应用概率统计　第二十五卷
第三期　2009年6月

Chinese Journal of Applied Probability
and Statistics Vol.25 No.3 Jun. 2009

Estimating the Monotonic Link Function in Single-Index Models *

Yang Jianjuan
(*Department of Basic Science, Zhejiang Shuren University, Zhejiang, 310015*)

Abstract

In this paper, we propose to estimate the monotonic link function of the single-index model by I-spline approximation. On the basis of a consistent estimation of the projection direction, the consistency of the least square criterion with a penalty function is established. Simulations are carried out to compare the existing method and to evidence the efficacy of our propose approach.

Keywords: Dimension reduction, I-spline, least square estimate, penalty function, single-index models.

AMS Subject Classification: 62H15, 62G08, 62E17.

§1. Introduction

In this paper we consider the regression of a univariate response Y on a p-dimensional predictor vector $X=(X_1,\cdots,X_p)^T$, where T denotes the transposition operator. In a routine analysis we can decompose Y into a link function $m(X)$ and a noise variable ϵ, which is assumed to be orthogonal to X and satisfy $\mathsf{E}(\epsilon|X)=0$, that is, $Y=m(X)+\epsilon$. Quite often the dimensionality p of X is usually very large, which hinders the estimation of the link function $m(\cdot)$ by nonparametric methods such as kernel smoothing. Therefore, we may suppose that $m(X)=\psi(\beta^TX)$ to circumvent the well-known *curse of dimensionality*, where β is an unknown vector which needs to be estimated from the available data, and the link-function $\psi(\cdot)$ may be nonlinear and unspecified. This is the so-called single index model. Now we arrive at a semi-parametric model which is more flexible on one hand and, on the other hand, avoids the curse of dimensionality one faces in fully in nonparametric models. The estimator of β as well as the link function ψ in this so-called single-index model has been extensively studied, among others, by Li and Duan[1] (1989), Duan and Li[2] (1991), Härdle et al.[3] (1993), Ichimura[4] (1993), Hristache et al.[5] (2001), Härdle and

*The research described here was supported by a NSF grant from National Natural Science Foundation of China (10701035), a special fund for young teachers in Shanghai universities (79001320).

Received July 4, 2007. Revised November 13, 2007.

Stoker[6] (1989) and Zhu et al.[7] (2008). In this context, a special case of the single-index model is considered:

$$Y = h(\beta^T X) + \epsilon, \tag{1.1}$$

where $h(\cdot)$ is a strictly monotonic transformation of the linear combination $\beta^T X$. This model is very general, including the linear model, logistic model and probit model, etc. However, little research work has been devoted to this model. Therefore, we study the estimation of this model in this work.

The rest of this paper is organized as follows. In Section 2, we illustrate that the estimation of the projection direction β. When a consistent estimation of β is available, we then investigate in Section 3 the I-spline approximation to the link function under the least square regression measure with a penalty function. The consistency of this criterion is also established in this section. Simulations are carried out in Section 4 to compare the existing methods and to evidence the efficacy of the I-spline approximation. All proofs are postponed to the Appendix.

§2. Estimation of the Projection Direction β

In this section, we will introduce a dimension-reduction method for estimating the projection index β. Consider first fitting the linear model $Y = b^T X + e$ by choosing the estimate b to minimize an objective function $\mathsf{E}(Y - b^T X)^2$. Li and Duan[1] (1989) have shown that the least square estimator

$$b = [\mathsf{Var}\,(X)]^{-1}\mathsf{Cov}\,(X, Y)$$

is a consistent estimate of $k\beta$ for some constant k if $\mathsf{E}(X|\beta^T X)$ is linear in $\beta^T X$. It follows from Eaton[8] (1986) that this linearity condition will hold for all β if X has an elliptically contoured distribution, say, the normal or student distribution. If $h(\cdot)$ is strictly monotonic, then the constant k is not zero. That is, the least square estimation b is a consistent estimator of β up to a nonzero constant scale. Throughout this article, the least square estimator b of β will be used. Therefore, we have

$$\mathsf{E}(Y|b^T X) = \mathsf{E}[(h(\beta^T X) + \epsilon)|b^T X] = h(\beta^T X) + \mathsf{E}(\epsilon|\beta^T X) = h(\beta^T X), \tag{2.1}$$

which implies that we can estimate the link function $h(\beta^T X)$ based on the least square estimator b. Denote by Z the variate $b^T X$, then only the univariate spline approximation or kernel smoother is needed, and hence the curse of dimensionality is avoided here. Notice that (2.1) also indicates that the plot $\{(b^T X, Y)\}$ may be a practically useful tool for visualizing an appropriate transformation $h(\cdot)$. This is also found in Zhu et al.[7] (2008).

