

递推局部线性回归估计及其应用

陈性敏^{1†}, 高超²

(1. 大连理工大学 数学科学学院, 辽宁 大连 116024; 2. 北京信息控制研究所, 北京 100037)

摘要: 在非参数统计中, 局部多项式回归是重要的工具, 然而以往研究的算法基本都是非递推的. 本文研究递推的局部线性回归估计及其应用. 首先推导出递推算法, 给出了回归函数及其导函数的非参数估计. 在一定的条件下, 证明了算法的强一致性. 并且通过仿真例子研究了它在非线性条件异方差模型的回归函数估计和非线性ARX(non-linear autoregressive system with exogenous inputs, NARX)系统辨识中的应用.

关键词: 局部线性回归; 递推辨识; 核估计; 强一致性; 非线性ARX系统

中图分类号: TP273 **文献标识码:** A

Recursive local linear regression estimation and its applications

CHEN Xing-min^{1†}, GAO Chao²

(1. School of Mathematical Sciences, Dalian University of Technology, Dalian Liaoning 116024, China;

2. Beijing Institute of Information and Control, Beijing 100037, China)

Abstract: In nonparametric statistics, local polynomial regression is one of the most important tools. However, almost the previous work is based on nonrecursive algorithms. We investigate the recursive local linear regression estimation. The recursive algorithms are derived for the nonparametric estimation of the regression function and its derivative. Strong consistency of the estimates is established under reasonable conditions. The applications to estimation of the regression model with nonlinear conditional heteroskedasticity and identification of the nonlinear ARX (NARX) system are demonstrated by numerical simulation.

Key words: local linear regression; recursive identification; kernel estimation; strong consistency; nonlinear ARX systems

1 引言(Introduction)

局部加权最小二乘回归的思想由来已久. 在非参数回归框架下, 所谓的局部多项式回归最先由文献[1]提出并系统地加以研究. 由于它本质上是加权最小二乘算法而具有计算效率上的优势, 而且它不仅能改进边界误差效应, 还有优于经典核回归估计的渐近性质^[2], 因此受到了越来越多的关注. 经过文献[2]的系统而全面的研究, 局部多项式方法已成为最重要的非参数工具之一.

相比非递推或离线算法, 递推的算法能在新的数据来临之时在线更新估计, 从实际的观点来看, 这类算法既节约了计算时间又节省了内存, 因此在时间序列和系统控制中是经常遇到和适用广泛的. 关于递推非参数回归估计的研究, 将核估计与随机逼近结合起来得到递推的算法首先由文献[3]所注意到并应用到回归函数的估计中, 因此这种算法也称为Révész算法. 文献[4]首先意识到其在系统控制中

的重要性将之应用于分类. 文献[5]给出了递推核回归估计一致性的充要条件. 对于回归变量是稳定Markov模型的情形, 递推核回归估计的逐点一致性、重对数律、中心极限定理和一致强收敛性均已建立, 见参考文献[6]及其有关文献. 文献[7]意识到了基于随机逼近的核回归递推估计在块导向系统等非线性系统的辨识中的重要作用, 进一步见文献[8-11]. 然而从递推的核回归估计到一般的递推多项式拟合回归估计, 其收敛性、渐近性等统计性质的研究并不容易.

在应用中, 传统的独立性假设并不能满足需要, 于是相依数据的回归估计受到了越来越多的关注. 对相依数据的局部多项式回归估计是由文献[12-14]所完成的, 均方收敛性、一致强收敛性及其收敛速度均已建立. 文献[15]将局部线性估计应用到非线性ARX(nonlinear autoregressive system with exogenous inputs, NARX)系统的辨识并证明了局部

线性估计器是渐近最小均方误差估计. 遗憾的是, 对相依数据的局部多项式回归估计的这些结果都是非递推的. 对一元回归, 递推局部多项式回归估计的均方收敛性和渐近正态性由文献[16]所解决. 对多元回归问题, 递推的算法及其收敛性、渐近性等仍是未解决的难题.

认识到基于核方法的局部多项式回归估计本质上是加权最小二乘算法. 对多元回归问题, 通过局部线性回归, 本文推导出了回归函数及其导函数的递推非参数估计. 在较一般的条件下, 证明了算法的强一致性.

2 问题描述(Problem formulation)

研究回归模型

$$y_k = f(\phi_k) + v_k \tag{1}$$

或

$$f(\phi) = E[y_k | \phi_k = \phi]. \tag{2}$$

不失一般性, 假设输出 y_k 和噪声 v_k 均为标量, 而 ϕ 是回归向量, 其维数为 $d(d \geq 1$ 是正整数); $f: \mathbb{R}^d \rightarrow \mathbb{R}$ 是多元回归函数; $\{v_k\}$ 又称为新息序列, 常假设是与 $\{\phi_k\}$ 独立的零均值随机过程.

无论是在统计、经济金融中, 还是在系统控制中, 回归模型(1)或(2)是经常遇到而又应用广泛的. 例如对应NARX系统^[10], 回归向量为

$$\phi_k = [y_{k-1} \ \cdots \ y_{k-n_y} \ u_{k-1} \ \cdots \ u_{k-n_u}]^T,$$

这里 u_k 是系统的输入, $n_y \geq 1$ 和 $n_u \geq 1$ 分别是系统的阶, 此时 $d = n_y + n_u$.

又如在随机最优控制中, 控制律的求解最终归结为解HJB方程, 而这往往难以有显式解, 因而条件期望的估计无论是理论研究还是实际应用都是根本重要的^[17].

在经济金融的时间序列分析的研究中, 还经常遇到非线性条件异方差的情形^[18], 此时 $v_k = g(\phi_k)e_k$, 其中常常假设 $\{e_k\}$ 是独立同分布(iid)的.

本文感兴趣的是如何从数据 $\{\phi_k, y_k\}_{k=1}^N$ 递推地估计回归函数 $f(\cdot)$ (即对任意给定的 ϕ , 估计 $f(\phi)$ 的值).

3 局部线性回归估计(Local linear regression estimation)

对一给定的 ϕ , 未知函数 f 可以表示成Taylor展开:

$$f(\psi) = f(\phi) + (\psi - \phi)^T \frac{\partial f(\psi)}{\partial \psi} \Big|_{\psi=\phi} + \frac{1}{2}(\psi - \phi)^T H(\phi)(\psi - \phi) + \text{h.o.t.}, \tag{3}$$

其中: h.o.t.(higher order terms)表示高阶项, 当 $\psi - \phi$ 足

够小时可忽略; $H(\phi) = \frac{\partial^2 f(\psi)}{\partial \psi^2} \Big|_{\psi=\phi}$ 是函数 f 在 ϕ 处的Hessian矩阵, 假设 $\det(H(\phi)) \neq 0$. 记 $\beta_0^* = f(\phi)$, $\beta_1^* = \frac{\partial f(\psi)}{\partial \psi} \Big|_{\psi=\phi}$ 分别为函数 f 在 ϕ 点的值和梯度.

基于上述表示, 局部线性估计器在 ϕ 附近用下式来逼近未知函数 f :

$$\hat{f}(\psi) = \beta_0^* + \beta_1^{*T}(\psi - \phi).$$

考虑系统(1), 本文进行下面的局部线性回归: 对于系数 β_0 和 β_1 , 最小化

$$Q_k(\theta) \triangleq \sum_{t=1}^k \{y_t - (\beta_0 + \beta_1^T(\phi_t - \phi))\}^2 \omega_{k,t}, \tag{4}$$

其中 $\theta \triangleq [\beta_0 \ \beta_1^T]^T$. 核权重选为

$$\omega_{k,t} = \frac{1}{k} K_{h_t}(\phi_t - \phi) = \frac{1}{k h_t^d} K\left(\frac{\phi_t - \phi}{h_t}\right), \tag{5}$$

这里 K 是 \mathbb{R}^d 上的核函数, 它是一个正的对称Borel可测函数 $K(\cdot): (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \rightarrow (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$,

$$K_{h_t}(\varphi) \triangleq h_t^{-d} K(h_t^{-1} \varphi),$$

$\{h_t\}$ 是窗宽序列. 在递推核估计中, 核函数和窗宽序列的选取是关键与难点, 它们的选取及作用如后面K1)–K3)和H1)–H3)所述.

令 $\hat{\beta}_0$ 和 $\hat{\beta}_1$ 为式(4)的局部最小二乘估计, 则函数 f 在点 ϕ 处的局部线性估计值为 $\hat{f}(\phi) = \hat{\beta}_0$, 记 $\hat{\theta} \triangleq [\hat{\beta}_0 \ \hat{\beta}_1^T]^T$.

