Optimal piecewise locally linear modelling

C. J. Harris*, X. Hong, M. Feng
ISIS Group, Dept of Electronics and Computer Science
University of Southampton, Southampton SO17 1BJ, UK

ABSTRACT

Associative memory networks such as Radial Basis Functions, Neurofuzzy and Fuzzy Logic used for modelling nonlinear processes suffer from the curse of dimensionality (COD), in that as the input dimension increases the parametrzisation, computation cost, training data requirements, etc. increase exponentially. Here a new algorithm is introduced for the construction of a Delaunay input space partitioned optimal piecewise locally linear models to overcome the COD as well as generate locally linear models directly amenable to linear control and estimation algorithms. The training of the model is configured as a new mixture of experts network with a new fast decision rule derived using convex set theory. A very fast simulated reannealing (VFSR) algorithm is utilised to search a global optimal solution of the Delaunay input space partition. A benchmark non-linear time series is used to demonstrate the new approach.

Keywords: Delaunay input space partition, convex set, very fast simulated reannealing.

1. INTRODUCTION

Nonlinear systems can be approximated by piecewise locally linear models via local linearisation[1,2]. Not only can a multiple local modelling approach be more efficient in capturing the real system dynamics than a single global nonlinear model[1,3], but also piecewise locally linear models are advantageous in that well-known linear control and filtering methods are directly applicable to the identified model. In a piecewise linear model, the input space is partitioned into a set of (usually orthogonal) regions. The local models that operate on these regions are identified separately and the system output is based on a compilation of the local models to obtain a good global approximation of the real system.

The approximation performance of a piecewise linear model is strongly influenced by how the input space is partitioned. Two main disadvantages of a lattice partition of input space are the curse of dimensionality, and the problem of discontinuities resulting in large approximation error possibly leading to instability. To overcome these problems, a mixture of expert network (MEN) has been introduced to form smooth piecewise linear models using a Delaunay partition of the input space[4], a popular domain partition method in numerical analysis which is efficient in forming irregular regions[5]. One of the main advantages of the MEN algorithm in dynamical system modelling is that the local models are formulated in a global form and consequently can be trained adaptively and on-line. At each time step, each local linear model is selected to be trained using a Min/Max decision rule in the mixture of experts network. However the application of the method is restricted due to an *a prior* assumption of known convexity of the output surface which determines the decision rules that are used to train the network. For a wider class of nonlinear system with multimodal output surfaces, global optimisation techniques may provide viable approaches to obtain the most appropriate Delaunay partitioned input space which is necessary to obtain smooth piecewise linear models with optimal approximation properties. A very fast simulated re-annealing (VFSR) method has been introduced for

*Correspondence: Email: cjh@ecs.soton.ac.uk; WWW: http://www.isis.ecs.soton.ac.uk; Tel: +44 (0)1703 592353; Fax: +44 (0)1703 594498

global optimisation of non-convex cost functions[6,7], and has been effectively applied in nonlinear signal processing applications[8]. In this paper, a new optimal piecewise local linear modelling algorithm is introduced. The algorithm consists of two training phases, global optimisation of input space partition using the VFSR procedure and local linear model training using conventional gradient descent such as the normalised least mean squares (NLMS) methods[9] with the introduction of a new mixture of experts network using a new decision rule derived from convex set theory. The VSFR is applied with a global optimisation cost function to find an optimal set of vertices of Delaunay triangulation that forms a set of subsets of input space in order to obtain an optimal piecewise linear model. The new MEN is formed with a new simple and fast decision rule, for the gate control in training each local model, which is derived using convex set theory and is applicable to high dimensional Delaunay input space partition. In consequence the advantage of fast on-line properties of MEN is extended to unknown output surface convexity problems.

The paper is orgainised as follows; within the framework of the MEN model, Section 2 presents the training algorithm of Delaunay input space partitioned piecewise linear models with the introduction of the new decision rule derived from convex set theory. Section 3 describes the application of VFSR algorithm to achieve a global smooth piecewise linear model via the optimisation of input space partition. Numerical examples are presented in Section 4 and the conclusions is given in Section 5.

