

# A Data Forward Stepwise Fitting Algorithm Based on Orthogonal Function System

Han-Ju Li<sup>1,a</sup>

<sup>1</sup> Dongguan Power Supply Bureau, Guangdong Power Grid Co., Ltd. Dongguan 523008, P.R.China  
99959828@qq.com

**Abstract.** Data fitting is the main method of functional data analysis, and it is widely used in the fields of economy, social science, engineering technology and so on. Least square method is the main method of data fitting, but the least square method is not convergent, no memory property, big fitting error and it is easy to over fitting. Based on the orthogonal trigonometric function system, this paper presents a data forward stepwise fitting algorithm. This algorithm takes forward stepwise fitting strategy, each time using the nearest base function to fit the residual error generated by the previous base function fitting, which makes the residual mean square error minimum. In this paper, we theoretically prove the convergence, the memory property and the fitting error diminishing character for the algorithm. Experimental results show that the proposed algorithm is effective, and the fitting performance is better than that of the least square method and the forward stepwise fitting algorithm based on the non-orthogonal function system.

## 1 Introduction

Data fitting or data approximation is an important and common technique in data mining [1]. Least squares method (LSM) is widely used in data fitting. But LSM has the following drawbacks: the solution may not exist; the number of parameters to be far less than the number of data, otherwise prone to over fitting; not successive approximation algorithm, the fitting error does not decrease with increasing complexity; do not have memory property, when increasing the base function, need to retraining parameters. In recent years, the paper [2-5] puts forward the method of piecewise linear fitting. In this method, the original data sequence is matched with the key points of the data sequence, which is used to compress the original sequence in order to get smaller storage and computation cost. However, this method cannot approach the original data sequence arbitrarily, and the fitting error is large. The orthogonal function system is an important set of functions. After completion, it can become the base of Hilbert space, which can be used to carry out the research of Fourier analysis, function approximation and data fitting. The orthogonal function system is easy to construct, and can be obtained by the implementation of the Schmidt orthogonalization method. The paper [6] gives some important orthogonal function systems. Because of the linear independence of the orthogonal function system, it can be used to extract the information of each vertical direction without overlapping. So the orthogonal function system is very suitable for data fitting. Based on LSM, the paper [7] uses

a class of Orthogonal Spline Function GF system to fit the scattered data and the paper [8] uses the wavelet function to fit the discrete data. Based on the popular method of surface rendering, the paper [9] applies a complete orthogonal function system to fit the point cloud data. Because of the excellent properties of orthogonal function system, this paper adopts the linear combination of complete orthogonal trigonometric function on  $L^2 [0, 1]$  to fit the original data. However, the algorithm idea of this paper is different from the previous literature, but adopts the forward stepwise fitting strategy, and extracts the maximum residual information in each step. The algorithm proposed in this paper has some excellent properties, such as convergence, memory property, and gradual approximation.

## 2 Algorithm and Property

Set the data sequence is  $\{y_1, y_2, \dots\}$ , given a positive integer  $M$  such that  $N+M$  is odd, the data sequence is mapped to the plane of the data set

$$\left\{ \left( \frac{1}{N+M}, y_1 \right), \left( \frac{2}{N+M}, y_2 \right), \dots, \left( \frac{i}{N+M}, y_i \right), \dots, \left( \frac{N}{N+M}, y_N \right) \right\}. \quad (1)$$

Let  $f(x)$  be the original function of (1), i.e.

$$y_i = f\left(\frac{i}{N+M}\right), i=1, 2, \dots, N.$$

Take the following orthogonal function system

<sup>a</sup> Corresponding author: 99959828@qq.com

$\{1, \cos 2\pi x, \sin 2\pi x, \cos 2\pi 2x, \sin 2\pi 2x, \dots\}$   
 in  $L^2[0, 1]$  and use the additive model

$$f_n(x) = a_0 + \sum_{k=1}^n (a_k \cos 2\pi kx + b_k \sin 2\pi kx) \quad (2)$$

of the above orthogonal function system to fit  $f(x)$ . The  $n$  in (2) reflects the complexity of the model, in general, the bigger the  $n$ , the better the fitting, however, the more complex the model, prone to over fitting. The problem now is to give (1), how to learn the parameters of (2) and determine the appropriate complexity.

