

一种基于 RKHS 及半参数理论的非线性充分降维方法

崔文泉, 刘波

(中国科学技术大学管理学院统计与金融系, 安徽合肥 230026)

摘要:提出了一种基于再生核 Hilbert 空间(reproducing kernel Hilbert space, RKHS)及半参数理论的非线性充分降维新方法——广义半参数核切片逆回归(generalized semiparametric kernel sliced inverse regression, generalized semi-KSIR 或 GSKSIR). 该方法将经典的半参数方法拓展至感兴趣参数为无穷维的广义半参数方法, 将半参数模型推广到不仅冗余参数为无穷维而且感兴趣参数也可为无穷维的广义半参数模型情形, 推导出相应的广义冗余切平面之正交补空间, 进而构造了降维方向的估计方程, 并由 RKHS 理论及正则化方法完成相应目标函数的求解, 求得具有优良性质的非线性充分降维子空间的估计, 并且新方法不需要切片逆回归(SIR)与核切片逆回归(KSIR)等方法所要求的基本的线性设计条件(linear design condition, LDC), 适用性较广. 最后进行了统计模拟研究, 显示了新方法在有限样本下具有良好表现.

关键词:再生核; 半参方法; 冗余切平面; 切片逆回归; 非线性充分降维

中图分类号: O212.7 **文献标识码:** A doi:10.3969/j.issn.0253-2778.2016.11.004

2010 Mathematics Subject Classification: Primary 62G8; Secondary 62N99

引用格式: 崔文泉, 刘波. 一种基于 RKHS 及半参数理论的非线性充分降维方法[J]. 中国科学技术大学学报, 2016, 46(11): 989-906.

CUI Wenquan, LIU Bo. A RKHS-based semiparametric approach to nonlinear dimension reduction[J]. Journal of University of Science and Technology of China, 2016, 46(11): 989-906.

A RKHS-based semiparametric approach to nonlinear dimension reduction

CUI Wenquan, LIU Bo

(Department of Statistics and Finance, School of Management, University of Science and of Technology of China, Hefei 230026, China)

Abstract: A nonlinear dimension reduction method, the generalized semiparametric kernel sliced inverse regression (GSKSIR for short), was proposed, developed based on the theory of reproducing kernel Hilbert Space (RKHS) and the semiparametric method. The method extends the classical semiparametric method into a more generalized semiparametric domain, and is capable of handling infinite dimensional interested a parameter spaces. With this method, both spaces of nuisance parameters and parameters of interests can be infinitely dimensional, the corresponding generalized nuisance tangent space orthogonal complement was derived, estimation equation for the purpose of dimension reduction was constructed, and optimization of the target function could be achieved based on RKHS theory and regularization method, which leads to a nonlinear estimated sufficient reduced dimension subspace with efficient properties.

收稿日期: 2016-03-17; 修回日期: 2016-06-05

基金项目: 安徽省自然科学基金(1308085MA02), 中国科学院知识创新计划(KJ CX3-SYW-S02)资助.

作者简介: 崔文泉(通讯作者), 男, 1964年生, 博士. 研究方向: 数理统计. E-mail: wqcui@ustc.edu.cn

Furthermore, this new method does not impose the linearity design conditions (LDC) required by methods such as the sliced inverse regression (SIR) and the kernel SIR, and so on, and thus, is more general and can be more widely applied. Finally, a Monte Carlo simulation was conducted, and the results demonstrate the excellent finite sample properties of this new method.

Key words: reproducing kernel; semiparametric methods; nuisance tangent space; sliced inverse regression; nonlinear dimension reduction

0 引言

随着大数据时代的到来,高维复杂数据统计分析的需求越来越广泛而深入,如来自生命科学技术中的高维数据、文本挖掘与搜索引擎中的高维数据、高光谱成像中的高维数据等等不同领域的数据分析. 尽管高维数据中包含了人们感兴趣的海量信息,但直接处理高维数据十分困难与棘手,所以如何在尽量没有信息损失的情况下将高维数据转化为可处理的低维模型就显得很有意义. 目前,在有监督学习的框架下,处理具有稀疏性结构的高维数据降维的方法可主要分为变量选择和充分降维 (sufficient dimension reduction, SDR) 两大类. 变量选择假定响应变量只与少数的协变量有关,剩下的协变量对于响应变量不造成任何影响,目的在于选出这些少数的重要变量达到降低维数的目的.

经典的充分降维讨论的是具有线性结构的模型. 假定响应变量 Y 依赖于 p 维协变量 X 的 q ($q < p$) 个线性组合,存在 q 个 p 维向量 $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_q$, 在给定 $X^T(\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_q) \triangleq X^T \mathbf{B}$ 时, Y 与 X 相互独立,即:

$$F(y | X) = F(y | X^T \mathbf{B}), \forall y \in \mathcal{R} \quad (1)$$

式中, $F(y | \cdot) = P(Y \leq y | \cdot)$ 是 Y 的条件分布函数. 记 $S_{Y|X} = \text{span}\{\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_q\}$ 为 SDR 中心子空间^[3], 是所有 SDR 子空间的交. 充分降维的目的在于估计 $S_{Y|X}$, 将 p 维输入 X 降到具有 q 维特征结构的 $X^T \mathbf{B}$. 针对模型(1)已有不少研究工作,其中包括著名的切片逆回归 (sliced inverse regression, SIR) 方法^[9], Cook 等^[17] 提出的切片平均方差估计 (sliced average variance estimation, SAVE) 方法和 Li 等^[18] 提出的海赛主方向 (direction regression, DR) 方法等. 对于模型(1)的上述方法,为了方法实现上的需要,常用如下的假设:线性设计条件 (linear design condition, LDC)

$$E(X | X^T \mathbf{B}) = \mathbf{P}X$$

和常数条件方差假设

$$\text{Cov}(X | X^T \mathbf{B}) = \mathbf{Q}$$

其中, $\mathbf{P} = \mathbf{B}(\mathbf{B}^T \mathbf{B})^{-1} \mathbf{B}^T$ 和 $\mathbf{Q} = \mathbf{I}_p - \mathbf{P}$, 分别为向由 \mathbf{B} 的列张成的线性子空间和正交补空间的正交投影阵. 如 SIR 和 SAVE 的前提是上述两个假设均成立. 文献[10]利用半参数理论提出半参数充分降维方法,该方法不依赖于上述两个假设条件,并可构造相应于 SIR 方法、SAVE 和 DR 的半参数方法,如半参数 SIR (semi-SIR, 或 SSIR) 方法.

模型(1)在具有线性结构的模型中有着广泛的适用性,包含了诸多具有线性结构的模型. 如:线性回归模型、广义线性回归模型、投影寻踪模型、单指标模型、神经网络模型等.

