Novel Kernel Orthogonal Partial Least Squares for Dominant Sensor Data Extraction

BO-WEI CHEN
Department of Electrical Engineering, National Sun Yat-sen University, Kaohsiung 80424, Taiwan
e-mail: bo-wei.chen@mail.nsysu.edu.tw

This work was supported in part by the Ministry of Science and Technology, Taiwan, under Grant 107-2218-E-110-013-MY3, and in part by the 2019 NVIDIA Data Science Grants.

ABSTRACT Orthogonal Partial Least Squares (OPLS) methods are aimed at finding the dominant factors from predictor variables that can maximize cross-covariance between the factors themselves and response variables while a high correlation between them should also be satisfied at the same time. Compared with discriminant analysis like Linear Discriminant Analysis, OPLS simultaneously considers covariance maximization and data fitting. However, unlike discriminant analysis that focuses on between-group discriminability, OPLS concentrates on cross-covariance that already contains no discriminant information. This deepens the difficulty of finding effective dominant factors. To rectify such a drawback of OPLS, this study proposes 1) successively orthogonal deflation in constrained noisy subspace and 2) isotropic space transform for enhancing OPLS. The former explores successively orthogonal projective vectors in subspace and iteratively updates the weighted signal space. The latter converts the dimensions with unequal influences into those with equal ones for correcting distortions. The two proposed rectifications are implemented in three types of Maximum Covariance Analysis (MCA) for examining the gradually changing functionalities, respectively — i) Successive Subspace-MCA, ii) Isotropic Subspace-MCA, and iii) Successive Isotropic Subspace-MCA. Experiments on open datasets were carried out to compare the proposed approaches with the baseline. The experimental results showed that the proposed rectifications maximized cross-covariance while fitting data well, thereby substantiating the effectiveness of the proposed idea.

INDEX TERMS Orthogonal partial least squares, partial least squares, maximum covariance analysis.

I. INTRODUCTION
In meteorological research, data captured by sensor networks usually involve multivariables. Heterogeneous data like temperature, pressure, and moisture are collected and compiled together in a dataset. These multivariate sensing data are frequently used as a source for reasoning observed atmospheric phenomena, e.g., sea surface temperature, aerosols, and precipitation. At present, a great deal of statistical research has been developed, such as Ordinary Least Squares (OLS), Factor Analysis, and Analysis of Variance. Among those statistical reasoning methods, Maximum Covariance Analysis (MCA) is widely applied for finding dominant factors from multivariate data, such that harvested weather data can be correlated with observed atmospheric phenomena. MCA discovers the maximum covariance between the dominant hidden components extracted from both predictor variables and response variables. For instance, meteorologists often conducted MCA to analyze dominant coupled variables that affected sea surface temperature [1], as shown in Fig. 1. Thus, meteorologists can discover which factors have the major influence on sea surface temperature. In statistics, there are many types of factors, such as discriminant factors and redundancy factors [2]. Unlike other factors, dominant factors particularly refer to variables with maximum variance. Compared with discriminant analysis for extracting discriminant factors (e.g., Fisher Linear Discriminant Analysis (LDA), [3], and [4]), finding dominant factors requires considerations for the tradeoff between maximum variance and high correlations [5]. The Partial Least Squares (PLS) method is an approach dedicated to such a purpose.

A. PARTIAL LEAST SQUARES
PLS regression is a variation of multilinear regression. Unlike typical regression (e.g., OLS), PLS embeds MCA into least squares concepts. Typical regression separates the hidden
component discovery from the main process, which does not particularly highlight maximum covariance between the hidden components of both predictor variables and response variables. Moreover, typical regression emphasizes finding a correlation between predictors and responses. PLS [6]–[10] jointly searches projection vectors for both independent variables and dependent responses when computing cross-covariance, so that the extracted hidden components of the independent variables and those of the dependent responses have maximum covariance. In other words, the hidden components of both predictor variables and response variables are simultaneously considered while PLS creates a link to map the hidden components of predictor variables onto those of response variables, rather than focusing on finding the hidden components on one side only.

To discover the hidden components on both sides, several studies have been devoted to PLS, including Nonlinear Iterative Partial Least Squares (NIPALS), PLS1, PLS2, SIMPLS [11], and probabilistic PLS [12], which were based on L2 norm. Recent research on PLS has evolved into other types of metrics, such as L1 norm [13], manifolds [14], [15], and correntropy [16]. In this study, L2 norm is focused. For PLS in classification problems, PLS-Discriminant Analysis (PLS-DA) is widely used, and PLS-DA is a term for PLS classification in contrast to PLS regression [17]–[19]. The response variables in PLS classification become categorical variables, or frequently called label information in pattern recognition. Literature [20] has examined the methodological difference between PLS-DA and LDA, subsequently proving that there was a connection between PLS-DA and LDA in their parameters. Despite such a connection, they are not equivalent because the objectives are basically different.

B. ORTHOGONAL PARTIAL LEAST SQUARES
Orthogonal Partial Least Squares (OPLS), or equivalently, Orthogonal Projection to Latent Structures, was derived from PLS [21], [22]. It is a special case of PLS problems. The difference between OPLS and PLS is that the hidden component analysis in OPLS is applied only to predictor variables, whereas PLS jointly works on both predictor variables and response variables. Thus, only the projection vectors from predictor variables are extracted. Like PLS regression and PLS-DA, OPLS also develops two versions. One is OPLS regression, and the other is OPLS-DA [23] for classification problems. The connection between OPLS-DA and LDA was investigated in [20]. Recent studies [24], [25] gave a clear investigation into the eigenstructures of PLS and OPLS from the perspective of eigenvalue decomposition.

