

SeMicro-PR 2018

Métodos de treinamento para modelos de modulador IQ em conjunto com amplificador de potência baseados em redes neurais

Luiza Beana C. Freire, Eduardo G. Lima

Grupo de Circuitos e Sistemas Integrados (GICS) – Departamento de Engenharia Elétrica, Curitiba, Brasil
luiza.chipans@gmail.com

Resumo—Este artigo pretende analisar três modelos de um sistema conjunto amplificador de potência/Modulador IQ baseados em redes neurais. Na literatura foi proposto um modelo de amplificador de potência em conjunto com modulador IQ utilizando um método de treinamento iterativo com o algoritmo de treinamento backpropagation. Porém esse método separa o sistema a ser modelado em dois sistemas parciais e pode não modelar de forma precisa o sistema completo. Em vista disso, espera-se que um método que modele o sistema de forma única seja mais efetivo. Esse método global, todavia, não permite o uso do algoritmo backpropagation, e outro algoritmo deve ser utilizado. Pela facilidade de implementação foi utilizado o algoritmo de treinamento de ajuste de curvas com mínimos quadrados não lineares (MQO). Os resultados demonstram que o método iterativo com backpropagation possui vantagens chegando a valores de -43dB de NMSE, em comparação os que utilizam o MQO alcançaram valores piores de NMSE com -24dB para o iterativo e -32dB para o global. Contudo, o método global mostrou ser superior ao método iterativo quando comparado utilizando o mesmo algoritmo de treinamento. É indicado dessa forma utilizar o método iterativo com backpropagation, porém para casos onde é necessário utilizar outro método que não o backpropagation é indicado utilizar o método de treinamento global.

I. INTRODUÇÃO

Há uma crescente demanda por maiores taxas de dados em equipamentos móveis [1]. Porém a largura de banda disponível para transmissão de dados permanece a mesma. Uma maneira de utilizar de forma mais eficiente a banda, ou seja, alocar mais dados em um mesmo espaço, é implementando esquemas de modulação como a multiplexação por divisão de frequências ortogonais (OFDM) [2], que utilizam sinais de envoltórias não constantes. Esses tipos de modulação fazem com que o comportamento não linear dos amplificadores de potência (PA) piore. A característica de linearidade é importante na medida que contém o sinal na sua banda alocada sem degradar o próprio sinal e os sinais em bandas adjacentes.

Os amplificadores são componentes situados nos transmissores. Eles têm como papel aumentar a potência do

sinal a ser transmitido pela antena. Para operarem com alta linearidade os PAs precisam comprometer sua eficiência e vice-versa. A eficiência também é uma característica desejável, pois aumenta a autonomia de aparelhos móveis e diminui custos de operação de estações base.

A fim de melhorar o compromisso entre linearidade e eficiência no PA são utilizadas técnicas de linearização. A técnica mais utilizada é a de pré distorção digital (DPD) [3]. Ela distorce o sinal de entrada do PA com o inverso da característica não linear, pretendendo assim que se anulem mutuamente, mantendo apenas a informação do sinal a ser amplificado.

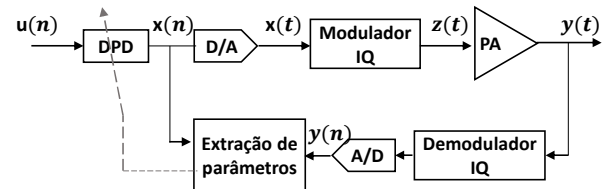


Fig. 1. Diagrama de blocos do sistema de pré-distorção digital.

O sinal processado pelo DPD é em banda base, ou seja, não modulado, e como o sinal processado pelo PA é em banda passante, modulado, é necessário o uso de moduladores e demoduladores I/Q para passar o sinal de um sistema para o outro, como na fig. 1. Além disso, os índices n e t retratam sinais digitais e analógicos respectivamente.

O modulador ideal IQ tem como papel modular um sinal portador com os componentes em fase (I) e em quadratura (Q) do sinal a ser transmitido. A entrada do modulador é o sinal complexo banda base $x(t) = x_I(t) + jx_Q(t)$, e o sinal de saída é o sinal portadora modulado por envoltória [4]:

$$\begin{aligned} z(t) &= \Re\{[x_I(t) + jx_Q(t)][\cos(2\pi f_c t) \\ &\quad + j\text{sen}(2\pi f_c t)]\} \\ &= x_I(t) \cos(2\pi f_c t) - x_Q(t) \text{sen}(2\pi f_c t), \end{aligned} \quad (1)$$

sendo f_c a frequência da portadora.

