

SubXPCA versus PCA: A Theoretical Investigation

Atul Negi

Dept. of CIS, University of Hyderabad, India, Email: atulcs@uohyd.ernet.in

VijayaKumar Kadappa

Dept. of MCA, Vasavi College of Engg., Hyderabad, India, Email: kadappakumar@gmail.com

Abstract

Principal Component Analysis (PCA) is a widely accepted dimensionality reduction technique that is optimal in a MSE sense. PCA extracts ‘global’ variations and is insensitive to ‘local’ variations in subpatterns. Recently, we have proposed a novel approach, SubXPCA, which was more effective computationally than PCA and also effective in computing principal components with both global and local information across subpatterns. In this paper, we show the near-optimality of SubXPCA (in terms of summarization of variance) by proving analytically that ‘SubXPCA approaches PCA with increase in number of local principal components of subpatterns.’ This is demonstrated empirically upon CMU Face Data.

1. Introduction

Principal Component Analysis (PCA) is concerned with summarizing the *variance-covariance structure* using a few linear combinations of the original set of d variables (features) [5]. The classical PCA is an optimal linear dimensionality reduction scheme in terms of mean squared error. However, classical PCA suffers from large time complexity ($O(N.d^2)$, where N is the number of training patterns, d generally tends to become large in real problems) to compute the covariance matrix. Many approaches such as neural network based PCA methods [6], 2DPCA based methods [9] reduce computational complexity as compared to classical PCA method. However, these methods are based on whole-patterns, which are suitable for global feature extraction like classical PCA, and they may not perform well if local variations are prominent [8]. It is known that PCA performs global feature extraction that retains information based upon covariances between every pair of original features. However, global features are inef-

fective when variations are confined to subpatterns, (i.e. local variations). For example, in faces a smile is a local variation confined to a sub region and is distributed over several patterns.

Feature partitioning based PCA (FP-PCA) methods such as SubPCA [1], Region-based PCA [7], Modular PCA [4], EigenRegions [3], etc were proposed in the literature that perform well in capturing local variations. These FP-PCA methods have reduced time complexity due to processing of subpatterns. The difficulty with FP-PCA methods is that they loose out global information and show poor summarization of variance due to lack of considering the cross subpattern covariances. Therefore, we need to factor computational efficiency with extraction of both local and global feature properties while computing PCs.

Recently, we proposed a novel FP-PCA method, SubXPCA [8], which extracts both global and local feature variations. SubXPCA is computationally more efficient as compared to PCA, and also effective in classification accuracy. In this paper, our focus is to show analytically the closeness of SubXPCA to classical PCA in terms of capturing variance and hence effectiveness of dimensionality reduction. A comparison is also done with a FP-PCA method, SubPCA, upon CMU Face data. In the following, section 2 reviews SubXPCA, its summarization of variance is studied in section 3. Experimental results are in section 4 and we conclude in section 5.

2 The SubXPCA Method

Here we review SubXPCA [8] method briefly.

1. *Partitioning step:* Divide every d -dimensional pattern, \mathbf{X}_i ($i = 1, 2, \dots, N$) into k (≥ 2) equally-sized subpatterns, $\{\mathbf{X}_i^1, \mathbf{X}_i^2, \dots, \mathbf{X}_i^k\}$. Each subpattern is of size u , where $u = \lfloor \frac{d}{k} \rfloor$. For a given pattern, \mathbf{X}_i , the j^{th} subpattern, \mathbf{X}_i^j , is given by $(\mathbf{X}_i^j)_{u \times 1} = (x_{il}, x_{i(l+1)}, \dots, x_{i(l+u-1)})^T$, where $l = (j-1).u + 1$,

$1 \leq j \leq k$.

2. *Grouping step:* We pick-up j^{th} subpattern corresponding to every pattern, \mathbf{X}_i ; $i = 1, 2, \dots, N$, and form j^{th} subpattern group, \mathbf{P}^j , which is given by the matrix, $[\mathbf{P}^j]_{N \times u} = [\mathbf{X}_1^j \mathbf{X}_2^j \dots \mathbf{X}_N^j]^T$. Let $(\bar{\mathbf{X}}^j)_{u \times 1}$ be the mean of N subpatterns, $(\mathbf{P}^j)_{N \times u}$. Let (\mathbf{P}_M^j) be the mean-subtracted version of (\mathbf{P}^j) .

