-T2,” *Electronics Letters*, vol. 48, pp. 566–568, May 2012.
- [19] J. Lee, S. Jeon, J. Kim, and Y. W. Suh, “Adaptive HPA linearization technique for practical ATSC DTV system,” *IEEE Transactions on Broadcasting*, vol. 59, pp. 376–381, June 2013.
- [20] K. Freiburger, M. Wolkerstorfer, H. Enzinger, and C. Vogel, “Digital predistorter identification based on constrained multi-objective optimization of WLAN standard performance metrics,” in *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 862–865, May 2015.
- [21] C. T. Chen, C. J. Li, J. Y. Du, T. S. Horng, J. K. Jau, J. Y. Li, P. K. Horng, and D. S. Deng, “Power amplifier linearization using baseband digital predistortion for WiMAX applications,” in *2008 Asia-Pacific Microwave Conference*, pp. 1–4, Dec 2008.

- [22] C. D. Presti, D. F. Kimball, and P. M. Asbeck, “Closed-loop digital predistortion system with fast real-time adaptation applied to a handset WCDMA PA module,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, pp. 604–618, March 2012.
- [23] “Presentation on Digital Predistortion of Power Amplifiers.” <http://literature.agilent.com/litweb/pdf/5989-9105EN.pdf>, 2016. [Online; acessado em 25 de Agosto de 2016].
- [24] H. Raab, P. Asbeck, P. B. Kenington, Z. B. Popovich, N. Potheary, J. F. Sevic, and N. O. Sokal, “RF and microwave power amplifier and transmitter technologies,” *High Frequency Electronics*, pp. 22–36, 2003.
- [25] “Ettus Research – USRP X310 Software Defined Radio Specifications.” <https://www.ettus.com/product/details/X310-KIT>, 2017. [Online; acessado em 07 de Dezembro de 2017].
- [26] P. Gilabert, G. Montoro, and E. Bertran, “On the Wiener and Hammerstein models for power amplifier predistortion,” in *2005 Asia-Pacific Microwave Conference Proceedings*, vol. 2, pp. 4 pp.–, Dec 2005.
- [27] A. Zhu, J. C. Pedro, and T. J. Brazil, “Dynamic deviation reduction-based Volterra behavioral modeling of RF power amplifiers,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 54, pp. 4323–4332, Dec 2006.
- [28] H. Paaso and A. Mämmelä, “Comparison of direct learning and indirect learning predistortion architectures,” *ISWCS’08 - Proceedings of the 2008 IEEE International Symposium on Wireless Communication Systems*, pp. 309–313, 2008.
- [29] Z. Wang, W. Chen, and G. Su, “Low computational complexity digital predistortion for broadband power amplifiers,” in *2015 Asia-Pacific Microwave Conference (APMC)*, vol. 1, pp. 1–3, Dec. 2015.
- [30] Y. Khawam, L. Albasha, and H. Mir, “Accurate and low complexity polynomial and neural network models for PA digital pre-distortion,” in *2016 16th Mediterranean Microwave Symposium (MMS)*, pp. 1–4, Nov. 2016.
- [31] C. Nader, P. N. Landin, W. V. Moer, N. Bjorsell, and P. Handel, “Performance evaluation of peak-to-average power ratio reduction and digital pre-distortion for OFDM based systems,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 59, pp. 3504–3511, Dec. 2011.
- [32] L. Guan and A. Zhu, “Optimized low-complexity implementation of least squares based model extraction for digital predistortion of RF power amplifiers,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 60, no. 3 PART 1, pp. 594–603, 2012.