Given complexity  $n$ , we need to learn  $2n+1$  parameters, denoted by  $\gamma_n = (a_0, a_1, b_1, \dots)$ . Let  $\mathbf{R}_{n+1}$  denote the residual vector of  $f(x)$  and  $f_n(x)$  in (1), i.e.

$$\mathbf{R}_{n+1} = \left( y_1 - f_n\left(\frac{1}{N+M}\right), y_2 - f_n\left(\frac{2}{N+M}\right), \dots, y_n - f_n\left(\frac{N}{N+M}\right) \right). \quad (3)$$

LSM is to obtain the parameters such that the mean square error  $\text{MSE}(\mathbf{R}_{n+1})$  minimum, i.e.

$$\gamma_n = \arg \min_{\gamma_n} \text{MSE}(\mathbf{R}_{n+1}) = \arg \min_{\gamma_n} \frac{1}{N} \sum_{i=1}^N \left( y_i - f_n\left(\frac{i}{N+M}\right) \right)^2.$$

In this paper, the forward stepwise fitting strategy is adopted. Defining basis functions

$$\beta_k(x) = a_k \cos 2\pi kx + b_k \sin 2\pi kx, \quad k = 0, 1, \dots, \text{ where } \beta_0(x) = a_0.$$

The idea of the algorithm is to use the basis function  $\beta_k(x)$  to fit the residual vector  $\mathbf{r}_k$  produced by the last basis function  $\beta_{k-1}(x)$ , which makes the mean square error  $\text{MSE}(\mathbf{r}_k - \beta_k)$  minimum, where

$$\beta_k = \left( \beta_k\left(\frac{1}{N+M}\right), \beta_k\left(\frac{2}{N+M}\right), \dots, \beta_k\left(\frac{N}{N+M}\right) \right), \quad \mathbf{r}_k = (r_{k1}, r_{k2}, \dots).$$

and  $\mathbf{r}_0 = (y_1, y_2, \dots)$ .

That is to solve the following optimization problems

$$(a_k, b_k) = \arg \min_{a_k, b_k} \text{MSE}(\mathbf{r}_k - \beta_k) = \arg \min_{a_k, b_k} \frac{1}{N} \sum_{i=1}^N \left( r_{ki} - \beta_k\left(\frac{i}{N+M}\right) \right)^2.$$

In the following we solve the above optimization problem.

If  $k = 0$ ,  $\mathbf{r}_0 = (y_1, y_2, \dots)$ ,  $\beta_0 = (a_0, a_0, \dots)$ , then we have

$$a_0 = \arg \min_{a_0} \text{MSE}(\mathbf{r}_0 - \beta_0) = \arg \min_{a_0} \frac{1}{N} \sum_{i=1}^N (y_i - a_0)^2 = \frac{1}{N} \sum_{i=1}^N y_i. \quad (4)$$

If  $k \geq 1$ , denote  $L(a_k, b_k) = \text{MSE}(\mathbf{r}_k - \beta_k)$ , let

$$\frac{\partial L(a_k, b_k)}{\partial a_k} = 0, \quad \frac{\partial L(a_k, b_k)}{\partial b_k} = 0,$$

and solve the above equations, and then we have

$$a_k = \frac{\left| \begin{array}{cc} \sum_{i=1}^N r_{ki} \cos \frac{2\pi ki}{N+M} & \sum_{i=1}^N \cos \frac{2\pi ki}{N+M} \sin \frac{2\pi ki}{N+M} \\ \sum_{i=1}^N r_{ki} \sin \frac{2\pi ki}{N+M} & \sum_{i=1}^N \sin^2 \frac{2\pi ki}{N+M} \end{array} \right|}{\left| \begin{array}{cc} \sum_{i=1}^N \cos^2 \frac{2\pi ki}{N+M} & \sum_{i=1}^N \cos \frac{2\pi ki}{N+M} \sin \frac{2\pi ki}{N+M} \\ \sum_{i=1}^N \cos \frac{2\pi ki}{N+M} \sin \frac{2\pi ki}{N+M} & \sum_{i=1}^N \sin^2 \frac{2\pi ki}{N+M} \end{array} \right|}, \quad (5)$$