3. *Local feature extraction step:* For every subpattern group, \mathbf{P}^j , where $j = 1, 2, \dots, k$, repeat the following steps (a)-(d): (a) Compute local covariance matrix, $(\mathbf{C}^j)_{u \times u}$. (b) Compute eigenvalues (λ_p^j) and eigenvectors (\mathbf{e}_p^j) , where $p = 1, 2, \dots, u$, of \mathbf{C}^j . (c) Select r ($\leq u$) eigenvectors corresponding to the first r largest eigenvalues obtained in the preceding step. Let $[\mathbf{E}^j]$ be the matrix of r eigenvectors selected in this step. (d) Extract r local PCs by projecting \mathbf{P}_M^j onto \mathbf{E}^j as follows. Let \mathbf{R}^j be the reduced data in this step and is given as follows. $[\mathbf{R}^j]_{N \times r} = [\mathbf{P}_M^j]_{N \times u} \cdot [\mathbf{E}^j]_{u \times r}$

$$[\mathbf{R}^j]_{N \times r} = \begin{bmatrix} (\mathbf{X}_1^j - \bar{\mathbf{X}}^j)^T \cdot \mathbf{E}^j \\ (\mathbf{X}_2^j - \bar{\mathbf{X}}^j)^T \cdot \mathbf{E}^j \\ \vdots \\ (\mathbf{X}_N^j - \bar{\mathbf{X}}^j)^T \cdot \mathbf{E}^j \end{bmatrix} = \begin{bmatrix} (\mathbf{Y}_1^j)^T \\ (\mathbf{Y}_2^j)^T \\ \vdots \\ (\mathbf{Y}_N^j)^T \end{bmatrix},$$

where \mathbf{Y}_i^j is the locally-reduced version of $\mathbf{X}_i^j - \bar{\mathbf{X}}^j$, the j^{th} mean-subtracted subpattern of \mathbf{X}_i .

4. *Combining locally-extracted features step:*

(a) Form locally-reduced pattern, \mathbf{Y}_i by concatenating locally-reduced subpatterns, $(\mathbf{Y}_i^j)_{r \times 1}$, $\forall j = 1, 2, \dots, k$, as shown by $(\mathbf{Y}_i)_{k.r \times 1} = [(\mathbf{Y}_i^1)^T, (\mathbf{Y}_i^2)^T, \dots, (\mathbf{Y}_i^k)^T]^T$. (b) Perform global feature extraction using cross subpattern covariances of $\mathbf{Y} = \{\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_N\}$, obtained in the preceding step: (i) Compute global covariance matrix, $(\mathbf{C}^g)_{(k.r) \times (k.r)}$ for the data \mathbf{Y} . (ii) Compute eigenvalues (λ_s) and eigenvectors (\mathbf{e}_s) of \mathbf{C}^g , where $s = 1, 2, \dots, (k.r)$. (iii) Select w ($\leq k.r$) eigenvectors corresponding to first w largest eigenvalues obtained in the preceding step. Let $[\mathbf{E}^g]$ be the matrix of w eigenvectors selected in this step. (iv) Extract w global PCs by projecting \mathbf{Y} onto \mathbf{E}^g . Let \mathbf{Z} be the data obtained after projection in this step and is given as $[\mathbf{Z}]_{N \times w} = [\mathbf{Y}]_{N \times k.r} \cdot [\mathbf{E}^g]_{k.r \times w}$. We finally obtained $[\mathbf{Z}]_{N \times w}$ which is the reduced form of $(\mathbf{X})_{N \times d}$ and \mathbf{Z} is used for subsequent tasks such as classification.

3 A Theoretical Investigation on SubX-PCA

It is known that, PCA is an optimal linear scheme in terms of summarization of variance. In this section, we study the summarization of variance of SubXPCA. Here we use the notation of section 2.