2. THE NEW DECISION RULE FOR THE TRAINING OF PIECEWISE LINEAR MODEL

The mixture of experts network(MEN)[4], depicted in Fig.1, is a layered structure. The first layer of the network represents the individual expert who sends a report to the manager. The manager monitors and evaluates all the reports to form a final opinion of the system output and how the network is to be trained, by operating the gate g_i of each expert using some decision rules. In the piecewise linear modelling problem, the experts correspond to a set of local linear models. Consider a system $y = f(\mathbf{x})$ where y denotes system output, $\mathbf{x}$ is system input vector with an appropriate dimension and $f(\bullet)$ is described by a piecewise linear model. Denote the input space S: $\mathbf{x} \in S \subset \Re^n$. A piecewise linear model consists of a set of m local linear models with a set of m operating ranges $S^{(1)}$, $S^{(2)}$, $\cdots$, $S^{(m)}$ which satisfy

$$S^{(1)} \bigcup S^{(2)} \bigcup \cdots \bigcup S^{(m)} = S \tag{1}$$

and

$$S^{(j)} \bigcap S^{(k)} = \oslash \quad \text{when } j \neq k \tag{2}$$

where $\oslash$ denotes empty set. The model is written as

$$\hat{y}_i(\mathbf{x}) = [1 \quad \mathbf{x}^T]\hat{\Theta}^{(i)}, \text{ IF } \mathbf{x} \in S^{(i)}\ , \ i = 1, \cdots, m \tag{3}$$

where $\hat{\Theta}^{(i)} \in \Re^{n+1}$ denotes parameter estimate vector used in ith local linear model.

The output of the network is given by

$$\hat{y}(\mathbf{x}) = \sum_{i=1}^{m} g_i \hat{y}_i(\mathbf{x}) \tag{4}$$

with

$$g_i = \begin{cases} 1 & \text{IF } \mathbf{x} \in S^{(i)} \\ 0 & \text{otherwise} \end{cases}$$

The mixture of experts network is advantageous in the application of dynamical system modelling and time series where the training of the network is carried out in an adaptive and on-line processing manner if some fast decision rules for the control of the gate g_i is available.

Consider the operating ranges $\{S^{(i)}\}$ is formed using Delaunay triangulation (eg. a triangle in 2D, a tetrahedron in 3D) as $S^{(i)} = \mathcal{H}\{\mathbf{v}_0^{(i)}, \mathbf{v}_1^{(i)}, \cdots, \mathbf{v}_n^{(i)}\}$, where $\{\mathbf{v}_0^{(i)}, \mathbf{v}_1^{(i)}, \cdots, \mathbf{v}_n^{(i)}\}$ denote the vetices of $\{S^{(i)}\}$, by allowing some points freely within or on the hull of the input space to form these vertices, as illustrated in Fig.2. Each Delaunay triangulation connects $n+1$ vertices in the n-dimensional space. The free parameters determining the positions of these free points can be written as a vector $\omega \in \Re^p$ with an appropriate dimension p, which can be optimally determined using the VFSR algorithm described in Section 3. Note that with present method, the transformation from ω to $\{\mathbf{v}_0^{(i)}, \mathbf{v}_1^{(i)}, \cdots, \mathbf{v}_n^{(i)}\}$ must be explicitly defined by the modeller subject to the neighbour information

