Note on response dimension reduction for multivariate regression

Jae Keun Yoo[1,a]

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

Abstract

Response dimension reduction in a sufficient dimension reduction (SDR) context has been widely ignored until Yoo and Cook (*Computational Statistics and Data Analysis*, **53**, 334–343, 2008) founded theories for it and developed an estimation approach. Recent research in SDR shows that a semi-parametric approach can outperform conventional non-parametric SDR methods. Yoo (*Statistics: A Journal of Theoretical and Applied Statistics*, **52**, 409–425, 2018) developed a semi-parametric approach for response reduction in Yoo and Cook (2008) context, and Yoo (*Journal of the Korean Statistical Society*, 2019) completes the semi-parametric approach by proposing an unstructured method. This paper theoretically discusses and provides insightful remarks on three versions of semi-parametric approaches that can be useful for statistical practitioners. It is also possible to avoid numerical instability by presenting the results for an orthogonal transformation of the response variables.

Keywords: conditional mean, multivariate regression, response dimension reduction, semi-parametric model, sufficient dimension reduction

1. Introduction

Multivariate regression $\mathbf{Y} \in \mathbb{R}^r | \mathbf{X} \in \mathbb{R}^p$ is popular in many science fields when analyzing repeated measures, longitudinal data and time series data, where $r \geq 2$ and $p \geq 2$. In regression, sufficient dimension reduction (SDR) replaces the p-original predictors with lower-dimensional linearly transformed predictors without loss of information with respect to selected aspects of $\mathbf{Y} | \mathbf{X}$. Accordingly, SDR methods are focused on the reduction of $\mathbf{X}$, not $\mathbf{Y}$. However, a proper dimension reduction of the responses can facilitate statistical analysis by avoiding the curse of dimensionality in multivariate regression. If the reduction of multi-dimensional response variables is needed to be done, it is natural to follow the notion of SDR, which is to replace the responses with a lower-dimensional linearly transformed one, without loss of information.

Yoo and Cook (2008) previously provide theoretical foundation and inference procedure for the response dimension reduction in SDR context. In the seminal work, two types of response dimension reduction subspaces are newly defined without losing information on $E(\mathbf{Y} | \mathbf{X})$ and proposed an approach to estimate the response reduction subspaces. The estimation approach is non-parametrical and does not assume any specific regression models.

Cook (2007) recently showed that a semi-parametric approach in SDR can outperform non-parametric methods. Following the idea, Yoo (2018) proposed two versions of semi-parametric response dimension reduction approaches, called "principal response reduction (PRR)" and "principal

[1] Department of Statistics, Ewha Womans University, 52, Ewhayeodae-gil, Seodaemun-gu, Seoul 03760, Korea.
 E-mail: peter.yoo@ewha.ac.kr

fitted response reduction (PFRR)" in the context of Yoo and Cook (2008). Yoo (2018) confirms that the two semi-parametric approaches have potential advantages in the response dimension reduction over Yoo and Cook (2008). Next, Yoo (2019) developed "unstructured PFRR (UPFRR)", which do not assume the structure of the covariance matrix of the random error vectors in Yoo (2018) in the estimation. The advantage of Yoo (2019) is a possibility of equivariant or invariant full-rank transformation. Yoo (2019) also provides good guidelines to choose either PRR or PFRR. Therefore, the semi-parametric response reduction would be complete when including the unstructured fitted response reduction.

This paper provides insightful remarks on three semi-parametric approaches in order to clearly distinguish differences among the three approaches. In addition, theoretical results for the orthogonal transformation of the response variables in the response reduction are derived for the three semi-approaches that include Yoo and Cook (2008). Normally, the full-rank transformation of the response variables are incompatible to the response reduction subspaces, not like the case for the predictors. However, in the case of the orthogonal transformation, some similar results to the predictor case can hold.

The organization of the paper is as follows. Section 2 briefly introduces the short review on Yoo and Cook (2008) and the three semi-parametric response reduction approaches. The following section is devoted to two remarks on the three approaches. Section 4 is devoted to showing the results on the orthogonal transformation. Section 5 summarizes the work.

For notational convenience, we define that $\mathcal{S}(\mathbf{B})$ stands for a subspace spanned by the columns of $\mathbf{B} \in \mathbb{R}^{p \times r}$ and that $\mathbf{\Sigma}_x$ is the covariance matrix for a random vector $\mathbf{X} \in \mathbb{R}^p$. And, all proofs are given in Appendix not to interrupt reading flow.

2. Review of response reduction

2.1. Non-parametric response reduction

For a multivariate regression of $\mathbf{Y} \in \mathbb{R}^r | \mathbf{X} \in \mathbb{R}^p$, suppose that there exists a $q \times r$ matrix $\mathbf{L}$ to have the smallest possible rank among all possible matrices to satisfy

$$E(\mathbf{Y}|\mathbf{X}) = E\left\{\mathbf{P}^{\mathrm{T}}_{\mathbf{L}(\mathbf{\Sigma}_y)}\mathbf{Y}|\mathbf{X}\right\}, \tag{2.1}$$

where $q \leq r$ and $\mathbf{P}_{\mathbf{L}(\mathbf{\Sigma}_y)} = \mathbf{L}(\mathbf{L}^{\mathrm{T}}\mathbf{\Sigma}_y\mathbf{L})^{-1}\mathbf{L}^{\mathrm{T}}\mathbf{\Sigma}_y$ is the orthogonal projection operator relative to the inner product $< v_1, v_2 >_{\mathbf{\Sigma}_y} = v_1^{\mathrm{T}}\mathbf{\Sigma}_y v_2$.