Lemma 1 $(\mathbf{Y}_i)_{k.r \times 1} = (\mathbf{V})_{k.r \times d}^T \cdot (\mathbf{X}_i - \bar{\mathbf{X}})_{d \times 1}$. Here, \mathbf{Y}_i is the locally-reduced pattern of \mathbf{X}_i ; $\bar{\mathbf{X}}$ is the mean of patterns, \mathbf{X} ; \mathbf{V} is the combined matrix of selected $k.r$ local eigenvectors which is given by

$$(\mathbf{V})_{d \times k.r} = \begin{bmatrix} (\mathbf{E}^1)_{u \times r} & (\mathbf{0})_{u \times r} & \dots & (\mathbf{0})_{u \times r} \\ (\mathbf{0})_{u \times r} & (\mathbf{E}^2)_{u \times r} & \dots & (\mathbf{0})_{u \times r} \\ \vdots & \vdots & \ddots & \vdots \\ (\mathbf{0})_{u \times r} & (\mathbf{0})_{u \times r} & \dots & (\mathbf{E}^k)_{u \times r} \end{bmatrix}$$

Proof 1 The locally-reduced pattern, $(\mathbf{Y}_i)_{k.r \times 1}$ (Steps 3-4 of section 2) is obtained as follows.

$$\begin{aligned} (\mathbf{Y}_i)_{k.r \times 1} &= \begin{bmatrix} (\mathbf{E}^1)_{r \times u}^T \cdot (\mathbf{X}_i^1 - \bar{\mathbf{X}}^1)_{u \times 1} \\ (\mathbf{E}^2)_{r \times u}^T \cdot (\mathbf{X}_i^2 - \bar{\mathbf{X}}^2)_{u \times 1} \\ \vdots \\ (\mathbf{E}^k)_{r \times u}^T \cdot (\mathbf{X}_i^k - \bar{\mathbf{X}}^k)_{u \times 1} \end{bmatrix} \\ &= \begin{bmatrix} (\mathbf{E}^1)_{u \times r} & \dots & (\mathbf{0})_{u \times r} \\ (\mathbf{0})_{u \times r} & \dots & (\mathbf{0})_{u \times r} \\ \vdots & \vdots & \vdots \\ (\mathbf{0})_{u \times r} & \dots & (\mathbf{E}^k)_{u \times r} \end{bmatrix}^T \cdot \begin{bmatrix} \mathbf{X}_i^1 - \bar{\mathbf{X}}^1 \\ \mathbf{X}_i^2 - \bar{\mathbf{X}}^2 \\ \vdots \\ \mathbf{X}_i^k - \bar{\mathbf{X}}^k \end{bmatrix} \\ &= (\mathbf{V})_{k.r \times d}^T \cdot \begin{bmatrix} \mathbf{X}_i^1 - \bar{\mathbf{X}}^1 \\ \vdots \\ \mathbf{X}_i^k - \bar{\mathbf{X}}^k \end{bmatrix}_{d \times 1} \\ &\Rightarrow (\mathbf{Y}_i)_{k.r \times 1} = (\mathbf{V})_{k.r \times d}^T \cdot [\mathbf{X}_i - \bar{\mathbf{X}}]_{d \times 1}. \end{aligned}$$

Lemma 2 $\mathbf{V}_{d \times d} \cdot \mathbf{V}_{d \times d}^T = \mathbf{I}_d = \mathbf{V}_{d \times d}^T \cdot \mathbf{V}_{d \times d}$. Here, \mathbf{V} is the combined matrix of all d local eigenvectors of $\{\mathbf{P}^j\}$, $j = 1, 2, \dots, k$, which is given by

$$(\mathbf{V})_{d \times d} = \begin{bmatrix} (\mathbf{E}^1)_{u \times u} & (\mathbf{0})_{u \times u} & \dots & (\mathbf{0})_{u \times u} \\ (\mathbf{0})_{u \times u} & (\mathbf{E}^2)_{u \times u} & \dots & (\mathbf{0})_{u \times u} \\ \vdots & \vdots & \ddots & \vdots \\ (\mathbf{0})_{u \times u} & (\mathbf{0})_{u \times u} & \dots & (\mathbf{E}^k)_{u \times u} \end{bmatrix}$$