基于模型(1)的充分降维研究获得的是具有线性结构的降维特征,然而在大量的现实问题中,线性结构往往是人们的一种理想或近似假定,降维结构更多的呈现出非线性特征,甚至是高度非线性的结构,发展有效的非线性降维方法是非常重要的研究课题. 非线性 SDR 将模型(1)推广到非线性结构情形,即

$$F(y | X) = F(y | U(X)), \forall y \in \mathcal{R} \quad (2)$$

式中, $U(X) = (u_1(X), \dots, u_q(X))$. 非线性结构自身的复杂性增大了对其进行充分降维研究的难度,使得相应的理论方法的研究进展较为缓慢.

已有若干建立在 RKHS 理论基础上的非线性充分降维方法的研究,将 SIR 发展为处理非线性充分降维的核切片逆回归 (kernel SIR, 或 KSIR) 是一类重要的方法,见文献[19-21]. 文献[5]将 KSIR 推广到具有删失机制的生存数据的充分降维研究中,方法实现简洁表现良好.

本文首次利用 RKHS 及半参数理论,提出一种处理非线性结构下的充分降维问题的新方法——广义半参数核切片逆回归 (generalized semiparametric kernel sliced inverse regression, GSKSIR) 方法,克服了 KSIR 等方法对于线性条件均值假设和常数条件方差假设的依赖,并将半参数方法推广到冗余参数与感兴趣参数皆可为无穷维的情形,适用范围更加广阔.

本文我们利用半参数理论推导出广义冗余切平面之正交补空间,由此构造降维方向的估计方程,接

着建立适当的等距同构映射,将对连续型无穷维函数的求解转化为对离散型数列函数的求解,并利用 RKHS 理论及正则化方法完成对目标函数的求解,得到 SDR 中心子空间的估计.

1 非线性充分降维的广义半参方法 GSKSIR

本文是基于 RKHS 理论提出半参非线性充分降维方法 GSKSIR 的,这里对 RKHS 的相关内容进行简单的回顾与介绍,更多具体的内容读者可参阅文献[1,12].

设 \mathcal{H}_R 是定义在 $\mathcal{X} \subset \mathcal{P}$ 上的 Hilbert 函数空间,若二元函数 $R(s, t), \forall s, t \in \mathcal{X}$, 满足条件: $R_s(t) = R(s, t)$ 是 t 的函数,且对 $\forall s \in \mathcal{X}$ 有 $R_s(t) \in \mathcal{H}_R$; 对 $\forall s \in \mathcal{X}, f \in \mathcal{H}_R$ 有 $f(s) = \langle f, R(s, \cdot) \rangle$ (再生性质). 则称 \mathcal{H}_R 是再生核 Hilbert 空间, R 称为再生核. 由再生核理论知再生核 R 由再生核 Hilbert 空间 \mathcal{H}_R 唯一确定.

若 \mathcal{H}_R 是可分的 Hilbert 空间, $R(s, t)$ 是其再生核, $\{\phi_j\}_{j=1}^{\infty}$ 是 \mathcal{H}_R 的标准正交基, 则对应的 R 有谱分解:

$$R(s, t) = \sum_{j=1}^{\infty} \lambda_j \phi_j(s) \phi_j(t), \forall s, t \in \mathcal{X}$$

且

$$\lim_{n \rightarrow \infty} \left\| R(s, t) - \sum_{j=1}^n \lambda_j \phi_j(s) \phi_j(t) \right\| = 0.$$

定义映射 $\phi: \mathcal{X} \rightarrow L_2$, 使得对 $\forall s \in \mathcal{X}$ 有

$$\phi(s) = (\sqrt{\lambda_1} \phi_1(s), \sqrt{\lambda_2} \phi_2(s), \dots)^T.$$

对 $\forall s, t \in \mathcal{X}$ 定义内积:

$$\langle \phi(s), \phi(t) \rangle = \phi(s)^T \phi(t) = R(s, t).$$

令 \mathcal{H} 为 $\text{span}\{\phi(s), \forall s \in \mathcal{X}\}$ 的完备化后所对应的 Hilbert 空间, 构建映射 $\tau: \mathcal{H}_R \rightarrow \mathcal{H}$, 使得

$$\tau(R_s(\cdot)) = \tau(R(s, \cdot)) = \phi(s), \forall s \in \mathcal{X}$$

且满足对 $\forall s, t \in \mathcal{X}$ 有

$$\langle \tau(R_s(\cdot)), \tau(R_t(\cdot)) \rangle_{\mathcal{H}_R} = \langle \phi(s), \phi(t) \rangle_{\mathcal{H}} = R(s, t).$$

则 τ 是 \mathcal{H}_R 到 \mathcal{H} 上的等距同构映射.

$\phi(X)$ 是取值于 \mathcal{H} 的随机元, 一般地, 对取值于 \mathcal{H} 的随机元 Z , 假设 $E\|Z\|^2 < \infty$, 则 Z 的数学期望定义为 $E(Z) \in \mathcal{H}$, 且满足 $\langle a, E(Z) \rangle = E\langle a, Z \rangle$, 对所有的 $a \in \mathcal{H}$ 成立. 定义 Z 的协方差算子为

$$\text{Cov}(Z) = E[(Z - E(Z)) \otimes (Z - E(Z))]$$

其中运算 \otimes 意义为: 对任意的 $a, b \in \mathcal{H}$ 有

$$(a \otimes b)f = \langle b, f \rangle a, \forall f \in \mathcal{H}.$$

假设响应变量 Y 与协变量 X 满足模型(2), 记

(X, Y) 的样本容量为 1 的样本的似然函数改为

$$\eta_1(X) \eta_2(Y, U(X)) \quad (3)$$

式中, η_1 为 X 的概率密度函数或分布律或者两者的混合, 视 X 的各分量是连续型随机变量还是离散型随机变量而定. 类似地, η_2 是给定 $U(x)$ 条件下 Y 的条件概率密度函数或条件分布率. 由 RKHS 的再生性质, 存在 $\beta_j \in \mathcal{H} (j=1, \dots, q)$, 使得

$$u_j(X) = \langle u_j, R(X, \cdot) \rangle_{\mathcal{H}_R} = \langle \beta_j, \phi(X) \rangle_{\mathcal{H}} \triangleq \beta_j^T \phi(X), \quad j = 1, \dots, q.$$

故

$$U(X) = (\langle \beta_1, \phi(X) \rangle_{\mathcal{H}}, \dots, \langle \beta_q, \phi(X) \rangle_{\mathcal{H}})^T = (\beta_1, \dots, \beta_q)^T \phi(X) \triangleq \beta^T \phi(X),$$

