Communications for Statistical Applications and Methods
2023, Vol. 30, No. 2, 179–189

https://doi.org/10.29220/CSAM.2023.30.2.179
Print ISSN 2287-7843 / Online ISSN 2383-4757

Two variations of cross-distance selection algorithm in hybrid sufficient dimension reduction

Jae Keun Yoo[1,a]

[a]Department of Statistics, Ewha Womans University, Korea

Abstract

Hybrid sufficient dimension reduction (SDR) methods to a weighted mean of kernel matrices of two different SDR methods by Ye and Weiss (2003) require heavy computation and time consumption due to bootstrapping. To avoid this, Park *et al.* (2022) recently develop the so-called cross-distance selection (CDS) algorithm. In this paper, two variations of the original CDS algorithm are proposed depending on how well and equally the covk-SAVE is treated in the selection procedure. In one variation, which is called the larger CDS algorithm, the covk-SAVE is equally and fairly utilized with the other two candiates of SIR-SAVE and covk-DR. But, for the final selection, a random selection should be necessary. On the other hand, SIR-SAVE and covk-DR are utilized with completely ruling covk-SAVE out, which is called the smaller CDS algorithm. Numerical studies confirm that the original CDS algorithm is better than or compete quite well to the two proposed variations. A real data example is presented to compare and interpret the decisions by the three CDS algorithms in practice.

Keywords: basis-adaptive selection, cross-distance selection, hybrid dimension reduction, sufficient dimension reduction, trace correlation

1. Introduction

Sufficient dimension reduction (SDR) for a regression of $Y \in \mathbb{R}^1|\mathbf{X} \in \mathbb{R}^p$ pursues the dimension reduction of predictors without losing information on the conditional distribution of $Y|\mathbf{X}$. Its primary interest is to replace the p-dimensional predictors $\mathbf{X}$ with a lower-dimensional projection $\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}$, which has the following equivalence:

$$Y \perp\!\!\!\perp \mathbf{X} \,|\, \boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}\,, \tag{1.1}$$

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

The d-dimensional column space of $\boldsymbol{\eta}$ to satisfy (1.1) is called a dimension reduction subspace. The main goal of SDR is to restore the intersection of all possible dimension reduction subspaces, which is called the *central subspace* $\mathcal{S}_{Y|\mathbf{X}}$. If $\mathcal{S}_{Y|\mathbf{X}}$ exists, its construction guarantees that it is unique and minimal. Readers are recommended to see Cook (1998a) for comprehensive discussion for the conditions to guarantee the existence of $\mathcal{S}_{Y|\mathbf{X}}$. Hereafter, $\boldsymbol{\eta}$ and d are denoted as an orthonormal basis andthe structural dimension of $\mathcal{S}_{Y|\mathbf{X}}$.

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).

[1] Corresponding Author: Department of Statistics, Ewha Womans University, 11-1 Daehyun-Dong Seodaemun-Gu, Seoul 03760, Korea. E-mail: peter.yoo@ewha.ac.kr

Published 31 March 2023 / journal homepage: http://csam.or.kr

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

Methods	Kernel matrices
Sliced inverse regression (SIR; Li, 1991)	$\mathbf{M}_{\text{SIR}} = \text{cov}\{E(\mathbf{Z} \mid Y)\}$
Sliced average variance estimation (SAVE; Cook and Weisberg, 1991)	$\mathbf{M}_{\text{SAVE}} = E\{\mathbf{I}_p - \text{cov}(\mathbf{Z} \mid Y)\}^2$
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 $\mathbf{K}_q = \{\text{cov}(\mathbf{Z}, W), \text{cov}(\mathbf{Z}, W^2), \ldots, \text{cov}(\mathbf{Z}, W^q)\}$
Directional regression (DR; Li and Wang, 2007)	$\mathbf{M}_{\text{DR}} = E\{E(\mathbf{Z}\mathbf{Z}^{\text{T}} \mid Y)\}^2 + 2E\{E(\mathbf{Z} \mid Y)E(\mathbf{Z}^{\text{T}} \mid Y)\}^2$ $+2E\{E(\mathbf{Z}^{\text{T}} \mid Y)E(\mathbf{Z} \mid Y)E(\mathbf{Z} \mid Y)E(\mathbf{Z}^{\text{T}} \mid Y)\} - 2\mathbf{I}_p$.

