Directionally Paired Principal Component Analysis 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

Contributions

- Propose Directionally Paired Principal Component Analysis (DP-PCA): optimal linear model for estimating coupled yet partially observable variables
  - Directly minimizes prediction errors rather than maximizing $\mathrm{cov}/\mathrm{corr}$
  - Lower prediction errors compared to existing linear cross-decomposition methods (PLS/CCA [1, 2])

[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.

• Propose Directionally Paired Principal Component Analysis (DP-PCA): optimal linear model for estimating coupled yet partially observable variables
• Directly minimizes prediction errors rather than maximizing cov/corr
• Lower prediction errors compared to existing linear cross-decomposition methods (PLS/CCA [1, 2])

Dimension reduction and PCA

- Complexity in raw data

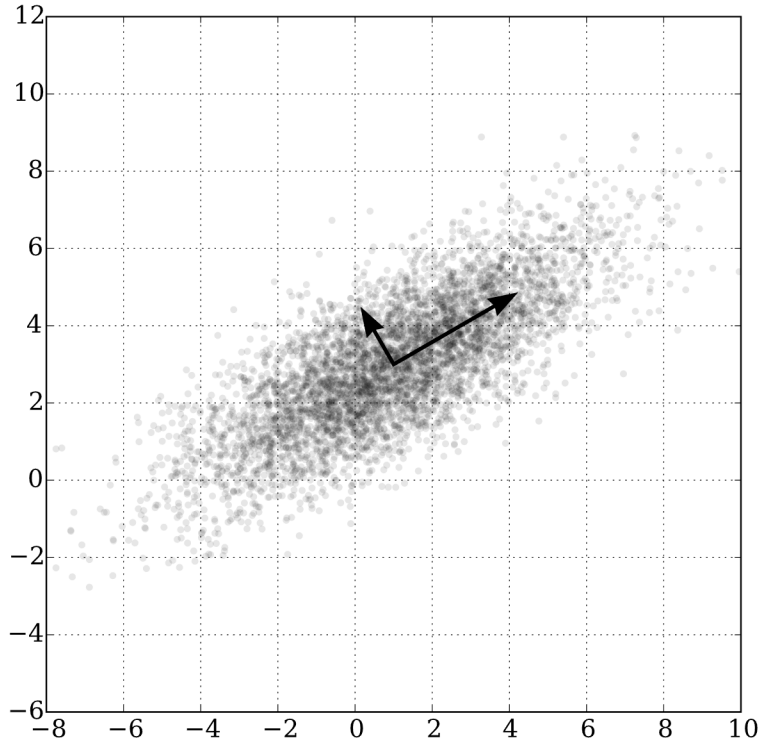
Dimension of useful feature < dimension of data

- PCA: single-variable set
  - Maximize variance

$$\mathbf{w}_{(1)} = \arg\max \left\{ \frac{\mathbf{w}^T \mathbf{X}^T \mathbf{X} \mathbf{w}}{\mathbf{w}^T \mathbf{w}} \right\}$$

  - Minimize reconstruction error

$$\varepsilon(A, U) = \frac{1}{N} \sum_{n=1}^{N} \|\mathbf{x}_n - U\mathbf{a}_n\|^2$$



PCA of a multivariate Gaussian distribution.
(source: Wikipedia)

Objective: coupled, partially observable data

- Training

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

$$\begin{pmatrix} U \\ V \end{pmatrix}$$

- Testing

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

$$\left.\begin{matrix} U \\ V \end{matrix}\right\}$$

Proposed DP-PCA

- Least squares formulation

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

- Optimality conditions

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

(Derivation in Section II.)