注1 著名的Nadaraya-Watson估计(简记为NW估计)实际上是通过优化下列加权最小二乘问题

$$\hat{\beta}_0 = \arg \min_{\beta_0} \sum_{t=1}^k \{y_t - \beta_0\}^2 K_H(\phi_t - \phi),$$

而得到

$$\hat{\beta}_0 = \frac{\sum_{t=1}^k K_H(\phi_t - \phi) y_t}{\sum_{i=1}^k K_H(\phi_i - \phi)} \tag{6}$$

其中 H 为窗宽矩阵, 一般是对称的非奇异阵, 如常用

$$H = \text{diag}\{h, \dots, h\}, \quad K_H(\varphi) \triangleq \frac{1}{\det(H)} K(H^{-1} \varphi).$$

通常地, H 根据样本总量 N 而取. 这是统计中最常用的非递推算法. 而在递推算法中 H 常取为关于 t 的矩阵序列. 例如 $H_t = \text{diag}\{h_t, \dots, h_t\}$, $\{h_t\}$ 是所谓的窗宽序列. 此时, 由NW估计式(6)派生的递推算法又分为半递推和全递推^[7, 19]. 所谓半递推的算法, 分别用下面两个随机逼近算法来估计回归向量 $\{\phi_k\}$ 的平稳密度函数 $q(\phi)$ 和函数 $g(\phi) = f(\phi)q(\phi)$,

$$\begin{cases} \tilde{q}_k(\phi) = \tilde{q}_{k-1}(\phi) - a_k(\tilde{q}_{k-1}(\phi) - K_{h_k}(\phi_k - \phi)), \\ \tilde{g}_k(\phi) = \tilde{g}_{k-1}(\phi) - a_k(\tilde{g}_{k-1}(\phi) - K_{h_k}(\phi_k - \phi)y_k), \end{cases} \tag{7}$$

进而得到回归函数 $f(\phi)$ 的估计

$$\tilde{f}_k(\phi) = \frac{\tilde{g}_k(\phi)}{\tilde{q}_k(\phi)}, \tag{8}$$

其中 $\{a_k\}$ 是随机逼近算法中的步长. 在这个算法中 $\{a_k\}$,

$\{h_k\}$ 及核函数 K 的选取是关键.

而全递推的算法则只用一个随机逼近算法即可估计回归函数 $f(\phi)$,

$$\bar{f}_k(\phi) = \bar{f}_{k-1}(\phi) - b_k K_{h_k}(\phi_k - \phi)(\bar{f}_{k-1}(\phi) - y_k), \tag{9}$$

其中 $\{b_k\}$ 是随机逼近算法中的步长. 此时 $\{b_k\}, \{h_k\}$ 及核函数 K 的选取是关键.

注 2 一般的 p 阶递推局部多项式回归估计, 通过对系数 $\{b_\alpha, |\alpha| = 0, 1, \dots, p\}$ 进行下列的加权最小二乘优化而得:

$$\sum_{t=1}^k \{y_t - \sum_{0 \leq |\alpha| \leq p} b_\alpha (\phi_t - \phi)^\alpha\}^2 \omega_{k,t}, \tag{10}$$

其中

$$\alpha = [\alpha_1 \ \dots \ \alpha_d]^\top \in \mathbb{N}_0^d$$

为 d 元的多指标. 当 $p > 1$ 时关于多指标的符号表示及运算, 及系数 $\{b_\alpha, |\alpha| = 0, 1, \dots, p\}$ 的表示方式, 参看文献[12-13]. 这里 $\omega_{k,t}$ 如式(5)所定义. 其中的关键是将系数 $\{b_\alpha, |\alpha| = 0, 1, \dots, p\}$ 表示成向量, 并把数据 $\{\phi_k, y_k\}$ 相应地排列. 由于参数是线性的, 因此按照下面的思想也容易推导出相应的递推算法.

4 递推算法(Recursive algorithm)

为便于说明基本思想, 以线性情形为例, 推导局部线性回归估计的递推算法. 记

$$\Phi_k = \begin{bmatrix} 1 & (\phi_1 - \phi)^\top \\ \vdots & \vdots \\ 1 & (\phi_k - \phi)^\top \end{bmatrix}, Y_k = \begin{bmatrix} y_1 \\ \vdots \\ y_k \end{bmatrix},$$

$$W_k = \text{diag}\{\omega_{k,1}, \dots, \omega_{k,k}\},$$

则极小化准则函数(4)可以写成矩阵的形式:

$$\hat{\theta}_k \triangleq \arg \min_{\theta} Q_k(\theta) = \arg \min_{\theta} (Y_k - \Phi_k \theta)^\top W_k (Y_k - \Phi_k \theta). \tag{11}$$

利用最小二乘原理可以求得

$$\hat{\theta}_k = (\Phi_k^\top W_k \Phi_k)^{-1} \Phi_k^\top W_k Y_k \triangleq S_k^{-1} T_k, \tag{12}$$

其中: $S_k \triangleq \Phi_k^\top W_k \Phi_k$ 是 $(d+1) \times (d+1)$ 的矩阵

$$S_k = \begin{bmatrix} \sum_{t=1}^k \omega_{k,t} & \sum_{t=1}^k \omega_{k,t} (\phi_t - \phi)^\top \\ \sum_{t=1}^k \omega_{k,t} (\phi_t - \phi) & \sum_{t=1}^k \omega_{k,t} (\phi_t - \phi) (\phi_t - \phi)^\top \end{bmatrix},$$

$T_k \triangleq \Phi_k^\top W_k Y_k$ 是 $(d+1) \times 1$ 的矩阵

$$T_k = \begin{bmatrix} \sum_{t=1}^k \omega_{k,t} y_t \\ \sum_{t=1}^k \omega_{k,t} (\phi_t - \phi) y_t \end{bmatrix}.$$

定义 $\mathbf{x}_k \triangleq [1 \ (\phi_k - \phi)^\top]^\top$, 注意到问题(11)的求解式(12)实际上是加权最小二乘算法, 因此可以写成递推的形式:

$$\begin{cases} \hat{\theta}_k = \hat{\theta}_{k-1} + a_k P_{k-1} \mathbf{x}_k (y_k - \mathbf{x}_k^\top \hat{\theta}_{k-1}), \\ P_k = P_{k-1} - a_k P_{k-1} \mathbf{x}_k \mathbf{x}_k^\top P_{k-1}, \\ a_k = [\lambda_k^{-1} + \mathbf{x}_k^\top P_{k-1} \mathbf{x}_k]^{-1}, \end{cases} \tag{13}$$

式中 $\lambda_k = K_{h_k}(\phi_k - \phi)$.

注 3 对于一般的 p 阶局部多项式回归估计, 也可推导出与式(13)类似的算法. 例如对一元 p 阶局部多项式回归估计($p \in \mathbb{N}, d = 1$), 对任意给定的 $\phi \in \mathbb{R}$, 记

$$\mathbf{x}_k \triangleq [1 \ \phi_k - \phi \ (\phi_k - \phi)^2 \ \dots \ (\phi_k - \phi)^p]^\top.$$

令

$$\hat{\alpha} = [\hat{\alpha}_0 \ \hat{\alpha}_1 \ \hat{\alpha}_2 \ \dots \ \hat{\alpha}_p]^\top$$

为

$$[f(\phi) \ f'(\phi) \ \frac{f''(\phi)}{2!} \ \dots \ \frac{f^{(p)}(\phi)}{p!}]^\top$$

的估计, 则优化式(10)变成

$$\hat{\alpha}_k = \arg \min_{\alpha} \sum_{t=1}^k [y_t - \alpha^\top \mathbf{x}_t] \omega_{k,t}, \tag{14}$$

其中:

$$\omega_{k,t} = \frac{1}{kh_t} K\left(\frac{\phi_t - \phi}{h_k}\right),$$

$$\alpha = [\alpha_0 \ \alpha_1 \ \alpha_2 \ \dots \ \alpha_p]^\top.$$

同样地, 上式的求解实际上是加权最小二乘算法, 可以递推地计算出

$$\begin{cases} \hat{\alpha}_k = \hat{\alpha}_{k-1} + a_k P_{k-1} \mathbf{x}_k (y_k - \mathbf{x}_k^\top \hat{\alpha}_{k-1}), \\ P_k = P_{k-1} - a_k P_{k-1} \mathbf{x}_k \mathbf{x}_k^\top P_{k-1}, \\ a_k = [\lambda_k^{-1} + \mathbf{x}_k^\top P_{k-1} \mathbf{x}_k]^{-1}, \\ \lambda_k = \frac{1}{h_k} K\left(\frac{\phi_k - \phi}{h_k}\right). \end{cases} \tag{15}$$

对于多元情形, 考虑到随着多项式的阶数增加, 要估计的参数急剧增多, 因此实际应用中一般局部线性估计已够用了^[20], 而且在一定条件下它还是渐近最小均方误差估计^[15].

