Research Article

CHEMOMETRICS

Received: 4 April 2014, Revised: 15 October 2014, Accepted: 24 October 2014, Published online in Wiley Online Library: 26 November 2014

(wileyonlinelibrary.com) DOI: 10.1002/cem.2691

High-throughput data dimension reduction via seeded canonical correlation analysis

Yunju Im, HeyIn Gang and Jae Keun Yoo*

Canonical correlation analysis (CCA) is one of popular statistical methodologies in multivariate analysis, especially, in studying relation of two sets of variables. However, if sample sizes are smaller than the maximum of the dimensions of two sets of variables, it is not plausible to construct canonical coefficient matrices due to failure of inverting sample covariance matrices. In this article, we develop a two step procedure of CCA implemented in such situation. For this, seeded dimension reduction is adapted into CCA. Numerical studies confirm the approach, and two real data analyses are presented. Copyright © 2014 John Wiley & Sons, Ltd.

Keywords: canonical correlation analysis; large *p* small *n*; multivariate analysis; seeded dimension reduction

1. INTRODUCTION

Canonical correlation analysis (CCA) is a method to measure an association between two sets of variables and focuses on seeking pairs of linear combinations from two sets of variables by maximizing the Pearson correlation between two sets of variables. The pairs of linear combination and their correlations are called canonical variates and canonical correlations, respectively. A few pairs of canonical variates are believed to be able to represent the relation between the original sets of variables and their variabilities. We briefly introduce the major concepts and statistics of CCA.

Suppose that we are interested in measures of association between two sets of variables of $\mathbf{X} \in \mathbb{R}^p$ and $\mathbf{Y} \in \mathbb{R}^r$. Define that $\mathrm{cov}(\mathbf{X}) = \boldsymbol{\Sigma}_x > 0$, $\mathrm{cov}(\mathbf{Y}) = \boldsymbol{\Sigma}_y > 0$, $\mathrm{cov}(\mathbf{X}, \mathbf{Y}) = \boldsymbol{\Sigma}_{xy}$, and $\mathrm{cov}(\mathbf{Y}, \mathbf{X}) = \boldsymbol{\Sigma}_{yx}$. For two linear combinations of $\mathbf{X}$ and $\mathbf{Y}$, saying $U = \mathbf{a}^\mathsf{T}\mathbf{X}$ and $V = \mathbf{b}^\mathsf{T}\mathbf{Y}$, it is obtained that $\mathrm{var}(U) = \mathbf{a}^\mathsf{T}\boldsymbol{\Sigma}_x\mathbf{a}$, $\mathrm{var}(V) = \mathbf{b}^\mathsf{T}\boldsymbol{\Sigma}_y\mathbf{b}$ and $\mathrm{cov}(U, V) = \mathbf{a}^\mathsf{T}\boldsymbol{\Sigma}_{xy}\mathbf{b}$, where $\mathbf{a} \in \mathbb{R}^{p\times 1}$ and $\mathbf{b} \in \mathbb{R}^{r\times 1}$.

We seek to find $\mathbf{a}$ and $\mathbf{b}$ to maximize Pearson-correlation between U and V:

$$\mathrm{cor}(U, V) = \frac{\mathbf{a}^\mathsf{T}\boldsymbol{\Sigma}_{xy}\mathbf{b}}{\sqrt{\mathbf{a}^\mathsf{T}\boldsymbol{\Sigma}_x\mathbf{a}}\sqrt{\mathbf{b}^\mathsf{T}\boldsymbol{\Sigma}_y\mathbf{b}}} \tag{1}$$

In CCA, such $\mathbf{a}$ and $\mathbf{b}$ are constructed based on the following criteria:

(1) The first canonical variate pair ($U_1 = \mathbf{a}_1^\mathsf{T}\mathbf{X}, V_1^\mathsf{T} = \mathbf{b}_1^\mathsf{T}\mathbf{Y}$) is constructed from the maximization of (1)
(2) At the $k \geq 2$ step, the kth canonical variate pair ($U_k = \mathbf{a}_k^\mathsf{T}\mathbf{X}, V_k = \mathbf{b}_k^\mathsf{T}\mathbf{Y}$) is constructed from the maximization of (1) with restriction that $\mathrm{var}(U_k) = \mathrm{var}(V_k) = 1$ and (U_k, V_k) are uncorrelated with the previous $(k-1)$ canonical variate pairs.
(3) Repeat Steps 1 and 2 until $k = \min(p, r)$.
(4) Select the first d pairs of (U_k, V_k) to represent the relationship between $\mathbf{X}$ and $\mathbf{Y}$.

Then the pairs $(\mathbf{a}_i, \mathbf{b}_i)$ are acquired as follows: $\mathbf{a}_i = \boldsymbol{\Sigma}_x^{-1/2}\boldsymbol{\psi}_i$ and $\mathbf{b}_i = \boldsymbol{\Sigma}_y^{-1/2}\boldsymbol{\phi}_i$ for $i = 1, \ldots, q$, where $(\boldsymbol{\psi}_1, ..., \boldsymbol{\psi}_q)$ and $(\boldsymbol{\phi}_1, ..., \boldsymbol{\phi}_q)$ are the eigenvectors of $\boldsymbol{\Sigma}_x^{-1/2}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_y^{-1}\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_y^{-1/2}$ and $\boldsymbol{\Sigma}_y^{-1/2}\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_x^{-1/2}$ with the corresponding common nonzero ordered-eigenvalues of $\rho_1^{*2} \geq \cdots \geq \rho_q^{*2} \geq 0$, respectively. The matrices of $\mathbf{M}_x = (\mathbf{a}_1, \ldots, \mathbf{a}_q)$ and $\mathbf{M}_y = (\mathbf{b}_1, \ldots, \mathbf{b}_q)$ are called canonical coefficient matrices.

Hereafter, the CCA by the decompositions of $\boldsymbol{\Sigma}_x^{-1/2}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_y^{-1}\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_y^{-1/2}$ and $\boldsymbol{\Sigma}_y^{-1/2}\boldsymbol{\Sigma}_{yx}\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_x^{-1/2}$ will be called the *standard CCA*. For more details regarding the standard CCA, readers may refer to Johnson and Wichern [1].

One major problem arises with a sample size n less than or equal to $\max(p, r)$, because the sample covariance matrices of $\boldsymbol{\Sigma}_x$ and $\boldsymbol{\Sigma}_y$ are not invertible. To overcome this, various penalized approaches were proposed. Parkhomenko *et al.* [2] suggested a CCA to be carried out by penalizing the left and right singular vectors in the covariance matrix of $\boldsymbol{\Sigma}_{xy}$. Waaijenborg *et al.* [3] converted the standard CCA to regression forms and adopted a penalized method called elastic net (Zou and Hastie [4]). And, a sparse CCA proposed by Le Cao *et al.* [5] adapted sparse partial least squares into CCA. Witten *et al.* [6] applied a penalized matrix decomposition into CCA. One common thing in the various versions of the sparse CCAs is to overcome the matrix inversion problem and to make parts of canonical coefficient matrices zeros. For this, tuning parameters are required, which are typically determined by cross-validation procedures. Therefore, when $\min(p, r)$ is large, the sparse CCA methods often turn out to computationally intensive in practice.

