Influence of Parameters of Modeling and Data Distribution for Optimal Condition on Locally Weighted Projection Regression Method

Farhad Asadi, Mohammad Javad Mollakazemi, Aref Ghafouri

Abstract**—**Recent research in neural networks science and neuroscience for modeling complex time series data and statistical learning has focused mostly on learning from high input space and signals. Local linear models are a strong choice for modeling local nonlinearity in data series. Locally weighted projection regression is a flexible and powerful algorithm for nonlinear approximation in high dimensional signal spaces. In this paper, different learning scenario of one and two dimensional data series with different distributions are investigated for simulation and further noise is inputted to data distribution for making different disordered distribution in time series data and for evaluation of algorithm in locality prediction of nonlinearity. Then, the performance of this algorithm is simulated and also when the distribution of data is high or when the number of data is less the sensitivity of this approach to data distribution and influence of important parameter of local validity in this algorithm with different data distribution is explained.

Keywords**—**Local nonlinear estimation, LWPR algorithm, Online training method.

I. Introduction

DESPITE powerful and computationally fast methods in statistical learning or in computational methods in neuroscience, nonlinear function approximation in high-dimensional data with disordered distribution remains an ongoing active research. Many dynamical systems are available in engineering application or other related science such as neuroscience or bio informatics that system has many states which evolve irregular in space and time and for modeling these systems we need high dimensional data But in other situation some dynamical systems need online modeling of dynamic processes such as modeling for learning control, particularly in the field of high-dimensional movement systems or in movement primitives science [1], [2]. Nonparametric regression modeling, such as LWPR has more flexible outline for approximating unknown nonlinearities. One of the disadvantages of local models is that they are discontinuous. A small variation in the input vector can change the nearest trajectory of data series. This problem can be improved by giving the furthest data trajectories less impact on the local model. Local linear models can be constructed using weighted linear regression where the furthest trajectories accept the lowest weights.

For example two important examples of using this algorithm in recent publication are considered here. First, in sensory motor or motor control of robotics, this approach is very usable. In this system the purpose of learning algorithm is the processing the inverse dynamic of robots while produced motion has stable and robust computational features. Another important example is online learning and especially control of autonomous airplane that at this dynamical system we need to develop stable adaptive control law and also it has a fast time evolution dynamics. Difficulty of this modeling is in formulization and programming of controller because in step of the decomposition of system to subsystems we face to important challenge that many subsystems have unknown parameters that vary during the operational points of system and variation of this parameters is high and by using this method we can extract the main and important features of this parameters [3]-[5]. Consequently, decomposition of system to subsystems must be fast and computational accurate for achieving function approximation. Also LWPR methods are very useful when there is limited knowledge about the model complexity in duration of modeling.

Generally, there are two different processes for learning method that in the first method we fit nonlinear functions globally, in other words by expansions of input to predefined function and next linear combination of the expanded inputs. In the second method, we fit nonlinear functions locally in state space, usually by using spatially simple low order functions and then automatically adjusting the number of local models [6], [7].

The current approaches in statistical learning have concentrated on the first method. Gaussian process regression or locally Gaussian function approximation [8], [9] are the exemplary cases. These approaches in spite of their convergence features are not suitable for online learning in high space dimensional data. The first disadvantage of these approaches is that they require a predefined determination for demonstration. The second disadvantage is that with adding data in its database their accurate and convergence is not ensured. For example, Gaussian processes regression are very expensive in time for real time learning because it need to program the complete joint distribution of data and next with

Farhad Asadi is master student of mechanical engineering in dynamic and control, K. N. Toosi, Tehran, Iran (phone: +989357430949, e-mail: f.asadi@mail.kntu.ac.ir).
Mohammad Javad Mollakazemi is master student of mechanical engineering in dynamic and control, K. N. Toosi, Tehran (phone: +989192762663, e-mail: molakazemi.javad@yahoo.com).
Aref Ghafouri is graduated student of mechanical engineering in vehicle powertrain, K. N Toosi, Tehran (phone: +989122159954, e-mail: aref.ghafouri@yahoo.com)

adding noise in input data the accuracy of the algorithm is affected adversely [10], [11].

In this paper, the overall formulation of method is presented and then different data series and performance of algorithm for facing in distribution change is explained. When the data series have low perturbed distribution data, it should be checked that this method is applicable or not. Finally, the influence of local validity parameter in modeling of our data is explained.

II. Formulization and Iteration of the Proposed Method

Real time online learning has a three major constraint .firstly the learning and prediction should be very fast because learning needs to take place at a frequency of 20-200 HZ and prediction in range of 200HZ up to 5KHZ. Secondly the learning system needs to be capable at dealing with high dimensional and irrelevant data and thirdly the data arrives as a continuous stream, thus, the model has to be continuously adapted to new training examples over time.