其中, $\beta = (\beta_1, \dots, \beta_q)$, 则式(3)可化简为

$$\eta_1(X) \eta_2(Y, \beta^T \phi(X)) \quad (4)$$

与半参数模型^[2,13]不同, 这里感兴趣参数 β 为无穷维, 冗余参数 η_1, η_2 为无穷维. 本文所提方法首次将感兴趣参数为有限维的半参数方法推广到无穷维感兴趣参数情形. 我们将冗余参数和感兴趣参数均为无穷维的模型称为广义半参数模型 (generalized semiparametric model, GSM). 由经典半参数理论知, 可由冗余切平面的正交补空间构造感兴趣参数的估计方程, 且参数估计具有相合性和渐进正态性等优良性质. 类似的, 本文将对 GSM(4) 构造相应的广义冗余切平面正交补空间 Λ^\perp , 对任意函数 $g(Y, U(X))$ 和 $\alpha(X)$, 可得 Λ^\perp 的函数有如下形式:

$$[g(Y, U(X)) - E\{g(Y, U(X)) | U(X)\}] \times [\alpha(X) - E(\alpha(X) | U(X))].$$

或者

$$[g(Y, \beta^T \phi(X)) - E\{g(Y, \beta^T \phi(X)) | \beta^T \phi(X)\}] \times [\alpha(X) - E(\alpha(X) | \beta^T \phi(X))].$$

因此无穷维感兴趣参数 β 的总体意义下的方程为

$$E\{[g(Y, \beta^T \phi(X)) - E\{g(Y, \beta^T \phi(X)) | \beta^T \phi(X)\}] \otimes [\alpha(X) - E(\alpha(X) | \beta^T \phi(X))]\} = 0.$$

特别地, 取 $g(Y, \beta^T \phi(X)) = E(\phi(X) | Y)$, $\alpha(X) = \phi^T(X)$, 可得总体意义下 GSKSIR 方程:

$$E\{[E(\phi(X) | Y) - E(E(\phi(X) | Y) | \beta^T \phi(X))] \times [\phi(X) - E(\phi(X) | \beta^T \phi(X))]^T\} = 0 \quad (5)$$

记

$$E(\phi(X) | Y) = \xi(Y),$$

$$E(E(\phi(X) | Y) | \beta^T \phi(X)) =$$

$$\zeta(\beta^T \phi(X)), E(\phi(X) | \beta^T \phi(X)) = \theta(\beta^T \phi(X)),$$

并记它们的核估计分别为 $\hat{\xi}(Y)$, $\hat{\zeta}(\beta^T \phi(X))$, $\hat{\theta}(\beta^T \phi(X))$. 给定样本观测数据: $\{(x_i, y_i)\}_{i=1}^n$, 可得式(5)的经验形式的估计方程为

$$\frac{1}{n} \sum_{i=1}^n [(\hat{\xi}(y_i) - \hat{\zeta}(\beta^T \phi(x_i))) \cdot (\phi(x_i) - \hat{\theta}(\beta^T \phi(x_i)))^T] = 0 \quad (6)$$

将式(6)整理成如下形式(推导见附录):

$$\Phi \tilde{\mathbf{K}}_{h_1}^{(1)} (\mathbf{I}_n - \tilde{\mathbf{K}}_{h_2}^{(2)}) (\mathbf{I}_n - (\tilde{\mathbf{K}}_{h_2}^{(2)})^T) \Phi^T = 0 \quad (7)$$

式中,

$$\begin{aligned} \Phi &= (\phi(x_1), \dots, \phi(x_n)), \\ \tilde{\mathbf{K}}_{h_1}^{(1)} &= \left[\frac{K_{h_1}^{(1)}(y_i, y_j)}{\sum_{l=1}^n K_{h_1}^{(1)}(y_l, y_j)} \right]_{1 \leq i, j \leq n}, \\ \tilde{\mathbf{K}}_{h_2}^{(2)} &= \left[\frac{K_{h_2}^{(2)}(\hat{\beta}^T \phi(x_i), \hat{\beta}^T \phi(x_j))}{\sum_{l=1}^n K_{h_2}^{(2)}(\hat{\beta}^T \phi(x_l), \hat{\beta}^T \phi(x_j))} \right]_{1 \leq i, j \leq n} = \\ &= \left[\frac{K_{h_2}^{(2)}(C^T R_i, C^T R_j)}{\sum_{l=1}^n K_{h_2}^{(2)}(C^T R_l, C^T R_j)} \right]_{1 \leq i, j \leq n} \end{aligned} \quad (8)$$

式中, $h_1, K_{h_1}^{(1)}$ 分别是 $E(\phi(X) | Y)$ 非参数核估计的窗宽和核函数; $h_2, K_{h_2}^{(2)}$ 分别是 $E(\phi(X) | \mathbf{B}^T \phi(X))$ 非参数估计的窗宽和核函数; $C = (c_1, \dots, c_q)$, $c_j \in \mathbb{R}^n$, $j \in \{1, \dots, q\}$, R_j 为对应于输入 $\{x_j\}_{i=1}^n$ 的 Gram 矩阵的第 j 列. $\phi(X)$ 是无穷维的. 估计方程(7)及(8)的等式成立见附录.

由于 $\phi(X)$ 是无穷维的, 因此方程(7)不能直接用于进行方程求解, 由下节的定理 2, 可将方程(7)转化为与之等价的如下方程:

$$\mathbf{R} \tilde{\mathbf{K}}_{h_1}^{(1)} (\mathbf{I}_n - \tilde{\mathbf{K}}_{h_2}^{(2)}) (\mathbf{I}_n - (\tilde{\mathbf{K}}_{h_2}^{(2)})^T) \mathbf{R} = 0 \quad (9)$$

式中, \mathbf{R} 是对应 $\{x_i\}_{i=1}^n$ 的 Gram 矩阵.

这样, 我们将对广义半参数非线性充分降维的求解问题转化为对方程(9)进行求解问题, 这样可以利用正则化方法对求解问题进行处理, 进而得到非线性充分降维方法的算法实现.

2 定理及算法实现

2.1 相关定理

冗余切平面及其正交补空间在半参数方法及理论研究中发挥重要的作用, 本文方法是建立在广义半参数模型(2)相应的广义冗余切平面之正交补空间(generalized nuisance tangent space orthogonal

complement, GNTSOC)基础上的. 下面定理 1 给出模型(2)的 GNTSOC.

定理 1 对于广义半参数模型(2), 记 Λ^\perp 为该模型的 GNTSOC, 则

$$\Lambda^\perp = \{f(Y, X) - E(f | U(X), Y); E(f | U(X), \forall f)\},$$

其中, $U(X) = (u_1(X), \dots, u_q(X))^T$, $u_j \in \mathcal{H}_R$ ($j = 1, \dots, q$), \mathcal{H}_R 为一再生核是 R 的 RKHS.

定理 2 方程(7)与方程(9)是等价的, 其中方程(9)中 $\mathbf{R} = \Phi^T \Phi = (R(x_i, x_j))_{1 \leq i, j \leq n}$.

