Designing various component analysis at will

Akisato Kimura, Hitoshi Sakano, Hirokazu Kameoka
NTT Communication Science Laboratories
akisato@ieee.org

Masashi Sugiyama
Tokyo Institute of Technology
sugi@cs.titech.ac.jp

Abstract

This paper provides a generic framework of component analysis (CA) methods introducing a new expression for scatter matrices and Gram matrices, called Generalized Pairwise Expression (GPE). This expression is quite compact but highly powerful: The framework includes not only (1) the standard CA methods but also (2) several regularization techniques, (3) weighted extensions, (4) some clustering methods, and (5) their semi-supervised extensions. This paper also presents quite a simple methodology for designing a desired CA method from the proposed framework: Adopting the known GPEs as templates, and generating a new method by combining these templates appropriately.

1 Introduction

Component analysis (CA) is traditional, quite simple but might be one of the powerful tools to obtain a hidden structure embedded in the data. Recent reports showed its effectiveness for several tasks in computer vision and pattern recognition. Principal component analysis (PCA), Fisher discriminant analysis (FDA), multiple linear regression (MLR), and canonical correlation analysis (CCA) are well known as standard CA methods [1]. They can be formulated as a generalized eigenvalue problem of a scatter matrix or an augmented matrix composed of several scatter matrices [3, 4]. Kernel CA methods as kernelized extensions of those standard methods have been also developed to deal with non-vector samples and non-linear analysis, which can be formulated as a generalized eigenvalue problem of Gram matrices, instead of scatter matrices [3, 4]. Kernel CA often needs some regularization techniques such as ℓ_2-norm regularization to inhibit overfitting and Laplacian regularization [2] to fit underlying data manifolds smoothly. In addition, improvements of robustness against outliers and separately distributed samples (e.g. locality preserving projection (LPP) [8] and local FDA (LFDA) [11]) and their extensions to semi-supervised analysis [12, 2] have been considered.

Although a lot of CA methods and several trials to unify these methods have been presented so far [3, 4, 5, 6], freely designing a tailor-made method of CA for a specific purpose or domain still remains an open problem. Until now, researchers have had to choose one of the existing methods that seems best to address the problem of interest, or had to laboriously develop a new analysis method tailored specifically for that purpose.

In view of the above discussions, this paper provides a new expression of scatter matrices and Gram matrices, which we call *generalized pairwise expression (GPE)* to make it easy to design a new CA method with desired properties. The methodology is quite simple: **adopting the above mentioned special cases as templates, and generating a new method by combining these templates appropriately**. This characteristics has not been discussed yet in any previous researches to our best knowledge. It is also possible to individually select and arrange samples for calculating the scatter matrices of the methods to be combined, which enables us to extend CA methods to semi-supervised ones and multi-modal ones.

2 Generalized pairwise expression

Consider two sample sets $\boldsymbol{X} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{N_x}\}$ and $\boldsymbol{Y} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N, \boldsymbol{y}_{N_x+1}, \ldots, \boldsymbol{y}_{N_x+N_y-N}\}$ with N_x and N_y samples and d_x and d_y dimensions. For brevity, suppose that both of the sample sets $\boldsymbol{X}$ and $\boldsymbol{Y}$ are centered on the origin, and samples with the same suffix are paired. A pair $(\boldsymbol{X}, \boldsymbol{Y})$ is separated into the following two types: A *complete set* $(\boldsymbol{X}_C, \boldsymbol{Y}_C)$ so that every sample pair $(\boldsymbol{x}_n, \boldsymbol{y}_n)$ co-occurs, and *incomplete sets* $\boldsymbol{X}_I$ and $\boldsymbol{Y}_I$ so that every sample $\boldsymbol{x}_n$ (resp. $\boldsymbol{y}_n$) cannot find the co-occurring sample. Unless otherwise stated, we assume $N_x = N_y = N$.