$$b_k = \frac{\left| \begin{array}{cc} \sum_{i=1}^N r_{ki} \sin \frac{2\pi ki}{N+M} & \sum_{i=1}^N \cos \frac{2\pi ki}{N+M} \sin \frac{2\pi ki}{N+M} \\ \sum_{i=1}^N r_{ki} \cos \frac{2\pi ki}{N+M} & \sum_{i=1}^N \cos^2 \frac{2\pi ki}{N+M} \end{array} \right|}{\left| \begin{array}{cc} \sum_{i=1}^N \cos^2 \frac{2\pi ki}{N+M} & \sum_{i=1}^N \cos \frac{2\pi ki}{N+M} \sin \frac{2\pi ki}{N+M} \\ \sum_{i=1}^N \cos \frac{2\pi ki}{N+M} \sin \frac{2\pi ki}{N+M} & \sum_{i=1}^N \sin^2 \frac{2\pi ki}{N+M} \end{array} \right|}. \quad (6)$$

Using (5) and (6), we obtain the following algorithm.

**Algorithm A** (Data forward stepwise fitting algorithm based on orthogonal trigonometric function, OTFS-DFSFA).

**Input:** data sequence  $\{y_1, y_2, \dots\}$  and Complexity  $n$ ;

**Output:**  $f_n(x) = a_0 + \sum_{k=1}^n (a_k \cos 2\pi kx + b_k \sin 2\pi kx)$ .

**Step1.** Mapping data sequence into (1);

**Step2.** According to (4), calculate  $a_0$ , get  $\beta_0(x) = a_0$  and calculate  $\mathbf{r}_1 = \mathbf{r}_0 - \beta_0$ ;

**Step3.** For  $k = 1, 2, \dots$ ,

- 1) According to (5) and (6), calculate  $a_k$  and  $b_k$ , and get basis function  $\beta_k(x) = a_k \cos 2\pi kx + b_k \sin 2\pi kx$ ;
- 2) Update residual vector  $\mathbf{r}_{k+1} = \mathbf{r}_k - \beta_k$ ;

**Step4.** Output  $f_n(x) = \sum_{k=0}^n \beta_k(x)$ .

Algorithm A has the following properties.

**Proposition B.**  $\mathbf{R}_{n+1} = \mathbf{r}_{n+1}$ .

**Proof.** It follows from Step3 2) in Algorithm A that

$$\begin{aligned} \mathbf{r}_{n+1}(i) &= \mathbf{r}_n(i) - \beta_n(i) = \mathbf{r}_{n-1}(i) - \beta_{n-1}(i) - \beta_n(i) \\ &= \dots \\ &= y_i - \left( a_0 + \beta_1\left(\frac{i}{N+M}\right) + \beta_2\left(\frac{i}{N+M}\right) + \dots + \beta_{n-1}\left(\frac{i}{N+M}\right) \right) \\ &= y_i - f_n\left(\frac{i}{N+M}\right) = \mathbf{R}_{n+1}(i), \quad i = 1, 2, \dots \end{aligned}$$

**Proposition C.** The coefficient vector  $\gamma_{n-1}$  of (2) with complexity  $n-1$  and the coefficient vector  $\gamma_n$  of (2) with complexity  $n$  satisfy

$$\gamma_{n-1}(i) = \gamma_n(i), \quad i = 1, 2, \dots, 2n-1.$$

**Proof.** It follows from (5), (6) and step2, step3 in Algorithm A.

Proposition C shows that the algorithm has memory property. The model of high complexity inherits corresponding position coefficient of the model of low complexity. Note that LSM does not have such property.

