Computational Statistics and Data Analysis 176 (2022) 107562

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

www.elsevier.com/locate/csda

On cross-distance selection algorithm for hybrid sufficient dimension reduction

Yujin Park, Kyongwon Kim *, Jae Keun Yoo *

Department of Statistics, Ewha Womans University, Seoul 120-750, Republic of Korea

ARTICLE INFO

Article history:
Received 24 February 2022
Received in revised form 26 June 2022
Accepted 26 June 2022
Available online 4 July 2022

Dedicated to the memory of Xiangrong Yin, my great and greater academic elder brother. I shall always appreciate your inspiration and your desire to make me a better teacher and researcher. Every moment with you has been joyful, and I will never forget our time together. Peace be with you in Heaven.

Keywords:
Covariance methods
Directional regression
Hybrid dimension reduction
Sliced average variance estimation
Sliced inverse regression
Sufficient dimension reduction

ABSTRACT

Given the extensive development of a variety of sufficient dimension reduction (SDR) methodologies, Ye and Weiss (2003) proposed a hybrid SDR method combining two pre-existing SDR methods. In particular, they used a bootstrap approach to select a proper weight. Since bootstrapping is computationally intensive and time-consuming, the hybrid reduction approach has not been widely used, although it is more accurate than conventional single SDR methods. To overcome these deficits, we propose a novel cross-distance selection algorithm. Similar to the bootstrapping method, the proposed selection algorithm is data-driven and has a strong rationale for its performance. The numerical studies demonstrate that the chosen hybrid method from our proposed algorithm offers a good estimation quality and reduces the computing time dramatically at the same time. Furthermore, our real data analysis confirms that the proposed selection algorithm has potential advantages with its practical usefulness over the existing bootstrapping method.

1. Introduction

Sufficient dimension reduction (SDR) for regression is a statistical methodology designed to reduce the dimension of predictors $\mathbf{X} \in \mathbb{R}^p$ without losing information on $Y|\mathbf{X}$, where $\mathbf{Y} \in \mathbb{R}$. Its main purpose is to replace the p-dimensional original predictors $\mathbf{X}$ with a lower-dimensional linearly transformed predictors $\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}$ without loss of information on $Y|\mathbf{X}$, which is equivalently stated as follows:

$$Y \perp\!\!\!\perp \mathbf{X}|\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}, \tag{1}$$

where $\perp\!\!\!\perp$ stands for independence, $\eta \in \mathbb{R}^{p\times d}$, and $d \leq p$.

A subspace spanned by the columns of $\boldsymbol{\eta}$ that satisfies (1) is called a dimension reduction subspace. If the support of X is convex, the intersection of all dimension reduction subspaces is again a dimension reduction subspace, called a central subspace and denoted $\mathcal{S}_{Y|\boldsymbol{X}}$. See Yin et al. (2008) for more details. If $\mathcal{S}_{Y|\boldsymbol{X}}$ exists, it is unique and the smallest among all

* Corresponding authors.
E-mail addresses: kimk@ewha.ac.kr (K. Kim), peter.yoo@ewha.ac.kr (J.K. Yoo).

https://doi.org/10.1016/j.csda.2022.107562
0167-9473/

Table 1
Kernel matrices of the five sufficient dimension reduction methods.

Methods	Kernel matrices
Sliced Inverse Regression (SIR; Li, 1991)	$\mathbf{M}_{\text{SIR}} = \text{cov}\{E(\mathbf{Z}\|Y)\}$
Sliced Average Variance Estimation (SAVE; Cook and Weisberg, 1991)	$\mathbf{M}_{\text{SAVE}} = E\{\mathbf{I}_p - \text{cov}(\mathbf{Z}\|Y)\}^2$
Directional Regression (DR; Li and Wang, 2007)	$\mathbf{M}_{\text{DR}} = E\{E(\mathbf{Z}\mathbf{Z}^{\text{T}}\|Y)\}^2 + 2E\{E(\mathbf{Z}\|Y)E(\mathbf{Z}^{\text{T}}\|Y)\}^2 + 2E\{E(\mathbf{Z}^{\text{T}}\|Y)E(\mathbf{Z}\|Y)E(\mathbf{Z}\|Y)E(\mathbf{Z}^{\text{T}}\|Y)\} - 2\mathbf{I}_p.$
Covariance method (covk; Yin and Cook, 2002)	$\mathbf{M}_{\text{covk}} = \mathbf{K}_q\mathbf{K}_q^{\text{T}}$, where $W = (Y - E(Y))/\sqrt{\text{var}(Y)}$ and $\mathbf{K}_q = \{\text{cov}(\mathbf{Z}, W), \text{cov}(\mathbf{Z}, W^2), \ldots, \text{cov}(\mathbf{Z}, W^q)\}$
Principal Hessian Direction (pHd; Li, 1992, Cook (1998b))	$\mathbf{M}_{\text{pHd}} = E\{Y - E(Y)\}\mathbf{Z}\mathbf{Z}^{\text{T}}$

possible dimension reduction subspaces. We recommend Cook (1998a) for a comprehensive discussion of $\mathcal{S}_{Y|\boldsymbol{X}}$. Hereafter, $\boldsymbol{\eta}$ and d stand for an orthonormal basis for $\mathcal{S}_{Y|\boldsymbol{X}}$, and its structural dimension, respectively.

Classical SDR methods that still dominate among the many are sliced inverse regression (SIR; Li, 1991), sliced average variance estimation (SAVE; Cook and Weisberg, 1991), principal Hessian directions (pHd; Li, 1992; Cook, 1998b), covariance method (covk; Yin and Cook, 2002), and directional regression (DR; Li and Wang, 2007). These methods are formulated as a generalized eigen-decomposition problem of the kernel matrix. Recently Ye and Weiss (2003) has defined a new class of dimension reduction methods that combine two existing SDR methods. For example, let $\mathbf{M}_{\text{SIR}}$ and $\mathbf{M}_{\text{SAVE}}$ be the kernel matrices constructed by SIR and SAVE, respectively, where a kernel matrix is one whose columns span $\mathcal{S}_{Y|\boldsymbol{X}}$. Then, the columns of a weighted sum of form $\alpha\mathbf{M}_{\text{SIR}} + (1-\alpha)\mathbf{M}_{\text{SAVE}}$ also span $\mathcal{S}_{Y|\boldsymbol{X}}$ for $0 \le \alpha \le 1$, and hence this mixture constructs a class of SDR methodologies. The combination of two SDR methods is called *hybrid sufficient dimension reduction*. To select an optimal value of α, the average distances between the original-sample and bootstrap-sample estimates of $\mathcal{S}_{Y|\boldsymbol{X}}$ are computed for a set of α, and we choose α to have the minimum average distance. This bootstrap approach will be discussed in more detail in Section 2.

It turns out that a hybrid of different SDR methods often provides a better and more reliable dimension reduction, but this is often overlooked in practice due to the computational intensity of choosing the two SDR methods and computing α.

The main purpose of the paper is to introduce a practical approach for selecting a proper hybrid SDR method and achieving computational efficiency compared to the bootstrap method (Ye and Weiss, 2003). In this paper, we present a novel algorithm to select an appropriate hybrid method. The key idea of the proposed selection algorithm is to measure distances between the two different hybrid methods for the candidate weights, and to choose the weights to achieve the smallest discrepancy. Furthermore, to choose one among hybrid methods, we adopt a basis-adaptive algorithm from Yoo (2018). This proposed a method to choose the best SDR method among four SDR methods, namely SIR, SAVE, PHD, and covk, in a linear regression problem. The basic idea is that if the estimated SDR subspaces from the two methods have a high correlation, they can be treated as a reasonable approach to recovering a central subspace, even though they may not be the best method. A trace correlation (Hooper, 1959) is used as a correlation measure between two subspaces. The trace correlation ranges from 0 to 1, a larger value signifying a higher correlation. The mathematical details of the trace correlation are presented in Section 2.3. Yoo (2018) first determine the pair of methods with the highest correlation among the four methods. Next, the combination of methods that do not contain one of the methods in the largest correlation pair is ruled out. They then find a couple of methods with the highest correlation among the pairs except the best one in the first step. Finally, the method that is included in both the best combination and the pair from the last step is chosen. This approach still maintains methodological comprehensiveness for the hybrid reduction while dramatically reducing the computing time.

