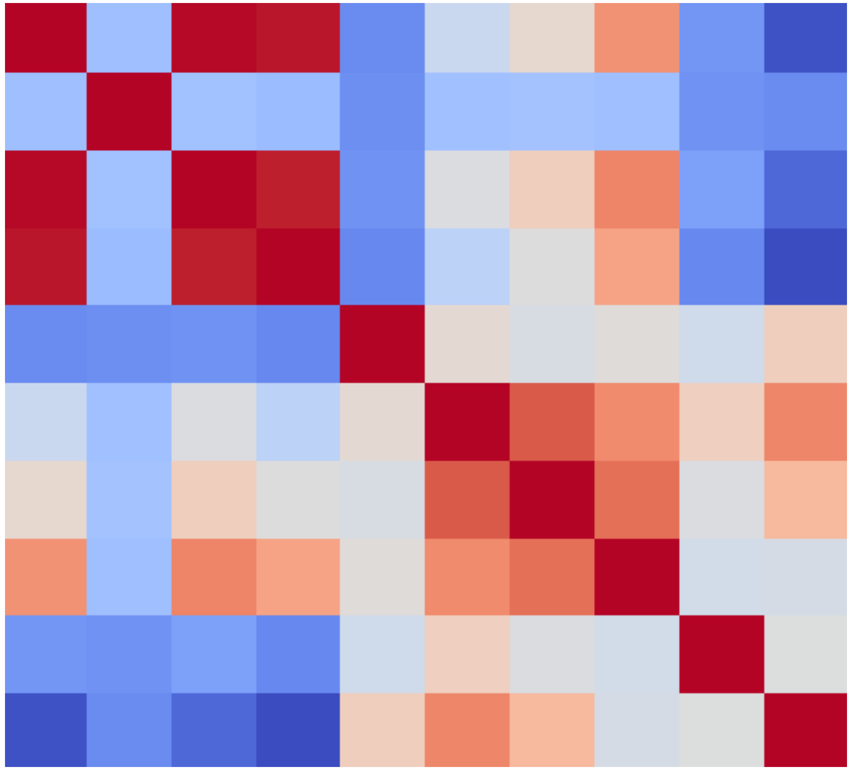
2.160 Identification, Estimation, and Learning

Part 1 Regression

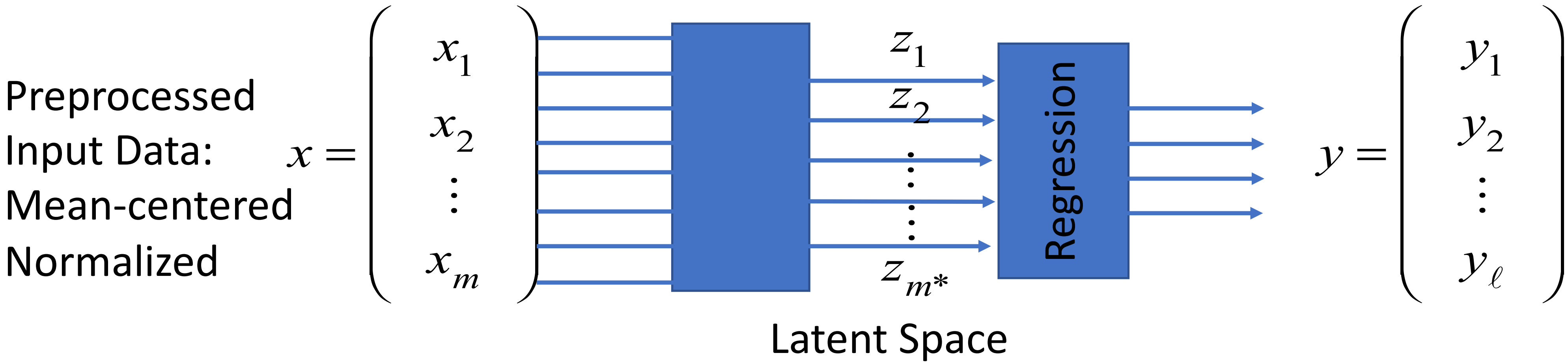


Lecture 6

Partial Least Squares Regression

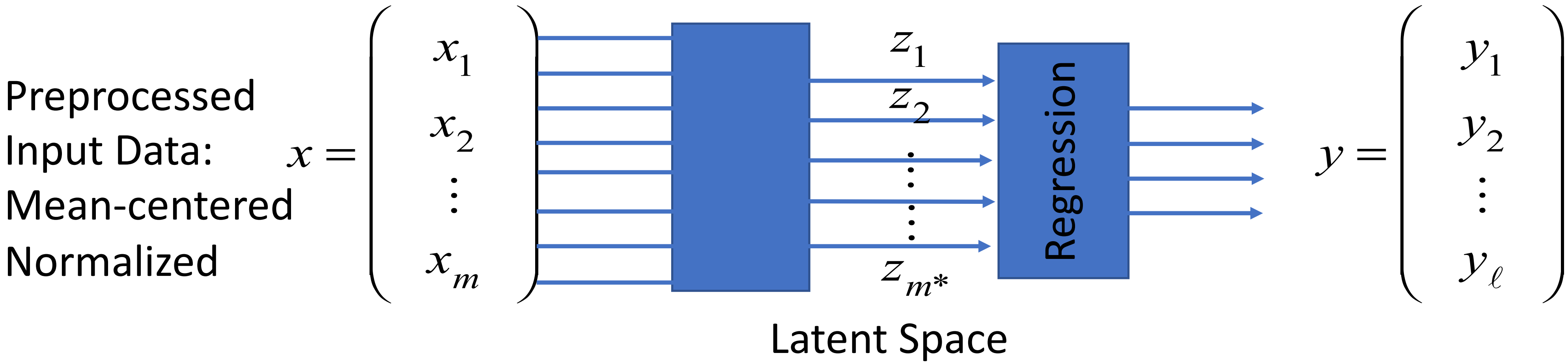
H. Harry Asada
Department of Mechanical Engineering
MIT