Many CA methods developed so far involve the following optimization problem:

$$\boldsymbol{w}^{(opt)} \quad := \quad \arg\max_{\boldsymbol{w} \in \mathcal{R}^d} (\boldsymbol{w}^\top \overline{\boldsymbol{C}} \boldsymbol{w})(\boldsymbol{w}^\top \underline{\boldsymbol{C}} \boldsymbol{w})^{-1},$$

where $\overline{\boldsymbol{C}}$ and $\underline{\boldsymbol{C}}$ are symmetric matrices with some statistical natures. Roughly speaking, $\overline{\boldsymbol{C}}$ encodes the quantity that we want to increase, and $\underline{\boldsymbol{C}}$ corresponds to the quantity that we want to decrease. The above optimization can be converted to the following generalized eigenvalue problem via Lagrangian multipliers:

$$\overline{\boldsymbol{C}}\boldsymbol{w} \;=\; \lambda\underline{\boldsymbol{C}}\boldsymbol{w},$$

The eigenvector $\boldsymbol{w}_k$ $(k = 1, \ldots, r)$ of this generalized eigenvalue problem gives a solution of the original CA.

When addressing CA methods, we often deal with the following type of second-order statistics as an extension of scatter matrices, since it is convenient to describe the relation between pairs of features:

$$\boldsymbol{S}_{\boldsymbol{Q},xy} \;:=\; \sum_{n=1}^{N}\sum_{m=1}^{N} Q_{n,m}(\boldsymbol{x}_n - \boldsymbol{x}_m)(\boldsymbol{y}_n - \boldsymbol{y}_m)^\top,$$

where $\boldsymbol{Q}$ is an $N \times N$ non-negative symmetric matrix. A typical example is the scatter matrix $\boldsymbol{S}_{xy} = N^{-1}\sum_{n=1}^{N}\boldsymbol{x}_n\boldsymbol{y}_n^\top$. Let $\boldsymbol{D}_{\boldsymbol{Q}}$ be the $N \times N$ diagonal matrix with $D_{\boldsymbol{Q},n,n} = \sum_{m=1}^{N} Q_{n,m}$, and $\boldsymbol{L}_{\boldsymbol{Q}} = \boldsymbol{D}_{\boldsymbol{Q}} - \boldsymbol{Q}$ (Laplacian). Then, $\boldsymbol{S}_{\boldsymbol{Q},xy}$ can be expressed as

$$\boldsymbol{S}_{\boldsymbol{Q},xy} \;=\; \boldsymbol{X}\boldsymbol{L}_{\boldsymbol{Q}}\boldsymbol{Y}^\top.$$

The above expression is called the *pairwise expression (PE)* of the second-order statistics $\boldsymbol{S}_{\boldsymbol{Q},xy}$[12].

Here, we extend it to the following expression introducing an additional matrix independent of $\boldsymbol{Q}$:

$$\boldsymbol{S}_{\boldsymbol{Q},xy} \;:=\; \boldsymbol{X}\boldsymbol{L}_{\boldsymbol{Q},1}\boldsymbol{Y}^\top + \boldsymbol{L}_2,$$

where $\boldsymbol{L}_{\boldsymbol{Q},1}$ is a $N \times N$ positive semi-definite matrix, and $\boldsymbol{L}_2$ is a $d_x \times d_y$ matrix.[1] We call this expression the *generalized pairwise expression (GPE)* of the second-order statistics $\boldsymbol{S}_{\boldsymbol{Q},xy}$. The first and second terms are called the *data term* and *bias term*, respectively.

We can derive the following fundamental lemmas of GPE from the definition, if the number N of samples is sufficiently large:

1. If $\boldsymbol{A}$ is a GPE and $\beta > 0$, then $\beta\boldsymbol{A}$ is also a GPE.
2. If $\boldsymbol{A}$ and $\boldsymbol{B}$ are $d_x \times d_y$ GPEs, then $\boldsymbol{A} + \boldsymbol{B}$ is also a GPE with the same size.
3. If $\boldsymbol{A}$ is a $d_x \times d_y$ GPE with columns, and $\boldsymbol{B}$ is a $d_y \times d_z$ GPE, then $\boldsymbol{AB}$ is a $d_x \times d_z$ GPE.