If equation (2.1) holds, then predictors $\mathbf{X}$ influences the components of the conditional mean $E(\mathbf{Y}|\mathbf{X})$ only through $\mathbf{P}_{\mathbf{L}(\mathbf{\Sigma}_y)}$. This directly implies that lower-dimensional linear projection onto $\mathcal{S}(\mathbf{L})$ can replace the original r-dimensional response $\mathbf{Y}$ without loss of information on $E(\mathbf{Y}|\mathbf{X})$. In Yoo and Cook (YC) (2008), this type of response dimension reduction is defined as linear response reduction for $E(\mathbf{Y}|\mathbf{X})$.

Next, it is assumed that there exists a $k \times k$ matrix $\mathbf{K}$ satisfying that

$$E(\mathbf{Y}|\mathbf{X}) = E\left\{E\left(\mathbf{Y}|\mathbf{X}, \mathbf{K}^{\mathrm{T}}\mathbf{Y}\right)|\mathbf{X}\right\} = E\left\{E\left(\mathbf{Y}|\mathbf{K}^{\mathrm{T}}\mathbf{Y}\right)|\mathbf{X}\right\}, \tag{2.2}$$

where $k \leq r$ and $\mathbf{K}$ is not equal to the identity matrix.

$E(\mathbf{Y}|\mathbf{K}^{\mathrm{T}}\mathbf{Y})$ in the last equivalence of (2.2) is a function of $\mathbf{K}^{\mathrm{T}}\mathbf{Y}$; therefore, $E(\mathbf{Y}|\mathbf{X})$ is equivalently expressed as $E\{g(\mathbf{K}^{\mathrm{T}}\mathbf{Y})|\mathbf{X}\}$ for some function $g(\cdot)$. Then, another dimension reduction of $\mathbf{Y}$ is accomplished, if $k < r$, and this response reduction is called a conditional response reduction for $E(\mathbf{Y}|\mathbf{X})$.

The subspace spanned by columns of $\mathbf{L}$ and $\mathbf{K}$ are then called a 'response dimension reduction subspace'.

Yoo and Cook (2008) show that $\mathcal{S}(\mathbf{K}) \subseteq \mathcal{S}(\mathbf{L})$ for $\mathbf{L}$ in (2.1) and $\mathbf{K}$ in (2.2) and that the two subspaces are equal, if satisfying: A1. $E(\mathbf{Y}|\mathbf{K}^{\mathrm{T}}\mathbf{Y} = a)$ is linear in a. Condition A1 is called the linearity condition. The condition is satisfied, if $\mathbf{Y}$ is elliptically distributed. According to Hall and Li (1993), the condition is expected to hold in a reasonable approximation. If condition A1 fails, $\mathbf{Y}$ can be transformed for normality. Under condition A1, Yoo and Cook (2008) propose $\Sigma_y^{-1} \mathrm{cov}(\mathbf{Y}, \mathbf{X})\Sigma_x^{-1}$ to recover $\mathbf{L}$ and $\mathbf{K}$.

2.2. Semi-parametric response reduction

2.2.1. Principal response reduction

A semi-parametric response reduction approach starts with the following multivariate regression with assuming $E(\mathbf{Y}) = 0$ and $E(\mathbf{X}) = 0$ without loss of generality:

$$\mathbf{Y} = \Gamma\boldsymbol{\nu}_x + \boldsymbol{\varepsilon}, \tag{2.3}$$

where $\Gamma \in \mathbb{R}^{r \times d}$ with $\Gamma^{\mathrm{T}}\Gamma = \mathbf{I}_d$ and $d \leq r$, $\boldsymbol{\varepsilon} \sim N(0, \Sigma)$ and $\mathrm{cov}(\boldsymbol{\nu}_x, \boldsymbol{\varepsilon}) = 0$. In addition, $\boldsymbol{\nu}_x$ is a d-dimensional random function of $\mathbf{X}$ with a positive definite sample covariance and $\sum_{\mathbf{X}=x} \boldsymbol{\nu}_x = 0$. Supposing that $\boldsymbol{\nu}_x = \mathbf{X}$, model (2.3) is equal to a multivariate linear regression.

One important assumption required for the response reduction is that $\mathcal{S}(\Gamma)$ is an invariant and reducing subspace of Σ, which guarantee that $\Sigma = \Gamma\Omega\Gamma^{\mathrm{T}} + \Gamma_0\Omega_0\Gamma_0^{\mathrm{T}}$, where $\Gamma_0 \in \mathbb{R}^{r \times (r-d)}$ with $\Gamma_0^{\mathrm{T}}\Gamma_0 = \mathbf{I}_{r-d}$ and $\Gamma_0^{\mathrm{T}}\Gamma = 0$, $\Omega = \Gamma^{\mathrm{T}}\Sigma\Gamma$ and $\Omega_0 = \Gamma_0^{\mathrm{T}}\Sigma\Gamma_0$.

Under model (2.3), Yoo (2018) shows that $E(\mathbf{Y}|\mathbf{X}) = E(\mathbf{P}_{\Gamma(\Sigma_y)}^{\mathrm{T}}\mathbf{Y}|\mathbf{X})$, where $\mathbf{P}_{\Gamma(\Sigma_y)} = \Gamma(\Gamma^{\mathrm{T}}\Sigma_y\Gamma)^{-1}$ $\Gamma^{\mathrm{T}}\Sigma_y$ is the orthogonal projection operator relative to the inner product $< v_1, v_2 >_{\Sigma_y} = v_1^{\mathrm{T}}\Sigma_y v_2$. The original response $\mathbf{Y}$ can be reduced through Γ without loss of information of $E(\mathbf{Y}|\mathbf{X})$.

