Statistica Sinica **25** (2015), 743-758
doi:http://dx.doi.org/10.5705/ss.2013.113

DIRECTION ESTIMATION IN SINGLE-INDEX REGRESSIONS VIA HILBERT-SCHMIDT INDEPENDENCE CRITERION

Nan Zhang and Xiangrong Yin

University of Georgia and University of Kentucky

Abstract: In this article, we use a Hilbert-Schmidt Independence Criterion to propose a new method for estimating directions in single-index models. This approach enjoys a model free property and requires no link function to be smoothed or estimated. Further, we propose a permutation test to check whether the estimated single-index is sufficient. The sampling distribution of our estimator is established. Finite sample performance of proposed estimates is examined through simulation studies and compared with two well-established methods: the refined Minimum Average Variance Estimation method (rMAVE, Xia et al. (2002)) and the Estimating Function Method (EFM, Cui, Härdle, and Zhu (2011)). A New Zealand Horse Mussels data set is analyzed via our approach to demonstrate the efficacy of our proposed approach.

Key words and phrases: Central subspace, Hilbert-Schmidt independence criterion, permutation test, single-index models, sufficient dimension reduction.

1. Introduction

Dimension reduction plays an important role in high-dimensional statistical modeling. The general goal is to infer the conditional distribution of the response given the reduced predictors without loss of regression information. Single-index models, as a special case of sufficient dimension reduction with one dimension, refer to regression problems where the regression information can be fully described by a single linear combination of the predictors. Single-index models have been widely applied in such disciplines as biostatistics, economics, and medicine.

There is a huge literature on estimating single-index models via a nonparametric approach, for instance Härdle, Hall, and Ichimura (1993), Ichimura (1993), and Hristache et al. (2001). We consider estimation from the standing of sufficient dimension reduction. Many sufficient dimension reduction methods have proven useful in coefficient estimation for single-index models, such as the refined minimum average function method (rMAVE, Xia et al. (2002)) and the expected likelihood based method that minimizes a Kullback-Leiblier distance by Yin and Cook (2005). As well, there are methods specially designed to estimate the single-index coefficients, for example, the average derivative method (ADE, Härdle and Stoker (1989)) and the estimating function method (EFM, Cui, Härdle, and Zhu (2011)). These methods often involve kernel smoothing techniques and require such assumptions as smooth link functions and at least one continuous predictor.

In this article, we study the independence between the response and the predictors via a Hilbert-Schmidt Independence Criterion (HSIC) and develop a new method for estimating directions in single-index models. The article is organized as follows. Section 2 describes our method, including motivation, theoretical results, estimation algorithm, and testing procedure. Section 3 contains simulation studies as well as a data analysis, followed by a short discussion in Section 4. Proofs are provided in the Appendix, while longer derivations are arranged in a supplementary file.

2. Methodology

We study a general statistic, based on HSIC that measures the independence between random variables, for estimating the coefficients in single-index models.

2.1. Definitions

Let $\mathbf{X}$ be a $p \times 1$ vector, and Y be a univariate response. The ultimate goal of sufficient dimension reduction is to search a number of linear combinations of $\mathbf{X}$, say $\beta^T\mathbf{X}$, where β is a $p \times d$ matrix, $d \leq p$, such that Y depends on $\mathbf{X}$ only through $\beta^T\mathbf{X}$, $Y \perp\!\!\!\perp \mathbf{X} | \beta^T\mathbf{X}$, where $\perp\!\!\!\perp$ means independence. The column space of β forms a dimension reduction subspace (Li (1991); Cook (1996)). The central subspace, $\mathcal{S}_{Y|\mathbf{X}}$, is defined as the intersection of all dimension reduction subspaces when itself is a dimension reduction subspace (Cook (1996)). The dimension of $\mathcal{S}_{Y|\mathbf{X}}$, denoted by $\dim(\mathcal{S}_{Y|\mathbf{X}}) = d$, is called the structural dimension. The existence and uniqueness of the central subspace has been shown by Cook (1996) and Yin, Li, and Cook (2008) under mild conditions. We assume the central subspace exists, and consider the special case of $d = 1$, the single-index models. We assume that η is a $p \times 1$ basis vector, spanning the central subspace, with β a generic $p \times 1$ vector.

Most traditional approaches estimate the single-index through a pre-specified model of $Y|\beta^T\mathbf{X}$, which makes the single-index $\beta^T\mathbf{X}$ best related to Y. Intuitively, a measure of independence could help to identify such an index. Indeed, HSIC is one such. It is used to measure the independence between random variables X and Y without pre-specifying any models (Sejdinovic et al. (2012)). As with correlation, low magnitudes in HSIC suggest weak relations. Thus, the linear combination $\beta^T\mathbf{X}$ that maximizes HSIC is the single-index that most relates to Y. For univariate X and Y, HSIC characterizes the distance between the joint distribution $\mathbf{P} := \mathbf{P}_{X,Y}$ and the product of the marginal $\mathbf{Q} := \mathbf{P}_X\mathbf{P}_Y$ as

$$H(X,Y) = \int |f_P(t,s) - f_Q(t,s)|^2 W(t,s)dtds, \tag{2.1}$$

where f_P and f_Q are the characteristic functions of $\mathbf{P}$ and $\mathbf{Q}$, respectively. Applying Bochner's theorem, Gretton et al. (2008), Gretton, Fukumizu, and Spiperumbudur (2009) showed that $H(X,Y)$ can be rewritten as

$$\begin{aligned} H(X,Y) = & \, \mathrm{E}[K(X-X')L(Y-Y')] + \mathrm{E}[K(X-X')]\mathrm{E}[L(Y-Y')] \\ & -2\mathrm{E}\{\mathrm{E}[K(X-X')|X]\mathrm{E}[L(Y-Y')|Y]\}, \end{aligned} \tag{2.2}$$

where X' and Y' denote independent copies of X and Y, and $K(\cdot)$ and $L(\cdot)$ are positive definite kernel functions. See Gretton et al. (2008), Gretton, Fukumizu, and Spiperumbudur (2009) for the restrictions on the choices of $W(t,s)$, and hence $K(\cdot)$ and $L(\cdot)$ so that the equivalence of (2.1) and (2.2) holds. We refer to Kankainen (1995) for some specific conditions and choices for $W(t,s)$.

