Directionally Paired Principal Component Analysis (DP-PCA) for Bivariate Estimation Problems

Yifei Fan,[1] Navdeep Dahiya,[1] Samuel Bignardi,[1] Romeil Sandhu,[2] and Anthony Yezzi[1]

[1]School of Electrical and Computer Engineering, Georgia Institute of Technology; [2]Computer Science Department, Stony Brook University

ABSTRACT

We propose Directionally Paired Principal Component Analysis (DP-PCA), an optimal linear dimension-reduction model for estimating coupled yet partially observable variable sets.

- Directly minimizes prediction errors rather than maximizing cov/corr
- Lower prediction errors compared to existing linear cross-decomposition methods (PLS/CCA [1, 2])

OBJECTIVE

Data: N data samples,

Observable: M_1 dimensions,

Unobservable (at testing): M_2 dimensions,

M_1, M_2 are too large for direct regression

Training

$$\begin{bmatrix} x_{1,1}^{\text{train}} & & x_{1,N}^{\text{train}} \\ \vdots & \cdots & \vdots \\ x_{M_1,1}^{\text{train}} & & x_{M_1,N}^{\text{train}} \\ \hdashline y_{1,1}^{\text{train}} & & y_{1,N}^{\text{train}} \\ \vdots & \cdots & \vdots \\ y_{M_2,1}^{\text{train}} & & y_{M_2,N}^{\text{train}} \end{bmatrix} \xrightarrow[\text{reduction}]{\text{dimension}} \underset{\begin{pmatrix} U \\ V \end{pmatrix}}{\begin{bmatrix} a_{1,1}^{\text{train}} & \cdots & a_{1,N}^{\text{train}} \\ \vdots & \ddots & \vdots \\ a_{L,1}^{\text{train}} & \cdots & a_{L,N}^{\text{train}} \end{bmatrix}}$$

Testing

$$\underset{U,\ V}{[\mathbf{x}]_{M_1 \times N}^{\text{test}}} \xrightarrow[\text{reduction}]{\text{dimension}} [\mathbf{a}]_{L \times N}^{\text{test}} \xrightarrow[\text{transform}]{\text{inverse}} \begin{bmatrix} [\hat{\mathbf{x}}]_{M_1 \times N}^{\text{test}} \\ [\hat{\mathbf{y}}]_{M_2 \times N}^{\text{test}} \end{bmatrix}$$

PROPOSED DP-PCA

Least squares formulation (*optimalY*)

$$\varepsilon_Y(U, V) = \frac{1}{N} \sum_{n=1}^{N} \|\mathbf{y}_n - VU^T \mathbf{x}_n\|^2$$

Optimality conditions

$$\begin{cases} XX^T U = XY^T V (V^T V)^{-1} \\ V = YX^T U (U^T XX^T U)^{-1} \end{cases}$$

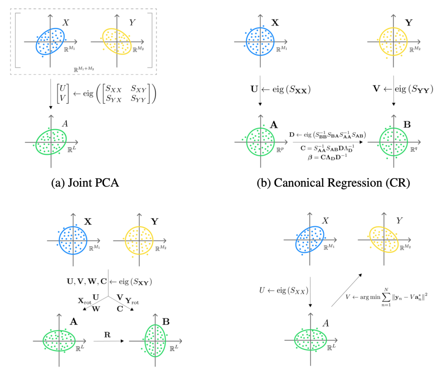
(Derivation in Section II.)

Solution steps:

1. Solve eigenvalue problem on the $N \times N$ matrix $YY^T: Y^T YZ = ZD$
2. Solve $X^T U = Z$ for U. (Z with size $N \times L$ contains L eigenvectors.)
3. Plug in optimality condition for V

(Dependently Coupled PCA (DP-PCA): obtain U via PCA on X. Concurrent work [3].)

