Early stopping criteria for adaptive training of dynamic nonlinear behavioural models

Méabh Loughman, Ronan Farrell and John Dooley

Department of Electronic Engineering, Maynooth University, Co. Kildare

Abstract—As the physical makeup of cellular basestations evolve into systems with multiple parallel transmission paths the effort involved in modelling these complex systems increases considerably. One task in particular which contributes to signal distortion on each signal path, is the power amplifier. In power amplifier (PA) modelling, Recursive Least Squares (RLS) has been used in the past to train Volterra models with memory terms. The Volterra model is widely used for modelling of PAs. In this paper we present a comparison of the stability performance for a PA model during training for various model memory lengths, model orders of non linearity and signal sample rates. This examination provides a technique to avoid instability occurring during the adaptive training of dynamic nonlinear behavioural models.

I. Introduction

Radio transceivers are increasing in complexity as a result of a more demanding and technology driven population [1]. There are a number of basestation architectures proposed to meet the demand for such increases in data rates. Power Amplifiers (PA's) are a crucial component in the architecture of a transceiver chain, responsible for increasing the power of a signal to ensure accurate and reliable transmission. PA's display non linear behaviour - accuracy and efficiency are a historical trade off. Distributed arrays of PAs are needed for many of the proposed architectures for future wireless networks. Behavioural modelling of PA's is an essential technique of testing PA behaviours under various circumstances. Extracting accurate models reduces the cost of hardware implementation testing.

Behavioural modelling of PA's is crucial in RF system modelling. One main objective in modelling these systems is to identify the most computationally efficient structure which can accurately characterise the behaviour. The Volterra series calculates each interaction of its inputs up to a defined order of non-linearity, as the number of inputs or the order of non-linearity increases the number of coefficients increases rapidly. Rules of thumb have been used in order to quantify the non-linear orders and methods such as Lipschitz numbers can be employed to dimension the memory depth needed for the PA model [2]. Similarly, more compact model structures derived from the Volterra model, such as the memory polynomial model or modified Volterra series, can be capable of modelling complex non-linear systems with memory. A system with finite order of non-linearity with finite memory depth can be described in the time domain compactly by equation 1.

$$y(n) = \sum_{p=1}^{P} y_p(n) \qquad (1)$$

Where,

$$y_p(n) = \sum_{i_1=0}^{N-1} \cdots \sum_{i_P=0}^{N-1} h_p(i_1, \cdots, i_p) \prod_{i=1}^{p} x(n - i_r)$$

Where $x(n)$ and $y(n)$ is the input and output signal to the system respectively. $h_p(i_1, ..., ip)$ represents the filter coefficient expansion utilising, p, the highest order for the non-linearity of the Volterra series expansion. N represents the maximum memory tap length chosen [3].

Small increments in either the order of nonlinearity or memory depth can greatly increase the number of model coefficients, which in turn can severely reduce the computational efficiency of the model as the nonlinear order or memory depth of the system to be modelled increases. However, the opposite scenario where the nonlinear order or memory depth can be reduced will greatly reduce the computational effort required. As research into wireless communication networks progresses beyond 5G standards, researchers must be aware of how all characteristics may be modelled in order to maximise system efficiency, reducing latencies in PA characterisation and implementation of adaptable digital pre-distortion is also desired. One drawback with adaptive techniques is the instability of the coefficients during the training of the model.

The following aspects are considered in this paper: a Volterra model's ability to capture the input to output relationship of experimentally measured PA signals with a suitable accuracy. Secondly, the effect on the model accuracy when higher and lower sample rates are used for the signals. Finally the effect of memory length on the model's training stability when an adaptive training routine is employed.