In this algorithm, we use projection regression technique that first decomposes high dimensional data. The major difficulty of this method lies in the selection of efficient projections and number of selected projections. The first important assumption is that data is distributed locally perfect. The learning problems considered here assume the standard regression problem that has form as (1):

$$Y = f(x) + \in \tag{1}$$

where x is n-dimensional input space and y is scalar variable. In this method, when the data space is low, even low order polynomials can be employed in order to model this local data. For nonlinear function approximation, the main concept of our learning system, in locally weighted projection regression (LWPR), is to find approximations by means of effective linear models. We should find number of local models K , the parameters βk, the region of validity, called receptive field (RF), parameterized as a distance metric D_k and is evident in (2):

$$\omega_K = eXP(-\tfrac{1}{2}\ X - C_K\ ^T D_k\ x - C_K\ ) \tag{2}$$

So with given this model each linear model calculates the estimation of output variable in local domain of data then the total output of the learning system is the normalized weighted mean of all K linear local models [12]-[14].

At this algorithm the centers c_k of the local model remain fixed during the computation [10]. Also, it should be said that iterative strategy of the proposed approach is mentioned in [10], [11]. The important contribution of this article is related to the ability of method for information processing in different time series data. Another main point is that if all the input variables are statistically independent and have equal variance then the projection direction will be computed easier. Schematic iteration rules are plotted in Fig. 1, from [10], that automatically allocate new locally linear models as needed. An outline of the final LWPR algorithm and parameters of algorithm are shown in Tables I and II.

TABLE I
LEGEND OF INDEXES AND SYMBOLS USED FOR LWPR [10]

Symbol	Legend
M	Number of training data
N (dim. of **x**)	Input dimensionality
$k = (1:K)$	Number of local models
$r = (1:R)$	Number of local projections
$\{\mathbf{x}i, yi\}$	Training data
pr	Regressed input space
X, Z	representations of input and projected data
W	Activation of data (**x**, y) one local model centered at **c**
W	Weight matrix **W** ≡ diag{$w1$,..., wM}
Wn	Sum of weights w seen by the local model
Br_i	component of slope of the local linear model
a_n^r	Sufficient statistics for covergence

TABLE II
COMPLETE LWPR ALGORITHM [10]

• Initialize the LWPR with no receptive field (RF)
• For every new training sample (x,y):
– For k = 1 to K (number of receptive fields):
∗ Calculate the activation
∗ Update projections and regression and distance metric
∗ Check if number of projections needs to be increased
If no RF was activated by more than $wgen$;
∗ Create a new RF with R = 2, c = x, D = Ddef

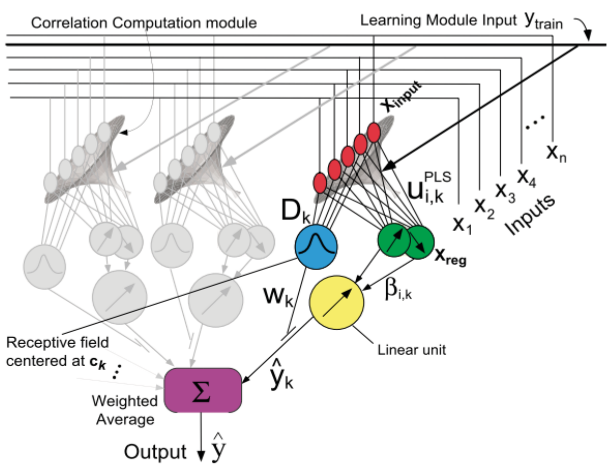

Fig. 1 Schematic diagram of algorithm for [10]

III. Data Characteristics of Local Modeling of Time Series

The following sections provide an evaluation of our proposed LWPR learning algorithm over one dimensional noisy time Series. For practical reasons it is assumed that the time series is generated by a nonlinear dynamic system governed that are evident in (3) and (4):

$$x_{t+1} = h\ x_t, u_t \tag{3}$$

$$y_t = g\,(x_t) \tag{4}$$

International Journal of Mechanical, Industrial and Aerospace Sciences
ISSN: 2517-9950
Vol:8, No:10, 2014

where $h\ x_t, u_t$ and $g\ (x_t)$ are nonlinear functions with continuous derivatives. Six artificial data sets are used to illustrate the local estimation of nonlinearity. These data series are plotted in Figs. 2-4 with different characteristics.