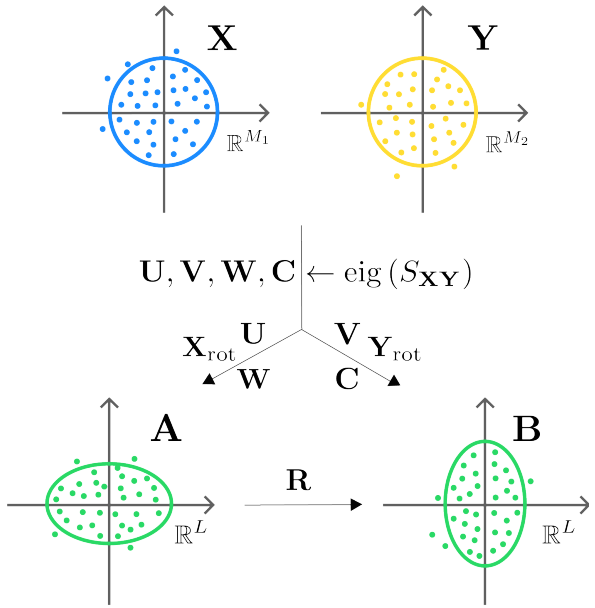
- Solution steps

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

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

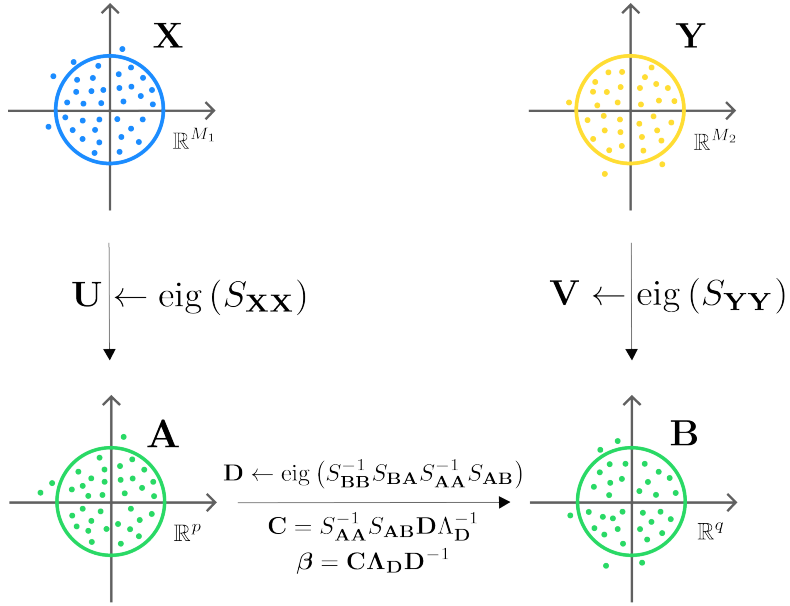
[3] N. Dahiya, Y. Fan, S. Bignardi, R. Sandhu, and A. Yezzi, "Dependently Coupled Principal Component Analysis for Bivariate Inversion Problems," *2020 25th International Conference on Pattern Recognition*. IEEE, 2020.

Comparison with related approaches



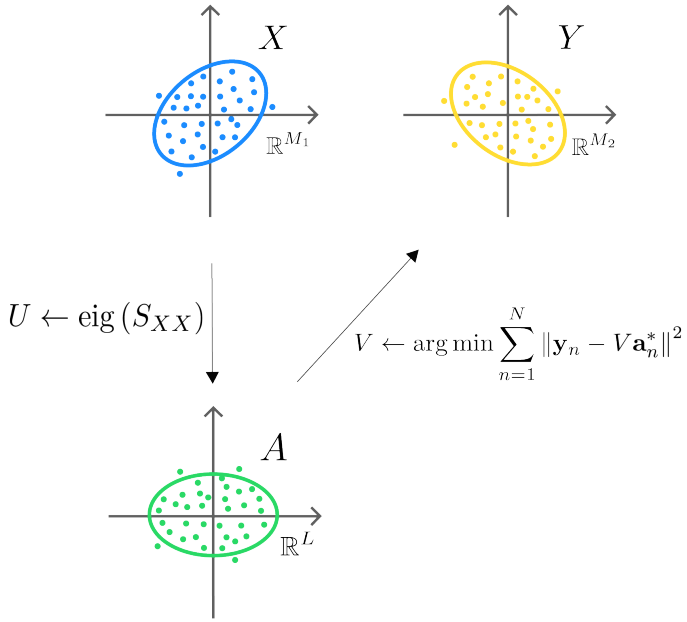
Partial Least Square Regression (PLSR)