Definition 1. The HSIC covariance between random variables $\beta^T\mathbf{X}$ and Y is the nonnegative number of $\sqrt{H}$,

$$\begin{aligned} H(\beta^T\mathbf{X},Y) = & \, \mathrm{E}[K(\beta^T(\mathbf{X}-\mathbf{X}'))L(Y-Y')] + \mathrm{E}[K(\beta^T(\mathbf{X}-\mathbf{X}'))]\mathrm{E}[L(Y-Y')] \\ & -2\mathrm{E}\{\mathrm{E}[K(\beta^T(\mathbf{X}-\mathbf{X}'))|\beta^T\mathbf{X}]\mathrm{E}[L(Y-Y')|Y]\}. \end{aligned} \tag{2.3}$$

Following the arguments in Gretton et al. (2008), Gretton, Fukumizu, and Spiperumbudur (2009) and Kankainen (1995), for certain $W(t,s)$, $H(\beta^T\mathbf{X},Y)$ characterizes the distance between $\mathbf{P} := \mathbf{P}_{\beta^T\mathbf{X}Y}$ and $\mathbf{Q} := \mathbf{P}_{\beta^T\mathbf{X}}\mathbf{P}_Y$, and can be written as $H(\beta^T\mathbf{X},Y) = \int |f_P(t,s) - f_Q(t,s)|^2 W(t,s)dtds$. Clearly, $H(\beta^T\mathbf{X},Y) \geq 0$, and $H(\beta^T\mathbf{X},Y) = 0$ if and only if $\beta^T\mathbf{X}$ and Y are independent.

The HSIC covariance is a generalization of covariance in the sense that $H(\beta^T\mathbf{X},Y) = 0$ characterizes the independence of $\beta^T\mathbf{X}$ and Y. There are many choices for weights $W(\cdot)$, variously the kernels K and L in (2.1). We adopt the kernel choice from Kankainen (1995): for univariate variables X and Y,

$$K := \exp\left(\frac{-\|X-X'\|^2}{2\sigma_X^2}\right) \quad \text{and} \quad L := \exp\left(\frac{-\|Y-Y'\|^2}{2\sigma_Y^2}\right).$$

One can develop a variety of properties for H, but our purpose is to use this measure for estimating the coefficients of the single-index. There is a recent method in Sheng and Yin (2013) that makes similar use of equation (2.1), but with a different weight, as in Székely, Rizzo, and Bakirov (2007) and Székely and Rizzo (2009). Their goal is the same as ours, but leads to a different theory, and different asymptotic properties. Another approach developed in Fukumizu, Bach, and Jordan (2004, 2009) using reproducing kernel Hilbert spaces for kernel dependence measures via covariance operators. This is related to our proposed method.

2.2. Property

We show that $H(\beta^T\mathbf{X}, Y)$ can be used for estimating the coefficients of the single-index, and assume $(\mathbf{X}, Y)$ has finite first two moments.

Let Σ_X be the covariance matrix of $\mathbf{X}$, assumed nonsingular. Let (η, α) form a $p \times p$ matrix such that $(\eta, \alpha)^T\Sigma_X(\eta, \alpha) = \mathbf{I}$.

Proposition 1. *Assume the support of $\mathbf{X} \in \mathbb{R}^p$, S, is a compact set, and that η spans the central subspace, $\eta^T\Sigma_X\eta = 1$. If $\eta^T\mathbf{X} \perp\!\!\!\perp \alpha^T\mathbf{X}$, then $\eta = \arg\max_{\beta^T\Sigma_X\beta=1} H(\beta^T\mathbf{X}, Y)$.*

In general, the support of $\mathbf{X}$ is not compact. But Yin, Li, and Cook (2008, Proposition 11) showed that, as long as a compact set S is large enough, then $S_{Y|\mathbf{X}_s} = S_{Y|\mathbf{X}}$, where $\mathbf{X}_s$ is $\mathbf{X}$ restricted onto S. We can restrict to a compact set S for simplicity. Proposition 1 states that maximizing $H(\beta^T\mathbf{X}, Y)$ over β under $\beta^T\Sigma_X\beta = 1$ recovers the single-index.

When $\mathbf{X}$ is normal, the condition $\eta^T\mathbf{X} \perp\!\!\!\perp \alpha^T\mathbf{X}$ is satisfied, but normality is not necessary. For example, if $X = (X_1, X_2, \ldots, X_p)$, with $X_1 \perp\!\!\!\perp (X_2, \ldots, X_p)$ and $\eta = (1, 0, \ldots, 0)$, then $\eta^T\mathbf{X} \perp\!\!\!\perp \alpha^T\mathbf{X}$. In general, a distribution with the "linearity condition" does not necessarily satisfy such condition. Hall and Li (1993) showed that when p is large, the independence condition holds asymptotically. We take the condition as not very restrictive; this is supported in our simulations later. Since $H(\eta^T\mathbf{X}, Y) = H(-\eta^T\mathbf{X}, Y)$, η and $-\eta$ are both solutions that span the same space, we select η to have its first nonzero element positive. The proof of Proposition 1 is in the Appendix.

2.3. Estimation

Following Kankainen (1995), the sample version of HSIC for univariate variables (X, Y) with (X_i, Y_i) being an i.i.d. sample, $i = 1, \cdots, n$, is

$$H_n = \frac{1}{n^2}\sum_{i,j} K_{ij}L_{ij} - \frac{2}{n^3}\sum_{i,j,k} K_{ij}L_{ik} + \frac{1}{n^4}\sum_{i,j,k,l} K_{ij}L_{kl}, \tag{2.4}$$

where

$$K_{ij} := \exp\left(\frac{-\|X_i - X_j\|^2}{2\sigma_X^2}\right) \quad \text{and} \quad L_{kl} := \exp\left(\frac{-\|Y_k - Y_l\|^2}{2\sigma_Y^2}\right).$$

Consider a random sample $(\mathbf{X}, Y) = \{(\mathbf{X}_i, Y_i) : i = 1, \ldots, n\}$ of n i.i.d. random vectors $(\mathbf{X}, Y)$. If $\hat{\Sigma}_X$ and $\hat{\sigma}_Y$ denote the sample covariance matrix and sample variance of $\mathbf{X}$ and Y, respectively, the sample HSIC can be defined as follows.

Definition 2. The sample HSIC covariance statistic H_n is

$$H_n(\beta^T\mathbf{X}, Y) = \frac{1}{n^2}\sum_{i,j} K_{ij}L_{ij} - \frac{2}{n^3}\sum_{i,j,k} K_{ij}L_{ik} + \frac{1}{n^4}\sum_{i,j,k,l} K_{ij}L_{kl},$$

where

$$K_{ij} := \exp\left(\frac{-(\beta^T(\mathbf{X}_i - \mathbf{X}_j))^2}{2\beta^T\hat{\Sigma}_X\beta}\right) \quad \text{and} \quad L_{kl} := \exp\left(\frac{-\|Y_k - Y_l\|^2}{2\hat{\sigma}_Y^2}\right).$$

2.4. Asymptotic properties

Proposition 2. *Under the assumptions in Proposition* 1*, if* $\eta_n = \arg\max_{\beta^T\hat{\Sigma}_X\beta=1} H_n(\beta^T\mathbf{X}, Y)$*, then* η_n *converges in probability to* η *as* $n \to \infty$*.*

Here we consider only η_n with first nonzero element positive. Due to the property of the chosen kernel (scale-free), no constraint on the support of $\mathbf{X}$ is needed.