注 4 一般的递推 p 阶局部多项式回归估计, 即使当 $p = 0$ 时, 该算法与半递推或全递推的RW算法仍是不同的.

本文要设计的核函数是一个正的对称Borel可测函数 $K(\cdot) : (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \rightarrow (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$, 满足以下条件:

K1) $\sup_{\mathbf{x} \in \mathbb{R}^d} K(\mathbf{x}) < \infty$;

K2) $K(\cdot) \in L^1(\mathbb{R}^d)$, $\int_{\mathbb{R}^d} K(\mathbf{x}) d\mathbf{x} = 1$, $\|\mathbf{x}\|^2 K(\mathbf{x}) \in L^1(\mathbb{R}^d)$, $\|\mathbf{x}\|^{2+d} K(\mathbf{x}) \xrightarrow{\|\mathbf{x}\| \rightarrow \infty} 0$;

K3) 对 $\delta > 2$, $K^\delta(\cdot) \in L^1(\mathbb{R}^d)$, $\|\mathbf{x}\|^{2\delta} K^\delta(\mathbf{x}) \in L^1(\mathbb{R}^d)$, $\|\mathbf{x}\|^{2\delta+d} K^\delta(\mathbf{x}) \xrightarrow{\|\mathbf{x}\| \rightarrow \infty} 0$.

例 1 如Gauss核函数

$$K(\mathbf{x}) = (2\pi)^{-d/2} e^{-\mathbf{x}^\top \mathbf{x} / 2}$$

满足条件K1)–K3).

窗宽序列 $\{h_k\}$ 按照下面条件选取:

H1) $h_k > 0, \lim_{k \rightarrow \infty} h_k = 0, kh_k^{2d} \uparrow \infty;$

H2) 对 $\delta > 2$ 有 $\sum_{k=1}^{\infty} (kh_k^{(1-1/\delta)d})^{-2} < \infty;$

H3) $j = 0, 1, 2, 3, \vartheta_j \triangleq \lim_{k \rightarrow \infty} \frac{1}{k} \sum_{t=1}^k (\frac{h_t}{h_k})^{dj} < \infty.$

例 2 满足条件H1)–H3)的窗宽序列是容易选取的, 比如取 $h_k = O(k^{-\gamma})$, 则 $\vartheta_j = 1/(1 - \gamma dj)$. 因此只要选取 γ 使得 $\gamma < 1/(dj)$ 即可.

5 强一致性(Strong consistency)

5.1 模型假设(Assumptions)

对回归模型(1)或(2), 为分析算法(13)的强一致性, 需要以下条件:

A1) 回归函数 f 有连续的二阶导数并是 Lipschitz 的, 进一步其 Hessian 矩阵 $H(\phi)$ 被一个正定矩阵 C 界定: $|\mathbf{x}^T H(\phi) \mathbf{x}| \leq \mathbf{x}^T C \mathbf{x}, \forall \mathbf{x} \in \mathbb{R}^d.$

A2) 过程 $\{\phi_k\}$ 是有渐近平稳分布为 π 的 α -混合 (α -mixing) 过程, $\pi(d\phi) = q(\phi)d\phi, q$ 是正连续函数, 且混合系数 $\alpha_\phi(k)$ 满足: 对某 $\delta > 2$ 和 $\varepsilon > 0$, 有

$$\sum_{k=1}^{\infty} \{ \log k (\log \log k)^{1+\varepsilon} [\alpha_\phi(k)]^{1-2/\delta} \cdot \sum_{t=k}^{\infty} (th_t^{(1-1/\delta)d})^{-2} \} < \infty. \quad (16)$$

A3) 噪声 $\{v_k\}$ 是零均值 iid 序列, $E[|v_k|^{2\delta}] < \infty$, 在 \mathbb{R} 上有正的和一致连续的概率密度函数 $p_v(\cdot)$, 并且与 ϕ_k 独立.

注 5 过程 $\{\phi_k\}$ 是满足式(16)的有平稳分布的 α -混合过程, 这个条件是关键而又宽松的, 在实际中很多系统都满足这一条件. 比如对开环 NARX 系统, 当非线性函数 f 有压缩的线性增长限制时, 其回归向量 ϕ_k 是几何遍历的 α -混合过程, 并且 α -混合系数指数衰减^[10].

5.2 算法的强一致性(Strong consistency of estimates)

定理 1 设回归模型(1)或(2)满足A1)–A3), 若核函数满足K1)–K3), 其窗宽选取满足H1)–H3), 则递推的局部线性回归估计(13)有下列的强一致性:

$$H_k(\hat{\theta}_k - \theta^*) \xrightarrow[k \rightarrow \infty]{a.s.} 0, \quad (17)$$

其中: $H_k = \text{diag}\{1, h_k, \dots, h_k\}, \theta^*$ 为真值(即 f 在 ϕ 点的值和梯度).

若条件A2)进一步加强为:

A2') 除了式(16)加强为

$$\sum_{n=1}^{\infty} \log n (\log \log n)^{1+\varepsilon} [\alpha_\phi(n)]^{1-2/\delta} < \infty, \quad (18)$$

A2)的其他条件成立.

则有更精细的结果:

定理 2 设回归模型(1)或(2)满足A1), A2')和A3), 若核函数满足K1)–K3), 其窗宽选取满足H1)–H3), 则递推的局部线性回归估计(13)的强收敛性有如下速度:

$$H_k(\hat{\theta}_k - \theta^*) = o\left(\frac{\log k (\log \log k)^{1+\varepsilon}}{kh_k^{2(1-1/\delta)d}}\right)^{1/2} + O(h_k^2), \text{ a.s.} \quad (19)$$

因此, 若窗宽选取

$$h_k = O\left(\frac{\log k (\log \log k)}{k}\right)^{\frac{1}{4+2(1-1/\delta)d}}, \quad (20)$$

则递推的局部线性回归估计(13)有收敛速度

$$H_k(\hat{\theta}_k - \theta^*) = o\left(\frac{(\log \log k)^{\varepsilon/2}}{k \left(\frac{\log k (\log \log k)}{k}\right)^{\frac{1}{4+2(1-1/\delta)d}}}\right). \quad (21)$$

注 6 由定理1–2可知, 递推局部线性回归估计(13)给出了回归函数的强一致估计. 对一元回归的递推局部多项式回归估计, 类似定理1–2的结果由 Vilar-Fernández 等获得, 但是他们需要对 $\{\phi_k, y_k\}$ 的联合分布以及 ϕ_t 和 $\phi_{t+s} (s \geq 1)$ 的联合分布有苛刻的要求, 并不适合本文考虑的情形. 本文的强一致性结果仅需要非线性函数 $f(\cdot)$ 、回归过程 $\{\phi_k\}$ 和回归噪声 $\{v_k\}$ 分别满足条件A1)–A3), 这是相对容易检验的.

6 应用例子(Applications)

6.1 对非线性条件异方差模型回归函数的递推估计(Recursive estimation of regression function with nonlinear conditional heteroskedasticity)

考虑下面的非线性条件异方差模型:

$$y_k = f(\phi_k) + g(\phi_k)e_k, \quad (22)$$

其中过程 $\{\phi_k\}$ 是 ARMA 过程:

$$\begin{aligned} A(z)\phi_k &= B(z)\epsilon_k, \\ A(z) &= 1 - 1.5z + 0.7z^2, \\ B(z) &= 1 + z + 0.5z^2, \end{aligned}$$

z 是后移算子, $z\phi_k = \phi_{k-1}; \epsilon_k \in \mathcal{U}[0, 0.5]$ 和 $e_k \in \mathcal{N}(0, 0.01)$ 分别是服从均匀分布和正态分布的 iid 序列. 非线性函数

$$f(x) = \sin(2x) + 2 \exp(-16x^2).$$

其导函数为

$$f'(x) = 2 \cos(2x) - 64x \exp(-16x^2).$$

非线性函数 $g(x) = \exp(0.2x)$.