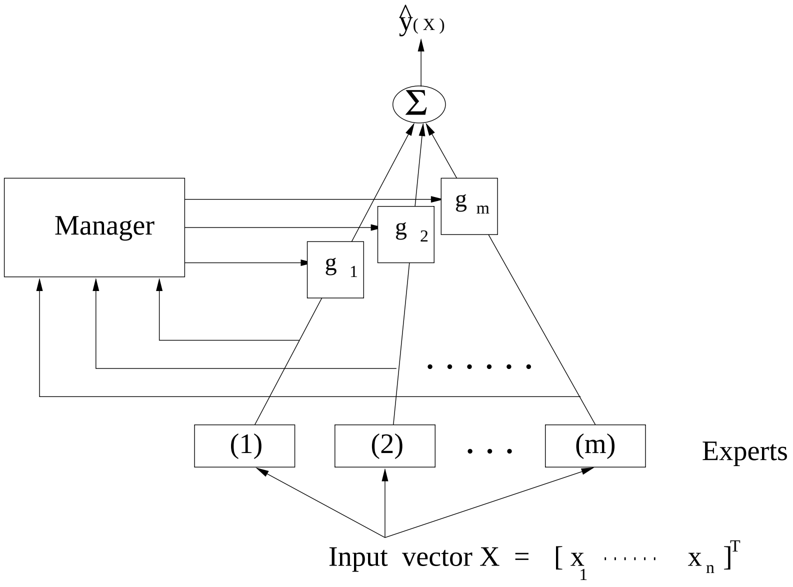

Figure 1: Mixture of experts network

produced when he or she defines and labels these Delaunay triangulations satisfying Eq.(1) and (2). These neighbour information involve a few constraints, such as, some of the vertices use the natural vertices of the input domain, some of the free point are constrained on the hull of input domain, and the vertices of different Delaunay triangulations use a coincident point. As a consequece of these constraints, the dimension size of free parameter ω can be reduced to a small number. For example, the input partition of Fig.2 involves only 8 free parameters. Hence the derivation of $\{\mathbf{v}_0^{(i)}, \mathbf{v}_1^{(i)}, \cdots, \mathbf{v}_n^{(i)}\}$ from ω would not incur much difficulty because it is always advised to use as small number of free points as possible in order to overcome the curse of dimensionality. For this purpose it is also recommended that preprocessing of variable selection is used to reduce n, the dimension of input vector $\mathbf{x}$, prior to modelling for very complicated systems using conventional algorithms such as projection pursuit, principal component analysis(PCA), canonical analysis or some special model construction algorithm[10,11,12].

It can be shown from convex set theory that any point $\mathbf{x}$ within $S^{(i)}$ is a unique convex combination of $\{\mathbf{v}_0^{(i)}, \mathbf{v}_1^{(i)}, \cdots, \mathbf{v}_n^{(i)}\}$ as[13]

$$\mathbf{x} = \sum_{j=0}^{n} \lambda_j^{(i)} \mathbf{v}_j^{(i)} \tag{5}$$

where $\lambda_j^{(i)} \geq 0$, $\sum_{j=0}^{n} \lambda_j^{(i)} = 1$. Denote $\Lambda^{(i)} = [\lambda_1^{(i)}, \cdots, \lambda_n^{(i)}]^T \in \Re^n$ and substitute $\lambda_0^{(i)} = 1 - \sum_{j=1}^{n} \lambda_j^{(i)}$ into Eq.(5), yielding

$$\Lambda^{(i)} = \left[\tilde{V}^{(i)}\right]^{-1} (\mathbf{x} - \mathbf{v}_0^{(i)}) \tag{6}$$

where $\tilde{V}^{(i)} = [\mathbf{v}_1^{(i)} - \mathbf{v}_0^{(i)}, \cdots, \mathbf{v}_n^{(i)} - \mathbf{v}_0^{(i)}] \in \Re^{n \times n}$. Hence the IF statement in Eq.(3) can be represented through m experts opinions given by Eq.(6). The manager chooses the gate g_i by using a new decision rule, that is, to check the solutions given by m set of $\Lambda^{(i)}$, $i = 1, \cdots, m$ and to select the only one that satisfies the constraints $\lambda_j^{(i)} \in [0, 1)$ for $1 \leq j \leq n$ and $\sum_{j=1}^{n} \lambda_j^{(i)} \in [0, 1)$, and then let $g_i = 1$.

The training of local linear models can be carried out adaptively in a on-line manner within a global framework using the mixture of experts network, and summarised as follows.

1. Determine the input space domain (the convex set S) constrained within the maximum and minimum value of the components of the input vector $\mathbf{x}$ by using data preprocessing. Given an appropriate number of points which located within or on the hull of the input space to form vertices of a set of m subranges (simplexes, Delaunay triangulation), denoted as $S^{(i)} = \mathcal{H}\{\mathbf{v}_0^{(i)}, \mathbf{v}_1^{(i)}, \cdots, \mathbf{v}_n^{(i)}\}$. Corresponding to each $S^{(i)}$, given a set of initial parameters for m local linear models $\hat{\Theta}^{(i)} \in \Re^{n+1}$, $i = 1, 2, \cdots, m$.