The organization of the paper is as follows. In Section 2, the hybrid dimension estimation and bootstrapping are introduced. Section 3 is devoted to developing a new selection algorithm. Numerical studies and a real data application are presented in Section 4, and we summarize our work in Section 5.

2. Hybrid sufficient dimension reduction

2.1. Candidate methodologies

The five SDR methods in the previous section are considered candidate SDR approaches for the foundational ingredients in the hybrid method. Here, we supplement covk and DR methods in addition to the three methods, SIR, pHd, and SAVE, which are used in Ye and Weiss (2003) (Table 1). For notational simplicity, let $\Sigma = \text{cov}(\mathbf{X})$, $\mathbf{Z} = \Sigma^{-1/2}(\mathbf{X} - E(\mathbf{X}))$ and $\mathcal{S}(\mathbf{B})$ be a subspace spanned by the columns of $\mathbf{B} \in \mathbb{R}^{p\times q}$.

Estimating $\mathcal{S}_{Y|\boldsymbol{X}}$ using the five SDR methods can be formulated as eigendecomposition of the kernel matrices because it can be shown that $\Sigma^{-1/2}\mathcal{S}(\mathbf{M}_\bullet) \subseteq \mathcal{S}_{Y|\boldsymbol{X}}$, where $\mathbf{M}_\bullet$ denotes any one of the kernel matrices in Table 1.

For SIR, SAVE, and DR, the construction of sample versions of $E(\mathbf{Z}|Y)$ and $E(\mathbf{Z}\mathbf{Z}^{\text{T}}|Y)$ is straightforward, when Y is categorical. If $\mathbf{Y}$ is many-valued or continuous, it is categorized by dividing its range into h slices. $\mathbf{M}_{\text{DR}}$ is presented as a

combination of $E(\mathbf{Z}|Y)$ and $E(\mathbf{Z}\mathbf{Z}^{\mathrm{T}}|Y)$, which are fundamental parts of SIR and SAVE, respectively. Therefore, DR can be seen as incorporating a piece of information from SIR and SAVE.

The relation between the methods is as follows. According to Li and Wang (2007), DR can estimate $\mathcal{S}_{Y|\boldsymbol{X}}$ exhaustively and spans the same subspace as SAVE under some conditions. Both SAVE and DR estimate $\mathcal{S}_{Y|\boldsymbol{X}}$ more comprehensively than the other three methodologies. Proposition 4 in Yin and Cook (2002) and Corollary 1 in Ye and Weiss (2003) show that $\mathcal{S}(\mathbf{M}_{\mathrm{covk}}) \subseteq \mathcal{S}(\mathbf{M}_{\mathrm{SIR}})$ and $\mathcal{S}(\mathbf{M}_{\mathrm{SIR}}) \subseteq \mathcal{S}(\mathbf{M}_{\mathrm{SAVE}})$, respectively. So, the following relationships among $\mathbf{M}_{\mathrm{covk}}$, $\mathbf{M}_{\mathrm{SIR}}$, $\mathbf{M}_{\mathrm{SAVE}}$, and $\mathbf{M}_{\mathrm{DR}}$ are directly established:

$$\mathcal{S}(\mathbf{M}_{\mathrm{covk}}) \subseteq \mathcal{S}(\mathbf{M}_{\mathrm{SIR}}) \subseteq \mathcal{S}(\mathbf{M}_{\mathrm{SAVE}}) = \mathcal{S}(\mathbf{M}_{\mathrm{DR}}). \tag{2}$$

No specific relationships between $\mathbf{M}_{\mathrm{pHd}}$ and the other kernel matrices are known at present.

2.2. Candidate methodologies for hybrid dimension reduction

The basic philosophy of the hybrid of two SDR methodologies is to improve the estimation of $\mathcal{S}_{Y|\boldsymbol{X}}$ by overcoming the deficits of each method. Therefore, a combination of similar methods is not desirable. The sample behaviors of SIR and covk are known to be similar according to Yin and Cook (2002) and Yoo (2009). The two methods of pHd and SAVE are mainly used when the regression is symmetric. Thus, we rule out a hybrid of SIR and covk and that of SAVE and pHd. Since DR shares information with both SIR and SAVE, as discussed in Section 2.1, DR will be combined with covk alone. Following these guidelines, we consider the following five hybrid candidates:

(1) SIR and SAVE: $\alpha\mathbf{M}_{\mathrm{SIR}} + (1-\alpha)\mathbf{M}_{\mathrm{SAVE}}$;
(2) SIR and pHd: $\alpha\mathbf{M}_{\mathrm{SIR}} + (1-\alpha)\mathbf{M}_{\mathrm{pHd}}$;
(3) covk and DR: $\alpha\mathbf{M}_{\mathrm{covk}} + (1-\alpha)\mathbf{M}_{\mathrm{DR}}$;
(4) covk and SAVE: $\alpha\mathbf{M}_{\mathrm{covk}} + (1-\alpha)\mathbf{M}_{\mathrm{SAVE}}$;
(5) covk and pHd: $\alpha\mathbf{M}_{\mathrm{covk}} + (1-\alpha)\mathbf{M}_{\mathrm{pHd}}$

A hybrid candidate is equivalent to either SIR or covk when α is equal to 1, while it becomes either of SAVE, pHd, or DR with $\alpha = 0$.

2.3. Bootstrap method

In general, minimum variance is preferred when developing an unbiased estimator. Ye and Weiss (2003) proposed the mean of the distance between a subspace of a sample kernel matrix and a bootstrap sample kernel matrix to measure the variability of an estimator and determine the best combination from a class of SDR methods. Ye and Weiss (2003) combined two candidate matrices from $\mathbf{M}_{\mathrm{SIR}}$, $\mathbf{M}_{\mathrm{SAVE}}$, $\mathbf{M}_{\mathrm{y-pHd}}$, $\mathbf{M}_{\mathrm{y^2-pHd}}$ and selected an optimal α based on the grid search method to improve estimation accuracy. Therefore, in the bootstrap approach, it is essential to measure a distance between two subspaces. Ye and Weiss (2003) adopted the vector correlation coefficient q (Hotelling, 1936) and the trace correlation u (Hooper, 1959). Define two k-dimensional subspaces of $\mathcal{S}(\mathbf{A} \in \mathbb{R}^{p\times k})$ and $\mathcal{S}(\mathbf{B} \in \mathbb{R}^{p\times k})$ such that $\mathbf{A}^{\mathrm{T}}\mathbf{A} = \mathbf{I}_k$ and $\mathbf{B}^{\mathrm{T}}\mathbf{B} = \mathbf{I}_k$. Then, let ρ_i^2, $i = 1, \ldots, k$, be the ordered eigenvalues of $\mathbf{B}^{\mathrm{T}}\mathbf{A}\mathbf{A}^{\mathrm{T}}\mathbf{B}$. Then, the vector correlation q and the trace correlation r are defined as follows:

$$q^2(\mathbf{A},\mathbf{B}) = \prod_{i=1}^{k} \rho_i^2 \quad \text{and} \quad u^2(\mathbf{A},\mathbf{B}) = \frac{1}{k}\sum_{i=1}^{k} \rho_i^2.$$

The values of $q^2(\mathbf{A},\mathbf{B})$ and $u^2(\mathbf{A},\mathbf{B})$ vary from 0 to 1, and $q^2(\mathbf{A},\mathbf{B})$ and $u^2(\mathbf{A},\mathbf{B})$ are equal to 1, if the two subspaces of $\mathcal{S}(\mathbf{A})$ and $\mathcal{S}(\mathbf{B})$ are equivalent. To transform a concept of correlation (higher means closer) to one of distances (smaller means closer), we consider the following:

$$q_D(\mathbf{A},\mathbf{B}) = 1 - \sqrt{q^2(\mathbf{A},\mathbf{B})} \quad \text{and} \quad u_D(\mathbf{A},\mathbf{B}) = 1 - \sqrt{u^2(\mathbf{A},\mathbf{B})}.$$

Hereafter, we use u_D as the distance measure between two subspaces, and it will be called *trace correlation distance*.