In this paper, we propose an approach of CCA applicable with $n \leq \max(p, r)$ by adapting a seeded dimension reduction

* *Correspondence to: Jae Keun Yoo, Department of Statistics, Ewha Womans University, Seoul 120-750, Korea*
E-mail: peter.yoo@ewha.ac.kr

Department of Statistics, Ewha Womans University

(Cook *et al.* [7]). Different from the sparse CCAs, the proposed method does not make parts of canonical coefficient matrices zeros intentionally. Instead, seed matrices, constructed from $\boldsymbol{\Sigma}_{xy}$ and $\boldsymbol{\Sigma}_{yx}$, are iteratively projected into the marginal covariance matrices of $\boldsymbol{\Sigma}_x$ and $\boldsymbol{\Sigma}_y$. By this, the proposed method can be enjoyed relatively less intensiveness in computation.

This paper is organized as follows. We briefly discuss seeded dimension reduction in Section 2. Section 3 is devoted to associate CCA with seeded dimension reduction for developing seeded canonical correlation analysis. Numerical studies and real data analysis are presented in Section 4. In Section 5, we summarize our works.

We will define notations frequently used. For a matrix $\mathbf{B} \in \mathbb{R}^{p\times q}$ and a subspace $\mathcal{S}$ of $\mathbb{R}^p$, $\mathbf{B}\mathcal{S}$ denotes the set of $\{\mathbf{B}x : x \in \mathcal{S}\}$. A subspace $\mathcal{S}(\mathbf{b})$ represents a subspace spanned by the columns of $\mathbf{b}$. For a symmetric and positive definite matrix $\boldsymbol{\Sigma}$, a $\boldsymbol{\Sigma}$ inner-product in $\mathbb{R}^p$ is defined as $< \mathbf{a}, \mathbf{b} >_{\boldsymbol{\Sigma}} = \mathbf{a}^{\mathsf{T}}\boldsymbol{\Sigma}\mathbf{b}$. An orthogonal projection operator onto $\mathcal{S}(\mathbf{b})$ relative to $< \mathbf{a}, \mathbf{b} >_{\boldsymbol{\Sigma}}$ will be defined as $\mathbf{b}(\mathbf{b}^{\mathsf{T}}\boldsymbol{\Sigma}\mathbf{b})^{\dagger}\mathbf{b}^{\mathsf{T}}\boldsymbol{\Sigma}$, where † stands for the Moore-Penrose inverse (Searle [8]). And, the true rank of $\mathbf{M}_x$ and $\mathbf{M}_y$ will be denoted as d.

2. SEEDED DIMENSION REDUCTION

The primary purpose of sufficient dimension reduction (SDR) in regression of $\mathbf{Y} \in \mathbb{R}^r|\mathbf{X} \in \mathbb{R}^p$ is to replace the original p-dimensional predictors $\mathbf{X}$ with a lower dimensional linear projection without loss of information on the conditional distribution of $\mathbf{Y}|\mathbf{X}$. In other words, SDR seeks to search $\boldsymbol{\alpha} \in \mathbb{R}^{p\times q}$ such that

$$\mathbf{Y} \perp\!\!\!\perp \mathbf{X}|\boldsymbol{\alpha}^{\mathsf{T}}\mathbf{X}, \tag{2}$$

where $\perp\!\!\!\perp$ represents independence and $q \le p$.

Subspaces spanned by the columns of $\boldsymbol{\alpha}$ satisfying (2) are called dimension reduction subspaces, and the minimal subspace among them is called the central subspace $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$. Naturally, the estimation of $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$ is the main stream in the context of SDR.

In general, many SDR methods require the inversion of the covariance matrix $\boldsymbol{\Sigma}_x$. However, with sample sizes $n \le p$, the inversion of the sample version of $\boldsymbol{\Sigma}_x$ is not plausible, and hence the SDR methods may not be applicable.

In order to overcome the problematic issue, Cook *et al.* [7] proposed a method without matrix inversion. For this to be carried out, a seed matrix $\boldsymbol{\nu} \in \mathbb{R}^{p\times q}$ should be defined so that $\mathcal{S}(\boldsymbol{\nu}) \subseteq \boldsymbol{\Sigma}_x\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$. In order to avoid complications, it is assumed that $\mathcal{S}(\boldsymbol{\nu}) = \boldsymbol{\Sigma}_x\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$.

Suppose that there is a known subspace $\mathcal{M}_{\mathbf{Y}|\mathbf{X}}$ of $\mathbb{R}^p$ containing $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$. This indicates that $\boldsymbol{\Sigma}_x^{-1}\mathcal{S}(\boldsymbol{\nu}) \subseteq \mathcal{M}_{\mathbf{Y}|\mathbf{X}}$. Let $\mathbf{P}_{\mathcal{M}_{\mathbf{Y}|\mathbf{X}}(\boldsymbol{\Sigma}_x)} = \mathbf{R}(\mathbf{R}^{\mathsf{T}}\boldsymbol{\Sigma}_x\mathbf{R})^{-1}\mathbf{R}^{\mathsf{T}}\boldsymbol{\Sigma}_x$ be an orthogonal projection operator onto $\mathcal{M}_{\mathbf{Y}|\mathbf{X}}$ relative to $< \mathbf{a}, \mathbf{b} >_{\boldsymbol{\Sigma}_x}$, where $\mathbf{R}$ is a $p \times q$ matrix such that $\mathcal{S}(\mathbf{R}) = \mathcal{M}_{\mathbf{Y}|\mathbf{X}}$.

The previous discussion directly implies the following equivalences:

$$\begin{aligned}\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\nu} &= \mathbf{P}_{\mathcal{M}_{\mathbf{Y}|\mathbf{X}}(\boldsymbol{\Sigma}_x)}\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\nu} = \mathbf{R}(\mathbf{R}^{\mathsf{T}}\boldsymbol{\Sigma}_x\mathbf{R})^{-1}\mathbf{R}^{\mathsf{T}}\boldsymbol{\Sigma}_x\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\nu} \\ &= \mathbf{R}(\mathbf{R}^{\mathsf{T}}\boldsymbol{\Sigma}_x\mathbf{R})^{-1}\mathbf{R}^{\mathsf{T}}\boldsymbol{\nu}.\end{aligned} \tag{3}$$

According to the last equivalence of (3), the columns of $\mathbf{R}(\mathbf{R}^{\mathsf{T}}\boldsymbol{\Sigma}_x\mathbf{R})^{-1}\mathbf{R}^{\mathsf{T}}$ span $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$, but the inversion of $\boldsymbol{\Sigma}_x$ is not required. If $\mathbf{R}^{\mathsf{T}}\boldsymbol{\Sigma}_x\mathbf{R}$ is not invertible, $(\mathbf{R}^{\mathsf{T}}\boldsymbol{\Sigma}_x\mathbf{R})^{\dagger}$ is applied instead.

To estimate $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$ through $\mathbf{R}(\mathbf{R}^{\mathsf{T}}\boldsymbol{\Sigma}_x\mathbf{R})^{-1}\mathbf{R}^{\mathsf{T}}\boldsymbol{\nu}$, we need to identify the matrix $\mathbf{R}$, whose column subspace is large enough to enclose $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$ and small enough to estimate $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$ from available data. In order to find the matrix $\mathbf{R}$, iterative projections of $\boldsymbol{\nu}$ onto $\boldsymbol{\Sigma}_x$ were suggested by Cook *et al.* [7]:

$$\mathbf{R}_u \equiv (\boldsymbol{\nu}, \boldsymbol{\Sigma}_x\boldsymbol{\nu}, \boldsymbol{\Sigma}_x^2\boldsymbol{\nu}, \ldots, \boldsymbol{\Sigma}_x^{u-1}\boldsymbol{\nu}),\ u = 1, 2, \ldots, u^*. \tag{4}$$

We call the letter u in (4) a termination index of the projections. It is noted that $\mathcal{S}(\mathbf{R}_{u-1}) \subseteq \mathcal{S}(\mathbf{R}_u)$ for any $u \ge 2$. Because $\mathcal{S}(\mathbf{R}_u)$ forms a nondecreasing sequence, it is important to select a proper termination index u, large enough to guarantee $\mathcal{S}(\mathbf{R}_u) = \mathcal{M}_{\mathbf{Y}|\mathbf{X}}$ and small enough to capture $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$. Recently, Yoo [9] suggests bootstrap coefficients of variations to determine the termination index, which does not require any asymptotics and is implemented in a simple way.