The primary interest is then placed onto the estimation of Γ in model (2.3). The maximum likelihood estimation approach is a natural choice, because the normal distribution of $\boldsymbol{\varepsilon}$ is assumed. Letting $\hat{\Sigma}_y$ be the usual moment estimator of Σ_y, Yoo (2018) shows that the maximum likelihood estimator (MLE) of Γ is a set of the eigenvectors corresponding to the first d largest eigenvalues of $\hat{\Sigma}_y$. This dimension reduction under model (2.3) is called PRR.

2.2.2. Principal fitted response reduction

The PRR utilizes the marginal information on $\mathbf{Y}$ without incorporating $\mathbf{X}$. This might be somewhat strange, because the mean function $E(\mathbf{Y}|\mathbf{X})$ is a function of $\mathbf{X}$, not $\mathbf{Y}$. To overcome this issue, we set $\boldsymbol{\nu}_x = \boldsymbol{\psi}\mathbf{f}_x$:

$$\mathbf{Y} = \Gamma\boldsymbol{\psi}\mathbf{f}_x + \boldsymbol{\varepsilon}, \tag{2.4}$$

where $\boldsymbol{\psi}$ is an unknown $d \times q$ matrix, and $\mathbf{f}_x \in \mathbb{R}^q$ is a known q dimensional vector-valued function of the predictor with $\sum_x \mathbf{f}_x = 0$. The following notations are defined for convenience.

$\mathbb{Y}$: the $n \times r$ data matrix for the responses

$\mathbb{X}$: the $n \times p$ data matrix for the predictors

$\mathbb{F}$: an $q \times n$ matrix constructed by stacking $\mathbf{f}_x^{\mathrm{T}}$ and $\mathbf{P}_{\mathbb{F}} = \mathbb{F}(\mathbb{F}^{\mathrm{T}}\mathbb{F})^{-1}\mathbb{F}^{\mathrm{T}}$

$\hat{\Sigma}_{\mathrm{fit}} = \mathbb{Y}^{\mathrm{T}}\mathbf{P}_{\mathbb{F}}\mathbb{Y}/n$ and $\hat{\Sigma}_{\mathrm{res}} = \hat{\Sigma}_y - \hat{\Sigma}_{\mathrm{fit}}$

Yoo (2018) uses $\mathbf{X}$, $\mathbf{X}^2 \exp(\mathbf{X})$, their combinations and the cluster indicator of $\mathbf{X}$ acquired from the K-means clustering algorithm as the candidates of $\mathbf{f}_x$. If setting $\mathbf{f}_x = \mathbf{X}$, $\mathbf{P}_{\mathbb{F}}\mathbb{Y}$ is equal to the ordinary least squares, and hence $\hat{\mathbf{\Sigma}}_{\text{fit}}$ is the regression product sums of square.

Under model (2.4), the MLE of $\mathbf{\Gamma}$ does not have a close form. The likelihood over $\mathbf{\Gamma}$ is as follows.

$$\mathcal{L}(\mathbf{\Gamma}, \mathbf{\Gamma}_0) = -\frac{n}{2} \log \left| \mathbf{\Gamma}_0^{\mathrm{T}} \hat{\mathbf{\Sigma}}_y \mathbf{\Gamma}_0 \right| - \frac{n}{2} \log \left| \mathbf{\Gamma}^{\mathrm{T}} \hat{\mathbf{\Sigma}}_{\text{res}} \mathbf{\Gamma} \right|.$$

Therefore, the MLE of $\mathbf{\Gamma}$ depend on both $\hat{\mathbf{\Sigma}}_y$ and $\hat{\mathbf{\Sigma}}_{\text{res}}$. Yoo (2018) recommends a sequential selection algorithm among a set of all the eigenvectors of $\hat{\mathbf{\Sigma}}_y$, $\hat{\mathbf{\Sigma}}_{\text{fit}}$, and $\hat{\mathbf{\Sigma}}_{\text{res}}$, following the suggestion in Cook (2007; Section 6.2). This approach to estimate $\mathbf{\Gamma}$ is called PFRR.

2.2.3. Unstructured principal fitted response reduction

In model (2.3), we assume that $\boldsymbol{\varepsilon} \sim N(0, \mathbf{\Sigma} > 0)$ and $\text{cov}(\boldsymbol{\nu}_x, \boldsymbol{\varepsilon}) = 0$:

$$\mathbf{Y} = \mathbf{\Gamma}\boldsymbol{\nu}_x + \boldsymbol{\varepsilon}. \tag{2.5}$$

The difference between models (2.3) and (2.5) is the structure of $\mathbf{\Sigma}$ along with $\mathbf{\Gamma}$. In model (2.5), the structure that $\mathbf{\Sigma} = \mathbf{\Gamma}\mathbf{\Omega}\mathbf{\Gamma}^{\mathrm{T}} + \mathbf{\Gamma}_0\mathbf{\Omega}_0\mathbf{\Gamma}_0^{\mathrm{T}}$ is no longer assumed.