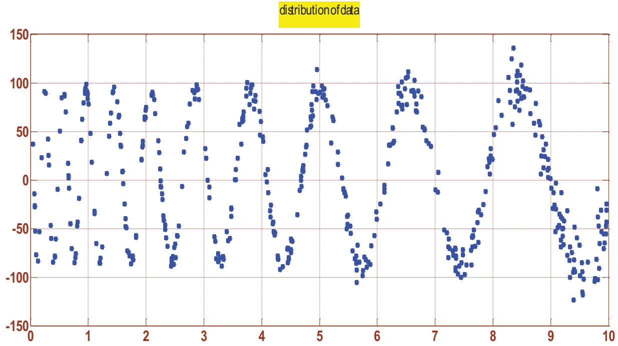

Fig. 2 Artificial data series with little variance

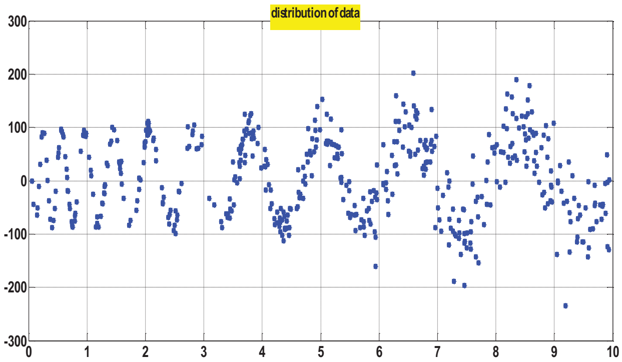

Fig. 3 Artificial data series with more variance

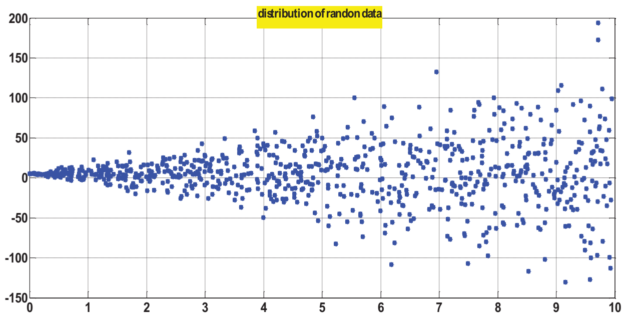

Fig. 4 Artificial data series with chaotic behavior

The decomposition of data results from the main hypothesis those nearby input points are likely to have similar weight values.

IV. Performance of LWPR Method in Locally Estimation with Perturbed Data

Generally, noise-free time series created by the dynamic system in (3) will encompass periodic oscillations solution. Region of validity should be chosen to cover several of these oscillations in DS. Thus, points that related the same local region are informative for algorithm for this local region. In algorithm after calculating the proximity between the new data point and all available centers, the data point will be involved to the nearest local model. This threshold for keeping data in local region is so important for both one and two dimensional data series. Firstly, the estimation of data for Fig. 2 is plotted in Fig. 5. In this simulation it is evident that at first section of data the estimation has poor performance. Estimation in this local region of data is plotted more clearly in Fig. 11 and estimation in last section of data is plotted in Fig. 12. In last region the estimation of data has better performance .Optimal value of region of validity is generally small for noise-free time series. However, short time series with more noise data are often too short to estimate because the components of *xt* closest in time to the estimation are given exponentially more weight like results of Fig. 11. However, for local models estimation algorithm can use the same *k* nearest data points as the basis for all estimation. The region of nearest data points that is considered in algorithm is plotted in Fig. 6. This figure shows that correlation between data points has important influence in estimation of nearest point and region of validity .So data series is perturbed with more noise and then estimation of local nonlinearity is plotted in Fig. 7. It is evident from this figure that at first section of data because data has more noise with uncorrelated data this estimation has poor performance and vice versa for last section of data.

When dynamics of system has irregular evolution spatially the problem of nearest points is important and also determining of parameter of region of validity and distance metric is depend to the sampling rate of data. It is especially suitable for chaotic systems where neighboring states are known to deviate exponentially with time. Estimation of chaotic data is plotted in Fig. 8 and also when more noise is inputted to local regions, the estimation is plotted in Fig. 9 and finally the region of nearest data points that is considered in algorithm is plotted in Fig. 10. Number of local models will increase if previously unknown sections of the state space are visited but also correlation between data point has influence to this locality estimation. Model accuracy is sensitive to the number of neighboring data points which makes this parameter a good applicant for locally and globally optimization.

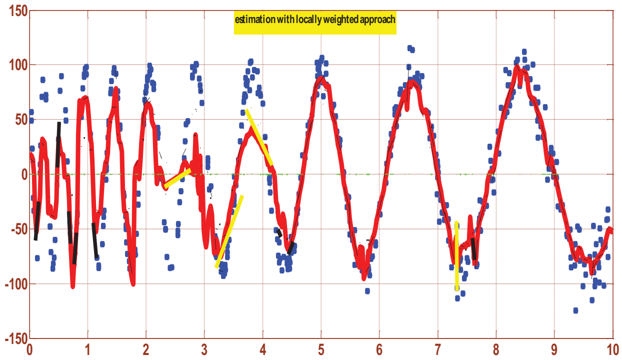

Fig. 5 Estimation of data series with increasing the parameter of region of validity