- [33] F. Antonio, W. Hamdy, P. Heidmann, J. Heizer, N. Kasturi, D. P. Oses, and C. Riddle, “A novel adaptive predistortion technique for power amplifiers,” in *Vehicular Technology Conference, 1999 IEEE 49th*, vol. 2, pp. 1505–1509 vol.2, July 1999.
- [34] B. D. Laki and C. J. Kikkert, “Adaptive digital predistortion for wideband high crest factor applications based on the WACP optimization objective: A conceptual overview,” *IEEE Transactions on Broadcasting*, vol. 58, pp. 609–618, Dec. 2012.
- [35] B. D. Laki and C. J. Kikkert, “Adaptive digital predistortion for wideband high crest factor applications based on the WACP optimization objective: An extended analysis,” *IEEE Transactions on Broadcasting*, vol. 59, pp. 136–145, 03 2013.
- [36] K. Freiburger, M. Wolkerstorfer, H. Enzinger, and C. Vogel, “Digital predistorter identification based on constrained multi-objective optimization of WLAN standard performance metrics,” in *2015 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 862–865, May 2015.
- [37] P. Kenington, *High-linearity RF Amplifier Design*. Artech House microwave library, Artech House, 2000.
- [38] D. Y. C. Lie, J. Tsay, T. Hall, T. Nukala, and J. Lopez, “High-efficiency silicon rf power amplifier design; current status and future outlook,” in *2016 IEEE International Symposium on Radio-Frequency Integration Technology (RFIT)*, pp. 1–4, Aug 2016.
- [39] T. Huber, R. Quay, and W. Bösch, “New concept for power compression improvement of GaN cascodes in broadband power amplifiers,” vol. 27, pp. 590–592, June 2017.
- [40] G. Montoro, P. Gilabert, J. Berenguer, and E. Bertran, “Digital predistortion of envelope tracking amplifiers driven by slew-rate limited envelopes,” in *Microwave Symposium Digest (MTT), 2011 IEEE MTT-S International*, pp. 1–1, June 2011.
- [41] J. Wood, *Behavioral Modeling and Linearization of RF Power Amplifiers*. Artech House Microwave Library, Artech House, 2014.
- [42] A. Fallis, *Simulation of Communication Systems: Modeling, Methodology and Techniques*, vol. 53. 2013.
- [43] R. Raich, “Orthogonal polynomials for complex gaussian processes,” *IEEE Transactions on Signal Processing*, vol. 52, pp. 2788–2797, Oct 2004.
- [44] R. Raich, H. Qian, and G. T. Zhou, “Orthogonal polynomials for power amplifier modeling and predistorter design,” *IEEE Transactions on Vehicular Technology*, vol. 53, no. 5, pp. 1468–1479, 2004.

- [45] H. Ku and J. S. Kenney, “Behavioral modeling of nonlinear rf power amplifiers considering memory effects,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 51, pp. 2495–2504, Dec 2003.
- [46] J. Vuolevi, T. Rahkonen, and J. Manninen, “Measurement technique for characterizing memory effects in RF power amplifiers,” in *Radio and Wireless Conference, 2000. RAWCON 2000. 2000 IEEE*, pp. 195–198, 2000.
- [47] P. Alper, “A consideration of the discrete Volterra series,” *Automatic Control, IEEE Transactions on*, vol. 10, pp. 322–327, Jul 1965.
- [48] L. Ding, G. T. Zhou, D. R. Morgan, Z. Ma, J. S. Kenney, J. Kim, and C. R. Giardina, “Memory polynomial predistorter based on the indirect learning architecture,” in *Global Telecommunications Conference, 2002. GLOBECOM '02. IEEE*, vol. 1, pp. 967–971 vol.1, Nov 2002.
- [49] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*. Melbourne, FL, USA: Krieger Publishing Co., Inc., 2006.
- [50] G. H. Golub and C. F. Van Loan, *Matrix Computations (3rd Ed.)*. Baltimore, MD, USA: Johns Hopkins University Press, 1996.
- [51] U. Ascher and C. Greif, *A First Course on Numerical Methods*. Computational Science and Engineering, Siam, 2011.
- [52] S. Haykin and B. Widrow, *Least-Mean-Square Adaptive Filters*. Adaptive and Cognitive Dynamic Systems: Signal Processing, Learning, Communications and Control, Wiley, 2003.
- [53] S. Haykin, *Adaptive Filter Theory*. Prentice-Hall information and system sciences series, Prentice Hall, 1996.
- [54] B. Farhang-Boroujeny, *Adaptive filters : theory and applications*. Chichester, New York, Weinheim: John Wiley & Sons, 1998.
- [55] P. Young, *Recursive Estimation and Time-Series Analysis: An Introduction for the Student and Practitioner*. Springer Berlin Heidelberg, 2011.
- [56] Y. Ma, S. He, Y. Akaiwa, and Y. Yamao, “An open-loop digital predistorter based on memory polynomial inverses for linearization of RF power amplifier,” *International Journal of RF and Microwave Computer-Aided Engineering*, vol. 21, no. 5, pp. 589–595, 2011.
- [57] C. Eun and E. J. Powers, “A new Volterra predistorter based on the indirect learning architecture,” *IEEE Trans. Signal Processing*, vol. 45, no. 1, pp. 223–227, 1997.