II. The Power Amplifier Model

Under certain operating conditions PAs exhibit dynamic and non linear operation. As such they may be closely modelled using non linear filters [4], [5], due to their similarity in behaviour. In order to model a nonlinear PA with memory
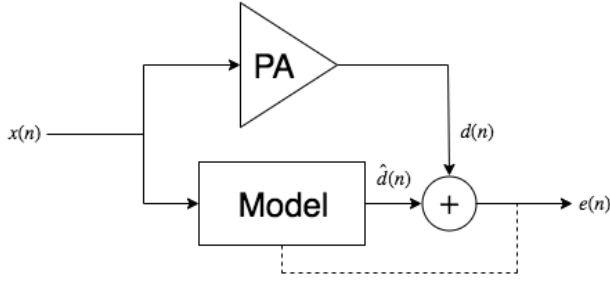
Figure 1. PA model

effects, a suitable order of non-linearity, memory length and a training routine must be selected. In order to test a cumbersome adaptive non-linear filter model, the Volterra series was selected for this work and the adaptive training routine to learn the model coefficients is Recursive Least Squares (RLS).

The operation of a digital filter is to take input signal samples $x(n)$, combine the current sample and previous samples, weight all of them and sum the resulting products. Similarly a non linear filter contains weights $\boldsymbol{H}(n)$, which are multiplied by the selection of input samples and the various combinations of the samples. The objective is for the output $\hat{d}(n)$ to be adequately comparable to the desired output signal $d(n)$. The system alters parameters of the model, continuously updating for each new sample point, to minimise the error function given by $e(n) = d(n) - \hat{d}(n)$.

A. Training algorithm and figure of merit

RLS was chosen as the training method alongside the Volterra model for this paper [6]. The exponentially weighted RLS algorithm can be adequately described in terms of its cost function as RLS filter coefficients, $\vec{\boldsymbol{H}}(n)$ adapt based on the cost function $J(n)$, which can be seen in equation 2.

$$J(n) = \sum_{k=1}^{n} \lambda^{(n-k)} (d(k) - \vec{\boldsymbol{H}}^T(n)\boldsymbol{X}(k))^2 \qquad (2)$$

Where λ is an exponentially weighted factor, $0 < \lambda < 1$, controlling the convergence speed of the function, referred at times as the forgetting factor. RLS filter coefficients are determined such that the weighted average of the squared estimation error is minimised from time $k = 1$ to $k = n$ [6].

Root Mean Squared Error (RMSE) was chosen as the figure of merit for this investigation due to its computational efficiency. RMSE can be calculated by equation 3.

$$RMSE = \sqrt{\left| (d(n) - \hat{d}(n)) \right|^2} \qquad (3)$$

RMSE is a commonly used figure of merit which is used in this work to provide a measure for the goodness of fit between a theoretical model and experimentally measured signal. It has been used previously in studies in which the relative performance of neural networks is compared [7] [8].

III. THEORETICAL COMPUTATION

Eigenvectors are vectors that have both direction and magnitude. The direction of the eigenvectors tend towards the point of convergence [9], between two data points, or sampled points in the case of this paper. Once the point of convergence, or minimum error, is reached the eigenvectors and values may become oscillatory [10], continuing to attempt to point in the direction of largest variance. In this paper we use observations of this relationship as an indicator of the onset of instability during the training of model coefficients.

The input signal training data length, before the onset of instability, is reduced when more memory taps are added to the model. The decrease in the length of the training before the onset of instability can be traced to $\boldsymbol{C}(n)$. $\boldsymbol{C}(n)$ is iteratively updated using $\boldsymbol{C}(n) = \lambda \boldsymbol{C}(n-1) + \boldsymbol{R}_{xx}$

Where $\boldsymbol{C}^{-1}(n)$ represents the auto-correlation matrix of the input signal, and $\boldsymbol{P}(n)$ is the cross-correlation vector of the input signal $x(n)$ and desired signal $d(n)$, both are exponentially weighted correlation matrices that are updated continuously to provide consecutive values of $\vec{\boldsymbol{H}}(n)$ [6]. For every update of $H(n)$, the more memory taps that are present in the model being trained, the more eigenvectors of $\boldsymbol{C}(n)$ are summed together. As the training routine advances to reach the minimum error $e(n)$ at a certain point the summation of the eigenvectors will tend towards the next area of greatest variance which will be an unstable condition.