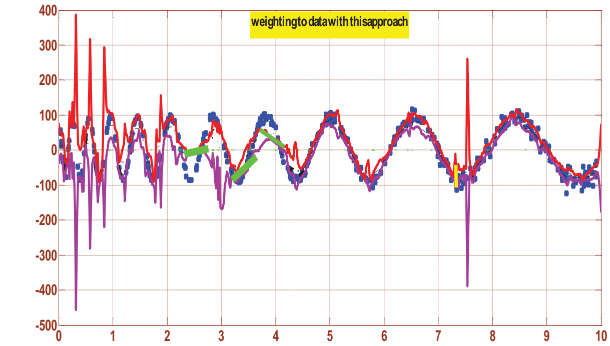

Fig. 6 Region of nearest data points that is considered in algorithm

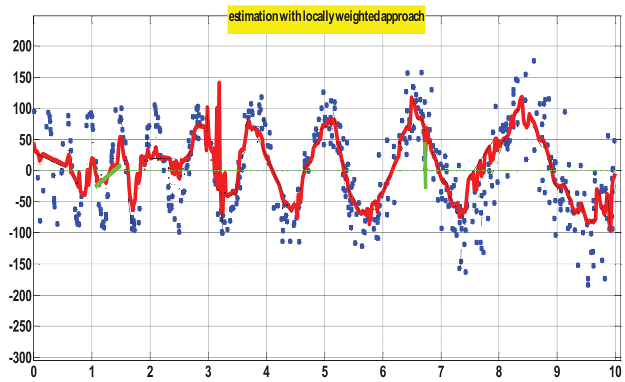

Fig. 7 Estimation of data series with increasing variance level

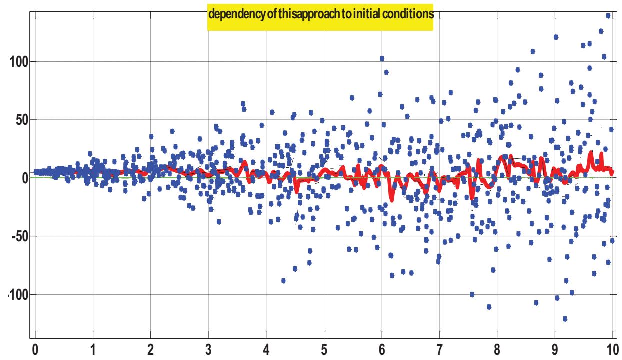

Fig. 8 Estimation of data series with chaotic behavior

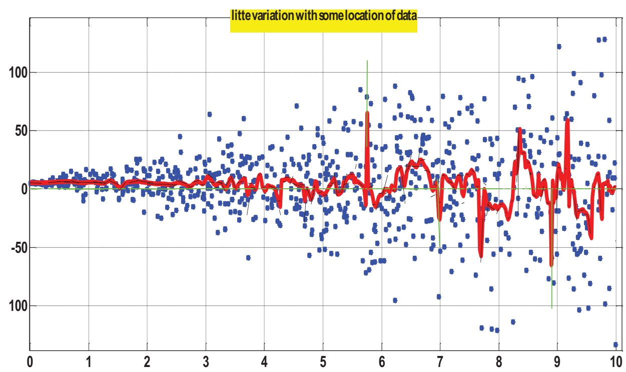

Fig. 9 Estimation of data series with chaotic behavior by increasing variance level

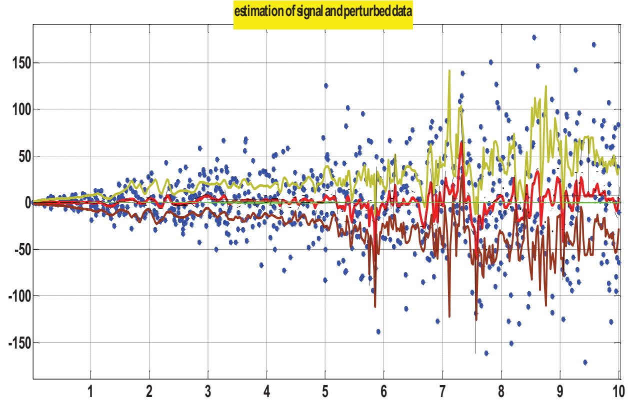

Fig. 10 Region of nearest data points that is considered in algorithm

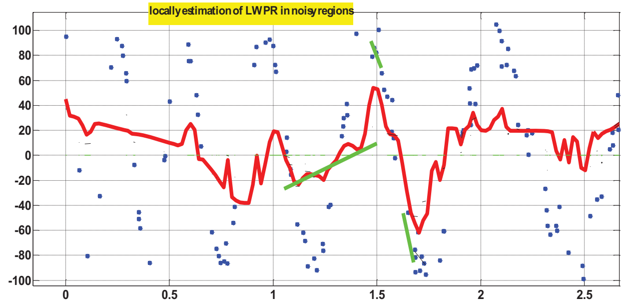

Fig. 11 Estimation of data in uncorrelated and noisy section of data

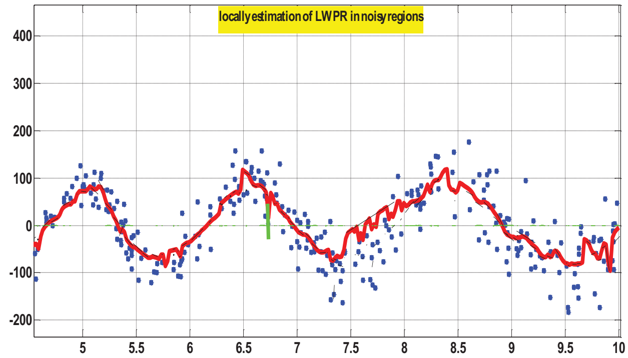

Fig. 12 Estimation of data in correlated but noisy section of data

V. Empirical Evaluation in Two Dimensional Data

For many nonlinear modeling like neural networks obtaining adaptive parameters typically needs too much computation over structural parameters. In contrast, LWPR method can be modeled and evaluated efficiently over several of the model parameters which make the choice of the parameters to be estimated especially in perturbed data. The method for estimation with decomposition of data is plotted Fig. 13.