仿真采用了蒙特卡罗法进行模拟, 样本总量为 200, 计算了 $N = 500$ 步, 其中核函数 K 选用 Gauss 核,

窗宽取 $h_k = k^{-0.3}$.

分别取 $p = 0, 1, 2$, 利用递推的局部 p 阶多项式回归估计, 对样本进行平均, 得到函数 $f(x)$ 在 $N = 500$ 时的最终估计效果如图1所示, 相应的估计误差曲线如图2所示.

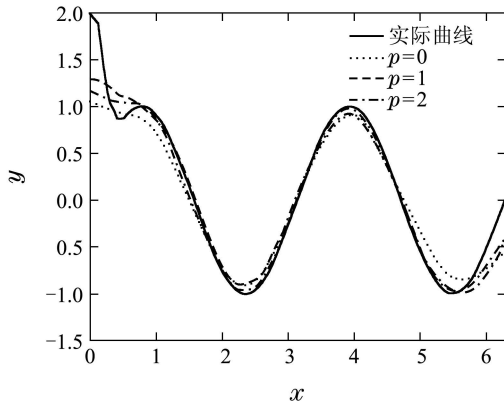


图1 $p = 0, 1, 2$ 时函数 $f(x)$ 的最终估计效果比较
Fig. 1 Estimates for $f(x)$ with respect to $p = 0, 1, 2$

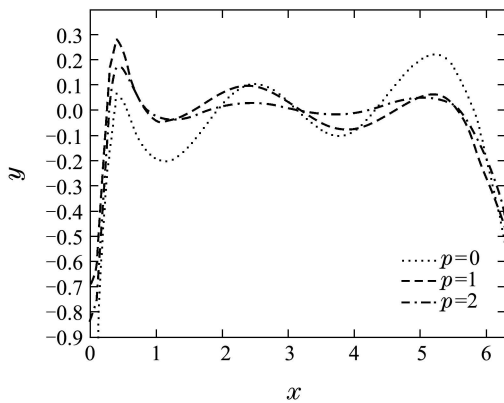


图2 $p = 0, 1, 2$ 时函数 $f(x)$ 的估计误差曲线
Fig. 2 Estimation errors of $f(x)$ with respect to $p = 0, 1, 2$

当 $p = 1, 2$ 时, 还得到函数 $f(x)$ 的导函数 $f'(x)$ 在 $N = 500$ 时的最终估计效果如图3所示, 相应的估计误差曲线如图4所示.

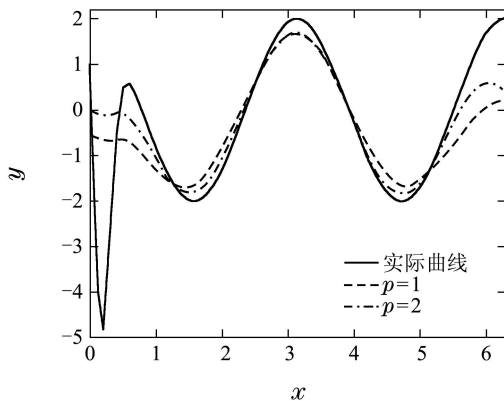


图3 $p = 1, 2$ 时导函数 $f'(x)$ 的最终估计效果比较
Fig. 3 Estimates for $f'(x)$ with respect to $p = 1, 2$

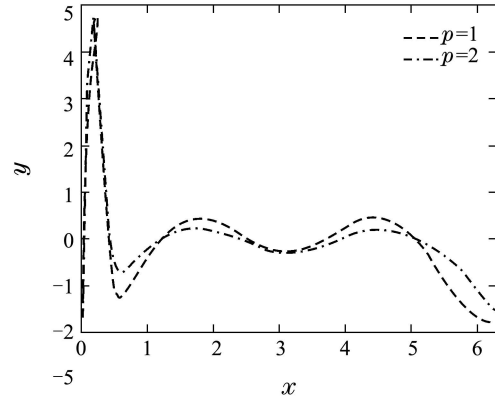


图4 $p = 1, 2$ 时导函数 $f'(x)$ 的估计误差曲线
Fig. 4 Estimation errors of $f'(x)$ with respect to $p = 1, 2$

可见, 随着 p 的增大, 递推局部多项式回归估计确能改进估计精度和边界误差, 有较大的优越性.

6.2 在 NARX 系统辨识中的应用 (Application to identification of NARX systems)

考虑 NARX(1, 1) 系统:

$$y_k = f(\phi_k) + v_k, \quad (23)$$

其中: $\phi_k = [y_{k-1} \ u_{k-1}]^T$, 输入 $u_k \in \mathcal{N}(0, 1)$, 噪声 $v_k \in \mathcal{N}(0, 1)$ 均为 iid 序列, 初值 $y_0 \in \mathcal{N}(0, 1)$, $u_0 \in \mathcal{N}(0, 1)$. 非线性函数为 $f(y, u) = \frac{1}{2}y \sin(\frac{\pi}{2}y) + 2u - u^3 + 4$.

计算了 $N = 3000$ 步, 利用不同的算法所得效果比较如下:

利用半递推的 RW 算法(7)–(8)所得非线性函数 $f(\phi)$ 的最终估计效果如图5所示. 其中取随机逼近算法的步长为 $a_k = k^{-1}$, 窗宽序列为 $h_k = k^{-0.167}$.

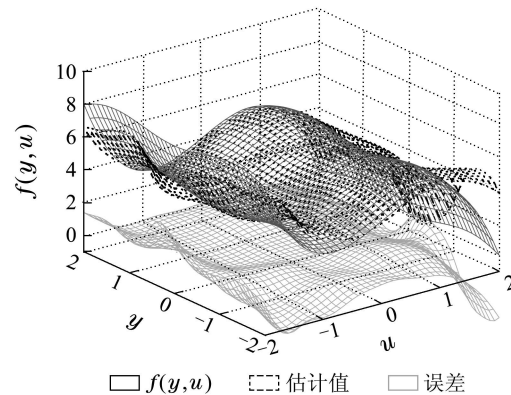


图5 利用半递推算算法(7)–(8)所得 $f(\phi)$ 的最终估计效果
Fig. 5 Estimate for $f(\phi)$ given by algorithms (7)–(8)

利用全递推的 RW 算法(9), 所得非线性函数 $f(\phi)$ 的估计并不收敛. 例如取随机逼近算法的步长为 $b_k = k^{-0.68}$, 窗宽序列为 $h_k = k^{-0.167}$, 最终估计效果如图6所示. 利用递推的局部线性回归估计(13), 所得最终的估计效果如图7所示. 算法(13)中窗宽序列选取 $h_k = k^{-0.167}$.

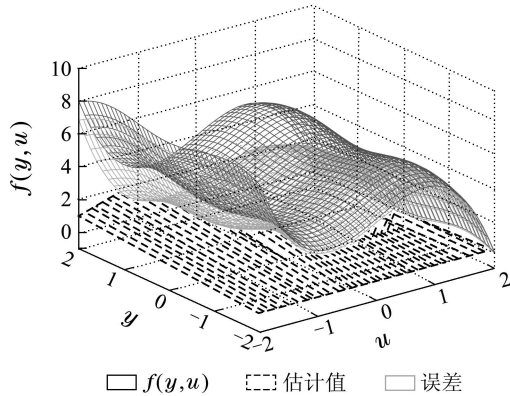


图 6 利用全递推算算法(9)所得 $f(\phi)$ 的最终估计效果
Fig. 6 Estimate for $f(\phi)$ given by algorithm (9)

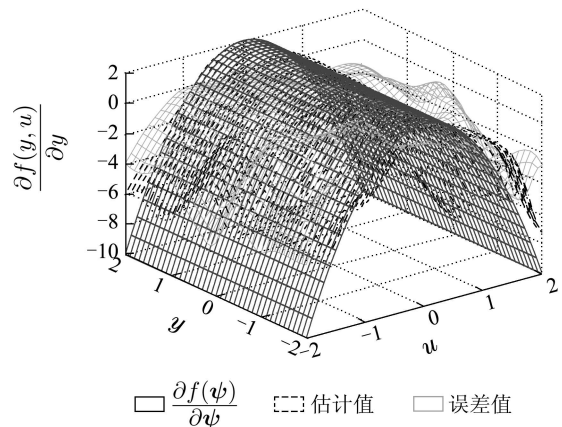


图 9 利用递推局部线性估计算法(13)所得 $\frac{\partial f(y, u)}{\partial u}$ 的最终估计效果
Fig. 9 Estimate for $\frac{\partial f(y, u)}{\partial u}$ given by algorithm (13)