为了实现方程(9)的求解, 使得算法实现具有稳定性, 类似文献[6, 8, 16], 对方程(9)进行一种正则化方法处理, 构造如下的目标函数:

$$A(C) \triangleq || \mathbf{R} \tilde{\mathbf{K}}_{h_1}^{(1)} (\mathbf{I}_n - \tilde{\mathbf{K}}_{h_2}^{(2)}) (\mathbf{I}_n - (\tilde{\mathbf{K}}_{h_2}^{(2)})^T) \mathbf{R} || + \lambda \cdot T(C^T C) \quad (10)$$

对于给定的 λ , 求解使式(10)值最小的 \hat{C} .

2.2 算法实现

下面给出依据目标函数(9)所得到的非线性充分降维的广义半参方法具体的算法步骤.

① 初始化: 对于给定的 λ , 由给定的样本 $\{x_i, y_i\}_{i=1}^n$, 计算 Gram 矩阵 $\mathbf{R} = (R(x_i, x_j))_{1 \leq i, j \leq n}$. 迭代次数 $t = 0$, 取初值 $\hat{C}^{(0)}$, 此处的初值可以选择 KSIR 所得的 \hat{C} .

② 对 $\forall i \in \{1, 2, \dots, m\}, j \in \{1, 2, \dots, q\}$ 计算

$$\hat{C}_{ij+}^{(t)} = \hat{C}^{(t)} + \delta e_{ij},$$

$$\hat{C}_{ij-}^{(t)} = \hat{C}^{(t)} - \delta e_{ij},$$

$$d_{ij} = (\|A(\hat{C}_{ij+}^{(t)})\|^2 - \|A(\hat{C}_{ij-}^{(t)})\|^2) / (2\delta).$$

其中, e_{ij} 是与 $\hat{C}^{(t)}$ 维数相同的矩阵, 除第 i 行第 j 列为 1 外其余位置均为 0. 矩阵 $\mathbf{D}_{n \times q} \triangleq (d_{ij})_{1 \leq i \leq n, 1 \leq j \leq q}$, 有

$$\text{Vec}(\mathbf{D}) = \frac{\partial \{\|A(\hat{C})\|^2\}}{\partial \{\text{Vec}(\hat{C})\}} \Big|_{\hat{C}^{(t)}}.$$

③ 对 $\forall i, k \in \{1, 2, \dots, n\}; j, l \in \{1, 2, \dots, q\}$ 计算

$$\hat{C}_{ijkl++}^{(t)} = \hat{C}^{(t)} + \delta e_{ij} + \delta e_{kl},$$

$$\hat{C}_{ijkl+-}^{(t)} = \hat{C}^{(t)} - \delta e_{ij} + \delta e_{kl},$$

$$\hat{C}_{ijkl+-}^{(t)} = \hat{C}^{(t)} + \delta e_{ij} - \delta e_{kl},$$

$$\hat{C}_{ijkl--}^{(t)} = \hat{C}^{(t)} - \delta e_{ij} - \delta e_{kl},$$

$g_{ij,kl} = (\|r(\hat{C}_{ijkl+}^{(t)})\|^2 - \|r(\hat{C}_{ijkl-}^{(t)})\|^2 - \|r(\hat{C}_{ijkt+}^{(t)})\|^2 + \|r(\hat{C}_{ijkt-}^{(t)})\|^2)/(4\delta^2)$.
 现令矩阵 $\mathbf{G}_{nq \times nq} = (g_{ij,kl})_{1 \leq i,k \leq n, 1 \leq j,l \leq q}$, 有

$$\mathbf{G} = \frac{\partial^2 \{ \|A(\hat{\mathbf{C}})\|^2 \}}{\partial \{ \text{Vec}(\hat{\mathbf{C}}) \} \partial \{ \text{Vec}(\hat{\mathbf{C}}) \}'} \Big|_{\hat{\mathbf{C}}^{(t)}}$$

④ 利用 $\hat{\mathbf{C}}^{(t)}$ 以及求得的矩阵 \mathbf{D} 与 \mathbf{G} 更新 $\hat{\mathbf{C}}$ 为 $\hat{\mathbf{C}}^{(t+1)}$. 首先对更新后的 $\hat{\mathbf{C}}^{(t+1)}$ 进行拉直运算:

$$\text{Vec}(\hat{\mathbf{C}})^{(t+1)} = \text{Vec}(\hat{\mathbf{C}}^{(t)}) - \mathbf{G}^{-1} \text{Vec}(\mathbf{D})$$

由此可得 $\hat{\mathbf{C}}^{(t+1)} = (a_1, \dots, a_q)$. 其次为了得到满足条件 $\hat{\mathbf{C}}^T \mathbf{R} \hat{\mathbf{C}} = \mathbf{I}_q$ 的 $\hat{\mathbf{C}}^{(t+1)}$, 需要对 $\hat{\mathbf{C}}^{(t+1)}$ 的每一列做处理, 即对 $\forall j \in \{1, 2, \dots, q\}$, 令

$$\bar{a}_j^{(t+1)} = \frac{a_j^{(t+1)}}{\sqrt{(a_j^{(t+1)})^T \mathbf{R} a_j^{(t+1)}}},$$

最终得到的 $\hat{\mathbf{C}}^{(t+1)} = (\bar{a}_1, \dots, \bar{a}_q)$ 即为所求的更新值. 先给定收敛的正数 ϵ , 若

$$\| \text{Vec}(\hat{\mathbf{C}}^{(t+1)}) - \text{Vec}(\hat{\mathbf{C}}^{(t)}) \|^2 \leq \epsilon,$$

则迭代停止, 否则令 $t = t + 1$ 返回第一步继续迭代. 由数据驱动方法求出适当的 λ . 本文利用交叉验证法对 λ 进行优选.

3 数值模拟

本节对 GSKSIR 方法进行模拟研究, 针对不同的情形, 分别与非线性降维的 KSIR 及半参数充分降维的 SSIR 方法进行比较, 显示出新方法在有限样本时的良好表现.

3.1 非线性降维情形

分别考虑满足 LDC 条件和不满足 LDC 条件的如下两个非线性模型.

模型 1(不满足 LDC 条件) 假定因变量 Y 满足

$$Y = \frac{\sum_{i=1}^6 X_i^2}{1 + \sum_{i=1}^6 X_i^2} + \epsilon, \epsilon \sim N(0, 1),$$

其中, $(X_1, X_2, X_3, X_4, X_5, X_6)^T$ 由下面方式生成: $(X_1, X_2)^T$ 服从多元正态分布 $N(\mathbf{0}, \mathbf{\Sigma})$, 其中 $\mathbf{\Sigma} = (\sigma_{ij})_{2 \times 2}, \sigma_{ij} = 0.5^{|i-j|}$; $X_3 = |X_1 + X_2| + |X_2| \epsilon_1$; $X_4 = |X_1 + X_2|^2 + |X_2| \epsilon_2$; X_5 服从参数为 $\frac{e^{X_2}}{1 + e^{X_2}}$ 的两点分布; X_6 服从参数为 $\Phi(X_2)$ 的两点分布, 其中

$\Phi(\cdot)$ 为标准正态分布的分布函数. $\{\epsilon_1, \epsilon_2\}$ 独立同分布于标准正态分布.