Next, we select candidate SDR methods and construct their corresponding sample kernel matrices $\hat{\mathbf{M}}_\bullet$. From the original sample $(Y_i, \mathbf{X}_i)$, $i = 1, \ldots, n$, generate N bootstrap samples $(Y_i^b, \mathbf{X}_i^b)$, $b = 1, \ldots, N$, and obtain bootstrap sample kernel matrices $\hat{\mathbf{M}}_\bullet^b$. For each method, compute $q_D^b(\hat{\mathbf{M}}_\bullet, \hat{\mathbf{M}}_\bullet^b)$ and ${u_D}^b(\hat{\mathbf{M}}_\bullet, \hat{\mathbf{M}}_\bullet^b)$ for $b = 1, \ldots, N$. Ye and Weiss (2003) discussed that a best method among the candidates should have smaller distances between $\mathcal{S}(\hat{\mathbf{M}}_\bullet)$ and $\mathcal{S}(\hat{\mathbf{M}}_\bullet^b)$ than the others. Accordingly, for each method, we compute the average distances of $\bar{q}_D$ and $\bar{u}_D$ using $\bar{q}_D = \frac{1}{N}\sum_{b=1}^{N} q_D^b(\hat{\mathbf{M}}_\bullet, \hat{\mathbf{M}}_\bullet^b)$ and $\bar{u}_D = \frac{1}{N}\sum_{b=1}^{N} {u_D}^b(\hat{\mathbf{M}}_\bullet, \hat{\mathbf{M}}_\bullet^b)$. The remaining step is then to select the method to give the smallest $\bar{q}_D$ and $\bar{u}_D$.

Table 2
Values of α achieving the minimum average of u_D and the corresponding average from the six simulated model: $\bar{u}_D$ is the average of u_D.

Method	Model 1		Model 2		Model 3		Model 4		Model 5		Model 6	
	α	$\bar{u}_D$	α	$\bar{u}_D$	α	$\bar{u}_D$	α	$\bar{u}_D$	α	$\bar{u}_D$	α	$\bar{u}_D$
SIR-SAVE	0.9	0.027	0.4	0.071	0.6	0.106	0.8	0.151	0.8	0.225	0.8	0.191
SIR-pHd	1.0	0.028	0.3	0.079	0.8	0.107	1.0	0.164	0.9	0.260	1.0	0.202
covk-DR	0.9	0.028	0.1	0.064	0.4	0.105	0.8	0.104	0.9	0.227	0.9	0.117
covk-SAVE	0.8	0.031	0.1	0.069	0.7	0.143	0.8	0.113	0.9	0.230	0.8	0.117
covk-pHd	1.0	0.031	0.0	0.080	0.9	0.158	1.0	0.118	1.0	0.238	1.0	0.123

3. Cross-distance approach for hybrid dimension reduction

3.1. Narrowing selection of hybrid methods

The theoretical relationships in (2) indicate that SAVE and DR are mostly preferred because of their comprehensive estimation of $\mathcal{S}_{Y|\boldsymbol{X}}$, but it is not guaranteed that these two methods yield better estimation results in the finite sample case than the others, which have their own methodological strengths and weaknesses. Therefore, narrowing the field of five hybrid candidates based on these theoretical relations is not desirable. Instead, we study the sample behaviors of the five hybrid candidates for the following six numerical models:

Model 1: $Y|\mathbf{X} = X_1 + 0.5\varepsilon$;
Model 2: $Y|\mathbf{X} = X_1^2 + 0.5\varepsilon$;
Model 3: $Y|\mathbf{X} = X_1 + X_2^2 + 0.5\varepsilon$;
Model 4: $Y|\mathbf{X} = X_1 + X_1X_2 + 0.5\varepsilon$;
Model 5: $Y|\mathbf{X} = X_1 + X_1^2 + X_1X_2 + 0.5\varepsilon$;
Model 6: $Y|\mathbf{X} = X_1 + 0.5\exp(X_2)\varepsilon$.

For each model, 10-dimensional predictors $\mathbf{X} = (X_1, \ldots, X_{10})^{\mathrm{T}}$ with random error ε were independently generated from $N(0, 1)$. The sample sizes were 100 for all models, and each model was iterated 1000 times. The value of α varied over the set of $(0, 0.1, 0.2, \ldots, 0.8, 0.9, 1.0)$.

The central subspace is spanned by $(1, 0, 0, \ldots, 0)^{\mathrm{T}}$ for Models 1-2 and by the two columns of $((1, 0, 0, \ldots, 0), (0, 1, 0, \ldots, 0))^{\mathrm{T}}$ for Models 3-6. In Model 1, SIR and covk are known to have a clear advantage over SAVE, DR, and pHd. On the other hand, for Model 2, SAVE and pHd estimate $\mathcal{S}_{Y|\boldsymbol{X}}$ better than SIR, covk, and DR because of their intrinsic symmetric structure. Model 4 is a first-order interaction regression, while Models 2, 3, and 5 are second-order polynomial regressions. Moreover, Model 6 has a heteroscedasticity property. Models 3, 5 and 6 are considered in Ye and Weiss (2003) to bring the necessity of the hybrid approach for sufficient dimension reduction.

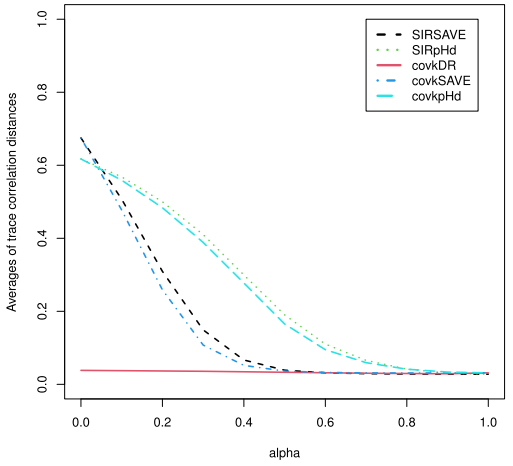
The above six simulated models do not cover all possible regression models. However, they have been widely used not only to teach linear regression but also to compare how well SDR methodologies estimate $\mathcal{S}_{Y|\boldsymbol{X}}$ in the literature (Li, 1991; Cook and Weisberg, 1991; Yin and Cook, 2002; Li and Wang, 2007; Yoo, 2009). So, the comparison of the five hybrid methods through the six artificial models would be quite representative for investigating the asymptotic behaviors of hybrid methods. As a summary measure of the numerical studies, the averages of the trace correlation distances between the true basis matrix of $\mathcal{S}_{Y|\boldsymbol{X}}$ and its estimate were computed and are reported in Fig. 1 under the true structural dimension. According to Fig. 1, SIR-SAVE and covk-SAVE methods clearly outperform SIR-pHd and covk-pHd, respectively. Henceforth, the two pHd-related hybrid methods are ruled out of consideration in our approach. There is no clear winner between SIR-SAVE and covk-SAVE. The SIR-SAVE, covk-DR and covk-SAVE estimate $\mathcal{S}_{Y|\boldsymbol{X}}$ fairly well with appropriate choices of α in most models, although covk-DR shows most stable estimation results than the other two.

The optimal value of α accomplishes the minimum average of u_D and its corresponding average for the six simulated models is reported in Table 2. According to Table 2, the two pHd-related hybrid methods often have the extreme values of α such as 0.0 and 1.0. On the other hand, none of SIR-SAVE, covk-DR, and covk-SAVE have such values. This implies that pHd would not have good cooperation with SIR and covk, and not tend to provide additional information to them. This confirms the validity of the elimination of SIR-pHd and covk-pHd from consideration. The theoretical relations between the remaining SIR-SAVE, covk-DR, and covk-SAVE are established by the following proposition.