For OPLS, after the hidden components of predictor variables are found, there is another process called data fitting, which employs OLS to map the hidden components onto response variables. Although data fitting involves computing the weight vector of OLS, the process requires participation of a projection matrix (i.e., a set of projection vectors). Searching such a projection matrix is actually separated from computation of OLS weight vectors. Because two processes are independent, the problem becomes how to find the maximum correlations/covariance between hidden components and response variables while hidden components should be effectively discovered at the same time. OPLS solves such a problem by calculating a cross-covariance matrix \( S_{XY} \) between predictor variables (e.g., \( \mathbf{X} = [x_1, \ldots, x_N] ; \mathbf{X} \in \mathbb{R}^M \); \( N \) is the number of samples; \( M \) signifies the dimensionality) and response variables (e.g., \( \mathbf{y} = [y_1, \ldots, y_N] ; \mathbf{y} \in \mathbb{R}^N \)) prior to the hidden component analysis. This is the fundamental requirement of typical OPLS problems that differ from discriminant analysis. Several variants [26], [27] of OPLS have been developed at present. Reference [26] utilized the objective of OPLS but integrated it into other concepts, such as Nonnegative Matrix Factorization, nonnegative constraints, orthogonal Procrustes solutions [28], [29], and sequential deflation. The purpose was to find a set of projection matrices based on those algorithms. Then, the system used those projection matrices as filter banks to generate dominant features. A similar application was developed in [30]. More recently, [27] proposed using normalized and redundancy-free measurement projection [31] based on Singular Value Decomposition to replace typical OPLS processes.

There is a clear difference between OPLS-based hidden component analysis and discriminant analysis (e.g., LDA). The former maximizes cross-covariance matrices, whereas the latter is aimed at maximizing between-group discriminability. It is worth noting that discriminant analysis cannot be directly applied to OPLS-based or PLS-based hidden

---

**FIGURE 1.** Example of sea surface temperature (SST) near the US Atlantic Ocean. Various colors indicate different SST. MCA is often conducted on meteorological data to extract which dominant factors (i.e., dominant predictor variables) affect SSTs (i.e., observed weather phenomena or response variables). Both maximum covariance and well data fitting are required in MCA. Image source: www.ospo.noaa.gov.
component analysis despite the success in discriminant pattern recognition. This is because OPLS-based hidden component analysis should maximize the cross-covariance matrix $S_{XY}$ rather than between-group matrices.

C. CHALLENGES

The first challenge is how to find dominant components to fit response variables while maximization of cross-covariance matrices is achieved. In other words, fitting data well without decreasing covariance too much is the purpose of OPLS. Discriminant analysis like feature selection, dimensional reduction (e.g., Principle Component Analysis (PCA), and LDA), and recent manifold learning (e.g., [32]) may work well for data fitting, but they did not simultaneously maximize cross-covariance matrices, which deviates the target of OPLS.

Secondly, a cross-covariance matrix is actually a mixture of within-group data and between-group data. Directly using the concept of maximizing the between-group discriminability (e.g., via discriminant analysis) is not applicable. This is because it violates the requirements of PLS and OPLS owing to different objectives.

Thirdly, unequal variance may exist in different types of predictor variables during data collection, e.g., heterogeneous data (e.g., vehicular/human/animal temperature) collected at different places (e.g., indoors/outdoors) and in various time (e.g., days/nights). Unequal distributional variance may result in biased data fitting in OPLS. Either PLS or OPLS still concentrated on original feature space that suffered from heterogeneous variance.

D. CONTRIBUTIONS

In light of the aforementioned problems, this study proposes the two types of rectifications to improve OPLS: 1) Successively orthogonal deflation in constrained noisy subspace and 2) isotropic space transform. Additionally, these two rectifications are also the innovation factors of this study. The details are elaborated as follows.

- Successively orthogonal deflation in constrained noisy subspace: In OPLS, its objective is maximization of cross-covariance matrices, and its implementation is actually trace-norm optimization of the weighted signal space (i.e., $S_{\Psi\Psi} \rightarrow S_{\Psi\Psi} S_{YY}^{-1/2}$). Such weighted signal space is computed based on the signal space (i.e., $S_{YY}$) and the noise space (i.e., $S_{\Psi\Psi}$) in terms of signal-to-noise ratios. However, the eigen-structural analysis of OPLS simply relies on the entire weighted signal space. After projection vectors are extracted, only the weighted signal space is deflated. Nevertheless, $S_{\Psi\Psi}$ is still the same. To correct such a drawback, the proposed successively orthogonal deflation updates the feature space $S_{\Psi\Psi}$ every time the weighted signal space is changed while the targeted maximization of cross-covariance matrices still remains intact.

- Isotropic space transform: Isotropic space refers to space with equal dimensional variance. As mentioned earlier in the challenges, heterogeneous data collected at different places (e.g., indoors/outdoors) and in various time (e.g., days/nights) may result in unequal distributional variance. This may subsequently cause biased estimations of cross-covariance maximization and data fitting for OPLS. Isotropic space transform corrects the distorted components in feature space affected by unequal variance and remaps them into isotropic vector space, where predictor dispersion is normalized. For multivariables, especially when heterogeneous sensors are used to capture data, unequal variance easily exists in datasets. Isotropic space transform can offset such a bias.