- Paired loadings $\mathbf{P}, \mathbf{Q}$ and weights $\mathbf{W}, \mathbf{C}$ by maximizing covariance of standardized data
- Regression in subspaces



Canonical Regression (CR)

- Paired orthonormal bases $\mathbf{U}, \mathbf{V}$ by max variance of standardized data
- Regression and maximizing covariance terms in subspaces



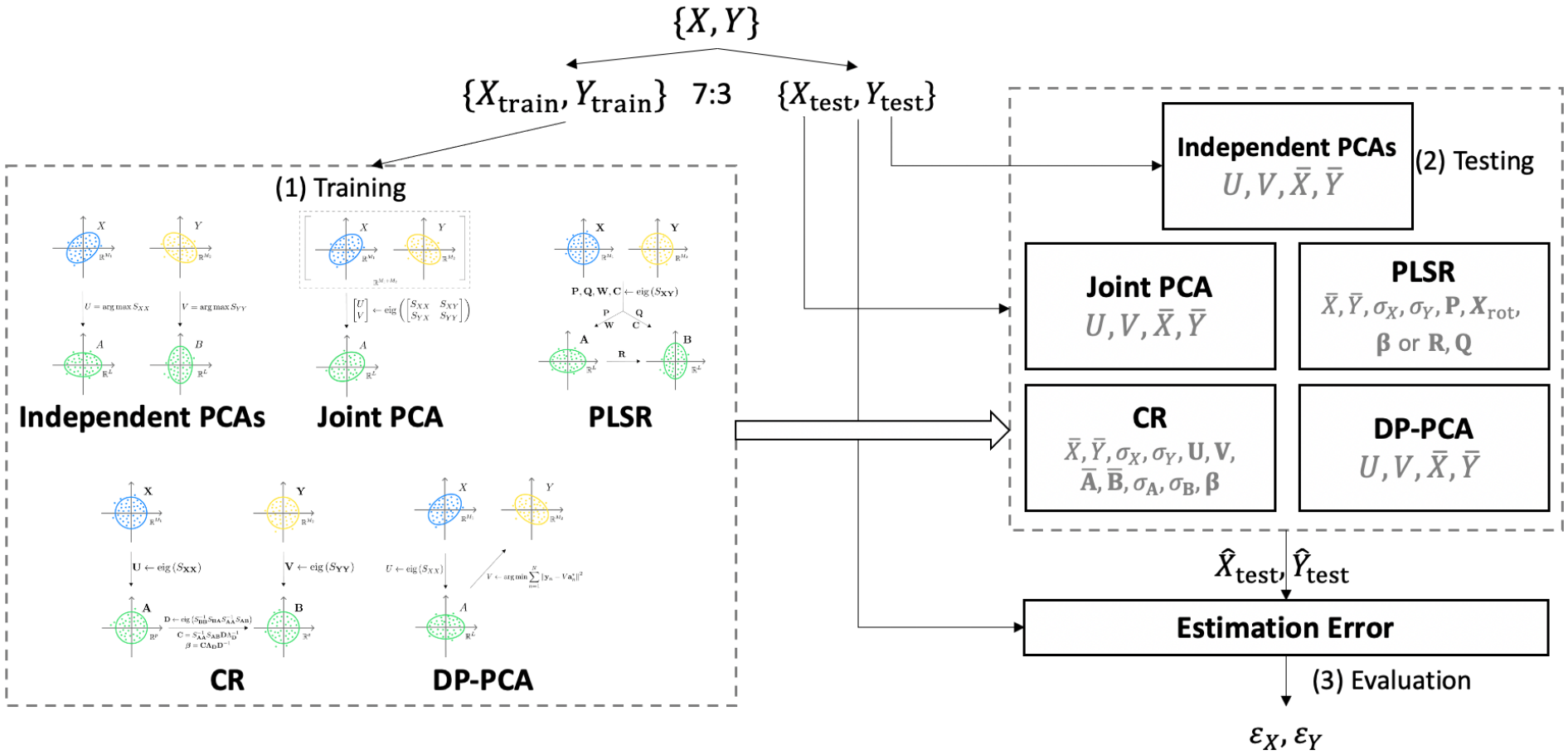
Proposed DP-PCA

- Orthonormal basis U for X by maximizing variance
- Paired basis V for Y by minimizing reconstruction error under shared coefficient $\mathbf{a}_n^*$