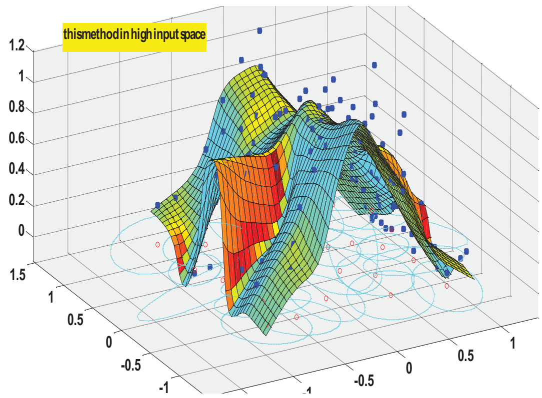

Fig. 13 Artificial data series with chaotic behavior

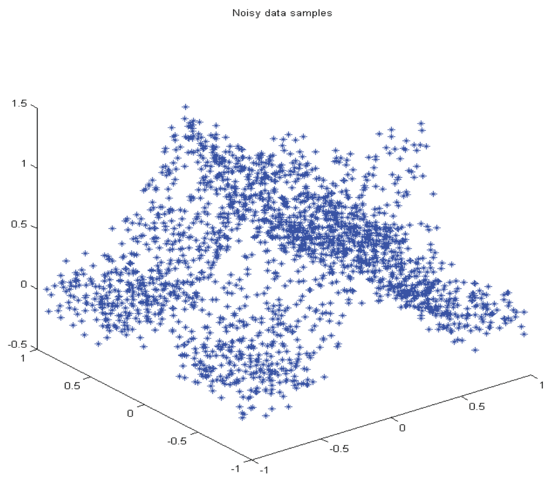

Fig. 14 Distribution of noisy data samples (n=2500)

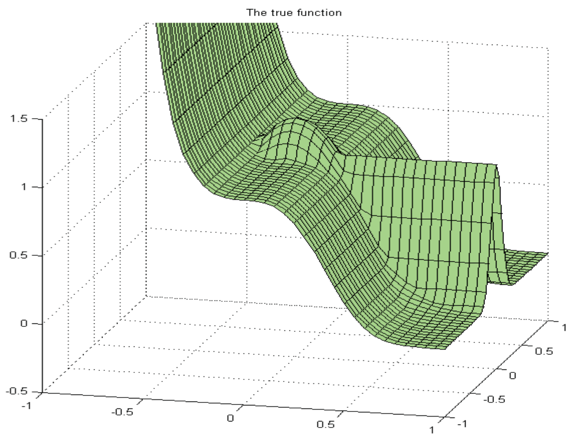

Fig. 15 The fitted and approximation function for our data

The following sections provide an evaluation of our proposed LWPR learning algorithm over a range of artificial and real-world data sets. Firstly, we evaluate this algorithm with very large data which is affected by noise and it have n=2500 data samples that this distribution is plotted in Fig. 14. Next, the true function approximation of this algorithm is shown in Fig. 15. Then, the decomposition of data with local model is plotted in Fig. 16. The runtime of this algorithm is not so much and is about 1.3 seconds that this shows the ability of this method for higher input dimension and for better computational source [15].

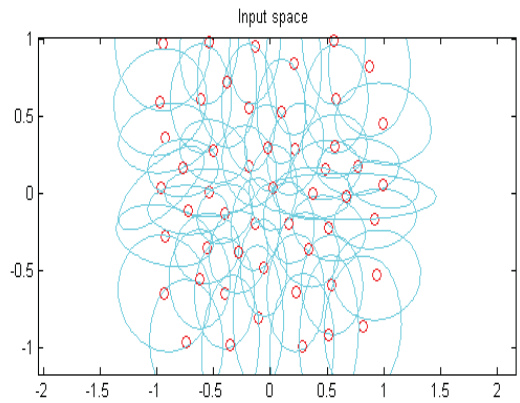

Fig. 16 The decomposition of input space to local model

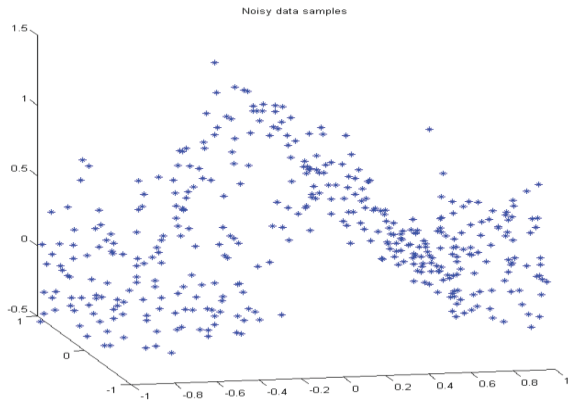

Fig. 17 Distribution of noisy data samples (n=200)

