A New Breed of CPCA (Constrained Principal Component Analysis)

Yoshio Takane
Department of Psychology
McGill University

PLS-09, Beihang University (BUAA), Beijing, September 7, 2009

CPCA: A Brief Review.

- Takane & Shibayama (1991);
  Takane & Hunter (2001).
- Combines regression analysis and PCA
  in a unified framework.
- $\mathbf{X}$: a main data matrix.
  $\mathbf{G}$: a matrix of row side predictor variables
  (e.g., demographic information about subjects).
  $\mathbf{H}$: a matrix of column side predictor variables
  (e.g., descriptor variables about stimuli).

CPCA: A Brief Review.

- External Analysis: Decompose $\mathbf{X}$ by $\mathbf{G}$ and/or $\mathbf{H}$.
  (a) $\mathbf{X} = \mathbf{P}_G\mathbf{X} + \mathbf{Q}_G\mathbf{X}$,
  (b) $\mathbf{X} = \mathbf{X}\mathbf{P}_H + \mathbf{X}\mathbf{Q}_H$,
  (c) $\mathbf{X} = \mathbf{P}_G\mathbf{X}\mathbf{P}_H + \mathbf{P}_G\mathbf{X}\mathbf{Q}_H + \mathbf{Q}_G\mathbf{X}\mathbf{P}_H + \mathbf{Q}_G\mathbf{X}\mathbf{Q}_H$,
  where $\mathbf{P}_G$ and $\mathbf{P}_H$ are the orthogonal projectors onto Sp($\mathbf{G}$) and Sp($\mathbf{H}$), respectively, and $\mathbf{Q}_G$ and $\mathbf{Q}_H$ are their orthogonal complements.
- Internal Analysis: Apply SVD to terms in the above decompositions.

Motivation for New CPCA.

- Decomposition (a) takes $\mathbf{X}$ out of Sp($\mathbf{X}$), and Decomposition (b) does not yield a columnwise orthogonal decomposition.
- We develop orthogonal decompositions of $\mathbf{X}$ that stay inside Sp($\mathbf{X}$).

Four Basic Decompositions of Orthogonal Projector $\mathbf{P}_X$.

- Let $\mathbf{X}^* = \mathbf{X}(\mathbf{X}'\mathbf{X})^+$ (where $^+$ indicates the Moore-Penrose inverse) denote the matrix of dual bases of $\mathbf{X}$. and let $\mathbf{K}$ span the ortho-complement space of Sp($\mathbf{H}$) within Sp($\mathbf{X}'$) (the row space of $\mathbf{X}$), and let $\mathbf{L}$ span the ortho-complement space of Sp($\mathbf{X}'\mathbf{G}$) within Sp($\mathbf{X}'$).
- Decomposition (A): $\mathbf{P}_X = \mathbf{P}_{XH} + \mathbf{P}_{X^*K}$.
- Decomposition (B): $\mathbf{P}_X = \mathbf{P}_{X^*H} + \mathbf{P}_{XK}$.
- Decomposition (C): $\mathbf{P}_X = \mathbf{P}_{XX'G} + \mathbf{P}_{X^*L}$.
- Decomposition (D): $\mathbf{P}_X = \mathbf{P}_{P_XG} + \mathbf{P}_{XL}$.
- The two terms on the righthand sides are orthogonal in the $\mathbf{I}$ metric.

Analogous Decompositions of Data Matrix $\mathbf{X}$.

