ROBUST PARAMETER TRACKING THROUGH REGIONAL FORGETTING

Robert Shorten, Andreas Schütte, A.D.Fagan†*

Daimler-Benz Research, Intelligent Systems Group, Alt-Moabit 91b, 10559 Berlin, Germany
E-mail:shorten@DBresearch-berlin.de.

ABSTRACT

The recursive least squares (RLS) algorithm with exponential forgetting (λRLS) is perhaps the best known and most widely used algorithm for tracking the time varying parameters of a linear regression model. The implicit assumption in using the λRLS algorithm is that the information is uniformly distributed over the time horizon. Frequently this assumption does not hold and serious difficulties can be encountered when using many model structures. These include convergence of the parameters to local system or noise characteristics and output bursting, i.e. a large error when the operating point changes. In this paper several simple alternatives to the standard λRLS algorithm are proposed. The proposed algorithms extend the idea of a sliding window by quantising the whole input space.These algorithms considerably reduce the risk of forgetting useful information and eliminate the possibility of output bursting by relating the adaptation capabilities of the algorithm to the amount of input stimulation. Simulation results confirm the efficacy of our approach.

1. Introduction

The problem of tracking a time varying system consists of specifying suitable model structures and the development of parameter estimation algorithms which can track the model parameters in a suitable fashion. Due to increases in low cost computational power and memory, the λRLS [1] [2] algorithm has become an attractive alternative to the popular least mean squares LMS algorithm. Under certain conditions the λRLS algorithm offers fast convergence to the optimal parameter set for linear regression structures. Examples of such model structures include standard linear models, NARMAX models [3], polynomial models, and radial basis function (RBF) networks with predetermined hidden layer parameters [4]. The λRLS algorithm is

$$\begin{aligned} \theta(t) &= \theta(t-1) + P(t)\psi(t)\epsilon(t), \\ \epsilon(t) &= y(t) - \theta^T(t-1)\psi(t), \\ P(t)^{-1} &= \lambda P(t-1)^{-1} + \psi(t)\psi(t)^T \end{aligned} \qquad (1)$$

where ϵ is the prediction error, P is the update gain matrix, ψ is the vector of regressors, and θ is the vector of model parameters. The term λ is called the forgetting factor. It is used to discard old information. With λ equal to 1 no information is lost.

Unfortunately the λRLS algorithm will only work well subject to certain assumptions. The most important of these are that older data is in some sense less reliable than more recent data. This is exactly the case if the system depends on the system state. If this assumption is not valid and the model structure is non-local, i.e. linear models, RBF's with wide basis functions, then serious problems can be encountered. These include convergence to local system behaviour due to local input signal excitation and bursting due to an exponential growing of the P matrix, i.e. a sudden output burst due to change in operating point [5].. These problems have been considered for the case of linear systems from the point of view of adapting the standard λRLS algorithm [6] [7] [8] [9] [10], and by [4] for the RBF structure.

In this paper we propose a number of algorithms which conceptually reduce the risk of forgetting useful information by retaining stimuli from across the input space. Essentially this means that the restriction on the shape of the window of past values is relaxed. These algorithms also eliminate the possibility of output bursting.

2. Definitions and Basic Assumptions

In the following discussion the system being modelled is referred to as the non-linearity. The class of systems considered is depicted in figure 1. All feedback variables are assumed to be known and the space consisting of the external inputs and feedback outputs is termed the input space. A uniformly exciting input signal is defined to be an input signal capable of exciting all regions of the input space that have been modelled. It is recognised that when using

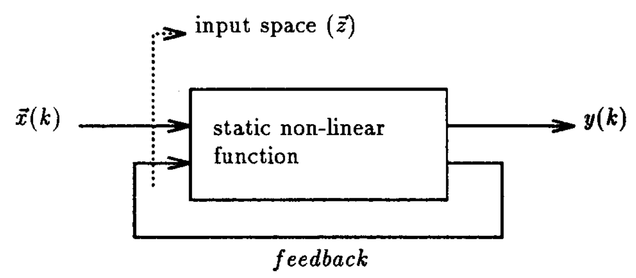

Figure 1: Modelling approach

*Also member of DSP group, University College Dublin, Belfield, Dublin 4, Ireland

†Member of DSP group, Belfield, Dublin 4, University College Dublin, Ireland

0-7803-2431-5/95 $4.00 © 1995 IEEE

the λRLS algorithm the nature of the input signal is of the utmost importance. Even when the non-linearity is time invariant the standard λRLS algorithm can fail if the input signal is not uniformly exciting. In this paper the most general case of a time varying non-linearity with non-uniform excitation is considered. In order to develop a solution to the problem the following assumptions are made:

- The non-linearity is modelled using a linear regression model,
$$Y = A\theta, \tag{2}$$
where Y is a Lx1 vector of system outputs, θ is the Kx1 parameter vector and A is a matrix whose rows consist of each regressor vector.
- It is assumed that the model structure is sufficiently complex and general to capture the changing non-linearity.
- It is also assumed that the time changes of the system are slow, i.e., the changes in the system are much slower than the excitation of the input space. This effectively ensures the time varying parameters are of a type which can be tracked.

3. Proposed Algorithms

In this section two approaches for the least squares fitting of a data set to a given model are proposed. These are motivated by the desire to restrict the amount of forgetting by relating the adaptive capabilities to the degree of system excitation, i.e, only forget in areas where new information is arriving.

3.1 Regionalised fitting algorithm

In developing this algorithm it is recognised that if at each time step enough information exists concerning the overall state of the system then the parameter set could be estimated using
$$\theta(k) = A^{+}Y, \tag{3}$$
where $\theta(k)$ is the estimated parameter set at time k, A^{+} the pseudoinverse of the regressor matrix, and Y is a matrix of system outputs. In the sequel β denotes a learning rate parameter. We propose a general algorithm, consisting of an output and input adaptation phase, as follows:

- *Initialisation Phase:*
  1. Vector quantise the input space to form a set of input space representatives (ISR).
  2. Select an initial output for each ISR.
- *Adaptation Phase:*
  1. At each time step update the ISR (r_i) nearest to the input vector. Use the current output to update the output associated with the i'th representative
  2. Use all representatives to calculate A and Y in equation 3.
  3. Calculate $\theta(k)$ using equation 3

How the input space representatives and their associated outputs are updated determines the plasticity and robustness of the algorithm.

3.1.1 ISR Output Adaptation

Two methodes for updating the outputs associated with each representative were investigated.

1. *Local exponential forgetting of ISR outputs*
The representative outputs are given by
$$\hat{y}_i(k) = \beta\hat{y}_i(k-1) + (1-\beta)y(k), \tag{4}$$
where $\hat{y}_i(k)$ denotes the output associated with ISR i at time $t = k$ and $y(k)$ denotes the most recent measurement. Note $\beta = 1$ corresponds to output replacement.
2. *Local windowing*
In this case a window in time of size M is created for each representative. At each time step the current output replaces the output M time steps in the past. The output representative is then given by
$$\hat{y}_i(k) = \frac{1}{M}\sum_{j=1}^{M} y_j, \tag{5}$$

Local windowing is robust with respect to noise but is not as elastic as local exponential forgetting.

3.1.2 Adaptation of ISR locations

Two strategies for updating the locations of the input space representatives at each time step were investigated.

1. *Fixed representatives*
The representatives remain fixed. This technique is advantageous since only the Y vector in equation 3 needs to be updated.
2. *Local windowing of ISR Locations*
In this case the representatives are updated according to
$$\vec{r}_i(k) = \frac{1}{M}\sum_{j=1}^{M} \vec{z}_j, \tag{6}$$
3. *Local Exponential Forgetting of ISR locations*
In this case the representatives are updated according to,
$$\vec{r}_i(k) = \beta\vec{r}_i(k-1) + (1-\beta)(z(k) - \vec{r}_i(k-1)), \tag{7}$$
where the $\vec{r}_i$ denotes the i'th input space representative and $z(k)$ is the input closest to representative i at time $t = k$.

The local forgetting algorithm results in a more elasic structure. However care shold be paid to ensure that the representatives do not drift toward one part of the input space. In this case the development of a more sophisticated adaptation algorithm may be neccessary. Also when using a local windowing technique with exponential forgetting of the ISR locations/outputs, particular care should be taken to assign correct learning rates.

3.1.3 Adding new represantitives

In the case where the need to introduce new representatives arises, this can be accomplished by replacing the representative which was not been updated for the longest time interval with the new representative. Robustness with respect to noise can be achieved by only inserting the representative point into the regressor after it has a certain number of data points associated with it.