Proposition 3. *Under the assumptions in Proposition* 1 *and in the supplement file, if* $\eta_n = \arg\max_{\beta^T\hat{\Sigma}_X\beta=1} H_n(\beta^T\mathbf{X}, Y)$*, then* $\sqrt{n}(\eta_n - \eta) \to N(0, V_{11})$*, where* V_{11} *is a covariance matrix defined in the Appendix.*

An explicit formula for V_{11} is to be derived, and used for calculating confidence intervals for the coefficients in the single-index later. Proofs of Propositions 2 and 3 are in the Appendix.

2.5. Algorithm

Our goal is to find the estimator η_n of η,

$$\eta_n = \arg\max_{\beta^T\hat{\Sigma}_X\beta=1} H_n(\beta^T\mathbf{X}, Y). \tag{2.5}$$

We propose a global search algorithm. There are multiple ways to obtain initial estimates, for example, one could adopt such well-known dimension reduction methods as *SIR* in (Li (1991)), *pHd* (Li (1992)), or *OPG* (Xia et al. (2002)). But we propose a choice of initial estimates based on HSIC.

2.5.1. Initial choice: approximation of Hessian matrix SVD

We propose the initial estimate based on the Hessian matrix of $H_n(\beta^T\mathbf{X}, Y)$ at (2.4). We first standardize $\mathbf{X}$, taking $\mathbf{Z} = \Sigma_X^{-1/2}(\mathbf{X} - \mathrm{E}(\mathbf{X}))$, such that $\Sigma_{\mathbf{Z}} = \mathbf{I}$, the $p \times p$ identity matrix. Then the direction of single-index in $\mathbf{X}$-scale is $\eta =$

$\Sigma_X^{-1/2}\eta_Z$ where η_Z is the single-index direction in $\mathbf{Z}$-scale. The sample version of $\mathbf{Z}$ is $\hat{\mathbf{Z}}_i = \hat{\Sigma}_X^{-1/2}(\mathbf{X}_i - \bar{\mathbf{X}}))$, where $\hat{\Sigma}_X$ and $\bar{\mathbf{X}}$ are the sample covariance matrix and mean of $\mathbf{X}_i$ for $i = 1, \cdots, n$, respectively.

Step 1 : Construct the $p \times p$ matrix $Hess_n(\hat{\mathbf{Z}}, Y)$, where

$$\begin{aligned} &Hess_n(\hat{\mathbf{Z}}, Y) \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} e^{-\frac{1}{2}(\hat{\mathbf{Z}}_i-\hat{\mathbf{Z}}_j)^T(\hat{\mathbf{Z}}_i-\hat{\mathbf{Z}}_j)} \cdot ((\hat{\mathbf{Z}}_i - \hat{\mathbf{Z}}_j)(\hat{\mathbf{Z}}_i - \hat{\mathbf{Z}}_j)^T - \mathbf{I}) \cdot e^{-\frac{\|Y_i-Y_j\|^2}{2\hat{\sigma}_Y^2}} \\ &\quad -\frac{2}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} e^{-\frac{1}{2}(\hat{\mathbf{Z}}_i-\hat{\mathbf{Z}}_j)^T(\hat{\mathbf{Z}}_i-\hat{\mathbf{Z}}_j)} \cdot ((\hat{\mathbf{Z}}_i - \hat{\mathbf{Z}}_j)(\hat{\mathbf{Z}}_i - \hat{\mathbf{Z}}_j)^T - \mathbf{I}) \cdot e^{-\frac{\|Y_i-Y_k\|^2}{2\hat{\sigma}_Y^2}} \\ &\quad +\frac{1}{n^4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} e^{-\frac{1}{2}(\hat{\mathbf{Z}}_i-\hat{\mathbf{Z}}_j)^T(\hat{\mathbf{Z}}_i-\hat{\mathbf{Z}}_j)} \cdot ((\hat{\mathbf{Z}}_i-\hat{\mathbf{Z}}_j)(\hat{\mathbf{Z}}_i-\hat{\mathbf{Z}}_j)^T-\mathbf{I}) \cdot e^{-\frac{\|Y_k-Y_l\|^2}{2\hat{\sigma}_Y^2}}; \end{aligned}$$

Step 2 : Take the Singular Value Decomposition (SVD), $Hess_n^* = U\Sigma V$;

Step 3 : Take the first column of U matrix as the initial estimate of $\beta_{Z,0}$.

Here $Hess_n(\hat{\mathbf{Z}}, Y)$ is a naive one-step approximation to the Hessian matrix of $H_n(\beta^T\hat{\mathbf{Z}}, Y)$, say, $Hess_n(\beta^T\hat{\mathbf{Z}}, Y)$, with $\mathbf{I}$ replacing $\beta\beta^T$ in all related terms of $Hess_n(\beta^T\hat{\mathbf{Z}}, Y)$, since β is unknown.

2.5.2. Global optimization

We propose a global optimization algorithm based on (2.5), but in $\mathbf{Z}$-scale. Here the constraint is $\beta_Z^T\beta_Z = 1$. We maximize $\beta_{Z,n} = \arg\max_{\beta_Z^T\beta_Z=1} H_n(\beta_Z^T\mathbf{Z}, Y)$, then transform the solution back to $\mathbf{X}$-scale.

Step 1 : Select the initial estimate $\beta_{Z,0}$ as in Section 2.5.1.

Step 2 : Construct $n \times n$ kernel matrices K and L, with entries

$$K_{ij} = \exp\left(\frac{-(\beta_Z^T(\hat{\mathbf{Z}}_i - \hat{\mathbf{Z}}_j))^2}{2\beta_Z^T\beta_Z}\right) \text{ and } L_{kl} = \exp\left(-\frac{(Y_k - Y_l)^2}{2\hat{\sigma}_Y^2}\right).$$

Step 3 : Obtain an estimate $\beta_{Z,n} = \arg\max_{\beta_Z^T\beta_Z=1} H_n(\beta_Z^T\mathbf{Z}, Y)$ based on *optim*, available in *R*.

Step 4 : Return to Step 2 and iteratively optimize the target function to converge to obtain $\beta_{Z,n}$. The estimate is $\eta_n = \hat{\Sigma}_X^{-1/2}\beta_{Z,n}$.

Maximization is carried out iteratively through the general purpose optimization function *optim*, available in *R*, that implements the Broyden-Fletcher-

Goldfarb-Shanno (BFGS) method (Bonnans et al. (2006)). Our codes in R are available upon request.

2.6. Testing $d = 1$

Utilizing HSIC, we consider a permutation test to verify the single-index assumption, $H_0 : Y \perp\!\!\!\perp \mathbf{Z} | \eta_Z^T \mathbf{Z}$. For testing the HSIC distance, d, is zero, or $Y \perp\!\!\!\perp \mathbf{Z}$ (which implies that $Y \perp\!\!\!\perp \beta^T \mathbf{Z}$ for any β), we propose the following procedure.

