COMPARISON OF NEURAL NETWORK NATURAL AND ORDINARY GRADIENT ALGORITHMS FOR SATELLITE DOWN LINK IDENTIFICATION

F. Langlet[1], H. Abdulkader[1], D. Roviras[1], L. Lapierre[2], and F. Castanié[1]

[1]TéSA, 2 rue Camichel, BP 7122, 31071 Toulouse cedex 7, France
Email: fabien.langlet@tesa.prd.fr
[2]CNES, BP i 2012, 18 av. E. Belin, 31055 Toulouse France

ABSTRACT

In this paper, we present a neural network architecture that belongs to the multi-layer perceptron family, associated with two different algorithms: the ordinary gradient and the natural gradient, we compare performances of those algorithms. The identification of a non-normalized power amplifier yielded to the introduction of an additional weight in the classical multi-layer perceptron structure. The application of this network is space telecommunications: identification of satellite communication channels, and especially the down link. This link is made up with two elements. The first one is a high power amplifier (non-linearity). The second one is a filter (memory).

1. INTRODUCTION

Neural networks (NN) application in space communications is becoming more and more important[4][1]. The non-linear computation of NN, their learning ability, their auto-organization and their parallel implementation allow them to be well adapt to solve complicated problems in telecommunications. Differents authors have used NN to solve non-linear problems in digital communications. For example, the multi-layer perceptron (MLP) has been used to identify non-linear channels[3]. This work has been motivated by the fact that the MLP is known as the 'universal approximation theorem': it has been simultaneously shown by several researchers in 1989 that an MLP with one hidden layer was able to uniformly approximate any function defined on a compact space if it has enough neurons on the hidden layer[2]. Another kind of NN is the radial basis function network (RBF). Many fundamental theoretical results about RBF were proved such as the 'universal approximation' property. Those networks have been used in UMTS channels equalization[1]. In this paper we realize the identification of a down link channel that is made up with a solid state power amplifier (SSPA) followed by a filter. The SSPA is not normalized, so that the input is not of the same order as the output.

We use a NN that belongs to the MLP family to identify separately each element of the channel. Two different algorithms are associated to that NN : the ordinary gradient and the natural gradient. In section 2 we present the system to identify. In section 3, we present the NN structure and its associated algorithms. The simulations results are shown in section 4.

2. THE SYSTEM TO IDENTIFY

The system to identify is the cascade of two elements: the SSPA and the filter. The main goal of communication satellite payloads is to provide a radio relay for links between earth stations. In order to exploit on-board resources with maximal efficiency, the payload equipments are often operated near their saturation points. This is particularly true for active components such as high power amplifiers (HPA). Various models have been applied for the characterization of HPA non-linear behaviors. HPA usually have finite bandwidth. If this bandwidth is larger than the signal bandwidth, then the non-linearity is memoryless and can be written as a complex voltage gain depending on the signal amplitude $G(\rho) = A(\rho).e^{j\Phi(\rho)}$. In our case the HPA is a SSPA. When using an SSPA aboard the satellite, it occurs two kinds of distortions: amplitude distortion (AM/AM conversion) and phase distortion (AM/PM conversion). Those two distortions are shown respectively on Fig. 1a and Fig. 1b. The "Centre National d'Etudes Spatiales" (CNES) gave us those data.

The main goal of the filter is to eliminate intermodulation products created by the SSPA. Filter characteristics are as follow: it is a band pass filter with 6 poles, centered on the carrier frequency $f_0 = 8200MHz$, the band pass is $60MHz$. Our simulations are made in base band so we work with the equivalent low pass filter.

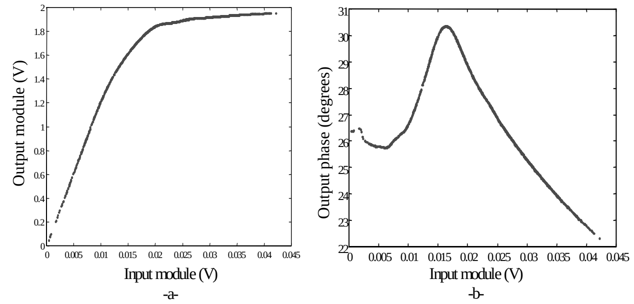

Fig. 1. AM/AM and AM/PM Conversion

3. THE NN STRUCTURE AND ITS ASSOCIATED ALGORITHMS

3.1. Structure

The NN is made up two parts (Fig. 2). It is a mimetic structure. The first one has the same architecture than the SSPA, it is non-linear (NLN) and computes the AM/AM and AM/PM conversion with two sub-networks. Each sub-network has 15 neurons in the hidden layer and one neuron in the output layer. The second part presents the architecture of a linear filter (LN). It is a FIR filter with 60 complex coefficients . The weight W_0 is a simple gain. The goal of that weight is to normalize the SSPA (i.e. input and output amplitudes for the AM/AM conversion of the SSPA are of the same order). All the weights of the NN are initialized with random values around zero.