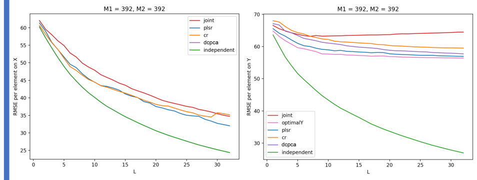
Comparison with Related Approaches



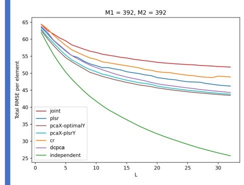
(a) Joint PCA (b) Canonical Regression (CR) (c) Partial Least Squares Regression (PLSR) (d) proposed DP-PCA

EVALUATION

Reconstruction and Prediction Errors



Reconstruction errors on X_{test} Prediction errors on Y_{test}



Total errors on $\{X_{\text{test}}, Y_{\text{test}}\}$

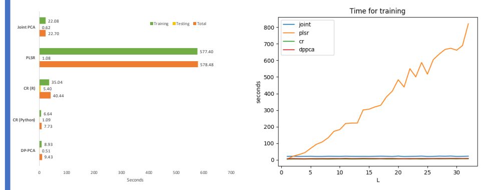
MNIST ($M_1 = M_2 = 392$)

Independent PCA: lower bound (infeasible)

PLSR [1], joint PCA, CR [4]

Execution Time and Storage

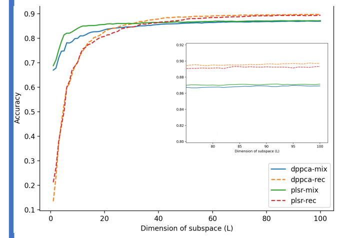
DP-PCA: smaller and comparatively fast



(a) Execution time for 100 runs ($M_1 = M_2 = 128$, $L = 32$) (b) Training time for 100 runs ($M_1 = M_2 = 128$, $L = 1$ to 32)

Method	Results to be stored after training
JPCA	$\bar{X}, \bar{Y}$: mean values of training data (size M_1 and M_2); U, V: bases for X and Y (size $M_1 \times L$ and $M_2 \times L$).
PLSR	$\bar{\mathbf{X}}, \bar{\mathbf{Y}}$: mean values of training data (size M_1 and M_2); $\sigma_{\mathbf{X}}, \sigma_{\mathbf{Y}}$: std. of training data (size M_1 and M_2); $\mathbf{U}, \mathbf{X}_{\text{rot}}$: loadings and rotations for $\mathbf{X}$ (both size $M_1 \times L$); One of the following: (1) $\mathbf{V}$: loadings for $\mathbf{Y}$ (size $M_2 \times L$) and $\mathbf{R}$: regression matrix between $\mathbf{A}$ and $\mathbf{B}$ (size $L \times L$), (2) $\boldsymbol{\beta} = \mathbf{VR}$: regression coefficients (size $M_2 \times M_1$).
CR	$\bar{\mathbf{X}}, \bar{\mathbf{Y}}$: mean values of training data (size M_1 and M_2); $\sigma_{\mathbf{X}}, \sigma_{\mathbf{Y}}$: std. of training data (size M_1 and M_2); $\mathbf{U}, \mathbf{V}$: bases for $\mathbf{X}$ and $\mathbf{Y}$ (size $M_1 \times L$ and $M_2 \times L$); $\bar{\mathbf{A}}, \bar{\mathbf{B}}$: mean values in the subspace (size L); $\sigma_{\mathbf{A}}, \sigma_{\mathbf{B}}$: standard deviation in the subspace (size L); $\boldsymbol{\beta}$: correlation coefficient between $\mathbf{A}$ and $\mathbf{B}$ (size $L \times L$).
DP-PCA	same as those in the J(oint) PCA

Classification on Reconstructed data



MNIST: 2nd Half pixels are missing as Y.

Mix: $\{X_{\text{test}}, \hat{Y}_{\text{test}}\}$

Rec: $\{\hat{X}_{\text{test}}, \hat{Y}_{\text{test}}\}$

DP-PCA > PLSR

CONCLUSIONS

Optimal solutions when estimating coupled yet partially observable data using linear models:

- With two sets of bases: DP-PCA fully optimized for unobservables (Y)
- With a single set of bases: Dependently Coupled PCA (DC-PCA) fully optimized for observables (X).

REFERENCES

[1] J. A. Wegelin et al., "A survey of partial least squares (pls) methods, with emphasis on the two-block case," *University of Washington, Tech. Rep*, 2000.

[2] H.Hotelling,"Relations between two sets of variates." *Biometrika*, vol. 28, no. 3/4, pp. 321–377, 1936.

[3] N. Dahiya, Y. Fan, S. Bignardi, R. Sandhu, and A. Yezzi, "Dependently Coupled Principal Component Analysis for Bivariate Inversion Problems," *2020 25th International Conference on Pattern Recognition*. IEEE, 2020.

[4] rdrr.io/github/jmhewitt/telefit/man/cca.predict.html