Popular SDR methods among many should be sliced inverse regression (SIR) (Li, 1991), sliced average variance estimation (SAVE) (Cook and Weisberg, 1991), principal hessian directions (Li, 1992; Cook, 1998b), covariance method (covk) (Yin and Cook, 2002), and directional regression (DR) (Li and Wang, 2007). In addtion Ye and Weiss (2003) propose a new class of dimension reduction methods to combine two SDR methods. For example, let $\mathbf{M}_{\text{SIR}}$ and $\mathbf{M}_{\text{SAVE}}$ denote kernel matrices constrcuted by SIR and SAVE to estimate $\mathcal{S}_{Y|\mathbf{X}}$, respectively. Then, their weighted mean of $\alpha\mathbf{M}_{\text{SIR}} + (1 - \alpha)\mathbf{M}_{\text{SAVE}}$ also can estimate $\mathcal{S}_{Y|\mathbf{X}}$ for $0 \leq \alpha \leq 1$. In Park *et al.* (2022), this type of sufficient dimension reduction method is called *hybrid suffcient dimension reduciton*. It is essential to select one hybrid SDR method among many candidates along with its good value of α, and Ye and Weiss (2003) developed a bootsrap approach by computing the average distances between the original-sample estimates and the bootstrap-sample estimates of $\mathcal{S}_{Y|\mathbf{X}}$ for various possible values of α.

However, the bootstrap approach requires heavy computation and much time-consumption to choose one hybrid SDR method along with a good value of α. To overcome this, recently, Park *et al.* (2022) propose the so-called *cross-distance selection* (CDS) algorithm with choosing a proper hybrid SDR method and its good α simultaneously. This CDS algorithm will be explained in detail in later section.

The main purpose of the paper is to propose two variations of the original CDS algorithm by Park *et al.* (2022), depending on how a combination of covk and SAVE is treated in the selection procedure. The two variations are to choose one hybrid SDR method from larger or smaller sets of pairs of hybrid SDR methods. This deserves to be investigated, because the original algorithm can be potentially improved in both sides of the accuracy in the estimation of $\mathcal{S}_{Y|\mathbf{X}}$ and the fitting time.

The organization of the paper is as follows. In Section 2, the hybrid dimension reduction and the cross-distance selection algorithm are introduced. Section 3 is devoted to develop its two additional variations. Numerical studies and real data application are presented in Section 4. We summarize our work in Section 5.

For notational conveniences, it is defined that $\mathbf{\Sigma} = \text{cov}(\mathbf{X})$ and $\mathbf{Z} = \mathbf{\Sigma}^{-1/2}(\mathbf{X} - E(\mathbf{X}))$. And, a notation $\mathcal{S}(\mathbf{B})$ is defined as a subspace spanned by the columns of $\mathbf{B} \in \mathbb{R}^{p\times q}$.

2. Cross-distance selection in hybrid suffcient dimension reduction

2.1. Hybrid sufficient dmension reduction

A hybrid sufficient dimension reduction method is a weighted mean of the two different kernel matrices constructed by the four popular SDR methodologies given in Table 1.

According to Park *et al.* (2022), the underlying philosophy of hybriding two SDR methodologies is to estimate $\mathcal{S}_{Y|\mathbf{X}}$ better by overcoming weakness for each one to have. The SIR and covk are not

combined, because their asymptotic estimation behaviors are known to be similar according to Yin and Cook (2002) and Yoo (2009). Since DR has partial information of both SIR and SAVE as Table 1 indicated, DR is combined neither SIR nor SAVE and is combined with covk alone. Finally, Park *et al.* (2022) suggest the next three hybrid candidates:

$$(1)\ \alpha\mathbf{M}_{\text{SIR}} + (1-\alpha)\,\mathbf{M}_{\text{SAVE}};\ \ (2)\ \alpha\mathbf{M}_{\text{covk}} + (1-\alpha)\,\mathbf{M}_{\text{DR}};\ \ (3)\ \alpha\mathbf{M}_{\text{covk}} + (1-\alpha)\,\mathbf{M}_{\text{SAVE}}.$$

For each hybrid candidate, if α is equal to 0 or 1, then it is reduced to one of the single kernels of SIR, SAVE, covk and DR. In this hybriding dimension reduction, the choice of one of the three methods and the determination of a proper value of α are essential. The selection of the hybrid methods clearly depend on the choice of α. Also, the determination of α will be changed according to what method is finally selected. This indicates that the two selections must be done simultaneously, not one after another.

Ye and Weiss (2003) propose a bootstrap approach to the simultaneous selection, changing $\alpha = 0, 0.1, \ldots, 0.8, 0.9, 1$. Distances between the original and bootstrap sample kernels are computed, and choose the method and α to give the smallest average distance. This is straightforward and easy to implement in practice, but it has the critical deficit of time-consuming in computational efficiency.