To gain some insights why this response plot works, we consider, for any joint distribution of (X, Y), the problem of finding $h(\cdot)$ such that the least square distance $\mathsf{E}[Y-\psi(X^Tb)]^2$ is minimized. Let

$$h^* = \arg\max_{\psi\in H_2} \mathsf{E}[Y-\psi(X^Tb)]^2.$$

Theorem 2.1 Suppose that X and Y are non-degenerate random variables such that $\mathsf{E}(X^2)$ and $\mathsf{E}(Y^2)$ are finite. Then for any given b, the optimal choice h^* is $\mathsf{E}(Y|X^Tb)$.

This theorem, together with (2.1), shows that the least square estimator h^* is also a consistent estimator of $h(\cdot)$.

§3. I-Spline Approximation of the Link Function $h(\cdot)$

When a consistent estimator of β is obtained, there are many nonparametric smoothers available to estimate the link function. In the present context, we should, however, restrict our special attention to the spline approximation of the monotonic link function $h(\cdot)$, as the ordinary B-spline transformation (among these, see Eubank[9] (1999) for more details) may not satisfy the monotonicity. Herein we introduce a monotonic I-spline with a penalty function. Differing from Zhu et al.[7] (2008) in which the link is approximated by maximizing the covariance, we suggest to estimate the link by minimizing the least square measure. The algorithm is a modification of Ramsay[10] (1988). Specifically, consider the I-spline of order 2 based on the knots mesh $\{t_j\}$ with $c=t_0<t_1<\cdots<t_J<t_{J+1}=d$. The basis function $\pi(\cdot)=(B_0,\cdots,B_{J+1})^T$ is defined through B_k as $B_0(z_i)\equiv 1$ and

$$\begin{aligned}
B_1(Z) &= \frac{(Z-t_0)^2}{(t_1-t_0)^2} I(t_0\le Z\le t_1) + I(Z\ge t_1);\\
B_k(Z) &= \frac{(Z-t_{k-1})^2}{(t_{k-1}-t_{k-2})(t_k-t_{k-2})} I(t_{k-2}\le Z\le t_{k-1})\\
&\quad + \Big[1-\frac{(Z-t_{k-1})^2}{(t_k-t_{k-1})(t_k-t_{k-2})}\Big] I(t_{k-1}\le Z\le t_k), \qquad k=1,\cdots,J;\\
B_{J+1}(Z) &= \frac{(Z-t_{J-1})^2}{(t_J-t_{J-1})^2} I(t_{J-1}\le z_i\le t_J).
\end{aligned}$$

Let $h(\cdot)=\theta^T\pi(\cdot)$. First we consider the transformation for each component $Z=b^TX$ with its i.i.d. copies $\{z_i=b^Tx_i, i=1,\cdots,n\}$. When the sample points $\{x_i,y_i\}$ are available, we replace b by its least square estimator $\widehat{\beta}$, and then use the j/J-th quartile of $\widehat{z}_i=x_i^T\widehat{\beta}$ as t_j. Because we want a monotonic transformation, we here consider the minimizer of the discrepancy subject to monotonicity on $h(\cdot)$. The monotonicity can be ensured by $\theta\ge 0$ (see Ramsay[10] (1988) for details).