- Since $\mathbf{P}_X\mathbf{X} = \mathbf{X}$, by applying the above decompositions to $\mathbf{X}$, we obtain decompositions of $\mathbf{X}$:
- By applying (A),
  $\mathbf{X} = \mathbf{P}_{XH}\mathbf{X} + \mathbf{P}_{X^*K}\mathbf{X} = \mathbf{X}\mathbf{P}'_{SH/S^+} + \mathbf{X}\mathbf{P}'_{K/S^+}$,
  where $\mathbf{S} = \mathbf{X}'\mathbf{X}$, and $\mathbf{P}_{A/M} = \mathbf{A}(\mathbf{A}'\mathbf{M}\mathbf{A})^-\mathbf{A}'\mathbf{M}$ is the orthogonal projector onto Sp($\mathbf{A}$) in metric $\mathbf{M}$.
- By applying (B),
  $\mathbf{X} = \mathbf{P}_{X^*H}\mathbf{X} + \mathbf{P}_{XK}\mathbf{X} = \mathbf{X}\mathbf{P}'_{H/S^+} + \mathbf{X}\mathbf{P}'_{SK/S^+}$.
- By applying (C),
  $\mathbf{X} = \mathbf{P}_{TG}\mathbf{X} + \mathbf{P}_{X^*L}\mathbf{X} = \mathbf{X}\mathbf{P}'_{SX'G/S^+} + \mathbf{X}\mathbf{P}'_{L/S^+}$.
- By applying (D),
  $\mathbf{X} = \mathbf{P}_{P_XG}\mathbf{X} + \mathbf{P}_{XL}\mathbf{X} = \mathbf{X}\mathbf{P}'_{X'G/S^+} + \mathbf{X}\mathbf{P}'_{SL/S^+}$.

Decompositions of $\mathbf{X}$ Continued.

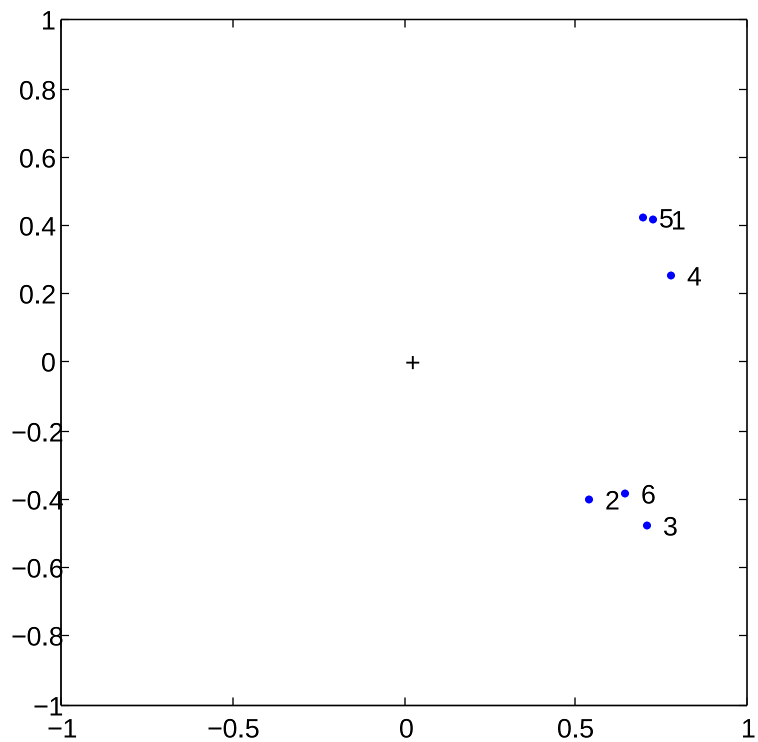
- The effects of the righthand side projectors can be transferred to the left hand sides (i.e., $\mathbf{X}\mathbf{P}_{X'} = \mathbf{X}$, where $\mathbf{P}_{X'}$ is the orthogonal projector onto $\mathrm{Sp}(\mathbf{X}')$).
- The two terms in the middle and the righthand sides of the above four decompositions are term by term equal.
- All of the terms in the above decompositions are in $\mathrm{Sp}(\mathbf{X})$.
- Each term in the above decompositions may be subject to SVD.

Data Decomposition (A).

- Not entirely new. Used by Guttman (1944, 1952)
- In the context of factor extraction by he group centroid method, where the matrix **H** looked something like