2.2. The original cross-distance selection algorithm

Park *et al.* (2022) propose an algorithm to select one of the hybrid methods and a proper value of α simultaneously with avoiding heavy computing time and keeping the competitive accuracy in the estimation of $\mathcal{S}_{Y|\mathbf{X}}$, comparing the bootstrapping by Ye and Weiss (2003). A basic idea of the selection algorithm by Park *et al.* (2022) is to constrain the two hybrid SDR methods among the three along with their own good values for αs. For this, the following quantity is computed:

$$r_D^d\left(\alpha_i, \alpha_j\right) = 1 - \sqrt{\frac{1}{d}\text{trace}\left[\hat{\boldsymbol{\eta}}_d^{\ddagger}\left(\alpha_j\right)^{\text{T}}\left(\hat{\boldsymbol{\eta}}_d^{\dagger}\left(\alpha_i\right)\hat{\boldsymbol{\eta}}_d^{\dagger}(\alpha_i)^{\text{T}}\right)\hat{\boldsymbol{\eta}}_d^{\ddagger}\left(\alpha_j\right)\right]},$$

where $\hat{\boldsymbol{\eta}}_d^{\dagger}(\alpha_i)$ and $\hat{\boldsymbol{\eta}}_d^{\ddagger}(\alpha_j)$ stand for the d-dimensional estimates of $\hat{\boldsymbol{\eta}}$ from the two hybrid methods and $\alpha_i, \alpha_j = 0,\ 0.1,\ldots,\ 0.9,\ 1.0$ A smaller value of $r_D^d(\alpha_i, \alpha_j)$ means that $\mathcal{S}(\hat{\boldsymbol{\eta}}_d^{\dagger}(\alpha_i))$ and $\mathcal{S}(\hat{\boldsymbol{\eta}}_d^{\ddagger}(\alpha_j))$ gets closer. If $r_D^d(\alpha_i, \alpha_j) = 0$, the two subspace of $\mathcal{S}(\hat{\boldsymbol{\eta}}_d^{\dagger}(\alpha_i))$ and $\mathcal{S}(\hat{\boldsymbol{\eta}}_d^{\ddagger}(\alpha_j))$ are the same.

The *original cross-distance selection* algorithm proposed by Part *et al.* (2022) is as follows:

Algorithm 1 : Original cross-distance selection

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 principal hessian direction (Li, 1992). 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.

In the algorithm, the BAS stands for a basis adaption selection algorithm by Yoo (2018). It recommends one SDR method mostly adapted to a data among the four candidate SDR methods. In the

original CDS algorithm, the BAS for SIR, SAVE, covk and principal hessian direction (Li, 1992) is implemented to consider either (SIR-SAVE & covk-DR) or (covk-SAVE & covk-DR). This possibly reduces unnecessary distance computing.

For instance, suppose that the BAS for SIR, SAVE, covk and principal hessian direction returns covk, Then, the distances between covk-SAVE and covk-DR are calculated, and good values of α for covk-DR method are finally recommended. Therefore, the distance computing between SIR-SAVE and covk-DR is ruled out. If the BAS for SIR, SAVE, covk and principal hessian direction returns any other one except covk, the pair of SIR-SAVE and covk-DR is considered, and good values of α for both hybrid SDR methods are sought. The third step selects one of the two hybrid SDR methods and its good α by running the BAS with SIR, SAVE, covk and DR.

The common part of the CDS algorithm to the Ye-Weiss bootstrapping algorithm is to compute distances between the kernel matrices. However, the key difference between the CDS and bootstrapping algorithms is placed onto the targets to compute the distances. In Ye and Weiss (2003), after picking one hybrid method and one value of α, the distances between the original sample kernel matrix and bootstrap sample kernel matrices are calculated B times, where B indicates the number of bootstrap samples. On the contrary, in Park *et al.* (2022), the distances are measured between two different hybrid methods for various values of α of each one. This latter approach dramatically reduces the number of fittings, because bootstrap samples are not necessary.

According to Part *et al.* (2022), the number of fits for the bootstrapping is $d_{\max} \times 27 \times B$, while the CDS requires $d_{\max} \times 21$, and the difference is $d_{\max} \times (27 \times B - 21)$. This indicates that the bootstrapping requires more time than the CDS without any exception.

3. Two more cross-distance selection algorithms

3.1. Larger cross-distance selection algorithm

In the original CDS algorithm, the covk-SAVE is ruled out for the final choice. If the BAS in the second step does not return covk, the covk-SAVE is not needed at all. Since the covk-SAVE may represent the data best, this should be unfair and may lead less accuracy.

The key in the second step of the original CDS is to yield the two hybrid SDR methods and their good αs. In this new algorithm, the BAS is not implemented in the second step, and, instead, for all pairs of the three candidate hybrid SDR methods, the $r_D^d(\alpha_i, \alpha_j)$ is computed, and choose the pair to minimize $r_D^d(\alpha_i, \alpha_j)$s along with the minimizers of α_i and α_j. Therefore, the covk-SAVE is equally treated with the other two candidates. This minimized pair of the hybrid SDR methods will be called *initial hybrid pair*. Then, run the BAS for SIR, SAVE, covk and DR for the final determination.