Latent Modeling

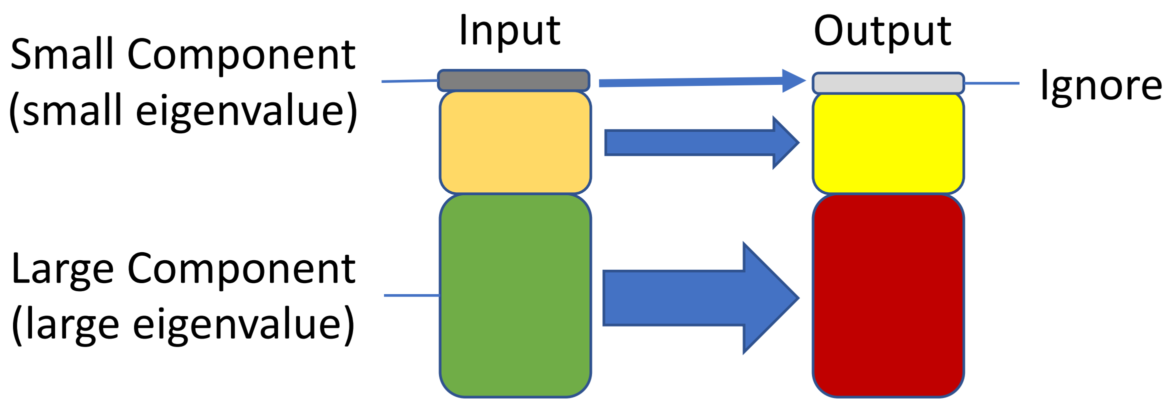


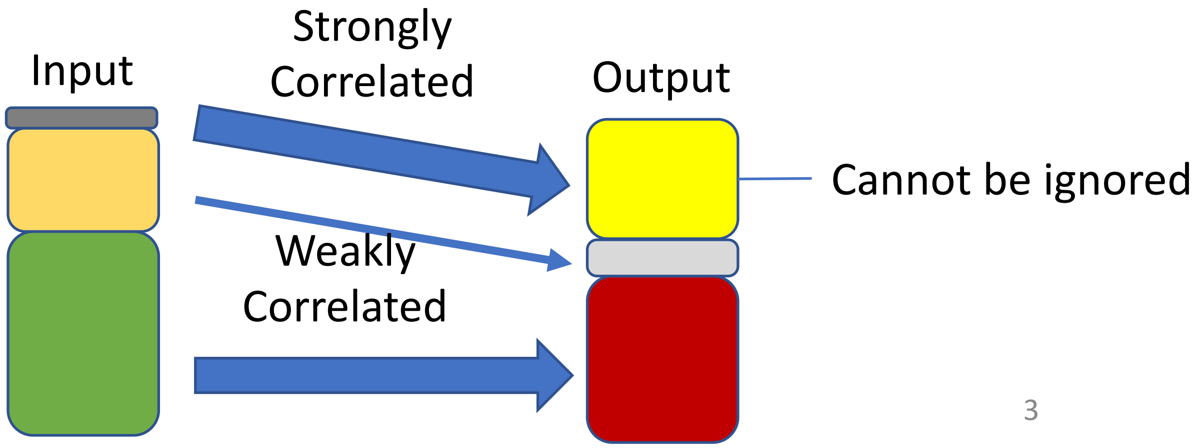
- ❑ Principal Component Regression: Characterizes the input data space, reduces the input dimension based on Principal Component Analysis, and regresses on principal components.

Latent Modeling

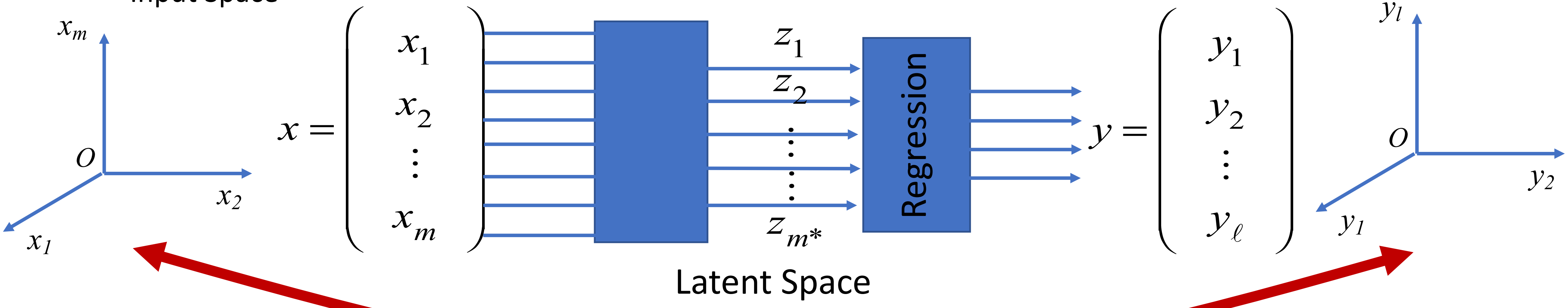


- ❑ Principal Component Regression: Characterizes the input data space, reduces the input dimension based on Principal Component Analysis, and regresses on principal components.
- ❑ Caveat! Small principal components, which are ignored, may be highly correlated with outputs. Those components must not be neglected.





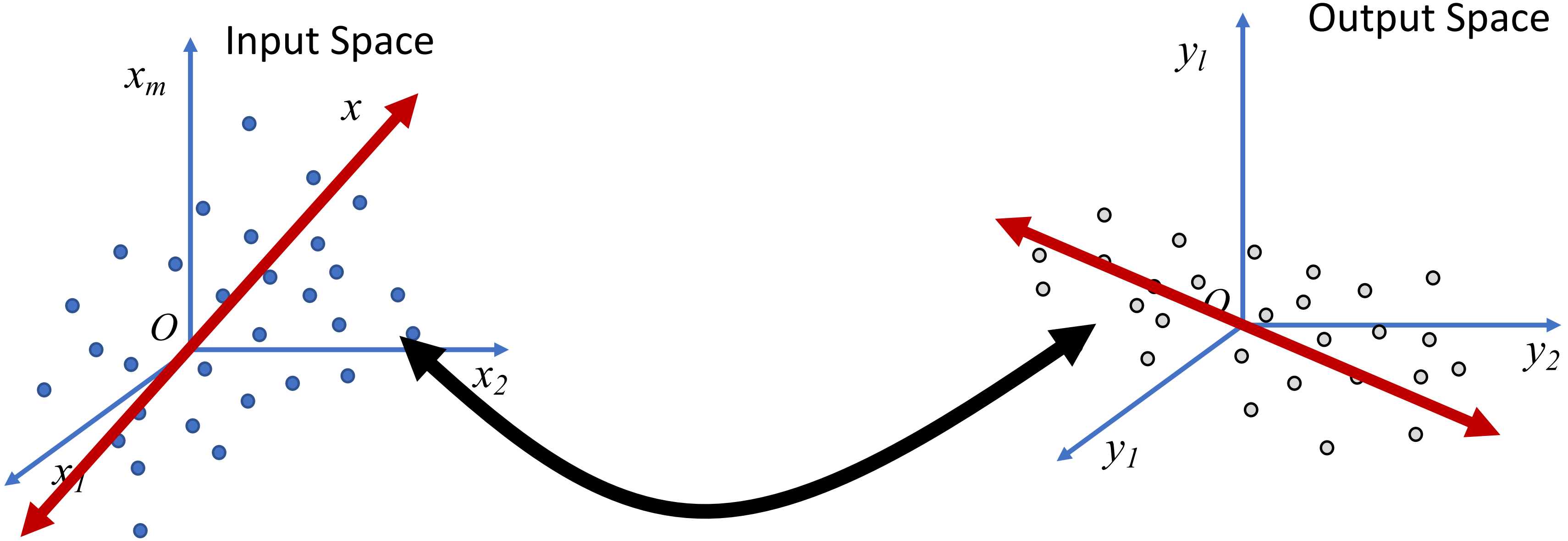
Latent Modeling



- ❑ Principal Component Regression: Characterizes the input data space, reduces the input dimension based on Principal Component Analysis, a … principal components.
- ❑ Caveat! Small principal components, which are ignored, may be highly correlated with outputs. Those components must not be neglected.
- ❑ Components having significant correlation with outputs must be involved in the latent space.
- ❑ This requires to analyze both input and output spaces, rather than characterizing the input space alone.
- ❑ Multiple Outputs: Unlike single output regressions, we often need to estimate multiple outputs, which may be correlated.
- ❑ This lecture will discuss the latent space modeling based on input – output correlation analysis.