3. SEEDED CANONICAL CORRELATION ANALYSIS

3.1. Development

To begin with, we need to see why the seeded dimension reduction method can be adapted to CCA, by investigating the relation between ordinary least squares and canonical coefficient matrices.

Recall the definitions of the canonical coefficient matrices $\mathbf{M}_x$ and $\mathbf{M}_y$ from the introduction. According to Lee and Yoo [10], it has been shown that

$$\mathcal{S}(\mathbf{M}_x) = \mathcal{S}(\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}_{xy}) \text{ and } \mathcal{S}(\mathbf{M}_y) = \mathcal{S}(\boldsymbol{\Sigma}_y^{-1}\boldsymbol{\Sigma}_{yx}), \tag{5}$$

where $\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}_{xy}$ and $\boldsymbol{\Sigma}_y^{-1}\boldsymbol{\Sigma}_{yx}$ are the ordinary least squares coefficient matrices of $\mathbf{Y}|\mathbf{X}$ and $\mathbf{X}|\mathbf{Y}$, respectively.

The relation in (5) directly indicates that the information on $\mathbf{M}_x$ and $\mathbf{M}_y$ is exhaustively restored through $\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}_{xy}$ and $\boldsymbol{\Sigma}_y^{-1}\boldsymbol{\Sigma}_{yx}$. Another interpretation of (5) is that one of possible basis matrices for $\mathcal{S}(\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}_{xy})$ can become $\mathbf{M}_x$. Therefore, once any basis matrices of $\mathcal{S}(\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}_{xy})$ and $\mathcal{S}(\boldsymbol{\Sigma}_y^{-1}\boldsymbol{\Sigma}_{xy})$ are known, $\mathbf{M}_x$ and $\mathbf{M}_y$ can be restored through some orthogonalization procedures.

If $\max(p, r) \ge n$, $\mathbf{M}_x$ and $\mathbf{M}_y$ cannot be estimated through the standard CCA due to failure of inverting the sample versions of $\boldsymbol{\Sigma}_x$ and $\boldsymbol{\Sigma}_y$. Meanwhile, the basis matrices of $\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}_{xy}$ and $\boldsymbol{\Sigma}_y^{-1}\boldsymbol{\Sigma}_{yx}$ are estimable through seeded dimension reduction with seed matrices of $\boldsymbol{\Sigma}_{xy}$ and $\boldsymbol{\Sigma}_{yx}$. Therefore, although the direct estimation of $\mathbf{M}_x$ and $\mathbf{M}_y$ is not plausible through the standard CCA with $\max(p, r) \ge n$, they can be indirectly restored by estimating basis matrices of $\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}_{xy}$ and $\boldsymbol{\Sigma}_y^{-1}\boldsymbol{\Sigma}_{yx}$.

Let $\mathbf{M}_{x,0}$ and $\mathbf{M}_{y,0}$ be matrices resulted from the seeded dimension reduction application with seed matrices of $\boldsymbol{\Sigma}_{xy}$ and $\boldsymbol{\Sigma}_{yx}$. Then the relation in (5) directly indicates that

$$\mathcal{S}(\mathbf{M}_x) = \mathcal{S}(\mathbf{M}_{x,0}) \text{ and } \mathcal{S}(\mathbf{M}_y) = \mathcal{S}(\mathbf{M}_{y,0}). \tag{6}$$

Next, the original two sets of $\mathbf{X}$ and $\mathbf{Y}$ are replaced with $\mathbf{M}_{x,0}^{\mathsf{T}}\mathbf{X}$ and $\mathbf{M}_{y,0}^{\mathsf{T}}\mathbf{Y}$. Because the ranks of $\mathbf{M}_x$ and $\mathbf{M}_y$, denoted as d, are less than $\min(p, r)$, these replacements result in dimension reductions of $\mathbf{X}$ and $\mathbf{Y}$, and it will be called *initialized CCA*, hereafter. It should be noted that the initialized CCA is guaranteed to be equally informative to the standard CCA application by (6).

Although the ranks of $\mathbf{M}_{x,0}^{\mathsf{T}}\mathbf{X}$ and $\mathbf{M}_{y,0}^{\mathsf{T}}\mathbf{Y}$ are theoretically less than $\min(p,r)$, $\mathbf{M}_{x,0}$ and $\mathbf{M}_{y,0}$ are still $p \times r$ and $r \times p$ matrices, respectively. Also, $\mathbf{M}_{x,0}$ and $\mathbf{M}_{y,0}$ may not be satisfied with the orthonormality required in the standard CCA. Therefore, the CCA dimension reductions are completed by the standard CCA application with $\mathbf{M}_{x,0}^{\mathsf{T}}\mathbf{X}$ and $\mathbf{M}_{y,0}^{\mathsf{T}}\mathbf{Y}$, which yields the final canonical coefficient matrices of $\mathbf{M}_x$ and $\mathbf{M}_y$. This procedure will be called *finalized CCA*. To perform the finalized CCA in practice, we need to have that d is relatively smaller than n. It is not heavy in practice, because this is the main reason for dimension reduction.

The proposed two-step CCA procedure done by the initialized and finalized CCAs will be called *seeded canonical correlation analysis*. Two implementations of the seeded CCA will be presented in next section.

3.2. Implementation

Recall that $\mathbf{M}_{x,0}$ and $\mathbf{M}_{y,0}$ are resulted from the initialized CCA application. The replacements of $\mathbf{X}$ and $\mathbf{Y}$ by $\mathbf{M}_{x,0}^{\mathsf{T}}\mathbf{X}$ and $\mathbf{M}_{y,0}^{\mathsf{T}}\mathbf{Y}$ cause initial dimension reductions, which are theoretically equivalent to the standard CCA. In implementation perspective, however, the dimensions of $\boldsymbol{\Sigma}_{xy}$ and $\boldsymbol{\Sigma}_{yx}$ used in the initialized CCA can affect the accuracies in estimation of $\mathbf{M}_{x,0}$ and $\mathbf{M}_{y,0}$ and numerical stability in computation. Hence, if the dimensions of $\boldsymbol{\Sigma}_{xy}$ and $\boldsymbol{\Sigma}_{yx}$ are large, $\boldsymbol{\Sigma}_{xy}$ and $\boldsymbol{\Sigma}_{yx}$ should be adjusted small enough to yield $\mathbf{M}_{x,0}$ and $\mathbf{M}_{y,0}$ equally informative to the standard CCA. Depending on the size of $\min(p,r)$, two versions of seeded CCA are proposed.