3.2 Regionalised λRLS Algorithm

In this section a variant on the standard λRLS algorithm is presented. The basic idea is to restrict the amount of forgetting by decomposing the P^{-1} matrix into a number of submatrices as follows

$$P^{-1} = \alpha I + \sum_{i=1}^{i=N} B_i. \tag{8}$$

Each of the matrices B_i are associated with a region of the input space. At each time step only the B_i associated with the current input vector is updated according to

$$B_i(k) = \lambda B_i(k-1) + \psi(k)\psi(k)^T, \tag{9}$$

thus ensuring that only the part of the P matrix associated with the current innovation is updated. This has the effect of slowing the rate of adaptation if the input signal spends any amount of time in a local region. Bursting is avoided by ensuring that the P^{-1} converges to $\frac{1}{\alpha}I$.

In addition the local P matrix is available at any time. Note that equation 8 can be written in terms of P and solved recursively. Partitioning the P matrix corresponds to quantising the error space over which the least squares optimisation is carried out. It should also be noted that equation 8 is proposed as an alternative for determining the parameters of a global model. It offers no advantage over a true local structure (RBF) with local learning [11]. Similar ideas termed "directional forgetting" have been presented in [8] [9] [10].

4. Examples

Results demonstrate that the algorithms presented above significantly reduce the risk of forgetting useful information. Furthermore the algorithms retain their ability to adapt by relating adaptation to input space excitation.

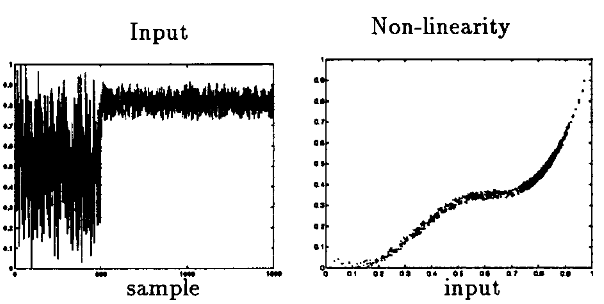

Figure 2: Input signal and non-linearity

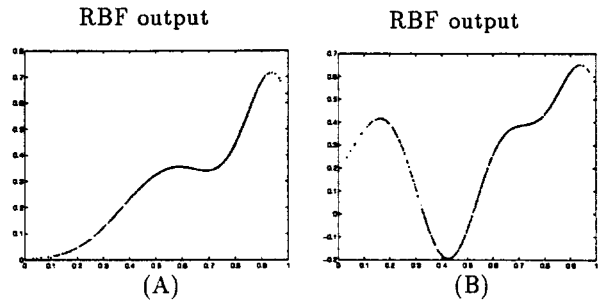

Figure 3: RBF trained with λRLS ($\lambda = 0.99$). A) Output after 500 samples, B) Output after 1500 samples.

4.1 Example 1

The effect of using a λRLS trained RBF network to learn a non-linearity is shown in figure 3. The non-linearity is described by

$$y = (x-0.5)^3 + v, \tag{10}$$

where v is uniform white noise $U(-0.1, 0.1)$ and the output was normalised to lie in the interval $(0,1)$. The input is uniformly exciting white noise for the first 500 samples and local for the next 1000 samples. The input signal and non-linearity are shown in figure 2. The RBF network had 5 units. The same network was trained using a fixed representative mesh (2 per basis function) and equation 6 used to adapt the output assigned to each input representative. The resulting network outputs can be seen in figure 4 to be more robust with respect to information forgetting.

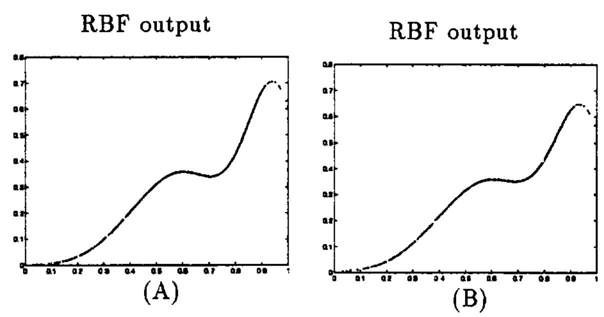

Figure 4: RBF trained with proposed algorithm. A) Output after 500 samples, B) Output after 1500 samples.

4.2 Example 2

In this example the regionalised λRLS algorithm was used to fit a linear model of the form

$$\hat{y}(k+1) = \theta_1 y(k) + \theta_2 u(k), \tag{11}$$

to data generated by,

$$y(k+1) = 0.35y(k) + u(k) + 0.25u(k)^2 + v, \tag{12}$$

where v is $U(-0.1, 0.1)$. The input signal u is uniform white noise and is depicted in figure 5.