Nesse processo o sinal sofre distorções devido ao desbalanceamento IQ e desvio da corrente contínua (DC offset). Pode-se realizar a compensação das distorções causadas pelo modulador real previamente ou então compensá-las em conjunto com as distorções do PA. A vantagem do último caso é não ter de utilizar mais um sistema dedicado a essa função, poupando recursos.

Para realizar o DPD é necessário inicialmente modelar a característica não linear inversa do PA em conjunto com o modulador. Para isso existem diversos modelos que podem ser usados baseados nos mais diversos métodos: séries de Volterra, polinomiais e redes neurais [5] são alguns exemplos dos métodos mais estudados. As redes neurais têm grande capacidade de modelagem de sistemas não lineares [6] e possuem algumas vantagens em relação às séries de Volterra por necessitarem de menos parâmetros e possuírem também computação menos complexa que os modelos baseados em polinômios [7]. Existem diversas arquiteturas de redes neurais, a mais utilizada é o *perceptron* de três camadas (TLP) [8]. Este trabalho irá tratar da modelagem inversa do comportamento do PA em conjunto com o modulador utilizando modelos baseados em TLP.

Os modelos podem ser modificados para melhor atender ao comportamento do sistema modelado. Os modelos de redes neurais possuem diversas características que podem ser analisadas nesse sentido: sua estrutura (entradas, saídas, quantidade de redes, etc), o tipo de sinal processado pela rede (se real ou complexo), função de ativação, seu método de treinamento, o algoritmo de treinamento, etc.

Este trabalho tem como objetivo analisar dois diferentes algoritmos de treinamento, bem como dois métodos distintos para o treinamento da rede. O modelo com treinamento iterativo foi proposto em [9], e será comparado com um modelo com treinamento global proposto neste trabalho. Para o treinamento iterativo será usado ainda dois métodos: um que foi idealizado especificamente para redes neurais, *backpropagation*, [10] e um método de ajuste de curvas com mínimos quadrados não linear [11].

II. REDES NEURAIS

O diagrama de blocos da arquitetura de rede neurais utilizada está na fig. 2 e possui E entradas, R neurônios na camada escondida e S saídas. As entradas do TLP são multiplicadas pelos pesos (w) e somadas aos bias (b^l), o sinal resultante é submetido à função de ativação (F) e por fim multiplicado pelos pesos de saída (h) e somados aos bias de saída (b^o).

Com relação à natureza do sinal, o sistema modelado processa números complexos e a rede neural utilizada processa números reais, dessa forma deve-se decompor o sinal complexo em componentes reais.

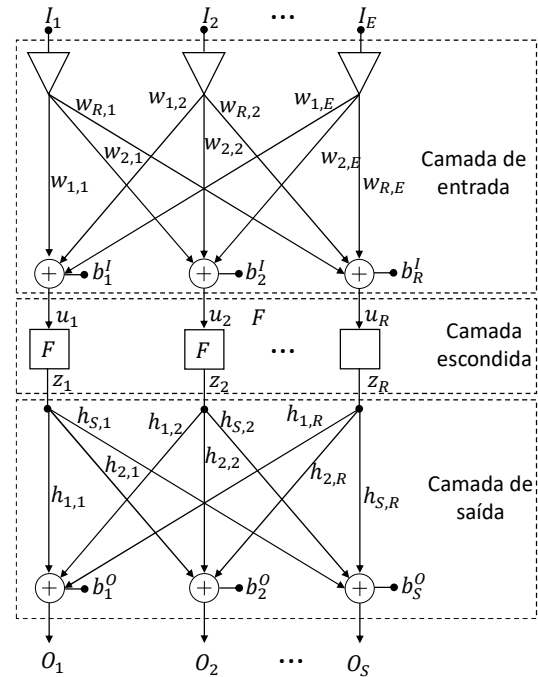


Fig. 2. Diagrama de blocos da rede neural TLP.

A função de ativação F utilizada é a função tangente hiperbólica dada por:

$$F = \tanh = \frac{e^{2u} - 1}{e^{2u} + 1} \quad (2)$$

Como descrito em [12] o sinal processado pela rede neural não deve possuir a informação da fase a fim de não gerar contribuições fora da banda.

A ideia principal do método *backpropagation* é retro propagar o erro da rede para os pesos, começando na camada de saída e indo em direção às camadas anteriores. Dessa forma se calculam novos valores para os pesos para então ser gerado um novo valor de erro. Esse ciclo continua até o valor de erro chegar em um valor desejado ou não haver variações significativas no valor. Esse método é específico para se obter os coeficientes de uma rede neural.