The underlying philosophy for the final determination is the methodological similarity between the initial hybrid pair and the BAS recommendation. For example, suppose that the initial hybrid pair and the choice by BAS are (SIR-SAVE, covk-DR) and SIR, respectively. Since the SIR-SAVE in the initial hybrid pair and the BAS recommendation shares SIR commonly, the SIR-SAVE is the final decision.

The sharing may not occur, however. When the initial hybrid pair and the BAS recommendation are (covk-DR, covk-SAVE) and SIR, respectively, there is no shared method. Then, it is reasonable to select covk-DR over cok-SAVE in the case, because the SIR is methodologically more kin to DR than SAVE. Suppose the initial hybrid pair and the BAS suggestion are (SIR-SAVE, covk-SAVE) and DR, respectively. Then, the SIR-SAVE will be the final decision over covk-SAVE, because the DR is methodologically more related with SIR than covk.

Another case that we need to consider is that the BAS recommendation is shared in both of the

Table 2: Final decision rule for the second step with cooperating the BAS recommendation

		Combination of the two hybrid methods		
		(SIR-SAVE, covk-DR)	(SIR-SAVE, covk-SAVE)	(covk-DR, covk-SAVE)
BAS	SIR	SIR-SAVE	SIR-SAVE	covk-DR
	SAVE	SIR-SAVE	*Bernoulli(0.5)*	covk-SAVE
	covk	covk-DR	covk-SAVE	*Bernoulli(0.5)*
	DR	covk-DR	SIR-SAVE	covk-DR

initial hybrid pair. Suppose that the initial hybrid pair and the BAS recommendation are (SIR-SAVE, covk-SAVE) and SAVE, respectively. The SAVE is shared in both SIR-SAVE and covk-SAVE. Another case happens with (covk-SAVE, covk-DR) and covk. We do not have any reasonable methodological guidance about the selection between the two, and the initial hybrid pair is indifferent to the BAS recommendation. So, the final determination will be randomly done. A random variable is generated from Bernoulli distribution with success probability 0.5. The value is equal to one, then the covk-SAVE is selected over SIR-SAVE and covk-DR. This final decision rule is summarized in Table 2. In Table 2, a notation of *Bernoulli (0.5)* stands for the random determination with success of covk-SAVE. So, this variation of the original CDS algorithm will be called *larger cross-distance selection* algorithm, and it is summarized as follows:

Algorithm 2 : Larger cross-distance selection

1. The first step is the same as the original CDS algorithm.
2. Caculate $r_D^d(\alpha_i, \alpha_j)$ between all pairs of the hybrid methods over the grids for α_i, α_j, and d. Then, pick the initial two hybrid SDR methods to minimize $r_D^d(\alpha_i, \alpha_j)$.
3. Run the BAS algorithm with SIR, SAVE, DR, and covk. Then, report one of the two initial hybrid methods along with its suggested α in Step 2 with cooperating the BAS recommendation in Table 2.

The total number of fitting required in the larger CDS algorithm is equal to $d_{\max} \times 32$ ($= d_{\max} \times (9 \times 2 + 8 + \binom{4}{2})$). The first 9×2 is for (SIR-SAVE, covk-DR) and (SIR-SAVE, covk-SAVE). The center 8 is for (covk-SAVE, covk-DR). The last $\binom{4}{2}$ is for BAS with SIR, SAVE, covk and DR.

3.2. Smaller cross-distance selection algorithm

It is questioned whether or not the covk-SAVE is really necessary for the original CDS algorithm, because it is never finally chosen anyhow. It is enough to consider only one pair of SIR-SAVE and covk-DR as hybrid SDR candidate methods with completely ruling covk-SAVE out. If so, the second step in the original CDS algorithm is no longer needed, and the third step alone should be enough. This algorithm will be called *smaller cross-distance selection* algorithm.

Algorithm 3 : Smaller cross-distance selection

1. The first step is the same as the original CDS algorithm.
2. Minimize $r_D^d(\alpha_i, \alpha_j)$ only between SIR-SAVE and covk-DR over the grids for α_i, α_j, and d. Run BAS with SIR, SAVE, covk and DR. If SIR and SAVE are recommended, SIR-SAVE is selected. Otherwise, covk-DR is chosen.

The total number of fitting for the smaller CDS algorithm is equal to $d_{\max} \times 15$ ($= d_{\max} \times (9 + \binom{4}{2})$).