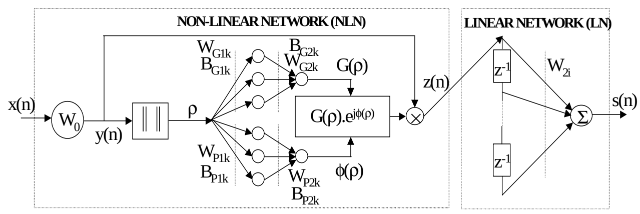

Fig. 2. NN architecture

3.2. Ordinary gradient algorithm

The learning ordinary gradient (OGD) algorithm is given by:

·**Forward phase**:

$$y(n) = W_0 \cdot x(n)$$
$$G(\rho) = \sum_{k=1}^{N_G} w_{G2k} \cdot f(w_{G1k} \cdot \rho + b_{G1k}) + b_{G2}$$
$$\phi(\rho) = \sum_{k=1}^{N_P} w_{P2k} \cdot f(w_{P1k} \cdot \rho + b_{P1k}) + b_{P2}$$

The NLN output is:

$$z = G(\rho) \cdot e^{i\phi(\rho)} \cdot y$$

where $f(.) = \tanh(.)$, is the non-linearity of one neuron and $\rho = \|y\|$.

We present to the LN the vector:

$$\mathbf{z}(n) = [z(n), z(n-1), ..., z(n-p+1)]^T,$$

where X^T is the transpose vector and p is the size of the LN. The LN output is as follows:

$$s(n) = \sum_{i=1}^{p} w_{2i}(n) \cdot z(n-i+1)$$

with complex coefficients w_{2i}, because of the complex impulse response.

The learning criterion to minimize is the square error criterion (where d is the desired output).

$$\|e\|^2 = \|d - s\|^2$$

·**Backpropagation**:

$\delta^1 = e_I + i.e_Q$

$\delta^2 = w_{21I}.e_I + w_{21Q}.e_Q + i.(-w_{21Q}.e_I + w_{21I}.e_Q)$

$\delta^G = 2 \cdot (\delta_I^2 \cdot (\cos(\phi(\rho)) \cdot y_I - \sin(\phi(\rho)) \cdot y_Q) + \delta_Q^2 \cdot (\sin(\phi(\rho)) \cdot y_I + \cos(\phi(\rho)) \cdot y_Q))$

$\delta^P = 2 \cdot G(\rho) \cdot (\delta_I^2 \cdot (-\sin(\phi(\rho)) \cdot y_I - \cos(\phi(\rho)) \cdot y_Q) + \delta_Q^2 \cdot (\cos(\phi(\rho)) \cdot y_I - \sin(\phi(\rho)) \cdot y_Q))$

·**Update of the coefficients**:

$w_{2i}(n+1) = w_{2i}(n) + \mu_{w_2}(n) \cdot e(n) \cdot z^*(n-i+1)$

$w_{G2k}(n+1) = w_{G2k}(n) + \mu_{w_{G2}}(n) \cdot x_{G1k}(n) \cdot \delta_{G2}(n)$

$w_{P2k}(n+1) = w_{P2k}(n) + \mu_{w_{P2}}(n) \cdot x_{P1k}(n) \cdot \delta_{P2}(n)$

$w_{G1k}(n+1) = w_{G1k}(n) + \mu_{w_{G1}}(n) \cdot w_{G2k}(n) \cdot \delta_{G2}(n) \cdot \rho(n) \cdot f'(w_{G1k}(n) \cdot \rho \cdot b_{G1k}(n))$

$w_{P1k}(n+1) = w_{P1k}(n) + \mu_{w_{P1}}(n) \cdot w_{P2k}(n) \cdot \delta_{P2}(n) \cdot \rho(n) \cdot f'((w_{P1k}(n) \cdot \rho \cdot b_{P1k}(n))$

b_{G1k} et b_{P1k} are respectively the bias of the first layer of gain ($G(\rho)$) and phase ($\phi(\rho)$) sub layers.