- Find the estimate $\beta_{Z,n}$ as in Section 2.5.2;
- Permutate Y a total of B times and calculate the indices $H_{n,i}(Y^{(i)}, \beta_{Z,n}^T \hat{\mathbf{Z}}), i = 1, \ldots, B$;
- If $H_n(Y, \beta_{Z,n}^T \hat{\mathbf{Z}})$ is greater than the 95% quantiles of $H_{n,i}(Y^{(i)}, \beta_{Z,n}^T \hat{\mathbf{Z}}), i = 1, \ldots, B$, infer $d \geq 1$; otherwise $d = 0$.

To test $d = 1$, we use a result of Cook (1998, Proposition 4.6):

$$(Y, \beta^T \mathbf{Z}) \perp\!\!\!\perp \beta^{\perp,T} \mathbf{Z} \Rightarrow Y \perp\!\!\!\perp \beta^{\perp,T} \mathbf{Z} | \beta^T \mathbf{Z}.$$

We test the left-hand side here to infer an upper-bound for the goal of testing the right-hand side. More details about this permutation test can be found in Cook and Yin (2001) and Yin and Cook (2002). We propose the following procedure.

- Form the orthogonal matrix $(\beta_{Z,n}, \beta_{Z,n}^{\perp})$;
- Calculate the index $H_n((Y, \beta_{Z,n}^T \hat{\mathbf{Z}}), \beta_{Z,n}^{\perp,T} \hat{\mathbf{Z}})$;
- Permutate and calculate indices $H_{n,i}((Y, \beta_{Z,n}^T \hat{\mathbf{Z}})^{(i)}, \beta_{Z,n}^{\perp,T} \hat{\mathbf{Z}})$, for $i = 1, \ldots, B$;
- If $H_n((Y, \beta_{Z,n}^T \hat{\mathbf{Z}}), \beta_{Z,n}^{\perp,T} \hat{\mathbf{Z}})$ is greater than the 95% quantiles of $H_{n,i}((Y, \beta_{Z,n}^T \hat{\mathbf{Z}})^{(i)}, \beta_{Z,n}^{\perp,T} \hat{\mathbf{Z}})$, $i = 1, \ldots, B$, infer $d \geq 2$; otherwise $d = 1$.

Our simulation studies (Section 3.2) suggest that this algorithm (a modified version, incorporating sample size) is efficient in computation and reliable in determining the true dimensionality of the central subspace.

Several papers have discussed the asymptotic distributions of the empirical estimates related to H_n, specifically for the independence test of $\mathbf{X}$ and Y. For instance, Theorem 2 in Gretton et al. (2008) states a result relating to the solutions of the eigenvalue problem depending on the unknown distribution of $\mathbf{X}$ and Y. We find that this distribution of a complex form that cannot be evaluated directly, and it is not useful for estimating d. Kankainen (1995) suggests approximating this null distribution as a two-parameter Gamma distribution in hypothesis testing. This is a straightforward approximations of an infinite sum of Chi-squared random variables, but the asymptotic distribution is not helpful in developing a test to determine d either.

Table 1. Model 1: Entries are $\Delta(\beta, \eta_n)$ calculated from 100 replicates.

	Design (A)			**Design (B)**			**Design (C)**		
n	rMAVE	EFM	HSIC	rMAVE	EFM	HSIC	rMAVE	EFM	HSIC
100	0.6667	0.5563	0.1600	0.8826	0.8865	0.3673	0.8887	0.7970	0.4045
	±0.2670	±0.4444	±0.0944	±0.1757	±0.2633	±0.2621	±0.1692	±0.3426	±0.3201
200	0.3924	0.4329	0.1007	0.7242	0.7539	0.1951	0.7397	0.4974	0.1326
	±0.2561	±0.4570	±0.0260	±0.2648	±0.3961	±0.1219	±0.2933	±0.4466	±0.0486
400	0.1859	0.1251	0.0710	0.5101	0.6194	0.1245	0.2867	0.2354	0.0797
	±0.1144	±0.2872	±0.0155	±0.2936	±0.4638	±0.0731	±0.2627	±0.3554	±0.0210

3. Numerical Studies

In this section, we present some results of simulation studies and the analysis of a data set. For simulation studies, we chose two well-established dimension reduction methods: the refined Minimum Average Variance Estimation method (rMAVE, Xia et al. (2002)) and the Estimating Function Method (EFM, Cui, Härdle, and Zhu (2011)) to compare the finite sample performance with our proposed HSIC method. Xia's Matlab code and Cui's R code were used. We used the following measure (Li, Zha, and Chiaromonte (2005)) to evaluate accuracy: $\Delta(\mathcal{S}_1, \mathcal{S}_2) = \|P_{\mathcal{S}_1} - P_{\mathcal{S}_2}\|$, where $\| \cdot \|$ stands for the maximum singular value of a matrix; $\mathcal{S}_1$ and $\mathcal{S}_2$ are two q-dimensional subspace of $\mathbb{R}^p$; $P_{\mathcal{S}_1}$ and $P_{\mathcal{S}_2}$ are the orthogonal projections onto $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively.

3.1. Simulations

We used four single-index models. For each model, $\beta = (2, 1, 0, 0, 0, 0, 0, 0, 0, 0)^T/\sqrt{5}$ is the true direction, sample sizes were $n = 100$, $n = 200$ and $n = 400$, and ran 100 replicates with $p = 10$ (results, not reported, were similar with 200 and 500 replicates). Three designs on predictors were used for each model to cover a variety of model assumptions. Design (A) had predictors $\mathbf{X} \sim N(0, \mathbf{I}_{10})$; Design (B) had non-normal predictors; Design (C) concentrated on discrete predictors.

Model 1: Mean function model. We took oscillating mean function model discussed in Cui, Härdle, and Zhu (2011): $Y = \sin(3\pi/4 \cdot \beta^T\mathbf{X}) + \epsilon$ with $\epsilon \sim N(0, 0.2^2)$. Here Design (B) had $X_1 \sim t(5)$, $X_2 \sim F(4, 10)$, $X_3 \sim \chi^2(5)$, $X_4 \sim N(-8, 4)$, and $X_j \sim N(0, 1)$, $j = 5, \ldots, 10$; Design (C) had $X_1 \sim Binomial(10, 0.2)$ and $X_j \sim Poisson(1), j = 2, \ldots, 10$. The simulation results listed in Table 1 show that the HSIC approach has a much better performance, consistently across designs.

Model 2: Variance function model. We took the constant mean but variance function model discussed by Yin and Cook (2005): $Y = 0.2(\beta^T\mathbf{X})^2\epsilon$ with ϵ is standard normal. Here Design (B) was $X_1 \sim N(0, 1)$, $X_2 \sim t(5)$, $X_3 \sim$

Table 2. Model 2: Entries are $\Delta(\beta, \eta_n)$ calculated from 100 replicates.