3.2.1. *Case 1:* $\min(p,r)$ *fairly small*

The first case is that $\min(p,r)$ is fairly smaller than n. For simplicity and without loss of generality, we assume that $r < p$, and the maximum number of pairs of canonical variates is equal to r. Then the initialized CCA is carried out with $\hat{\boldsymbol{\Sigma}}_{xy}$ and $\hat{\boldsymbol{\Sigma}}_{yx}$ as the seed matrices, and $\mathbf{X}$ alone is replaced with $\hat{\mathbf{M}}_{x,0}^{\mathsf{T}}\mathbf{X}$. Next, by the finalized CCA with $\hat{\mathbf{M}}_{x,0}^{\mathsf{T}}\mathbf{X}$ and the original $\mathbf{Y}$, the CCA reduction is completed.

3.2.2. *Case 2:* $\min(p,r)$ *large compared with n*

The second case is that $\min(p,r)$ is not fairly small compared with n, or even larger than n. Then, as the seed matrices in the initialized CCA, the sets of the eigenvectors corresponding to m largest eigenvalues of $\hat{\boldsymbol{\Sigma}}_{xy}$ and $\hat{\boldsymbol{\Sigma}}_{yx}$ are used. The verification of these replacements is summarized in the next lemma.

Lemma 1
Assume that $d < \min(p,r)$, where d stand for the true ranks of $\text{rank}(\mathbf{M}_x)$ and $\text{rank}(\mathbf{M}_y)$. Let $\boldsymbol{\nu}_x \in \mathbb{R}^{p\times d}$ and $\boldsymbol{\nu}_y \in \mathbb{R}^{r\times d}$ represent the eigenvectors corresponding to the d largest eigenvalues of $\boldsymbol{\Sigma}_{xy}$ and $\boldsymbol{\Sigma}_{yx}$, respectively. Then, $\mathcal{S}(\mathbf{M}_x) = \mathcal{S}(\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\nu}_x)$ and $\mathcal{S}(\mathbf{M}_y) = \mathcal{S}(\boldsymbol{\Sigma}_y^{-1}\boldsymbol{\nu}_y)$.

Proof: We will prove only that $\mathcal{S}(\mathbf{M}_x) = \mathcal{S}(\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\nu}_x)$, because, by applying the same arguments, $\mathcal{S}(\mathbf{M}_y) = \mathcal{S}(\boldsymbol{\Sigma}_y^{-1}\boldsymbol{\nu}_y)$ is established. By relation (5), $\text{rank}(\boldsymbol{\Sigma}_x^{-1}\boldsymbol{\Sigma}_{xy}) = d$, which directly implies that $\text{rank}(\boldsymbol{\Sigma}_{yx}) = d$. Therefore, we have $\mathcal{S}(\boldsymbol{\Sigma}_{yx}) = \mathcal{S}(\boldsymbol{\nu}_x)$ by the construction of $\boldsymbol{\nu}_x$. Then we have $\mathcal{S}(\mathbf{M}_x) = \boldsymbol{\Sigma}_x^{-1}\mathcal{S}(\boldsymbol{\Sigma}_{yx}) = \boldsymbol{\Sigma}_x^{-1}\mathcal{S}(\boldsymbol{\nu}_x)$. This completes the proof.□

Denote the two sets of the eigenvectors corresponding to the m largest eigenvalues of $\hat{\boldsymbol{\Sigma}}_{xy}$ and $\hat{\boldsymbol{\Sigma}}_{yx}$ as $\hat{\boldsymbol{\nu}}_x \in \mathbb{R}^{p\times m}$ and $\hat{\boldsymbol{\nu}}_y \in \mathbb{R}^{r\times m}$, respectively. Through the initialized CCA with $\hat{\boldsymbol{\nu}}_x$ and $\hat{\boldsymbol{\nu}}_y$ as seed matrices, $\hat{\mathbf{M}}_{x,0}$ and $\hat{\mathbf{M}}_{y,0}$ are constructed, and $\hat{\mathbf{M}}_{x,0}^{\mathsf{T}}\mathbf{X}$ and $\hat{\mathbf{M}}_{y,0}^{\mathsf{T}}\mathbf{Y}$ replace the original sets of $\mathbf{X}$ and $\mathbf{Y}$. Again, the CCA reduction is completed by the finalized CCA with $\hat{\mathbf{M}}_{x,0}^{\mathsf{T}}\mathbf{X}$ and $\hat{\mathbf{M}}_{y,0}^{\mathsf{T}}\mathbf{Y}$.

Here, it is important to choose m large enough to have $\mathcal{S}(\hat{\boldsymbol{\Sigma}}_{yx}) = \mathcal{S}(\hat{\boldsymbol{\nu}}_x)$ and $\mathcal{S}(\hat{\boldsymbol{\Sigma}}_{xy}) = \mathcal{S}(\hat{\boldsymbol{\nu}}_y)$ but small enough to estimate $\mathbf{M}_x$ and $\mathbf{M}_y$ more accurately and to have numerical stability. To select a proper value of m without avoiding complexity, we adopt two simple and widely accepted ways in practice. One is graphical determination by a scree plot for eigenvalue of $\hat{\boldsymbol{\Sigma}}_{xy}$, and the other is the number of eigenvalues whose sum is to cover 60% or above of the total variations of $\hat{\boldsymbol{\Sigma}}_{xy}$.

4. NUMERICAL STUDIES AND DATA ANALYSIS

4.1. Numerical studies

To confirm the effectiveness of the proposed method, three simulation examples are presented. In all examples, we consider $n = 100$, and the number of iteration is always 100. The termination index u^* is fixed at 3 for simplicity.

Example 1
The first example is the case that $\min(p,r)$ is fairly small. Define $\boldsymbol{\eta}$ to have its first and last of 20% of elements equal to $(0.4)^{-1/2}$ and all other elements are equal to 0. The first set $\mathbf{X}$ of p variables was generated from $MN(\mu, \boldsymbol{\Sigma})$, where MN stands for multivariate normal distribution, $\mu = 0$ and $\boldsymbol{\Sigma} = \text{diag}(\lambda)$. In the example, we considered either 10 or 100 for p and two choices for λ: $\lambda_1 = (2/3)\text{diag}(2,\ldots,2,1,\ldots,1)^{\mathsf{T}}$ with equal multiplicity between 1 and 2 and $\lambda_2 = (2/p+1)(p,(p-1),\ldots,1)^{\mathsf{T}}$.

The second set $\mathbf{Y}$ was constructed in three different cases: (1) $Y_1 = \boldsymbol{\eta}^{\mathsf{T}}\mathbf{X} + \varepsilon_1$ and $Y_2 = \boldsymbol{\eta}^{\mathsf{T}}\mathbf{X} + \varepsilon_2$; (2) $Y_1 = \boldsymbol{\eta}^{\mathsf{T}}\mathbf{X} + \varepsilon_1$ and $Y_2 = (\boldsymbol{\eta}^{\mathsf{T}}\mathbf{X})^3 + \varepsilon_2$; (3) $Y_1 = (\boldsymbol{\eta}^{\mathsf{T}}\mathbf{X})^3 + \varepsilon_1$ and $Y_2 = (\boldsymbol{\eta}^{\mathsf{T}}\mathbf{X})^3 + \varepsilon_2$, where $\varepsilon_i \overset{i.i.d}{\sim} N(0,1) \perp\!\!\!\perp \mathbf{X}$ for $i = 1,2$.