Comparison on storage requirement

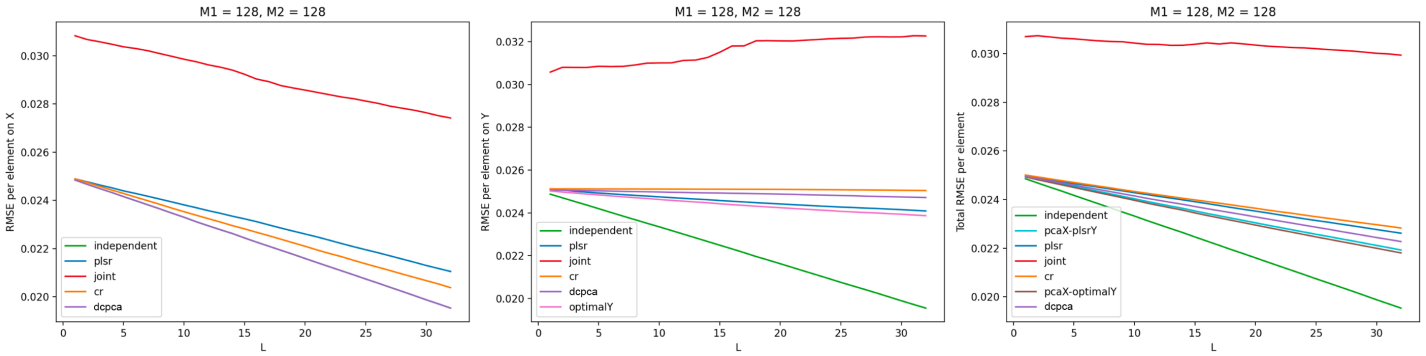
Method	Results to be stored after training
Joint PCA	$\bar{X}, \bar{Y}$: mean values of training data (with size M_1 and M_2); U, V: bases for X and Y (with size $M_1 \times L$ and $M_2 \times L$) .
PLSR	$\bar{X}, \bar{Y}$: mean values of training data (with size M_1 and M_2); σ_X, σ_Y: standard deviation of training data (with size M_1 and M_2); $\mathbf{P}, \mathbf{X}_{\text{rotations}}$: loadings and rotations for $\mathbf{X}$ (two matrices both with size $M_1 \times L$); Either (1) $\boldsymbol{\beta}$: regression coefficient for predicting $\mathbf{Y}$ from $\mathbf{X}$ (with size $M_2 \times M_1$) or (2) $\mathbf{R}$: regression matrix between $\mathbf{A}$ and $\mathbf{B}$ (with size $L \times L$), $\mathbf{Q}$: loadings for $\mathbf{Y}$ (with size $M_2 \times L$).
CR	$\bar{X}, \bar{Y}$: mean values of training data (with size M_1 and M_2); σ_X, σ_Y: standard deviation of training data (with size M_1 and M_2); $\mathbf{U}, \mathbf{V}$: loadings for X and Y (with size $M_1 \times L$ and $M_2 \times L$) $\bar{\mathbf{A}}, \bar{\mathbf{B}}$: mean values in the subspace (with size L) $\sigma_{\mathbf{A}}, \sigma_{\mathbf{B}}$: standard deviation in the subspace (with size L) $\boldsymbol{\beta}$: correlation coefficient between $\mathbf{U}$ and $\mathbf{V}$ (with size $L \times L$)
DP-PCA	Same as those in Joint PCA

Evaluation via reconstruction and prediction



Experiment result: synthetic data

- Multivariate Gaussian distribution w/ random mean and covariance



(a) Reconstruction error on observable X_{test}

(b) Prediction error on unobservable Y_{test}

(c) Total reconstruction error

Figure 6.5: Evaluation on dimension reduction via data reconstruction and prediction of coupled synthetic data. $N = 10^4$, $M_1 = M_2 = 128$, $L = 1$ to 32. Horizontal axis: dimension L of the target subspace (i.e., budget); vertical axis: reconstruction/prediction error.