Proof. The 1st and 2nd claims can be easily proved, so we concentrate to prove the 3rd one. First, let us define

$$\boldsymbol{A} := \boldsymbol{X}\boldsymbol{L}_{A1}\boldsymbol{Y}^\top + \boldsymbol{L}_{A2}, \quad \boldsymbol{B} := \boldsymbol{Y}\boldsymbol{L}_{B1}\boldsymbol{Z}^\top + \boldsymbol{L}_{B2},$$

where $\boldsymbol{L}_{A1}$ (resp. $\boldsymbol{L}_{B1}$) is a positive semi-definite matrix with d_x (resp. d_y) rows and d_y (resp. d_z) columns, and $\boldsymbol{L}_{A2}$ (resp. $\boldsymbol{L}_{B2}$) is a $d_x \times d_y$ (resp. $d_y \times d_z$) non-negative matrix. Then, we obtain

$$\begin{aligned}\boldsymbol{AB} &= (\boldsymbol{X}\boldsymbol{L}_{A1}\boldsymbol{Y}^\top + \boldsymbol{L}_{A2})(\boldsymbol{Y}\boldsymbol{L}_{B1}\boldsymbol{Z}^\top + \boldsymbol{L}_{B2}),\\ &= \boldsymbol{X}(\boldsymbol{L}_{A1}\boldsymbol{Y}^\top\boldsymbol{Y}\boldsymbol{L}_{B1})\boldsymbol{Z}^\top + (\boldsymbol{L}_{A2}\boldsymbol{Y})\boldsymbol{L}_{B1}\boldsymbol{Z}^\top\\ &\quad +\boldsymbol{X}\boldsymbol{L}_{A1}(\boldsymbol{Y}^\top\boldsymbol{L}_{B2}) + \boldsymbol{L}_{A2}\boldsymbol{L}_{B2}.\end{aligned}$$

Here, we can find some matrices $\boldsymbol{L}_{Ci}$ $(i = 1, 2, 3)$ satisfying $\boldsymbol{L}_{C1} = \boldsymbol{L}_{A1}\boldsymbol{Y}^\top\boldsymbol{Y}\boldsymbol{L}_{B1}$, $\boldsymbol{X}\boldsymbol{L}_{C2} = \boldsymbol{L}_{A2}\boldsymbol{Y}$, and $\boldsymbol{L}_{C3}\boldsymbol{Z}^\top = \boldsymbol{Y}^\top\boldsymbol{L}_{B2}$, if $N > \max(d_x, d_y, d_z)$, which implies that for some matrices $\boldsymbol{L}_{D1}$ and $\boldsymbol{L}_{D2}$

$$\begin{aligned}\boldsymbol{AB} &= \boldsymbol{X}\boldsymbol{L}_{C1}\boldsymbol{Z}^\top + \boldsymbol{X}\boldsymbol{L}_{C2}\boldsymbol{L}_{B1}\boldsymbol{Z}^\top\\ &\quad +\boldsymbol{X}\boldsymbol{L}_{A1}\boldsymbol{L}_{C3}\boldsymbol{Z}^\top + \boldsymbol{L}_{A2}\boldsymbol{L}_{B2}\\ &= \boldsymbol{X}(\boldsymbol{L}_{C1} + \boldsymbol{L}_{C2}\boldsymbol{L}_{B1} + \boldsymbol{L}_{A1}\boldsymbol{L}_{C3})\boldsymbol{Z}^\top + \boldsymbol{L}_{A2}\boldsymbol{L}_{B2}\\ &= \boldsymbol{X}\boldsymbol{L}_{D1}\boldsymbol{Z}^\top + \boldsymbol{L}_{D2}.\end{aligned}$$

which means $\boldsymbol{AB}$ is also a GPE. □

Thanks to the above properties, **various CA methods can be easily designed by simply combining (i.e. adding, scaling and multiplying) GPEs of existing methods with desired properties.** The rest of the problem is to reveal GPE of existing methods and to present several significant examples that would bring us quite an important hint when constructing new CA methods.

3 Reviewing CA methods via GPE

The GPEs of the standard CA methods are listed in Table 1. Several detailed derivations can be seen in [3, 4]. Instead, this paper provides several significant examples that would be quite an important hint when constructing new CA methods.