	Design (A)			**Design (B)**			**Design (C)**		
n	rMAVE	EFM	HSIC	rMAVE	EFM	HSIC	rMAVE	EFM	HSIC
100	0.7993	0.8437	0.2901	0.7998	0.8894	0.2900	0.8029	0.9140	0.3677
	±0.1825	±0.1409	±0.1471	±0.1800	±0.1524	±0.1419	±0.1621	±0.1103	±0.1266
200	0.7531	0.8165	0.1763	0.7372	0.8590	0.2029	0.8145	0.8995	0.2320
	±0.1767	±0.1577	±0.0445	±0.1851	±0.1622	±0.1202	±0.1660	±0.0997	±0.0640
400	0.7281	0.8009	0.1208	0.7130	0.8721	0.1214	0.7437	0.8774	0.1587
	±0.1823	±0.1785	±0.0305	±0.1955	±0.1476	±0.0300	±0.1692	±0.1235	±0.0396

Table 3. Model 3: Entries are $\Delta(\beta, \eta_n)$ calculated from 100 replicates.

	Design (A)			**Design (B)**			**Design (C)**		
n	rMAVE	EFM	HSIC	rMAVE	EFM	HSIC	rMAVE	EFM	HSIC
100	0.5783	0.7239	0.6099	0.5187	0.6661	0.5129	0.5717	0.6612	0.5336
	±0.1396	±0.1846	±0.1961	±0.1384	±0.1990	±0.1793	±0.1034	±0.1813	±0.1586
200	0.4117	0.4674	0.4263	0.3760	0.4367	0.3533	0.4202	0.4900	0.4088
	±0.1097	±0.1721	±0.1400	±0.0883	±0.1921	±0.1119	±0.0868	±0.1285	±0.0380
400	0.2929	0.2578	0.2764	0.2587	0.2431	0.2374	0.3437	0.4090	0.4014
	±0.0713	±0.0790	±0.0716	±0.0658	±0.0697	±0.0614	±0.0859	±0.0838	±0.0917

$Gamma(9, 0.5)$, $X_4 \sim F(5, 12)$, and $X_j \sim N(0, 1)$, $j = 5, \ldots, 10$; Design (C) was $X_j \sim Poisson(1)$, $j = 1, 2, 3, 4$, and $X_j \sim N(0, 1)$, $j = 5, \ldots, 10$. Results in Table 2 show that the HSIC approach has the best performance across all sample sizes and settings considered. The advantage becomes more obvious as sample size increases. The EFM and rMAVE methods do not work well since the mean function they focused on is mostly noise, and the main available information lies in variance function. However, dMAVE (Xia (2007)) performs well in this case; it is not reported here.

Model 3: Categorical response model. We took the binary response model discussed by Cui, Härdle, and Zhu (2011):

$$P(Y_i = 1|\mathbf{X}) = \frac{\exp\{g(\beta^T \mathbf{X})\}}{[1 + \exp\{g(\beta^T \mathbf{X})\}]},$$
$$g(\beta^T \mathbf{X}) = \frac{\exp(5\beta^T \mathbf{X} - 2)}{1 + \exp(5\beta^T \mathbf{X} - 3)} - 1.5.$$

In Design (B), X_j, $j = 1, \ldots, 10$, were uniform $U(-2, 2)$. In Design (C), X_j, $j = 1, \ldots, 10$, were Bernoulli with probability of success 0.5. Results are shown in the Table 3. Here all methods have comparable results.

Accuracy of the Asymptotic Variance. To examine the performance of the asymptotic variance of the proposed estimator, we report results for Model 1

Table 4. Performance of the Asymptotic Variance of Model 1 with $n = 100$.

		$\hat{\eta}_1$	$\hat{\eta}_2$	$\hat{\eta}_3$	$\hat{\eta}_4$	$\hat{\eta}_5$
Design (A)	SE	0.0034	0.0078	0.0116	0.0078	0.0074
	η_n	0.8705	0.4769	-0.0597	-0.0588	0.0133
	SE(η_n)	0.0866	0.0684	0.0860	0.1036	0.0920
	95% C.I.	(0.7008, 1.0402)	(0.3428, 0.6110)	(-0.2283, 0.1088)	(-0.2618, 0.1443)	(-0.1671, 0.1936)
		$\hat{\eta}_6$	$\hat{\eta}_7$	$\hat{\eta}_8$	$\hat{\eta}_9$	$\hat{\eta}_{10}$
	SE	0.0063	0.0076	0.0094	0.0061	0.0088
	η_n	-0.0188	0.0214	-0.0246	-0.0603	0.0506
	SE(η_n)	0.0894	0.1095	0.0858	0.1079	0.1136
	95% C.I.	(-0.1941, 0.1564)	(-0.1932, 0.2360)	(-0.1927, 0.1436)	(-0.2718, 0.1512)	(-0.1720, 0.2733)

SE: standard error of the 100 variances for each estimated parameter based on 100 replicates.

For randomly generated data, η_n, SE(η_n), and 95% C.I. represent, respectively, the direction estimate, its standard error, and 95% confidence interval.

Design (A) in Table 4 for sample size $n = 100$. Sample variance was calculated for each of the 100 replicates using the derived asymptotic variance formula. Standard error (SE) of these variances for each estimated parameter is reported to show the stability of the variance estimates. The direction estimate η_n, its standard error (SE(η_n)), and the associated 95% confidence interval from randomly generated data are also reported in the table. The results show that our variance estimates are quite stable at $n = 100$. We expect results are improved when sample size increases. The results for other models and designs have shown similar patterns, and are not reported.

Model 4: Classic linear model. We took the simple linear regression model $Y = \beta^T\mathbf{X} + 0.2\epsilon$, where ϵ is a standard normal random variable. The Design (B) was $X_1 \sim t(5)$, $X_2 \sim F(4, 10)$, $X_3 \sim \chi^2(5)$, $X_4 \sim N(-8, 4)$, and $X_j \sim N(0, 1)$, $j = 5, \ldots, 10$; Design (C) was $X_1 \sim$ Binomial$(10, 0.2)$, and $X_j \sim$ Poisson(1), $j = 2, \ldots, 10$. The purpose here was to see, in a classic linear model, where the least squares method (LS) has the best performance, how our method performs. As shown in Table 5, the LS method has the best performance and, not surprisingly, EFM has the second best; it was designed for single-index models and in particular for the mean direction. rMAVE, which is also designed to recover directions in the mean function, is the third best. Here HSIC doesn't lose much in estimation accuracy. Indeed, all four methods have quite comparable results. Table 6 shows the variance performance in Model 4, indicating that our asymptotic variance estimates are stable in the classic linear model setting.