Since there are no set limits for the number of memory taps which can be added to the nonlinear model, care must be taken to avoid the onset of instability during training.

The signals examined in Figure 5 were that of the same order of nonlinearity, only differing in memory length. The memory tap lengths used in the models trained cause an un-rectifable oscillation, and this is a demonstration of the reduction of the input training signal length before instability. Table III depicts the maximum signal lengths before instability for a range of memory tap lengths. The non linear order of the model and sampling frequency was maintained for the results shown in Table III. It is evident from the plot that there exists a relationship between the input signal training length reduction to the memory tap length. The influence of increasing the number of memory taps is much greater than that caused by increasing the order of non linearity of the model seen in Table I.

Where $\boldsymbol{R}_{xx}$ is the input signal auto-correlation matrix given by $\boldsymbol{R}_{xx}$, or, $E\{x(n)x^T(n)\}$. With a memory length of one, the auto-correlation matrix takes into account one previous

value of $x(n)$ and $x(n-1)$. Using N memory taps leads to an accelerated convergence of the minimum error.

Aggregating the new values of $\boldsymbol{C}(n)$ causes summation of the eigenvalues, which represent the variance of the data along the eigenvector directions. When the minimum error value, or point of convergence, is reached the eigenvalues continue to alter the values of $\vec{\boldsymbol{H}}(n)$ to orient to the direction of greatest variance. As the eigenvalues are aggregated, this leads to a sharp increase in the error signal $e(n)$ toward the next area of greatest variance, to a unit vector scalar multiple of the largest eigenvector, which converges to ∞ [11].

$$\boldsymbol{C}(n) = \sum_{1}^{N} \sum_{k=1}^{n} \lambda^{(n-k)} x(n) x^T(n)$$

A limit for convergence can be set as:

$$0 < \sum_{M=1}^{N} T\{\boldsymbol{R}_{xx}\} < \frac{2}{\lambda_L} \quad (4)$$

Where $T\{\boldsymbol{R}_{xx}\}$ is the summation of the diagonal eigenvalues of the auto-covariance matrices representing the convergence parameters of $\hat{d}(n)$. λ_L is the largest eigenvalue associated with the model equation 4 ensures that the eigenvectors remain oriented toward the initial area of greatest convergence, the minimum error, not permitting re-orientation to a unit vector scalar multiple of the largest eigenvector.

IV. EXPERIMENTAL RESULTS

This study utilises a wideband code division multiple access (WCDMA) 5MHz bandwidth single carrier signal sent through a Doherty PA at 2.6GHz. The input signal is set in order to drive the PA into saturation and both the input and output are sampled at 30.72 MHz. Figure 2 presents the transmitted and received signals by the Volterra model used in this investigation, given by equation 1.

Training an accurate PA model requires a training signal of adequate length. While the training signal length needs to be long enough to allow the algorithm to converge on values for model coefficients, the model may become unstable during training. Figure 3 displays the problematic nature of training a model while attempting to use too many sample points. The model depicted in Figure 3 was of order two and three memory taps. The model error estimate $e(n)$ becomes unstable after approximately $13,800$ time samples.

Data provided for this experiment was procured by training the model using an increasing signal length until it becomes unstable. The number of samples was then reduced by one hundred sample points until the model became stable again, RMSE was calculated at that point. Investigation was focused on identifying the signal training length for which the training becomes unstable. Noise is present in the measured signals
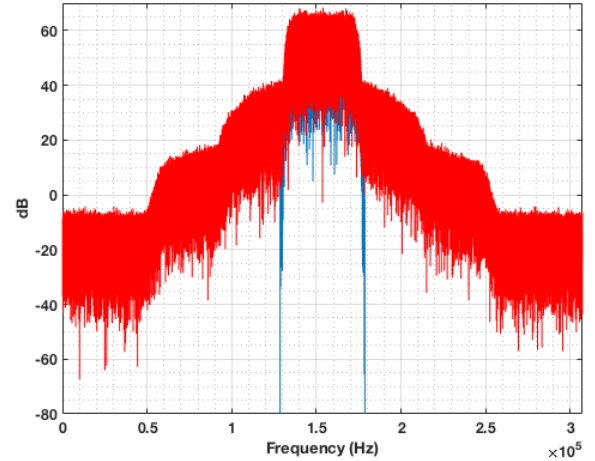