Table 3: The number of fits required for the three CDS algorithms

Original CDS	Larger CDS	Smaller CDS
$d_{\max} \times 21$	$d_{\max} \times 32$	$d_{\max} \times 15$

3.3. Short remarks on the three CDS algorithms

The difference among the three CDS algorithms is based on how much the information by covk-SAVE is untilized in the selection procedure. The larger CDS algorithm treats the three hybrid SDR methods equally to pick an initial hybrid pair. In the original CDS, it is utilized only when the BAS recommends covk. The smaller CDS completely rules covk-SAVE out from the beginning. So, the asymptotic estimation performance of each CDS algorithm totally depend on how well the covk-SAVE represent data and can estimate $\mathcal{S}_{Y|\mathbf{X}}$.

The numbers of fits required for the three CDS algorithms are summarized in Table 3. The larger CDS algorithm requires two times more than the smaller one. For example, if setting $d_{\max} = 4$, the smaller CDS requires 60 fitting, while the larger one does 128. The difference in fits between the larger and smaller CDSs is 68, which is bigger than the total number of the smaller CDS.

4. Numerical studies and real data analysis

4.1. Numerical studies

For numerical studies, six simulated models considered in Park *et al.* (2022) are investigated to compare the estimation of performances of the three CDS algorithms.

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 : $Y|\mathbf{X} = X_1 + X_1X_2 + 0.5\varepsilon$.

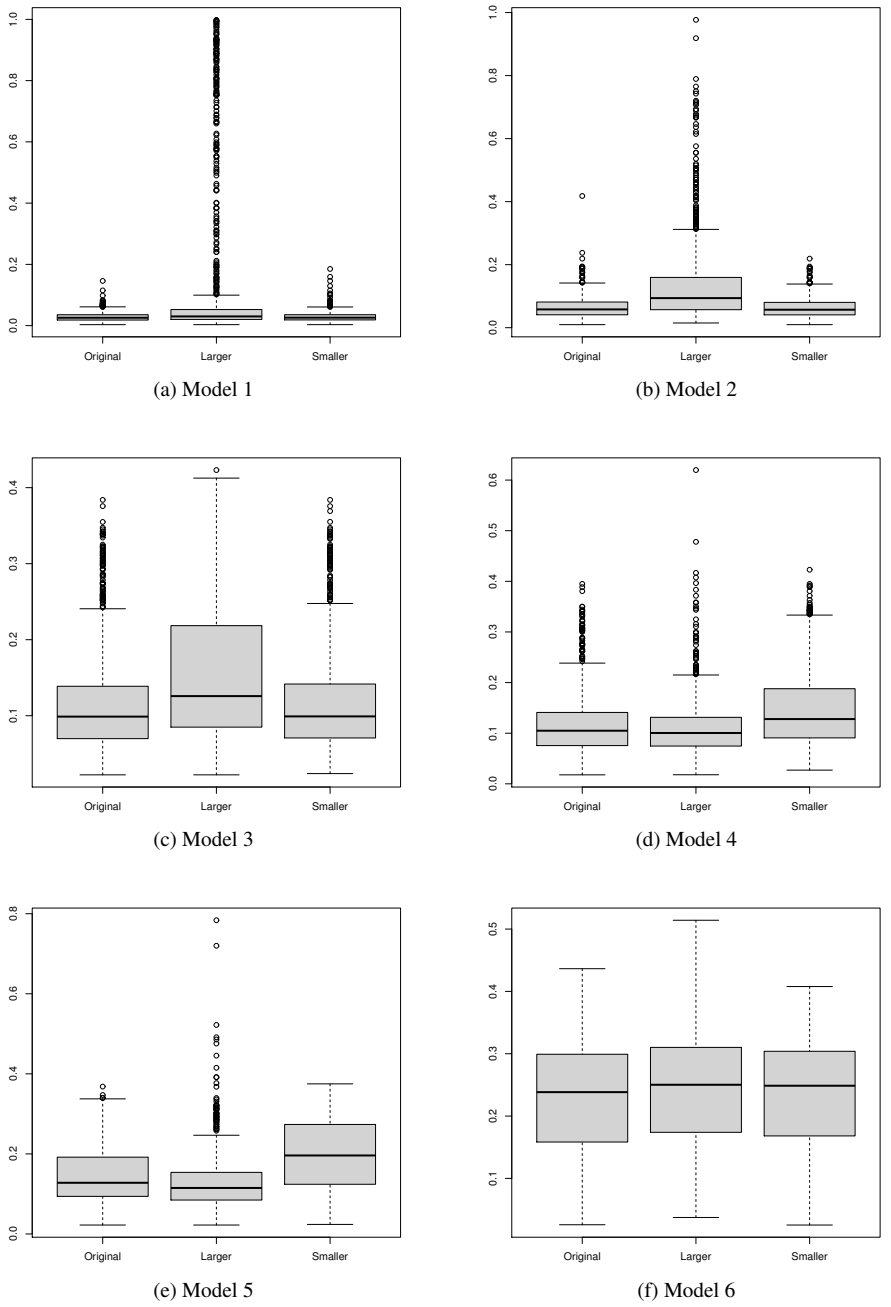
Model 5 : $Y|\mathbf{X} = X_1 + X_1^2 + X_1X_2 + 0.5\varepsilon$.