3.2. Permutation results

To illustrate how the proposed permutation test performs in predicting the true dimensionality ($d = 1$) of central subspace, we conducted simulation studies using the models of the previous section. We found that in the models that

Table 5. Model 4: Entries are $\Delta(\beta, \eta_n)$ calculated from 100 simulated samples.

	Design (A)				**Design (B)**				**Design (C)**			
n	LS	rMAVE	EFM	HSIC	LS	rMAVE	EFM	HSIC	LS	rMAVE	EFM	HSIC
100	0.0596	0.0908	0.0615	0.1083	0.0544	0.0819	0.0554	0.1347	0.0582	0.0938	0.0600	0.1188
	±0.0163	±0.0240	±0.0160	±0.0288	±0.0168	±0.0252	±0.0165	±0.0407	±0.0142	±0.0253	±0.0145	±0.0283
200	0.0421	0.0576	0.0431	0.0736	0.0377	0.0510	0.0381	0.0889	0.0416	0.0580	0.0423	0.0847
	±0.0109	±0.0147	±0.0109	±0.0203	±0.0097	±0.0136	±0.0097	±0.0262	±0.0098	±0.0151	±0.0100	±0.0201
400	0.0298	0.0360	0.0299	0.0507	0.0266	0.0359	0.0270	0.0651	0.0307	0.0396	0.0309	0.0576
	±0.0079	±0.0093	±0.0078	±0.0123	±0.0067	±0.0083	±0.0069	±0.0200	±0.0069	±0.0094	±0.0070	±0.0156

Table 6. Performance of the Asymptotic Variance of Model 4 with $n = 100$.

		$\hat{\eta}_1$	$\hat{\eta}_2$	$\hat{\eta}_3$	$\hat{\eta}_4$	$\hat{\eta}_5$
Design (A)	SE	0.0019	0.0032	0.0030	0.0025	0.0029
	η_n	0.9016	0.4219	-0.0173	0.0143	0.0620
	SE(η_n)	0.0673	0.1003	0.1004	0.0918	0.0857
	95% C.I.	(0.7697, 1.0336)	(0.2254, 0.6185)	(-0.2140, 0.1795)	(-0.1656, 0.1943)	(-0.1060, 0.2300)
		$\hat{\eta}_6$	$\hat{\eta}_7$	$\hat{\eta}_8$	$\hat{\eta}_9$	$\hat{\eta}_{10}$
	SE	0.0032	0.0033	0.0028	0.0028	0.0026
	η_n	-0.0181	-0.0346	-0.0511	-0.0062	-0.0219
	SE(η_n)	0.0883	0.0892	0.1104	0.1132	0.1288
	95% C.I.	(-0.1911, 0.1550)	(-0.2094, 0.1403)	(-0.2675, 0.1653)	(-0.2281, 0.2156)	(-0.2743, 0.2306)

SE: standard error of the 100 variances for each estimated parameter based on 100 replicates.

For randomly generated data, η_n, SE(η_n), and 95% C.I. represent, respectively, the direction estimate, its standard error, and 95% confidence interval.

we simulated, the proposed permutation test provided accurate estimates of d when the underlying distribution of $\mathbf{X}$ was normal (Design (A)). Under Design (B) and Design (C), our permutation test tended to be conservative, in the sense that it tended to predict a higher dimensionality ($d > 1$). For $n = 400$ compared to $n = 100$, we saw less accuracy of the estimated d: for Designs (B) and (C), the percentages of correctly estimated d had a decreasing trend as sample size increased. For our test, a smaller sample size, say 100, may be best for prediction of the true dimensionality of central subspace. We propose a modified version of our test, when $n > 100$. We randomly select 100 observations, for m times without replacement, and record the estimated dimension as d_j, $j = 1, \ldots, m$. Then, the estimated d has the largest frequency among the d_j's. In our simulations, we set $m = 20$. Table 7 indicates that the modified test provides reliable estimates for the cases considered in Designs (A), (B), and (C). Entries are percentages of the estimated dimension calculated from 100 replicates, and the number of permutations for each replicate was 200.

3.3. Data

We analyzed a New Zealand Horse Mussels data set; it has been discussed

Table 7. Permutation test results. Percentages of estimated dimensionality for single-index models. Entries are calculated from 100 replicates. Number of permutations is 200.

		Design (A)			Design (B)			Design (C)		
	n	$d=0$	$d=1$	$d>1$	$d=0$	$d=1$	$d>1$	$d=0$	$d=1$	$d>1$
Model 1	50	0	98	2	4	93	3	0	98	2
	100	0	97	3	0	87	13	0	94	6
	200	0	100	0	0	98	2	0	99	1
	400	0	100	0	0	99	1	0	100	0
Model 2	50	3	94	3	1	94	5	2	94	4
	100	0	95	5	2	96	2	0	90	10
	200	0	100	0	0	100	0	0	91	1
	400	0	100	0	0	99	1	0	100	0
Model 3	50	2	97	1	0	100	0	1	89	10
	100	0	100	0	0	100	0	0	80	20
	200	0	100	0	0	100	0	0	98	2
	400	0	100	0	0	100	0	0	100	0

by Cook (1998). This data set contains 201 observations collected at 5 sites in the Marlborough Sounds of the Northeast of New Zealand's South Island. The response variable is muscle mass M, the edible portion of the mussel, in grams. The quantitative predictors of interests are shell length L, shell width W, each in mm, and shell mass S in grams. We used the transformation of the predictors that was suggested by Cook (1998): $\mathbf{X} = (L, W^{0.36}, S^{0.11})$.

Noticing that shell length L is on a larger scale than the other predictors, we standardized all predictors into $\mathbf{Z}$-scale, $\mathbf{Z} = (Z_L, Z_W, Z_S)$, with mean 0 and unit variance. Our permutation test indicated that a single-index model would be appropriate to model these data. The normalized single direction estimated by HSIC approach was $\hat{\beta}_H = (0.1897, -0.0604, 0.9800)^T$. A 2D scatter plot of M versus single-index $\hat{\beta}_H^T\mathbf{Z}$ suggests a second degree polynomial model, Fit1, (Fit1$= 14.16 + 8.05(\hat{\beta}_H^T\mathbf{Z}) + 1.40(\hat{\beta}_H^T\mathbf{Z})^2$). Fit1 in Figure 1 shows a good fit of the single-index model. Model diagnostic plots (not reported), including residual and QQ plots, indicate no significant evidence that normality assumption is violated. We conclude that the edible portion of the mussel, is positively related to the single-index, while single-index is dominated by the mussel's shell mass: increases in shell mass tend to grow the edible portion of the mussel polynomially.