Proof 2 $\mathbf{V}_{d \times d} \cdot \mathbf{V}_{d \times d}^T =$

$$\begin{aligned} &\begin{bmatrix} (\mathbf{E}^1)_{u \times u} \cdot (\mathbf{E}^1)_{u \times u}^T & \dots & (\mathbf{0})_{u \times u} \\ (\mathbf{0})_{u \times u} & \dots & (\mathbf{0})_{u \times u} \\ \vdots & \vdots & \vdots \\ (\mathbf{0})_{u \times u} & \dots & (\mathbf{E}^k)_{u \times u} \cdot (\mathbf{E}^k)_{u \times u}^T \end{bmatrix}_{d \times d} \\ &= \begin{bmatrix} (\mathbf{I}_{u \times u}) & (\mathbf{0})_{u \times u} & \dots & (\mathbf{0})_{u \times u} \\ (\mathbf{0})_{u \times u} & (\mathbf{I}_{u \times u}) & \dots & (\mathbf{0})_{u \times u} \\ \vdots & \vdots & \ddots & \vdots \\ (\mathbf{0})_{u \times u} & (\mathbf{0})_{u \times u} & \dots & (\mathbf{I}_{u \times u}) \end{bmatrix}_{d \times d} \quad (\text{because} \\ &(\mathbf{E}^j) \cdot (\mathbf{E}^j)^T = \mathbf{I}, \text{ i.e. the orthonormality of eigenvectors [5]}) \end{aligned}$$

$\Rightarrow \mathbf{V}_{d \times d} \cdot \mathbf{V}_{d \times d}^T = \mathbf{I}_{k.u \times k.u} = \mathbf{I}_{k.u}$

$\Rightarrow \mathbf{V}_{d \times d} \cdot \mathbf{V}_{d \times d}^T = \mathbf{I}_d$ (Step-1 of section 2, $d = k.u$)

Similarly we can prove $\mathbf{V}_{d \times d}^T \cdot \mathbf{V}_{d \times d} = \mathbf{I}_{d \times d}$.

Lemma 3 $(\mathbf{C}^g)_{k.r \times k.r} = \mathbf{V}_{k.r \times d}^T \cdot \mathbf{C}_{d \times d} \cdot \mathbf{V}_{d \times k.r}$. Here, $(\mathbf{C}^g)_{k.r \times k.r}$ is the cross subpattern covariance

matrix of $\mathbf{Y} = \{\mathbf{Y}_1, \dots, \mathbf{Y}_N\}$; \mathbf{V} is the combined matrix of selected ($k.r$) local eigenvectors as given in Lemma 1 and \mathbf{C} is the covariance matrix of $\mathbf{X}_{N \times d}$.

Proof 3 We compute the cross subpattern covariance matrix, \mathbf{C}^g , of the locally-reduced patterns, \mathbf{Y} as follows.

$$\begin{aligned} (\mathbf{C}^g)_{k.r \times k.r} &= \frac{1}{N} \cdot \sum_{i=1}^N [(\mathbf{Y}_i)_{k.r \times 1} \cdot (\mathbf{Y}_i^T)_{1 \times k.r}] \\ &= \frac{1}{N} \cdot \sum_{i=1}^N [(\mathbf{V}^T \cdot (\mathbf{X}_i - \bar{\mathbf{X}})) \cdot (\mathbf{V}^T \cdot (\mathbf{X}_i - \bar{\mathbf{X}}))^T] \text{ (from Lemma 1)} \\ &= \frac{1}{N} \cdot \sum_{i=1}^N [(\mathbf{V}^T \cdot (\mathbf{X}_i - \bar{\mathbf{X}})) \cdot ((\mathbf{X}_i - \bar{\mathbf{X}})^T \cdot \mathbf{V})] \\ &= \mathbf{V}^T \cdot \frac{1}{N} \cdot \sum_{i=1}^N [(\mathbf{X}_i - \bar{\mathbf{X}}) \cdot (\mathbf{X}_i - \bar{\mathbf{X}})^T] \cdot \mathbf{V} \\ &\Rightarrow \mathbf{C}^g = \mathbf{V}^T \cdot \mathbf{C} \cdot \mathbf{V}. \end{aligned}$$

Theorem 1 $\lambda_i(SX) = \lambda_i(P)$; $\forall i \in \{1, 2, \dots, d\}$, if $r = u$ (or equivalently if $k.r = d$). In other words, SubXPCA shows the same summarization of variance as classical PCA method if the number of local eigenvectors selected per subpattern is equal to the subpattern size. Here, $\lambda_i(SX)$ and $\lambda_i(P)$ are the i^{th} largest eigenvalues obtained by SubXPCA and PCA respectively.