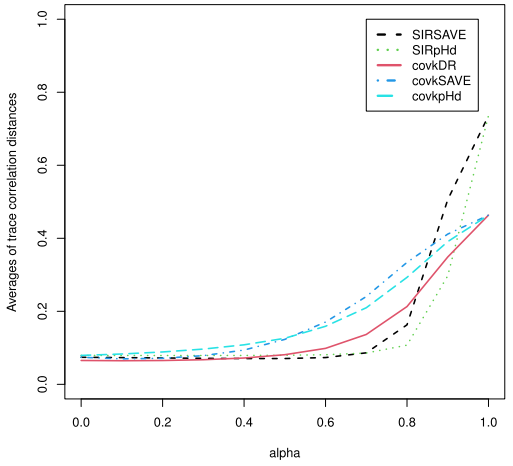
Proposition 1. *Relations between SIR-SAVE, covk-DR, and covk-SAVE are as follows.*

a. $\mathcal{S}(\alpha\mathbf{M}_{\mathrm{covk}} + (1-\alpha)\mathbf{M}_{\mathrm{DR}}) = \mathcal{S}(\alpha\mathbf{M}_{\mathrm{covk}} + (1-\alpha)\mathbf{M}_{\mathrm{SAVE}})$.
b. *For* $\alpha = 1$, $\mathcal{S}(\alpha\mathbf{M}_{\mathrm{covk}} + (1-\alpha)\mathbf{M}_{\mathrm{DR}}) = \mathcal{S}(\mathbf{M}_{\mathrm{covk}}) \subseteq \mathcal{S}(\alpha\mathbf{M}_{\mathrm{SIR}} + (1-\alpha)\mathbf{M}_{\mathrm{SAVE}}) = \mathcal{S}(\mathbf{M}_{\mathrm{SIR}})$.
c. *For* $0 \le \alpha < 1$, $\mathcal{S}(\alpha\mathbf{M}_{\mathrm{covk}} + (1-\alpha)\mathbf{M}_{\mathrm{DR}}) = \mathcal{S}(\alpha\mathbf{M}_{\mathrm{SIR}} + (1-\alpha)\mathbf{M}_{\mathrm{SAVE}}) = \mathcal{S}(\mathbf{M}_{\mathrm{SAVE}}) = \mathcal{S}(\mathbf{M}_{\mathrm{DR}})$.

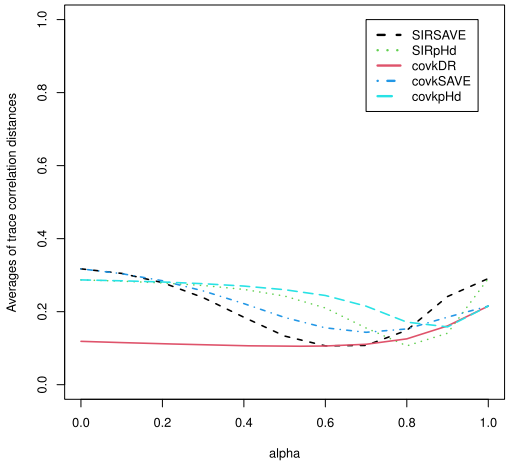
Proof. The proof can be found in the Appendix. □



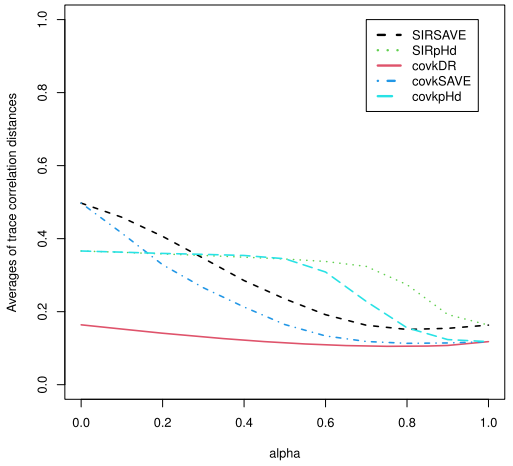
(a) Model 1



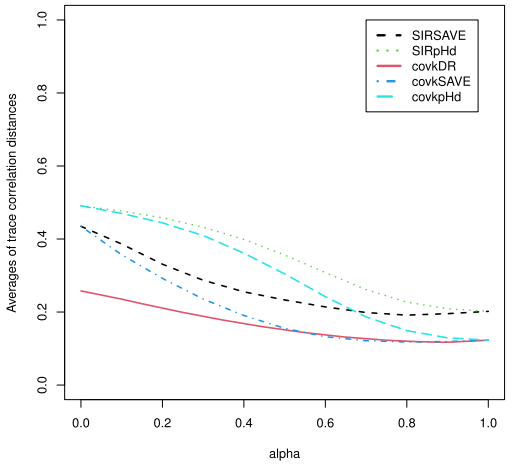
(b) Model 2



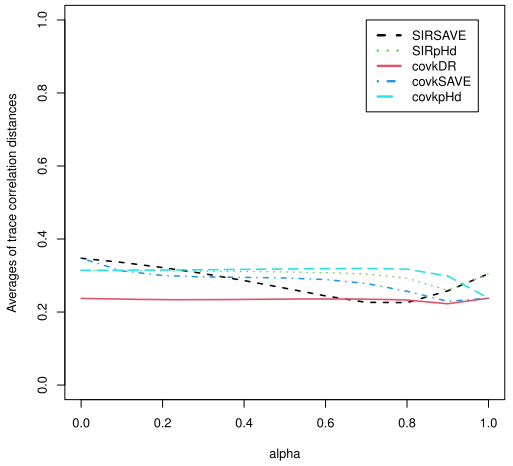
(c) Model 3



(d) Model 4



(e) Model 5



(f) Model 6

Fig. 1. Averages of trace correlation distance u_D for Models 1–6 for various values of α.

Table 3
Cross table of αs for SIR-SAVE and covk-DR.

		covk-DR	
		Good α (G)	Bad α (B)
SIR-SAVE	Good α (G)	G1	G2
	Bad α (B)	G3	G4

Based on these numerical studies and the theoretical relation, the five hybrid methodologies are reduced to SIR-SAVE, covk-DR and covk-SAVE, and this narrowing of focus does not sacrifice the benefits of the larger group of all five hybrid methods.

3.2. Cross-distance selection of α

For further discussion, we define $\hat{\mathbf{M}}_{\mathrm{SS}}(\alpha)=\alpha\hat{\mathbf{M}}_{\mathrm{SIR}}+(1-\alpha)\hat{\mathbf{M}}_{\mathrm{SAVE}}$, $\hat{\mathbf{M}}_{\mathrm{cD}}(\alpha)=\alpha\hat{\mathbf{M}}_{\mathrm{covk}}+(1-\alpha)\hat{\mathbf{M}}_{\mathrm{DR}}$, and $\hat{\mathbf{M}}_{\mathrm{cS}}(\alpha)=\alpha\hat{\mathbf{M}}_{\mathrm{covk}}+(1-\alpha)\hat{\mathbf{M}}_{\mathrm{SAVE}}$, and that $\hat{\boldsymbol{\eta}}_d^{\mathrm{SS}}(\alpha)$, $\hat{\boldsymbol{\eta}}_d^{\mathrm{cD}}(\alpha)$, and $\hat{\boldsymbol{\eta}}_d^{\mathrm{cS}}(\alpha)$ are the d-dimensional estimate of η from $\hat{\mathbf{M}}_{\mathrm{SS}}(\alpha)$, $\hat{\mathbf{M}}_{\mathrm{cD}}(\alpha)$, and $\hat{\mathbf{M}}_{\mathrm{cS}}(\alpha)$, respectively, where $\hat{\mathbf{M}}_{\bullet}(\alpha)$ is the sample kernel matrix constructed from each hybrid method with a weight α.

Since α determines the weights of the two SDR methods in a hybrid approach, the criterion of selecting α is essential. Here, we apply a grid search process. The basic and reasonable philosophy to select a good α is placed on finding α which the hybrid method with α can estimate $\mathcal{S}_{Y|\boldsymbol{X}}$ better than the other candidates for α in a grid. Under this philosophy, the following minimization is the key idea for the choices of α:

$$\underset{\alpha_i,\alpha_j}{\operatorname{argmin}}\, r_D^d(\alpha_i,\alpha_j)=||\hat{\boldsymbol{\eta}}_d^{\dagger}(\alpha_i)-\hat{\boldsymbol{\eta}}_d^{\ddagger}(\alpha_j)||^2,\ \ \alpha_i \text{ and } \alpha_j=0,0.1,\ldots,0.9,1.0,$$

where $r_D^d(\alpha_i,\alpha_j)=u_D(\hat{\mathbf{M}}_{\dagger}(\alpha_i),\hat{\mathbf{M}}_{\ddagger}(\alpha_j))$ with a structural dimension d and $\hat{\mathbf{M}}_{\dagger}(\alpha_i),\hat{\mathbf{M}}_{\ddagger}(\alpha_j)$ represent the kernel matrices of the two hybrid methods $\dagger$ and $\ddagger$, respectively, $||\bullet||$ represents the norm, and $\hat{\boldsymbol{\eta}}_d^{\dagger}(\alpha_i)$ and $\hat{\boldsymbol{\eta}}_d^{\ddagger}(\alpha_j)$ stand for the d-dimensional estimates of $\hat{\boldsymbol{\eta}}$ from the two hybrid methods.