No caso do método MQO, ele é adequado para ajustes de dados não linear. Tem como objetivo minimizar a soma dos quadrados do sinal de erro (diferença entre os dados observados e calculados). Esse método é utilizado de forma genérica e pode ser aplicado a problemas diversos [13].

III. MODELOS DE PA E MODULADOR BASEADOS EM REDES NEURAIS

Modelos matemáticos visam representar o comportamento de um sistema real através de parâmetros, funções e relações. Nesse trabalho é utilizado um modelo caixa-preta, onde a entrada e a saída do sistema são inseridas na rede TLP e não se tem a priori nenhuma outra informação sobre o sistema. O objetivo é diminuir o erro entre a saída da rede e a saída real do sistema adequando seus parâmetros.

O sistema a ser modulado é caracterizado na fig. 1. A entrada do modelo é a saída do PA $y(n)$ e a saída do sistema é a entrada do modulador, $x(n)$.

Num cenário não ideal onde a resposta do modulador varia de forma não linear com desbalanceamentos de fase (φ) e de amplitude (α) sua saída é dada por:

$$z(n) = \Re\{[x_I(t) + jx_Q(t)][\cos(2\pi f_c n) - (\alpha + 1)j\text{sen}(2\pi f_c n + \varphi)]\}, \quad (3)$$

Como o modelo utiliza sinais em banda base, a representação banda base de (3) é dada por:

$$\tilde{z}(n) = b_1 \tilde{x} + b_2 \tilde{x}^*, \quad (4)$$

sendo $()^*$ a indicação de complexo conjugado, b_1 e b_2 coeficientes complexos relacionados aos fatores de desbalanceamento. Considerando o modelo inverso e trocando de papel a saída e a entrada, o modelo resultante é:

$$\tilde{x}(n) = (f_1 |\tilde{y}| + f_2 |\tilde{y}|) e^{+j\theta_{y(n)}} + (f_3 |\tilde{y}| + f_4 |\tilde{y}|) e^{-j\theta_{y(n)}} \quad (5)$$

onde f_1 e f_2 são funções não lineares, $|\tilde{y}|$ e $\theta_{y(n)}$ são o módulo e a fase do sinal de entrada da rede $y(n)$ respectivamente.

A. Modelo de treinamento iterativo

O modelo iterativo usando TLP de valores reais foi apresentado em [8]. Nesse método são utilizadas duas redes TLP. A entrada das duas redes é o módulo da saída do PA. A rede 1 (NN1) gera duas saídas, os componentes real (g_1) e imaginário (g_2) de um sinal complexo que é multiplicado por $e^{+j\theta_{y(n)}}$ para recuperar a fase, que é retirada do sinal a ser processado pela rede. A rede 2 (NN2) gera duas saídas, os componentes real (g_3) e imaginário (g_4) de um sinal complexo que é multiplicado por $e^{-j\theta_{y(n)}}$ para recuperar a fase, que é retirada do sinal a ser processado pela rede.

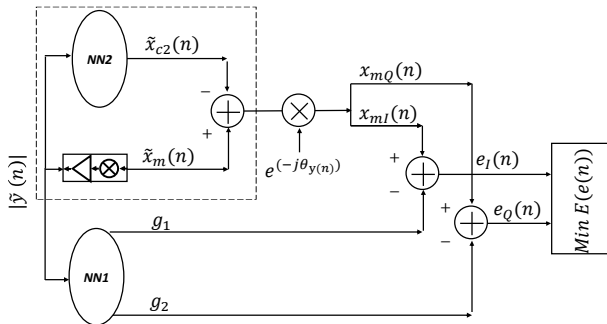


Fig. 3. Diagrama de blocos das iterações ímpares.

Cada rede deve gerar um sinal parcial de $\tilde{z}(n)$, ou seja, uma rede deve gerar o sinal $b_1 \tilde{x}$ e a outra rede o sinal $b_2 \tilde{x}^*$. Espera-se dessa forma aumentar a acurácia da rede, pois esses sinais são muito distintos para serem modelados por uma única rede. Porém, na prática, não se tem acesso a esses sinais, mas sim à soma desses sinais. Para simular esse efeito então é proposto um treinamento iterativo, onde o sinal desejado da NN1 é o sinal real do PA subtraído do sinal gerado pela rede 2 retirada a informação da fase. Da mesma forma para a NN2, o sinal desejado é o sinal do PA subtraído da NN1, retirada a

informação da fase. No treinamento esse processo de subtração é repetido até o erro global ser minimizado.