Since we consider the algorithm with this constraint, then the solution of θ is not simple. However, the minimization problem is a quadratic problem and can be resolved without much difficulty, although such a transformation may not be strictly monotonic because some optimal values of the components of θ may be zero. It is clear that we can restrict the first derivative to be bounded away from 0. To achieve this, the following criterion with a penalty function can be used. Let

$$\mathbf{D}_{\theta,\beta}(h(Z)) = \mathsf{E}[Y - h(Z)]^2 + 2\alpha \sum_{i=0}^{J+1} \log(1+\theta_i) \tag{3.1}$$

and $\theta = (\theta_0, \cdots, \theta_{J+1})$ is the minimizer of $\mathbf{D}_{\theta,\beta}(h)$ over all $\theta = (\theta_0, \cdots, \theta_{J+1})$. Once we have data points $\{(x_1, y_1), \cdots, (x_n, y_n)\}$, we can estimate b first, and then write $z_i = x_i^T b$ and $\widehat{z}_i = x_i^T \widehat{b}$. Then the estimators of $\mathbf{D}_{\theta,\beta}(h(Z))$, $\widehat{\mathbf{D}}_{\theta,\beta}(h(z))$, and $\widehat{\mathbf{D}}_{\theta,\widehat{\beta}}(h(\widehat{z}))$ are separately defined by

$$\widehat{\mathbf{D}}_{\theta,\beta}(h(z)) = \frac{1}{n} \sum_{k=1}^{n} [y_k - h(z_k)]^2 + 2\alpha \sum_{i=0}^{J+1} \log(1+\theta_i), \tag{3.2}$$

$$\widehat{\mathbf{D}}_{\theta,\widehat{\beta}}(h(\widehat{z})) = \frac{1}{n} \sum_{k=1}^{n} [y_k - h(\widehat{z}_k)]^2 + 2\alpha \sum_{i=0}^{J+1} \log(1+\theta_i). \tag{3.3}$$

Clearly, the final estimator $\widehat{\theta} = (\widehat{\theta}_0, \cdots, \widehat{\theta}_{J+1})$ is the minimizer of the corresponding $\widehat{\mathbf{D}}_{\theta}(h(\widehat{z}))$ over all $\theta = (\theta_0, \cdots, \theta_{J+1})$. We can also obtain the convergence of $\min \widehat{\mathbf{D}}_{\theta,\beta}(h(z))$ to $\min \mathbf{D}_{\theta,\beta}(h(z))$. The result is as follows.

Theorem 3.1 Assume that $J^{3/2} = o(\sqrt{n})$ and the fourth moments of X is finite. Then for any given β, we have

$$\left| \min_{\theta} \widehat{\mathbf{D}}_{\theta,\beta}(h(z)) - \min_{\theta} \mathbf{D}_{\theta,\beta}(h(Z)) \right| = O_p(J^{3/2}/\sqrt{n}). \tag{3.4}$$

Compare this with the convergence rate in Zhu et al.[7] (2007), the convergence rate under the least square criterion is much faster than that by maximizing the correlation coefficient in Zhu et al.[7] (2008). This shows that estimating the link function under the least square criterion is much more efficient.

For the I-spline, we do not have a closed-form solution for θ. However, adding the penalty function can ease the computation burden for the quadratic problem. The point is illustrated in the following. Note that

$$\frac{\mathrm{d}\mathbf{D}_{\theta,\widehat{\beta}}(h(\widehat{z}))}{\mathrm{d}\theta} = \frac{2}{n} \sum_{k=1}^{n} \pi(\widehat{z}_k)\pi^T(\widehat{z}_k)\theta - \frac{2}{n} \sum_{k=1}^{n} \pi(\widehat{z}_k) y_k - \frac{2\alpha}{1+\theta}$$

where $1/(1+\theta) = [1/(1+\theta_0), \cdots, 1/(1+\theta_{J+1})]^T$. From this, when letting the above equation be 0, we can derive the solution of θ. Specifically, let $\Pi_{(J+1)\times(J+1)} = \mathsf{E}[\pi(Z)\pi^T(Z)]$, $F_{(J+1)\times 1} = \mathsf{E}[\pi(Z)Y]$, and their corresponding estimation

$$\widehat{\Pi}_{(J+1)\times(J+1)} = \frac{1}{n} \sum_{k=1}^{n} \pi(\widehat{z}_k)\pi^T(\widehat{z}_k) \qquad \text{and} \qquad \widehat{F}_{(J+1)\times 1} = \frac{1}{n} \sum_{k=1}^{n} \pi(\widehat{z}_k) y_k.$$