We explain how the minimization would work with SIR-SAVE and covk-DR. The squared distance $||\hat{\boldsymbol{\eta}}_d^{\mathrm{SS}}(\alpha_i)-\hat{\boldsymbol{\eta}}_d^{\mathrm{cD}}(\alpha_j)||^2$ has the following expansion for the true basis $\boldsymbol{\eta}$:

$$||\hat{\boldsymbol{\eta}}_d^{\mathrm{SS}}(\alpha_i)-\hat{\boldsymbol{\eta}}_d^{\mathrm{cD}}(\alpha_j)||^2\geq||\hat{\boldsymbol{\eta}}_d^{\mathrm{SS}}(\alpha_i)-\boldsymbol{\eta}||^2+||\hat{\boldsymbol{\eta}}_d^{\mathrm{cD}}(\alpha_j)-\boldsymbol{\eta}||^2-2||\hat{\boldsymbol{\eta}}_d^{\mathrm{SS}}(\alpha_i)-\boldsymbol{\eta}||\times||\hat{\boldsymbol{\eta}}_d^{\mathrm{cD}}(\alpha_j)-\boldsymbol{\eta}||. \tag{3}$$

The squared distance $||\hat{\boldsymbol{\eta}}_d^{\mathrm{SS}}(\alpha_i)-\hat{\boldsymbol{\eta}}_d^{\mathrm{cD}}(\alpha_j)||^2$ in (3) is minimized not only by minimizing $||\hat{\boldsymbol{\eta}}_d^{\mathrm{SS}}(\alpha_i)-\boldsymbol{\eta}||^2$ and $||\hat{\boldsymbol{\eta}}_d^{\mathrm{cD}}(\alpha_j)-\boldsymbol{\eta}||^2$ but also by maximizing $||\hat{\boldsymbol{\eta}}_d^{\mathrm{SS}}(\alpha_i)-\boldsymbol{\eta}||\times||\hat{\boldsymbol{\eta}}_d^{\mathrm{cD}}(\alpha_j)-\boldsymbol{\eta}||$. The bootstrapping approach by Ye and Weiss (2003) is to select α to minimize $||\hat{\boldsymbol{\eta}}_d^{\mathrm{SS}}(\alpha_i)-\boldsymbol{\eta}||^2$ or $||\hat{\boldsymbol{\eta}}_d^{\mathrm{cD}}(\alpha_j)-\boldsymbol{\eta}||^2$ separately. This is the critical distinguishing feature of the proposed novel selection approach compared to the boostrapping of Ye and Weiss (2003).

Now, we classify α into the four candidate groups for SIR-SAVE and covk-DR, which is reported in Table 3. The group "Good α" in Table 3 means that α in the group estimates $\mathcal{S}_{Y|\boldsymbol{X}}$ well. For example, for Model 3, the values of 0.5, 0.6, and 0.7 would be in "Good α" for SIR-SAVE and those of 0.0, 0.1, …, 0.7 should be in "Good α" for covk-DR. Then, it is must be desirable that the minimizers of $||\hat{\boldsymbol{\eta}}_d^{\mathrm{SS}}(\alpha_i)-\hat{\boldsymbol{\eta}}_d^{\mathrm{cD}}(\alpha_j)||^2$ should be chosen in G1 of Table 3, but the minimization does not guarantee this because of the cross-product distance $||\hat{\boldsymbol{\eta}}_d^{\mathrm{SS}}(\alpha_i)-\boldsymbol{\eta}||\times||\hat{\boldsymbol{\eta}}_d^{\mathrm{cD}}(\alpha_j)-\boldsymbol{\eta}||$.

For G1 and G2, G1 is expected to be chosen over G2 by investigating the difference of $r_D^d(\mathrm{G},\mathrm{G})-r_D^d(\mathrm{G},\mathrm{B})$:

$$\begin{aligned}
r_D^d(\mathrm{G},\mathrm{G})-r_D^d(\mathrm{G},\mathrm{B}) &= ||\hat{\boldsymbol{\eta}}_d^{\mathrm{SS}}(\mathrm{G})-\boldsymbol{\eta}||^2+||\hat{\boldsymbol{\eta}}_d^{\mathrm{cD}}(\mathrm{G})-\boldsymbol{\eta}||^2-2\langle\hat{\boldsymbol{\eta}}_d^{\mathrm{SS}}(\mathrm{G})-\boldsymbol{\eta},\hat{\boldsymbol{\eta}}_d^{\mathrm{cD}}(\mathrm{G})-\boldsymbol{\eta}\rangle\\
&\quad-||\hat{\boldsymbol{\eta}}_d^{\mathrm{SS}}(\mathrm{G})-\boldsymbol{\eta}||^2-||\hat{\boldsymbol{\eta}}_d^{\mathrm{cD}}(\mathrm{B})-\boldsymbol{\eta}||^2+2\langle\hat{\boldsymbol{\eta}}_d^{\mathrm{SS}}(\mathrm{G})-\boldsymbol{\eta},\hat{\boldsymbol{\eta}}_d^{\mathrm{cD}}(\mathrm{B})-\boldsymbol{\eta}\rangle\\
&= ||\hat{\boldsymbol{\eta}}_d^{\mathrm{cD}}(\mathrm{G})-\boldsymbol{\eta}||^2-||\hat{\boldsymbol{\eta}}_d^{\mathrm{cD}}(\mathrm{B})-\boldsymbol{\eta}||^2\\
&\quad+2\langle\hat{\boldsymbol{\eta}}_d^{\mathrm{SS}}(\mathrm{G})-\boldsymbol{\eta},\hat{\boldsymbol{\eta}}_d^{\mathrm{cD}}(\mathrm{B})-\boldsymbol{\eta}-(\hat{\boldsymbol{\eta}}_d^{\mathrm{cD}}(\mathrm{G})-\boldsymbol{\eta})\rangle\\
&\leq ||\hat{\boldsymbol{\eta}}_d^{\mathrm{cD}}(\mathrm{G})-\boldsymbol{\eta}||^2-||\hat{\boldsymbol{\eta}}_d^{\mathrm{cD}}(\mathrm{B})-\boldsymbol{\eta}||^2\\
&\quad+2||\hat{\boldsymbol{\eta}}_d^{\mathrm{SS}}(\mathrm{G})-\boldsymbol{\eta}||\times||\hat{\boldsymbol{\eta}}_d^{\mathrm{cD}}(\mathrm{B})-\boldsymbol{\eta}-(\hat{\boldsymbol{\eta}}_d^{\mathrm{cD}}(\mathrm{G})-\boldsymbol{\eta})||\\
&\leq ||\hat{\boldsymbol{\eta}}_d^{\mathrm{cD}}(\mathrm{G})-\boldsymbol{\eta}||^2-||\hat{\boldsymbol{\eta}}_d^{\mathrm{cD}}(\mathrm{B})-\boldsymbol{\eta}||^2\\
&\quad+2||\hat{\boldsymbol{\eta}}_d^{\mathrm{SS}}(\mathrm{G})-\boldsymbol{\eta}||\times(||\hat{\boldsymbol{\eta}}_d^{\mathrm{cD}}(\mathrm{B})-\boldsymbol{\eta}||+||(\hat{\boldsymbol{\eta}}_d^{\mathrm{cD}}(\mathrm{G})-\boldsymbol{\eta})||).
\end{aligned} \tag{4}$$