·**Iteration until convergence**.

3.3. Natural gradient algorithm

The learning of the linear part has the same scheme as above. The NLN is learned by the natural gradient which is a natural choice in a neural network manifold[5]. In this paper we consider the neural network (NN) as two sub-NNs. So we train each sub-NN separately in order to reduce the complexity of computation.

Natural gradient (NGD) for training NNs can be done by two methods: global and behavioral methods. It has been proven by simulation that using the natural gradient accelerates the convergence and reduces the residual mean square error (MSE) significantly.

The natural gradient rule takes into account the geometrical property of the NLN manifold which is a Riemannian Manifold. It has the forme

$$\widetilde{\nabla} L = A^{-1} \nabla L$$

where A^{-1} is the inverse of the Fisher Information Matrix (FIM) and ∇L is the ordinary gradient descent of the cost function $L = \|e(n)\|^2$. In the case of Euclidean manifolds $A = I_d$ (the identity matrix), so $\widetilde{\nabla} L = \nabla L$. The FIM can be calculated from ∇L by the rule:

$$A = E\left[\frac{\nabla L}{\nabla \theta}\left(\frac{\nabla L}{\nabla \theta}\right)^T\right]$$

where θ is the vector of the sub-network coefficients $\theta = \left[W_1^T, B^T, W_2^T\right]^T$.

4. SIMULATION RESULTS

In this section we present identification results of the considered downlink channel (Fig. 3). Those simulations are made in base band. The learning signal is a uniformly distributed white noise.

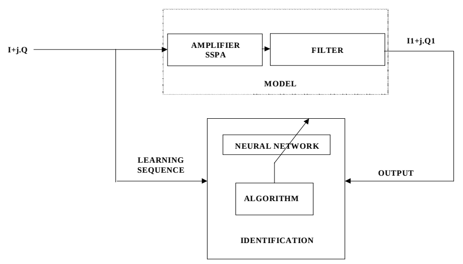

Fig. 3. Identification of a down link channel

We have observe experimentally that it is possible to identify this system only if we normalize the SSPA with the weight W_0. This is true for both used algorithms (OGD or NGD). Fig. 4 shows the mean square error (MSE) evolution of the I component with and without the weight W_0 for OGD. If we use that weight ($W_0 = 80$) we obtain at the end of convergence an $SNR = 30$ dB between the learning and the error signals. If we do not use it ($W_0 = 1$) then $SNR = 11$ dB. So, the weight W_0 decreases the MSE convergence by 19 dB and increases the convergence speed eff iciency.

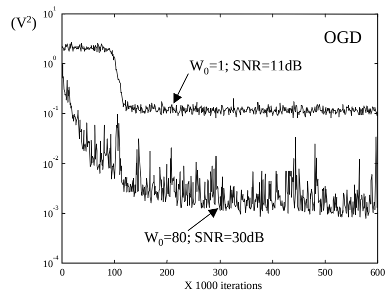

Fig. 4. Mean square error evolution with and without W_0

Fig. 5 shows the MSE evolution ($W_0 = 80$) of the I component for both algorithms: the OGD and the NGD. At the end of the learning sequence we obtain for the OGD: $SNR = 30$ dB. With the NGD, we obtain: $SNR = 54$ dB. So, the NGD decreases the MSE by 24 dB. This is the first advantage of the NGD.

The second advantage of the NGD is the convergence speed. Indeed, on Fig. 5 we see that OGD need around 250.10^4 iterations to converge whereas the NGD has nearly converge in around 50.10^4 iterations.

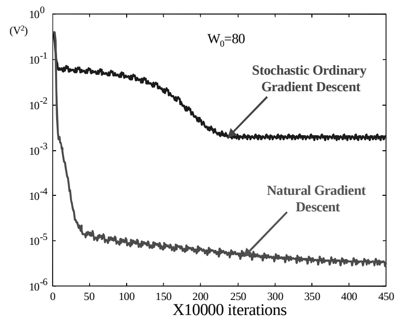

Fig. 5. Mean square error evolution with OGD and NGD

Fig. 6 shows the normalized AM/AM conversion. In Fig. 6a (Fig. 6b), we superpose the SSPA and the NN with NGD (OGD). It is clear when we look both figures that the identification is better with the NGD algorithm.