Independent: lower bound, best possible

Joint PCA: worst as Y_{test} unobservable

CR: sub-optimal for both X_{test} and Y_{test}

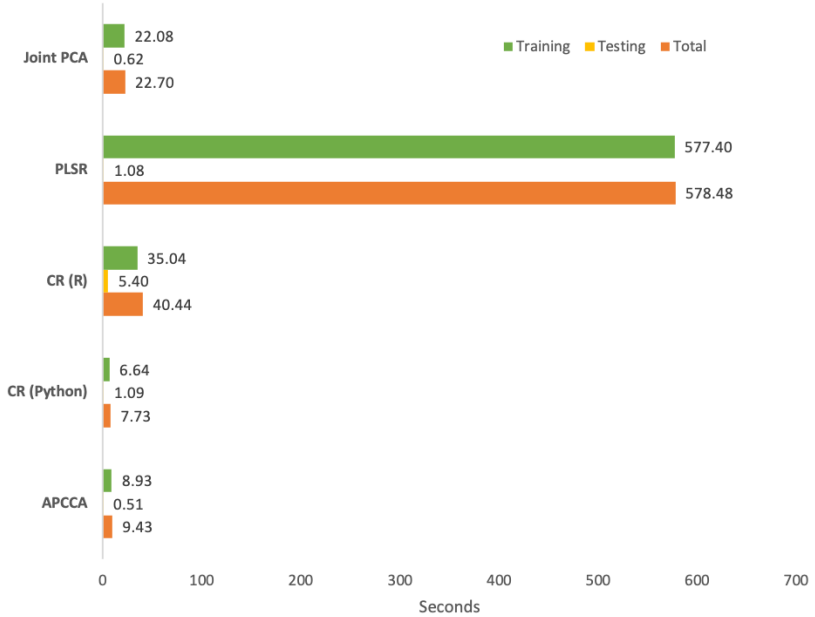
PLSR: sacrificing X_{test} for better Y_{test}

DC-PCA: best X_{test} and total, descent Y_{test}

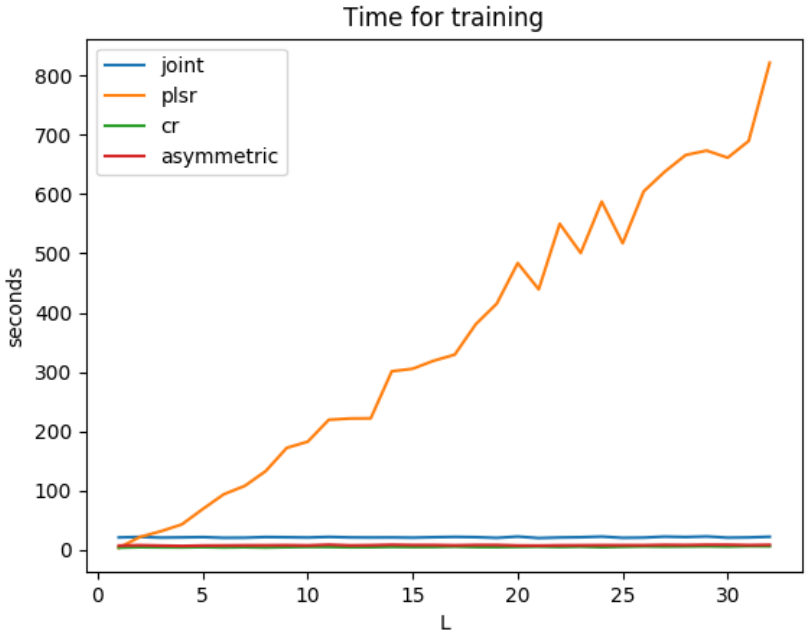
Best combination: PCA for X_{test}, and optimal Y mode of DP-PCA for Y_{test}