In the last inequality of (4), the first part $||\hat{\boldsymbol{\eta}}_d^{\mathrm{cD}}(\mathrm{G})-\boldsymbol{\eta}||^2-||\hat{\boldsymbol{\eta}}_d^{\mathrm{cD}}(\mathrm{B})-\boldsymbol{\eta}||^2$ should be negative, because $\hat{\boldsymbol{\eta}}_d^{\mathrm{cD}}(\mathrm{G})$ estimates $\mathcal{S}_{Y|\boldsymbol{X}}$ better than $\hat{\boldsymbol{\eta}}_d^{\mathrm{cD}}(\mathrm{B})$. Since both $\hat{\boldsymbol{\eta}}_d^{\mathrm{cD}}(\mathrm{G})$ and $\hat{\boldsymbol{\eta}}_d^{\mathrm{SS}}(\mathrm{G})$ estimate $\mathcal{S}_{Y|\boldsymbol{X}}$ well, it would be expected that $\mathcal{S}(\hat{\boldsymbol{\eta}}_d^{\mathrm{cD}}(\mathrm{G}))\approx\mathcal{S}(\hat{\boldsymbol{\eta}}_d^{\mathrm{SS}}(\mathrm{G}))$. If so, we have $||\hat{\boldsymbol{\eta}}_d^{\mathrm{cD}}(\mathrm{G})-\boldsymbol{\eta}||\approx||\hat{\boldsymbol{\eta}}_d^{\mathrm{SS}}(\mathrm{G})-\boldsymbol{\eta}||\approx 0$. Therefore, $r_D^d(\mathrm{G},\mathrm{G})-r_D^d(\mathrm{G},\mathrm{B})$ would be negative, the values in α in G1 are normally selected over those in G2. The same rationale can be applied to the comparison between G1 and G3.

A rough condition for selecting α in G1 over G2 and G3 is that $\mathcal{S}(\hat{\boldsymbol{\eta}}_d^{\dagger}(\mathrm{G})) \approx \mathcal{S}(\hat{\boldsymbol{\eta}}_d^{\ddagger}(\mathrm{G}))$. This indicates that both hybrid methods to compute $r_D^d(\alpha_i, \alpha_j)$ need to yield homogeneous estimation results. To satisfy this, we consider two cases of $r_D^d(\alpha_i, \alpha_j)$ between SIR-SAVE and covk-DR and between covk-DR and covk-SAVE. It is well known that SIR and covk yield similar estimation results (Yoo, 2018), and DR has information about both SIR and SAVE. The cross-distance between SIR-SAVE and covk-SAVE is ruled out, because large differences in α for SAVE can lead to the outcome that $\mathcal{S}(\hat{\boldsymbol{\eta}}_d^{\mathrm{SS}}(\mathrm{G}))$ is distant from $\mathcal{S}(\hat{\boldsymbol{\eta}}_d^{\mathrm{cS}}(\mathrm{G}))$. Other than this, another poor selection of α can arise in G4. The values of α in G4 can be minimizers due to methodological similarity caused by extreme values of α. For example, the choice of $\alpha = 1$ reduces SIR-SAVE and covk-DR to SIR and covk, respectively, and their estimates should be similar as discussed above. So, $r_D^d(\alpha_i, \alpha_j)$ may possibly be minimized at $\alpha_i = 1$ and $\alpha_j = 1$ regardless of how well these methods estimate $\mathcal{S}_{Y|\boldsymbol{X}}$. To address this issue, we remove the two extreme values $\alpha = 0$ and $\alpha = 1$. According to Table 2, the three hybrid methods do not include either $\alpha = 0$ or $\alpha = 1$ for the best values of α. Furthermore, for covk-DR and covk-SAVE, $r_D^d(\alpha_i, \alpha_j)$ will be smaller for larger values of α_i and α_j due to the same reason. We will therefore consider $\alpha = 0.1, \ldots, 0.8$, excluding 0.9, for covk-DR and covk-SAVE. In Models 1, 5, and 6, selecting $\alpha = 0.9$ provides the best estimation accuracy for the covk-DR method, but the differences in accuracy from $\alpha = 0.8$ are 0.0004, 0.0102, and 0.0031, respectively. The covk-SAVE has the best α at 0.9 in Model 5, and the difference with $\alpha = 0.8$ is 0.028. The elimination of $\alpha = 0.9$ does not affect the estimation accuracy and will not diminish the novelty and advantage of the hybrid dimension reduction.

Note that the distance $r_D^d(\alpha_i, \alpha_j)$ depends on α_i and α_j as well as the structural dimension d. Therefore, the distance should be computed for $d = 1, 2, \ldots, m$, where m is user-selected. Since d is unknown in practice, it is necessary to pick the minimizers for each of $d = 1, 2, \ldots, m$.

In Model 3, which Ye and Weiss (2003) brought the necessity of hybrid dimension reduction, the best values of α for SIR-SAVE and covk-DR are 0.6 and 0.4, respectively, according to Table 2. If we use the proposed selection approach for one simulated-data, SIR-SAVE and covk-DR select 0.6 and 0.5 as their proper α. For SIR-SAVE, the CDS determines the optimal value correctly, and, for covk-DR, the difference between $\alpha = 0.4$ and $\alpha = 0.5$ is 0.005, which is essentially ignorable. This result is attractive and persuasive.

Based on the reasoning discussed above, a two-step selection algorithm is proposed as follows. The first step is initiated to determine one of two pairs of (SIR-SAVE, covk-DR) and (covk-SAVE, covk-DR). The essential method in the two pairs is covk, and hence it is necessary to evaluate how well covk represents data. For this investigation, the so-called basis adaption selection algorithm (Yoo, 2018; BAS) is implemented for covk, SIR, SAVE and pHd, instead of DR. Since DR is related to both SIR and SAVE, its inclusion in BAS will prevent a fair evaluation of covk in the data. If covk is selected by BAS, then $r_D^d(\alpha_i, \alpha_j)$ between covk-SAVE and covk-DR are computed to search for good α_i and α_j. If not, $r_D^d(\alpha_i, \alpha_j)$ between SIR-SAVE and covk-DR are minimized.

The second step is to select one of the two hybrid methods determined in the first step. For (covk-SAVE, covk-DR), covk-DR with its chosen α will be the representative hybrid method, because various numerical studies have shown that covk-DR estimates $\mathcal{S}_{Y|\boldsymbol{X}}$ better than covk-SAVE. In the case of (SIR-SAVE, covk-DR), BAS is conducted for SIR, SAVE, covk and DR. If either SIR or SAVE is recommended, SIR-SAVE with its suggested α will be chosen, and covk-DR will be, otherwise.

Since this proposed selection approach is based on computing the cross-distances between SIR-SAVE, covk-DR, and covk-SAVE, it will be called *cross-distance selection* (CDS) algorithm. The formal cross-distance selection algorithm is as follows.

Algorithm Cross-distance selection algorithm

1. Fix the maximum value $d_{\max}$ of d, which is less than p. Here, we set $d_{\max}$ to 4. Since d turns out to be equal to one or two in many SDR application, $d_{\max} = 4$ should suffice in practice.
2. Run the BAS algorithm with SIR, SAVE, covk and pHd. If covk is suggested, $r_D^d(\alpha_i, \alpha_j)$ between covk-SAVE and covk-DR is minimized over the grids for α_i, α_j, and d. Then, covk-DR with its suggested α and d is fitted, and the dimension reduction is terminated. There will be no further step. Otherwise, $r_D^d(\alpha_i, \alpha_j)$ between SIR-SAVE and covk-DR is minimized.
3. If $r_D^d(\alpha_i, \alpha_j)$ between SIR-SAVE and covk-DR is minimized in the previous step, run BAS with SIR, SAVE, covk and DR. If SIR and SAVE are recommended, SIR-SAVE with its suggested α and d is fitted. Otherwise, covk-DR with its suggested α and d is fitted.