**Proposition D.**  $\text{MSE}(\mathbf{R}_{n+1})$  is a decreasing function on  $n$ .

**Proof.** By Proposition B we have  $\text{MSE}(\mathbf{R}_{n+1}) = \text{MSE}(\mathbf{r}_{n+1})$ . By Step3 2) in Algorithm A we have

$$\text{MSE}(\mathbf{r}_{n+1}) = \text{MSE}(\mathbf{r}_n - \beta_n),$$

where  $\beta_n$  is obtained by Step3 1) in Algorithm A, i.e.

$$\begin{aligned} (a_n, b_n) &= \operatorname{argmin}_{a_n, b_n} \operatorname{MSE}(\mathbf{r}_n - \boldsymbol{\beta}_n) = \operatorname{argmin}_{a_n, b_n} \frac{1}{N} \sum_{i=1}^N \left( r_i - \beta_n \left( \frac{i}{N+M} \right) \right)^2 \\ &= \operatorname{argmin}_{a_n, b_n} \frac{1}{N} \sum_{i=1}^N \left( r_i - a_n \cos \frac{2\pi ki}{N+M} - b_n \sin \frac{2\pi ki}{N+M} \right)^2 \end{aligned}$$

Therefore

$$\begin{aligned} \operatorname{MSE}(\mathbf{r}_{n+1}) &= \operatorname{MSE}(\mathbf{r}_n - \boldsymbol{\beta}_n) = \frac{1}{N} \sum_{i=1}^N \left( r_i - a_n \cos \frac{2\pi ki}{N+M} - b_n \sin \frac{2\pi ki}{N+M} \right)^2 \\ &\leq \frac{1}{N} \sum_{i=1}^N \left( r_i - 0 \cdot \cos \frac{2\pi ki}{N+M} - 0 \cdot \sin \frac{2\pi ki}{N+M} \right)^2 = \operatorname{MSE}(\mathbf{r}_n). \end{aligned}$$

Note that LSM does not have Proposition D.

### 3 Convergence of the Algorithm

**Theorem E.** Algorithm A is convergent.

**Proof.** It suffices to prove that the denominator of formula (5) and (6) is not equal to 0, i.e.

$$\left( \sum_{i=1}^N \cos^2 \frac{2\pi ki}{N+M} \right) \left( \sum_{i=1}^N \sin^2 \frac{2\pi ki}{N+M} \right) - \left( \sum_{i=1}^N \cos \frac{2\pi ki}{N+M} \sin \frac{2\pi ki}{N+M} \right)^2 \neq 0.$$

In fact, by Cauchy inequality we have

$$\left( \sum_{i=1}^N \cos^2 \frac{2\pi ki}{N+M} \right) \left( \sum_{i=1}^N \sin^2 \frac{2\pi ki}{N+M} \right) - \left( \sum_{i=1}^N \cos \frac{2\pi ki}{N+M} \sin \frac{2\pi ki}{N+M} \right)^2 \geq 0. \quad (8)$$

So we only need to prove that the equality sign of (8) not occur. A necessary and sufficient condition for the equality sign of (8) occur is

$$\cos \frac{2\pi ki}{N+M} = 0, i=1,2,\dots \quad \text{or} \quad \sin \frac{2\pi ki}{N+M} = K \cos \frac{2\pi ki}{N+M}, i=1,2,\dots, \quad (9)$$

Where  $K$  is a constant. First it follows from the first item of (9) that there exists integer  $m_i$  such that

$$\frac{2\pi ki}{N+M} = m_i \pi + \frac{\pi}{2}, i=1,2,\dots$$

This implies  $4ki = (2m_i + 1)(N+M)$ ,  $i=1,2,\dots$ . However, the above equation is not valid. In fact, the left side of the equation is an even number, and the right is an odd number. So  $\cos \frac{2\pi ki}{N+M} \neq 0$  for some  $i=1,2,\dots$ . So the first item of (9) is not established. Second it follows from the second item of (9) that

$$\tan \left( \frac{2\pi ki}{N+M} \right) = \frac{\sin \frac{2\pi ki}{N+M}}{\cos \frac{2\pi ki}{N+M}} = K, i=1,2,\dots$$

The function  $\tan \left( \frac{2\pi kx}{N+M} \right)$  is periodic with  $\frac{N+M}{2k}$ , and is strictly monotonically increasing function in a cycle. It follows from  $\tan \left( \frac{2\pi ki}{N+M} \right) = \tan \left( \frac{2\pi k(i+1)}{N+M} \right)$  that there exists an integer  $m$  such that  $\frac{2\pi k(i+1)}{N+M} - \frac{2\pi ki}{N+M} = \frac{(N+M)m}{2k}$ , i.e.  $4\pi k^2 = (N+M)^2 m$ . This is a contradiction because the left of the upper equation is an irrational number, and the right is an integer. Therefore, (9) does not hold and (8) cannot take equal sign, so the denominator of (5) and (6) is not zero.