Model 6 : $Y|\mathbf{X} = X_1 + 0.5\exp(X_2)\varepsilon$.

For all six models, 10-dimensional predictors $\mathbf{X} = (X_1, \ldots, X_{10})^{\mathrm{T}}$ were commonly used. And the predictors and and a random error ε were independently sampled from $N(0, 1)$. Each model was generated 1000 times with the sample sizes 100. The weight α in the three hybrid SDR methods varies in a set of $(0.1, 0.2, \ldots, 0.8, 0.9)$.

Models 1–2 has one structural dimension, whose central subspace is spanned by $(1, 0, 0, \ldots, 0)^{\mathrm{T}}$. On the other hand, for all the other four models, $\mathcal{S}_{Y|\mathbf{X}}$ is commonly spanned by the two columns of $((1, 0, 0, \ldots, 0),(0, 1, 0, \ldots, 0))^{\mathrm{T}}$, so the structural dimension is equal to two.

It is known that SIR and covk have a clear advantage over SAVE and DR in estimating $\mathcal{S}_{Y|\mathbf{X}}$ for Model 1. Model 2 is a classic example for usefulness of SAVE, and both SIR and covk fails for the model. The DR can be used for Model 2, but SAVE estimates $\mathcal{S}_{Y|\mathbf{X}}$ even better. Models 3–5 are forms of usual polynomial regression with interactions. Model 6 has a representative example for test heteroscedasticity in many regression textbooks. Especially, Ye and Weiss (2003) used Models 3–6 to initiate the necessity of the hybrid approach for sufficient dimension reduction.



(a) Model 1 (b) Model 2

(c) Model 3 (d) Model 4

(e) Model 5 (f) Model 6

Figure 1: *Boxplots for trace correlation distance u_D for Models 1–6: Original, the original CDS; larger, the larger CDS; smaller, the smaller CDS.*

According to Park *et al.* (2002), the above six models do not represent all possible regression models. However, they have been widely adopted not only to teach linear regression but also to compare how well SDR methodologies estimate $\mathcal{S}_{Y|\mathbf{X}}$ in the literature.

To measure how well $\mathcal{S}_{Y|\mathbf{X}}$ is estimated, we compute a trace correlation (Hooper, 1959):

$$u^2(\mathbf{A}, \mathbf{B}) = \frac{1}{k}\sum_{i=1}^{k}\rho_i^2,$$

where $\mathbf{A} \in \mathbb{R}^{p\times k}$ and $\mathbf{B} \in \mathbb{R}^{p\times k}$ are orthonormal basis matrices for k-dimensional subspaces of $\mathcal{A}$ and $\mathcal{B}$, respectively, and ρ_i^2, $i = 1, \ldots, k$, stands for the ordered eigenvalues of $\mathbf{B}^{\mathrm{T}}\mathbf{A}\mathbf{A}^{\mathrm{T}}\mathbf{B}$. The values of $u^2(\mathbf{A}, \mathbf{B})$ chages from 0 to 1 like absolute Pearson's correlation coefficient. The trace correlation $u^2(\mathbf{A}, \mathbf{B})$ becomes 1, if and only if the two subspaces of $\mathcal{A}$ and $\mathcal{B}$ are identical. To covert a correlation (higher, closer) to a distance (smaller, closer), the following trace correlation distance u_D is considered:

$$u_D(\mathbf{A}, \mathbf{B}) = 1 - \sqrt{u^2(\mathbf{A}, \mathbf{B})}.$$

Actually, $u^2(\mathbf{A}, \mathbf{B})$ is an equivalent quantity to $r_D^d(\alpha_i, \alpha_j)$, and a different notation is used for distinguishing where it is applied and for more focusing on the weight α.

As a summary of the simulation studies, boxplots of u_D computed from the six models are reported in Figure 1. According to Figure 1, the larger CDS algorithm has the largest variation than the other two. Although the originial CDS algorithm does not dominate the larger and smaller ones in the estimation of $\mathcal{S}_{Y|\mathbf{X}}$, it is better than or well compete than the other two. Especially, as the regression has more complex mean and variance functions such as Models 4–6, the smaller CDS algorithm gets more inferior. This implies that the covk-SAVE may provide the more additional accuracy in estimating $\mathcal{S}_{Y|\mathbf{X}}$ than conidering just one pair of SIR-SAVE and covk-DR alone. When a regression has relatively less complicated mean function, like Models 1–3, the larger CDS algorithm has less accuracy and more variability than the original and smaller CDS ones.