Execution time

- Experience with publicly available implementations



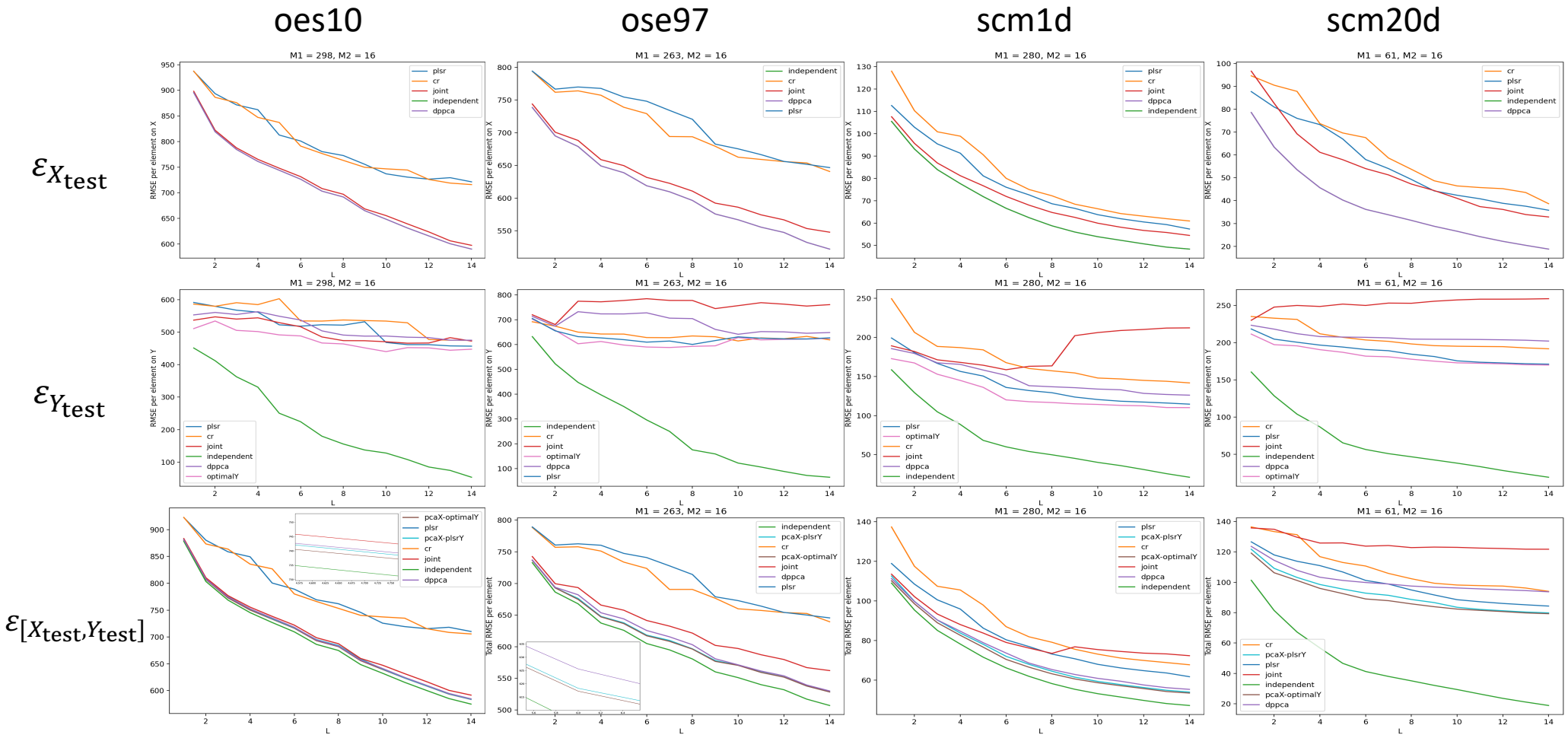
(a) Execution time for 100 runs
($M_1 = M_2 = 128$, $L = 32$)



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

Joint PCA: https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html
PLSR: https://scikit-learn.org/stable/modules/generated/sklearn.cross_decomposition.PLSRegression.html
CR (R): https://rdrr.io/github/jmhewitt/telefit/man/cca.predict.html
CR (Python): https://gist.github.com/thelittlekid/89630241f5b90a838a7b583a5836d350