### 4 Model Selection

The canonical method of model selection is regularization method. Regularization usually has the following form:

$$\min_{f \in \Omega} \frac{1}{N} \sum_{i=1}^N L(y_i, f(x_i)) + \lambda J(f), \quad (10)$$

Where the first term measure the capacity of model fitting the training data, and the second is the regularization term, which reflects the complexity of the model structure. The relative importance of model fitting accuracy and structure complexity is controlled by regularization parameter  $\lambda \geq 0$ . Data fitting is usually transformed into an optimization problem with loss function and regularization term, in which parameter  $\lambda$  play a tradeoff between loss and over fitting risk. The literature [10] has carried out a thorough study on this problem.

Given complexity  $n$ , because of the memory property of Algorithm A, Algorithm A can learn  $n$  models, and the complexity of these models from 1 to  $n$ . The question now is how to choose the right model from these models. According to (10), the following specific regularization forms are adopted.

$$\min_{k \in \{1,2,\dots\}} \frac{1}{N} \sum_{i=1}^N \left( y_i - f_k \left( \frac{i}{N+M} \right) \right)^2 + \lambda \|\boldsymbol{\gamma}_k\|_2^2, \quad (11)$$

Where  $\boldsymbol{\gamma}_k = (a_0, a_1, b_1, \dots)$  and  $\|\boldsymbol{\gamma}_k\|_2^2 = a_0^2 + \sum_{i=1}^k (a_i^2 + b_i^2)$ . According to Proposition B and Proposition D, the first term of (11) is equal to  $\operatorname{MSE}(\mathbf{r}_{k+1})$ , and decreases as the  $k$  increases. When the complexity is changed from  $k$  to  $k+1$ , we analyze the change of the two terms of (11). The first item of (11) decrease from  $\operatorname{MSE}(\mathbf{r}_{k+1})$  to  $\operatorname{MSE}(\mathbf{r}_{k+2})$ , the decrease is  $\operatorname{MSE}(\mathbf{r}_{k+1}) - \operatorname{MSE}(\mathbf{r}_{k+2})$ , and the second item of (11) increase from  $\lambda \|\boldsymbol{\gamma}_k\|_2^2$  to  $\lambda \|\boldsymbol{\gamma}_{k+1}\|_2^2$ , and the increase is  $\lambda (a_{k+1}^2 + b_{k+1}^2)$ . If

$$\operatorname{MSE}(\mathbf{r}_{k+1}) - \operatorname{MSE}(\mathbf{r}_{k+2}) \leq \lambda (a_{k+1}^2 + b_{k+1}^2)$$

Then when the complexity is changed from  $k$  to  $k+1$ , the value of (11) becomes larger. According to the above discussion and the memory property of Algorithm A, the following over fitting criteria are obtained:

If complexity  $k$  satisfies

$$\operatorname{MSE}(\mathbf{r}_k) - \operatorname{MSE}(\mathbf{r}_{k+1}) \leq \lambda (a_k^2 + b_k^2), \quad (12)$$

Then  $f_m(x)$  is over fitting for all  $m \geq k$ , where  $\lambda$  can be adjusted according to practical problems.

### 5 Experiments

#### 5.1 Experimental data

The data used in the experiment comes from the Dongguan statistical survey information network, which records the monthly electricity load in Dongguan from

January 2011 to May 2016 (unit KWH), with a total of 65 records. Take and get the following data set.