$$\mathbf{H} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

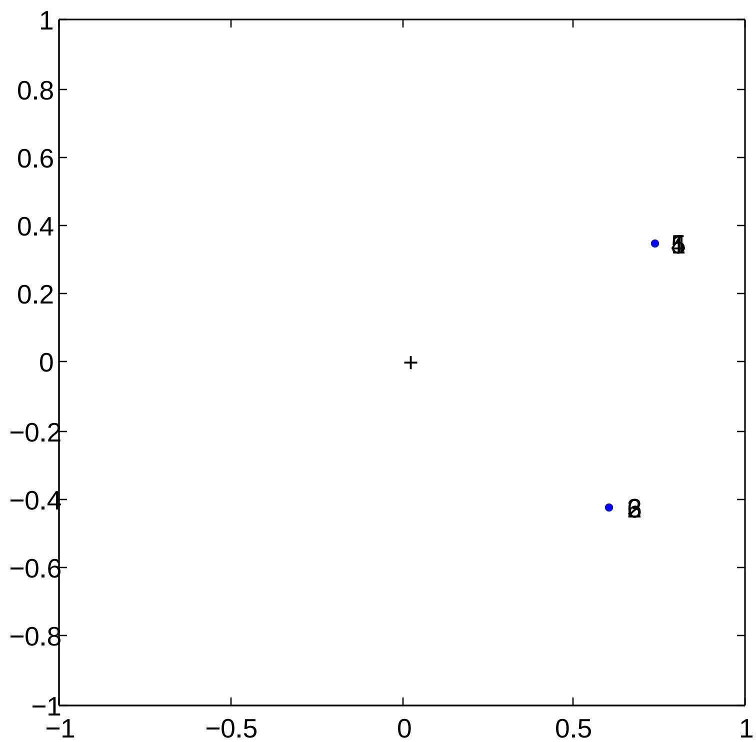
with 6 observed variables, which split into two groups, the first 3 defining Factor 1, and the last 3 Factor 2.



1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1
−1
−0.5
0
0.5
1
5
1
4
2
6
3

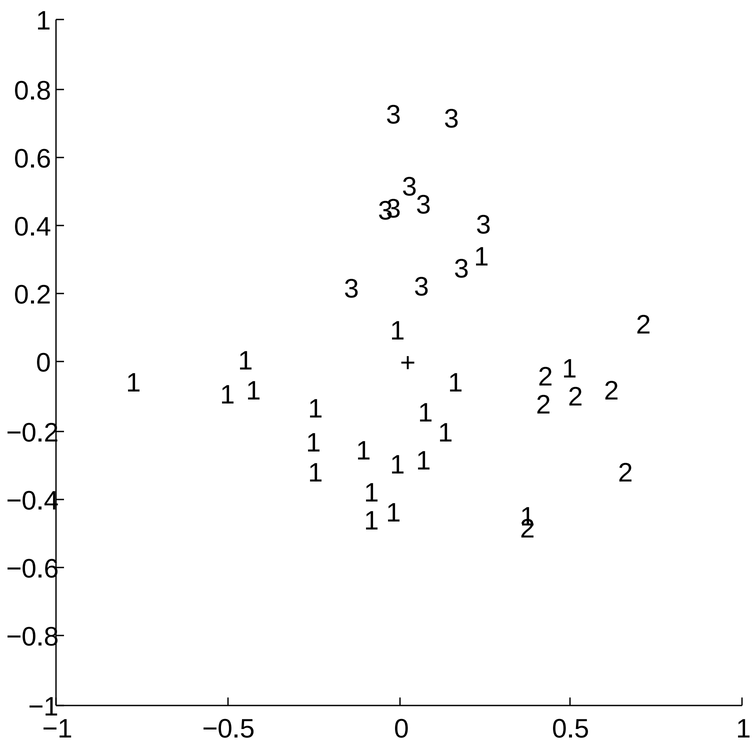
Data Decomposition (B).

- Can be derived from (A) by exchanging $\mathbf{X}$ and $\mathbf{X}^*$.
- Two possible motivations:
  (1) ML estimation in Growth Curve Models (Potthoff and Roy, 1964).
  (2) The duality relationship between $\mathbf{X}$ and $\mathbf{X}^*$. The matrix $\mathbf{H}$ applied to $\mathbf{X}$ regulates the weights applied to $\mathbf{X}$ to derive linear composites of $\mathbf{X}$, while the $\mathbf{H}$ applied to $\mathbf{X}^*$ regulates the covariance between $\mathbf{X}$ and the linear composites (e.g., if $\mathbf{H}$ is a constant vector, $\mathbf{XH}$ obtains a linear composite of $\mathbf{X}$ with equal weights, whereas $\mathbf{X}^*\mathbf{H}$ obtains a linear composite of $\mathbf{X}$ having equal covariances with all variables in $\mathbf{X}$).