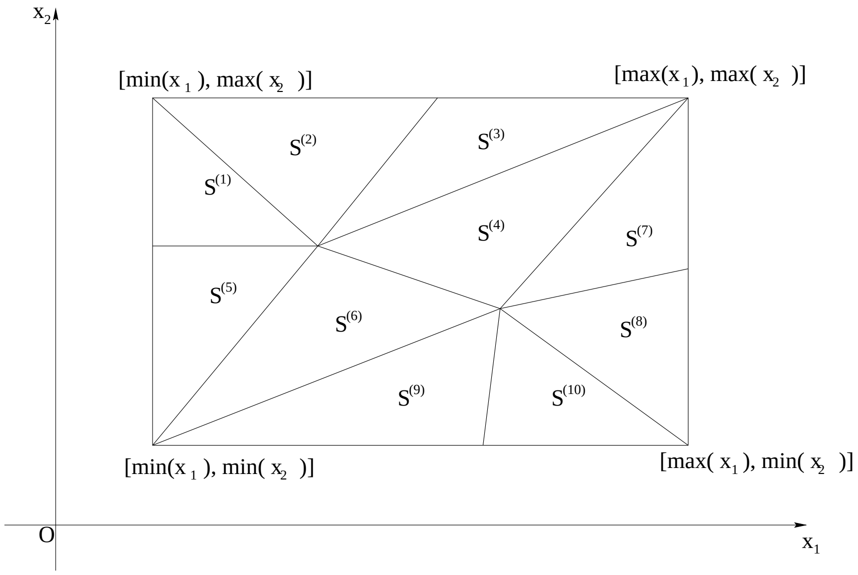

Figure 2: Delaunay partition of input space

2. At time step t, the training sample is given as $\{y(t), \mathbf{x}(t)\}$. Each expert opinion is formed using the solution of Eq.(6) (where $\mathbf{x}$ is replaced by $\mathbf{x}(t)$) and report to the manager. The manager select the ith model that satisfies the constraints $\lambda_j^{(i)} \in [0,1)$ for $1 \leq j \leq n$ and $\sum_{j=1}^{n} \lambda_j^{(i)} \in [0,1)$, and then let $g_i = 1$, $g_k = 0$ for $k \neq i$.

3. The output of the network is produced and the network weights is adjusted using the NLMS algorithm[9] as

$$\begin{aligned} \hat{y}(t) &= [1 \quad \mathbf{x}(t)^T]\hat{\Theta}^{(i)}(t-1) \\ \hat{\Theta}^{(i)}(t) &= \hat{\Theta}^{(i)}(t-1) + \eta \frac{[1 \quad \mathbf{x}^T(t-1)]^T e(t)}{c + [1 \quad \mathbf{x}^T(t-1)][1 \quad \mathbf{x}^T(t-1)]^T} \\ e(t) &= y(t) - \hat{y}(t) \end{aligned} \tag{7}$$

where η is a preset learning rate, c is an arbitrarily small positive constant.

4. The overall performance the network can be measured using a mean square error (MSE) as $J = \sum_{t=1}^{N} e^2(t)$. The network output is a highly complicated nonlinear function of the positions of the vertices which determines how the input space is partitioned. The optimization of J with respect to the vertices position to obtain an optimal piecewise linear model can be realised using the VFSR algorithm given in next section.

3. VFSR FOR OPTIMAL DELAUNAY INPUT SPACE PARTITION

The VFSR algorithm is a global optimisation method which is modified from the traditional simulated annealing (SA) algorithm to overcome the disadvantage of slow convergence of SA but maintains the advantage of the simplicity of SA. The VFSR is very suitable for constrainted optimization problems, such as to search for an optimal parameter vector of limited dimension within a constraint space. In the present problem the free parameter vector ω forming the free vertices of the Delaunay triangulation is the parameter vector chosen to be optimized. One implementation of VFSR algorithm that was described in[8] will be used in the present work. The algorithm is as follows:

1. In the initialisation, an initial $\omega \in \Re^p$ is randomly generated within a feasible set as $\Omega : L_i \leq \omega_i \leq H_i$. The values of L_i and H_i correspond to some $\min[x_j]$ and $\max[x_j]$ respectively. the initial temperature of the acceptance probability function, $T_c(0)$, is set to $J(\omega)$, and the intial temperatures of the parameter generating probability functions, $T_i(0), 1 \leq i \leq p$, are set to 1.0. A user defined control parameter c in annealing is given, and the annealing times k_i for $1 \leq i \leq p$ and k_c, are all set to 0.