Furthermore, we write the the (i,j)-th element of the matrix Π by $\widehat{a}_{ij}$, and the j-th element of the vector F by $\widehat{b}_j$. For each $0 \le j \le J+1$, θ_j is the solution of the following equation:

$$(\widehat{a}_{jj}\theta_j - \widehat{b}_j)(1+\theta_j) + \sum_{i\neq j} \widehat{a}_{ji}\theta_i(1+\theta_j) - \alpha = 0,$$

where $\widehat{a}_{jj} > 0$. We can obtain the final solutions by an iterative algorithm because each θ_j is only related to all other components θ_i. The following theorem states the converge of this algorithm.

Theorem 3.2 Suppose that all marginal density functions of X_l, $l = 1,\cdots,p$ are bounded and positive on $[c,d]$. Choosing $\alpha < (\lambda_{\min}\Pi^2)^{1/2}$, the above algorithm converges in probability, where $\lambda_{\min}(\Pi^2)$ stands for the smallest eigenvalue of the matrix $\Pi^T\Pi$.

To perform the above approximation, one must decide on the number of knots J based on the available data point $\{(x_i,y_i), i = 1,\cdots,n\}$. Here, we use the modified Bayesian Information Criterion (BIC) (Schwarz[11] (1978)) proposed by Zhu, Zhu and Li[12] (2007). That is, the optimal choice of k_n, denoted by $k_{n,\text{opt}}$, is defined by

$$k_{n,\text{opt}} = \arg\min_{k_n}\Big[\log(\widehat{\sigma}^2(k_n)) + (k_n+3)\frac{\max\{\log n, 3\}}{n}\Big], \tag{3.5}$$

where $\widehat{\sigma}^2(k_n)$ is the sum of the squares of the residuals that were obtained in (1.1). Obviously, some other model selection criterion can be used in the place of (3.5). For instance, the modified Akaike type information criteria, as proposed by Fujikoshi and Satoh[13] (1997) are an alternative (McQuarrie and Trai[14] (1998) provide a comprehensive discussion). Although we make no claim that the modified BIC proposed by Zhu, Zhu and Li[12] (2007) are always the best in this context, our experience indicates that it works very well across a variety of situations.

§4. Empirical Study

Simulations are carried out in this section to evidence the performance of I-spline approximation under the least square measure for the single-index model with monotone link function. Zhu et al.[7] (2008) studied the same problem but the estimation of the link function is obtained by maximizing the correlation coefficient $\text{Corr}^2[h(Z),Y]$, while in this paper the approximation is based on the least square criterion. To facilitate the comparison, we use the same model as Zhu et al.[7] (2008):

$$Y = \mathrm{e}^{\beta^T X} + \sigma\varepsilon, \tag{4.1}$$

where X and ε are independent. $X = (X_1,\cdots,X_p)$ is p-dimensional and ε is from $N(0,1)$. In the simulation, p vaies from 4 through 10, and $\beta = (1,\cdots,1)^T/\sqrt{p}$, the corresponding

p-dimensional vector. To examine the impact of the variance σ^2, we choose $\sigma = 0.3$, 0.5, 1, 2, 5, 10, and $X = (X_1, \cdots, X_p)$ is drawn from $N(0, I_p)$, where I_p is a $p \times p$ identity matrix. In each case, we generate 100 data sets each of size 100. We now apply the methods above and in Zhu et al.[7] (2008). We used penalty function defined in (3.1) to do the I-spline approximation with the knots determined by the modified BIC type criterion introduced in Zhu et al.[7] (2008). In the following two tables (Table 2 is cited from Zhu et al.[7] (2008) for ease of illustration), we report the mean of the multiple correlation coefficients, which we call the empirical correlation coefficient.