Because $\min(p,r) = 2$ in the example, the estimation of $\mathbf{M}_x$ was focused rather than that of $\mathbf{M}_y$. Then $\mathbf{M}_x$ was estimated by the first case of the seeded CCA with $\hat{\boldsymbol{\Sigma}}_{xy} \in \mathbb{R}^{p\times 2}$, which resulted in $\hat{\mathbf{M}}_x \in \mathbb{R}^{p\times 2}$. To measure the accuracy of the estimation of $\mathbf{M}_x$, the averages of $|r|$s were computed. The notation $|r|$ stands for the absolute value of R^2 from a regression of $\boldsymbol{\eta}^{\mathsf{T}}\mathbf{X}|\hat{\mathbf{M}}_x^{\mathsf{T}}\mathbf{X}$. If $\hat{\mathbf{M}}_x$ estimates $\boldsymbol{\eta}$ well, the averages should be close to one. The results are presented in Table I.

Example 2
In the example, the dimension of $\mathbf{X}$ was fixed at $p = 500$ and was generated from $MN(0,\boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = (2/3)\text{diag}(2,\cdots,2,1,\ldots,1)^{\mathsf{T}}$ with equal multiplicity between 1 and 2.

For the second set $\mathbf{Y}$, its dimension r varied among 10, 20, 50, 100, and 200. Define that $\mu_1 = 1+X_1$ and $\mu_2 = 2+X_2$. Then $\mathbf{Y}$ was generated as follows: $Y_1 = \mu_1+\varepsilon_1$; $Y_2 = \mu_2+\varepsilon_2$; $Y_3 = \mu_1+\mu_2+\varepsilon_3$; $Y_4 = \mu_1-\mu_2+\varepsilon_4$; $Y_r = \varepsilon_r$ for $r \geq 5$, where $\boldsymbol{\varepsilon} = (\varepsilon_1,\ldots,\varepsilon_r)^{\mathsf{T}}$ was independently generated from the standard normal distribution, and $\boldsymbol{\varepsilon} \perp\!\!\!\perp \mathbf{X}$.

In the example, $\mathbf{X}$ is associated with $\mathbf{Y}$ only through X_1 and X_2. Therefore, $\mathbf{M}_x$ is equal to $\boldsymbol{\eta} = \{(1,0,\ldots,0)^{\mathsf{T}},(0,1,0,\ldots,0)^{\mathsf{T}}\}$. On the other hand, $\mathbf{Y}$ is associated with $\mathbf{X}$ only through μ_1 and μ_2. Hence, $\mathbf{M}_y$ is equal to $\boldsymbol{\phi} = \{(1,0,\ldots,0)^{\mathsf{T}},(0,1,0,\ldots,0)^{\mathsf{T}}\}$.

Table I. Averages of $|r|$s computed from $\boldsymbol{\eta}^{\mathsf{T}}\mathbf{X}|\hat{\mathbf{M}}_x^{\mathsf{T}}\mathbf{X}$ for Example 1

	λ_1		λ_2	
models	$p = 10$	$p = 500$	$p = 10$	$p = 500$
$Y_1 = \beta^{\mathsf{T}}\mathbf{X} + \varepsilon_1, Y_2 = \beta^{\mathsf{T}}\mathbf{X} + \varepsilon_2$	0.9789	0.8198	0.9992	0.9421
$Y_1 = \beta^{\mathsf{T}}\mathbf{X} + \varepsilon_1, Y_2 = (\beta^{\mathsf{T}}\mathbf{X})^3 + \varepsilon_2$	0.9748	0.7915	0.9709	0.8025
$Y_1 = (\beta^{\mathsf{T}}\mathbf{X})^3 + \varepsilon_1, Y_2 = (\beta^{\mathsf{T}}\mathbf{X})^3 + \varepsilon_2$	0.9841	0.855	0.9898	0.928

Table II. Averages of $|r|$s for Example 2

r	$\boldsymbol{\eta}_1^{\mathsf{T}}\mathbf{X}\|\hat{\mathbf{M}}_x^{\mathsf{T}}\mathbf{X}$	$\boldsymbol{\eta}_2^{\mathsf{T}}\mathbf{X}\|\hat{\mathbf{M}}_x^{\mathsf{T}}\mathbf{X}$	$\boldsymbol{\phi}_1^{\mathsf{T}}\mathbf{Y}\|\hat{\mathbf{M}}_y^{\mathsf{T}}\mathbf{Y}$	$\boldsymbol{\phi}_2^{\mathsf{T}}\mathbf{Y}\|\hat{\mathbf{M}}_y^{\mathsf{T}}\mathbf{Y}$
10	0.8979	0.8901	0.8929	0.8852
20	0.8923	0.8924	0.8817	0.8831
50	0.8854	0.8862	0.8632	0.8637
100	0.8714	0.8676	0.8356	0.8312
200	0.8332	0.8274	0.7801	0.7828

Table III. Averages of $|r|$s for Example 3

r	$\boldsymbol{\eta}_1^{\mathsf{T}}\mathbf{X}\|\hat{\mathbf{M}}_x^{\mathsf{T}}\mathbf{X}$	$\boldsymbol{\eta}_2^{\mathsf{T}}\mathbf{X}\|\hat{\mathbf{M}}_x^{\mathsf{T}}\mathbf{X}$	$\boldsymbol{\phi}_1^{\mathsf{T}}\mathbf{Y}\|\hat{\mathbf{M}}_y^{\mathsf{T}}\mathbf{Y}$	$\boldsymbol{\phi}_2^{\mathsf{T}}\mathbf{Y}\|\hat{\mathbf{M}}_y^{\mathsf{T}}\mathbf{Y}$
10	0.8943	0.9912	0.9992	0.8927
20	0.8918	0.9910	0.9992	0.8905
50	0.8827	0.9911	0.9992	0.8810
100	0.8623	0.9910	0.9992	0.8606
200	0.8307	0.9908	0.9988	0.8341

Because $\min(p, r)$ is relatively large, the matrices of $\hat{\boldsymbol{\nu}}_x \in \mathbb{R}^{p\times 2}$ and $\hat{\boldsymbol{\nu}}_y \in \mathbb{R}^{r\times 2}$ were used in the initialized CCA, where $\hat{\boldsymbol{\nu}}_x$ and $\hat{\boldsymbol{\nu}}_y$ were sets of the eigenvectors corresponding to the first two largest eigenvalues of $\hat{\boldsymbol{\Sigma}}_{xy}$ and $\hat{\boldsymbol{\Sigma}}_{yx}$, respectively.

Let $|r_i^x|$ stand for the absolute value of R^2 from a regression of $\boldsymbol{\eta}_i^{\mathsf{T}}\mathbf{X}|\hat{\mathbf{M}}_x^{\mathsf{T}}\mathbf{X}$ for $i = 1, 2$. To measure the accuracy of the estimation of $\mathbf{M}_x$, the averages of $|r_i^x|$s were calculated for $i = 1, 2$. Similarly, for $\mathbf{M}_y$, we computed the averages of $|r_i^y|$s from a regression of $\boldsymbol{\phi}_j^{\mathsf{T}}\mathbf{X}|\hat{\mathbf{M}}_y^{\mathsf{T}}\mathbf{X}$ for $j = 1, 2$. The results are summarized in Table II.

Example 3

Example 2 was slightly modified by setting $\mu_1 = 1.5(5 + |X_1|)(X_2 + X_3)$ and $\mu_2 = 1 + X_1$, while the rest of the setting in Example 2 remained the same. The purpose of the example is to study how the existence of non-linear relationship in $\mathbf{X}$ affects the estimation of $\mathbf{M}_x$ and $\mathbf{M}_y$. In the example, $\mathbf{X}$ is associated with $\mathbf{Y}$ only through X_1 and $X_2 + X_3$, while the association of $\mathbf{Y}$ with $\mathbf{X}$ remains the same as Example 2. Hence, $\boldsymbol{\eta}$ alone is changed to $\{(1, 0, \cdots, 0)^{\mathsf{T}}, (0, 1, 1, 0, \cdots, 0)^{\mathsf{T}}\}$. The same statistics summarize the simulation studies, and they are reported in Table III.