The HSIC estimate was compared to the rMAVE, EFM, and SIR estimates reported in Cook (1998), denoted as $\hat{\beta}_M$, $\hat{\beta}_E$, and $\hat{\beta}_S$, respectively. We then investigated the pairwise correlations $Corr(\hat{\beta}_H^T\mathbf{Z}, \hat{\beta}_M^T\mathbf{Z})$, $Corr(\hat{\beta}_H^T\mathbf{Z}, \hat{\beta}_E^T\mathbf{Z})$, and $Corr(\hat{\beta}_H^T\mathbf{Z}, \hat{\beta}_S^T\mathbf{Z})$, as well as the distances Δ (Table 8) between the directions. The benchmark criterion proposed in Li, Wen, and Zhu (2008) was adopted for closeness of directions: the average of 10,000 distances, $\Delta_m(\beta_1, \beta_2) = \|P_{\beta_1} - P_{\beta_2}\|$,

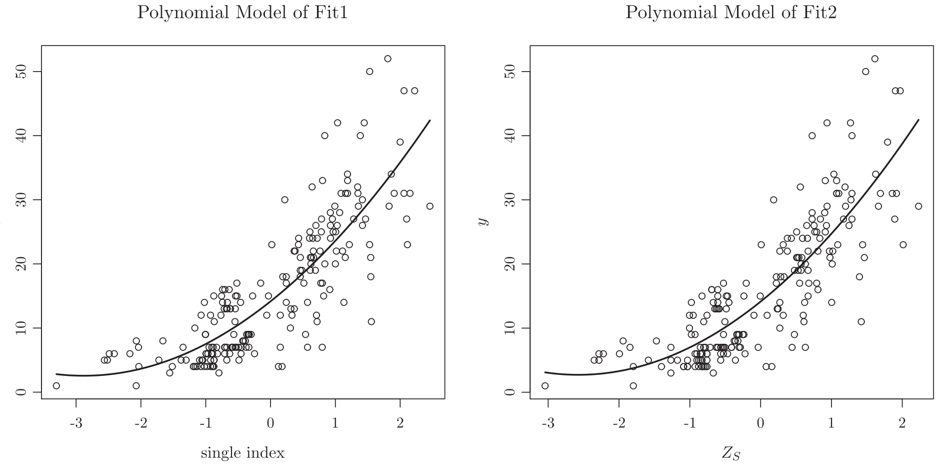

Figure 1. Second order polynomial fit of single-index model.

Table 8. Comparisons of Correlation and Distance.

Method	Correlation	Distance	Proportion
HSIC vs. rMAVE	1.0000	0.3778	0.0737
HSIC vs. EFM	0.9997	0.2700	0.0385
HSIC vs. SIR	1.0000	0.1352	0.0097

where β_1 and β_2 are randomly generated directions in $\mathbb{R}^3$ such that $\beta_1 \perp\!\!\!\perp \beta_2$. Benchmark distance has an average of 0.7848 with standard error 0.2243. The proportion of benchmark distances less or equal to a distance between two direction estimates in comparison are reported in Table 8. A small proportion indicates two estimates agree with each other in recovering the true direction. Both correlation and benchmark criteria suggest that rMAVE, EFM, and SIR methods obtain essentially the same direction as does HSIC.

The asymptotic variance formula derived earlier can be incorporated here to judge the significance of the predictors. Their standard errors are 0.3355, 0.4033, and 0.4471, respectively. The 95% confidence intervals for the standardized predictors are (-0.4679, 0.8473), (-0.8508, 0.7301), and (0.1037, 1.856), respectively, which suggests that only the standardized shell mass predictor, Z_S, is significant. A second degree polynomial model, Fit2, (Fit2$= 14.14 + 8.87Z_S + 1.72Z_S^2$), is fitted and plotted in the right panel of Figure 1. Compared to Fit1, we can see that two models perform very similar to each other. Leave-one-out cross validation was conducted on both models, with Residual Sum of Squares (RSS) 36.2 for Fit1 and 36.4 for Fit2. We conclude that the parsimonious model (Fit2) is our model, and that shell length L and shell width W are statistically insignificant in predicting muscle mass M, the edible portion of the New Zealand horse mussel.

4. Discussion

Our HSIC approach requires relatively weak conditions for estimating single-index models. There are many choices for the weights or kernel. We only choose Gaussian kernel in this paper. Simulation studies have shown its merits for models we studied, as it consistently provides stable results for normal predictors as well as robust estimates for non-normal and categorical cases. The proposed permutation test based on HSIC can be useful in practice. It provides a statistical justification for applying single-index methods. Although this paper focuses on single-index estimation, the idea can be extended to multiple-index models. This is currently under investigation.

Acknowledgements

The authors thank the Editor, an associate editor and two referees for their valuable comments, which greatly improved the paper. Yin's work was supported in part by a NSF grant 1205546.

Appendix

Proof of Proposition 1. Let η_0 be the projection of β onto η, $\eta_0 = c\eta$, where c is a scalar. If $\eta_0^{\perp} = \beta - \eta_0$, where the orthogonality '$\perp$' means $\eta_0^T \Sigma_X \eta_0^{\perp} = 0$, then $1 = \beta^T \Sigma_X \beta = c^2 + \eta_0^{\perp,T} \Sigma_X \eta_0^{\perp} \geq c^2$. Hence, $|c| \leq 1$. Now,

$$
\begin{aligned}
H(\beta^T \mathbf{X}, \mathbf{Y}) &= \int |\mathrm{E}e^{i<t,\beta^T \mathbf{X}>+i<s,Y>} - \mathrm{E}e^{i<t,\beta^T \mathbf{X}>}\mathrm{E}e^{i<s,Y>}|^2 dw \\
&= \int |\mathrm{E}\{e^{i<t,\beta^T \mathbf{X}>}[\mathrm{E}(e^{i<s,Y>}|\mathbf{X})]\} - \mathrm{E}e^{i<t,\beta^T \mathbf{X}>}\mathrm{E}e^{i<s,Y>}|^2 dw \\
&= \int |\mathrm{E}\{e^{i<t,(\eta_0^T+\eta_0^{\perp,T})\mathbf{X}>}[\mathrm{E}(e^{i<s,Y>}|\eta_0^T \mathbf{X})]\} \\
&\quad -\mathrm{E}e^{i<t,(\eta_0^T+\eta_0^{\perp,T})\mathbf{X}>}\mathrm{E}e^{i<s,Y>}|^2 dw \\
&= \int |\mathrm{E}\{e^{i<t,\eta_0^{\perp,T}\mathbf{X}>}\}\{\mathrm{E}e^{i<t,\eta_0^T \mathbf{X}>+i<s,Y>} - \mathrm{E}e^{i<t,\eta_0^T \mathbf{X}>}\mathrm{E}e^{i<s,Y>}\}|^2 dw \\
&= \int |\mathrm{E}\{e^{i<t,\eta_0^{\perp,T}\mathbf{X}>}\}|^2 |\mathrm{E}e^{i<t,\eta_0^T \mathbf{X}>+i<s,Y>} - \mathrm{E}e^{i<t,\eta_0^T \mathbf{X}>}\mathrm{E}e^{i<s,Y>}|^2 dw \\
&\leq \int |\mathrm{E}e^{i<t,\eta_0^T \mathbf{X}>+i<s,Y>} - \mathrm{E}e^{i<t,\eta_0^T \mathbf{X}>}\mathrm{E}e^{i<s,Y>}|^2 dw \\
&= H(\eta_0^T \mathbf{X}, \mathbf{Y}) = H(\eta^T \mathbf{X}, \mathbf{Y}).
\end{aligned}
$$