3.1 Semi-supervised local FDA (SELF)

Semi-supervised local FDA (SELF) [12] integrates localized FDA (LFDA) [11] as a supervised CA method and PCA as an unsupervised one. SELF is a typical example for designing CA methods via the GPE framework from the following two viewpoints:

1. Combining several CA methods via GPE,
2. Selecting samples to calculate data terms, a key to extend CA methods to semi-supervised ones.

SELF is effective when (1) we have a complete set $(\boldsymbol{X}_C, \boldsymbol{Y}_C)$, where each sample in $\boldsymbol{Y}$ represents a class ID of the paired sample in $\boldsymbol{X}$, and (2) an incomplete set $\boldsymbol{X}_I$ only exists, namely there are at least one unlabeled samples. In such cases, we can regularize the objective function of the LFDA using the additional

[1] In general, we do not have to explicitly consider the matrix $\boldsymbol{Q}$ for the following discussions.

Table 1. GPEs of standard methods

Method	$\overline{\boldsymbol{C}}$	$\underline{\boldsymbol{C}}$
PCA	$\boldsymbol{S}_{xx}$	$\boldsymbol{I}_{d_X}$
FDA	$\boldsymbol{S}_{xx}^{(b)}$	$\boldsymbol{S}_{xx}^{(w)}$
CCA	$\begin{bmatrix} \boldsymbol{0} & \boldsymbol{S}_{xy} \\ \boldsymbol{S}_{yx} & \boldsymbol{0} \end{bmatrix}$	$\begin{bmatrix} \boldsymbol{S}_{xx} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S}_{yy} \end{bmatrix}$
MLR	$\begin{bmatrix} \boldsymbol{0} & \boldsymbol{S}_{xy} \\ \boldsymbol{S}_{yx} & \boldsymbol{0} \end{bmatrix}$	$\begin{bmatrix} \boldsymbol{S}_{xx} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{d_y} \end{bmatrix}$
PCR[9]	$\begin{bmatrix} \boldsymbol{0} & \boldsymbol{S}_{\hat{x}y} \\ \boldsymbol{S}_{y\hat{x}} & \boldsymbol{0} \end{bmatrix}$	$\begin{bmatrix} \boldsymbol{S}_{\hat{x}\hat{x}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{d_y} \end{bmatrix}$
OPLS	$\boldsymbol{S}_{xy}\boldsymbol{S}_{xy}^{\top}$	$\boldsymbol{S}_{xx}$
Ridge regression	$\begin{bmatrix} \boldsymbol{0} & \boldsymbol{S}_{xy} \\ \boldsymbol{S}_{yx} & \boldsymbol{0} \end{bmatrix}$	$\begin{bmatrix} \boldsymbol{S}_{xx} + \delta \boldsymbol{I}_{d_x} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{d_y} \end{bmatrix}$
LPP[8]	$\boldsymbol{X}\boldsymbol{L}\boldsymbol{X}^{\top}$	$\boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^{\top}$
LFDA[11]	$\boldsymbol{S}_{\boldsymbol{Q},xx}^{(lb)}$	$\boldsymbol{S}_{\boldsymbol{Q},xx}^{(lw)}$

PCR: Principal component regression, OPLS: Orthogonal partial least-squares.
$\boldsymbol{S}_{xx}^{(b)}$ and $\boldsymbol{S}_{xx}^{(w)}$: Between-class and within-class scatter matrices of $\boldsymbol{X}$, $\hat{\boldsymbol{X}} = \boldsymbol{U}_K \Sigma_K \boldsymbol{V}_K^{\top}$: K-rank approximation of $\boldsymbol{X}$ by SVD, $\boldsymbol{I}_d$: $d \times d$ identity matrix, $\delta > 0$: constant, $\boldsymbol{S}_{\boldsymbol{Q}}^{(b)}$ and $\boldsymbol{S}_{\boldsymbol{Q}}^{(w)}$: between-class and within-class scatter matrices of $\boldsymbol{X}$ weighted by an $N \times N$ non-negative symmetric matrix.