- [58] J. F. Bonnans, J. C. Gilbert, C. Lemaréchal, and C. A. Sagastizábal, *Numerical Optimization: Theoretical and Practical Aspects (Universitext)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2006.
- [59] J. Nocedal and S. Wright, *Numerical optimization*. Springer series in operations research and financial engineering, New York, NY: Springer, 2. ed. ed., 2006.
- [60] C. T. Kelley, *Iterative methods for optimization*. Frontiers in applied mathematics, Philadelphia: SIAM, 1999.
- [61] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, “Convergence properties of the Nelder-Mead simplex method in low dimensions,” *SIAM Journal on Optimization*, vol. 9, no. 1, pp. 112–147, 1998.
- [62] D. F. Shanno, “Conditioning of Quasi-Newton Methods for Function Minimization,” *Mathematics of Computation*, vol. 24, pp. 647–656, July 1970.
- [63] U. Niesen, D. Shah, and G. W. Wornell, “Adaptive alternating minimization algorithms,” *IEEE Transactions on Information Theory*, vol. 55, no. 3, pp. 1423–1429, 2009.
- [64] R. v. Nee and R. Prasad, *OFDM for Wireless Multimedia Communications*. Norwood, MA, USA: Artech House, Inc., 1st ed., 2000.
- [65] J. E. Audet, Charles; Dennis, “Analysis of generalized pattern searches,” *SIAM Journal on Optimization*, vol. 13, pp. 889–903, Jan. 2002.
- [66] ITU, “ITU Recommendation – Spectrum limit masks for digital terrestrial television broadcasting,” *ITU-R BT.1206-1*, 2013.
- [67] F. Gantmacher, *The theory of matrices*. No. v. 1 in The Theory of Matrices, Chelsea Pub. Co., 1960.
- [68] H. Yanai, K. Takeuchi, and Y. Takane, *Projection Matrices, Generalized Inverse Matrices, and Singular Value Decomposition*. Statistics for Social and Behavioral Sciences, Springer New York, 2011.
- [69] J. E. Volder, “The CORDIC trigonometric computing technique,” *IRE Transactions on Electronic Computers*, vol. EC-8, pp. 330–334, Sep 1959.
- [70] L. Sundstrom, M. Faulkner, and M. Johansson, “Quantization analysis and design of a digital predistortion linearizer for RF power amplifiers,” *IEEE Transactions on Vehicular Technology*, vol. 45, pp. 707–719, Nov 1996.

-
- [71] “Arrow SoCKit Evaluation Board.” <http://rocketboards.org/foswiki/view/Documentation/ArrowSoCKitEvaluationBoard>, 2015. [Online; acessado em 9 de Dezembro de 2015].
- [72] X. Moronval, J. Bouny, and G. Bouisse, “Ultra wideband Doherty amplifier,” Oct. 21 2015. EP Patent App. EP20,140,175,618.
- [73] T. H. Lee, *Planar Microwave Engineering: A Practical Guide to Theory, Measurement, and Circuits*. Cambridge University Press, har/cdr ed., 2004.
- [74] Y. Rahmatallah and S. Mohan, “Peak-to-average power ratio reduction in OFDM systems: A survey and taxonomy,” *IEEE Communications Surveys Tutorials*, vol. 15, pp. 1567–1592, Fourth 2013.
- [75] D. Schreurs, M. O’Droma, A. Goacher, and M. Gadringer, *RF Power Amplifier Behavioral Modeling*. The Cambridge RF and Microwave Engineering Series, Cambridge University Press, 1 ed., 2008.
- [76] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1st ed., 1989.