Table 1. Monthly load data

time(x)	load(y)	time(x)	load(y)	time(x)	load(y)	time(x)	load(y)
1/66	47.06	18/66	53.56	34/66	53.09	50/66	26.47
2/66	36.17	19/66	54.32	35/66	50.46	51/66	48.80
3/66	32.31	20/66	58.16	36/66	50.60	52/66	53.32
4/66	44.61	21/66	61.04	37/66	39.80	53/66	59.02
5/66	46.10	22/66	53.85	38/66	34.12	54/66	65.30
6/66	48.07	23/66	50.23	39/66	51.20	55/66	69.44
7/66	52.13	24/66	42.14	40/66	50.95	56/66	69.33
8/66	56.05	25/66	51.01	41/66	58.20	57/66	61.98
9/66	59.34	26/66	44.25	42/66	63.31	58/66	56.03
10/66	50.88	27/66	26.21	43/66	70.87	59/66	53.55
11/66	48.76	28/66	46.67	44/66	68.62	60/66	52.96
12/66	45.21	29/66	47.67	45/66	64.08	61/66	49.41
13/66	45.02	30/66	58.59	46/66	56.73	62/66	28.66
14/66	29.76	31/66	63.33	47/66	51.17	63/66	53.54
15/66	39.81	32/66	62.87	48/66	51.92	64/66	52.67
16/66	46.19	33/66	57.99	49/66	50.66	65/66	62.43
17/66	47.09						

**5.2 Comparison of the fitting performance with the least square method**

First, this experiment compares the descending speed of MSE with LSM. Given complexity  $n$ , the fitting degree of the algorithm is measured by  $MSE(R_{n+1})$ , where  $R_{n+1}$  is determined by (3). Taking complexity from 1 to 45, we get the contrast graph of the MSE descending speed of the two algorithms. As can be seen from Figure 1,  $MSE(R_{n+1})$  of the LSM does not tend to zero as the complexity increases, and  $MSE(R_{n+1})$  of the forward stepwise fitting algorithm tends to zero as the complexity increases.

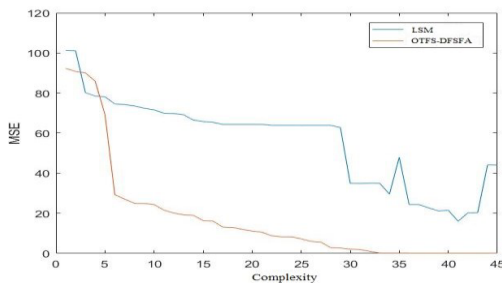
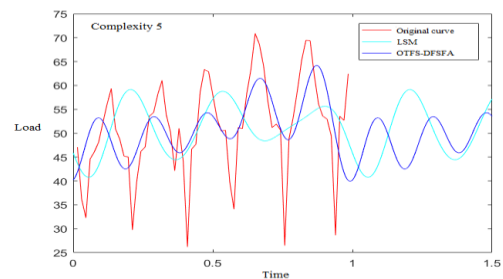
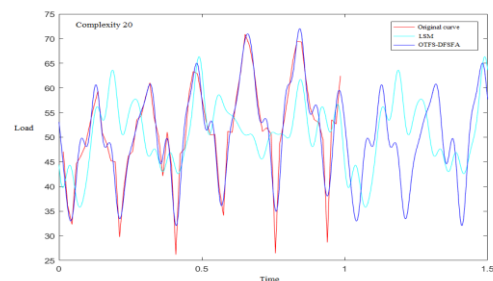


Figure 1. Comparison of the two algorithms for MSE descent speed.

Secondly, complexity 5 and 20 are taken respectively to compare the actual fitting effects of the two algorithms. The experimental results are shown in figure 2.



(a) Complexity 5



(b) Complexity 20

Figure 2. Fitting effect diagram for the two algorithms .

**5.3 Comparison of fitting performance with non-orthogonal function system**

This experiment adopts the following non-orthogonal function system

$$\{1, x, \cos 2\pi x, x^2, \cos 2\pi 2x, \dots, \dots\}$$

The forward stepwise fitting algorithm which is similar to algorithm A is adopted to compare the fitting effects of two different functional systems. First, compare

the MSE's descent rate, taking the complexity from 1 to 45, and get the contrast chart of the following MSE descent speed in figure3.