data $\boldsymbol{X}_I$. In detail, SELF integrates the GPE ($\boldsymbol{S}_{\boldsymbol{Q},C}^{(lb)}$, $\boldsymbol{S}_{\boldsymbol{Q},C}^{(lw)}$) of LFDA calculated only from the complete set $(\boldsymbol{X}_C, \boldsymbol{Y}_C)$ and the GPE $\boldsymbol{S}_{xx}$ of PCA calculated from all the samples $\boldsymbol{X}$, as follows:

$$\begin{aligned} \overline{\boldsymbol{C}}^{(SELF)} &= \beta \boldsymbol{S}_{\boldsymbol{Q},C}^{(lb)} + (1-\beta)\boldsymbol{S}_{xx}, \\ \underline{\boldsymbol{C}}^{(SELF)} &= \beta \boldsymbol{S}_{\boldsymbol{Q},C}^{(lw)} + (1-\beta)\boldsymbol{I}_{d_x}, \end{aligned}$$

where β is a scalar satisfying $0 \leq \beta \leq 1$. When $\beta = 1$, SELF is equivalent to LFDA with only the labeled samples $(\boldsymbol{X}_C, \boldsymbol{Y}_C)$. Meanwhile, when $\beta = 0$, SELF is equivalent to PCA with all samples in $\boldsymbol{X}$. In general, SELF inherits the properties of both LFDA and PCA, and their influences can be controlled by the scalar β.

3.2 Semi-supervised CCA (SemiCCA)

In a similar manner to SELF, we can derive a new extension of CCA that can deal with not only a complete set $(\boldsymbol{X}_C, \boldsymbol{Y}_C)$ and also incomplete sets $(\boldsymbol{X}_I, \boldsymbol{Y}_I)$. This extensive method is generally called SemiCCA [10]. The GPE of SemiCCA can be described as

$$\overline{\boldsymbol{C}}^{(sCCA)} = \beta \begin{bmatrix} \boldsymbol{0} & \boldsymbol{S}_{Cxy} \\ \boldsymbol{S}_{Cyx} & \boldsymbol{0} \end{bmatrix} + (1-\beta) \begin{bmatrix} \boldsymbol{S}_{xx} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S}_{yy} \end{bmatrix},$$

$$\underline{\boldsymbol{C}}^{(sCCA)} = \beta \begin{bmatrix} \boldsymbol{S}_{Cxx} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{S}_{Cyy} \end{bmatrix} + (1-\beta) \begin{bmatrix} \boldsymbol{I}_{d_x} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{d_y} \end{bmatrix}.$$

Table 2. GPEs of kernelized CA methods

Method	$\overline{\boldsymbol{C}}$	$\underline{\boldsymbol{C}}$
kPCA	$\boldsymbol{K}_x$	$\boldsymbol{I}_N$
kFDA	$\boldsymbol{K}_x^{(b)}$	$\boldsymbol{K}_x^{(w)}$
kCCA	$\begin{bmatrix} \boldsymbol{0} & \boldsymbol{K}_x\boldsymbol{K}_y \\ \boldsymbol{K}_y\boldsymbol{K}_x & \boldsymbol{0} \end{bmatrix}$	$\begin{bmatrix} \boldsymbol{K}_x^2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{K}_y^2 \end{bmatrix}$
kMLR	$\begin{bmatrix} \boldsymbol{0} & \boldsymbol{K}_x\boldsymbol{K}_y \\ \boldsymbol{K}_y\boldsymbol{K}_x & \boldsymbol{0} \end{bmatrix}$	$\begin{bmatrix} \boldsymbol{K}_x^2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_N \end{bmatrix}$
kCCA+ℓ_2[7]	$\begin{bmatrix} \boldsymbol{0} & \boldsymbol{K}_x\boldsymbol{K}_y \\ \boldsymbol{K}_y\boldsymbol{K}_x & \boldsymbol{0} \end{bmatrix}$	$\begin{bmatrix} \boldsymbol{K}_x^{(\ell_2)} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{K}_y^{(\ell_2)} \end{bmatrix}$
L-kCCA[2]	$\begin{bmatrix} \boldsymbol{0} & \boldsymbol{K}_x\boldsymbol{K}_y \\ \boldsymbol{K}_y\boldsymbol{K}_x & \boldsymbol{0} \end{bmatrix}$	$\begin{bmatrix} \boldsymbol{K}_x^{(L)} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{K}_x^{(L)} \end{bmatrix}$
LE, SC	$\boldsymbol{L}_x$	$\boldsymbol{D}_x$
LLE	$\boldsymbol{K}_x^{(LL)}\boldsymbol{K}_x^{(LL)\top}$	$\boldsymbol{I}_N$
NC, nSC	$\boldsymbol{D}_x^{-1/2}\boldsymbol{L}_x\boldsymbol{D}_x^{-1/2}$	$\boldsymbol{I}_N$