Yoo (2019) presents the relationship between $\mathbf{\Sigma}$ and $\mathbf{\Sigma}_y$ for the invariant condition so that $\mathcal{S}(\mathbf{\Sigma}\mathbf{\Gamma}) \subseteq \mathcal{S}(\mathbf{\Gamma})$ if and only if $\mathcal{S}(\mathbf{\Sigma}_y\mathbf{\Gamma}) \subseteq \mathcal{S}(\mathbf{\Gamma})$. That is, the invariant condition for $\mathbf{\Sigma}$ is equivalent to that for $\mathbf{\Sigma}_y$. Suppose that one estimates $\mathbf{\Gamma}$ through PRR or PFRR. The result then allows us to investigate the invariant condition of $\mathbf{\Gamma}$ through a usual moment estimator of $\mathbf{\Sigma}_y$. Then Yoo (2019) show that $E(\mathbf{Y}|\mathbf{X}) = E(\mathbf{P}_{\mathbf{\Gamma}(\mathbf{\Sigma}_y)}^{\mathrm{T}}\mathbf{Y}|\mathbf{X})$ for model (2.5), as long as the invariance of $\mathcal{S}(\mathbf{\Gamma})$ for $\mathbf{\Sigma}_y$ holds. So, from now on, the invariant condition of $\mathbf{\Gamma}$ will be put on for $\mathbf{\Sigma}_y$ in model (2.5).

To utilize the information of predictors in the estimation of $\mathbf{\Gamma}$, its fitted component model is constructed as:

$$\mathbf{Y} = \mathbf{\Gamma}\boldsymbol{\psi}\mathbf{f}_x + \boldsymbol{\varepsilon}. \tag{2.6}$$

Let $\mathbf{E}_d$ and $\mathcal{S}_d(\mathbf{E})$ stand for the first d largest eigenvectors of a matrix $\mathbf{E}_d$ and a subspace spanned by the columns of $\mathbf{E}_d$, respectively. Define that $\mathbf{B} = \hat{\mathbf{\Sigma}}^{-1/2}\hat{\mathbf{\Sigma}}_{\text{fit}}\hat{\mathbf{\Sigma}}^{-1/2}$, $\mathbf{B}_{\text{res}} = \hat{\mathbf{\Sigma}}_{\text{res}}^{-1/2}\hat{\mathbf{\Sigma}}_{\text{fit}}\hat{\mathbf{\Sigma}}_{\text{res}}^{-1/2}$, and $\mathbf{B}_y = \hat{\mathbf{\Sigma}}_y^{-1/2}\hat{\mathbf{\Sigma}}_{\text{fit}}\hat{\mathbf{\Sigma}}_y^{-1/2}$. Also, define that $\hat{\mathbf{\Lambda}} = (\hat{\lambda}_1, \dots, \hat{\lambda}_q)$ and $\hat{\mathbf{V}} = (\hat{\boldsymbol{\gamma}}_1, \dots, \hat{\boldsymbol{\gamma}}_q)$ be the ordered eigenvalues and corresponding eigenvectors of $\mathbf{B}_{\text{res}}$ Let $\hat{\mathbf{K}}_d = \text{diag}(0, \dots, 0, \hat{\lambda}_{d+1}, \dots, \hat{\lambda}_q)$. Then, under model (2.6), Yoo (2019) derives the following results:

(1) $\hat{\mathcal{S}}(\mathbf{\Gamma}) = \hat{\mathbf{\Sigma}}^{\frac{1}{2}}\mathcal{S}_d(\mathbf{B})$ or $\hat{\mathbf{\Gamma}} = \hat{\mathbf{\Sigma}}^{\frac{1}{2}}\mathbf{B}_d$.

(2) $\hat{\mathbf{\Sigma}} = \hat{\mathbf{\Sigma}}_{\text{res}} + \hat{\mathbf{\Sigma}}_{\text{res}}^{\frac{1}{2}}\hat{\mathbf{V}}\hat{\mathbf{K}}_d\hat{\mathbf{V}}^{\mathrm{T}}\hat{\mathbf{\Sigma}}_{\text{res}}^{\frac{1}{2}} = \hat{\mathbf{\Sigma}}_{\text{res}}^{\frac{1}{2}}\left(\mathbf{I}_r + \hat{\mathbf{V}}\hat{\mathbf{K}}_d\hat{\mathbf{V}}^{\mathrm{T}}\right)\hat{\mathbf{\Sigma}}_{\text{res}}^{\frac{1}{2}}$.

(3) $\mathcal{L}_{\text{UPFRR}}^d = -\frac{n}{2} \log \left| \hat{\mathbf{\Sigma}}_{\text{res}} \right| + \frac{n}{2} \sum_{i=d+1}^{q} \log \left(1 + \hat{\lambda}_i \right)$.

(4) $\hat{\mathcal{S}}(\mathbf{\Gamma}) = \hat{\mathbf{\Sigma}}^{\frac{1}{2}}\mathcal{S}_d(\mathbf{B}) = \hat{\mathbf{\Sigma}}_{\text{res}}^{\frac{1}{2}}\mathcal{S}_d(\mathbf{B}_{\text{res}}) = \hat{\mathbf{\Sigma}}_y^{\frac{1}{2}}\mathcal{S}_d(\mathbf{B}_y)$.

The response reduction though model (2.6) will be called "unstructured PFRR (UPFRR)".