Table 1 The empirical correlation coefficient between y and the I-Spline approximation $h(\cdot)$ by using the least square criterion in this paper

	$p=4$	$p=5$	$p=6$	$p=7$	$p=8$	$p=9$	$p=10$
$\sigma=0.3$	0.9570	0.9454	0.9353	0.9290	0.9286	0.9285	0.9258
$\sigma=0.5$	0.9567	0.9443	0.9351	0.9287	0.9284	0.9284	0.9257
$\sigma=1$	0.9541	0.9435	0.9339	0.9288	0.9283	0.9284	0.9256
$\sigma=2$	0.9473	0.9402	0.9333	0.9286	0.9283	0.9282	0.9255
$\sigma=5$	0.9126	0.8842	0.9246	0.9280	0.9279	0.9280	0.9253
$\sigma=10$	0.8255	0.9231	0.9048	0.9271	0.9277	0.9278	0.9252

Table 2 The empirical correlation coefficient between y and the I-Spline approximation $h(\cdot)$ by maximizing the correlation coefficients in Zhu et al.[7] (2007)

	$p=4$	$p=5$	$p=6$	$p=7$	$p=8$	$p=9$	$p=10$
$\sigma=0.3$	0.8761	0.8663	0.8577	0.8572	0.8551	0.8543	0.8568
$\sigma=0.5$	0.8634	0.8588	0.8599	0.8566	0.8553	0.8580	0.8561
$\sigma=1$	0.8684	0.8660	0.8578	0.8590	0.8554	0.8538	0.8573
$\sigma=2$	0.8679	0.8612	0.8658	0.8559	0.8545	0.8599	0.8596
$\sigma=5$	0.8394	0.8507	0.8535	0.8542	0.8538	0.8556	0.8565
$\sigma=10$	0.7881	0.8239	0.8425	0.8433	0.8454	0.8508	0.8558

From the above results, we can clearly see that the performance of the I-spline approximation is acceptable. For model (4.1), the approximation scheme introduced in this paper via minimizing the discrepancy function (results are shown in Table 1) is more efficient than that in Zhu et al.[7] (2007) via maximizing the correlation coefficient (results are shown in Table 2). We can also find that large noise deteriorates the accuracy of approximation, the correlation is reduced. As we can expect that the correlation decreased with the increase of the dimension p.

Appendix

Proof of Theorem 2.1 Note that for any given b,

$$\min_{\psi} \mathsf{E}[Y-\psi(X^Tb)]^2 = \mathsf{E}[Y-\mathsf{E}(Y|X^Tb)]^2 + \min_{\psi} \mathsf{E}[\mathsf{E}(Y|X^Tb)-\psi(X^Tb)]^2,$$

which implies the desired conclusion. $\square$

Proof of Theorem 3.1 It suffices to prove that

$$\begin{aligned} & \min_{\theta} |\widehat{\mathbf{D}}_{\theta,\beta}(h(z)) - \mathbf{D}_{\theta,\beta}(h(Z))| \\ = & \min_{\theta} |\widehat{\mathbf{C}}_{1\theta,\beta}(h) - \mathbf{C}_{1\theta,\beta}(h) - 2\widehat{\mathbf{C}}_{2\theta,\beta}(h) + 2\mathbf{C}_{2\theta,\beta}(h)| = O_p(J^{3/2}/\sqrt{n}), \end{aligned}$$

where $\widehat{\mathbf{C}}_{1\theta,\beta}(h) = \mathsf{E}_n[\theta^T\pi(Z)\pi^T(Z)\theta]$, $\mathbf{C}_{1\theta,\beta}(h) = \mathsf{E}[\theta^T\pi(Z)\pi^T(Z)\theta]$, $\widehat{\mathbf{C}}_{2\theta,\beta}(h) = \mathsf{E}_n[\theta^T\pi(Z)Y]$, and $\mathbf{C}_{2\theta,\beta}(h) = \mathsf{E}[\theta^T\pi(Z)Y]$.

Firstly, we will study the first part at the RHS. Note that

$$\min |\widehat{\mathbf{C}}_{1\theta,\beta}(h) - \mathbf{C}_{1\theta,\beta}(h)| \le (J+2)^2 \max_{k,k_1} |\mathsf{E}_n[B_k(Z)B_{k_1}(Z)] - \mathsf{E}[B_k(Z)B_{k_1}(Z)]|.$$

In the following, we will show that

$$\max_{k,k_1} I_{k,k_1} =: \max_{k,k_1} |B_k(Z)B_{k_1}(Z) - \mathsf{E}(B_k(Z)B_{k_1}(Z))| = O_p(1/\sqrt{nJ}). \tag{A.1}$$