For another summary of the numerical studies, Table 4 reports the average of αs selected by each CDS algorithm along with the best value of α given in Table 2 of Park *et al.* (2022). As seen in Table 4, the average of α selected by original CDS algorithm is the closest to the best one. It is observed that the average of α by the larger CDS algorithm is quite far from the best one with simpler regression mean functions, while it is so for the smaller CDS algorithm with more complex regresions. Also, the original CDS is competitive to the two other variations in the standard devitions of the selected α for the six models.

From the numerical studies, the covk-SAVE must not be ruled out but it should be carefully involved just like the original CDS algorithm. It is confirmed that the original CDS algorithm is arguably the best than the other two variations, and it is recommended as the default algorithm to use in practice.

4.2. Real data example: Abalone data

As a real data exmple, the analysis of abalone data, which is presented in Park *et al.* (2022), is illustrated. The main purpose of data is for predicting the age of abalone with their physical measurements. The data contain the following seven physical measurements: 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).

Table 4: Averages and standard devitaion in the parenthesis of α determined by the three CDS algorithms for each hybrid SDR methods: BEST, the best α reported in Table 2 of Park *et al.* (2022); original, the original CDS; larger, the larger CDS; smaller, the smaller CDS

	Model 1				Model 2			
	BEST	Original	Larger	Smaller	BEST	Original	Larger	Smaller
SIR-SAVE	0.9 (0.014)	0.782 (0.145)	0.466 (0.346)	0.764 (0.148)	0.4 (0.279)	0.245 (0.169)	0.135 (0.082)	0.245 (0.169)
covk-SAVE	0.8 (0.020)	N/A N/A	0.507 (0.303)	N/A N/A	0.1 (0.040)	N/A N/A	0.415 (0.259)	N/A NA
covk-DR	0.9 (0.016)	0.765 (0.163)	0.782 (0.102)	0.619 (0.319)	0.1 (0.031)	0.159 (0.153)	0.652 (0.246)	0.143 (0.116)
	Model 3				Model 4			
	BEST	Original	Larger	Smaller	BEST	Original	Larger	Smaller
SIR-SAVE	0.6 (0.060)	0.573 (0.198)	0.185 (0.198)	0.642 (0.183)	0.8 (0.079)	0.679 (0.141)	0.520 (0.302)	0.676 (0.142)
covk-SAVE	0.7 (0.075)	N/A N/A	0.666 (0.185)	N/A N/A	0.8 (0.051)	N/A N/A	0.697 (0.1654)	N/A NA
covk-DR	0.4 (0.057)	0.313 (0.280)	0.675 (0.247)	0.220 (0.202)	0.8 (0.045)	0.640 (0.273)	0.784 (0.092)	0.236 (0.197)
	Model 5				Model 6			
	BEST	Original	Larger	Smaller	BEST	Original	Larger	Smaller
SIR-SAVE	0.8 (0.080)	0.701 (0.145)	0.455 (0.331)	0.675 (0.149)	0.8 (0.086)	0.532 (0.194)	0.146 (0.130)	0.579 (0.172)
covk-SAVE	0.9 (0.048)	N/A /NA	0.738 (0.074)	N/A N/A	0.8 (0.081)	N/A N/A	0.628 (0.105)	N/A NA
covk-DR	0.9 (0.053)	0.637 (0.274)	0.777 (0.107)	0.258 (0.216)	0.9 (0.084)	0.405 (0.332)	0.778 (0.112)	0.152 (0.120)

the whole weight is eliminated due to avoiding multi-colinearity. The response is the age of abalone. For the analysis, we follow the suggestion by Park *et al.* (2022) for variable configurations, in which two outliers (the 1418^{th} and 2052^{th} observations) were removed before the reduction, and the three remaining weight predictors were transformed with square-root scale to satisfy the conditions in SIR, SAVE, covk and DR.

Then, the final result of the original and smaller CDS algorithms conicide with the selelction of covk-DR with $\alpha = 0.9$, which heavily depends on covk. In both, the other hybrid method in the pair is SIR-SAVE. For the original CDS algorithm applicaton, the first BAS application determines SIR, so the covk-SAVE was ruled out for further analysis. This is why the orignal and smaller CDS algorithms provide the same results. The second BAS application for the choice of SIR-SAVE and covk-DR suggests covk, so the covk-DR was finally recommended with $\alpha = 0.9$.

On the other hand, the larger CDS algorithm recommend covk-SAVE with $\alpha = 0.1$, which is quite close to SAVE. The initial hybrid pair is SIR-SAVE along with $\alpha = 0.3$, which is closer to SAVE than to SIR. Because of selecting covk by BAS, the covk-SAVE is the final decision.