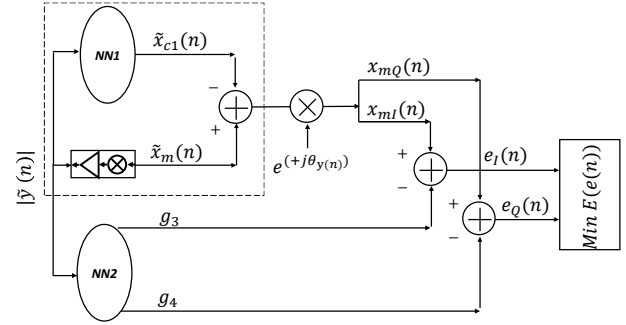


Fig. 4. Diagrama de blocos das iterações pares.

A fig. (3) e (4) mostram os diagramas de blocos das iterações ímpares e pares respectivamente. Nas iterações ímpares a saída da rede 2, $\tilde{x}_{c2}(n)$, é subtraída do sinal de saída medido do PA, $\tilde{x}_m(n)$ e retirada a fase do sinal de entrada. O sinal resultante é então decomposto em suas componentes real e imaginária. Esses sinais são comparados com as componentes real e imaginária da saída da NN1, gerando os sinais de erro a serem minimizados, $(e_I(n))$ e $(e_Q(n))$.

O processo se espelha nas iterações pares, onde o sinal de erro é gerado comparando o sinal desejado com as componentes real e imaginária da saída da NN2. Na primeira iteração o sinal $\tilde{x}_{c2}(n)$ é iniciado como zero.

B. Modelo de treinamento global

Diferentemente do treinamento iterativo, o treinamento global treina as duas redes em conjunto, ou seja, o algoritmo tem a informação das duas redes ao realizar o treinamento enquanto no treinamento iterativo as redes são treinadas separadamente, e só utilizam a informação das saídas que são subtraídas do sinal desejado. O diagrama de blocos do treinamento global é mostrado na fig. (5).

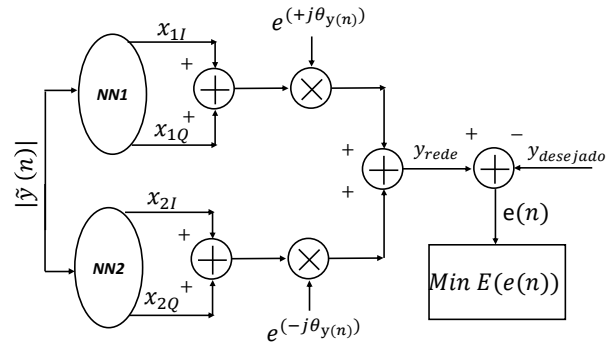


Fig. 5. Diagrama de blocos do treinamento global.

A NN1 gera o sinal real, $x_{1I}(n)$, e imaginário, $x_{1Q}(n)$, que é recomposto em um sinal complexo. Esse sinal então reincorpora a informação da fase. O mesmo processo ocorre com a NN2. Os sinais das duas redes são somados para formar o sinal completo a ser comparado com o sinal medido da saída do PA.

IV. RESULTADO DAS SIMULAÇÕES

Os modelos analisados foram implementados no ambiente de simulação MATLAB.

O sinal utilizado tem modulação LTE OFDMA com 10 MHz de largura de banda. A frequência de amostragem do sinal é 12,288 MHz e composto por 409600 amostras no tempo. O PA sob teste é um modelo polinomial de 5° ordem, seguido por um filtro passa-banda de 1,2288 GHz.

O modulador IQ sob teste tem desbalanceamentos de amplitude de 1,2 e de fase de $\pi/3$, além de uma frequência de portadora de 1,2288GHz. Para realizar a medição da envoltória é utilizado um demodulador IQ ideal, usando um filtro passa baixas de 122,88 MHz.

A medida de quantificação da acurácia dos modelos que será utilizada é o erro quadrático médio normalizado (NMSE) [14], em função do número de operações em ponto flutuante (FLOPs).

TABELA 1 – RESULTADO DE NMSE PARA OS MODELOS ANALISADOS

Método Iterativo	
Algoritmo	NMSE (dB)
<i>Backpropagation</i>	-43
Mínimos quadrados	-24
Método Global	
Mínimos quadrados	-32

Os resultados de NMSE para os três modelos podem ser visualizados na tabela 1. Os valores foram obtidos para 5 neurônios e 30 FLOPs (como a estrutura da rede foi a mesma com diferença apenas nas formas de treinamento o número de FLOPs é igual para os três modelos).