The rest of this paper is organized as follows. Sections II details typical OPLS. Section III then describes the proposed method. Next, Section IV summarizes the performance of the proposed method and the analytic results. Conclusions are finally drawn in Section V along with future recommendations.

II. PREVIOUS METHOD: OPLS

This section introduces typical OPLS and Kernel OPLS (KOPLS). Assume that there is a multivariate dataset $X = [x_1, \ldots, x_N]$ along with its response variables $y = [y_1, \ldots, y_N]$, where $X \in \mathbb{R}^M$, $y \in \mathbb{R}^1$, $M$ represents the number of dimensions, and $N$ denotes the number of samples. Let $J$ denote the degree of intrinsic space when data $X$ are transformed into $\Psi$ by a kernel function $\varphi()$, i.e., $X \in \mathbb{R}^M \rightarrow \Psi \in \mathbb{R}^J$. Therefore, KOPLS is expressed as

$$V^T \Psi = \Phi + B,$$

$$y = \Phi^T u + B',$$

where $T$ means the transpose operator, $V$ is an unknown $J \times D$ projection matrix (i.e., a set of projection vectors), $\Phi$ signifies dimensionally reduced data, and $D$ is the number of hidden dimensions $(D \leq J)$. Moreover, $B$ and $B'$ are residuals. To represent KOPLS in a concise form, (1) becomes

$$\min_u E_{KOPLS}(u) = \min_u \left\{ ||y - V^T \Psi u||_2^2 \right\},$$

where $|| \cdot ||^2$ calculates the $L_2$-norm distance. The objective of KOPLS is to minimize the least squares errors $E$ while finding a projection matrix $V$ and a $D$-by-1 weight vector $u$. A $L_2$-norm regularized version is obtained by adding a penalty term $\rho ||u||^2$ to the equation, where $\rho$ indicates the ridge parameter. Furthermore, $V$ should satisfy the following condition.

$$\max \text{Trace} \left\{ V^T S_{\Psi\Psi} S_{YY} V \right\}, \quad \text{s.t.} V^T S_{\Psi\Psi} V = I,$$

where $S_{\Psi\Psi} = \frac{1}{N}(\Psi \Psi^T - N \mu \Psi \mu^T) = (\frac{1}{N}) \Psi \Psi^T$. Besides, $C = I - (\frac{1}{N}) \epsilon \epsilon^T$ is the centering matrix, $\epsilon$ represents an $N$-by-1 vector with all the elements equal to ones, and $I$ denotes an identity matrix. For brevity, this study omits $I$ with different sizes.
To facilitate the subsequent discussions, the response variable \( y \) needs to be extended to a multivariate version \( Y \) by using one-hot encoding, where \( Y \) denotes an \( N \)-by-\( C \) matrix, with each row representing the response to the input \( \varphi(x_i) \). Also, in the following section, the multivariate versions of \( S_{\psi Y} \) and \( U \) are used to replace \( S_{\psi y} \) and \( u \) respectively. In classification problems, \( C \) refers to the number of classes.

### III. PROPOSED METHOD

#### A. OVERVIEW

As mentioned earlier in the contributions of the introduction, this study proposes two types of rectifications for improving OPLS — 1) Successively orthogonal deflation in constrained noisy subspace and 2) the isotropic space transform. These two rectifications are implemented in a series of related MCAs to demonstrate their gradually changing functionalities, respectively — i) Successive Subspace-MCA, ii) Isotropic Subspace-MCA, and iii) Successive Isotropic Subspace-MCA. These MCAs are all interconnected, with the same objective “finding coupled factors with maximum covariance” but with different strategies (i.e., the aforementioned two types of rectifications). Their interconnections are listed in Table 1. In typical OPLS, the objective is to discover dominant projection such that the cross-covariance matrix \( S_{\psi Y}S_{\psi Y}^T \) is maximized while \( S_{\psi Y} \) is constrained during trace-norm optimization. The three proposed MCA methods advance OPLS by introducing successive subspace deflation and the isotropic space transform \( \mathcal{I} \). The former finds successively orthogonal projective vectors in subspace, whereas the latter is aimed at converting the dimensions with unequal influences into those with equal ones in order to avoid biases in dimensions. As listed in the table, the proposed Successive Subspace-MCA iteratively considers successively orthogonal projective vectors from residuals compared with typical OPLS. The Isotropic Subspace-MCA explores dominant projection in isotropic space rather than in original feature space, where typical KOPLS searches. The Successive Isotropic Subspace-MCA combines the Successive Subspace-MCA and the Isotropic Subspace-MCA, where successively orthogonal projective vectors are explored in the residual isotropic subspace. In this study, these three proposed MCA methods along with the baseline are discussed.

### TABLE 1. Interconnections between the proposed methods and KOPLS.

| MCA                      | Maximization            | Constraint            |
|--------------------------|-------------------------|-----------------------|
| KOPLS                    | \( S_{\psi y}S_{\psi y}^T \) | \( S_{\psi y} \) |
| Proposed: Successive Subspace (SS) | \( S_{\psi y}S_{\psi y}^T \) | \( S_{\psi y} \) |
| Proposed: Isotropic Subspace (IS) | \( \mathcal{I}\{S_{\psi y}S_{\psi y}^T\} \) | \( \mathcal{I}(S_{\psi y}) \) |
| Proposed: Successive Isotropic Subspace (SIS) | \( \mathcal{I}\{S_{\psi y}S_{\psi y}^T\} \) | \( \mathcal{I}(S_{\psi y}) \) |

\( \mathcal{I} \) means the isotropic space transform, and \( i \) denotes the \( i \)-th iteration during a successively orthogonal projective search in the constrained residual subspace.