Proof 1 It is given that $k.r = d$ or $r = u$. It is to be noted that all the matrices we use here (i.e. $\mathbf{G}, \mathbf{\Lambda}^g, \mathbf{V}, \mathbf{C}, \mathbf{I}$) are of size $d \times d$. Let \mathbf{G} and $\mathbf{\Lambda}^g$ represent eigenvector matrix and diagonal eigenvalue matrix of \mathbf{C}^g respectively. We express the eigenvalue problem [5] of \mathbf{C}^g as $\mathbf{C}^g \cdot \mathbf{G} = \mathbf{G} \cdot \mathbf{\Lambda}^g$
 $\Rightarrow \mathbf{G}^T \cdot (\mathbf{C}^g) \cdot \mathbf{G} = \mathbf{\Lambda}^g$ (because $\mathbf{G}^T \cdot \mathbf{G} = \mathbf{G} \cdot \mathbf{G}^T = \mathbf{I}$)
 $\Rightarrow \mathbf{G}^T \cdot (\mathbf{V}^T \cdot \mathbf{C} \cdot \mathbf{V}) \cdot \mathbf{G} = \mathbf{\Lambda}^g$ (from Lemma 3)
 $\Rightarrow (\mathbf{G}^T \cdot \mathbf{V}^T) \cdot \mathbf{C} \cdot (\mathbf{V} \cdot \mathbf{G}) = \mathbf{\Lambda}^g$
 $\Rightarrow (\mathbf{V} \cdot \mathbf{G})^T \cdot \mathbf{C} \cdot (\mathbf{V} \cdot \mathbf{G}) = \mathbf{\Lambda}^g$
 $\Rightarrow (\mathbf{V} \cdot \mathbf{G}) \cdot (\mathbf{V} \cdot \mathbf{G})^T \cdot \mathbf{C} \cdot (\mathbf{V} \cdot \mathbf{G}) = (\mathbf{V} \cdot \mathbf{G}) \cdot \mathbf{\Lambda}^g$
 $\Rightarrow \mathbf{C} \cdot (\mathbf{V} \cdot \mathbf{G}) = (\mathbf{V} \cdot \mathbf{G}) \cdot \mathbf{\Lambda}^g$ [because $(\mathbf{V} \cdot \mathbf{G})^T \cdot (\mathbf{V} \cdot \mathbf{G}) = \mathbf{G}^T \cdot (\mathbf{V}^T \cdot \mathbf{V}) \cdot \mathbf{G} = \mathbf{G}^T \cdot (\mathbf{I}) \cdot \mathbf{G} = \mathbf{I}$ and $(\mathbf{V} \cdot \mathbf{G}) \cdot (\mathbf{V} \cdot \mathbf{G})^T = \mathbf{V} \cdot (\mathbf{G} \cdot \mathbf{G}^T) \cdot \mathbf{V}^T = \mathbf{V} \cdot \mathbf{I} \cdot \mathbf{V}^T = \mathbf{I}$. From Lemma 2, it is clear that $\mathbf{V}^T \cdot \mathbf{V} = \mathbf{V} \cdot \mathbf{V}^T = \mathbf{I}$]
Now the eigenvalue problem of \mathbf{C}^g is reduced to eigenvalue problem of \mathbf{C} .

Therefore \mathbf{C} has the same set of eigenvalues as \mathbf{C}^g .

Lemma 4 $\lim_{r \rightarrow u} [\mathbf{V}_{d \times k.r} \cdot \mathbf{V}_{k.r \times d}^T = \mathbf{I}_d]$. Here, $\mathbf{V}_{d \times k.r}$ is the combined matrix of $k.r$ ($\leq d$) local column eigenvectors of all subpatterns as given in Lemma 1; \mathbf{I}_d is the $d \times d$ Identity matrix.