V. CONCLUSÃO

Foram analisados 2 métodos de treinamento de modelos de PA em conjunto com modulador IQ baseados em TLP, bem como dois algoritmos de minimização de erro. O algoritmo de mínimos quadrados é um algoritmo genérico e pode ser utilizado para o ajuste de curva de diversas aplicações. Já o *backpropagation* é um algoritmo específico criado para ajuste dos pesos de uma rede neural. O método iterativo com *backpropagation* obteve NMSE de -43dB e o método iterativo com MQO obteve NMSE de -24dB, observa-se que algoritmos moldados para um problema podem perder generalidade, porém ganham em eficiência. Dessa forma, encoraja-se o uso do algoritmo *backpropagation* para métodos iterativos de modelagem de PA e modulador que utilizem redes TLP. Além disso observou-se que o tempo de treinamento para o algoritmo *backpropagation* foi consideravelmente menor que para o MQO. Na impossibilidade do uso do método *backpropagation*, utilizar o método de treinamento global com MQO, pois este demonstrou superioridade em relação ao treinamento iterativo com MQO, visto os valores de NMSE de -32dB para o global e -24dB para o iterativo.

AGRADECIMENTOS

Os autores gostariam de agradecer ao suporte financeiro provido pela Fundação Araucária de Apoio ao Desenvolvimento Científico e tecnológico do Estado do Paraná (FAPPR), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

REFERÊNCIAS

- [1] M. Agiwal ; A. Roy ; N. Saxena, “Next Generation 5G Wireless Networks: A Comprehensive Survey”, *IEEE Communications Surveys & Tutorials*, vol. 18, n. 3, pp. 1617 – 1655, 2016.
- [2] Mridula, A. S. Sappal, “Nonlinearities in Power Amplifier and its Remedies”, *International Journal of Electronics Engineering Research*, vol. 9, n. 6, pp. 883887, 2017.
- [3] A. Katz, “Linearization: reducing distortion in power amplifiers”, *IEEE Microwave Magazine*, vol. 2, n. 4, pp. 37-49, 2001.
- [4] J. J. Witt, “Modelling, Estimation and Compensation of Imbalances in Quadrature Transceivers”, tese de doutorado, Univ. Stellenbosch, Dep. of Electrical and Electronic Engineering, 2011.
- [5] J. C. Pedro and S. A. Maas, “A comparative overview of microwave and wireless power-amplifier behavioral modeling approaches,” *IEEE Trans. Microw. Theory Tech.*, vol. 53, no. 4, pp. 1150–1163, Apr. 2005.
- [6] S. Haykin, *Neural Networks: a Comprehensive Foundation*, 2nd edition. New Jersey: Prentice Hall, 1999.
- [7] T. Liu, S. Boumaiza, and F. M. Ghannouchi, “Dynamic behavioral modeling of 3G power amplifiers using real-valued time-delay neural networks”, *IEEE Trans. Microw. Theory Tech.*, vol. 52, no. 3, pp. 1025–1033, Mar. 2004.
- [8] T. Liu, S. Boumaiza, and F. M. Ghannouchi, “Dynamic behavioral modeling of 3G power amplifiers using real-valued time-delay neural networks,” *IEEE Trans. Microw. Theory Tech.*, vol. 52, no. 3, pp. 1025–1033, Mar. 2004.
- [9] L. B. C. Freire, B. T. Marcondes and E. G. Lima, “A Three-layer Perceptron Tailored for the Joint Modeling of I/Q Modulator Mismatches and Power Amplifier Distortions”, *33° Simpósio Sul de Microeletrônica proceedings*, Curitiba, 2018, pp. 2124.
- [10] Hagan, M.T., and M. Menhaj, “Training feed-forward networks with the Marquardt algorithm,” *IEEE Transactions on Neural Networks*, Vol. 5, No. 6, 1999, pp. 989–993, 1994.
- [11] Marquardt, D., “An Algorithm for Least-Squares Estimation of Nonlinear Parameters,” *SIAM Journal on Applied Mathematics*, Vol. 11, No. 2, June 1963, pp. 431–441.
- [12] E. G. Lima, T. R. Cunha, and J. C. Pedro, “A Physically Meaningful Neural Network Behavioral Model for Wireless Transmitters Exhibiting PM–AM/PM–PM Distortions”, *IEEE Trans. Microw. Theory Tech.*, vol. 59, no. 12, pp. 3512–3521, Dec. 2011.
- [13] P. Hansen, V. Pereyra and G. Scherer. “Least Squares Data Fitting with Applications”, Maryland, The Johns Hopkins University Press, 2013.
- [14] V. N. P. Dao and V. R. Vemuri. “A Performance Comparison of Different Back Propagation Neural Networks Methods in Computer Network Intrusion Detection”, *Differential Equations and Dynamical Systems*, vol 10, No 1&2, pp 201-214, Jan&April, 2002.