#### B. SUCCESSIVE SUBSPACE-MCA

The first one of the proposed MCA is Successive Subspace-MCA (SS-MCA). It enhances KOPLS methods that find coupled factors with maximum covariance. In KOPLS, the weighted signal space (i.e., \( S^{-1/2}_{\psi y}S_{\psi y}S_{\psi y}^T \) or equivalently \( S^{-1/2}_{\psi y}S_{\psi y}S_{\psi y}^T S^{-1/2}_{\psi y} \)) is based on the signal space (i.e., \( S_{\psi y}S_{\psi y}^T \)) and the noise space (i.e., \( S_{\psi y} \)), but the weighted signal space is deflated without any update on the noise space.

Let \( v_{SS} \) represent a \( J \times 1 \) projection vector, and

\[
\max_{i=1}^{D} \frac{\sum_{i} v_{SS}^T [i] S_{\psi y} S_{\psi y}^T v_{SS} [i]}{v_{SS}^T [i] S_{\psi y} S_{\psi y}^T v_{SS} [i],}
\]

s.t. \( \left\{ \begin{array}{l} v_{SS} [i] \perp v_{SS} [j], i \neq j \vspace{0.1cm} \end{array} \right. \)

which is equivalent to

\[
S_{\psi y}^{-1} [i] S_{\psi y} S_{\psi y}^T v_{SS} [i] = \lambda_{SS} [i] v_{SS} [i],
\]

where

\[
S_{\psi y} [i+1] = \left( I - v_{SS} [i] v_{SS}^T [i] \right) S_{\psi y} [i] \left( I - v_{SS} [i] v_{SS}^T [i] \right)
\]

means Hotelling deflation iterative process, and a basis is removed from subspace. Furthermore, \( i \) denotes the \( i \)-th iteration, and \( \lambda \) signifies the eigenvalue. Equation (6) updates the noise space when a new projection vector is found. Then, the process loops back to (5) and keeps searching another projection vector in the residual until \( D \) vectors are discovered. The advantage is that the signal space or the target of covariance maximization stays the same during matrix deflation, but the noise space (i.e., \( S_{\psi y} \)) is updated in each iteration with the change to the weighted signal space. The following descriptions also summarize the algorithm of the proposed SS-MCA and its rotational-invariant property.

- SS-MCA is rotational invariant

**Proof:**

Let \( \mathbb{R}_{SS} \) denote a \( J \)-by-\( J \) unitary rotation matrix and \( V_{SS} \) be a \( J \)-by-\( J \) matrix that consists of \( v_{ss} [1], \ldots, v_{ss} [D] \). It follows...
Algorithm SS-MCA

Input: \(X, Y, \text{ and } D\)

Output: \(U\) and \(V_{SS}\)

1. Compute \(S_{\Psi Y}, S_{\Psi \Psi}, \text{ and } \Psi\)
2. For \(i = 1: D\),
3. Compute \(v_{SS}[i]\) based on (5)
4. Compute \(S_{\Psi \Psi}[i + 1]\) based on (6)
5. Build \(V_{SS} = [v_{SS}[1], \ldots, v_{SS}[D]]\)
6. Solve \(U\) with \(\Psi, Y, V_{SS}\), and the multivariate version of (2)

that

\[
\begin{align*}
\text{max Trace} & \left\{ \frac{R_{SS} V_{SS} S_{\Psi \Psi} S_{\Psi} Y_{SS} R_{SS}}{R_{SS} V_{SS} S_{\Psi \Psi} S_{\Psi} Y_{SS} R_{SS}} \right\} \\
= & \text{max Trace} \left\{ \frac{V_{SS} S_{\Psi \Psi} S_{\Psi} Y_{SS} R_{SS}}{V_{SS} S_{\Psi \Psi} S_{\Psi} Y_{SS} R_{SS}} \right\} \\
= & \text{max Trace} \left\{ \frac{V_{SS} S_{\Psi \Psi} S_{\Psi} Y_{SS}}{V_{SS} S_{\Psi \Psi} S_{\Psi} Y_{SS}} \right\}.
\end{align*}
\]

(7)

C. ISOTROPIC SUBSPACE-MCA

The second one of the proposed MCA is Isotropic Subspace-MCA (IS-MCA). Unlike SS-MCA and KOPLS that ignore the unequal effect of each dimension, IS-MCA considers such influences and uses the isotropic space transform to convert the dimensions with different strength into equal strength. In other words, IS-MCA relies on the isotropic space transform, which corrects the distributional variance of each predicator variable. Let \(l\) represent the isotropic space transform, and IS-MCA is expressed as

\[
\begin{align*}
\text{max Trace} & \left\{ \frac{V_{IS}^T \{ S_{\Psi \Psi} S_{\Psi} Y_{IS} \} V_{IS}}{V_{IS}^T \{ S_{\Psi \Psi} S_{\Psi} Y_{IS} \} V_{IS}} \right\} \\
\text{s.t.} & \quad V_{IS}^T \{ S_{\Psi \Psi} \} V_{IS} = I \\
= & \text{max Trace} \left\{ \frac{V_{IS}^T \{ S_{\Psi \Psi} S_{\Psi} Y_{IS} \} V_{IS}}{V_{IS}^T \{ S_{\Psi \Psi} S_{\Psi} Y_{IS} \} V_{IS}} \right\}.
\end{align*}
\]

(8)

where \(l\) focuses on generating a transform vector \(F = S_{\Psi \Psi}^{-1/2}\), based on the constraint \(V_{IS}^T \{ S_{\Psi \Psi} \} V_{IS} = I\). Thus, \(l = S_{\Psi \Psi}^{-1/2} S_{\Psi \Psi} S_{\Psi} Y_{IS} S_{\Psi}^{-T/2}\), and \(l = S_{\Psi \Psi} S_{\Psi} Y_{IS}\) is transformed into \(S_{\Psi \Psi} S_{\Psi} Y_{IS} S_{\Psi}^{-T/2}\).