Table 1: Parameters and their dimensions in PFRR and UPFRR

	Γ	ψ	Σ	Total
PFRR	$(r-u)u$	qu	$u(u+1)/2 + (r-u)(r-u+1)/2$	$qu + r(r+1)/2$
UPFRR	$(r-u)u$	qu	$r(r+1)/2$	$(r-u)u + qu + r(r+1)/2$

PFRR = principal fitted response reduction; UPFRR = unstructured PFRR.

3. Two remarks on PRR, PFRR, and UPFRR

3.1. First remark on PRR and PFRR

Under PRR and PFRR, recall that $\Sigma = \Gamma\Omega\Gamma^T + \Gamma_0\Omega_0\Gamma_0^T$, where $\Gamma_0 \in \mathbb{R}^{r\times(r-d)}$ with $\Gamma_0^T\Gamma_0 = I_{r-d}$ and $\Gamma_0^T\Gamma = 0$, $\Omega = \Gamma^T\Sigma\Gamma$ and $\Omega_0 = \Gamma_0^T\Sigma\Gamma_0$. For PRR, the covariance matrix Σ cannot be restored because Ω is not estimable. In PFRR, Ω and Ω_0 can be estimated with $\hat{\Gamma}^T\hat{\Sigma}_{res}\hat{\Gamma}$ and $\hat{\Gamma}_0^T\hat{\Sigma}_y\hat{\Gamma}_0$, respectively. Therefore, a sample version of Σ is possibly constructed by $\hat{\Gamma}\hat{\Gamma}^T\hat{\Sigma}_{res}\hat{\Gamma}\hat{\Gamma}^T + \hat{\Gamma}_0\hat{\Gamma}_0^T\hat{\Sigma}_y\hat{\Gamma}_0\hat{\Gamma}_0^T$. It should be noted that the different sample quantities for $\hat{\Sigma}_{res}$ and $\hat{\Sigma}_y$ are used for Γ and Γ_0 and it is clearly observed that this does not coincide with its population structure.

The structural dimension d of Γ is assumed to be known in the estimation of Γ for PRR and PFRR. Since it is normally unknown, it should be estimated through a hypothesis test of $H_0 : d = m$ versus $H_1 : d = \min(q, r)$ for $m = 0, 1, \ldots, \min(q, r) - 1$. Since both PRR and PFRR use likelihood functions, a likelihood ratio test (LRT) should be a natural choice. For PFRR, the dimension estimation by LRT can be done with $\chi^2_{q(r-m)}$, while it is not plausible in PRR because Σ is not estimable.

3.2. Second remark on PFRR and UPFRR

Under UPFRR, Γ spans an invariant subspace of Σ, if and only if Γ spans an invariant subspace of Σ_y. This equivalence holds for PRR and PFRR, which is summarized in the following proposition.

Proposition 1. *Assume that model (2.3) holds. Then, $\mathcal{S}(\Sigma\Gamma) \subseteq \mathcal{S}(\Gamma)$ if and only if $\mathcal{S}(\Sigma_y\Gamma) \subseteq \mathcal{S}(\Gamma)$.*

According to Cook *et al.* (2007, Section 2.2), for a symmetric matrix $\mathbf{A}$, any invariant subspace of $\mathbf{A}$ is also a reducing subspace. Therefore, an invariant subspace of Σ becomes a reducing subspace. This implies that PFRR and UPFRR are the same model. The difference between the two models is if the structure of $\Sigma = \Gamma\Omega\Gamma^T + \Gamma_0\Omega_0\Gamma_0^T$ is kept in the estimation of Γ. The benefit to keep the structure, which is PFRR, is less parameters in the model than UPFRR. Table 1 summarizes the numbers of parameters of PFRR and UPFRR along with the difference $(r-u)u$. More parameters are a drawback of UPFRR, but it would not be a concern if the dimension of the response reduction subspace is not high. For example, suppose that $r = 4$ and $u = 1$ or $= 2$. Then the difference will be 3 for $u = 1$ and 4 for $u = 2$. However, the advantage of the approach is that UPFRR has a closed form of the estimator of Γ and obtains the equivariant transformation results for the responses.

4. Orthogonal transformation

The next proposition summarizes results of an orthogonal transformation of $\mathbf{Y}$ for response dimension reduction.

Proposition 2. *Consider an orthogonal transformation of $\mathbf{Y}$ such that $\mathbf{W} = \mathbf{O}^T\mathbf{Y}$ for an orthogonal matrix $\mathbf{O} \in \mathbb{R}^{r\times r}$, and let $\Gamma_w = \mathbf{O}^T\Gamma$.*
Assume that $E(\mathbf{Y}|\mathbf{X}) = \mathbf{P}_{\Gamma(\Sigma_y)}^T E(\mathbf{Y}|\mathbf{X})$. Then, the following statement holds.

(a) It holds that $E(\mathbf{W}|\mathbf{X}) = \mathbf{P}_{\Gamma_w(\Sigma_w)}^T E(\mathbf{W}|\mathbf{X})$.

Assume multivariate regression models in (2.3), (2.4), and (2.6).

(b) $\mathcal{S}(\Sigma_w \Gamma_w) \subseteq \mathcal{S}(\Gamma_w)$.

(c) An MLE of Γ_w is equal to $\mathbf{O}^T \hat{\Gamma}$ for PRR, PFRR, and UPFRR.