4.3 The Core Algorithm of Multi-Input, Multi-Output Partial Least Squares Regression

Partial Least Squares Regression is a latent modeling method for predicting a set of outputs in relation to a reduced order inputs. The basic idea is to find a low-dimensional set of input space variables that is most correlated with a given set of output data. It is to analyze data in both input space and output space.



4.3 The Core Algorithm of Multi-Input, Multi-Output Partial Least Squares Regression

Step 1. Find the directions of a pair of unit vectors, $v \in \Re^m$ in the input space and $w \in \Re^\ell$ in the output space, that maximizes the correlation between the projection of input vector onto the unit vector, $z = v^T x$, and that of the output vector, $r = w^T y$.

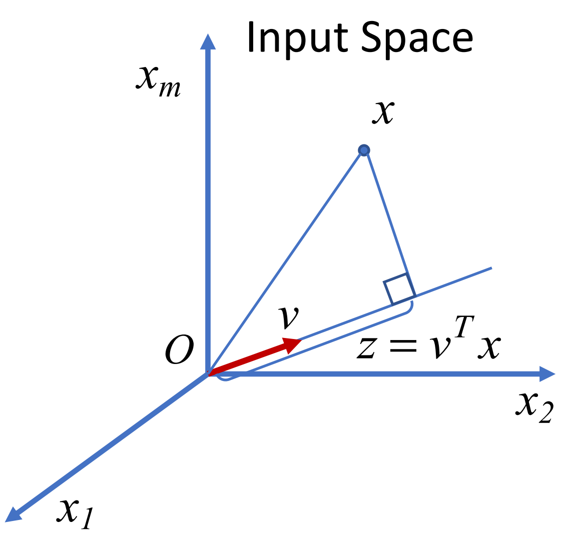
$$\max_{v,w} E[z \cdot r] = \max_{v,w} E[v^T x \cdot w^T y]$$

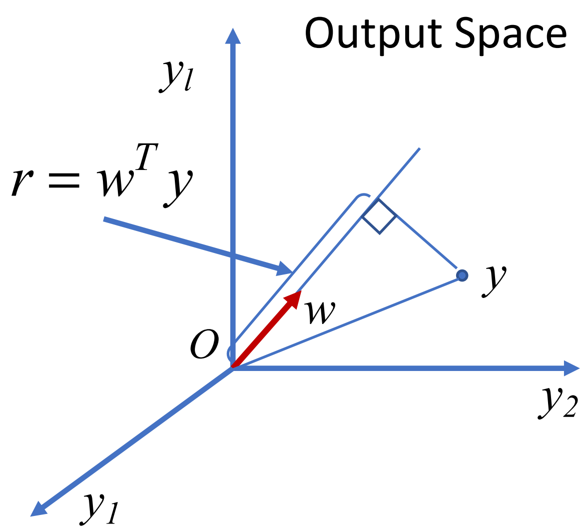
$$= \max_{v,w} v^T \underbrace{E[x\, y^T]}_{C_{XY}} w$$

where

$$|v| = 1, |w| = 1$$

Covariance of mean-centered random variables x and y.





- ❑ $z = v^T x$ is called the ***score*** of input x with respect to v, and $r = w^T y$ is called the ***score*** of output y with respect to w.

Recap: Covariance Matrix

$$C_{XY} = E[x\,y^T] = E\left[\begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}\begin{pmatrix} y_1 & \cdots & y_\ell \end{pmatrix}\right] = \begin{pmatrix} E[x_1 y_1] & \cdots & E[x_1 y_\ell] \\ \vdots & \ddots & \vdots \\ E[x_m y_1] & \cdots & E[x_m y_\ell] \end{pmatrix}$$

Note x and y are mean-centered and normalized.

Problem

$$\max_{v,w} v^T C_{XY} w \qquad \text{Subject to} \qquad |v| = 1, |w| = 1$$

Solution: Using two Lagrange multipliers for the two constraint equations,

$$(v^o, w^o) = \arg\max_{v,w} \left\{ \underbrace{v^T C_{XY} w \underbrace{- \frac{1}{2}\lambda_v (v^T v - 1) - \frac{1}{2}\lambda_w (w^T w - 1)}_{\text{2 Constraints}}}_{\text{A function of both } v \text{ and } w} \right\}$$

Solution: Using two Lagrange multipliers,

$$(v^o, w^o) = \arg\max_{v,w} \left\{ v^T C_{XY} w - \frac{1}{2}\lambda_v (v^T v - 1) - \frac{1}{2}\lambda_w (w^T w - 1) \right\}$$

The necessary conditions for v and w to maximize the correlation are:

$$\frac{\partial}{\partial v} = 0 \quad \Rightarrow \quad C_{XY} w - \lambda_v v = 0 \quad (35) \qquad \frac{\partial}{\partial w} = 0 \quad \Rightarrow \quad (C_{XY})^T v - \lambda_w w = 0 \quad (36)$$

Note that by definition: $(C_{XY})^T = C_{YX}$.

Quick clarification: Transpose of a Covariance Matrix

$$(C_{XY})^T = \begin{pmatrix} E[x_1 y_1] & \cdots & E[x_1 y_\ell] \\ \vdots & \ddots & \vdots \\ E[x_m y_1] & \cdots & E[x_m y_\ell] \end{pmatrix}^T = \begin{pmatrix} E[x_1 y_1] & \cdots & E[x_m y_1] \\ \vdots & \ddots & \vdots \\ E[x_1 y_\ell] & \cdots & E[x_m y_\ell] \end{pmatrix}$$

$$== \begin{pmatrix} E[y_1 x_1] & \cdots & E[y_1 x_m] \\ \vdots & \ddots & \vdots \\ E[y_\ell x_1] & \cdots & E[y_\ell x_m] \end{pmatrix} = E\left(\begin{pmatrix} y_1 \\ \vdots \\ y_\ell \end{pmatrix} \begin{pmatrix} x_1 & \cdots & x_m \end{pmatrix} \right) = C_{YX}$$

Or, simply

$$\left(\begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} \begin{pmatrix} y_1 & \cdots & y_\ell \end{pmatrix} \right)^T = \begin{pmatrix} y_1 & \cdots & y_\ell \end{pmatrix}^T \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}^T = \begin{pmatrix} y_1 \\ \vdots \\ y_\ell \end{pmatrix} \begin{pmatrix} x_1 & \cdots & x_m \end{pmatrix}$$

Solution: Using Lagrange multipliers,

$$(v^o, w^o) = \arg\max_{v,w} \left\{ v^T C_{XY} w - \frac{1}{2}\lambda_v (v^T v - 1) - \frac{1}{2}\lambda_w (w^T w - 1) \right\}$$

The necessary conditions for v and w to maximize the correlation are:

$$\frac{\partial}{\partial v} = 0 \quad \Rightarrow \quad C_{XY} w - \lambda_v v = 0 \quad (35) \qquad \frac{\partial}{\partial w} = 0 \quad \Rightarrow \quad (C_{XY})^T v - \lambda_w w = 0 \quad (36)$$

Note that by definition: $(C_{XY})^T = C_{YX}$.