2. The algorithm generates a new point in the parameter space with:

$$\omega_i^{new} = \omega_i^{old} + q_i(H_i - L_i) \text{ for } 1 \leq i \leq p \text{ and } \omega_i^{new} \in \Omega \tag{8}$$

where q_i is calculated as

$$q_i = sgn(v_i - 0.5)T_i(k_i)\left((1+\frac{1}{T_i(k_i)})^{|2v_i-1|} - 1\right) \tag{9}$$

and v_i is a uniformly distributed random variable in [0,1]. If ω_i^{new} is not in Ω it is discarded until a new $\omega_i^{new} \in \Omega$ is produced. The ω_i^{new} is the transformed to $\{\mathbf{v}_0^{(i)}, \mathbf{v}_1^{(i)}, \cdots, \mathbf{v}_n^{(i)}\}$ by the user defined labelling. The local linear models algorithm given in Section 2 is used to obtain a $J(\omega^{new})$ which is then evaluated and the acceptance probability of ω^{new} is given by

$$P_{accept} = \frac{1}{1 + \exp[(J(\omega^{new}) - J(\omega^{old}))/T_c(k_c)]} \tag{10}$$

A uniform randon variable P_{unif} is generated in [0,1]. If $P_{unif} \leq P_{accept}$, ω^{new} is accepted; otherwise it is rejected.

3. Denote ω^{best} as the best point amongst a predetermined number N_{accept} acceptance points. Reannealing procedure involves first calculating the sensitivities

$$s_i = \left| \frac{J(\omega^{best} + \mathbf{e}_i\delta) - J(\omega^{best})}{\delta} \right|, \text{ for } 1 \leq i \leq p \tag{11}$$

where δ is a small step size, and $\mathbf{e}_i \in \Re^p$ is a unit vector with ith element as 1, other elements as 0. Let $s_{max} = \max\{s_i, 1 \leq i \leq p\}$. Each parameter generating temperature T_i is scaled by a factor s_{max}/s_i and the annealing time k_i is reset

$$T_i(k_i) = \frac{s_{max}}{s_i}T_i(k_i), \qquad k_i = \left(-\frac{1}{c}\log[\frac{T_i(k_i)}{T_i(0)}]\right)^p \tag{12}$$

Similarly, $T_c(0)$ is reset to the value of the last accepted cost cost function, $T_c(k_c)$ is reset to $J(\omega^{best})$ and the annealing time k_c is rescaled accordingly

$$k_c = \left(-\frac{1}{c}\log[\frac{T_c(k_c)}{T_c(0)}]\right)^p \tag{13}$$

4. After every N_{genera} generated points, annealing takes places with

$$\begin{array}{lcl} k_i & = & k_i + 1 \\ T_i(k_i) & = & T_i(0)\exp(-ck_i^{\frac{1}{p}}) \end{array} \quad 1 \leq i \leq p$$

and

$$\begin{array}{lcl} k_c & = & k_c + 1 \\ T_c(k_i) & = & T_c(0)\exp(-ck_c^{\frac{1}{p}}) \end{array} \quad 1 \leq i \leq p$$

Otherwise goto step 2.

5. The algorithm is terminated if the parameters has remained unchanged for a few successive reannealings or a preset maximum number of cost function evaluations has been reached; Otherwise, goto step (2).

4. NUMERICAL EXAMPLE

Consider the benchmark nonlinear autoregressive (NAR) time series[14]

$$\begin{array}{rcl} y(t) & = & (0.8 - 0.5\exp(-y^2(t-1)))y(t-1) - (0.3 + 0.9\exp(-y^2(t-1)))y(t-2) \\ & & +0.1\sin(3.1415926y(t-1)) + e(t) \end{array} \tag{14}$$

where the noise $e(t)$ was a gaussian white sequence with mean zero and variance 0.02. 1000 data points were generated and the first 500 points were used as an estimation data set. The remaining data were used as a validation data set. The input vector is set as $\mathbf{x} = [y(t-1), y(t-2)]^T$. The input space partition is shown in Fig.3, where the rectangle ABCD represents the bounds of a slightly enlarged input space S. 4 points E,F,G,H are allowed to move along each side of the rectangular ABCD and point K is a free point within the input space, forming $m = 8$ trangulations $S^{(i)}$'s that can be determined using a free parameter vector ω of dimension size $p = 6$. The VFSR algorithm was applied to determine the parameter vector ω, or the position of points E, F, G, H and K, as plotted in Fig.3. At the this identification stage, the local linear models were trained using ordinary least squares to reduce computation time. 8 local linear models of Eq.(3) were obtained as