The third equality follows from the assumption $Y \perp\!\!\!\perp \mathbf{X}|\eta^T \mathbf{X}$, and $\eta_0 = c\eta$, where $|c| \leq 1$. The fourth equality follows from the assumption $\eta^T \mathbf{X} \perp\!\!\!\perp \alpha^T \mathbf{X}$, where (η, α) forms a $p \times p$ matrix such that $(\eta, \alpha)^T \Sigma_X (\eta, \alpha) = \mathbf{I}$. The last inequality follows from the characteristic function with equality if $|c| = 1$. Thus the maximum is achieved when $|c| = 1$. The last equality holds because H is scale invariant, due to the choice of kernel (Kankainen (1995)).

Proof of Proposition 2. If $G = (\mathbf{I}_p, 0)$, then $\eta_n = G\theta_n$ and $\eta = G\theta$. The conclusion follows from Lemma 3. Lemma 3 is proved in the supplementary file.

Proof of Proposition 3. Let $G = (\mathbf{I}_p, 0)$ be a $p \times (p+1)$ matrix, where $\mathbf{I}_p$ is a $p \times p$ identity matrix. Then $\eta_n = G\theta_n$ and $\eta = G\theta$. By Lemma 4, we have $\sqrt{n}(\eta_n - \eta) = \sqrt{n}G(\theta_n - \theta) \xrightarrow{D} N(0, V_{11})$, where $V_{11} = GVG^T$. Lemma 4 is proved in the supplementary file.

References

Bonnans, J., Gilbert, J., Lemaráchal, C. and Sagastizábal, C. A. (2006). *Numerical Optimization: Theoretical and Practical Aspects.* Springer-Verlag, Berlin.

Cook, R. D. (1996). Graphics for regressions with a binary response. *J. Amer. Statist. Assoc.* **91**, 983-992.

Cook, R. D. (1998). *Regression Graphics: Ideas for Studying Regressions through Graphics.* Wiley, Inc.

Cook, R. D. and Yin, X. (2001). Dimension reduction and visualization in discriminant analysis (with discussion). *Austral. N. Z. J. Statist.* **43**, 147-199.

Cui, X., Härdle, W. K. and Zhu, L. (2011). The EFM approach for single-index models. *Ann. Statist.* **39**, 1658-1688.

Fukumizu, K., Bach, F. R. and Jordan, M. I. (2004). Dimensionality reduction for supervised learning with reproducing kernel Hilbert spaces. *J. Machine Learning Research* **5**, 73-99.

Fukumizu, K., Bach, F. R. and Jordan, M. I. (2009). Kernel dimension reduction in regression. *Ann. Appl. Statist.* **37**, 1871-1905.

Gretton, A., Fukumizu, K. and Spiperumbudur, B. K. (2009). Discussion of: Brownian distance covariance. *Ann. Appl. Statist.* **3**, 1285-1294.

Gretton, A., Fukumizu, K., Teo, C. H., Song, L., Schëlkopf, B. and Smola, A. J. (2008). A kernel statistical test of independence. *Adv. Neural Information Processing Systems* **20**, 585-592.

Hall, P. and Li, K.-C. (1993). On almost linearity of low dimensional projections from high dimensional data. *Ann. Statist.* **21**, 867-889.

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

Härdle, W. and Stoker, T. (1989). Investigating smooth multiple regression by method of average derivatives. *J. Amer. Statist. Assoc.* **84**, 986-995.

Hristache, M. , Juditsky, A. Polzehl, J. and Spokoiny, V. (2001). Structure adaptive approach for dimension reduction. *Ann. Statist.* **29**, 1537-1566.

Ichimura, H. (1993). Semiparametric least squares (sls) and weighted sls estimation of single-index models. *J. Econom.* **58**, 71-120.

Kankainen, A. (1995). Consistency testing of total independence based on the empirical characteristic function. Ph.D. thesis, University of Jyväskylä.

Li, B., Wen, S. and Zhu, L. (2008). On a projective resampling method for dimension reduction with multivariate responses. *J. Amer. Statist. Assoc.* **103**, 1177-1186.

Li, B., Zha, H. and Chiaromonte, F. (2005). Contour regression: a general approach to dimension reduction. *Ann. Statist.* **33**, 1580-1616.

Li, K.-C. (1991). Sliced inverse regression for dimension reduction (with discussion). *J. Amer. Statist. Assoc.* **86**, 316-342.

Li, K.-C. (1992). On principal Hessian directions for data visualization and dimension reduction. *J. Amer. Statist. Assoc.* **87**, 1025-1039.

Sejdinovic, D., Gretton, A., Sriperumbudur, B. and Fukumizu, K. (2012). Hypothesis testing using pairwise distances and associated kernels. Proceedings of the 39th International Conference on Machine Learning, Edinburge, Scotland, UK.

Sheng, W. and Yin, X. (2013). Direction estimation in single-index models via distance covariance. *J. Multivariate Anal.* **122**, 148-161.

Székely, G. J. and Rizzo, M. L. (2009). Brownian distance covariance. *Ann. Appl. Statist.* **3**, 1236-1265.

Székely, G. J., Rizzo, M. L. and Bakirov, N. (2007). Measuring and testing independence by correlation of distance. *Ann. Statist.* **35**, 2769-2794.

Xia, Y. (2007). A constructive approach to the estimation of dimension reduction directions. *Ann. Statist.* **35**, 2654-2690.

Xia, Y., Tong, H., Li, W. and Zhu, L.-X. (2002). An adaptive estimation of dimension reduction space. *J. Roy. Statist. Soc. Ser. B* **64**, 363-410.

Yin, X. and Cook, R. D. (2002). Dimension reduction for the conditional kth moment in regression. *J. Roy. Statist. Soc. Ser. B* **64**, 159-175.

Yin, X. and Cook, R. D. (2005). Direction estimation in single-index regressions. *Biometrika* **92**, 371-384.

Yin, X., Li, B. and Cook, R. D. (2008). Successive direction extraction for estimating the central subspace in a multi-index regression. *J. Multivariate Anal.* **99**, 1733-1757.

Department of Statistics, University of Georgia, Athens, GA 30602, U.S.A.

E-mail: nanzhang@uga.edu

Department of Statistics, University of Kentucky, Lexington KY 40536-0082, U.S.A.

E-mail: yinxiangrong@uky.edu

(Received May 2013; accepted April 2014)