For simulation purposes the P matrix was partitioned into 4 submatrices defined uniformly over the space defined by the external input u.

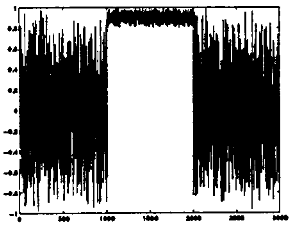

Figure 5: Input signal for example 2

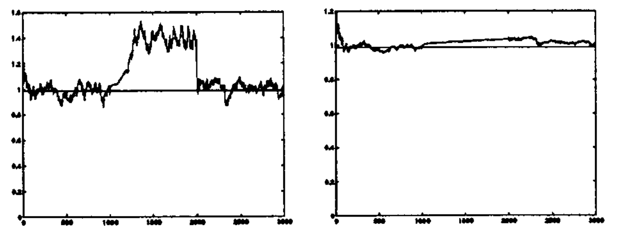

Figure 6: Evolution of θ_2.

The parameter set determined off-line using all the data yielded a model of the form

$$\hat{y}(k+1) = 0.4733y(k) + 0.9897u(k). \tag{13}$$

This parameter set is optimal in the sense that it yields the best linear model given all of the data. Application of the standard λRLS algorithm with $\lambda = 0.98$ yielded at the 2000'th sample a model of the form,

$$\hat{y}(k+1) = 0.3435y(k) + 1.3461u(k), \tag{14}$$

whereas the regionalised version of the algorithm yielded,

$$\hat{y}(k+1) = 0.4951y(k) + 1.0353u(k), \tag{15}$$

thus clearly demonstrating the robustness of the algorithm. The λRLS algorithm has converged to local uncertainty and the noise characteristics. This effect can be dangerous as it can lead to large prediction errors. It should be noted that the regionalised λRLS algorithm offers a conceptually simple alternative to the algorithms proposed in [9] [10]. For some applications the memory burden of the above algorithm may be great.

5. Conclusions

Several algortihms for the robust tracking of the parameters of a linear regression model have been presented in this paper. The authors suggest replacing the recursive least square algorithm with exponential forgetting, with recursive pseudoinverse algorithms. Clearly in many applications this is a possible since the main motivation for using the λRLS algorithm, i.e. limited memory, no longer applies. Within this framework various adaptation strategies can be applied, some of which are outlined above.

Acknowledgements

This work was in part sponsored by the European Union Human Capital and Mobility programme (HCM), contract number ERBCHBICT930711. The authors also wish to acknowledge the many helpful comments made by Dietmar Neumerkel and Daniel Sbabaro-Hofer.

6. References

[1] L. Ljung and S. Gunnarsson, "Adaptation and tracking in system identification - a survey," *Automatica*, vol. 26, no. 1, pp. 7–21, 1990.

[2] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification.* The MIT Press, 1983.

[3] S. Chen and S. Billings, "Representations of non-linear systems: The NARMAX model," *International Journal of Control*, vol. 49, no. 3, pp. 1013–1032, 1989.

[4] O. Dubois, N. Nicolas, and A. Billat, "Adaptive Neural Network Control of the Temperature in an Oven," in *Proceedings of the IEE Colloquium on Advances in Neural Networks for Control and Systems*, 1994.

[5] K. Astrom and B. Wittenmark, *Adaptive Control.* Addison-Wesley, 1989.

[6] P. Andersson, "Adaptive forgetting in recursive identification through multiple models," *International Journal of Control*, vol. 42, no. 5, pp. 1175–1193, 1985.

[7] L. Fortescue, T.R.and Kershenbaum and B. Ydsie, "Implementation of self-tuning regulators with variable forgrtting factors," *Automatica*, vol. 17, no. 6, pp. 831–835, 1981.

[8] R. Kulhavý and M. Kaŕný, "Tracking of slowly varying parameters by directional forgetting," in *Proceedings of IFAC 9th Triennial World Congress, Budapest, Hungary*, 1984.

[9] R. Kulhavý, "Restricted exponential forgetting in real-time identification," *Automatica*, vol. 23, no. 5, pp. 589–600, 1987.

[10] T. Hägglund, "Recursive estimation of slowly time-varying parameters," in *Proceedings of IFAC 1985, York, U.K.*, 1985.

[11] R. Murray-Smith, "Local model networks and local learning," in *Fuzzy Duisberg'94, pp. 404-409*, 1994.