Proposition 2(a) indicates that Γ_w is a basis matrix of a response dimension reduction subspace for a regression of $\mathbf{W}|\mathbf{X}$. Γ_w is re-written as $\mathbf{O}^{-1}\Gamma$ since the inverse matrix of $\mathbf{O}$ is $\mathbf{O}^T$. This coincides with for SDR for $\mathbf{X}$ (Cook, 1998, Proposition 6.3). Proposition 2(a) is not guaranteed for any non-singular transformation $\mathbf{A}^T\mathbf{Y}$ of $\mathbf{Y}$ because the structure of $E(\mathbf{Y}|\mathbf{X})$ is changed according to $\mathbf{A}$. One good choice for $\mathbf{O}$ should be the eigenvectors of cov($\mathbf{Y}$). Let Ω be a set of all eigenvectors of cov($\mathbf{Y}$). Then, the covariance matrix of $\Omega^T\mathbf{Y}$ becomes a diagonal matrix, not the identity matrix, so numerical instability can be avoided, if necessary. Suppose that a response dimension reduction is done for a regression of $\Omega^T\mathbf{Y}|\mathbf{X}$, and obtain $\hat{\Gamma}_\omega$ as its estimate. Then, the basis estimate for the response reduction of $\mathbf{Y}|\mathbf{X}$ is directly computed as $\Omega\hat{\Gamma}_\omega$.

Proposition 2(b) implies that Γ_w is an invariant subspace of Σ_w, so Γ_w is also an invariant space for the covariance matrix of $\mathbf{O}^T\varepsilon$. This directly indicates that Γ_w is a basis matrix of the response reduction.

According to Proposition 2(c), the basis matrix of the response reduction for the orthogonal transformation is estimated by the estimate before the transformation pre-multiplied by the orthogonal matrix, so the same result is derived as the Yoo-Cook response reduction in Proposition 2(a).

5. Discussion

SDR has been successful in high-dimensional data analysis when involving multi-dimensional responses; consequently, their dimension reduction can facilitate the data analysis and induce undiscovered scientific results. Following the notion of SDR, the response dimension reduction was founded in Yoo and Cook (2008) along with a proposed non-parametrical approach. Two semi-parametric approaches were recently developed in Yoo (2018) and showed that the latter has potential advantage in the estimation of response reduction subspace over the former. Yoo (2019) also completes the semi-parametric method by proposing an unstructured approach. In the paper, the three version of the semi-parametric approach are discussed theoretically and provide insightful remarks that are beneficial to usual statistical practitioners to employ the semi-proposed approach. The paper also presents the results on an orthogonal transformation of response variables for the seminal work of Yoo and Cook (2008) and the three semi-parametric approaches. It is shown that it is possible to avoid numerical instability in practice in the estimation of a basis of the response reduction subspace.

Acknowledgements

Appendix: Proofs

Proof of Proposition 1: ($\Rightarrow$) By construction of model (2.3), we have $\Sigma_y = \Gamma \text{cov}(\nu_x)\Gamma^T + \Sigma$, and so $\Sigma_y = \Gamma \text{cov}(\nu_x)\Gamma^T + \Gamma\Omega\Gamma^T + \Gamma_0\Omega_0\Gamma_0^T$. So, $\Sigma_y\Gamma = \Gamma \text{cov}(\nu_x) + \Gamma\Omega = \Gamma\{\text{cov}(\nu_x) + \Omega\}$. This directly implies that $\Sigma_y\Gamma$ is expressed as a linear combination of Γ, so we have $\mathcal{S}(\Sigma_y\Gamma) \subseteq \mathcal{S}(\Gamma)$. ($\Leftarrow$) Assume that $\mathcal{S}(\Sigma_y\Gamma) \subseteq \mathcal{S}(\Gamma)$. The assumption indicates that there is a $d \times d$ matrix η^* such that $\Sigma_y\Gamma = \Gamma\eta^*$.

Then we have that $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_y - \boldsymbol{\Gamma}\,\mathrm{cov}(\boldsymbol{\nu}_x)\boldsymbol{\Gamma}^{\mathrm{T}}$, so $\boldsymbol{\Sigma}\boldsymbol{\Gamma} = \boldsymbol{\Sigma}_y\boldsymbol{\Gamma} - \boldsymbol{\Gamma}\,\mathrm{cov}(\boldsymbol{\nu}_x) = \boldsymbol{\Gamma}(\boldsymbol{\eta}^* - \mathrm{cov}(\boldsymbol{\nu}_x))$. This indicates that $\boldsymbol{\Sigma}\boldsymbol{\Gamma}$ is expressed as a linear combination of $\boldsymbol{\Gamma}$, so we have $\mathcal{S}(\boldsymbol{\Sigma}\boldsymbol{\Gamma}) \subseteq \mathcal{S}(\boldsymbol{\Gamma})$. This completes the proof. $\qquad\square$

Proof of Proposition 2: Recall that $\mathbf{OO}^{\mathrm{T}} = \mathbf{O}^{\mathrm{T}}\mathbf{O} = \mathbf{I}_r$ and $\mathbf{W} = \mathbf{O}^{\mathrm{T}}\mathbf{Y}$. It is easily noted that $\boldsymbol{\Sigma}_w = \mathrm{cov}(\mathbf{W}) = \mathbf{O}^{\mathrm{T}}\boldsymbol{\Sigma}_y\mathbf{O}$.