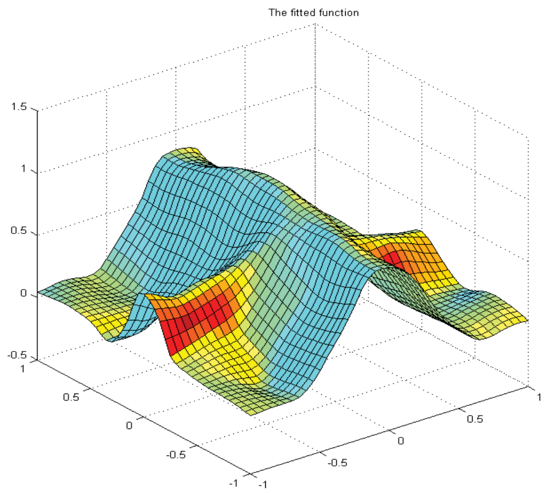

Fig. 18 The fitted and approximation function for our data

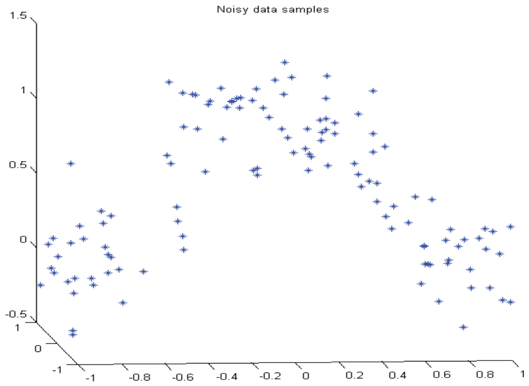

Fig. 19 Distribution of noisy data samples (n=100)

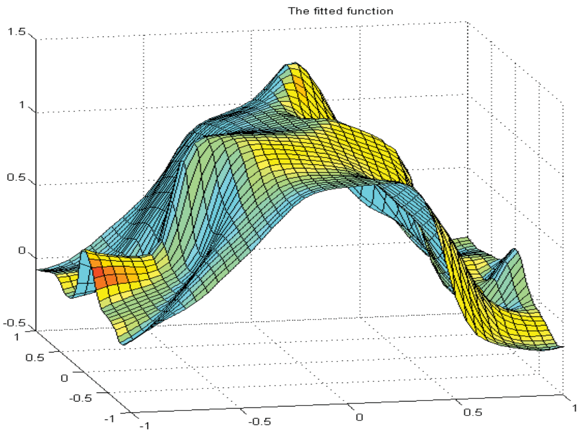

Fig. 20 The fitted and approximation function for our data

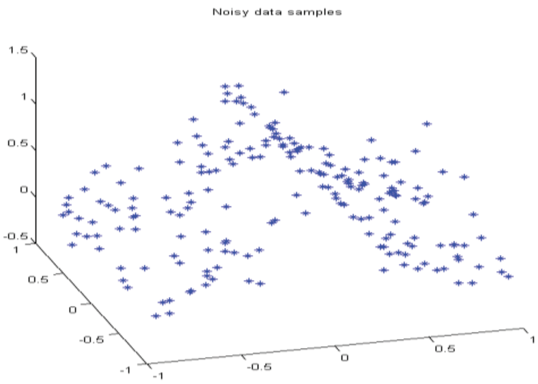

Fig. 21 Distribution of noisy data samples (n=100)

In the next simulation we use n=200 noisy dataset that is plotted in Fig. 17 and in another time series with keeping this distribution but with n=100 noisy data is designed that is plotted in Fig. 19. Fitted and approximation function for our data with first data is plotted in Fig. 18 and for second data is plotted in Fig. 20. It is evident from figures and also important contribution of this paper is that with reducing the number of data set while not changing the distribution of data, the fitted approximation is very sensitive to lower point data or disordered point in time series section that represent noise or imperfect measurement. For example, in Fig. 20 it is evident that this approximation is changed with some noisy prevalent data.

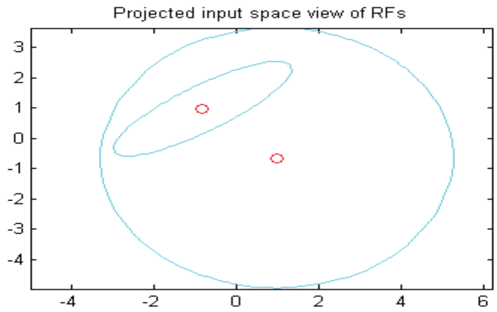

Fig. 22 The decomposition of input space to local model

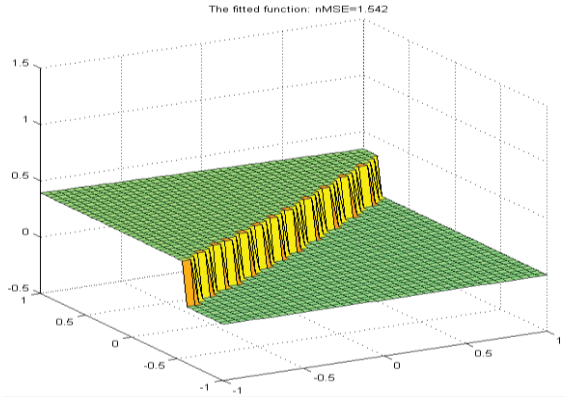

Fig. 23 The fitted and approximation function for our data

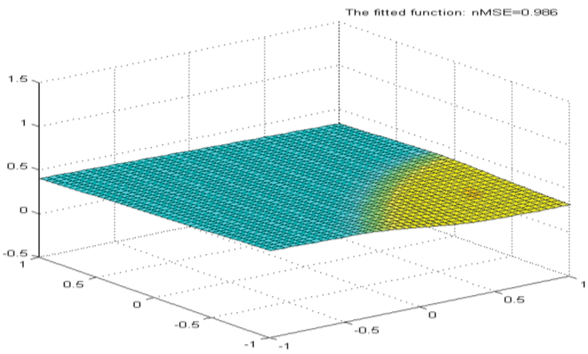

Fig. 24 The fitted and approximation function for our data with change in local validity