From (36), $w = \frac{1}{\lambda_w} C_{YX} v$. Substituting this in (35) yields. $C_{XY} C_{YX} v = \lambda_v \lambda_w v$

This implies that vector v is an eigen vector of matrix $C_{XY} C_{YX}$

Similarly, from (35) $v = \frac{1}{\lambda_v} C_{XY} w$. Substituting this into (36) yields $C_{YX} C_{XY} w = \lambda_v \lambda_w w$

This implies that vector w is an eigen vector of matrix $C_{YX} C_{XY}$

$$C_{XY}C_{YX}v = \lambda_v \lambda_w v \qquad\qquad C_{YX}C_{XY}w = \lambda_v \lambda_w w$$

Note that in both cases the eigenvalue is the same : $\lambda_v \lambda_w$

We can show that $\lambda_v = \lambda_w$

Pre-multiplying v^T to (35), $C_{XY}w - \lambda_v v = 0$

$$v^T C_{XY} w - \lambda_v v^T v = 0 \quad \therefore \lambda_v = v^T C_{XY} w$$

Pre-multiplying w^T to (36), $(C_{YX})^T v - \lambda_w w = 0$

$$w^T (C_{XY})^T v - \lambda_w w^T w = 0 \quad \therefore \lambda_w = w^T (C_{XY})^T v = v^T (C_{XY}) w = \lambda_v$$

$$\therefore \lambda_v = \lambda_w$$

$C_{XY}C_{YX}$ and $C_{YX}C_{XY}$ have the same eigenvalues. $\lambda_v = \lambda_w = \lambda$

Singular Value Decomposition --- Extension of Eigen Decomposition

A matrix $A \in \Re^{m\times\ell}$ can be decomposed to

$$A = VDW^T$$

where

$$V = \begin{bmatrix} v_1 & v_2 & \cdots & v_m \end{bmatrix} \in \Re^{m\times m}$$

v_i = the i-th left singular vector of matrix A
= the eigenvectors of matrix $AA^T \in \Re^{m\times m}$, which is a real, symmetric matrix having all real eigenvalues and eigen vectors.

$$W = \begin{bmatrix} w_1 & w_2 & \cdots & w_m \end{bmatrix} \in \Re^{\ell\times\ell}$$

w_i = the i-th right singular vector of matrix A
= the eigenvectors of matrix $A^T A \in \Re^{\ell\times\ell}$, which is a real, symmetric matrix having all real eigenvalues and eigen vectors.

$$D = \begin{pmatrix} s_1 & 0 & \cdots & 0 \\ 0 & s_2 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & s_\ell \\ 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{pmatrix} \in \Re^{m\times\ell}$$

: a rectangular diagonal matrix

s_i = the i-th singular value of matrix A
= the square root of the non-zero eigenvalue of matrix $AA^T \in \Re^{m\times m}$, or $A^T A \in \Re^{\ell\times\ell}$, both are real-symmetric, positive semi-definite matrices with non-negative eigenvalues.

Theorem

The unit vectors, v_0 and w_0, that maximize the correlation between input and output scores, $z = v^T x$ and $r = w^T y$, are the left and right singular vectors, respectively, associated with the largest singular value of the cross-correlation matrix C_{XY}.

$$C_{XY} = \begin{bmatrix} v_1 & \cdots\cdots \end{bmatrix} \begin{bmatrix} s_1 & \cdots & 0 \\ \vdots & * & 0 \\ 0 & 0 & \end{bmatrix} \begin{bmatrix} w_1^T \\ \vdots \\ \vdots \end{bmatrix}$$

The first right eigen vectors of matrix $C_{YX}C_{XY}; \ell \times \ell$

The first (largest) singular value

The first left eigen vectors of matrix $C_{XY}C_{YX}; m \times m$

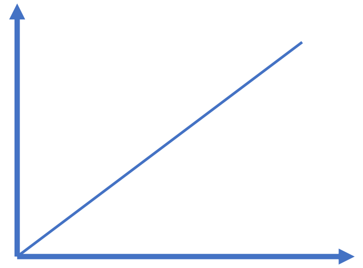
The square root of the largest eigenvalue of $C_{XY}C_{YX}$ or $C_{YX}C_{XY}$

$$s_1 \geq s_2 \geq \cdots$$

Step 2. Optimal Prediction with One Latent Variable

- In Step 1 we have found the component of the input space that is most correlated with the output;
- We now predict output y based on the score, $z = v^T x$

High-dimensional input vector $x \overset{}{\underset{z=v^T x}{\longmapsto}} z$ Input score: scalar



Parameters to determine

$$\hat{y} = \boxed{q}z$$

$$\hat{y} \in \Re^{\ell\times 1}$$

$$q \in \Re^{\ell\times 1}$$

Optimal Prediction

$$\hat{y} = q^o z$$

$$q^o = \arg\min_q E[|y - \hat{y}|^2]$$

It is conceivable that this optimal coefficient/parameter vector q^0 is in the same direction as unit vector w.
(This is a problem involved in the next assignment.)

Obtaining optimal q^0

$$\hat{y} = qz = qv^T x$$

$$
\begin{aligned}
E[(y-\hat{y})^T(y-\hat{y})] &= E[(y-qv^T x)^T (y-qv^T x)] \\
&= E[y^T y - 2y^T qv^T x + x^T vq^T qv^T x] \\
&= E[y^T y - 2v^T xy^T q + v^T xx^T vq^T q] \\
&= E[y^T y] - 2v^T E[xy^T]q + v^T E[xx^T]vq^T q
\end{aligned}
$$

Note that x and y are random variables, v and q are not.

The necessary conditions for optimality

$$\frac{d}{dq} = 0 \qquad -2\underbrace{E[yx^T]}_{C_{YX};\ell\times m}v + 2qv^T\underbrace{E[xx^T]}_{C_{XX};m\times m}v = 0$$

$$\therefore \quad q^o = \frac{C_{YX}v}{v^T C_{XX} v}$$

This is called ***Output Loading Vector***.

(We omit superscript o hereafter.)

Step 3. Deflation

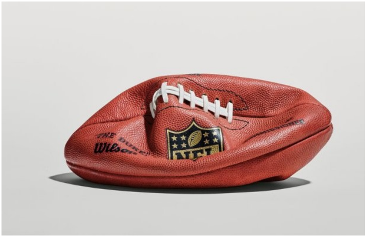
- We have found just one set of latent variables associated with the highest correlation between input and output.
- But, an accurate prediction cannot be obtained with just one set of latent variables. Now we want to find the components of the second and the third highest correlation.
- The singular Value decomposition of the cross-correlation matrix, however, does not directly give the second and the third most significant latent variables.

$$C_{XY} = \begin{bmatrix} v_1 & v_2 & \cdots \end{bmatrix} \begin{bmatrix} s_1 & \cdots & 0 \\ \vdots & s_2 & 0 \\ 0 & 0 & \ddots \end{bmatrix} \begin{bmatrix} w_1^T \\ w_2^T \\ \vdots \end{bmatrix}$$

These do not provide the latent variables that are second most significant (correlated).

- To overcome this difficulty, we have to go through the procedure called “Deflation’.