It is also worth noting that \(S_{\Psi \Psi} = S_{\Psi \Psi} S_{\Psi} S_{\Psi}^{-1} S_{\Psi} S_{\Psi}^{-1/2} S_{\Psi} S_{\Psi}^{-T/2}\). Accordingly, the distributional variance of each predicator variable becomes normalized. The difference between IS-MVA and SS-MVS is that both the signal space (i.e., \(S_{\Psi \Psi} S_{\Psi} Y_{IS}\)) and the noise space (i.e., \(S_{\Psi \Psi}\)) in IS-MVA are transformed into another type of space. After the projection matrix \(V_{IS}\) is found, one can use transform vector \(F = S_{\Psi \Psi}^{-1/2}\) to convert \(V_{IS}\) back to original feature space. It follows that

\[
W_{IS} = S_{\Psi \Psi}^{-T} V_{IS}.
\]

(9)

In brief, \(V_{IS}\) functions on isotropic space, whereas \(W_{IS}\) works on original feature space.

- In IS-MCA, \(V_{IS}\) and \(W_{IS}\) are generated from the data in isotropic space and original feature space respectively.

**Proof:**

\[
\begin{align*}
\text{max Trace} & \left\{ \frac{V_{IS}^T \{ S_{\Psi \Psi} S_{\Psi} Y_{IS} \} V_{IS}}{V_{IS}^T \{ S_{\Psi \Psi} S_{\Psi} Y_{IS} \} V_{IS}} \right\} \\
= & \text{max Trace} \left\{ \frac{V_{IS}^T \{ S_{\Psi \Psi} S_{\Psi} Y_{IS} \} V_{IS}}{V_{IS}^T \{ S_{\Psi \Psi} S_{\Psi} Y_{IS} \} V_{IS}} \right\}.
\end{align*}
\]

(10)

Like SS-MCA, the proposed IS-MCA is also rotational invariant. The following description shows the algorithm for the proposed IS-MCA.

D. SUCCESSIVE ISOTROPIC SUBSPACE-MCA

In the previous subsection “IS-MCA,” the signal space and the noise space are both transformed into equal-variance subspace. However, when IS-MCA searches dominant projection, the weighted signal space based on the signal space and the noise space is deflated without any updates on the noise space. To rectify such, the proposed SIS-MCA further integrates the concept of successively orthogonal projective vectors into IS-MCA. The following description details the proposed SIS-MCA. To simplify notations, let \(l(S_{\Psi \Psi})[i] = S_{\Psi \Psi}^{-1/2} S_{\Psi \Psi}[i] S_{\Psi \Psi}^{-T/2}\). It follows that

\[
\begin{align*}
\text{max} & \sum_{i=1}^{D} \frac{V_{SIS}^T[l] l(S_{\Psi \Psi} S_{\Psi} Y_{SIS}[i]) v_{SIS}[i]}{\sum_{i=1}^{D} v_{SIS}[i] l(S_{\Psi \Psi})[i] v_{SIS}[i]}, \\
\text{s.t.} & \quad v_{SIS}[i] \perp v_{SIS}[j], \quad i \neq j \\
& \quad v_{SIS}[i] S_{\Psi \Psi}[i] v_{SIS}[i] = 1.
\end{align*}
\]

(11)

Algorithm IS-MCA

Input: \(X, Y, \text{ and } D\)

Output: \(U\) and \(W_{IS}\)

1. Compute \(S_{\Psi \Psi}, S_{\Psi \Psi}, \text{ and } \Psi\)
2. Compute \(S_{\Psi \Psi}^{-1/2}\)
3. Compute \(l(S_{\Psi \Psi}) = S_{\Psi \Psi}^{-1/2} S_{\Psi \Psi} S_{\Psi \Psi}^{-T/2}\)
4. Compute \(l(S_{\Psi \Psi} S_{\Psi} Y_{SIS}) = S_{\Psi \Psi}^{-1/2} S_{\Psi \Psi} S_{\Psi} Y_{SIS} S_{\Psi \Psi}^{-T/2}\)
5. Compute \(V_{IS}\) based on (8)
6. Convert \(V_{IS}\) into \(W_{IS}\)
7. Solve \(U\) with \(\Psi, Y, W_{IS}\), and the multivariate version of (2)
B.-W. Chen: Novel Kernel OPLS for Dominant Sensor Data Extraction

Notably, cross-covariance and data fitting should be considered at the same time. When the both targets are satisfied, the method meets the requirement of MCA.

where

$$\mathcal{I}(S_{\Psi Y})[i]^{-1}\mathcal{I}(S_{\Psi Y}S_{\Psi Y}^T)v_{\text{SIS}}[i] = \lambda_{\text{SIS}}[i]v_{\text{SIS}}[i], \quad (12)$$

and

$$S_{\Psi}[i + 1] = \left(1 - w_{\text{SIS}}[i]w_{\text{SIS}}^T[i]\right)S_{\Psi}[i]$$

$$\times \left(1 - w_{\text{SIS}}[i]w_{\text{SIS}}^T[i]\right). \quad (14)$$

In each iteration where a projection vector is selected, the proposed SIS-MCA performs an update on its residual feature space by invoking (14). Hence, when the subsequent iteration (i.e., (12)) is launched, potential projection vectors are explored from the residual. The algorithm for the proposed SIS-MCA is summarized in the algorithm.