In this section we evaluate another important parameter of this method. These are the βk or the region of validity and also distance metric **D***k*. This parameters effect on searching the local linear for projection of data and the optimal value of this number in modeling is not observable and deterministic. That in two experimental data we showed this effect. Firstly, we examined the data sample with n=100 with two different distributions that are plotted in Figs. 21 and 26, then in each simulation different quantity for the local validity parameter in modeling is used. For example, the decomposition of input space to local model with two different local validities is plotted in Figs. 22 and 25. For other data distribution, this examination is plotted in Figs. 27 and 30.

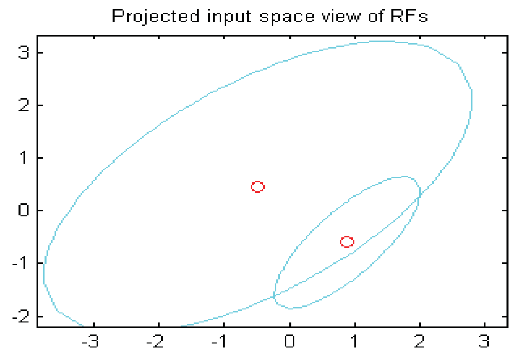

Fig. 25 The decomposition of input space to local model

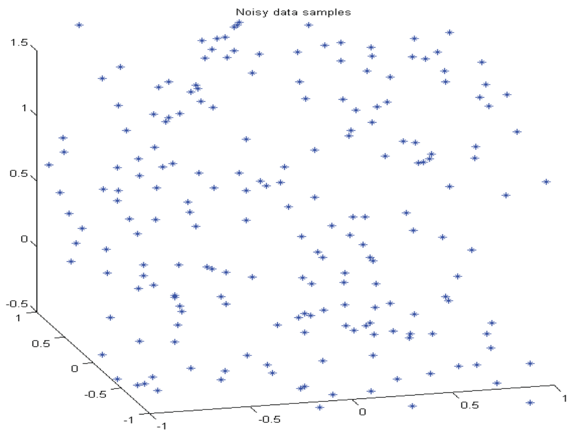

Fig. 26 Distribution of noisy data samples (n=100)

Finally approximation function on our data is plotted in Figs. 22 and 24 and then for second simulation is plotted in Figs. 28 and 29.

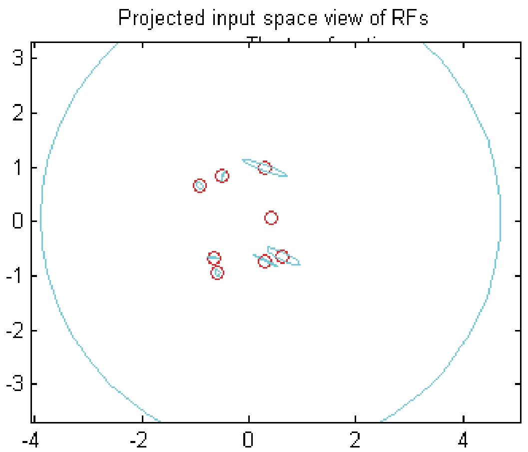

Fig. 27 The decomposition of input space to local model

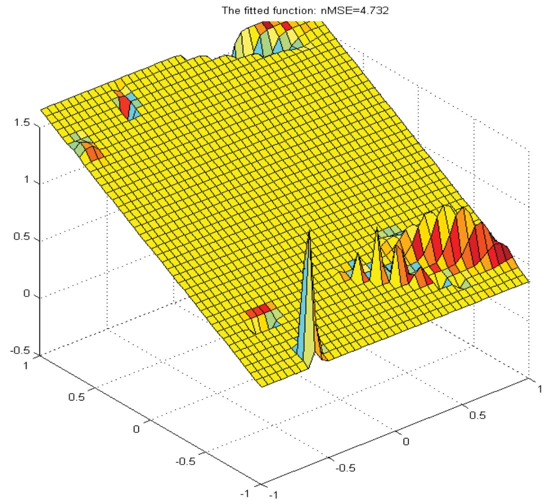

Fig. 28 The fitted and approximation function for our data

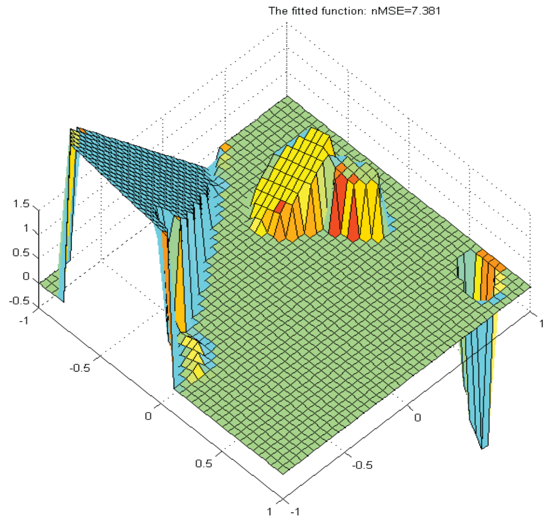

Fig. 29 The fitted and approximation function for our data with change in local validity