According to Table I, the performances of seeded CCA seem quite good for both $p = 10$ and 500. And, Table II shows that the performances are all notable, although the performances get worse as r increases. Compared with Example 2, the good performances in the estimation of both $\mathbf{M}_x$ and $\mathbf{M}_y$ are observed.

Overall, various numerical studies including these three examples support the proposed method well, and there will be no critical issues in use in practice.

4.2. Real data application

4.2.1. Near-infrared spectroscopy of biscuit doughs data

To illustrate the seeded CCA in practice, near-infrared spectroscopy of biscuit doughs data (Brown *et al.* [11]; NIR) was adopted. Quantitative NIR spectroscopy is used to analyze diverse compositions in food, drink, pharmaceutical products, and petrochemicals. The experiment of the biscuit dough data set was conducted to measure the suitability of NIR spectroscopy for analyzing the composition of biscuit. The percentages of four components of biscuits made by standard recipe-fat, sucrose, dry flour, and water were calculated from 72 biscuit samples. One set of variables, saying $\mathbf{Y} \in \mathbb{R}^4$, is composed of this four ingredients.

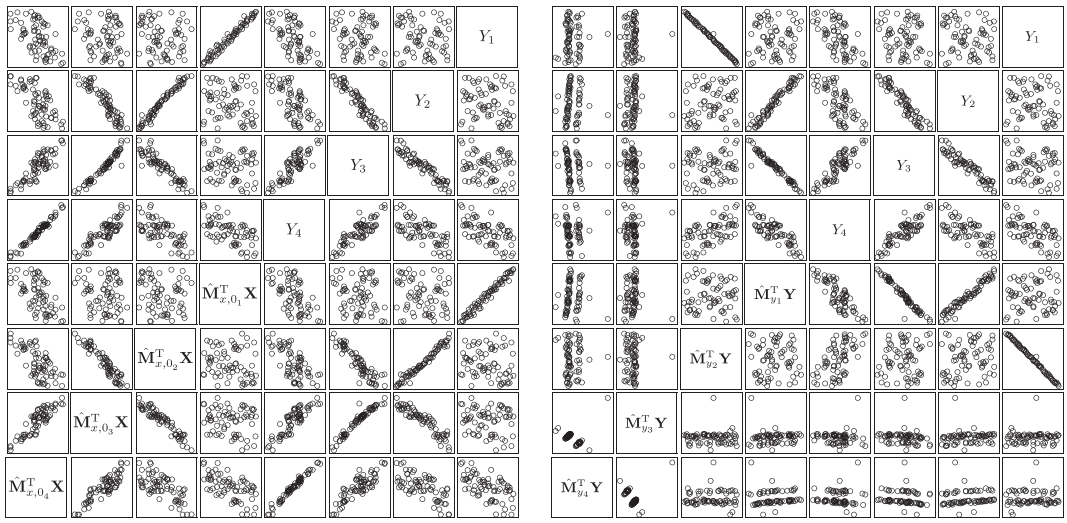
For each sample, there was a wavelength observed by spectroscopy and 700 different points of it had been measured, range from 1100 to 2498 nanometers (NM) at an interval of 2 nm. From this data, the first 140 and the last 49 wavelengths were removed in Brown *et al.* [11], because those figures hardly seemed to contain the useful information, and increased the interval to 4 nm. Consequently, the wavelength had its bounds from 1380 to 2400nm with 256 points, representing another set of variables $\mathbf{X} \in \mathbb{R}^{256}$.

The data were obtained from `ppls-package` of statistical language `R` and is named `cookie`. Because the 23th and 61st samples in the data set were suspected as outliers, they were deleted from the data set before analysis.

Due to $p = 256 > n = 72$, it is clear that the standard CCA is not applicable. Because $r = 4$ is reasonably small compared with n, the first case of the seeded CCA in Section 3.2.1 was implemented. Then $\hat{\mathbf{M}}_{x,0}$ was constructed through taking $\hat{\boldsymbol{\Sigma}}_{xy} \in \mathbb{R}^{256\times 4}$ as a seed matrix. The proper number of projection turned out to be $u^* = 2$ with simple graphical determination regarding the change in $\hat{\mathbf{R}}_u$ and $\hat{\mathbf{R}}_{u-1}$ (not reported).

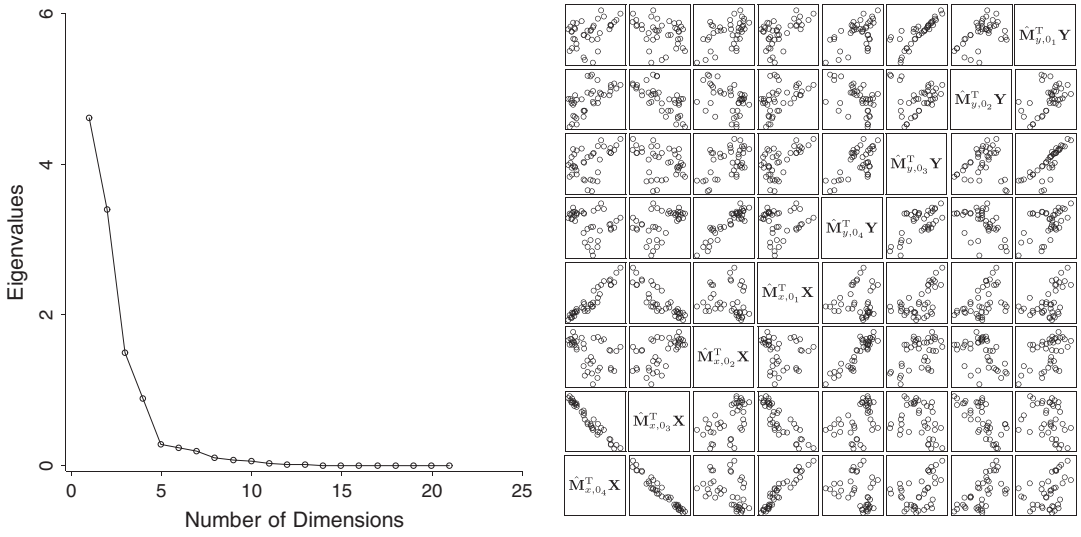
This initialized CCA resulted in the replacement of the original 256-dimensional $\mathbf{X}$ by the four-dimensional $\hat{\mathbf{M}}_{x,0}^{\mathsf{T}}\mathbf{X}$. For notational conveniences, let $\hat{\mathbf{M}}_{x,0_i}$ and $\hat{\mathbf{M}}_{x,0_{-i}}$ stand for the ith column of $\hat{\mathbf{M}}_{x,0}$ and the $\hat{\mathbf{M}}_{x,0}$ after removing the ith column, respectively. Similarly, $\hat{\mathbf{M}}_{y,0_i}$ and $\hat{\mathbf{M}}_{y,0_{-i}}$ are defined.