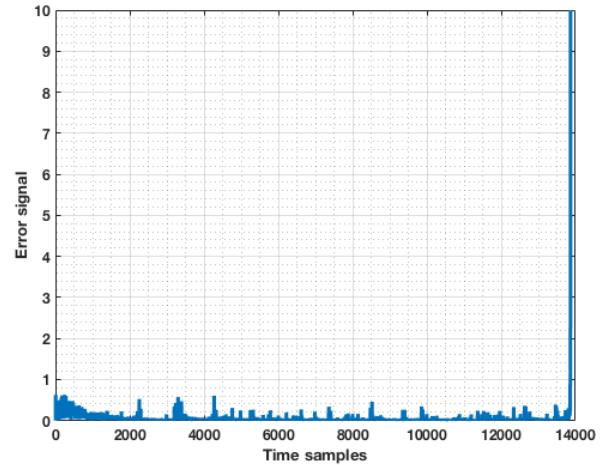Figure 2. Input and output signals sent through Doherty PA

Figure 3. Error from training versus time samples

however its effect on the overall performance of the model is negligible. In this work a limited set of experiments to transmit and receive signals from/to a PA were conducted. It is intended to carry out a more comprehensive set of experimental tests to obtain a more populated sample set than seen in 5.

A. Non linear order

Increasing the order of non-linearity of the model lead to a decrease in the length of the input training signal before the onset of instability during training. Figure 4 depicts the decrease in the training signal length due to non linear order, it shows the relationship for the model orders of nonlinearity from one to seven, and memory lengths, one to seven.

Table I shows the reduction of the maximum signal length needed to maintain stability with varying memory order.

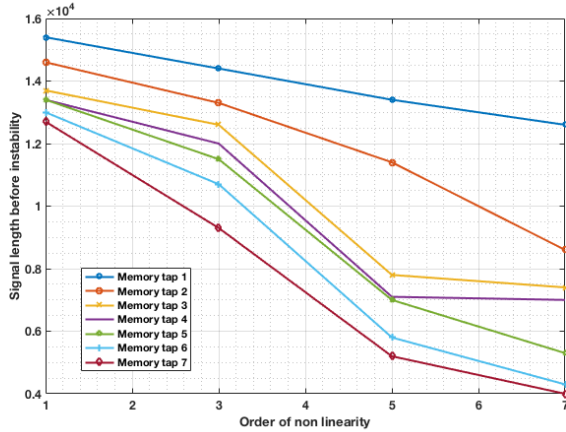| Non linear order	Maximum signal length(samples)
2	14100
3	12900
5	6600
7	6400

Figure 4. Maximum signal length before instability versus model order of non linearity

B. Sampling frequency investigation

Increasing and decreasing the sampling frequency altered the length of training signal before onset of instability. All sampling frequencies set in this work adhered to the Nyquist criterion. Table II presents the alteration in signal length stability utilising the same order, memory length and starting sample point for all of the sample rates tested.

Table II
SIGNAL LENGTH BEFORE INSTABILITY COMPARISON TO SAMPLING
FREQUENCY

Sample frequency(MHz)	Max signal length(samples)	RMSE
7.68	14200	0.0263
15.36	14100	0.0147
30.72	14200	0.0331
61.44	14100	0.0339
122.88	13800	0.0307

Table II presents the variation in the RMSE calculation between each signal sampling rate. Increasing the sample rate, intuitively, should reduce the error as the signal is sampled more frequently, this is not the case. Down-sampling the signal introduced a lower error at a more comparable training length.