- Proof of part (a): By the assumption, we have the following equivalences.

$$\begin{aligned} E(\mathbf{Y}|\mathbf{X}) = \mathbf{P}^{\mathrm{T}}_{\boldsymbol{\Gamma}(\boldsymbol{\Sigma}_y)}E(\mathbf{Y}|\mathbf{X}) &= \boldsymbol{\Sigma}_y\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\Sigma}_y\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^{\mathrm{T}}E(\mathbf{Y}|\mathbf{X}) \\ &= \mathbf{OO}^{\mathrm{T}}\boldsymbol{\Sigma}_y\mathbf{OO}^{\mathrm{T}}\boldsymbol{\Gamma}\left(\boldsymbol{\Gamma}^{\mathrm{T}}\mathbf{OO}^{\mathrm{T}}\boldsymbol{\Sigma}_y\mathbf{OO}^{\mathrm{T}}\boldsymbol{\Gamma}\right)^{-1}\boldsymbol{\Gamma}^{\mathrm{T}}\mathbf{OO}^{\mathrm{T}}E(\mathbf{Y}|\mathbf{X}) \\ &= \mathbf{O}\boldsymbol{\Sigma}_w\boldsymbol{\Gamma}_w\left(\boldsymbol{\Gamma}^{\mathrm{T}}_w\boldsymbol{\Sigma}_w\boldsymbol{\Gamma}_w\right)^{-1}\boldsymbol{\Gamma}^{\mathrm{T}}_w E\left(\mathbf{O}^{\mathrm{T}}\mathbf{Y}|\mathbf{X}\right) \\ &= \mathbf{O}\mathbf{P}^{\mathrm{T}}_{\boldsymbol{\Gamma}_w(\boldsymbol{\Sigma}_w)}E(\mathbf{W}|\mathbf{X}). \end{aligned}$$

By pre-multiplying $\mathbf{O}^{\mathrm{T}}$ to both sides in the last equivalence above, we have $E(\mathbf{W}|\mathbf{X}) = \mathbf{P}^{\mathrm{T}}_{\boldsymbol{\Gamma}_w(\boldsymbol{\Sigma}_w)}E(\mathbf{W}|\mathbf{X})$. This completes the proof.

- Proof of part (b): By the assumption of PRR, PFRR, and UPFRR given in (2.3), (2.4), and (2.6), respectively, we have the following equivalences.

$$\mathcal{S}(\boldsymbol{\Sigma}_y\boldsymbol{\Gamma}) \subseteq \mathcal{S}(\boldsymbol{\Gamma}) \Leftrightarrow \mathcal{S}\left(\mathbf{OO}^{\mathrm{T}}\boldsymbol{\Sigma}_y\mathbf{OO}^{\mathrm{T}}\boldsymbol{\Gamma}\right) \subseteq \mathcal{S}\left(\mathbf{OO}^{\mathrm{T}}\boldsymbol{\Gamma}\right) \Leftrightarrow \mathbf{O}\mathcal{S}(\boldsymbol{\Sigma}_w\boldsymbol{\Gamma}_w) \subseteq \mathbf{O}\mathcal{S}(\boldsymbol{\Gamma}_w).$$

By pre-multiplying $\mathbf{O}^{\mathrm{T}}$ to both sides in the last equivalence, we have $\mathcal{S}(\boldsymbol{\Sigma}_w\boldsymbol{\Gamma}_w) \subseteq \mathcal{S}(\boldsymbol{\Gamma}_w)$ and this completes the proof.

- Proof of part (c): We have that $\mathbf{O}^{\mathrm{T}}\mathbf{Y} = \mathbf{O}^{\mathrm{T}}\boldsymbol{\Gamma}\boldsymbol{\nu}_x + \mathbf{O}^{\mathrm{T}}\boldsymbol{\varepsilon}$, equivalently $\mathbf{W} = \boldsymbol{\Gamma}_w\boldsymbol{\nu}_x + \boldsymbol{\varepsilon}_w$ by letting $\boldsymbol{\varepsilon}_w = \mathbf{O}^{\mathrm{T}}\boldsymbol{\varepsilon}$. Under PRR and PFRR in (2.3) and (2.4), $\boldsymbol{\Sigma}_{\mathbf{O}} = \mathrm{cov}(\boldsymbol{\varepsilon}_w) = \mathbf{O}^{\mathrm{T}}(\boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^{\mathrm{T}})\mathbf{O} + \mathbf{O}^{\mathrm{T}}(\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}^{\mathrm{T}}_0)\mathbf{O}$. It is easily noted that $\boldsymbol{\Gamma}^{\mathrm{T}}_w(\mathbf{O}^{\mathrm{T}}\boldsymbol{\Gamma}_0) = 0$, so $\mathbf{O}^{\mathrm{T}}\boldsymbol{\Gamma}_0$ becomes the orthogonal complement of $\boldsymbol{\Gamma}_w$. Letting $\mathbf{O}^{\mathrm{T}}\boldsymbol{\Gamma} = \boldsymbol{\Gamma}_{w,0}$, $\boldsymbol{\Sigma}_{\mathbf{O}} = \boldsymbol{\Gamma}_w\boldsymbol{\Omega}\boldsymbol{\Gamma}^{\mathrm{T}}_w + \boldsymbol{\Gamma}_{w,0}\boldsymbol{\Omega}_0\boldsymbol{\Gamma}^{\mathrm{T}}_{w,0}$. Therefore, the conditions required in PRR and PFRR hold for $\mathbf{W}|\mathbf{X}$. It is also noted that $\hat{\boldsymbol{\Sigma}}_w = \mathbf{O}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_y\mathbf{O}$, $\hat{\boldsymbol{\Sigma}}_{\mathrm{fit},w} = \mathbf{O}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_{\mathrm{fit}}\mathbf{O}$, and $\hat{\boldsymbol{\Sigma}}_{\mathrm{res},w} = \mathbf{O}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_{\mathrm{res}}\mathbf{O}$. This implies that the largest eigenvectors of $\hat{\boldsymbol{\Sigma}}_w$, $\hat{\boldsymbol{\Sigma}}_{\mathrm{fit},w}$, and $\hat{\boldsymbol{\Sigma}}_{\mathrm{res},w}$ is the same as the largest eigenvectors of $\hat{\boldsymbol{\Sigma}}_y$, $\hat{\boldsymbol{\Sigma}}_{\mathrm{fit}}$, and $\hat{\boldsymbol{\Sigma}}_{\mathrm{res}}$ pre-multiplied by $\mathbf{O}^{\mathrm{T}}$, respectively.