If bootstrapping is used for the three hybrid methods with B, the total number of fittings is equal to $3 \times (B \times 9 \times d_{\max})$. If using a setting with $d_{\max} = 4$ and $B = 500$, this number turns out to be 43,200. By contrast, the total number of fitting steps in CDS is equal to $\binom{4}{2} \times d_{\max} + 2 \times 9 \times d_{\max} + \binom{4}{2} \times d_{\max}$. The first and last $\binom{4}{2} \times d_{\max}$ are for two applications of BAS. If $d_{\max} = 4$, it is equal to 120. With the setting, the bootstrap approach requires to run, at least, 360 times more fitting procedures than the proposed CDS. For Model 1 given in Section 3.1 on a desktop computer with Windows 10 64-bit operating system and Intel(R) Core(TM) m5-6Y54 CPU @1.10GHz 1.51GHz and 8GB RAM, the running times of the proposed CDS and the existing bootstrapping selection among SIR-SAVE, covk-SAVE, and covk-DR are 0.45 and 246.98 seconds, respectively. This confirms that the CDS has very strong advantage over the bootstrap method in computational efficiency.

4. Numerical studies and real data analysis

4.1. Numerical studies

To demonstrate the competitiveness of the cross-distance selection algorithm, we used the simulated models in Section 3.1 with the same settings. The proposed CDS is compared with SIR-SAVE, covk-DR, and covk-SAVE using the best α in Table 2. The boxplots of u_D computed for each model are presented in Fig. 2 with labels on the horizontal axis are as follows. "SIRSAVE0#", "covkDR0•", and "covk-SAVE0†" stand for SIR-SAVE, covk-DR, and covk-SAVE with $\alpha = 0.\#$, $\alpha = 0.\bullet$, and $\alpha = 0.\dagger$, respectively, from Table 2. "CDS" represents the fit by the hybrid SDR method with α chosen by the CDS algorithm. Lastly, "Boot" stands for the best fit by SIR-SAVE, covk-SAVE, and covk-DR selected by the bootstrapping criterion with 500 bootstrapping in Ye and Weiss (2003). In Fig. 2, the red lines represent the medians of u_Ds.

According to Fig. 2, it is observed that, for Models 1, 2, and 4, there is no significant difference between CDS, bootstrapping selection, and the best versions of SIR-SAVE, covk-DR, and covk-SAVE from Table 2. For Models 3 and 5, CDS clearly dominates the bootstrapping selection, while bootstrapping is the best method for Model 6. These numerical studies along with computing time considerations confirm the potential advantage of the cross-distance selection algorithm over the bootstrapping selection.

4.2. Real data analysis: abalone data

We apply CDS to data describing abalone (Triastcyn and Faltings, 2020). The age of abalone is determined by cutting the shell through the cone, staining it, and counting how many rings are observed through a microscope. This task requires a good deal of time. The age in years is the number of rings plus 1.5. To predict the age of abalone, their physical measurements, which are relatively easier to obtain, are used. For this, the following seven characteristics were measured and recorded: longest shell measurement (length, mm) shell measurement perpendicular to the longest shell measurement (diameter, mm) meat in shell (height, mm) whole abalone weight (whole weight, grams) meat weight (shucked weight, grams) gut weight after bleeding (viscera weight, grams), and shell weight after being dried (shell weight, grams). These measurements are used as the predictors. The total number of observations is 4177. The full data and a more detailed description can be found at

https://archive.ics.uci.edu/ml/machine-learning-databases/abalone/.

In the dataset, the 1418th and 2052th observations were suspected to be outliers based on a scatterplot matrix of predictors and responses, and they were removed in the analysis. Since $\sqrt{R^2}$ computed from a regression of the whole weight given the other weights was 0.995, the whole weight was eliminated from the analysis to avoid multi-collinearity, Next, to satisfy the conditions required in SIR, SAVE, covk, and DR, the three remaining weights were transformed to a square-root scale. With these settings of the data, the proposed CDS algorithm and the existing Ye and Weiss (2003) approach with 500 bootstrapping were applied to data with 10 slices for SIR and 5 slices for SAVE and DR.

To begin with, the computing time of the CDS and bootstrapping approaches are 3.84 seconds and 2865.32 seconds (about 47 minutes), respectively, which shows a clear computational advantage of our approach over Ye and Weiss (2003). The CDS algorithm suggests covk-DR with $\alpha = 0.9$, and covk-DR heavily depends on covk. The Ye and Weiss (2003) approach recommends SIR-SAVE with $\alpha = 1$, which is simply SIR. This is because the data have linear a relationship between the response and the predictors. To investigate the estimates of $\mathcal{S}_{Y|\boldsymbol{X}}$ between the two, the trace correlation distances are computed for $d = 1, 2$, and 3, which are 0.259, 0.008, and 0.042, respectively. According to Yoo (2018), SIR and covk often yield quite close estimates of $\mathcal{S}_{Y|\boldsymbol{X}}$, the structural dimension would be determined to be equal to 2, because the distances are the shortest under $d = 2$. Then, supposing that $d = 2$, the two estimates span essentially the same space, and there is no difference in the dimension reduction between the CDS and Ye and Weiss (2003) approaches. However, CDS has a clear benefit in computational efficiency over the Ye and Weiss (2003) approach. This data analysis confirms a potential advantage of the proposed CDS algorithm over the bootstrapping method.

5. Discussion

Hybrid sufficient dimension reduction (SDR) is a weighted sum of kernel matrices of two different SDR methods, which are proposed by Ye and Weiss (2003). A bootstrap approach was developed in Ye and Weiss (2003) to select not only a proper value of a weight required to combine two methods but also one among many candidate hybrid methods. Since bootstrapping is computationally intensive and time-consuming, the hybrid reduction has not been popular even though it is more accurate than the usual single SDR methods in the estimation of the central subspace.

Before developing a new algorithm, a handful of hybrid SDR methods were reduced to three good candidates throughout various representative simulated models. The survived hybrid methods are the following three combinations of SDR methods: sliced inverse regression (Li (1991)) and sliced average variance estimation (Cook and Weisberg (1991)); covariance method (Yin and Cook, 2002) and directional regression (Li and Wang, 2007); covariance method and sliced average variance estimation. For the three chosen hybrid methods, a CDS algorithm was proposed. The proposed selection algorithm