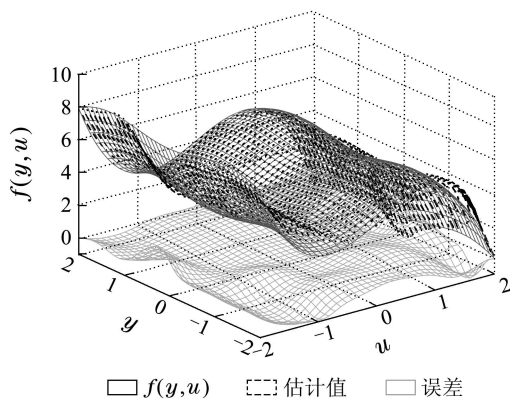


图 7 利用递推局部线性估计算法(13)所得 $f(\phi)$ 的最终估计效果
Fig. 7 Estimate for $f(\phi)$ given by algorithm (13)

从上面的比较得知, 相比半递推和全递推的RW算法, 局部线性回归估计能显著地改进边界误差, 有较大的优越性.

7 总结(Conclusions)

本文研究了局部线性回归估计, 导出了递推算算法, 从而给出了回归函数及其导函数的非参数估计. 在一定的条件下, 证明了算法的强一致性. 并研究了递推局部多项式回归估计在非线性和方差模型的回归函数估计和非线性ARX系统辨识中的应用.

在更一般的条件下算法的收敛性值得进一步研究. 进而, 研究递推局部多项式回归估计在非线性和自适应控制中的应用是个非常有趣的课题.

以上3种算法中核函数 K 均采用了 Gauss 核. 图中用实线描述真实曲面, 用短划线表示估计的曲面, 估计误差则用点线的曲面图置于其下. 其中本文将区间 $[-2, 2] \times [-2, 2]$ 分成了 40×40 小格.

此外, 从递推的局部线性回归估计(13)还得到了回归函数的偏导数估计, 效果如图8和图9所示.

参考文献(References):

- [1] STONE C J. Consistent nonparametric regression [J]. *The Annals of Statistics*, 1977, 5(4): 595 – 620.
- [2] FAN J, GIJBELS I. *Local Polynomial Modelling and Its Applications*[M]. Boca Raton, FL: Chapman and Hall/CRC, 1996.
- [3] RÉVÉSZ P. Robbins-Monro procedure in a Hilbert space and its application in the theory of learning processes. i [J]. *Studia Scientiarum Mathematicarum Hungarica*, 1973, 8: 391 – 398.
- [4] KRZYŻAK A, PAWLAK M. Almost everywhere convergence of a recursive regression function estimate and classification [J]. *IEEE Transactions on Information Theory*, 1984, 30(1): 91 – 93.
- [5] GREBLICKI W, PAWLAK M. Necessary and sufficient consistency conditions for a recursive kernel regression estimate [J]. *Journal of Multivariate Analysis*, 1987, 23(1): 67 – 76.
- [6] DUFLO M. *Random Iterative Models* [M]. Berlin/Heidelberg: Springer-Verlag, 1997.
- [7] GREBLICKI W. Stochastic approximation in nonparametric identification of Hammerstein systems [J]. *IEEE Transactions on Automatic Control*, 2002, 47(11): 1800 – 1810.
- [8] CHEN X M, CHEN H F. Recursive identification for MIMO Hammerstein systems [J]. *IEEE Transactions on Automatic Control*, 2011, 56(4): 895 – 902.
- [9] HU X L, CHEN H F. Strong consistency of recursive identification for Wiener systems [J]. *Automatica*, 2005, 41(11): 1905 – 1916.

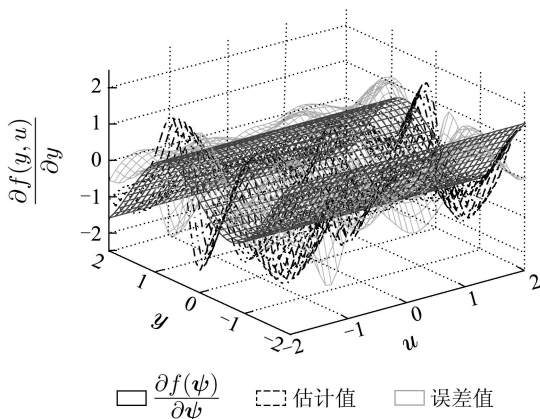


图 8 利用递推局部线性估计算法(13)所得 $\frac{\partial f(y, u)}{\partial y}$ 的最终估计效果
Fig. 8 Estimate for $\frac{\partial f(y, u)}{\partial y}$ given by algorithm (13)

[10] ZHAO W X, CHEN H F, ZHENG W X. Recursive identification for nonlinear ARX systems based on stochastic approximation algorithm [J]. *IEEE Transactions on Automatic Control*, 2010, 55(6): 1287 – 1299.

[11] HILGERT N, ROSSI V, VILA J P, et al. Identification, estimation, and control of uncertain dynamic systems: A nonparametric approach [J]. *Communications in Statistics — Theory and Methods*, 2007, 36(14): 2509 – 2525.

[12] MASRY E. Multivariate local polynomial regression for time series: Uniform strong consistency and rates [J]. *Journal of Time Series Analysis*, 1996, 17(6): 571 – 599.

[13] MASRY E. Multivariate regression estimation local polynomial fitting for time series [J]. *Stochastic Processes and their Applications*, 1996, 65(1): 81 – 101.

[14] MASRY E, FAN J. Local polynomial estimation of regression functions for mixing processes [J]. *Scandinavian Journal of Statistics*, 1997, 24(2): 165 – 179.

[15] BAI E W. Non-parametric nonlinear system identification: An asymptotic minimum mean squared error estimator [J]. *IEEE Transactions on Automatic Control*, 2010, 55(7): 1615 – 1626.

[16] VILAR-FERNÁNDEZ J A, VILAR-FERNÁNDEZ J M. Recursive estimation of regression functions by local polynomial fitting [J]. *Annals of the Institute of Statistical Mathematics*, 1998, 50(4): 729 – 754.

[17] BELOMESTNY D, KOLODKO A, SCHOENMAKERS J. Regression methods for stochastic control problems and their convergence analysis [J]. *SIAM Journal on Control and Optimization*, 2010, 48(5): 3562 – 3588.

[18] MASRY E, TJØTHEIM D. Nonparametric estimation and identification of nonlinear ARCH time series: Strong convergence and asymptotic normality [J]. *Econometric Theory*, 1995, 11(2): 258 – 289.

[19] GREBLICKI W, PAWLAK M. *Nonparametric System Identification* [M]. Cambridge: Cambridge University Press, 2008.

[20] FAN J, YAO Q. *Nonlinear Time Series: Nonparametric and Parametric Methods* [M]. New York: Springer-Verlag, 2003.

[21] MASRY E. Almost sure convergence of recursive density estimators for stationary mixing processes [J]. *Statistics & Probability Letters*, 1987, 5(4): 249 – 254.

[22] WHEEDEN R L, ZYGMUND A. *Measure and Integral: An Introduction to Real Analysis* [M]. New York: Marcel Dekker, 1977.