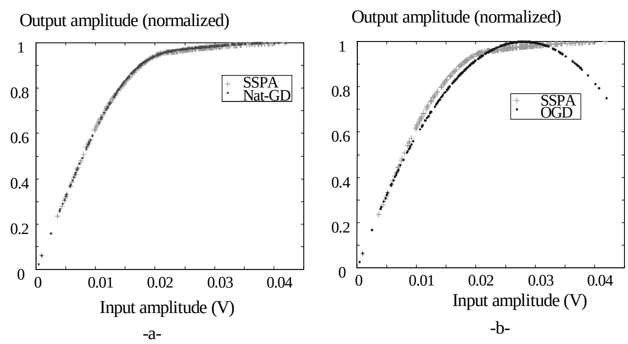

Fig. 6. Comparison of the AM/AM Identification with OGD and NGD

Fig. 7 shows the AM/PM conversion. It is also clear that the identification is better with the NGD algorithm (Fig. 7a).

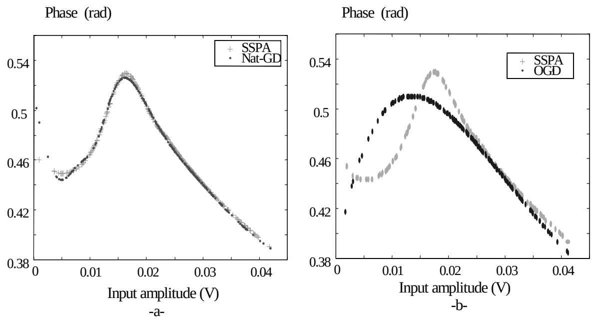

Fig. 7. Comparison of the AM/PM Identification with OGD and NGD

Fig. 8a and 8b present respectively the module and phase of the transfer function filter and LN with NGD (the same result is obtained by applying the OGD algorithm). The LN is well fitted to the filter especially between -100 MHz and 100 MHz.

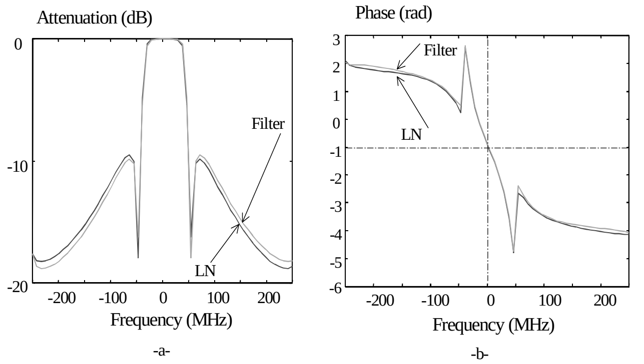

Fig. 8. Identification of the filter

5. CONCLUSION

We have identified a non-linearity (SSPA) followed by a memory (filter) with a NN that belongs to the MLP family, associated with two different algorithms (OGD and NGD). We have shown that for all algorithm used, the NN converge only if the SSPA is normalized by the weight W_0.

Moreover, we have shown that the NGD presents best results than OGD on two points: the SNR after convergence and the convergence speed. For the first point we get 54 dB with the NGD and 30 dB with the OGD. So, the identification of the SSPA is highly efficient with the NGD algorithm (Fig. 6 an Fig. 7). For the second point, the convergence speed of the NGD is around five times better than the OGD (Fig. 5).

Finally we identify separately each element of the channel. The NLN identify the SSPA and the LN identify the filter.

6. REFERENCES

[1] S. Bouchired, M. Ibnkahla, D. Roviras and F. Castanié, ”Neural network Equalization of Satellite Mobile Communication Channels”, Procs. of ACTS Mobile Communications Summit ’97, Aalborg (Danemark), Octobre 1997.

[2] S. Haykin, Neural Networks: a comprehensive foundation, IEEE Press, 1994.

[3] M. Ibnkahla, N. Bershad, J. Sombrin and F. Castanié, ”Neural network modelling and identification of nonlinear channels with memory: algorithms, applications and analytic models”, IEEE Trans. Signal Processing, Vol.46, N°5, pp.1208-1220, May 1998, Addendum to Special Issue on Applications of Neural Networks to Signal Processing.

[4] F. Langlet-Cauët, M. Ibnkahla and F. Castanié, ”Neural Network Hardware Implementation: Overview and Applications to Satellite Communications”, Proc. of DSP’98, Nordwick, Sept. 98.

[5] H.H. Yang and S.I. Amari, “Training Multi-Layer Perceptrons by Natural Gradient descent,” In ICONIP’97 Proceeding, new Zeland.