模型 2(满足 LDC 条件) 假定因变量 Y 满足:

$$Y = -\frac{\ln(U)}{\exp(\sum_{i=1}^6 X_i^2)}, U \sim U(0, 1)$$

其中, $\{X_i\}_{i=1}^6$ 独立同分布于标准正态分布.

下面在不同样本容量下分别用 GSSKIR, KSIR 及 SSIR 进行充分降维, 利用数据驱动获得各方法 tuning 参数的优选, 样本容量分别取 50, 100, 200, GSSKIR 和 KSIR 用多项式核进行再生核求解.

降维的估计方向与 $\sum_{i=1}^6 X_i^2$ 对应真方向之间的相关系数刻画了降维效果, 结果见表 1 和图 1. 模拟结果显示, 在非线性情形, 半参数充分降维的 SSIR 表现得很差. 对于新方法和 KSIR: 在满足 LDC 条件下, GSKSIR 和 KSIR 表现都很好, GSKSIR 比 KSIR 的更好一些, 说明在这种情形下用 KSIR 进行非线性降维效果还是不错的; 在不满足 LDC 条件时, GSKSIR 明显比 KSIR 表现得好, 说明新方法在这种情形下有很好的优势.

表 1 非线性结构下 3 种方法的对比

Tab. 1 Comparison of KSIR, GSKSIR and SSIR under non-linear structure

样本容量	KSIR	GSKSIR	SSIR
模型 1	50 0.828 5(0.135 9)	0.924 4(0.058 4)	0.337 0(0.187 6)
	100 0.877 3(0.064 4)	0.946 5(0.037 1)	0.299 9(0.158 3)
	200 0.907 2(0.034 6)	0.959 1(0.023 7)	0.313 9(0.140 6)
模型 2	50 0.910 1(0.033 5)	0.910 5(0.033 6)	0.158 0(0.117 5)
	100 0.933 8(0.024 2)	0.934 0(0.024 0)	0.108 3(0.081 0)
	200 0.955 1(0.014 8)	0.955 2(0.014 7)	0.085 1(0.063 5)

3.2 线性充分降维情形

模型 3 假定因变量 Y 满足:

$$Y = \frac{1}{\sqrt{6}} \sum_{i=1}^6 X_i + 0.5\epsilon, \epsilon \sim N(0, 1).$$

由此可得真实方向为 $v_1(X) = \sum_{i=1}^6 X_i$, 此时 $q = 1$. 假设协变量 X 具有如下结构: $(X_1, X_2)^T$ 服从多元正态分布 $N(\mathbf{0}, \mathbf{\Sigma})$, 其中 $\mathbf{\Sigma} = (\sigma_{ij})_{2 \times 2}, \sigma_{ij} = 0.5^{|i-j|}$; $X_3 = |X_1 + X_2| + |X_2| \epsilon_1$; $X_4 = |X_1 + X_2|^2 + |X_2| \epsilon_2$; X_5 服从参数为 $\frac{e^{X_2}}{1 + e^{X_2}}$ 的两点分布; X_6 服从参数为 $\Phi(X_2)$ 的两点分布, 其中 $\Phi(\cdot)$ 为标准正态分布的分布函数, $\{\epsilon_1, \epsilon_2\}$ 独立同分布于标准正态分布.

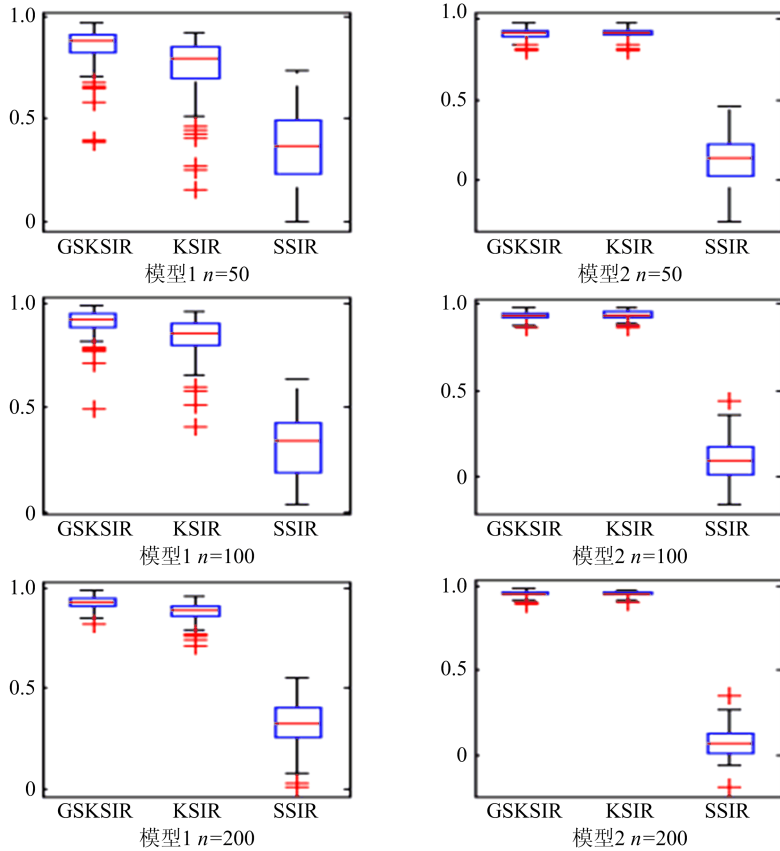


图 1 非线性结构下 3 种方法的对比的箱线图

Fig. 1 Box-plots for comparing KSIR, GSKSIR and SSIR under non-linear structure

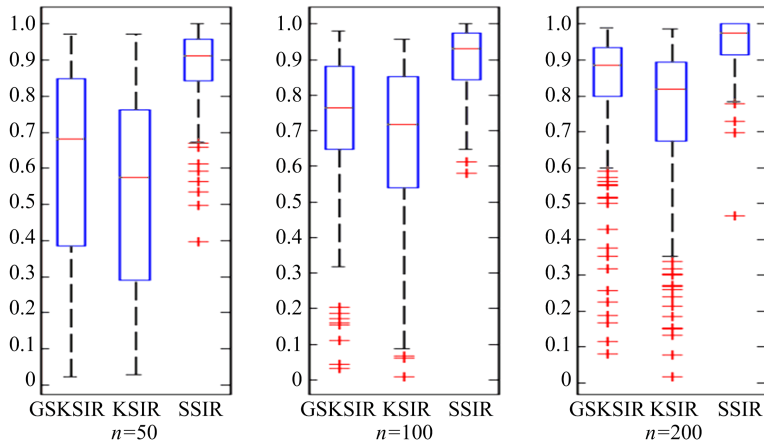


图 2 线性结构下 3 种方法的对比的箱线图

Fig. 2 Box-plots for comparing KSIR, GSKSIR and SSIR under linear structure

选用多项式核,以 KSIR 方法得到的结果为初值. 结果见表 2 和箱线图 2. 结果显示在不满足 LDC 的条件时,在线性结构的情况下 3 种方法的结果都比较好,其中 SSIR 的结果表现得最好. GSKSIR 的表现比专门针对于线性情形的 SSIR 略差,比 KSIR 要好一些,说明在线性结构下新方法有一定的适用性.