Proof 4 Step 1: We study the orthonormal property of $(\mathbf{V})_{d \times k.r}$. It is clear that $\mathbf{V}^T \cdot \mathbf{V} = \mathbf{I}_{k.r}$ because the eigenvectors are orthonormal. However, $\mathbf{V} \cdot \mathbf{V}^T \neq \mathbf{I}_d$ because $(\mathbf{V})_{d \times k.r}$ may not contain all d local eigenvectors ($k.r \leq d$) and the same is expressed as follows.
 $(\mathbf{V} \cdot \mathbf{V}^T)_{d \times d} =$

$$\begin{bmatrix} [\mathbf{E}^1 \cdot (\mathbf{E}^1)^T] & (\mathbf{0})_{u \times u} & \dots & (\mathbf{0})_{u \times u} \\ (\mathbf{0})_{u \times u} & [\mathbf{E}^2 \cdot (\mathbf{E}^2)^T]_{u \times u} & \dots & (\mathbf{0})_{u \times u} \\ \vdots & \vdots & \ddots & \vdots \\ (\mathbf{0})_{u \times u} & (\mathbf{0})_{u \times u} & \dots & [\mathbf{E}^k \cdot (\mathbf{E}^k)^T] \end{bmatrix}$$

where $[\mathbf{E}^j \cdot (\mathbf{E}^j)^T]_{u \times u} =$

$$\begin{bmatrix} \sum_{i=1}^r [e_{i1}^j \cdot e_{i1}^j] & \sum_{i=1}^r [e_{i1}^j \cdot e_{i2}^j] & \dots & \sum_{i=1}^r [e_{i1}^j \cdot e_{iu}^j] \\ \sum_{i=1}^r [e_{i2}^j \cdot e_{i1}^j] & \sum_{i=1}^r [e_{i2}^j \cdot e_{i2}^j] & \dots & \sum_{i=1}^r [e_{i2}^j \cdot e_{iu}^j] \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^r [e_{iu}^j \cdot e_{i1}^j] & \sum_{i=1}^r [e_{iu}^j \cdot e_{i2}^j] & \dots & \sum_{i=1}^r [e_{iu}^j \cdot e_{iu}^j] \end{bmatrix}$$

$j \in \{1, \dots, k\}$; e_{it}^j is the t^{th} loading of eigenvector \mathbf{e}_i^j .
Step 2: (i) Consider the first-type expression of $\mathbf{E}^j \cdot (\mathbf{E}^j)^T$ i.e. $(\sum_{i=1}^r [e_{iq}^j \cdot e_{iq}^j])$; $q \in \{1, 2, \dots, u\}$ and substitute $r = u$ in the expression.

After substituting $r = u$, we get $(\sum_{i=1}^u [e_{iq}^j \cdot e_{iq}^j])$; $\forall q \in \{1, 2, \dots, u\}$, which is equal to 1 (from orthonormal property of eigenvectors [5]).

(ii) Next consider the second-type expression of $\mathbf{E}^j \cdot (\mathbf{E}^j)^T$ i.e. $(\sum_{i=1}^r [e_{iq}^j \cdot e_{il}^j])$; $q, l \in \{1, 2, \dots, u\}$, $q \neq l$ and substitute $r = u$ in the expression.

After substituting $r = u$, we get $(\sum_{i=1}^u [e_{iq}^j \cdot e_{il}^j])$; $\forall q, l \in \{1, 2, \dots, u\}$, $q \neq l$, which is equal to 0 (from orthonormal property of eigenvectors [5]).

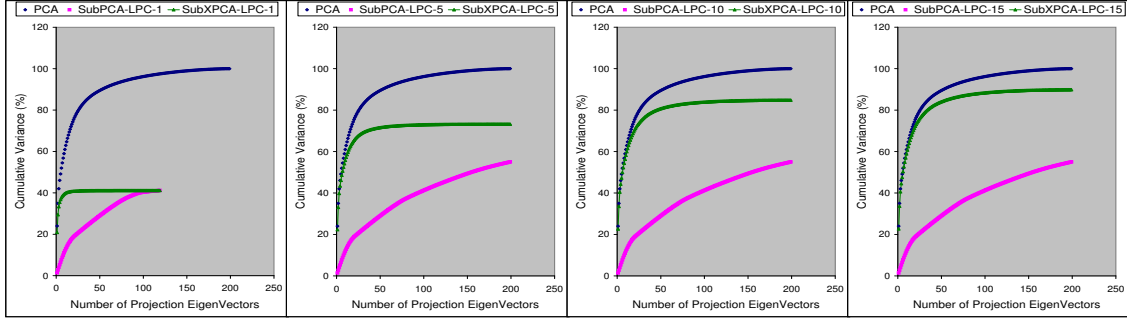
Thus, it is evident that the expression $(\sum_{i=1}^r [e_{iq}^j \cdot e_{iq}^j])$ moves closer to 1 as r increases. Similarly $(\sum_{i=1}^r [e_{iq}^j \cdot e_{il}^j])$ moves closer to 0 as r increases.
 $\Rightarrow \mathbf{E}^j \cdot [\mathbf{E}^j]^T \rightarrow \mathbf{I}_u$ as $r \rightarrow u$.