附录 定理的证明(Appendix Proof of theorems)

为了证明该算法的收敛性, 下面关于 $\hat{\theta}_k$ 的分解是重要的.

$$\hat{\theta}_k = S_k^{-1} \Phi_k^T W_k \begin{bmatrix} f(\phi_1) \\ f(\phi_2) \\ \vdots \\ f(\phi_k) \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_k \end{bmatrix} = S_k^{-1} \Phi_k^T W_k \left\{ \begin{bmatrix} 1 & (\phi_1 - \phi)^T \\ 1 & (\phi_2 - \phi)^T \\ \vdots & \vdots \\ 1 & (\phi_k - \phi)^T \end{bmatrix} \begin{bmatrix} f(\phi) \\ \frac{\partial f}{\partial \phi}(\phi) \end{bmatrix} + \frac{1}{2} \begin{bmatrix} (\phi_1 - \phi)^T H(\phi)(\phi_1 - \phi) \\ (\phi_2 - \phi)^T H(\phi)(\phi_2 - \phi) \\ \vdots \\ (\phi_k - \phi)^T H(\phi)(\phi_k - \phi) \end{bmatrix} \right\} +$$

$$S_k^{-1} \Phi_k^T W_k \begin{bmatrix} r(\phi_1 - \phi) \\ r(\phi_2 - \phi) \\ \vdots \\ r(\phi_k - \phi) \end{bmatrix} + S_k^{-1} \Phi_k^T W_k \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_k \end{bmatrix} = \begin{bmatrix} f(\phi) \\ \frac{\partial f}{\partial \phi}(\phi) \end{bmatrix} + \frac{1}{2} S_k^{-1} \Phi_k^T W_k \begin{bmatrix} (\phi_1 - \phi)^T H(\phi)(\phi_1 - \phi) \\ (\phi_2 - \phi)^T H(\phi)(\phi_2 - \phi) \\ \vdots \\ (\phi_k - \phi)^T H(\phi)(\phi_k - \phi) \end{bmatrix} + S_k^{-1} \Phi_k^T W_k \begin{bmatrix} r(\phi_1 - \phi) \\ r(\phi_2 - \phi) \\ \vdots \\ r(\phi_k - \phi) \end{bmatrix} + S_k^{-1} \Phi_k^T W_k \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_k \end{bmatrix}, \quad (24)$$

其中 $r(\phi_t - \phi)$ 为 f 在 ϕ 点做Taylor展开的余项:

$$r(\phi_t - \phi) = 2 \sum_{|\alpha|=2} \frac{(\phi_t - \phi)^\alpha}{\alpha!} \int_0^1 (1-w) \cdot \{ [D^\alpha f](\phi + w(\phi_t - \phi)) - [D^\alpha f](\phi) \} dw. \quad (25)$$

在这里使用了多个指标记号: 对 d 元的多指标 $\alpha = [\alpha_1 \cdots \alpha_d]^T \in \mathbb{N}_0^d$ 和向量 $\phi = [\phi_1 \cdots \phi_d]^T \in \mathbb{R}^d$, 定义运算 $|\alpha| = \sum_{i=1}^d \alpha_i$, $\phi^\alpha = \phi_1^{\alpha_1} \cdots \phi_d^{\alpha_d}$, $D^\alpha = \partial^\alpha = \frac{\partial^{\alpha_1}}{\partial \phi_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_d}}{\partial \phi_d^{\alpha_d}}$.

假设 θ^* 为真值(即 f 在 ϕ 点的值和梯度), 式(24)可记为

$$\hat{\theta}_k - \theta^* = \frac{1}{2} S_k^{-1} B_k + S_k^{-1} R_k + S_k^{-1} V_k, \quad (26)$$

其中:

$$S_k = \begin{bmatrix} \frac{1}{k} \sum_{t=1}^k K_{h_t}(\phi_t - \phi) & \frac{1}{k} \sum_{t=1}^k \mathcal{E} \\ \frac{1}{k} \sum_{t=1}^k K_{h_t}(\phi_t - \phi)(\phi_t - \phi) & \frac{1}{k} \sum_{t=1}^k \mathcal{F} \end{bmatrix},$$

$$T_k = \begin{bmatrix} \frac{1}{k} \sum_{t=1}^k K_{h_t}(\phi_t - \phi) y_t \\ \frac{1}{k} \sum_{t=1}^k K_{h_t}(\phi_t - \phi)(\phi_t - \phi) y_t \end{bmatrix},$$

$$B_k = \begin{bmatrix} \frac{1}{k} \sum_{t=1}^k \mathcal{E} H(\phi)(\phi_t - \phi) \\ \frac{1}{k} \sum_{t=1}^k \mathcal{F} H(\phi)(\phi_t - \phi) \end{bmatrix},$$

$$R_k = \begin{bmatrix} \frac{1}{k} \sum_{t=1}^k K_{h_t}(\phi_t - \phi) r(\phi_t - \phi) \\ \frac{1}{k} \sum_{t=1}^k K_{h_t}(\phi_t - \phi)(\phi_t - \phi) r(\phi_t - \phi) \end{bmatrix},$$

其中: $\mathcal{E} = K_{h_t}(\phi_t - \phi)(\phi_t - \phi)^T$, $\mathcal{F} = K_{h_t}(\phi_t - \phi)(\phi_t - \phi)(\phi_t - \phi)^T$, V_k 与 R_k 类似, 只是将其中求和的元 $r(\phi_t - \phi)$ 改为 v_t .

为了符号的简便, 记

$$Z_t^0 \triangleq K_{h_t}(\phi_t - \phi), \quad Z_t^1 \triangleq K_{h_t}(\phi_t - \phi)(\phi_t - \phi),$$

$$\begin{aligned} Z_t^2 &\triangleq K_{h_t}(\phi_t - \phi)(\phi_t - \phi)(\phi_t - \phi)^T, \\ Z_t^{H,2} &\triangleq K_{h_t}(\phi_t - \phi)(\phi_t - \phi)^T H(\phi)(\phi_t - \phi), \\ Z_t^{H,3} &\triangleq K_{h_t}(\phi_t - \phi)(\phi_t - \phi)(\phi_t - \phi)^T H(\phi)(\phi_t - \phi), \\ Z_t^{R,0} &\triangleq K_{h_t}(\phi_t - \phi)r(\phi_t - \phi), \\ Z_t^{R,1} &\triangleq K_{h_t}(\phi_t - \phi)(\phi_t - \phi)r(\phi_t - \phi), \\ Z_t^{V,0} &\triangleq K_{h_t}(\phi_t - \phi)v_t, \\ Z_t^{V,1} &\triangleq K_{h_t}(\phi_t - \phi)(\phi_t - \phi)v_t. \end{aligned}$$

下面分别分析 S_k, B_k, R_k 和 V_k , 其中主要利用Masry在1987年得出的关于 α -混合过程的a.s.收敛性定理(见文献[21]中定理3)、Bochner引理(见文献[22]中的定理9.9)、Toeplitz引理和Kronecker引理.

定理 3^[21] 假设 $\{X_k\}$ 是 α -混合过程, 其混合系数为 $\alpha(k), g_k$ 是 \mathbb{R}^d 上的一序列Borel可测函数. 令 $Z_k = g_k(X_k) - E[g_k(X_k)]$ 和 $S_n = \sum_{k=1}^n Z_k$. 若

$$\sum_{k=1}^{\infty} [E(|Z_k|^\delta)]^{\frac{2}{\delta}} < \infty, \tag{27}$$

$$\sum_{n=1}^{\infty} \{\log n (\log \log n)^{1+\varepsilon} [\alpha(n)]^{1-\frac{2}{\delta}} \sum_{k=n}^{\infty} [E(|Z_k|^\delta)]^{\frac{2}{\delta}}\} < \infty \tag{28}$$

对某 $\varepsilon > 0$ 和 $\delta > 2$ 成立, 则当 $n \rightarrow \infty$ 时, $\sum_{k=1}^n Z_k$ 几乎处处收敛到一个有限的极限.

引理 1 设回归过程 $\{\phi_k\}$ 满足条件A2), 若核函数满足K1)–K3), 其窗宽选取满足H1)–H3), 则

$$H_k^{-1} S_k H_k^{-1} \xrightarrow[k \rightarrow \infty]{\text{a.s.}} q(\phi) S, \quad h_k^{-2} H_k^{-1} B_k \xrightarrow[k \rightarrow \infty]{\text{a.s.}} q(\phi) B, \tag{29}$$

其中:

$$\begin{aligned} H_k &= \text{diag}\{1, h_k, \dots, h_k\}, \\ S &= \begin{bmatrix} 1 & \vartheta_1 \int \mathbf{x}^T K(\mathbf{x}) d\mathbf{x} \\ \vartheta_1 \int \mathbf{x} K(\mathbf{x}) d\mathbf{x} & \vartheta_2 \int \mathbf{x} \mathbf{x}^T K(\mathbf{x}) d\mathbf{x} \end{bmatrix}, \\ B &= \begin{bmatrix} \vartheta_2 \int \mathbf{x}^T H(\phi) \mathbf{x} K(\mathbf{x}) d\mathbf{x} \\ \vartheta_3 \int \mathbf{x} \mathbf{x}^T H(\phi) \mathbf{x} K(\mathbf{x}) d\mathbf{x} \end{bmatrix}, \end{aligned}$$

式中: $q(\phi)$ 是过程 ϕ_k 的平稳分布的概率密度函数 q 在 ϕ 点的值(见条件A2)), ϑ_j 是在窗宽的选取条件H3)中所用:

$$\vartheta_j \triangleq \lim_{k \rightarrow \infty} \frac{1}{k} \sum_{t=1}^k \left(\frac{h_t}{h_k}\right)^{dj} < \infty,$$

$j = 0, 1, 2, 3$. 由于核函数 K 是对称的, 则

$$\int \mathbf{x} K(\mathbf{x}) d\mathbf{x} = \int \mathbf{x}^T K(\mathbf{x}) d\mathbf{x} = 0,$$

从而 S 是正定的.