Next, to complete the CCA dimension reduction, the relationship between $\hat{\mathbf{M}}_{x,0}^{\mathsf{T}}\mathbf{X}$ and $\mathbf{Y}$ was inspected through in a scatterplot matrix in Figure 1a. The plot indicates that two pairs of $(\hat{\mathbf{M}}_{x,0_1}^{\mathsf{T}}\mathbf{X}, Y_1)$ and $\{\hat{\mathbf{M}}_{x,0_{-1}}^{\mathsf{T}}\mathbf{X}, (Y_2, Y_3, Y_4)\}$ have clear separation by the relations within each pair and between pairs. Along with this information, the finalized CCA was carried out with $\hat{\mathbf{M}}_{x,0}^{\mathsf{T}}\mathbf{X}$ and $\mathbf{Y}$, and denote $\hat{\mathbf{M}}_x \in \mathbb{R}^{4\times 4}$ and $\hat{\mathbf{M}}_y \in \mathbb{R}^{4\times 4}$ as the finalized canonical coefficient matrices for $\mathbf{X}$ and $\mathbf{Y}$. To withdraw more information on the relationship between $\hat{\mathbf{M}}_y^{\mathsf{T}}\mathbf{Y}$ and $\mathbf{Y}$, the scatterplot matrix in Figure 1(b) was constructed. From the plot, it can be observed that the first canonical variate $\hat{\mathbf{M}}_{y_1}^{\mathsf{T}}\mathbf{Y}$ has strong linear relationships with Y_2, Y_3 and Y_4, while the second one $\hat{\mathbf{M}}_{y_2}^{\mathsf{T}}\mathbf{Y}$ can be thought of as Y_1 itself.



a. $\hat{\mathbf{M}}_{x,0}^{\mathrm{T}}\mathbf{X}$ and $\mathbf{Y}$ b. $\hat{\mathbf{M}}_{y}^{\mathrm{T}}\mathbf{Y}$ and $\mathbf{Y}$

Figure 1. Scatterplot matrices for NIR data.



a. Scree Plot of $\hat{\mathbf{\Sigma}}_{xy}$ b. Scatterplot matrix for initialized CCA

Figure 2. Plots for CCA in nutrimouse data.

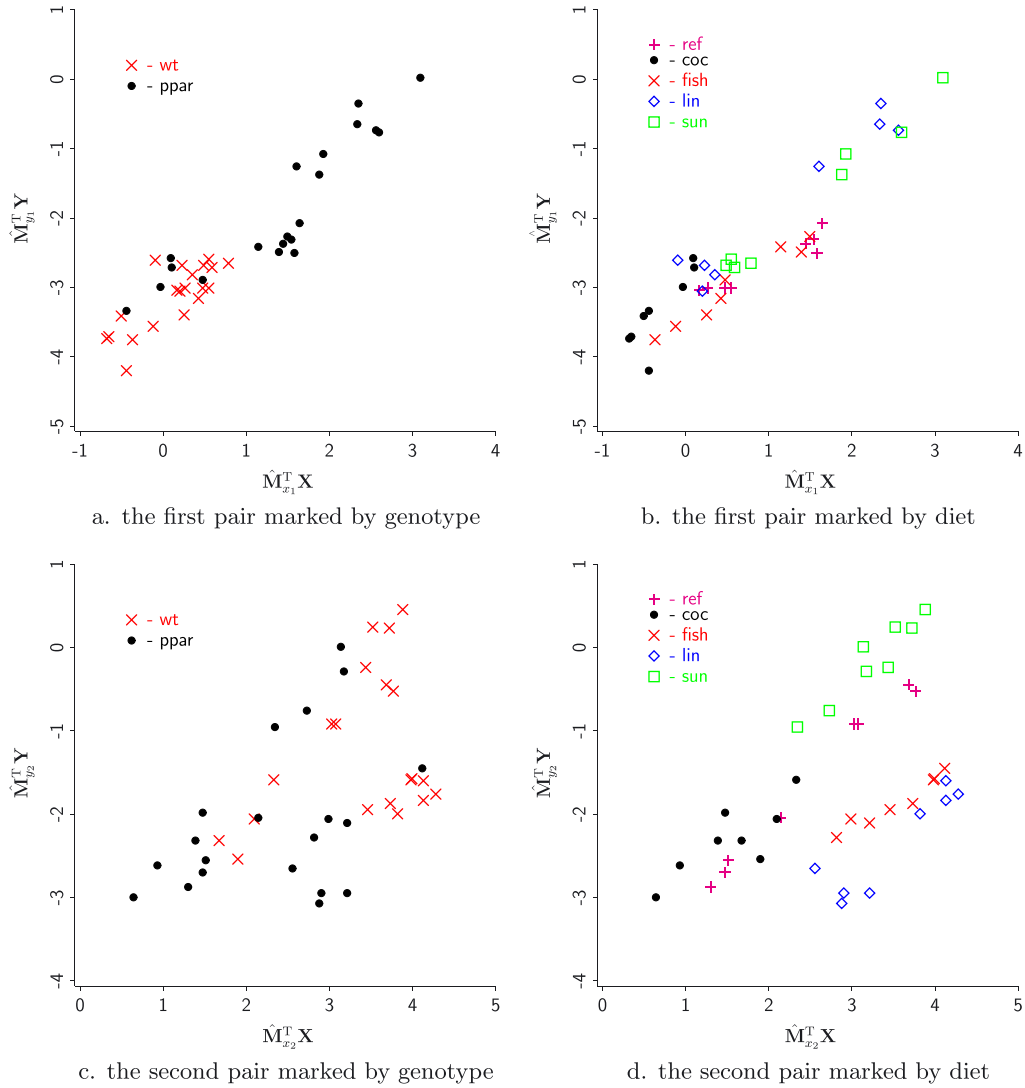
By the discussion, the two pairs of canonical variates of $(\hat{\mathbf{M}}_{x_1}^{\mathrm{T}}\mathbf{X}, \hat{\mathbf{M}}_{y_1}^{\mathrm{T}}\mathbf{Y})$ and $(\hat{\mathbf{M}}_{x_2}^{\mathrm{T}}\mathbf{X}, \hat{\mathbf{M}}_{y_2}^{\mathrm{T}}\mathbf{Y})$ should be concluded to be sufficient to summarize the relation between $\mathbf{Y}$ and $\mathbf{X}$. If one can do further analysis to investigate how the percentages of the four ingredient is changed given the wavelengths, s/he can apply multivariate linear regression of $(\hat{\mathbf{M}}_{y_1}^{\mathrm{T}}\mathbf{Y}, \hat{\mathbf{M}}_{y_2}^{\mathrm{T}}\mathbf{Y})|(\hat{\mathbf{M}}_{x_1}^{\mathrm{T}}\mathbf{X}, \hat{\mathbf{M}}_{x_2}^{\mathrm{T}}\mathbf{X})$.

4.2.2. *Nutrimouse data*

For the illustration purpose of the second case in Section 3.2.2, `nutrimouse` data (Martin *et al.* [12]) was used. The data were collected from a nutrigenomic study in 40 mouse (n = 40) by which investigated the effects of five regimens with contrasted fatty acid compositions on liver lipids and hepatic gene expression.

The following variables were used for two sets of variables: one set $\mathbf{X} \in \mathbb{R}^{120}$ was expressions of 120 genes measured in liver cells, acquired through microarray technology and selected among about 30 000 as potentially useful in the nutrition study. The other set $\mathbf{Y} \in \mathbb{R}^{21}$ was concentrations of 21 hepatic fatty acids (FA) measured by gas chromatography. Additionally, the 40 mouse are cross-classified based on two factors of genotype an diet: genotype: wild-type (WT) mice and PPAR*alpha* deficient mice (PPAR); diet: corn and colza oils (50/50, REF), hydrogenated coconut oil for a saturated FA diet (COC), sunflower oil for ω6 FA-rich diet (SUN), linseed oil for ω3-rich diet (LIN), and corn/colza/enriched fish oils (42.5/42.5/15, FISH). The data are publicly available in the `mixOmics-package` of statistical language `R`.