The proposed SIS-MCA has rotational invariance. This property can be verified by using the same concept mentioned in rotational-invariant SS-MCA. Therefore, the proof is omitted herein.

IV. EXPERIMENTAL RESULT

Experiments on six open datasets were carried out for evaluating performance. The information of these datasets is listed in Table 2. The first column shows the name. The rest columns signify the number of classes, samples, and dimensions (after kernel mapping), respectively. All the datasets were downloaded from the UCI machine learning repository.

Algorithm SIS-MCA

**Input:** X, Y, and D  
**Output:** U and W_{SIS}

1. Compute $S_{\Psi Y}$, $S_{\Psi Y}$, and $\Psi$
2. Compute $S_{\Psi}^{-1/2}$
3. Compute $\mathcal{I}(S_{\Psi Y})$ and $\mathcal{I}(S_{\Psi Y}S_{\Psi Y}^T)$
4. For $i = 1: D$,
   5. Compute $v_{\text{SIS}}[i]$ based on (12)
   6. Convert $v_{\text{SIS}}[i]$ into $w_{\text{SIS}}[i]$
   7. Compute $S_{\Psi}[i + 1]$ and $\mathcal{I}(S_{\Psi Y})[i + 1]$
   8. Build $W_{\text{SIS}} = [w_{\text{SIS}}[1], \ldots, w_{\text{SIS}}[D]]$
   9. Solve U with $\Psi$, Y, $W_{\text{SIS}}$, and the multivariate version of (2)

| Name   | #Classes | #Samples | #Dimensions |
|--------|----------|----------|-------------|
| Car    | 4        | 1728     | 84          |
| Abalone| 3        | 4177     | 165         |
| Cmc    | 3        | 1473     | 220         |
| Kr-vs-k| 17       | 28056    | 84          |
| Liver  | 2        | 345      | 84          |
| Mammography | 2        | 830      | 56          |
The experiment used 80.00% of the data for training and 20.00% of the data for testing. Ten-fold cross-validation was applied to the experiment. For kernels, as RBFs were inapplicable to intrinsic space due to infinite dimensions, third-ordered polynomial (poly3) kernels were used herein. For comparison, OPLS along with the proposed methods was tested. The proposed methods included SS-MCA, IS-MCA, and SIS-MCA; each implemented the proposed rectifications — 1) Successively orthogonal deflation in constrained noisy subspace and 2) the isotropic space transform. The criteria for benchmarks were a) fitting accuracy, b) covariance between the response variables and hidden components, and c) average covariance increases. It is worth noting that the objective of OPLS is fitting data well without decreasing covariance too much when one examines those benchmarks. This means that high accuracy with low covariance does not achieve the target of OPLS.

Fig. 2–Fig. 7 show the comparison between typical OPLS and the proposed methods. The horizontal axis denotes discovered hidden components, whereas the vertical axis represents fitting accuracy or covariance. Besides, only a half of the hidden components are displayed in the figures. For accuracy, observations showed that all the proposed methods were better than typical OPLS, especially when the number of hidden components increased or reached a half of the hidden components. The difference between OPLS and the proposed methods increased. Moreover, the accuracy of OPLS gradually degraded, whereas that of the proposed methods remained more stable compared with OPLS. For OPLS,
TABLE 3. Average covariance increase between the response variables and the hidden components.

| Algorithm          | Car   | Abalone | Cmc   | Kr-vs-k | Liver | Mammography |
|--------------------|-------|---------|-------|---------|-------|-------------|
| Baseline: KOPLS    | 0.00265 | 0.00026 | 0.00253 | 0.00166 | 0.01340 | 0.00298     |
| Proposed SS-MCA    | 0.00868 | 0.01022 | 0.00671 | 0.01489 | 0.02259 | 0.02204     |
| Proposed IS-MCA    | 0.00246 | 0.00028 | 0.00181 | 0.00214 | 0.00916 | 0.00253     |
| Proposed SIS-MCA   | 0.02356 | 0.00735 | 0.00760 | 0.02276 | 0.02469 | 0.00036     |

FIGURE 7. Upper: Accuracy rates based on the mammography dataset. Lower: Covariance between the response variables and the hidden components. Notably, cross-covariance and data fitting should be considered at the same time. When the both targets are satisfied, the method meets the requirement of MCA.