证 记

$$\begin{aligned} \kappa_0 &\triangleq 1, \quad \kappa_1 \triangleq \int \mathbf{x} K(\mathbf{x}) d\mathbf{x}, \quad \kappa_2 \triangleq \int \mathbf{x} \mathbf{x}^T K(\mathbf{x}) d\mathbf{x}, \\ \kappa_2^H &\triangleq \int \mathbf{x}^T H(\phi) \mathbf{x} K(\mathbf{x}) d\mathbf{x}, \quad \kappa_3^H \triangleq \int \mathbf{x} \mathbf{x}^T H(\phi) \mathbf{x} K(\mathbf{x}) d\mathbf{x}. \end{aligned}$$

将 $s_k^j \triangleq \frac{1}{k} \sum_{t=1}^k Z_t^j$ 按照下式分解为

$$\begin{aligned} h_k^{-j} s_k^j - \vartheta_j \kappa_j q(\phi) &= \\ h_k^{-j} (s_k^j - E[s_k^j]) + (h_k^{-j} E[s_k^j] - \vartheta_j \kappa_j q(\phi)) &\triangleq \\ \Delta_r + \Delta_d, \quad j = 0, 1, 2. \end{aligned} \tag{30}$$

同理可将 $b_k^l \triangleq \frac{1}{k} \sum_{t=1}^k Z_t^{H,l}$ 分解为

$$\begin{aligned} h_k^{-l} b_k^l - \vartheta_l \kappa_l^H &= \\ h_k^{-l} (b_k^l - E[b_k^l]) + (h_k^{-l} E[b_k^l] - \vartheta_l \kappa_l^H) &\triangleq \\ \Delta_{H,r} + \Delta_{H,d}, \quad l = 2, 3. \end{aligned} \tag{31}$$

由条件A2): 过程 $\{\phi_k\}$ 有渐近平稳分布 $\pi(d\phi) = q(\phi)d\phi$, 这意味着存在初始分布 π 使得过程 $\{\phi_k\}$ 是平稳的. 因此, 不失一般性, 假设过程 $\{\phi_k\}$ 是平稳的. 由条件K1), K2)和H1), 利用Bochner引理, 可求得极限:

$$\begin{aligned} E[Z_k^0] &= E[K_{h_k}(\phi_k - \phi)] = \\ \int_{\mathbb{R}^d} \frac{1}{h_k^d} K\left(\frac{\phi_k - \phi}{h_k}\right) q(\phi_k) d\phi_k &\xrightarrow[k \rightarrow \infty]{} q(\phi), \end{aligned} \tag{32}$$

$$\begin{aligned} h_k^{-1} E[Z_k^1] &= h_k^{-1} E[K_{h_k}(\phi_k - \phi)(\phi_k - \phi)] = \\ \int_{\mathbb{R}^d} \left(\frac{\phi_k - \phi}{h_k}\right) \frac{1}{h_k^d} K\left(\frac{\phi_k - \phi}{h_k}\right) q(\phi_k) d\phi_k &\xrightarrow[k \rightarrow \infty]{} q(\phi) \kappa_1. \end{aligned} \tag{33}$$

同理可得

$$\begin{aligned} h_k^{-2} E[Z_k^2] &\xrightarrow[k \rightarrow \infty]{} \kappa_2 q(\phi), \\ h_k^{-l} E[Z_k^{H,l}] &\xrightarrow[k \rightarrow \infty]{} \kappa_l^H q(\phi), \quad l = 2, 3. \end{aligned} \tag{34}$$

结合条件H3), 根据Toeplitz引理, 可得

$$\begin{aligned} h_k^{-j} E[s_k^j] &= \frac{1}{k h_k^j} \sum_{t=1}^k E[Z_t^j] = \\ \frac{1}{k} \sum_{t=1}^k \left(\frac{h_t}{h_k}\right)^j h_t^{-j} E[Z_t^j] &\xrightarrow[k \rightarrow \infty]{} \vartheta_j \kappa_j q(\phi), \quad j = 0, 1, 2, \\ h_k^{-l} E[b_k^l] &= \frac{1}{k h_k^l} \sum_{t=1}^k E[Z_t^{H,l}] = \\ \frac{1}{k} \sum_{t=1}^k \left(\frac{h_t}{h_k}\right)^l h_t^{-l} E[Z_t^{H,l}] &\xrightarrow[k \rightarrow \infty]{} \vartheta_l \kappa_l^H q(\phi), \quad l = 2, 3 \end{aligned} \tag{36}$$

成立, 因此 Δ_d 和 $\Delta_{H,d}$ 当 $k \rightarrow \infty$ 时趋于零.

根据条件K1), K3)和H1), 利用Bochner引理可得

$$\begin{aligned} E[(Z_k^0)^\delta] &= E[(K_{h_k}(\phi_k - \phi))^\delta] = \\ \int_{\mathbb{R}^d} \frac{1}{h_k^{d\delta}} K^\delta\left(\frac{\phi_k - \phi}{h_k}\right) q(\phi_k) d\phi_k &= \\ h_k^{-(\delta-1)d} \int_{\mathbb{R}^d} \frac{1}{h_k^\delta} K^\delta\left(\frac{\phi_k - \phi}{h_k}\right) q(\phi_k) d\phi_k &= \\ O(h_k^{-(\delta-1)d}). \end{aligned} \tag{37}$$

同理可得

$$E[\|Z_k^j\|^\delta] = O(h_k^{-(\delta-1)d+j\delta}), \quad j = 1, 2; \tag{38}$$

$$E[\|Z_k^{H,l}\|^\delta] = O(h_k^{-(\delta-1)d+l\delta}), \quad l = 2, 3. \tag{39}$$

记

$$Y_t^j \triangleq \frac{1}{th_t^j} (Z_t^j - E[Z_t^j]), j = 0, 1, 2;$$

$$Y_t^{H,l} \triangleq \frac{1}{th_t^j} (Z_t^{H,l} - E[Z_t^{H,l}]), l = 2, 3.$$

根据前面对 $E[\|Z_k^j\|^\delta]$ 和 $E[\|Z_k^{H,l}\|^\delta]$ 的阶的估计(37)–(39), 有

$$E[\|Y_t^j\|^\delta] = O\left(\frac{h_k^{-(\delta-1)+j\delta}}{t^\delta h_t^{j\delta}}\right) = O\left(\frac{1}{t^\delta h_t^{(\delta-1)}}\right), j = 0, 1, 2, \tag{40}$$

从而有

$$\{E[\|Y_t^j\|^\delta]\}^{2/\delta} = O((th_t^{(1-1/\delta)})^{-2}), j = 0, 1, 2. \tag{41}$$

同理有

$$\{E[\|Y_t^{H,l}\|^\delta]\}^{2/\delta} = O((th_t^{(1-1/\delta)})^{-2}), l = 2, 3. \tag{42}$$

因此, 由假设A2)和H3), 结合上面这些式子, [21]中定理3的条件满足, 从而当 $k \rightarrow \infty$ 时, $\sum_{t=1}^k Y_t^j$ a.s.收敛到有限的极限, $j = 0, 1, 2$; 当 $k \rightarrow \infty$ 时, $\sum_{t=1}^k Y_t^{H,l}$ a.s.收敛到有限的极限, $l = 2, 3$.

再根据Kronecker引理有

$$\Delta_r = \frac{1}{kh_k^j} \sum_{t=1}^k th_t Y_t^j \xrightarrow[k \rightarrow \infty]{\text{a.s.}} 0, j = 0, 1, 2, \tag{43}$$

和

$$\Delta_{H,r} = \frac{1}{kh_k^l} \sum_{t=1}^k th_t Y_t^{H,l} \xrightarrow[k \rightarrow \infty]{\text{a.s.}} 0, l = 2, 3. \tag{44}$$

因此, 引理得证.