The recommendation by the original and smaller CDS and the larger CDS is quite different. To investigate this in more detail, the trace-correlation distances u_D were computed between the basis estimators by the original and larger CDS algorithms for $d = 1, 2, 3$. The distances for $d = 1, 2, 3$ are 0.085, 0.037 and 0.161, respectively. For $d = 2$, covk-DR with $\alpha = 0.9$ and covk-SAVE with $\alpha = 0.1$ are quite similar to each other. According to Park *et al.* (2022), the consideration of $d = 2$ is reasonable, so the distance differences for $d = 1, 2, 3$ would be interpreted based on this. The difference in the basis estimates given $d = 1$ for covk-DR with $\alpha = 0.9$ and covk-SAVE with $\alpha = 0.1$

is partially due to potentially heavy dependence on covk and SAVE, respectively. Supposing that $d = 2$, the third basis estimates for covk-DR and covk-SAVE would be random as discussed in Yoo (2018) with overestimating d. Also, this implies the necessity of a careful testing procedure for the structural dimension d in the hybrid SDR method.

Since the importance of covk in the data coincides in both the original CDS selection of covk-DR with $\alpha = 0.9$ and the second BAS decision of covk, the further analysis of the data should be facilitated with the two-dimensionsional predictors from covk-DR with $\alpha = 0.9$ rather than covk-SAVE with $\alpha = 0.1$.

5. Discussion

Hybrid sufficient dimension reduction (SDR) methods to a weighted mean of kernel matrices of two different SDR methods by Ye and Weiss (2003) can provide more accuracy in estimating the central subspace than one single SDR method. However, due to heavy computation and time consumption required for bootstrapping, the hybrid SDR has not been used. To overcome this, Park *et al.* (2022) recently develop the so-called cross-distance selection (CDS) algorithm to select two SDR methods and their good weight α.

In the orginal CDS algorithm by Park *et al.* (2022), the covk-SAVE has been ruled out in the final decision, although it is partially utilized in the whole selection procedure. In the paper, two variations of the original CDS algorithm are proposed depending on how well and equally the covk-SAVE is treated in the selection procedure. In one variation, which is called the larger CDS algorithm, the covk-SAVE is equally and fairly utilized compared with the other two candiates of SIR-SAVE and covk-DR. But, for the final selection, a random selection should be necessary. On the other hand, SIR-SAVE and covk-DR are utilized with completely ruling covk-SAVE out, which is called the smaller CDS algorithm.

According to the numerical studies, as a regression is more complicated in the mean and variance functions, the smaller CDS algorithm is inferiror to the original and larger CDS ones, which of the two are quite similar. On the other hand, as the regression is simpler, the smaller and original CDS algorithms are similar to each other and better than the larger CDS one. So, these numerical studies indicate that covk-SAVE should be necessary, at least, somewhere in the whosle selection procedure, and it is confirmed that the original CDS algorithm utilizes the information of covk-SAVE fairly well. So, the original CDS algorithm is recommended to use as the default in practice.

From the real data example, a careful dimension determinaiton of the hybrid SDR methods must be developed, and the work in the direction is under considertion.

References

Cook RD (1998a). *Regression Graphics*, Wiley New York, New York.

Cook RD (1998b). Principal hessian directions revisited, *Journal of the American Statistical Association*, **93**, 84–94.

Cook RD and Weisberg S (1991). Discussion of "sliced inverse regression for dimension reduction" by Li KC, *Journal of the American Statistical Association*, **86**, 328–332.

Hooper JW (1959). Simultaneous equations and canonical correlation theory, *Econometrika*, **27**, 245–256.

Li B and Wang S (2007). On directional regression for dimension reduction, *Journal of the American Statistical Association*, **102**, 997–1008.

Li KC (1991). Sliced inverse regression for dimension reduction, *Journal of the American Statistical*

Association, **86**, 316–327.

Li KC (1992). On principal hessian directions for data visualization and dimension reduction: Another application of Stein's lemma, *Journal of the American Statistical Association*, **87**, 1025–1039.

Park Y, Kim K, and Yoo, JK (2022). On cross-distance selection algorithm for hybrid sufficient dimension reduction, *Computational Statistics and Data Analysis*, **176**, 1075627.

Ye Z and Weiss RE (2003). Using the bootstrap to select one of a new class of dimension reduction methods, *Journal of the American Statistical Association*, **98**, 968–979.

Yin X and Cook RD (2002). Dimension reduction for the conditional k^{th} moment in regression, *Journal of the Royal Statistical Society: Series B*, **64**, 159–175.

Yoo JK (2009). Partial moment-based sufficient dimension reduction, *Statistics and Probability Letters*, **79**, 450–456.

Yoo JK (2018). Basis-adaptive selection algorithm in dr-package, *The R Journal*, **10**, 124–132.

Received September 22, 2022; Revised October 23, 2022; Accepted October 30, 2022