Although it gained more accuracy rates when fewer hidden components were used in data fitting, the covariance between the response variables and the hidden components was not satisfactory. Notably, fitting data well without decreasing covariance too much is the purpose of OPLS. Closely examining Fig. 2–Fig. 7 reveals that either the proposed SS-MCA or the proposed SIS-MCA generated maximum covariance. This is because both SS-MCA and SIS-MCA utilized the proposed technique — Successively orthogonal deflation in constrained noisy subspace, which was a major rectification for enhancing OPLS. Such a rectification narrowed down the noisy space and enhanced signal-to-noise ratios. For IS-MCA, the reason that the proposed IS-MCA did not generate maximum covariance was that it simply used the isotropic space transform. IS-MCA could be regarded as OPLS in isotropic space. Despite such, the proposed IS-MCA still created a steppingstone for SIS-MCA because SIS-MCA inherited both of the advantages from SS-MCA (using successively orthogonal deflation) and IS-MCA (using the isotropic space transform). Table 3 summarizes the average covariance increase when more hidden components were added. As listed in the table, either SS-MCA or SIS-MCA yielded a higher increase of covariance, far better than OPLS. Based on the above experimental results from a) fitting accuracy, b) covariance, and c) average covariance increases, the proposed two types of enhancing rectifications — Successively orthogonal deflation in constrained noisy subspace and the isotropic space transform could improve OPLS. This indicated that the proposed method not only fitted data well but also maximized covariance.

V. CONCLUSION

OPLS is usually susceptible to low signal-to-noise ratios owing to its objective — Maximizing cross-covariance in the signal space (i.e., $S_{XY}S_{XY}^T$) while the noise space (i.e., $S_{YY}$) should also be minimized at the same time. Even worse, between-group information is not used when cross-covariance matrices are computed. Such a characteristic intrinsically limits the capability to find correlations between predictor and response variables. To enhance signal-to-noise ratios, this work proposes a novel OPLS method by introducing 1) successively orthogonal deflation in constrained noisy subspace and 2) isotropic space transform for rectifying low signal-to-noise ratios while maximizing cross-covariance. The former seeks any potential hidden components from residual feature space and provides updates on the weighted signal space (i.e., $S_{XY}S_{XY}^T/2S_{YY}S_{YY}^{-1/2}$), which OPLS ignores. The feature space (i.e., $S_{YY}$) and its residuals are iteratively deflated, and the result is reflected in the weighted signal space. The latter searches potential hidden components in isotropic space instead of feature space. Isotropic space provides an alternative perspective for successively orthogonal deflation. The two proposed rectifications were implemented and assessed in our SS-MCA, IS-MCA, and SIS-MCA, respectively.

Based on experimental results, observations showed that either SS-MCA (using successively orthogonal deflation) or SIS-MCA (using the isotropic space transform and successively orthogonal deflation) satisfied the objective. Cross-covariance was maximized while data were also fitted, far better than the baseline. As expected, IS-MCA did not achieve the objective. This is because it simply...
performed isotropic space transform, and its rest processes were the same as the baseline. Despite such, experimental results indicated that using the isotropic space transform and successively orthogonal deflation simultaneously was still conducive to finding dominant components. This is why SIS-MCA outperformed the other approaches in some results. Such findings indicated that the proposed rectifications were capable of enhancing OPLS effectively.

The future work will concentrate on a more effective eigenstructure, e.g., new isotropic space transform that utilizes between-group information while maximizing cross-covariance matrices at the same time. The proposed approach devised a transform vector that relied on $S_{Between}^{-1/2}$, but an attempt to use between-group matrices (i.e., $S_{Between}^{-1}$) and within-group matrices (i.e., $S_{Within}^{-1}$) did not yield better results than the proposed one. Finally, as an open issue to this topic, it is still a challenge to efficiently decide the optimum number of dominant components. A more efficient way will be developed to avoid iterative matrix deflation for rapid computation.

REFERENCES

[1] W.-T. Wang and H.-C. Huang, “Regularized spatial maximum covariance analysis,” *Environmetrics*, vol. 29, no. 2, p. e2481, Oct. 2017.

[2] M. K. Tippett, T. DelSole, S. J. Mason, and A. G. Barnston, “Regression-based methods for finding coupled patterns,” *J. Climate*, vol. 21, no. 17, pp. 4384–4398, Sep. 2008.

[3] Y. Yu, T. Mckelvey, and S.-Y. Kung, “A classification scheme for ‘high-dimensional-small-sample-size’ data using soda and ridge-SVM with microwave measurement applications,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Vancouver, BC, Canada, vol. 31, May 2013, pp. 3542–3546.

[4] S.-Y. Kung, “Discriminative component analysis for privacy protection and visualization of big data,” *Multimedia Tools Appl.*, vol. 76, no. 3, pp. 3999–4034, Oct. 2015.

[5] W. W. Hsieh, *Machine Learning Methods in the Environmental Sciences*, Cambridge, U.K.: Cambridge Univ. Press, Aug. 2009.

[6] J. Arenas-Garcia and G. Camps-Valls, “Efficient kernel orthonormalized PLS for remote sensing applications,” *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 10, pp. 2872–2881, Oct. 2008.

[7] H. Abdi, “Partial least squares regression and projection on latent structure regression (PLS regression),” *Wiley Interdiscip. Rev., Comput. Statist.*, vol. 2, no. 1, pp. 97–106, Jan. 2010.

[8] S.-M. Horng, Y.-Y. Lee, and C.-L. Wu, “A study of the paying behavior for subscribing social network sites,” *Comput. Commun.*, vol. 73, pp. 282–290, Jan. 2016.

[9] N. Zhao, Q. Xu, and H. Wang, “Marginal screening for partial least squares regression,” *IEEE Access*, vol. 5, pp. 14047–14055, 2017.

[10] F. Al-Obeidat, B. Spencer, and O. Alfandi, “Consistently accurate forecasts of temperature within buildings from sensor data using ridge and lasso regression,” *Future Gener. Comput. Syst.*, Apr. 2018.

[11] R. Rosipal and N. Krämer, “Overview and recent advances in partial least squares,” *Future Gener. Comput. Syst.*, vol. 20, no. 5, pp. 54–51.