引理 2 设回归模型(1)或(2)满足A1)–A3), 若核函数满足K1)–K3), 其窗宽选取满足H1)–H3), 则

$$h_k^{-(2+j)} \frac{1}{k} \sum_{t=1}^k Z_t^{R,j} \xrightarrow[k \rightarrow \infty]{\text{a.s.}} 0, j = 0, 1, \tag{45}$$

从而

$$h_k^{-2} H_k^{-1} R_k \xrightarrow[k \rightarrow \infty]{\text{a.s.}} 0. \tag{46}$$

证 将 $r_t^j \triangleq \frac{1}{k} \sum_{t=1}^k Z_t^{R,j}$ 按照下式分解为

$$h_k^{-(2+j)} r_t^j = h_k^{-(2+j)} (r_t^j - E[r_t^j]) + h_k^{-(2+j)} E[r_t^j] \triangleq R_r + R_d, j = 0, 1. \tag{47}$$

注意到将 $f(\phi_k)$ 在 ϕ 处的Taylor余项(25)代入 $Z_k^{R,j}$, $j = 0, 1$, 并求期望得

$$E[Z_k^{R,j}] = E[K_{h_k}(\phi_k - \phi)(\phi_k - \phi)^j r(\phi_k - \phi)] = 2 \int_{\mathbb{R}^d} \frac{1}{h_k^d} K\left(\frac{\phi_k - \phi}{h_k}\right) (\phi_k - \phi)^j q(\phi_k) \sum_{|\alpha|=2} \frac{(\phi_k - \phi)^\alpha}{\alpha!} \cdot \int_0^1 (1-w) \{ [D^\alpha f](\phi + w(\phi_k - \phi)) - [D^\alpha f](\phi) \} dw d\phi_k = 2h_k^{j+2} \int_{\mathbb{R}^d} \frac{1}{h_k^{d-1}} K(u) u^j \sum_{|\alpha|=2} \frac{u^\alpha}{\alpha!} I^\alpha(\phi, u) q(\phi + h_k u) du, \tag{48}$$

其中:

$$I^\alpha(\phi, u) \triangleq \int_0^1 (1-w) \{ [D^\alpha f](\phi + wh_k u) - [D^\alpha f](\phi) \} dw. \tag{49}$$

注意: 这里为了符号的简便, 对向量 $\phi \in \mathbb{R}^d$, 令 $\phi^0 = 1$, $\phi^1 = \phi$.

由A1)知, $|I^\alpha(\phi, u)| = O(h_k \|u\| \int_0^1 (1-w) dw)$, 因此

$$\|E[Z_k^{R,j}]\| = O\{h_k^{j+3} (\int_0^1 (1-w) dw) \cdot (\int_{\mathbb{R}^d} \frac{1}{h_k^{d-1}} K(u) u^j \sum_{|\alpha|=2} \frac{u^\alpha}{\alpha!} \|u\| q(\phi + h_k u) du)\}. \tag{50}$$

再次利用Bochner引理, 可得

$$E[Z_k^{R,j}] = O(h_k^{j+3}), j = 0, 1. \tag{51}$$

因此,

$$\begin{aligned} \frac{1}{h_k} R_d &= h_k^{-(3+j)} \left\| \frac{1}{k} \sum_{t=1}^k Z_t^{R,j} \right\| = O(h_k^{-(3+j)} \frac{1}{k} \sum_{t=1}^k E\|Z_t^{R,j}\|) = O\left(\frac{1}{k} \sum_{t=1}^k \left(\frac{h_t}{h_k}\right)^{3+j}\right), j = 0, 1, \end{aligned} \tag{52}$$

根据H3), 有

$$R_d = O(h_k) = o(1). \tag{53}$$

对 R_r 利用前面类似的分析, 可证得 $R_r \xrightarrow[k \rightarrow \infty]{\text{a.s.}} 0$.

引理 3 设回归模型(1)或(2)满足A1)–A3), 若核函数满足K1)–K3), 其窗宽选取满足H1)–H3), 则

$$h_k^{-j} \frac{1}{k} \sum_{t=1}^k Z_t^{V,j} \xrightarrow[k \rightarrow \infty]{\text{a.s.}} 0, j = 0, 1, \tag{54}$$

从而

$$H_k^{-1} V_k \xrightarrow[k \rightarrow \infty]{\text{a.s.}} 0. \tag{55}$$

证 记 $Y_t^{V,j} \triangleq \frac{1}{th_t^j} (Z_t^{V,j} - E[Z_t^{V,j}]), j = 0, 1$. 注意到 v_t 与 ϕ_t 是独立的, 由引理1的结论, 有

$$E[\|Y_t^{V,j}\|^\delta] = E[\|Y_t^j\|^\delta] E[\|v_t\|^\delta] = O\left(\frac{1}{t^\delta h_t^{(\delta-1)d}}\right), j = 0, 1, \tag{56}$$

因此, 定理3的条件满足, 从而, 当 $k \rightarrow \infty$ 时, $\sum_{t=1}^k Y_t^{V,j}$ a.s.收敛到有限的极限, $j = 0, 1$.

再根据Toeplitz引理有

$$\frac{1}{kh_k^j} \sum_{t=1}^k th_t Y_t^{V,j} \xrightarrow[k \rightarrow \infty]{\text{a.s.}} 0, j = 0, 1. \tag{57}$$

证 (定理1的证明) 根据前面对 $\hat{\theta}_k - \theta^*$ 的分解式(26), 其中:
 两边同时左乘 H_k , 可得

$$\begin{aligned} H_k(\hat{\theta}_k - \theta^*) &= \\ & \frac{1}{2}(H_k^{-1}S_kH_k^{-1})^{-1}(h_k^{-2}H_k^{-1}B_k)h_k^2 + \\ & (H_k^{-1}S_kH_k^{-1})^{-1}(h_k^{-2}H_k^{-1}R_k)h_k^2 + \\ & (H_k^{-1}S_kH_k^{-1})^{-1}(H_k^{-1}V_k) \triangleq \\ & \Upsilon_B + \Upsilon_R + \Upsilon_V. \end{aligned} \tag{58}$$

从引理1-3, 可以知道 $\Upsilon_B = O(h_k^2)$, $\Upsilon_R = o(h_k^2)$, $\Upsilon_V = o(1)$. 因此定理1得证.

证 (定理2的证明) 令

$$\rho_k = \left(\frac{kh_k^{2(1-1/\delta)d}}{\log k(\log \log k)^{1+\varepsilon}} \right)^{1/2}.$$

注意到式(58), 本文只需要证明

$$\rho_k h_k^{-j} \frac{1}{k} \sum_{t=1}^k Z_t^{V,j} \xrightarrow[k \rightarrow \infty]{\text{a.s.}} 0, \quad j = 0, 1. \tag{59}$$

事实上, 可将上式左端分解为

$$\begin{aligned} \rho_k h_k^{-j} \frac{1}{k} \sum_{t=1}^k Z_t^{V,j} &= \\ & \frac{1}{\varrho_k} \sum_{t=1}^k \varrho_t [(X_t^{V,j} - E[X_t^{V,j}]) + E[X_t^{V,j}]], \end{aligned} \tag{60}$$

$$\varrho_k = \frac{kh_k^j}{\rho_k}, \quad X_t^{V,j} = \frac{\rho_k}{kh_k^j} Z_t^{V,j}.$$

注意到 $E[X_t^{V,j}] = 0$.

类似定理1的证明, 可以得出

$$\begin{aligned} & \{E[\|X_t^{V,j}\|^\delta]\}^{2/\delta} = \\ & O\left(\frac{1}{\log t(\log \log t)^{1+\varepsilon}}\right)^2, \quad j = 0, 1. \end{aligned} \tag{61}$$

因此, 由假设A2')和式(61), 定理3的条件满足, 从而 $\sum_{t=1}^k X_t^{V,j}$ 当 $k \rightarrow \infty$ 时 a.s. 收敛到有限的极限, $j = 0, 1$.

再根据Toeplitz引理有

$$\frac{1}{\varrho_k} \sum_{t=1}^k \varrho_t X_t^{V,j} \xrightarrow[k \rightarrow \infty]{\text{a.s.}} 0, \quad j = 0, 1. \tag{62}$$

再次利用Kronecker引理可得定理2结论.

作者简介:

陈性敏 (1984-), 男, 讲师, 研究方向为系统辨识与控制、非参数统计、随机优化, E-mail: xingmin.chen@gmail.com;

高超 (1986-), 女, 博士研究生, 目前研究方向为分布参数系统、随机服务系统、泛函分析, E-mail: gaochao198604@126.com.