By this relation, for PRR, we directly have that $\hat{\boldsymbol{\Gamma}}_w = \mathbf{O}^{\mathrm{T}}\hat{\boldsymbol{\Gamma}}$.

For PFRR, we have the following likelihood function:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\Gamma}_w, \boldsymbol{\Gamma}_{w0}) &= -\frac{n}{2}\log\left|\boldsymbol{\Gamma}^{\mathrm{T}}_{w0}\hat{\boldsymbol{\Sigma}}_w\boldsymbol{\Gamma}_{w0}\right| - \frac{n}{2}\log\left|\boldsymbol{\Gamma}^{\mathrm{T}}_w\hat{\boldsymbol{\Sigma}}_{\mathrm{res},w}\boldsymbol{\Gamma}_w\right| \\ &= -\frac{n}{2}\log\left|(\mathbf{O}\boldsymbol{\Gamma}_{w0})^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_y(\mathbf{O}\boldsymbol{\Gamma}_{w0})\right| - \frac{n}{2}\log\left|(\mathbf{O}\boldsymbol{\Gamma}_w)^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_{\mathrm{res}}(\mathbf{O}\boldsymbol{\Gamma}_w)\right|. \end{aligned}$$

Therefore, $\boldsymbol{\Gamma}_w$ is maximized at $\hat{\boldsymbol{\Gamma}}$, so $\mathbf{O}\hat{\boldsymbol{\Gamma}}_w = \hat{\boldsymbol{\Gamma}}$. This directly implies that $\hat{\boldsymbol{\Gamma}}_w = \mathbf{O}^{\mathrm{T}}\hat{\boldsymbol{\Gamma}}$.

For UPFRR in the regression of $\mathbf{W}|\mathbf{X}$, we have

$$\begin{aligned} \mathbf{B}_w = \hat{\boldsymbol{\Sigma}}^{-\frac{1}{2}}_w\hat{\boldsymbol{\Sigma}}_{\mathrm{fit},w}\hat{\boldsymbol{\Sigma}}^{-\frac{1}{2}}_w &= \mathbf{O}^{\mathrm{T}}\mathbf{O}\hat{\boldsymbol{\Sigma}}^{-\frac{1}{2}}_w\mathbf{O}^{\mathrm{T}}\mathbf{O}\hat{\boldsymbol{\Sigma}}_{\mathrm{fit},w}\mathbf{O}^{\mathrm{T}}\mathbf{O}\boldsymbol{\Sigma}^{-\frac{1}{2}}_w\mathbf{O}^{\mathrm{T}}\mathbf{O} \\ &= \mathbf{O}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}^{-\frac{1}{2}}_y\hat{\boldsymbol{\Sigma}}_{\mathrm{fit}}\hat{\boldsymbol{\Sigma}}^{-\frac{1}{2}}_y\mathbf{O} \end{aligned}$$

Therefore, the largest eigenvectors of $\mathbf{B}_w$ is the same as those of $\mathbf{B}_y$ pre-multiplied by $\mathbf{O}^T$. Let $\mathbf{B}_{w,d}$ be the first largest d eigenvectors of $\mathbf{B}_w$. Since $\hat{\mathbf{\Gamma}}_w = \hat{\mathbf{\Sigma}}_w^{1/2}\mathbf{B}_{w,d} = \mathbf{O}^T\mathbf{O}\Sigma_w^{1/2}\mathbf{O}^T\mathbf{O}\mathbf{B}_{w,d} = \mathbf{O}^T\Sigma_y^{1/2}\mathbf{B}_{y,d} = \mathbf{O}^T\hat{\mathbf{\Gamma}}$. This completes the proof for part (c).

$\square$

References

Cook RD (1998). *Regression Graphics: Ideas for Studying Regressions Through Graphics*, Wiley, New York.

Cook RD (2007). Fisher lecture: dimension reduction in regression, *Statistical Science*, **22**, 1–26.

Cook RD, Li B, and Chiaromonte F (2007). Dimension reduction in regression without matrix inversion, *Biometrika*, **94**, 569–584.

Hall P and Li KC (1993). On almost linearity of low-dimensional projections from high-dimensional data, *Annals of Statistics*, **21**, 867–889.

Yoo JK (2018). Response dimension reduction: model-based approach, *Statistics: A Journal of Theoretical and Applied Statistics*, **52**, 409–425.

Yoo JK (2019). Unstructured principal fitted response reduction in multivariate regression, *Journal of the Korean Statistical Society* (in Press), https://doi.org/10.1016/j.jkss.2019.02.001

Yoo JK and Cook RD (2008). Response dimension reduction for the conditional mean in multivariate regression, *Computational Statistics and Data Analysis*, **53**, 334–343.