1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1
−1
−0.5
0
0.5
1

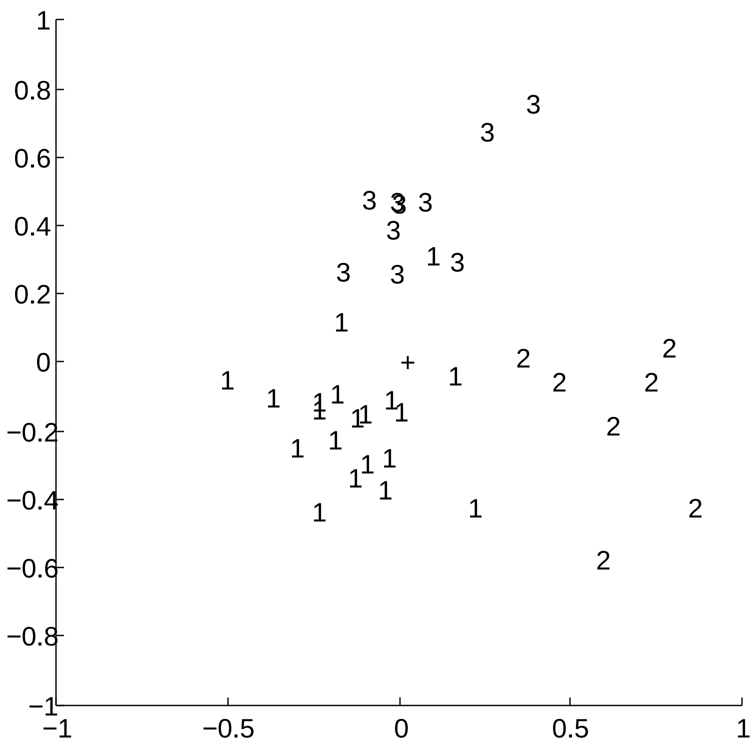
Data Decomposition (C).

- Obtained from (A) by replacing $\mathbf{H}$ by $\mathbf{X}'\mathbf{G}$.


1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1
−1
−0.5
0
0.5
1

Data Decomposition (D).

- Obtained from (B) by replacing $\mathbf{H}$ by $\mathbf{X}'\mathbf{G}$, and
- From (C) by exchanging $\mathbf{X}$ and $\mathbf{X}^*$.
- A special case of a decomposition given in Takane, Yanai, and Hwang (2006).
- The first term $\mathbf{P}_{P_X G}\mathbf{X}$ represents the part of $\mathbf{X}$ that is predictable form $\mathbf{G}$. (It represents a matrix of between-groups data matrix, if $\mathbf{G}$ indicates a matrix of indicator variables.)
- Rao (1964) presented a similar decomposition $\mathbf{X} = \mathbf{X}\mathbf{P}_{X'G} + \mathbf{X}\mathbf{Q}_{X'G}$. The two terms on the righthand side are not columnwise orthogonal.


1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1
−1
−0.5
0
0.5
1

Decomposition (AC).

- Decomposition (AC):

$$\mathbf{P}_X = \mathbf{P}_{XHH'X'G} + \mathbf{P}_{XH(H'X'XH)^- B} + \mathbf{P}_{X^*KK'X^{*\prime}G} + \mathbf{P}_{X^*K(K'X^{*\prime}X^*K)^- C},$$

  where $\mathbf{B}$ spans the ortho-complement space of Sp($\mathbf{H}'\mathbf{X}'\mathbf{G}$) within Sp($\mathbf{H}'\mathbf{X}'$), and $\mathbf{C}$ spans the ortho-complement space of Sp($\mathbf{K}'\mathbf{X}^{*\prime}\mathbf{G}$) within Sp($\mathbf{K}'\mathbf{X}^{*\prime}$).
- Applied to the left hand side of $\mathbf{X}$ to obtain the corresponding decomposition of $\mathbf{X}$. Its effect is again transferable to the righthand side. Each term in the decomposition of $\mathbf{X}$ may be subjected to SVD.

Three Other Decompositions.