By the Markov inequality

$$\max_{k,k_1} \mathsf{P}\{|I_{k,k_1}| > b\} \le \max_{k,k_1} \mathsf{E}(B_k(Z)B_{k_1}(Z))^2/(nb^2).$$

Note that $B_k(Z) \le cI_{(t_k,t_{k+1})}(Z)$ and then $\mathsf{E}(B_k(Z)) \le c/J$. Adding the condition that $\mathsf{E}Y^4 < \infty$, we can obtain that

$$\max_{k,k_1} \mathsf{E}(B_k(Z)B_{k_1}(Z))^2 \le \sqrt{c^4\mathsf{E}(I_{(t_k,t_{k+1})}(Z))}\sqrt{c^4\mathsf{E}(I_{(t_{k_1},t_{k_1+1})}(Z))} \le C/J.$$

Choosing b satisfies $1/b = o(\sqrt{nJ})$, we derive that

$$\mathsf{P}\Big\{\max_{k,k_1} |I_{k,k_1}| > b\Big\} = [\mathsf{P}\{|I_{k,k_1}| > b\}]^{(J+1)^2} \le C/(nJb^2) = o(1).$$

Therefore, we have

$$\min |\widehat{\mathbf{C}}_{1\theta,\beta}(h) - \mathbf{C}_{1\theta,\beta}(h)| = O_p(J^{3/2}/\sqrt{n}). \tag{A.2}$$

Now we turn to investigating the second term $\widehat{\mathbf{C}}_{2\theta,\beta}(h)$. Note that

$$\begin{aligned} \max I_{k,2} & =: \max |\widehat{\mathbf{C}}_{2\theta,\beta}(h) - \mathbf{C}_{2\theta,\beta}(h)| \\ & \le (J+2)\max_{k} |B_k(Z)Y - \mathsf{E}(B_k(Z)Y)|. \end{aligned}$$

For any $0 \le k \le J+1$, and any $b>0$ by the Markov inequality

$$\max_k \mathsf{P}\{|I_{k,2}|>b\} \le \max_k \mathsf{E}(B_k(Z)Y)^2/(nb^2).$$

Note that $B_k(Z) \le cI_{(t_k,t_{k+1})}(Z)$ and then $\mathsf{E}(B_k(Z)) \le c/J$. Adding the condition that $\mathsf{E}Y^4<\infty$, we can obtain that

$$\max_k \mathsf{E}(B_k(Z)Y)^2 \le \sqrt{c^4\mathsf{E}(I_{(t_k,t_{k+1})}(Z))}\sqrt{\mathsf{E}Y^4} \le C/\sqrt{J}.$$

Choosing b satisfies $1/b=o(\sqrt{n}/J^{1/4})$, we derive that

$$\mathsf{P}\Big\{\max_k |I_{k,2}|>b\Big\} \le \sum_k \max_k \mathsf{P}\{|I_{k,2}|>b\} \le C(\sqrt{J})/(nb^2)=o(1).$$

Therefore, we have

$$\min|\widehat{\mathbf{C}}_{2\theta,\beta}(h)-\mathbf{C}_{2\theta,\beta}(h)| = O_p(J^{5/4}/\sqrt{n}). \tag{A.3}$$

The proof is concluded from (A.2) and (A.3). □

Proof of Theorem 3.2 Without loss of difficulty, we can show that $\widehat{\Pi}$ converges in probability to Π. Let $\theta_i^{(l)}$ be the results from the lth step of the iterative algorithm. Therefore,

$$\begin{aligned}\Pi(\theta_i^{(l+1)}-\theta_i^{(l)}) &= \alpha\Big(\frac{1}{1+\theta_i^{(l)}}-\frac{1}{1+\theta_i^{(l-1)}}\Big)\\ &= \alpha c_i^{(l)}(\theta_i^{(l)}-\theta_i^{(l-1)}),\end{aligned}$$

and then

$$\begin{aligned}(\theta_i^{(l+1)}-\theta_i^{(l)}) &= \alpha\Pi^{-1}\Big(\frac{1}{1+\theta_i^{(l)}}-\frac{1}{1+\theta_i^{(l-1)}}\Big)\\ &= \alpha\Pi^{-1}c_i^{(k)}(\theta_i^{(l)}-\theta_i^{(l-1)}),\end{aligned}$$