Deflation



- ❑ Predicting output y based on the first set of latent variables, we have used some components involved in the data matrix;
- ❑ This component of data matrix must not be used for determining the second most significant component;
- ❑ We have to remove the components already used in the first round prediction, and examine the residual components that have the highest correlation with output.
- ❑ Output data

Original data

$$y' = y - \text{(the component used for the first round output prediction)}$$

Residue

- ❑ Using the output loading vector $\boldsymbol{q}$, the deflated output vector is given by

$$y' = y - zq$$

- ❑ The deflation of input data matrix is a bit more involved. Collectively, we can write

$$X' = X - (\text{components used in the 1st round})$$

Input Data Deflation: Finding the components used in the 1st round

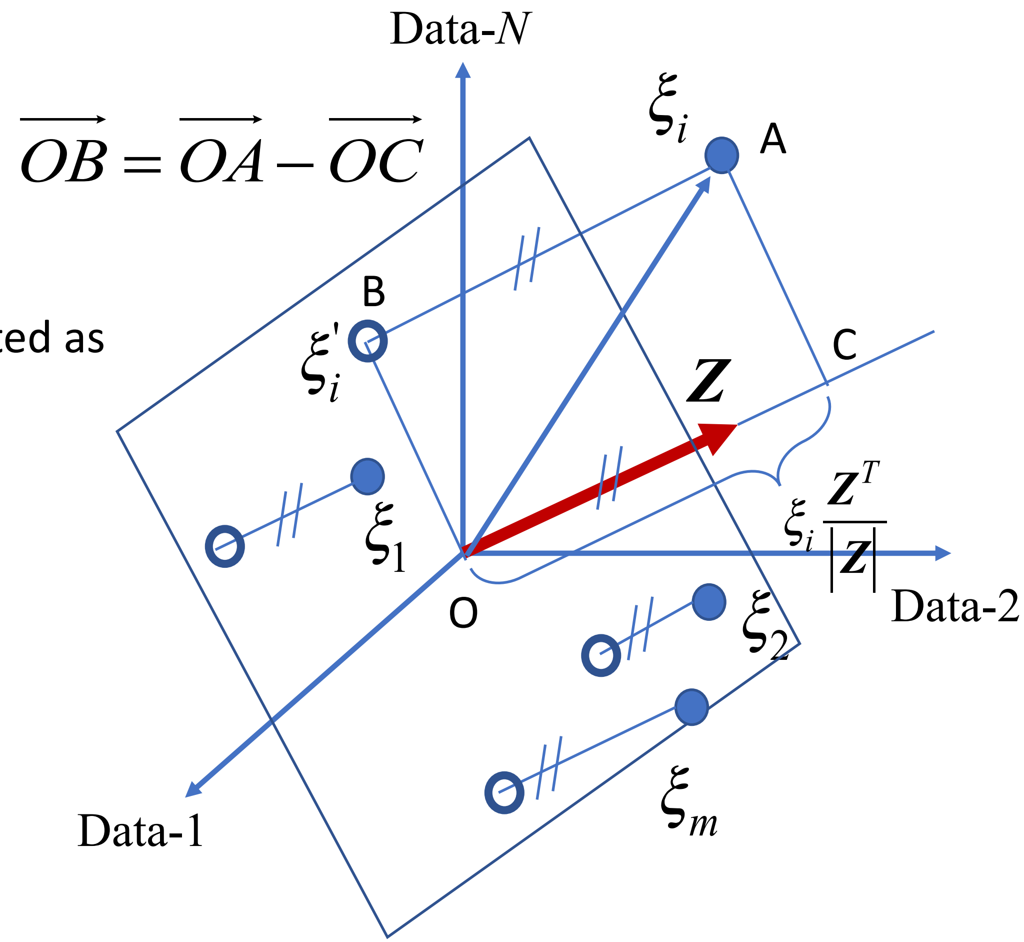
- The input data matrix can be written as a collection of row vectors,

$$X = \begin{pmatrix} \boldsymbol{x}^{(1)} & \cdots & \boldsymbol{x}^{(N)} \end{pmatrix} = \begin{pmatrix} x_1^{(1)} & \cdots & x_1^{(N)} \\ \vdots & \vdots & \vdots \\ x_m^{(1)} & \cdots & x_m^{(N)} \end{pmatrix} = \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_m \end{pmatrix}$$

- The score of each data point from $\boldsymbol{x}^{(1)}$ to $\boldsymbol{x}^{(N)}$ can be collectively represented as

$$z = v^T x \longrightarrow \begin{pmatrix} z^{(1)} & \cdots & z^{(N)} \end{pmatrix} = \boldsymbol{Z}$$

- Plot $\boldsymbol{Z}$ in N-dimensional space together with $\xi_1, \cdots, \xi_m$
- The direction of vector $\boldsymbol{Z}$ indicates the distribution of scores among the N data points at which the first round latent variables have extracted information from the original data for predicting the output.
- We need to delete these components already used in the first round.
- Projecting ξ_i onto the plane perpendicular to vector $\boldsymbol{Z}$ yields



$$\xi_i' = \xi_i - \underbrace{\xi_i \frac{\boldsymbol{Z}^T}{|\boldsymbol{Z}|}}_{\text{Magnitude}} \cdot \underbrace{\frac{\boldsymbol{Z}}{|\boldsymbol{Z}|}}_{\text{Direction}} = \xi_i \left(I - \frac{\boldsymbol{Z}^T \boldsymbol{Z}}{|\boldsymbol{Z}|^2} \right)$$

Or collectively, $X' = X \left(I - \dfrac{\boldsymbol{Z}^T \boldsymbol{Z}}{|\boldsymbol{Z}|^2} \right)$

Note that vectors $\boldsymbol{Z}$ and ξ_i are row vectors.

Input Loading Vector

- Let us rewrite $$X' = X\left(I - \frac{\boldsymbol{Z}^T \boldsymbol{Z}}{\left|\boldsymbol{Z}\right|^2} \right)$$
- Note that $\boldsymbol{Z} = \begin{pmatrix} z^{(1)} & \cdots & z^{(N)} \end{pmatrix} = \begin{pmatrix} v^T \boldsymbol{x}^{(1)} & \cdots & v^T \boldsymbol{x}^{(N)} \end{pmatrix} = v^T X$
- Therefore $\left|\boldsymbol{Z}\right|^2 = \boldsymbol{Z}\boldsymbol{Z}^T = v^T X X^T v \cong v^T C_{XX} v$
- The above deflated data matrix can be written as

$$X' = X\left(I - \frac{\boldsymbol{Z}^T \boldsymbol{Z}}{\left|\boldsymbol{Z}\right|^2} \right) = X - \frac{1}{v^T C_{XX} v} X\boldsymbol{Z}^T \boldsymbol{Z} = X - \frac{1}{v^T C_{XX} v} X X^T v v^T X$$

$$= X - \frac{1}{v^T C_{XX} v} C_{XX} v v^T X = \left(I - \frac{C_{XX} v v^T}{v^T C_{XX} v} \right) X$$

- For each column vector $\boldsymbol{x}' = \left(I - \frac{C_{XX} v v^T}{v^T C_{XX} v} \right) \boldsymbol{x} = \boldsymbol{x} - \frac{C_{XX} v}{v^T C_{XX} v} v^T \boldsymbol{x} = \boldsymbol{x} - p \cdot z$

where $p = \frac{C_{XX} v}{v^T C_{XX} v}$ Is called ***Input Loading Vector***.