[12] Q. Li, F. Pan, Z. Zhao, and J. Yu, “Process modeling and monitoring with incomplete data based on robust probabilistic partial least square method,” *IEEE Access*, vol. 6, pp. 10160–10168, Feb. 2018.

[13] Q. Jiang, X. Yan, H. Yi, and F. Gao, “Data-driven batch-end quality modeling and monitoring based on optimized sparse partial least squares,” *IEEE Trans. Ind. Electron.*, vol. 67, no. 5, pp. 4098–4107, May 2020.

[14] C. Zhang and Q. Liu, “Region constraint person re-identification via partial least square on riemannian manifold,” *IEEE Access*, vol. 6, pp. 17060–17066, Mar. 2018.

[15] H. Chen, Y. Sun, J. Gao, Y. Hu, and B. Yin, “Solving partial least squares regression via manifold optimization approaches,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 2, pp. 588–600, Feb. 2019.

[16] Y. Mou, L. Zhou, W. Chen, J. Fan, and X. Zhao, “Maximum correntropy criterion partial least squares,” *Optik*, vol. 165, pp. 137–147, Jul. 2018.

[17] H. Nocairi, E. M. Qannari, E. Vigneau, and D. Bertrand, “Discrimination on latent components with respect to patterns. Application to multilinear data,” *Comput. Statist. Data Anal.*, vol. 48, no. 1, pp. 139–147, Jan. 2005.

[18] R. G. Burreton and G. R. Lloyd, “Partial least squares discriminant analysis: Taking the magic away,” *J. Chemometrics*, vol. 28, no. 4, pp. 213–225, Mar. 2014.

[19] L. Tang, S. Peng, Y. Bi, P. Shan, and X. Hu, “A new method combining LDA and PLS for dimension reduction,” *PLS ONE*, vol. 9, no. 5, May 2014, Art. no. e96944.

[20] M. Barker and W. Rayens, “Partial least squares for discrimination,” *J. Chemometrics*, vol. 17, no. 3, pp. 166–173, 2003.

[21] H. Abdi and L. J. Williams, “Partial least squares methods: Partial least squares correlation and partial least square regression,” in *Computational Toxicology*, vol. 2, B. Reiffeld and A. N. Mayeno, Eds. Totowa, NJ, USA: Humana Press, 2013, pp. 549–579.

[22] G. Li, B. Liu, S. J. Qin, and D. Zhou, “Quality relevant data-driven modeling and monitoring of multivariate dynamic processes: The dynamic T-PLS approach,” *IEEE Trans. Neural Netw.*, vol. 22, no. 12, pp. 2262–2271, Dec. 2011.

[23] M. Bylesjo, M. Rantalainen, O. Cloarec, J. K. Nicholson, E. Holmes, and J. Trygg, “OPLS discriminant analysis: Combining the strengths of PLS-DA and SIMCA classification,” *J. Chemometrics*, vol. 20, nos. 8–10, pp. 341–351, 2007.

[24] J. Arenas-Garcia, K. B. Petersen, G. Camps-Valls, and L. K. Hansen, “Kernel multivariate analysis framework for supervised subspace learning: A tutorial on linear and kernel multivariate methods,” *IEEE Signal Process. Mag.*, vol. 30, no. 4, pp. 16–29, Jul. 2013.

[25] T. De Bie, N. Cristianini, and R. Rosipal, “Eigenproblems in pattern recognition,” in *Handbook of Geometric Computing: Applications in Pattern Recognition. Computer Vision, Neuralcomputing, and Robotics*, E. B. Corrochano, Ed. Berlin, Germany: Springer, 2005, pp. 129–167.

[26] S. Munoz-Romero, J. Arenas-Garcia, and V. Gomez-Verdejo, “Nonnegative OPLS for supervised design of filter banks: Application to image and audio feature extraction,” *IEEE Trans. Multimedia*, vol. 20, no. 7, pp. 1751–1766, Jul. 2018.

[27] Q. Jiang and X. Yan, “Quality-driven kernel projection to latent structure model for nonlinear process monitoring,” *IEEE Access*, vol. 7, pp. 74450–74458, Jun. 2019.

[28] N. T. Trendafilov, “On the €l PROCRUSTESproblem,” *Future Gener. Comput. Syst.*, vol. 19, no. 7, pp. 1177–1186, Oct. 2003.

[29] G. Aloor and L. Jacob, “Distributed wireless sensor network localization using stochastic proximity embedding,” *Comput. Commun.*, vol. 33, no. 6, pp. 745–755, Apr. 2010.

[30] G. R. K. Kumar and M. R. Reddy, “An orthonormalized partial least squares based spatial filter for SSVEP extraction,” in *Proc. Int. Conf. Inтелl. Hum. Comput. Interact.*, Allahabad, India, vol. 9, Dec. 2018, pp. 16–25.

[31] S. X. Ding, S. Yin, K. Peng, H. Hao, and B. Shen, “A novel scheme for key performance indicator prediction and diagnosis with application to an industrial hot strip mill,” *IEEE Trans Ind. Informat.*, vol. 9, no. 4, pp. 2239–2247, Nov. 2013.

[32] F. Luo, B. Du, L. Zhang, L. Zhang, and D. Tao, “Feature learning using spatial-spectral hypergraph discriminant analysis for hyperspectral image,” *IEEE Trans. Cybern.*, vol. 49, no. 7, pp. 2406–2419, Jul. 2019.