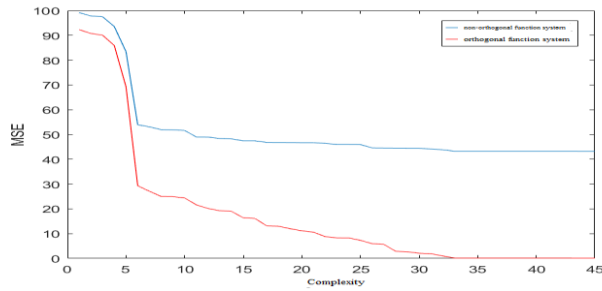
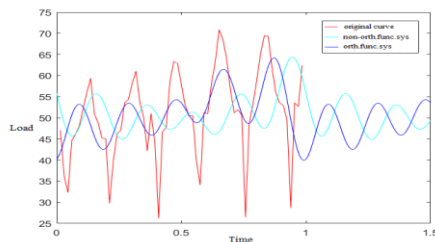
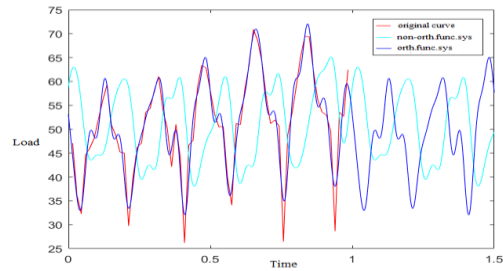


Figure 3. Comparison of the two functional systems for MSE descent speed.

Secondly, complexity 5 and 20 are taken respectively to compare the actual fitting effects of the two functional systems. The experimental results are shown in figure 4.



(a) Complexity 5



(b) Complexity 20

Figure 4. Fitting effect diagram for the two functional systems.

### 5.4 Model selection and over fitting decision

Taking complexity 34, by the memory property of the algorithm A, we get a total of 34 models with complexity from 1 to 34, and take  $\lambda=1/65$  to compute the regularization formula (11) and the over fitting condition (12) of each model. The calculation results are as follows.

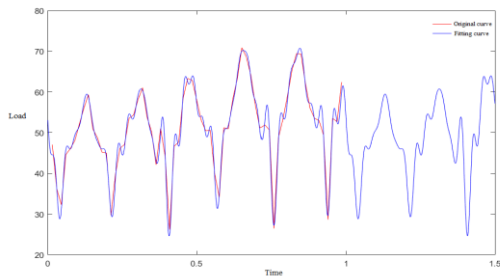
As can be seen from above table, if  $k=32$ , then  $MSE(\mathbf{r}_{k+1}) + \lambda \|\gamma_k\|_2^2$  is minimal. Therefore, the optimal complexity of the model is 32. If  $k=33$ , then

$$MSE(\mathbf{r}_k) - MSE(\mathbf{r}_{k+1}) < \lambda(a_k^2 + b_k^2).$$

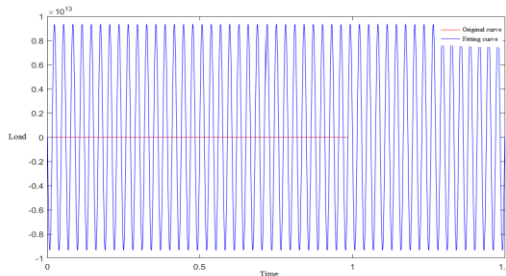
So,  $f_n(x)$  is over fitting for all  $n \geq 33$ . In the following the fitting effect of the models are given in figure 3 where complexity is 32 and 33.