Properties of the Deflated Input Data Matrix and the Input Loading Vector

We can show that, with the input loading vector p, the deflated data matrix becomes the smallest.

$$p^o = \arg\min_p E[|x'|^2] \leftarrow x' = x - zp = x - pv^T x = (I - pv^T)x$$

$$= \arg\min_p E[x^T (I - pv^T)^T (I - pv^T) x]$$

$$= \arg\min_p \left\{ E[x^T x] - 2p^T E[xx^T]v + p^T p v^T E[xx^T]v \right\}$$

Necessary conditions for min.

$$\frac{d}{dp} = 0 \qquad 2C_{XX}v + 2pv^T C_{XX}v = 0 \qquad \therefore \quad p^o = \frac{C_{XX}v}{v^T C_{XX} v}$$

This is the same as the input loading vector. Therefore, the loading vector minimizes the deflated data matrix. In other words, the 1st round latent variables have taken the most information.

Question:
Why is p not aligned with v?

If C_{XX} is the identity matrix, v and p are aligned. However, the data are distributed not uniformly over the input space:

$$C_{XX} \neq \kappa I$$

Input Deflation $x' = (I - p^o v^T)x = \left(I - \dfrac{C_{XX} v v^T}{v^T C_{XX} v} \right) x$ Output Deflation : $y' = y - zq^o$

Deflated Covariance and Cross-Covariance Matrices C'_{XX} and C'_{YX}

To compute the second set of latent variables, we need

$$C'_{XX} = E[x'(x')^T] \qquad C'_{YX} = E[y'(x')^T]$$

for the deflated input and output data.

$$\begin{aligned}
C'_{XX} &= E[(I - pv^T)xx^T(I - vp^T)] \quad \leftarrow \quad x' = (I - pv^T)x \\
&= (I - pv^T)E[xx^T](I - vp^T) \\
&= C_{XX} - pv^T C_{XX} - C_{XX}vp^T + pv^T C_{XX} vp^T \\
&= C_{XX} - pv^T C_{XX} \quad \leftarrow \quad pv^T C_{XX} v = C_{XX} v,\ p = \frac{C_{XX} v}{v^T C_{XX} v}
\end{aligned}$$

$C'_{XX} = (I - pv^T)C_{XX}$ Similarly, $C'_{YX} = C_{YX}(I - vp^T)$

Partial Least Squares Regression : Summary

The most significant m^* sets of latent variables are obtained recursively,

$$C_{XX}[1] = E[xx^T], \quad C_{YX}[1] = E[yx^T]$$

For k = 1 to m*

$$C_{XX}[k+1] = (I - p[k]v^T[k])C_{XX}[k];$$

$$C_{YX}[k+1] = C_{YX}[k](I - v[k]p^T[k]);$$

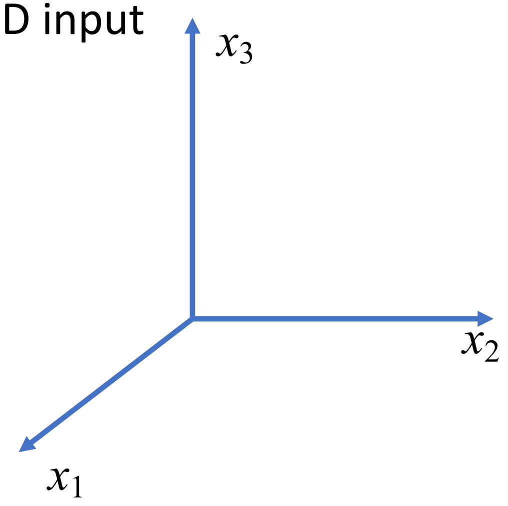
$$p[k] = \frac{C_{XX}[k]v[k]}{v[k]^T C_{XX}[k]v[k]} \qquad q[k] = \frac{C_{YX}[k]v[k]}{v[k]^T C_{XX}[k]v[k]}$$

With these loading vectors, x and y can be approximated to

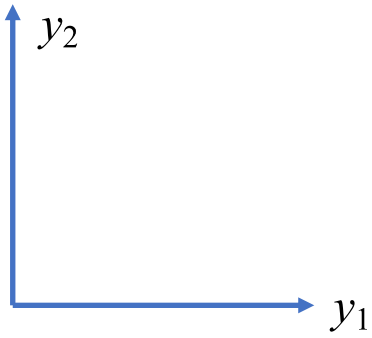
$$x = \sum_{k=1}^{m^*} z[k]p[k] + x[m^*] \qquad y = \sum_{k=1}^{m^*} z[k]q[k] + y[m^*]$$

Example

3D input

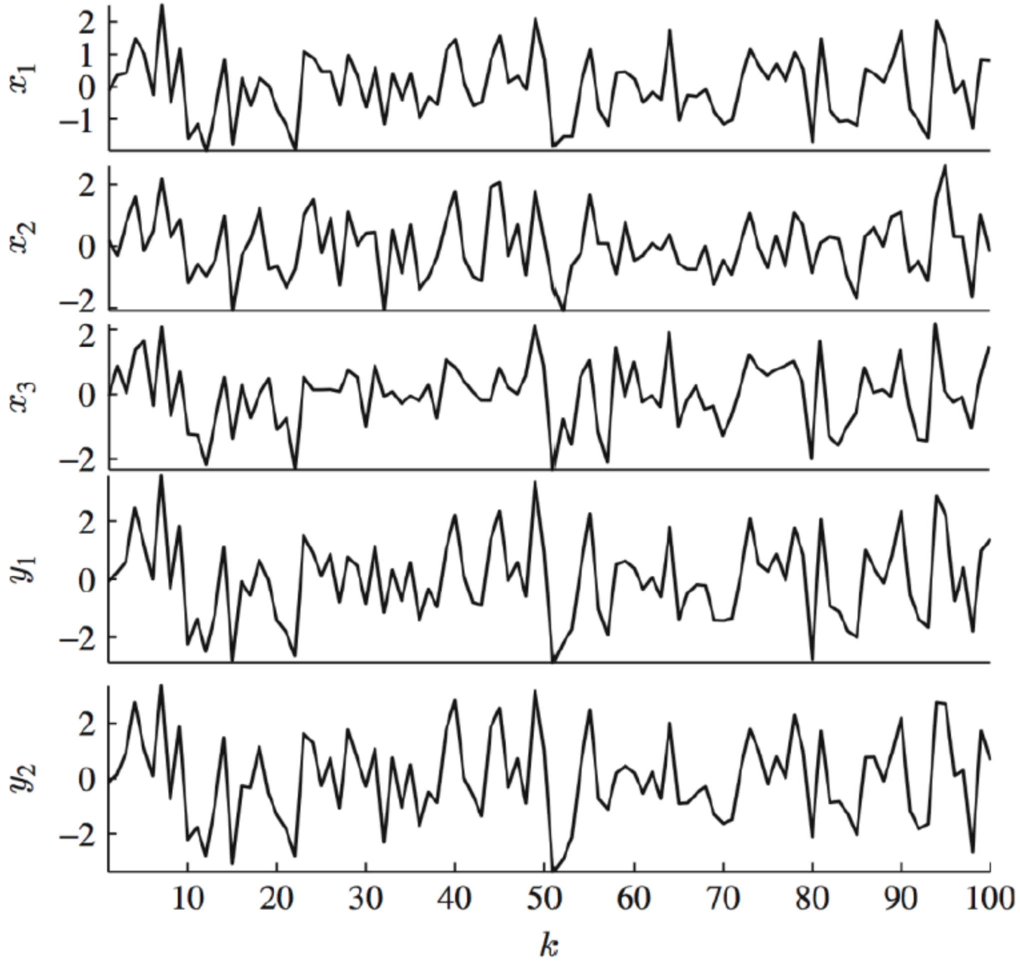


2D output



Sample Data (sample size = 100)