C. Memory tap length

Increasing the memory tap length of the model, it is notable that the length of the signal before instability reduces greatly. Figure 5 compares the memory tap length effect on the training
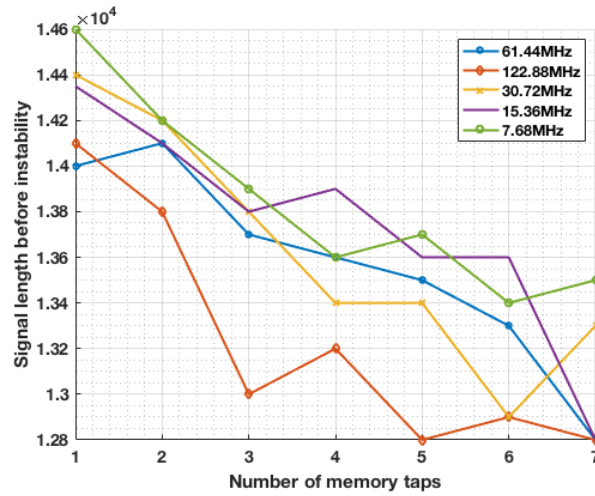

Figure 5. Maximum training signal length versus memory length

Table III
TRAINING SIGNAL LENGTH RELATIONSHIP TO MEMORY TAP LENGTH

Memory tap length	Maximum signal length (samples)
1	13400
2	11400
3	7800
4	7100
5	7000
6	5800
7	5200

signals length before instability. The RMSE of each signal varies but the number of training samples before the onset of instability decreases substantially.

V. CONCLUSION

In conclusion, this paper provides a definitive early stopping technique to avoid instability of a nonlinear dynamic model during training. The model parameter which has the largest influence on reducing the length of the training signal before the onset of instability, for adaptive nonlinear modelling, is the memory tap length. Model non-linear order and sampling frequency had an effect on the achievable RMSE but showed less variation in training signal length before instability than the model memory tap length.

VI. ACKNOWLEDGEMENT

REFERENCES

[1] K. Zheng, Z. Yang, K. Zhang, P. Chatzimisios, K. Yang, and W. Xiang, "Big data-driven optimization for mobile networks toward 5g," *IEEE Network*, vol. 30, no. 1, pp. 44–51, January 2016.

[2] J. Dooley, B. O'Brien, K. Finnerty, and R. Farrell, "Estimation of sparse memory taps for rf power amplifier behavioral models," *IEEE Microwave and Wireless Components Letters*, vol. 25, no. 1, pp. 64–66, Jan 2015.

[3] A. Zhu, J. Dooley, and T. J. Brazil, "Simplified volterra series based behavioral modeling of rf power amplifiers using deviation-reduction," in *2006 IEEE MTT-S International Microwave Symposium Digest*, June 2006, pp. 1113–1116.

[4] Z. Wang, J. Dooley, K. Finnerty, and R. Farrell, "Selection of compressed training data for rf power amplifier behavioral modeling," 09 2015.

[5] J. Dooley, B. O'Brien, and T. J. Brazil, "Behavioural modelling of rf power amplifiers using modified volterra series in the time domain," in *High Frequency Postgraduate Student Colloquium, 2004*, Sep. 2004, pp. 169–174.

[6] V. J. Mathews and G. L. Sicuranza, *Polynomial signal processing*. Wiley, 2000.

[7] J. Zhai, J. Zhou, L. Zhang, and W. Hong, "Behavioral modeling of power amplifiers with dynamic fuzzy neural networks," *IEEE Microwave and Wireless Components Letters*, vol. 20, no. 9, pp. 528–530, Sep. 2010.

[8] A. Zhu, M. Wren, and T. J. Brazil, "An efficient volterra-based behavioral model for wideband rf power amplifiers," in *IEEE MTT-S International Microwave Symposium Digest, 2003*, vol. 2, June 2003, pp. 787–790 vol.2.

[9] P. S. R. Diniz, *Adaptive filtering*. Springer, 2006.

[10] B. Farhang-Boroujeny, *Adaptive Filters: Theory and Applications, 2nd Edition*. John Wiley Sons, 2013.

[11] B. Steinberg, *Representation Theory of Finite Groups [recurso electrónico]*. Springer New York, 2012.