Table 2. Model selection calculation process

$k$	$MSE(\mathbf{r}_k) - MSE(\mathbf{r}_{k+1})$	$\lambda(a_k^2 + b_k^2)$	$MSE(\mathbf{r}_{k+1}) + \lambda \ \gamma_k\ _2^2$	$k$	$MSE(\mathbf{r}_k) - MSE(\mathbf{r}_{k+1})$	$\lambda(a_k^2 + b_k^2)$	$MSE(\mathbf{r}_{k+1}) + \lambda \ \gamma_k\ _2^2$
1	10.251	0.3114	93.416	18	0.1563	0.0047	16.5497
2	1.5381	0.0477	91.926	19	1.0096	0.0306	15.5708
3	0.7125	0.0219	91.235	20	0.8475	0.0258	14.7491
4	4.1539	0.1298	87.211	21	0.5502	0.0167	14.2156
5	16.685	0.5155	71.042	22	1.8056	0.0549	12.4648
6	39.905	1.2366	32.373	23	0.4944	0.0152	11.9856
7	2.2415	0.0686	30.201	24	0.0228	0.0007	11.9635
8	2.0925	0.0644	28.173	25	0.9756	0.0297	11.0176
9	0.0653	0.002	28.120	26	1.3258	0.0411	9.733
10	0.5375	0.0164	27.589	27	0.3139	0.0095	9.4286
11	2.7816	0.0866	24.894	28	2.7799	0.0844	6.733
12	1.4559	0.0441	23.482	29	0.2114	0.0065	6.5282
13	0.9028	0.0278	22.607	30	0.5797	0.0177	5.9661
14	0.2252	0.0070	22.389	31	0.1792	0.0056	5.7925
15	2.6388	0.0806	19.830	32	0.8889	0.0274	4.931
16	0.2561	0.0078	19.582	33	0.8519	1.35e+24	1.35e+24
17	2.9715	0.0908	16.701	34	2.93e-04	9.16e-06	1.35e+24



(a) Complexity 32



(b) Complexity 33

Figure 5. The best fitting effect diagram and the over fitting effect diagram.

3. Kevin B Pratt, Eugene Fink. Search for patterns in compressed time series. *Inter.J.Ima.Grap* 2,1 (2002): 89-106
4. Sanghyun Park, Dongwon Lee, Wesley W Chu. Fast retrieval of similar subsequences in long sequence data bases. *Proceedings of the 1999 Workshop on Knowledge and Data Engineering Exchange* (IEEE. Computer Society, Washington, 1999)
5. L.J Yu, J.T Li, Y.D Zhang. Fast picture and macro block level adaptive frame/field coding for H.264. *Proceedings of IEEE Asia Pacific Conference on Circuits and Systems* (IEEE Press, 2006)
6. R.X Song. *J.Sys.Sci& Math.Sci* 28, 8 (2008):949-960
7. Z.C Cai, W Chen. Least square approximation and analysis for scattered data based on orthogonal GF system. *Act.Scienti.Natur.Univ.Suny* 52, 5 (2013):73-77
8. X.K Liang. Fitting method of data based on wavelet function. *Com.Appl.Soft* 20, 4 (2003):3-4
9. Y.Y Liang, R.X Song. Complete orthogonal function system V and points cloud fitting. *J.Sys.Simu* 18, 8 (2006):2109-2113
10. L Wang, W.J Wang, G.X Jiang. Optimization for smoothing parameters in process of data fitting. *Com.Sci* 42, 9 (2015):226-229

## 6 Conclusion

In this paper, a data forward stepwise fitting algorithm based on orthogonal function system is proposed. The algorithm has memory property, and the time complexity of the algorithm is less than that of LSM when learning multiple models. The algorithm has gradual approximation property, and the fitting error decreases as the complexity of the model increases. The experimental results show that the performance of the proposed algorithm (MSE dropping speed and fitting accuracy) is better than that of LSM and the forward stepwise fitting algorithm based on the non-orthogonal function system. At the same time, the optimal model selection method and over fitting criterion are given. The experiment shows that the model selection method and the over fitting criterion are effective and have practical significance.

## Acknowledgment

This paper is supported by the innovation project (no: 031900KK52170031) of Guangdong Power Grid Co.

## References

1. J.O Ramsay, B.W Silverman. *Applied functional data analysis: methods and case studies* (Springer, New York, 2002)
2. T Pavlidis, S.L Horowitz. Segmentation of plane curves. *IEEE Trans.Com* 23, 8(1974): 860- 870

<sup>a</sup> Corresponding author: 99959828@qq.com