Data are generated with the linear model + Gaussian noise $g \sim N(0, 0.05I)$



True relationship

$$y = Bx + g$$

$$B = \begin{pmatrix} 0.341 & 0.534 & 0.727 \\ 0.309 & 0.836 & 0.568 \end{pmatrix}$$

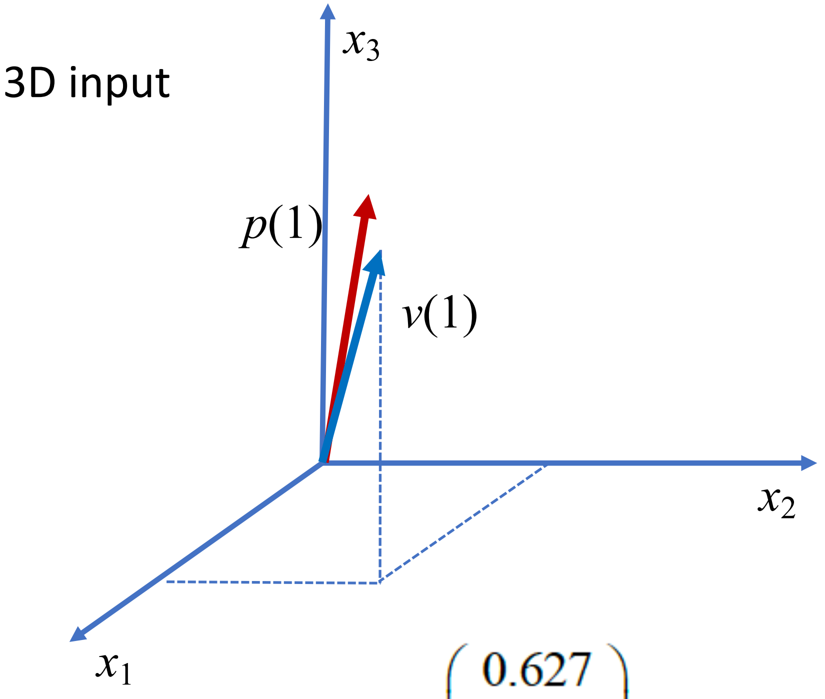
Input Covariance

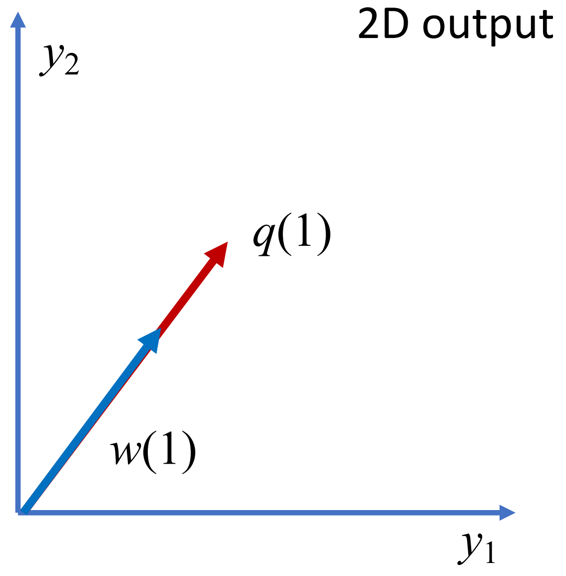
$$C_{XX} = \begin{pmatrix} 1 & 0.8 & 0.9 \\ 0.8 & 1 & 0.5 \\ 0.9 & 0.5 & 1 \end{pmatrix}$$

Input-Output Cross-Covariance

$$C_{YX} = \begin{pmatrix} 1.42 & 1.17 & 1.30 \\ 1.49 & 1.37 & 1.27 \end{pmatrix}$$

First Round Latent Variables



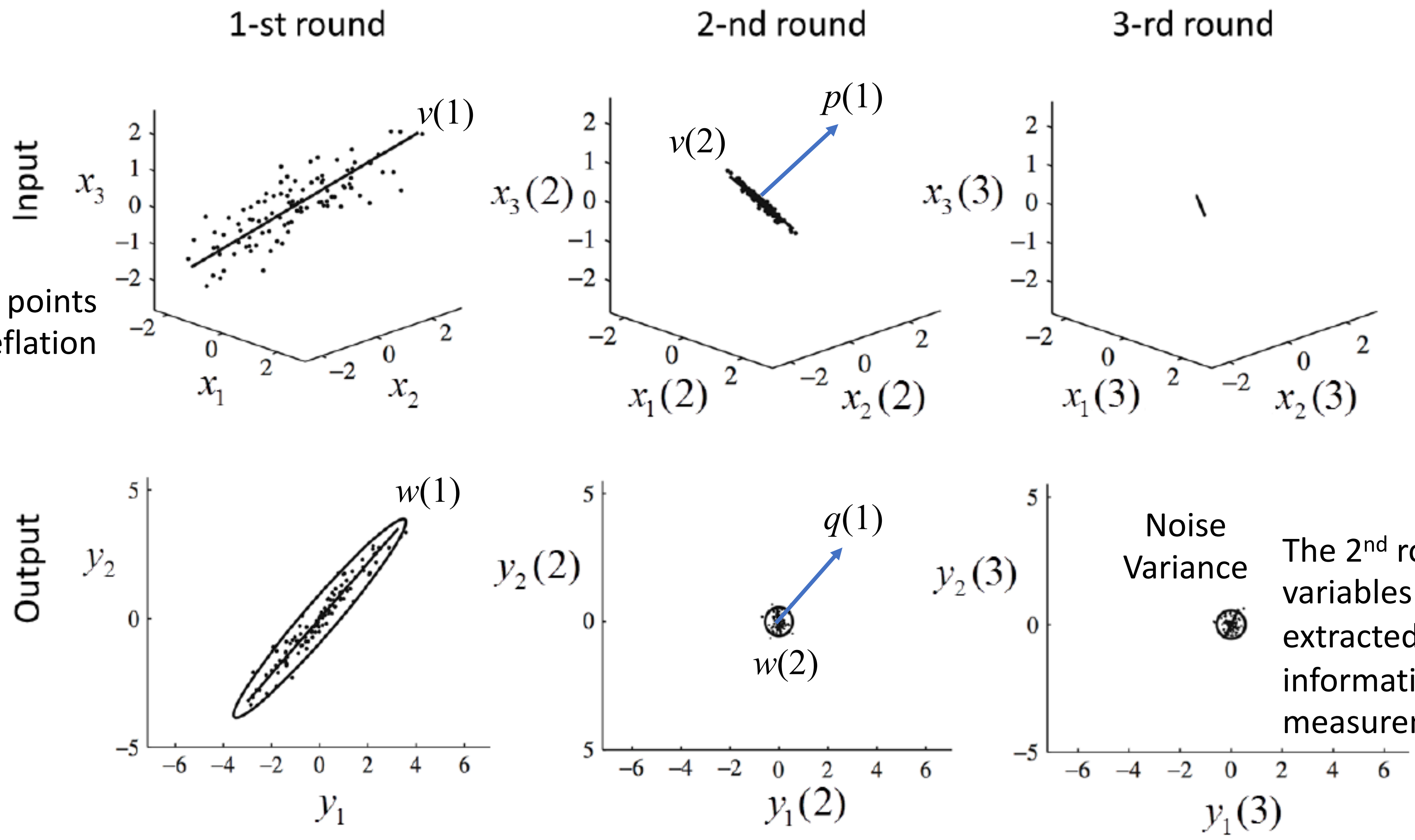


$$v(1) = \begin{pmatrix} 0.627 \\ 0.5515 \\ 0.550 \end{pmatrix}, \quad w(1) = \begin{pmatrix} 0.679 \\ 0.734 \end{pmatrix}, \quad p(1) = \begin{pmatrix} 0.631 \\ 0.536 \\ 0.561 \end{pmatrix}, \quad q(1) = \begin{pmatrix} 0.915 \\ 0.989 \end{pmatrix}$$