L-kCCA: Laplacian-regularized kernel CCA, $\boldsymbol{K}_x$, $\boldsymbol{K}_x$: Gram matrices, $\boldsymbol{K}_x^{(\ell_2)} = \boldsymbol{K}_x^2 + \delta_x \boldsymbol{K}_x$, $\boldsymbol{K}_x^{(L)} = \boldsymbol{K}_x^2 + \gamma_x \boldsymbol{R}_x$, $\boldsymbol{R}_x = \boldsymbol{K}_x \boldsymbol{L}_x \boldsymbol{K}_x$m, LE: Laplacian eigenmap, SC: Spectral clustering, NC: Normalized cuts, nSC: normalized SC, $\boldsymbol{K}_x^{(LL)} = \boldsymbol{I}_N - \boldsymbol{K}_x$

3.3 Kernelized extensions

A lot of methods in the GPE framework can be kernelized in a similar manner to the existing ones. The GPEs of major kernelized CA methods are listed in Table 1. By introducing kernelized expression, several methods for clustering and local embedding can be included in this framework, e.g. Laplacian eigenmap (LE), locally linear embedding (LLE), spectral clustering (SC) and normalized cuts (NC).

4 Designing new methods

Summarizing the discussions so far, we describe (1) GPEs of standard CA methods, (2) the way for integrating several GPEs and (3) some semi-supervised extensions by changing samples for calculating GPEs. This section shows that we can easily design new CA methods at will by replicating those steps.

Consider a problem of video categorization, where its training data includes image features $\boldsymbol{X}$, audio features $\boldsymbol{Y}$ and class indexes. Finding appropriate correlations of such three different modals would be still challenging. Here, we consider an integration (CFDA) of CCA and FDA which enables us to extract class-wise differences of multiple feature correlations as well as to achieve discriminative embedding simultaneously.

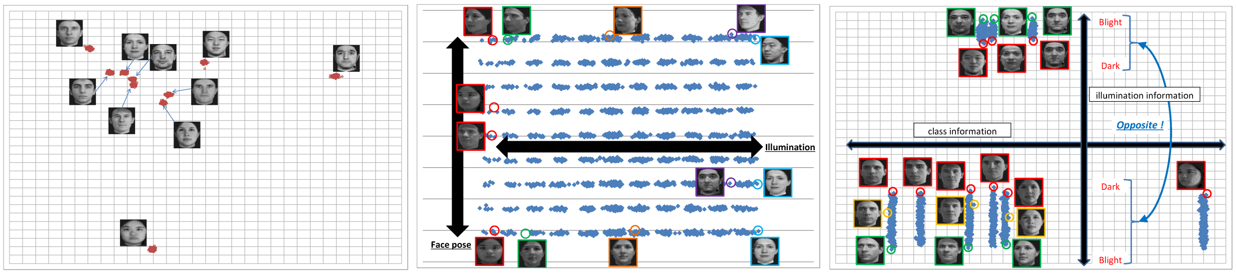

Figure 1. 2D embedding of MIT CBCL dataset. (Left) FDA with person IDs as classes, (middle) CCA with illuminations and face poses as side info, (right) CCA-FDA with person IDs as classes and illuminations and face poses as side information

CCA-FDA can be formulated as

$$\overline{\boldsymbol{C}}_{\boldsymbol{Q}}^{(CFDA)} = \beta \begin{bmatrix} \mathbf{0} & \boldsymbol{S}_{xy} \\ \boldsymbol{S}_{yx} & \mathbf{0} \end{bmatrix} + (1-\beta)\boldsymbol{S}_{\boldsymbol{Q}}^{(lb)},$$

$$\underline{\boldsymbol{C}}_{\boldsymbol{Q}}^{(CFDA)} = \beta \begin{bmatrix} \boldsymbol{S}_{xx} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{S}_{yy} \end{bmatrix} + (1-\beta)\boldsymbol{S}_{\boldsymbol{Q}}^{(lw)}.$$

Figure 1 shows an example how CCA-FDA works with MIT CBCL face dataset, which implies that CCA-FDA obtains a specific property combining CCA and FDA.