- Decomposition (AD):
  $\mathbf{P}_X = \mathbf{P}_{P_{XH}G} + \mathbf{P}_{XHB} + \mathbf{P}_{P_{X^*K}G} + \mathbf{P}_{X^*KC}$, where $\mathbf{B}$ and $\mathbf{C}$ are as defined in the previous slide.
- Decomposition (BC):

$$\mathbf{P}_X = \mathbf{P}_{X^*HH'X^{*\prime}G} + \mathbf{P}_{X^*H(H'X^{*\prime}X^*H)^-B} + \mathbf{P}_{XKK'X'G} + \mathbf{P}_{XK(K'X'XK)^-C},$$

  where $\mathbf{B}$ spans the ortho-complement space of $\mathrm{Sp}(\mathbf{H}'\mathbf{X}^{*\prime}\mathbf{G})$ within $\mathrm{Sp}(\mathbf{H}'\mathbf{X}^{*\prime})$, and $\mathbf{C}$ spans the ortho-complement space of $\mathrm{Sp}(\mathbf{K}'\mathbf{X}'\mathbf{G})$ within $\mathrm{Sp}(\mathbf{K}'\mathbf{X}')$.
- Decomposition (BD):
  $\mathbf{P}_X = \mathbf{P}_{P_{X^*H}G} + \mathbf{P}_{X^*HB} + \mathbf{P}_{P_{XK}G} + \mathbf{P}_{XKC}$ where $\mathbf{B}$ and $\mathbf{C}$ are as defined immediately above.

Ridge Operators: Definition.

- Ridge Operator:

$$\mathbf{R}_X(\lambda) = \mathbf{X}(\mathbf{X}'\mathbf{X} + \lambda \mathbf{P}_{X'})^{-}\mathbf{X}' = \mathbf{X}(\mathbf{X}'\mathbf{M}_X(\lambda)\mathbf{X})^{-}\mathbf{X}',$$

  where λ is the ridge parameter (a small positive value), and $\mathbf{M}_X(\lambda) = \mathbf{P}_X + \lambda(\mathbf{X}\mathbf{X}')^{+}$ is the ridge metric matrix (Takane and Yanai, 2007; Takane 2008).

- Many similar properties to orthogonal projectors. Implications for regularized PCA.

Two Basic Decompositions.

- Decomposition (D'):

$$\mathbf{R}_X(\lambda) = \mathbf{R}_{R_X(\lambda)G}(\lambda) + \mathbf{R}_{XL}(\lambda), \quad (1)$$

where $\mathbf{L}$ spans the ortho-complement space of Sp($\mathbf{X}'\mathbf{G}$) within Sp($\mathbf{X}'$). (The two terms on the righthand side are orthogonal with respect to $\mathbf{M}_X(\lambda)$.)

- Decomposition (A'):

$$\mathbf{R}_X(\lambda) = \mathbf{R}_{XH}(\lambda) + \mathbf{R}_{X^*K}(\lambda), \quad (2)$$

where $\mathbf{X}^* = \mathbf{X}(\mathbf{X}'\mathbf{M}_X(\lambda)\mathbf{X})^+$, and Sp($\mathbf{K}$) spans the ortho-complement space of Sp($\mathbf{H}$) within Sp($\mathbf{X}'$).

- Two other decompositions analogous to Decomposition (B) and (C).

Decompositions (D') and (A') Combined.

- Decomposition (AD'):
  $\mathbf{R}_X(\lambda) = \mathbf{R}_{R_{XH}(\lambda)G} + \mathbf{R}_{XHB}(\lambda) + \mathbf{R}_{R_{X*K}(\lambda)G}(\lambda) + \mathbf{R}_{X*KC}(\lambda)$,
  where $\mathbf{B}$ spans the ortho-complement space of $\mathrm{Sp}(\mathbf{H}'\mathbf{X}'\mathbf{G})$ within $\mathrm{Sp}(\mathbf{H}'\mathbf{X}')$, and $\mathbf{C}$ spans the ortho-complement space of $\mathrm{Sp}(\mathbf{K}'\mathbf{X}^{*'}\mathbf{G})$ within $\mathrm{Sp}(\mathbf{K}'\mathbf{X}^{*'})$. (The four terms in the decomposition are orthogonal with respect to $\mathbf{M}_X(\lambda)$.)

Concluding Remarks.

- A lot to explore.