Table 3: Parameter setting of simulated data for model comparison

Parameters	Design1	Design2
Number of observations	100	50
Number of predictors	20	40
Relevant number of predictors	5, 5, 7	15, 15, 9
Number of responses	6	5
Position of relevant components	1, 2, 3, 4, 6, 5	1, 2, 3, 4, 5, 7
Gamma (decay of eigenvalues)	0.9	0.4
Coefficient of Determination	0.8, 0.9, 0.7	0.8, 0.9, 0.5
Position of response components	1, 4, 2, 6, 3, 5	1, 4, 2, 5, 3
Number of test observations	1000	1000

rotation can be both restricted and unrestricted as discussed in previous section. The restricted rotation is carried out combining response vectors along with noise vector in a block-wise manner according to the users choice. Illustration in fig-…

Suppose, in our previous example, if response components are combined as – $\mathbf{W}_1, \mathbf{W}_4$, $\mathbf{W}_2$ and $\mathbf{W}_3, \mathbf{W}_5$. Here, any predictor variable is only relevant for $\mathbf{W}_1, \mathbf{W}_2$ and $\mathbf{W}_3$ while $\mathbf{W}_4$ and $\mathbf{W}_5$ are noise. The resulting response variables are $\mathbf{Y}_1 \ldots \mathbf{Y}_5$ where, the first and fourth response variable spans the same space as by the first response components $\mathbf{W}_1$ and noise component $\mathbf{W}_4$ and so on. Thus, the predictors and predictor space relevant for response component $\mathbf{W}_1$ is also relevant for response $\mathbf{Y}_1$ and $\mathbf{Y}_4$.

Implementation

Example of model comparison with simulated data from `simrel-m` package

In this section, `simrel-m` is implemented to simulate multi-response linear model dataset and use it to compare Principal Components Regression (PCR), Partial Least Squared Regression (PLS), Cannonically Powered Partial Least Squared Regression (CPPLS), Maximum Likelihood under Envelope Estimation and Ordinary Least Squared Regression (OLS) on the basis of their prediction ability of test observations. Here two design of parameters as in Table-3 are implemented for the simulation.

Here, the first design has large number of observations as compared to the number of variables while the number of variables in second design is nearly equals to its number of observations. Both the models have three informative response components from which first design generates 6 responses and the second design generates 5 responses. In addition, the eigenvalues of predictors decreases sharply in the first design than the second one. The prediction error are compared on the basis of 1000 test samples.

Prediction error is measured as mean squared error of prediction (MSEP) using the expression in equation~(16).

$$\text{MSEP}_{\text{train}} = \frac{\mathbf{Y}_{\text{train}} - \hat{\mathbf{Y}}_{\text{train}}}{n_{\text{train}}} \text{ and } \text{MSEP}_{\text{test}} = \frac{\mathbf{Y}_{\text{test}} - \hat{\mathbf{Y}}_{\text{test}}}{n_{\text{test}}}$$

where, $\hat{\mathbf{Y}}_{\text{train}} = \mathbf{X}_{\text{train}}\boldsymbol{\beta}$ and $\hat{\mathbf{Y}}_{\text{test}} = \mathbf{X}_{\text{test}}\boldsymbol{\beta}$

Comparison of Estimation Methods based on Prediction error

Figure-1 shows the performance of different models with components 1 to 10 ...

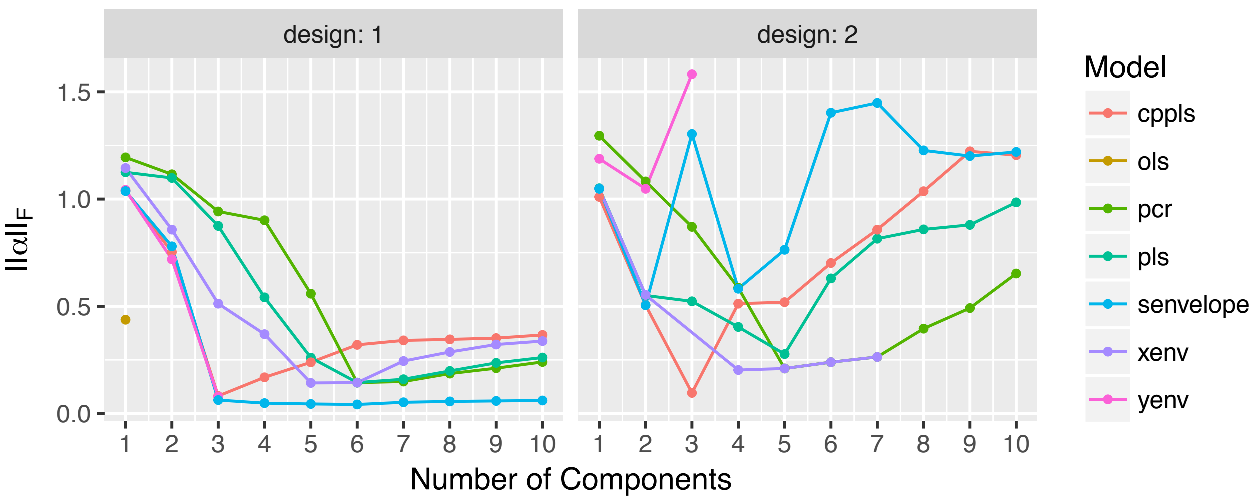

Figure 1: Model comparison based on test prediction error

Some more text and more and more

Appendix

Comprehensive Explanation of Notation

Following table present a comprehensive notation used in this paper.