Theorem 2 $\lim_{r \rightarrow u} \sum_{i=1}^{k.r} [\lambda_i(SX)] = \sum_{i=1}^{k.r} [\lambda_i(P)]$. In other words, the summarization of variance of 'SubXPCA' tends to the summarization of variance of 'classical PCA' as the number of local eigenvectors selected per subpattern tends to the subpattern size. Here, $\lambda_i(SX)$ and $\lambda_i(P)$ are the i^{th} largest eigenvalues obtained by SubXPCA and PCA methods respectively.

Proof 2 It is to be noted that $r \leq u$. Let $(\mathbf{G})_{k.r \times k.r}$ and $(\mathbf{\Lambda}^g)_{k.r \times k.r}$ represent eigenvector matrix and diagonal eigenvalue matrix of $\mathbf{C}^g_{k.r \times k.r}$ respectively. Eigenvalue problem of \mathbf{C}^g is expressed as follows [5].

$$\begin{aligned} (\mathbf{C}^g)_{k.r \times k.r} \cdot (\mathbf{G})_{k.r \times k.r} &= (\mathbf{G})_{k.r \times k.r} \cdot (\mathbf{\Lambda}^g)_{k.r \times k.r} \\ &\Rightarrow (\mathbf{G}^T)_{k.r \times k.r} \cdot (\mathbf{C}^g)_{k.r \times k.r} \cdot \mathbf{G}_{k.r \times k.r} = (\mathbf{\Lambda}^g)_{k.r \times k.r} \\ &\text{(because } \mathbf{G}^T \cdot \mathbf{G} = \mathbf{G} \cdot \mathbf{G}^T = \mathbf{I}) \\ &\Rightarrow (\mathbf{G}^T)_{k.r \times k.r} \cdot (\mathbf{V}_{k.r \times d}^T \cdot \mathbf{C}_{d \times d} \cdot \mathbf{V}_{d \times k.r}) \cdot \mathbf{G} = (\mathbf{\Lambda}^g) \\ &\text{(from Lemma 3, } \mathbf{C}^g = \mathbf{V}^T \cdot \mathbf{C} \cdot \mathbf{V}) \\ &\Rightarrow (\mathbf{G}^T \cdot \mathbf{V}^T) \cdot \mathbf{C} \cdot (\mathbf{V} \cdot \mathbf{G}) = (\mathbf{\Lambda}^g)_{k.r \times k.r} \\ &\Rightarrow (\mathbf{V} \cdot \mathbf{G})^T \cdot \mathbf{C} \cdot (\mathbf{V} \cdot \mathbf{G}) = (\mathbf{\Lambda}^g)_{k.r \times k.r} \\ &\Rightarrow (\mathbf{V} \cdot \mathbf{G}) \cdot (\mathbf{V} \cdot \mathbf{G})^T \cdot \mathbf{C} \cdot (\mathbf{V} \cdot \mathbf{G}) = (\mathbf{V} \cdot \mathbf{G}) \cdot (\mathbf{\Lambda}^g)_{k.r \times k.r} \\ &\Rightarrow [\mathbf{V}_{d \times k.r} \cdot \mathbf{V}_{k.r \times d}^T] \cdot \mathbf{C} \cdot (\mathbf{V}_{d \times k.r} \cdot (\mathbf{G})_{k.r \times k.r}) = \\ &[(\mathbf{V})_{d \times k.r} \cdot (\mathbf{G})_{k.r \times k.r}] \cdot (\mathbf{\Lambda}^g)_{k.r \times k.r} \text{ (because } \mathbf{G} \text{ is orthonormal)} \end{aligned}$$

It is to be noted that $\mathbf{V}_{d \times k.r} \cdot \mathbf{V}_{k.r \times d}^T \neq \mathbf{I}_d$ (because $\mathbf{V}_{d \times k.r}$ does not contain all d local eigenvectors),