where $c_i^{(l)}=-1/[(1+\theta_i^{(l)})(1+\theta_i^{(l-1)})]$. Note that $|c_i^{(l)}|$ is smaller than or equal to 1. Because Π is positive definite matrix and so is Π^{-1}. The largest eigenvalue $\lambda_{\max}((\Pi^T\Pi)^{-1})=1/\lambda_{\min}(\Pi^T\Pi)\le c<\infty$. Note that for any non-negative symmetric matrix B and unitary vector α, $\|B\alpha\|=\alpha^TBB\alpha\le\lambda_{\max}(BB)$ where $\lambda_{\max}$ stands for the largest eigenvalue of the matrix BB. Therefore, it is easy to see that

$$\|\theta_i^{(l+1)}-\theta_i^{(l)}\| \le \frac{\alpha}{(\lambda_{\min}(\Pi^T\Pi))^{1/2}}\|\theta_i^{(l)}-\theta_i^{(l-1)}\|.$$

If α is chosen to be smaller than $(\lambda_{\min}(\Pi^2))^{1/2}$, then the algorithm converges in probability.
□

References

[1] Li, K.C. and Duan, N.H., Regression analysis under link violation, *Ann. Statist.*, **17**(1989), 1009–1052.

[2] Duan, N.H. and Li, K.C., Slicing regression: a link-free regression method, *Ann. Statist.*, **19**(1991), 502–530.

[3] Härdle, W., Hall, P. and Ichimura, H., Optimal smoothing in single-index models, *Ann. Statist.*, **21**(1993), 157–178.

[4] Ichimura, H., Semi-parametric least squares (SLS) and weighted SLS estimation of single-index models, *J. Econometrics*, **58**(1993), 71–120.

[5] Hristache, M., Juditsky, A. and Spokoiny, V., Direct estimation of the index coefficients in a single-index model, *Ann. Statist.*, **29**(2001), 595–623.

[6] Härdle, W. and Stoker, T.M., Investigating smooth multiple regression by the method of average derivatives, *J. Amer. Statist. Assoc.*, **84**(1989), 986–995.

[7] Zhu, L.P., Yang, X.Y., Yu, Z. and Liu, X.R., An analysis of single-index model with monotonic link function, *Applied Mathematics: A Journal of Chinese Universities, Series B*, **23**(2007), 107–112.

[8] Eaton, M.L., A characterization of spherical distributions, *J. Multi. Analysis*, **20**(1986), 272–276.

[9] Eubank, R.L., *Nonparametric Regression and Spline Smoothing*, New York: Marcel Dekker, 1999.

[10] Ramsay, J.O., Monotone regression spline in action (with discussion), *Statist. Sci.*, **3**(1988), 425–461.

[11] Schwarz, G., Estimating the dimension of a model, *Ann. Statist.*, **6**(1978), 461–464.

[12] Zhu, L.X. Zhu, L.P. and Li, X., Transformed partial least squares for multivariate data, *Statist. Sinica.*, **17**(2007), 1657–1676.

[13] Fujikoshi, Y. and Satoh, K., Modified AIC and C_p in multivariate linear regression, *Biometrika*, **84**(1997), 707–716.

[14] McQuarrie, A.D.R. and Trai, C.L., *Regression and Time Series Model Selection*, Singapore: World Scientific, 1998.

[15] He, X. and Shen, L., Linear regression after spline transformation, *Biometrika*, **84**(1997), 474–481.

估计单指标模型的单调联系函数

杨 建 娟

(浙江树人大学基础部, 杭州, 310015)

本文我们研究了联系函数单调时单指标模型的模型估计问题. 基于投影方向的相合估计, 本文提出用I-样条的办法来估计联系函数, 并建立带惩罚函数的最小二乘准则的相合性. 通过模拟与现有的方法进行了对比, 表明我们的估计方法是非常有效的.

关键词: 降维, I-样条, 最小二乘估计, 惩罚函数, 单指标模型.

学科分类号: O212.2.