$$\begin{aligned}
\hat{y}_1(\mathbf{x}(t)) &= -0.0605 + 0.6845y(t-1) - 0.6636y(t-2), \text{ IF } \mathbf{x}(t) \in S^{(1)} \\
\hat{y}_2(\mathbf{x}(t)) &= -0.4503 + 1.1478y(t-1) - 0.7507y(t-2), \text{ IF } \mathbf{x}(t) \in S^{(2)} \\
\hat{y}_3(\mathbf{x}(t)) &= -0.4463 + 1.3563y(t-1) - 0.9220y(t-2), \text{ IF } \mathbf{x}(t) \in S^{(3)} \\
\hat{y}_4(\mathbf{x}(t)) &= -0.1022 + 0.9262y(t-1) - 1.1197y(t-2), \text{ IF } \mathbf{x}(t) \in S^{(4)} \\
\hat{y}_5(\mathbf{x}(t)) &= -0.1336 + 0.2907y(t-1) - 1.0535y(t-2), \text{ IF } \mathbf{x}(t) \in S^{(5)} \\
\hat{y}_6(\mathbf{x}(t)) &= 0.2329 + 0.8593y(t-1) - 0.6843y(t-2), \text{ IF } \mathbf{x}(t) \in S^{(6)} \\
\hat{y}_7(\mathbf{x}(t)) &= 0.2200 + 1.2104y(t-1) - 1.0669y(t-2), \text{ IF } \mathbf{x}(t) \in S^{(7)} \\
\hat{y}_8(\mathbf{x}(t)) &= 0.1853 + 0.2381y(t-1) - 1.0057y(t-2), \text{ IF } \mathbf{x}(t) \in S^{(8)}
\end{aligned} \tag{15}$$

The results of one-step ahead prediction using these local linear models are plotted in Fig.3 demonstrating the good predictive performance of the model. The VFSR algorithm in this application is still slow, however once an approprite input partition is obtained, the on-line processing using the algorithm introduced in Section 2 can be carried out successfully. The model validity test and the one-step ahead prediction over the validation data set $t = 800 \sim 900$ are plotted in Fig.4 demonstrating the model(15) is appropriate. The MSE of the validation data set $t = 501 \sim 1000$ is $(0.023)^2$.

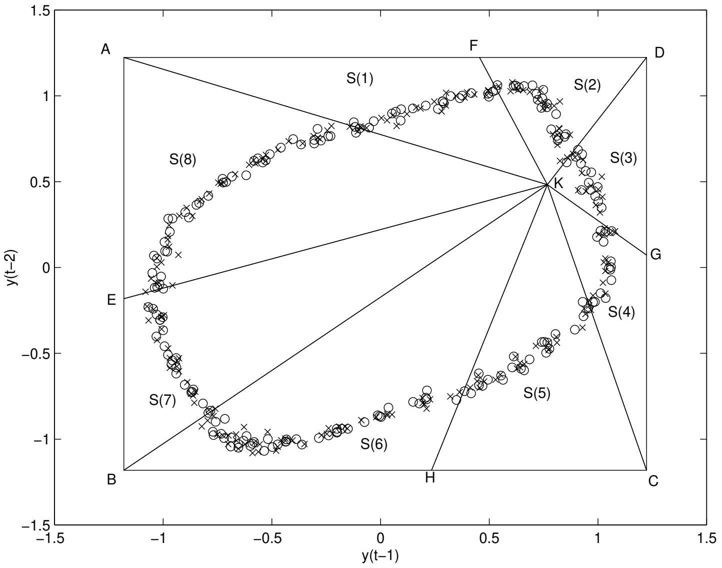

Figure 3: Piecewise locally linear modelling for numerical example (Circles: System measurements and Xs: one-step ahead predictions)

5. CONCLUSIONS

In this paper, a new algorithm has been introduced for the construction of Delaunay input space partitioned optimal piecewise locally linear model. Within the framework of a mixture of experts network, a new fast decision rule derived using convex set theory has been developed to train the model. A very fast simulated reannealing (VFSR) algorithm is utilised to search a global optimal solution for the Delaunay input space partition.