表 2 线性结构下 3 种方法的对比

Tab. 2 Comparison of KSIR, GSKSIR and SSIR under linear structure

样本容量	KSIR	GSKSIR	SSIR
模型 50	0.530 8(0.274 1)	0.602 1(0.268 5)	0.874 1(0.123 0)
模型 100	0.659 2(0.237 2)	0.716 8(0.229 2)	0.902 7(0.099 0)
模型 200	0.746 8(0.211 6)	0.830 8(0.171 6)	0.945 9(0.072 9)

附录

A.1 方程(6)的推导

方程(6)的推导如下:对 $\forall i=1,2,\dots,n$ 有

$$\hat{\xi}(y_i) = \sum_{j=1}^n \phi(x_j) K_{h_1}^{(1)}(y_j, y_i) / \sum_{j=1}^n K_{h_1}^{(1)}(y_j, y_i) =$$

$$\Phi \begin{pmatrix} \frac{K_{h_1}^{(1)}(y_1, y_i)}{\sum_{j=1}^n K_{h_1}^{(1)}(y_j, y_i)} \\ \vdots \\ \frac{K_{h_1}^{(1)}(y_n, y_i)}{\sum_{j=1}^n K_{h_1}^{(1)}(y_j, y_i)} \end{pmatrix},$$

$$\hat{\xi}(\beta^T \phi(x_i)) =$$

$$\sum_{j=1}^n \hat{\xi}(y_j) \frac{K_{h_2}^{(2)}(\beta^T \phi(x_j), \beta^T \phi(x_i))}{\sum_{j=1}^n K_{h_2}^{(2)}(\beta^T \phi(x_j), \beta^T \phi(x_i))} =$$

$$\Phi \tilde{K}_{h_1}^{(1)} \begin{pmatrix} \frac{K_{h_2}^{(2)}(\beta^T \phi(x_1), \beta^T \phi(x_i))}{\sum_{j=1}^n K_{h_2}^{(2)}(\beta^T \phi(x_j), \beta^T \phi(x_i))} \\ \vdots \\ \frac{K_{h_2}^{(2)}(\beta^T \phi(x_n), \beta^T \phi(x_i))}{\sum_{j=1}^n K_{h_2}^{(2)}(\beta^T \phi(x_j), \beta^T \phi(x_i))} \end{pmatrix},$$

$$\hat{\theta}(\beta^T \phi(x_i)) =$$

$$\sum_{j=1}^n \phi(x_j) \frac{K_{h_2}^{(2)}(\beta^T \phi(x_j), \beta^T \phi(x_i))}{\sum_{j=1}^n K_{h_2}^{(2)}(\beta^T \phi(x_j), \beta^T \phi(x_i))} =$$

$$\Phi \begin{pmatrix} \frac{K_{h_2}^{(2)}(\beta^T \phi(x_1), \beta^T \phi(x_i))}{\sum_{j=1}^n K_{h_2}^{(2)}(\beta^T \phi(x_j), \beta^T \phi(x_i))} \\ \vdots \\ \frac{K_{h_2}^{(2)}(\beta^T \phi(x_n), \beta^T \phi(x_i))}{\sum_{j=1}^n K_{h_2}^{(2)}(\beta^T \phi(x_j), \beta^T \phi(x_i))} \end{pmatrix}.$$

将表达式代入估计方程得

$$\sum_{i=1}^n [\hat{\xi}(y_i)(\phi(x_i) - \hat{\theta}(\beta^T \phi(x_i)))^T] =$$

$$(\hat{\xi}(y_1) \cdots \hat{\xi}(y_n)) \begin{pmatrix} (\phi(x_1) - \hat{\theta}(\beta^T \phi(x_1)))^T \\ \vdots \\ (\phi(x_1) - \hat{\theta}(\beta^T \phi(x_n)))^T \end{pmatrix} =$$

$$\Phi \tilde{K}_{h_1}^{(1)} [\Phi^T - (\tilde{K}_{h_2}^{(2)})^T \Phi^T] =$$

$$\Phi \tilde{K}_{h_1}^{(1)} (I_n - (\tilde{K}_{h_2}^{(2)})^T) \Phi^T.$$

同理可得

$$\sum_{i=1}^n [\hat{\xi}(\beta^T \phi(x_i))(\phi(x_i) - \hat{\theta}(\beta^T \phi(x_i)))^T] =$$

$$\Phi \tilde{K}_{h_1}^{(1)} \tilde{K}_{h_2}^{(2)} (I_n - (\tilde{K}_{h_2}^{(2)})^T) \Phi^T,$$

故有

$$\frac{1}{n} \sum_{i=1}^n [(\hat{\xi}(y_i) - \hat{\xi}(\beta^T \phi(x_i)))(\phi(x_i) - \hat{\theta}(\beta^T \phi(x_i)))^T] =$$

$$\Phi \tilde{K}_{h_1}^{(1)} (I_n - \tilde{K}_{h_2}^{(2)}) (I_n - (\tilde{K}_{h_2}^{(2)})^T) \Phi^T = 0.$$

由上面的结果可以得出方程(6).

A.2 方程(7)的推导

对 $\forall j=1,2,\dots,q$, 设 $\beta_j = \beta_j^{(0)} + \beta_j^{(1)}$, $\beta_j^{(0)} \in \mathcal{H}_0 = \text{span}\{\phi(x_1), \dots, \phi(x_n)\}$, $\beta_j^{(1)} \in \mathcal{H}_0^\perp$. 其中 $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_0^\perp$, \mathcal{H}_0^\perp 是 \mathcal{H}_0 的正交补空间. 存在 $c_j \in \mathbb{R}^R$ 使得

$$\beta_j^{(0)} = \sum_{i=1}^n c_{ij} \phi(x_i) = \Phi c_j.$$

因此有

$$\langle \beta_j, \phi(x_i) \rangle_{\mathcal{H}} = \langle \beta_j^{(0)} + \beta_j^{(1)}, \phi(x_i) \rangle_{\mathcal{H}} =$$

$$\langle \beta_j^{(0)}, \phi(x_i) \rangle_{\mathcal{H}} =$$

$$\sum_{i=1}^n c_{ij} \langle \phi(x_i), \phi(x_i) \rangle_{\mathcal{H}} =$$

$$\sum_{i=1}^n c_{ij} R(x_i, x_i)_{\mathcal{H}} = R(x_{\leq n}, x_i) c_j,$$

其中, $R(x_{\leq n}, x_i) = \begin{pmatrix} R(x_1, x_i) \\ \vdots \\ R(x_n, x_i) \end{pmatrix}$ 为对应 $\{x_i\}_1^n$ 的

Gram 矩阵 \mathbf{R} 的第 i 列, 记为 R_i . 由此, 可得

$$\beta^T \Phi(x_i) = \begin{pmatrix} \langle \beta_1, \phi(x_i) \rangle_{\mathcal{H}} \\ \vdots \\ \langle \beta_q, \phi(x_i) \rangle_{\mathcal{H}} \end{pmatrix} = \begin{pmatrix} c_1^T \\ \vdots \\ c_q^T \end{pmatrix} R_i = C^T R_i,$$

其中, $C = (c_1, \dots, c_q)$.