Integrating two methods within the kernelized GPE framework is not obvious, since a simple addition of Gram matrices is not a GPE. One example can be seen in a kernelized extension of SELF, called kernel SELF [12]. Remember that the original SELF integrates LFDA with labeled samples and PCA with all the samples (see Section 3.1), and it can be formulated by a localized between-class scatter matrix $\boldsymbol{S}_{\boldsymbol{Q},C}^{(lb)}$, localized within-class matrix $\boldsymbol{S}_{\boldsymbol{Q},C}^{(lw)}$ and the ordinary scatter matrix $\boldsymbol{S}_{xx}$. Kernel SELF can be formulated via their Laplacian matrices $\boldsymbol{L}_{\boldsymbol{Q},C}^{(lb)}$, $\boldsymbol{L}_{\boldsymbol{Q},C}^{(lw)}$, $\boldsymbol{L}_{xx}$, as follows:

$$\overline{\boldsymbol{C}}^{(kSELF)} = \boldsymbol{K}_x\{\beta \boldsymbol{L}_{\boldsymbol{Q},C}^{(lb)} + (1-\beta)\boldsymbol{L}_{xx}\}\boldsymbol{K}_x,$$

$$\underline{\boldsymbol{C}}^{(kSELF)} = \beta \boldsymbol{K}_x \boldsymbol{L}_{\boldsymbol{Q},C}^{(lb)} \boldsymbol{K}_x + (1-\beta)\boldsymbol{K}_x.$$

From this formulation, when dealing with kernelized CA, we have to explicitly derive GPEs of existing methods, and replace the data matrix into its Gram matrix.

5 Concluding remarks

This paper presented a new expression of scatter matrices and Gram matrices called generalized pairwise expression (GPE). The GPE not only provided a unified insight into various CA methods and their extensions, but also made it eacy to design new CA methods with desired properties. The methodology is quite simple: adopting GPEs of existing methods as templates, and combining (adding, scaling and multiplying those templates according to the properties you want.

The GPE framework covers a wide variety of CA methods, and thus the way we have presented in this paper for designing new methods is still one example. Developing more general guidelines would be one of the important future work. We will disclose the details just before the conference in arxiv.org.

References

[1] T. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley-Interscience, 2003.

[2] M. Blaschko+. Semi-supervised laplacian regularization of kernel canonical correlation analysis. In *Proc. ECML-PKDD2008*

[3] M. Borga+. A unified approach to PCA, PLS, MLR and CCA. Technical Report LiTH-ISY-R-1992, Linkoping University, 1997.

[4] T. De Bie+. Eigenproblems in pattern recognition. In *Handbook of Geometric Computing: Applications in Pattern Recognition, Computer Vision, Neuralcomputing, and Robotics*, Springer, 2005.

[5] F. De la Torre. A unification of component analysis methods. In *Handbook of Pattern Recognition and Computer Vision*, World Scientific Pub, 2010.

[6] F. De la Torre. A least-squares framework for component analysis. *IEEE Trans. PAMI*, 2012.

[7] D. Hardoon+. Canonical Correlation Analysis: An Overview with Application to Learning Methods. *Neural Computation*, 2004.

[8] X. He and P. Niyogi. Locality Preserving Projections. In *Proc. NIPS2003*.

[9] I. Jolliffe. A Note on the Use of Principal Components in Regression. *JRSS*, Series C (Applied Statistics), 1982.

[10] A. Kimura+. SemiCCA: Efficient semi-supervisedl learning of canonical correlations. In *Proc. ICPR2010*.

[11] M. Sugiyama. Dimensionality Reduction of Multimodal Labeled Data by Local Fisher Discriminant Analysis. *JMLR*, 2007.

[12] M. Sugiyama+. Semi-supervised local Fisher discriminant analysis for dimensionality reduction. *Machine learning*, 2010.