Because $\min(p, r) = 21$ and was relatively larger compared with $n = 40$, $\hat{\mathbf{v}}_x \in \mathbb{R}^{120\times m}$ and $\hat{\mathbf{v}}_y \in \mathbb{R}^{21\times m}$ of the eigenvectors corresponding to the m largest eigenvalues $\hat{\mathbf{\Sigma}}_{yx}$ and $\hat{\mathbf{\Sigma}}_{xy}$ were used as the seed matrices in the initialized CCA. By inspecting the scree plot of $\hat{\mathbf{\Sigma}}_{xy}$, reported in Figure 2(a), it was determined that $m = 4$.



a. the first pair marked by genotype

b. the first pair marked by diet

c. the second pair marked by genotype

d. the second pair marked by diet

Figure 3. Summary plots for nutrimouse data

Additionally, the sums of the first two largest and the first four largest eigenvalues of $\hat{\boldsymbol{\Sigma}}_{xy}$ accounts for 90.1% and 99.4% of the total variablity, respectively. With $\hat{\nu}_x$ and $\hat{\nu}_y$ setting $m = 4$, the initialized CCA yielded $\hat{\mathbf{M}}_{x,0}^{\mathsf{T}}\mathbf{X}$ and $\hat{\mathbf{M}}_{y,0}^{\mathsf{T}}\mathbf{Y}$, where $\hat{\mathbf{M}}_{x,0} \in \mathbb{R}^{120\times 4}$ and $\hat{\mathbf{M}}_{y,0} \in \mathbb{R}^{21\times 4}$. Before the finalized CCA, the relation of $\hat{\mathbf{M}}_{x,0}^{\mathsf{T}}\mathbf{X}$ and $\hat{\mathbf{M}}_{y,0}^{\mathsf{T}}\mathbf{Y}$ was investigated through a scatter plot matrix in Figure 2b. The plot indicates that $(\hat{\mathbf{M}}_{x,0_3}, \hat{\mathbf{M}}_{x,0_4})^{\mathsf{T}}\mathbf{X}$ and $(\hat{\mathbf{M}}_{y,0_3}, \hat{\mathbf{M}}_{y,0_4})^{\mathsf{T}}\mathbf{Y}$ are redundant, and hence $\hat{\mathbf{M}}_{x,0}$ and $\hat{\mathbf{M}}_{y,0}$ were additionally reduced to $(\hat{\mathbf{M}}_{x,0_1}, \hat{\mathbf{M}}_{x,0_2}) \in \mathbb{R}^{120\times 2}$ and $(\hat{\mathbf{M}}_{y,0_1}, \hat{\mathbf{M}}_{y,0_2}) \in \mathbb{R}^{21\times 2}$. So, it was concluded that two pairs of the canonical covariates could represent the relation of $\mathbf{X}$ and $\mathbf{Y}$. At last, to have $\hat{\mathbf{M}}_x^{\mathsf{T}}\mathbf{X}$ and $\hat{\mathbf{M}}_y^{\mathsf{T}}\mathbf{Y}$, the CCA reduction was completed by the finalized CCA with $(\hat{\mathbf{M}}_{x,0_1}, \hat{\mathbf{M}}_{x,0_2})^{\mathsf{T}}\mathbf{X}$ and $(\hat{\mathbf{M}}_{y,0_1}, \hat{\mathbf{M}}_{y,0_2})^{\mathsf{T}}\mathbf{Y}$.

In Figure 3(a)–(d), the scatter plots of the first and second pairs of the finalized canonical variates marked by genotype and diet of mouse used are reported as summary. According to Figure 3(a) and (c), the first canonical covariate is distinguished better by genotype than the second one, while the latter is better grouped for diet than the former in Figures 3(b) and (d).

5. Conclusion

Canonical correlation analysis is one of popular statistical methodologies in multivariate analysis, especially, in studying relation of two sets of variables. However, when sample sizes are smaller than the maximum of the dimensions of two sets of variables, it is plausible to implement canonical correlation analysis, because the sample covariance matrices are not invertible in practice.

In this article, by adopting seeded dimension reduction, a two-step canonical correlation procedure is proposed and is called seeded canonical correlation analysis. In the first step, the dimensions of the original two sets are initially reduced without losing information on the canonical correlation analysis perspective. Then, in the second step, the standard canonical correlation application with the initially reduced sets is carried out to complete the dimension reduction. Numerical studies confirm the approach, and two real data analyses are presented for illustration of the purpose.

It is believed that our works can make canonical correlation analysis more fruitful as statistical methodologies in

high-throughput data analysis. The codes for the seeded canonical correlation analysis are publicly available on the following website of the authors:
http://home.ewha.ac.kr/~yjkstat/publication.html.

Acknowledgements

The authors are also grateful to the associate editor and the referee for many insightful and helpful comments. The corresponding author, Jae Keun Yoo appreciates the Department of Statistics, University of Washington to wrap this research up during the visit in summer, 2013. For Jae Keun Yoo (the corresponding author), 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-2014R1A2A1A11049389). For Yunju Im and Heyin Gang, this work was supported by the BK21 Plus Project through the National Research Foundation of Korea (NRF) funded by the Korean Ministry of Education (22A20130011003). Currently, the author, Yunju Im is a PhD candidate of the graduate program at the Department of Statistics and Actuarial Science, University of Iowa.

REFERENCES

1. Johnson RA, Wichern DW. Applied Multivariate Statistical Analysis (6th edn). Pearson Prentice Hall: New Jersey, USA, 2007, 539–574.
2. Parkhomenko E, Tritchler D, Beyene J. Genome-wide sparse canonical correlation of gene expression with genotypes. *BMC Proc.* 2007; **1**: S119.
3. Waaijenborg S, Verselewel de Witt Hamer P, Zwinderman A. Quantifying the association between gene expressions and DNA-markers by penalized canonical correlation analysis. *Stat. Appl. Genet. Mol.* 2008; **7**: 3.
4. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J. R. Statist. Soc. B* 2005; **67**: 301–320.
5. Le Cao K, Pascal M, Robert-Granie C, Philippe B. Sparse canonical methods for biological data integration: application to a cross-platform study. *BMC Bioinformatics* 2009; **10**: 34.
6. Witten D, Tibshirani R, Hastie T. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics* 2009; **10**: 515–534.
7. Cook RD, Li B, Chiaromonte F. Dimension reduction in regression without matrix inversion. *Biometrika* 2007; **94**: 569–584.
8. Searle A. Matrix algebra useful for statistics. Wiley: New York, USA, 1982, 212–213.
9. Yoo JK. Advances in seeded dimension reduction: bootstrap criteria and extensions. *Comput. Stat. Data. An.* 2013; **60**: 70–79.
10. Lee K, Yoo JK. Canonical correlation analysis through linear modeling. *Aust. Nz. J. Stat.* 2014; **56**: 59–72.
11. Brown PJ, Fearn T, Vannucci M. Bayesian wavelet regression on curves with applications to a spectroscopic calibration problem. *J. Am. Stat. Assoc.* 2001; **96**: 398–408.
12. Martin PGP, Guillou H, Lasserre F, Djean S, Lan A, Pascussi J-M, San Cristobal M, Legrand P, Besse P, Pineau T. Novel aspects of PPAR-mediated regulation of lipid and xenobiotic metabolism revealed through a multrigenomic study. *Hepatology* 2007; **54**: 767–777.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web-site.