The new approach can partly overcome the problem of discontuity of lattice partitioned piecewise local linear models. Although the method is introduced in quite a general framework, the users are advised to apply the

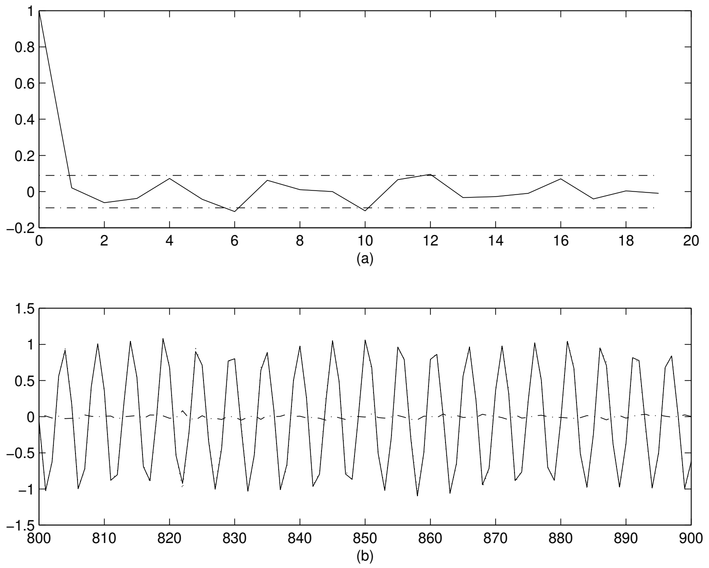

Figure 4: Model validation for numerical example; (a) Model validity test $\Phi_{ee}(\tau)$ and (b)one-step ahead prediction over the validation data set (Solid Line: System measurements, Dotted Line: one-step ahead prediction and Dashdot Line: model residual)

method to relatively low dimension problems due to the irregular shape of input space partitions and the curse of dimensionality, or use input variable selection preprocessing.

Whilst the VFSR algorithm is slow in finding optimal input space partition, this does not affect the effectiveness of the on-line algorithm described in Section 2 if an appropriate input partition had already been found. The new algorithm introduced in Section 2 may also be combined with other global optimisation approaches such as genetic algorithms, but these are likely to be even slower than VFSR.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial support of EPSRC and RACAL.

REFERENCES

1. R. Murray Smith and T. A. Johnansen, *Multiple Model Approaches to Modelling and Control*, Taylor and Francis, 1997.

2. S. A. Billings and W. S. F. Voon (1987). "Piecewise linear identification of non-linear systems". *Int. J. Control*, Vol. 46, pp215-235, 1987.

3. C. S. Hsu, *Cell-to-Cell Mapping: A Method of Global Analysis for Nonlinear Systems.* Springer-Verlag, 1987.

4. Z. Q. Wu and C. J. Harris, "A mixture expert network for piecewise linear modelling", *http://www. isis. ecs. soton. ac. uk, to be published.*, 1998.

5. S. H. Lo, "Delaunay triangulation of nonconvex plan domains", *Int. J. Numerical Methods in Engineering.* Vol. 28, pp2695-2707, 1989.

6. L. Ingber, "Very fast simulated re-annealing", *Mathematical and Computer Modelling.* Vol. 12, No.8, pp967-973, 1989.

7. L. Ingber, "Genetic algorithms and very fast simulated reanealling: a comparison", *Mathematical and Computer Modelling*. Vol. 16, No. 11, pp87-100, 1992.

8. S. Chen and B. L. Luk, "Adaptive simulated annealing for optimisation in nonlinear signal processing applications", Submitted for publication, 1998

9. M. Brown and C. J. Harris, *Neurofuzzy Adaptive Modelling and Control*. Prentice Hall, Hemel Hempstead, 1994.

10. K. M. Bossley, *Neurofuzzy Modelling Approaches in System Identification*. Ph.D Thesis, Dept of Electronics and Computer Science, University of Southampton, 1997

11. X. Hong, C. J. Harris and P. A. Wilson "Variable selection algorithm for the construction of MIMO operating point dependent neurofuzzy networks". Submitted for publication, 1998.

12. P. S. Huang, C. J. Harris and M. S. Nixon "A statistic approach for recognizing humans by gait using spatial-temporal templates", *Proceedings of IEEE International Conference on Image Processing*, Chicago, Illinois, USA, Vol.3, pp178-182, 1998.

13. J. Stoer and C. Witzgall, *Convexity and Optimization in Finite Dimensions I*. Springer-Verlag, 1970.

14. S. Chen and S. A. Billings, "Recursive hybrid algorithm for non-linear system identification using radial basis function networks", *Int. J. Control*, Vol. 55, pp1051-1070, 1992.