(a) 120 local PCs ($r=1$). (b) 200 local PCs ($r=5$) (c) 200 local PCs ($r=10$) (d) 200 local PCs ($r=15$)

Figure 1. Summarization of variance with varying no. of local PCs of CMU face data. Each image pattern is divided into 120 subpatterns (blocks). In Fig. (a) we choose 120 (1×120 PCs) local PCs. PCA uses first 200 PCs. In Figs. (b)-(d), we choose initially 600 (5×120 PCs), 1200 (10×120 PCs), 1800 (15×120 PCs) local PCs respectively. From the local PCs, we consider top 200 PCs. Note that SubXPCA's variance moves closer to PCA's variance as the number of local PCs increases, where as SubPCA's variance is lagging.

therefore the eigenvalue problem of \mathbf{C}^g is not reduced to the eigenvalue problem of \mathbf{C} .

From Lemma 4, we know that $\mathbf{V} \cdot \mathbf{V}^T$ tends to \mathbf{I}_d as r tends to u (or $k \cdot r$ tends to d).

$\Rightarrow \mathbf{C} \cdot (\mathbf{V} \cdot \mathbf{G}) \rightarrow (\mathbf{V} \cdot \mathbf{G}) \cdot \mathbf{\Lambda}^g$ as $r \rightarrow u$ (or $k \cdot r \rightarrow d$).

\Rightarrow As r tends to u , the eigenvalue problem of \mathbf{C}^g approaches to the eigenvalue problem of \mathbf{C} .

4 Experimental Results and Analysis

In this section, we demonstrate the properties of SubXPCA proposed in the previous section through our experimentation on CMU face data set [2]. CMU face data set [2] contains 624 face images (120×128) of 200 persons. We use 10 images per person generated randomly (a total of 200 images) for computing principal components.

From the Figs. 1(a)-1(d), it is clear that the summarization of variance of SubXPCA method moves closer to PCA method with increasing number of local principal components ($r = 1, 5, 10, 15; k = 120, u = 128$). The experiment using CMU face data set, demonstrates the theoretical properties (**Theorems 1–2**) proposed in the previous section. From the Figs. 1(a)-1(d), it is also clear that FP-PCA methods such as SubPCA show poor summarization of variance (i.e. lower dimensionality reduction) as compared to PCA and SubXPCA methods. The proportion of variances summarized by SubPCA and SubXPCA are computed with respect to the summation of first 200 eigenvalues obtained by PCA.

5 Concluding Remarks

We proved that SubXPCA incrementally summarizes the variance of classical PCA with increasing number of local PCs. The two methods coincide when the number of local PCs is equal to the subpattern size. This method is unique since other FP-PCA methods are incapable of approximating PCA's variance to such an extent. In a nutshell, SubXPCA method shows near-optimal summarization of variance (i.e. close to PCA) of the data by retaining the merits of FP-PCA methods.

References

- [1] S. Chen and Y. Zhu. Subpattern-based PCA. *Pattern Recognition*, 37:1081–1083, 2004.
- [2] CMU. CMU face data set. <http://kdd.ics.uci.edu/databases/faces/faces.data.html>.
- [3] C. Fredembach and et al. Eigenregions for image classification. *IEEE Trans. on PAMI*, 26:1645–1649, 2004.
- [4] R. Gottumukkal and V. K. Asari. An improved face reco. tech. based on mod. PCA approach. *PR Letters*, 2004.
- [5] R. A. Johnson and D. W. Wichern. *Applied Multivariate Statistical Analysis, Third edition*. PHI, 2001.
- [6] S. Nicole. Feedfor. neu. nets, for principal compo. extraction. *Compu. Stat. & Data Analysis*, 33, 2000.
- [7] S. A. Rizvi, T. N. Saadawi, and N. M. Nasrabadi. A clutter rejection technique for FLIR imagery using region based PCA. *Pattern Recognition*, 33, 2000.
- [8] K. VijayaKumar and A. Negi. SubXPCA and a generalized feature partitioning approach to PCA. *Pattern Recognition*, 41, No.4:1398–1409, April 2008.
- [9] J. Yang, D. Zhang, A. F. Frangi, and J. yu Yang. Two-dimensional PCA: A new appr. to appearance-based face repr. and recn. *IEEE Tran. on PAMI*, 26, 2004.