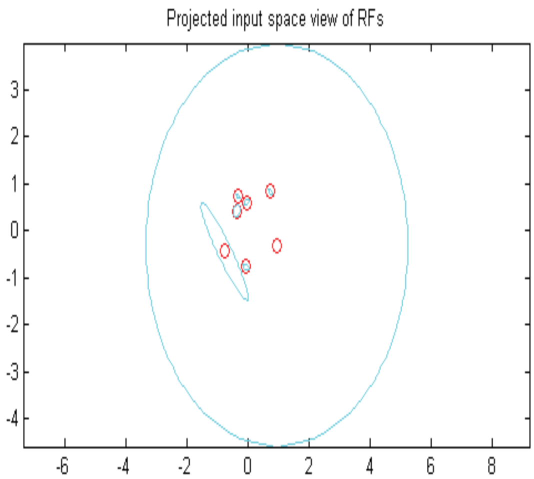

Fig. 30 The decomposition of input space to local model

Important change of our function is evident in simulation results so for modeling with this method we should think to this parameter very accurately for choosing their value and also it should be said here in some articles some notification for choosing this value is presented. Finally, the variation of fitted function with lower value of this parameter is high but in some situation we need choose the lower value for lower computational time and tradeoff between these factors for optimal approximation is important and necessary in this method.

VI. Conclusion

Nonlinear regression with spatially localized models remains one of the most data-efficient and computationally efficient methods for incremental learning with automatic determination of the model complexity. In this paper, the important notice of parameters in these modeling was explained with different data-sets. Furthermore, accuracy for choosing the parameter and limitation of the method was explained with simulation results.

References

[1] An, C. H., Atkeson, C., & Hollerbach, "*J. Model based control of a robot manipulator*". Cambridge, MA: MIT Press, 1988 .

[2] Atkeson, C., Moore, A., & Schaal, "*Locally weighted learning". Artificial Intelligence Review,* 11(4), 76–113, 1997.

[3] D'Souza, A., Vijayakumar, S., & Schaal, SAre, "*internal models of the entire body learnable" Society for Neuroscience Abstracts, 27, 2001.*

[4] Fahlman, S. E., & Lebiere, C, "The cascade-correlation learning architecture". In D. S. Touretzky (ed.), Advances in neural information processing systems, 2 (pp. 524–532). San Diego: Morgan-Kaufmann. 1990.

[5] Frank, I., & Friedman, "J. *A statistical view of some chemometric tools*". Technometrics, 35(2), 109–135. (1993).

[6] Friedman, J. H., & Stutzle, W. Projection pursuit regression. Journal of America Statistical Association, 76, 817–823,1981.

[7] Gelman, A. B., Carlin, J. S., Stern, H. S., & Rubin, D. B. Bayesian data analysis. London: Chapman and Hall,1995.

[8] Ghahramani, Z., & Beal, M. Variational inference for Bayesian mixtures of factor analysers. In S. A. Solla, T. K. Leen, & K.-R. Mu¨ ller (Eds.), Advances in neural information processing systems, 12 (pp. 449–455). Cambridge, MA: MIT Press, 2000.

[9] S. Schaal and C. G. Atkeson, *"Constructive incre mental learning from only local information," Neural Computation, vol. 10, pp. 2047-2084, 1998.*

[10] Sethu Vijayakumar, Aaron D'Souza and Stefan Schaal. "Incremental Online Learning in High Dimensions". Neural Computation, vol. 17, no. 12, pp. 2602-2634 (2005).

[11] Stefan Klanke, Sethu Vijayakumar and Stefan Schaal, *A library for Locally Weighted Projection Regression*, Journal of Machine Learning Research (JMLR), vol. 9, pp. 623-626 ,2008.

[12] Gibbs, M., & Mackay, D. J. C. *Efficient implementation of gaussian processes* (Tech. Rep.). Cambridge: Cavendish Laboratory,1997.

[13] C. Cortes and V. Vapnik, "Support vector networks," Machine Learning, vol. 20, pp. 273-297, 1995.

[14] V. Vapnik, S. Golowich, and A. Smola, "*Support vector method for function approximation, regression estimation, and signal processing*," in Advances in Neural Information Processing Systems 9, M. Mozer, M. I. Jordan, and T. Petsche, Eds. Cambridge, MA: MIT Press, 1996, pp. 281-287 and http://en.wikipedia.org/wiki?curid=4261562.

[15] Farhad Asadi, Mohammad javad Mollakazemi, "*Investigation of fluctuation locations and effect of data distribution in time series dynamical regimes*".. Accepted and oral presentation in ICBCBBE 2014: XII International Conference on Bioinformatics, Computational Biology and Biomedical Engineering, October, 27-28, 2014, Istanbul, Turkey.