Experiment result: multi-target regression



Experiment result: single-channel images

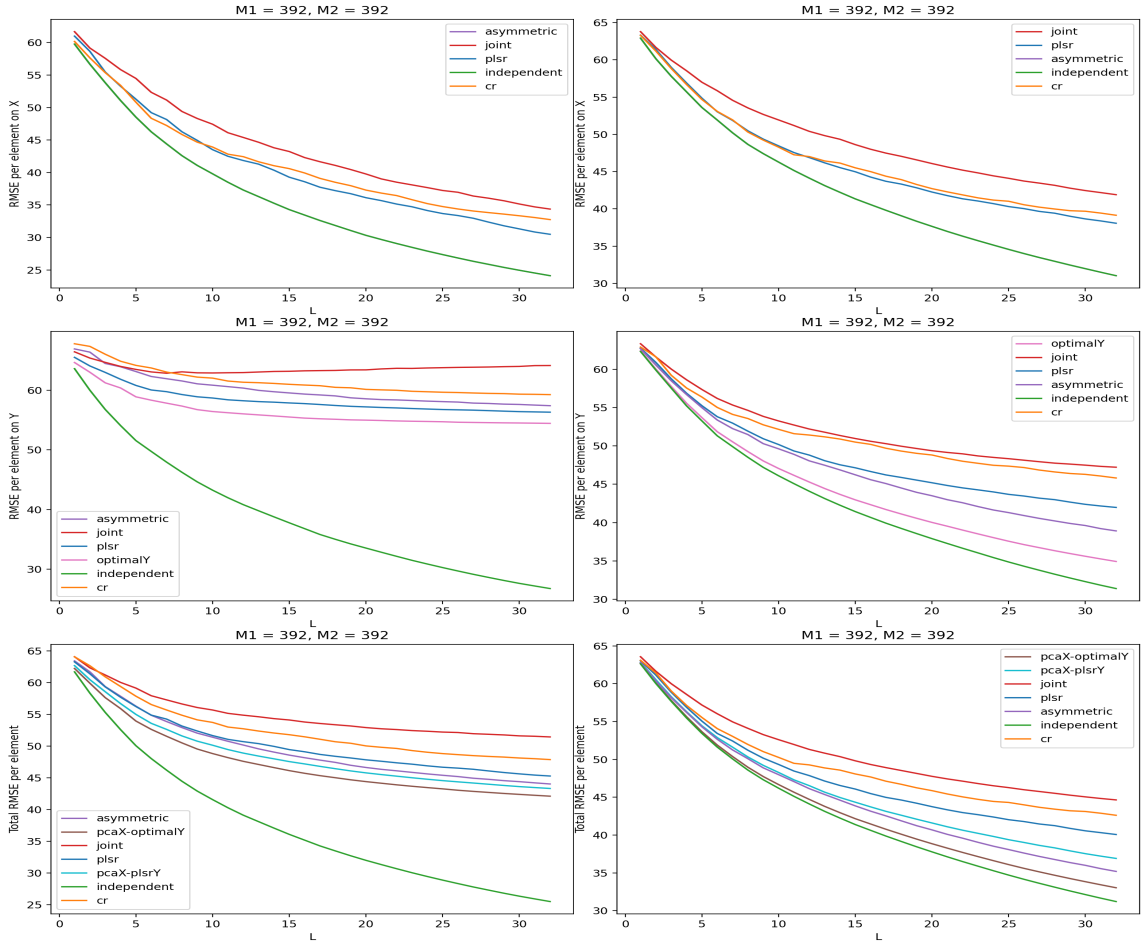
MNIST: sequential split

MNIST: random split

$\mathcal{E}_{X_{\text{test}}}$

$\mathcal{E}_{Y_{\text{test}}}$

$\mathcal{E}_{[X_{\text{test}}, Y_{\text{test}}]}$



Sequential split: according to the sequence of indices (i.e., top vs bottom)
Random split: randomly yet consistently across images

DP-PCA: for random split, achieve best results on all three under the budget of a single-pair bases.

Conclusions

- Best (combined) estimation of coupled yet partially observable data:
  - Standard PCA for the observable part X
  - Optimal Y mode of DP-PCA for the unobservable part Y

  - More storage requirement (two pairs of bases) and longer computation time

- DC-PCA: best single approach for overall estimation
  - Suitable when the unobservable are no more important than the observable
  - Lowest storage requirement and fast speed