A.3 定理 1 的推导

分别记 Λ_1 和 Λ_2 为对应的 η_1 和 η_2 的冗余切平面, 不难得

$$\left. \begin{aligned} \Lambda_1 &= \{f(X) : E(f) = 0, \forall f\}, \\ \Lambda_2 &= \{f(Y, U(X)) : \\ &E(f | X) = E(f | U(X)) = 0, \forall f\} \end{aligned} \right\} \quad (A1)$$

易知 $\Lambda_1 \perp \Lambda_2$, 因此(4)的冗余切平面 $\Lambda = \Lambda_1 \oplus \Lambda_2$. 对 $\forall f \in \Lambda_1$, $g \in \{f(Y, X) : E(f | X) = 0\}$, 则

$$E[f^T(X)g(Y, X)] = E\{E(f^T(X)g(Y, X) | X)\} =$$

$$E\{f^T(X)E[g(Y, X) | X]\} = 0,$$

所以

$$\{f(Y, X): E(f | X) = 0\} \in \Lambda_1^\perp \quad (A2)$$

反之, 对于 $\forall f(X) \in \Lambda_1$, 要使

$$Ef^T(X)E[g(Y, X) | X] = 0$$

必须 $E(g(Y, X) | X)$ 为常值函数, 有 $E[g(Y, X) | X] = 0$, 也即

$$\Lambda_1^\perp \subset \{f(Y, X): E(f | X) = 0\} \quad (A3)$$

由(A1)和(A2), 得

$$\Lambda_1^\perp = \{f(Y, X): E(f | X) = 0\}.$$

下面我们证明

$$\Lambda_2^\perp = \{f(Y, X): E(f | U(X), Y) \text{ 仅依赖于 } U(x)\}.$$

$$\forall f \in \Lambda_2,$$

$$g \in \mathcal{G} \triangleq \{f(Y, X): E[f(Y, X) | U(x), Y] \text{ 仅为 } U(x) \text{ 的函数}\},$$

则

$$\begin{aligned} E\{f^T(Y, U(X))g(Y, X)\} &= \\ E\{E[f^T(Y, U(X))g(Y, X) | U(X), Y]\} &= \\ E\{f^T(Y, U(X))E[g(Y, X) | U(X), Y]\} &= \\ E\{f^T(Y, U(X))h(U(X))\} &= \\ E\{E[f^T(Y, U(X))h(U(X)) | U(X)]\} &= \\ E\{h^T(U(X))E[f(Y, U(X)) | U(X)]\}, \end{aligned}$$

其中, $h(U(X)) = E[g(Y, X) | U(X), Y]$. 由式(11)知上式为 0, 故

$$\mathcal{G} \subset \Lambda_2^\perp.$$

另一方面, $\forall f(Y, X) \in \Lambda_2^\perp$, 记

$$g = E(f | U(X), Y) - E(f | U(X)) \quad (A4)$$

显然 $g \in \Lambda_2$, 因此

$$E(g^T f) = 0 \quad (A5)$$

又根据(A5), 有

$$\begin{aligned} E(g^T f) &= E\{g^T E[f | U(X), Y]\} = \\ Eg^T g + E\{g^T E[f | U(X)]\} &= \\ Eg^T g + E\{E[g^T E(f | U(X)) | U(X)]\} &= \\ Eg^T g \end{aligned} \quad (A6)$$

由(A5)和(A6)可得 $E(f | U(X), Y) = E(f | U(X))$, 即

$$\Lambda_2^\perp \subset \mathcal{G} \quad (A7)$$

故由(A3)和(A4)可得

$$\Lambda_2^\perp = \mathcal{G} = \{f(Y, X):$$

$$E(f(Y, X) | U(X), Y) \text{ 仅为 } U(X) \text{ 的函数}\}.$$

最后证明:

$$\Lambda^\perp = \Lambda_1^\perp \cap \Lambda_2^\perp =$$

$$\{f(Y, X) - E(f | U(X), Y): E(f | X) =$$

$$E(f | U(X)), \forall f\}.$$

记 $\mathcal{F} = \{f(Y, X) - E(f | U(X), Y): E(f | X) = E(f | U(X)), \forall f\}$, 显然 $\forall g \in \mathcal{F}$, 均有

$$E(g | U(x), Y) = 0,$$

那么 $\mathcal{F} \subset \Lambda_2^\perp$ 又因为

$$\begin{aligned} E\{f(Y, X) - E(f | U(X), Y) | X\} &= \\ E\{f(Y, X) | X\} - E\{E[f | U(X), Y] | X\} &= \\ E(f | X) - E(f | U(X)) &= 0, \end{aligned}$$

因此 $\mathcal{F} \subset \Lambda_1^\perp$, 即

$$\mathcal{F} \subset \Lambda^\perp \quad (A8)$$

对 $\forall f \in \Lambda^\perp$, 那么 $f \in \Lambda_1^\perp$, 则存在函数 a , 使得:

$$E(f | U(X), Y) = a(U(X)).$$

进而

$$\begin{aligned} a(U(X)) &= E(f | U(X) = U(x), Y) = \\ \frac{\int_{U(X)=U(x)} f(Y, X) \eta_1(X) \eta_2(Y | U(X)) d\mu(X)}{\int_{U(X)=U(x)} \eta_1(X) \eta_2(Y | U(X)) d\mu(X)} &= \\ \frac{\int_{U(X)=U(x)} f(Y, X) \eta_1(X) d\mu(X)}{\int_{U(X)=U(x)} \eta_1(X) d\mu(X)}. \end{aligned}$$

因此,

$$\begin{aligned} a(U(X)) &= \int a(U(X)) \eta_2(Y | U(X)) d\mu(Y) \\ \int \frac{\int_{U(X)=U(x)} f(Y, X) \eta_1(X) d\mu(X)}{\int_{U(X)=U(x)} \eta_1(X) d\mu(X)} \cdot \\ \eta_2(Y | U(X)) d\mu(Y) &= \\ \frac{\iint_{U(X)=U(x)} f(Y, X) \eta_1(X) \eta_2(Y | U(X)) d\mu(X) d\mu(Y)}{\int_{U(X)=U(x)} \eta_1(X) d\mu(X)} &= \\ \frac{\int_{U(X)=U(x)} \left[\int f(Y, X) \eta_2(Y | U(X)) d\mu(Y) \right] \eta_1(X) d\mu(X)}{\int_{U(X)=U(x)} \eta_1(X) d\mu(X)} \end{aligned} \quad (A9)$$

由 $f \in \Lambda_1^\perp$, 则

$$\begin{aligned} \int f(Y, X) \eta_2(Y | U(X)) d\mu(Y) &= \\ E(f(Y | X) | U(X)) &= 0, \end{aligned}$$

所以(A9)等于 0, 即 $a(U(x)) = 0$. 因此 Λ^\perp 中的元素具有 $f(Y, X) - E(f | U(X), Y)$ 的形式, 故

$$\Lambda^\perp \subset \mathcal{F} \quad (A10)$$

由(A8)和(A10), 定理 1 成立.