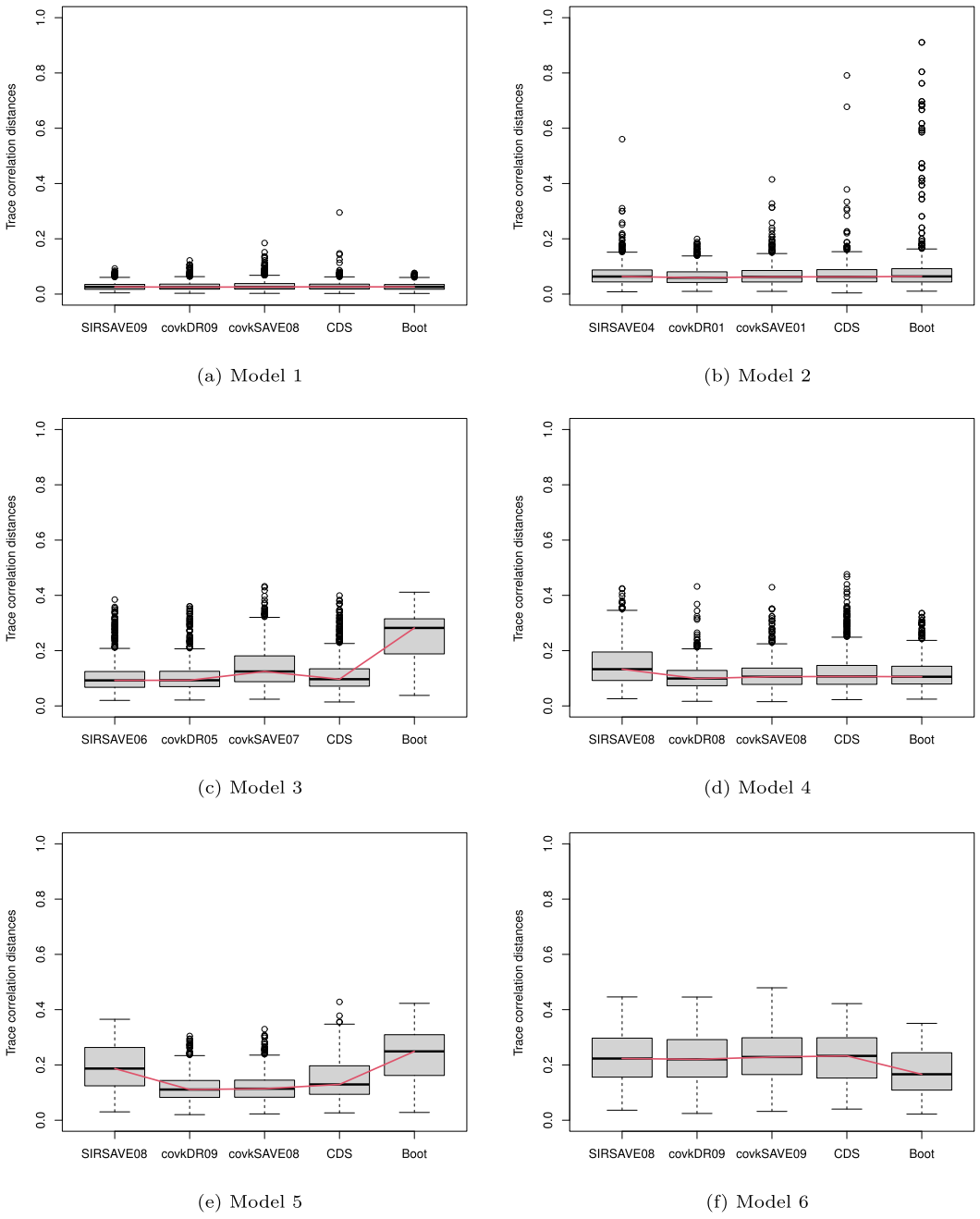


Fig. 2. Averages of trace correlation distance u_D for Models 1–6 for SIR-SAVE, covk-DR, covk-SAVE, CDS, and bootstrapping (Boot); Red line represents median of u_Ds. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

is data-driven like bootstrapping and has good reasoning about its performance. According to the numerical studies, the proposed algorithm often outperforms the bootstrap method (Ye and Weiss, 2003) in estimation accuracy and guarantees computational efficiency at the same time.

For future research, combining three or four SDR methods can be considered. This would possibly stress more methodological commonalities than methodological differences. Embracing the differences is why hybrid SDR methods estimate the central subspace better than the individual method. Therefore, the combination of three or four SDR methods may not be preferred to that of the two. Also, it will take much time to find proper weights, because all possible combinations of the weights are excessively numerous.

Acknowledgement

We would like to express our sincere appreciation to the Editor, the Associate Editor, and the two referees for their very informative reports, which helped us improve the manuscript.

For Jae Keun Yoo, this work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Korean Ministry of Education (NRF-2021R1F1A1059844). For Kyongwon Kim, this work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1F1A1046976). This research was also supported by the MSIT (Ministry of Science, ICT), Korea, under the High-Potential Individuals Global Training Program (RS-2022-00154879) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).

Appendix A. Proof of Proposition 1

To prove Proposition 1, we prove the following Lemma 5.1.

Lemma 5.1. *Suppose that* $\mathbf{A}$ *and* $\mathbf{B}$ *are semi-positive definite with* $\mathcal{S}(\mathbf{A}) \subseteq \mathcal{S}(\mathbf{B})$. *Then, for* $0 \leq \alpha < 1$, $\mathcal{S}(\mathbf{B}) = \mathcal{S}(\alpha\mathbf{A} + (1-\alpha)\mathbf{B})$.

Proof. For $\alpha = 0$, the equivalence trivially holds. For $0 < \alpha < 1$, Lemma 4 in Ye and Weiss (2003) guarantees that $\mathcal{S}(\mathbf{B}) = \mathcal{S}((1-\alpha)\mathbf{B}) \subseteq \mathcal{S}(\alpha\mathbf{A} + (1-\alpha)\mathbf{B})$. Suppose that $\nu \neq 0$ is orthogonal to $\mathcal{S}(\mathbf{B})$, that is $\nu^{\mathrm{T}}\mathbf{B} = 0$. The condition that $\mathcal{S}(\mathbf{A}) \subseteq \mathcal{S}(\mathbf{B})$ directly implies $\nu^{\mathrm{T}}\mathbf{A} = 0$. Consequently, $\nu^{\mathrm{T}}(\alpha\mathbf{A} + (1-\alpha)\mathbf{B}) = 0$ and we have $\mathcal{S}(\alpha\mathbf{A} + (1-\alpha)\mathbf{B}) \subseteq \mathcal{S}(\mathbf{B})$. Finally, we have $\mathcal{S}(\mathbf{B}) = \mathcal{S}(\alpha\mathbf{A} + (1-\alpha)\mathbf{B})$, and this completes the proof. □

Since $\mathcal{S}(\mathbf{M}_{\mathrm{covk}}) \subseteq \mathcal{S}(\mathbf{M}_{\mathrm{SIR}}) \subseteq \mathcal{S}(\mathbf{M}_{\mathrm{SAVE}}) = \mathcal{S}(\mathbf{M}_{\mathrm{DR}})$ in (2), Proposition 5.1 directly indicates that $\mathcal{S}(\alpha\mathbf{M}_{\mathrm{covk}} + (1-\alpha)\mathbf{M}_{\mathrm{SAVE}}) = \mathcal{S}(\mathbf{M}_{\mathrm{SAVE}}) = \mathcal{S}(\mathbf{M}_{\mathrm{DR}}) = \mathcal{S}(\alpha\mathbf{M}_{\mathrm{covk}} + (1-\alpha)\mathbf{M}_{\mathrm{DR}})$ for $0 \leq \alpha < 1$, and part (a) trivially holds for $\alpha = 1$. This completes proof of part (a). Proof of part (b) is straightforward. For part (c), Proposition 5.1 directly implies that $\mathcal{S}(\alpha\mathbf{M}_{\mathrm{SIR}} + (1-\alpha)\mathbf{M}_{\mathrm{SAVE}}) = \mathcal{S}(\mathbf{M}_{\mathrm{SAVE}}) = \mathcal{S}(\mathbf{M}_{\mathrm{DR}}) = \mathcal{S}(\alpha\mathbf{M}_{\mathrm{covk}} + (1-\alpha)\mathbf{M}_{\mathrm{DR}})$. This completes the proof of part (c). □

References

Cook, R.D., 1998a. Regression Graphics. John Wiley & Sons, New York.
Cook, R.D., 1998b. Principal Hessian directions revisited. J. Am. Stat. Assoc. 93 (441), 84–94.
Cook, R.D., Weisberg, S., 1991. Discussion of sliced inverse regression for dimension reduction. J. Am. Stat. Assoc. 86 (414), 328–332.
Hooper, J.W., 1959. Simultaneous equations and canonical correlation theory. Econometrica: J. Econometric Soc., 245–256.
Hotelling, H., 1936. Relations between two sets of variates. Biometrika 28 (3/4), 321–377.
Li, B., Wang, S., 2007. On directional regression for dimension reduction. J. Am. Stat. Assoc. 102 (479), 997–1008.
Li, K.-C., 1991. Sliced inverse regression for dimension reduction. J. Am. Stat. Assoc. 86 (414), 316–327.
Li, K.-C., 1992. On principal Hessian directions for data visualization and dimension reduction: another application of Stein's lemma. J. Am. Stat. Assoc. 87 (420), 1025–1039.
Triastcyn, A., Faltings, B., 2020. Bayesian differential privacy for machine learning. In: International Conference on Machine Learning, PMLR, pp. 9583–9592.
Ye, Z., Weiss, R.E., 2003. Using the bootstrap to select one of a new class of dimension reduction methods. J. Am. Stat. Assoc. 98 (464), 968–979.
Yin, X., Cook, R.D., 2002. Dimension reduction for the conditional kth moment in regression. J. R. Stat. Soc., Ser. B, Stat. Methodol. 64 (2), 159–175.
Yin, X., Li, B., Cook, R.D., 2008. Successive direction extraction for estimating the central subspace in a multiple-index regression. J. Multivar. Anal. 99 (8), 1733–1757.
Yoo, J.K., 2009. Partial moment-based sufficient dimension reduction. Stat. Probab. Lett. 79 (4), 450–456.
Yoo, J.K., 2018. Basis-adaptive selection algorithm in dr-package. R J. 10 (2), 124.