A.4 定理 2 的推导

(\Rightarrow) 由于 $\Phi^T \Phi = R$, 所以显然成立.

(\Leftarrow) 对 Φ 进行如下的谱分解

$$\Phi = UDV^T =$$

$$(u_1, \dots, u_q) \begin{pmatrix} \bar{D}_{q \times q} & O_{q \times (p-q)} \\ O_{(n-q) \times q} & O_{(n-1) \times (n-q)} \end{pmatrix} \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \bar{U} \bar{D} \bar{V}^T,$$

其中, $\bar{U} = (u_1, \dots, u_q)$, $\bar{V} = (v_1, \dots, v_q)$, 则

$$R = \Phi^T \Phi = (\bar{V} \bar{D} \bar{U}^T) (\bar{U} \bar{D} \bar{V}^T) = \bar{V} \bar{D}^2 \bar{V}^T.$$

因此方程(7)可以化为

$$\bar{V} \bar{D}^2 \bar{V}^T \tilde{\mathbf{K}}_{h_1}^{(1)} (\mathbf{I}_n - \tilde{\mathbf{K}}_{h_2}^{(2)}) (\mathbf{I}_n - (\tilde{\mathbf{K}}_{h_2}^{(2)})^T) \bar{V} \bar{D}^2 \bar{V}^T = 0.$$

由 $(\bar{D}^{-2} \bar{V}^T) \bar{V} \bar{D}^2 \bar{V}^T = \bar{V}^T$, $(\bar{V} \bar{D}^2 \bar{V}^T) \bar{V} \bar{D}^{-2} = \bar{V}$, 方程(7)可化为

$$\bar{V}^T \tilde{\mathbf{K}}_{h_1}^{(1)} (\mathbf{I}_n - \tilde{\mathbf{K}}_{h_2}^{(2)}) (\mathbf{I}_n - (\tilde{\mathbf{K}}_{h_2}^{(2)})^T) \bar{V} = 0,$$

因此

$$\bar{U} \bar{D} \bar{V}^T \tilde{\mathbf{K}}_{h_1}^{(1)} (\mathbf{I}_n - \tilde{\mathbf{K}}_{h_2}^{(2)}) (\mathbf{I}_n - (\tilde{\mathbf{K}}_{h_2}^{(2)})^T) \bar{V} \bar{D} \bar{U}^T =$$

$$\Phi \tilde{\mathbf{K}}_{h_1}^{(1)} (\mathbf{I}_n - \tilde{\mathbf{K}}_{h_2}^{(2)}) (\mathbf{I}_n - (\tilde{\mathbf{K}}_{h_2}^{(2)})^T) \Phi^T.$$

故得出(9)成立, 因此方程(7)与方程(9)等价.

参考文献(References)

- [1] ARONSZAJN N. Theory of reproducing kernels[J]. Transactions of the American Mathematical Society, 1950, 68(3): 337-404.
- [2] BICKEL P J, KLAASSEN C A J, RITOV Y, et al. Efficient and Adaptive Estimation for Semiparametric Models [M]. Baltimore, MD: Johns Hopkins University Press, 1993.
- [3] COOK R D. Regression Graphics: Ideas for Studying Regressions through Graphics[M]. New York: Wiley, 1998.
- [4] COOK R D, LI B. Dimension reduction for conditional mean in regression[J]. The Annals of Statistics, 2002, 30(2): 455-474.
- [5] CUI Wenquan, WU Chenglong. An approach to estimating nonlinear sufficient dimension reduction subspace for censored survival data[J]. Journal of University of Science and Technology of China, 2015, 45(9): 709-716.
- [6] FERRÉ L, VILLA N. Multilayer perceptron with functional inputs: An inverse regression approach[J]. Scandinavian Journal of Statistics, 2006, 33(4): 807-823.

- [7] FUKUMIZU K, BACH F R, JORDAN M I. Kernel dimension reduction in regression[J]. The Annals of Statistics, 2009, 37(4): 1 871-1 905.
- [8] LI L, YIN X. Sliced inverse regression with regularizations[J]. Biometrics, 2008, 64(1): 124-131.
- [9] LI K C. Sliced inverse regression for dimension reduction [J]. Journal of the American Statistical Association, 1991, 86(414): 316-327.
- [10] MA Y, ZHU L. A review on dimension reduction[J]. International Statistical Review, 2013, 81(1): 134-150.
- [11] MA Y, ZHU L. A semiparametric approach to dimension reduction [J]. Journal of the American Statistical Association, 2012, 107(497): 168-179.
- [12] SCHÖLKOPF B, SMOLA A J. Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond[M]. Cambridge, MA: MIT Press, 2001.
- [13] TSIATIS A A. Semiparametric Theory and Missing Data[M]. New York: Springer, 2006.
- [14] WU Q, LIANG F, MUKHERJEE S. Consistency of regularized sliced inverse regression for kernel models [R]. Durham, NC: Duke University; IL: University of Illinois Urbana-Champaign, 2008.
- [15] WU H M. Kernel sliced inverse regression with applications to classification [J]. Journal of Computational and Graphical Statistics, 2008, 17(3): 590-610.
- [16] ZHONG W, ZENG P, MA P, et al. RSIR: regularized sliced inverse regression for motif discovery [J]. Bioinformatics, 2005, 21(22): 4 169-4 175.
- [17] COOK R D, WEISBERG S. Discussion of "sliced inverse regression for dimension reduction"[J]. Journal of the American Statistical Association, 1991, 86: 28-33.
- [18] LI B, WANG S. On directional regression for dimension reduction [J]. Journal of the American Statistical Association, 2007, 102(479): 997-1 008.
- [19] Wu H M. Kernel sliced inverse regression with applications to classification [J]. Journal of Computational and Graphical Statistics, 2008, 17(3): 590-610.
- [20] KENJI F, FRANCIS R B, MICHAEL I J. Kernel dimension reduction in regression[J]. The Annals of Statistics, 2009, 37(4): 1 871-1 905.
- [21] WU Q, LIANG F, MUKHERJEE S. Kernel sliced inverse regression: regularization and consistency[J]. Abstract and Applied Analysis, 2013: 540725.