Singular Value Decomposition of the cross-correlation matrix

$$C_{XY} = \begin{bmatrix} v_1 & \cdots\cdots \end{bmatrix} \begin{bmatrix} s_1 & \cdots & 0 \\ \vdots & * & 0 \\ 0 & 0 & \end{bmatrix} \begin{bmatrix} w_1^T \\ \vdots \\ \vdots \end{bmatrix} \qquad p = \frac{C_{XX} v}{v^T C_{XX} v} \qquad q = \frac{C_{YX} v}{v^T C_{XX} v}$$

Partial Least Squares Regression



The 2nd round latent variables completely extracted all meaningful information, leaving only measurement noise.

The original data are deflated in the direction of $p(1)$ for input and that of $q(1)$ for output.

Applications of Partial Least Squares Regression (PLSR)

- ❑ Chemistry
- ❑ Biology
- ❑ Biomechanics
- ❑ Robotics
- ❑ Social Science

Applications of Partial Least Squares Regression (PLSR)

- Chemistry
- Biology
- Biomechanics
- Robotics
- Social Science

Leader-Follower Approach

Two-Person Demonstration

The robot arms are back-drivable.

Robot = Follower

Human = Leader

Observe

two-human task execution

Transfer

The identified laws to the robot

Extract

Dynamic Coordination

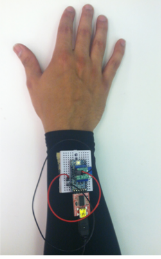
Control Laws

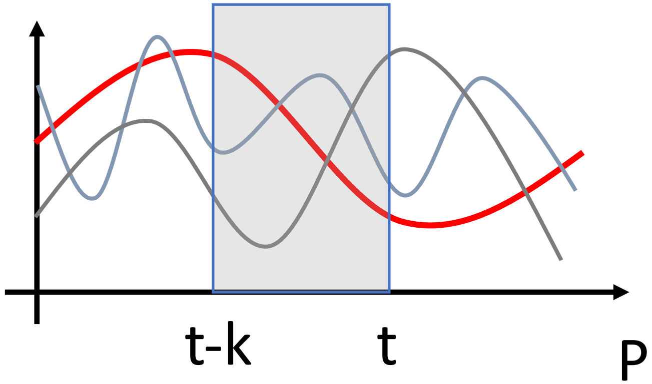
Extracting Coordinated Control Laws from teaching data by using

Partial Least Squares Regression

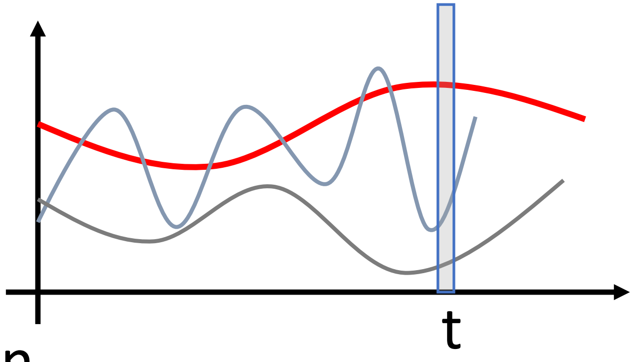
Human 1 hand trajectories

Human 2 (Robot) hand trajectories





Prediction



Input

$\boldsymbol{x}$ = (3 axes of acceleration and
3 axes of angular velocity
of both hands at time ***t***;
------------------- at time ***t-1***;

------------------- at time ***t-k***)

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}^1 \\ \vdots \\ \mathbf{x}^N \end{pmatrix} = \boxed{100 \times 360}$$

High-dimensional input space

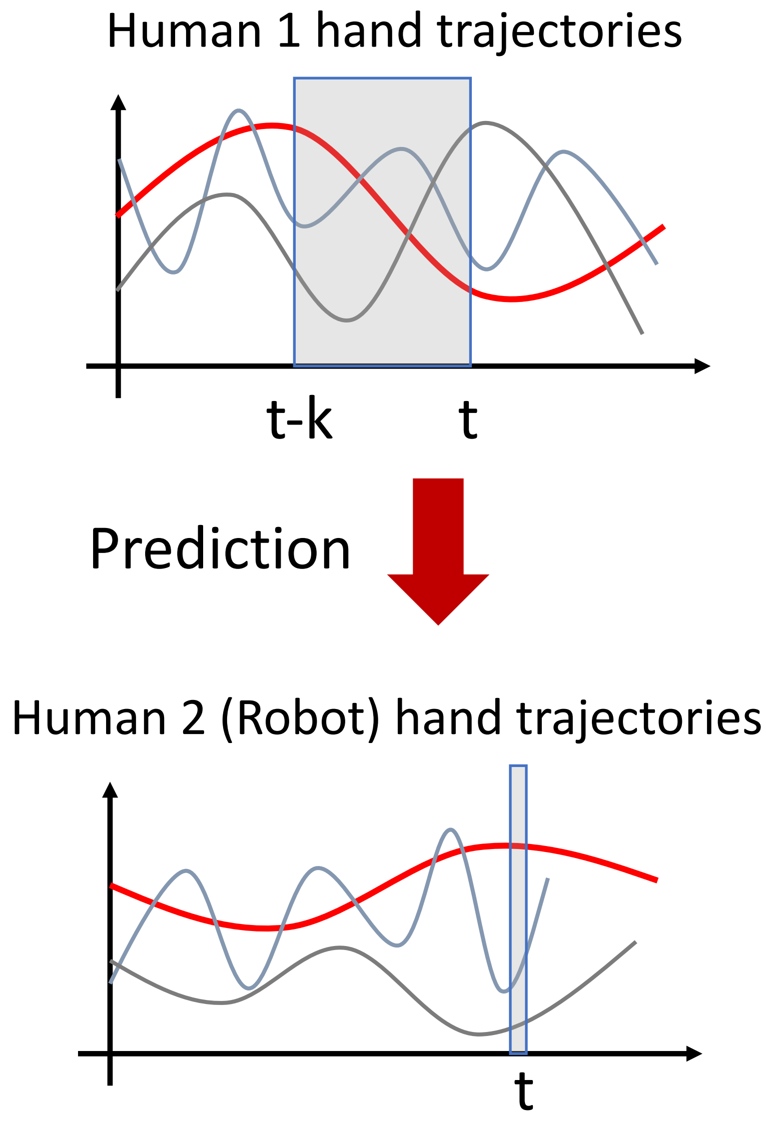
Output

$\boldsymbol{y}$ = (4 joint angles of right robot arm
4 joint angles of left robot arm at time ***t***)

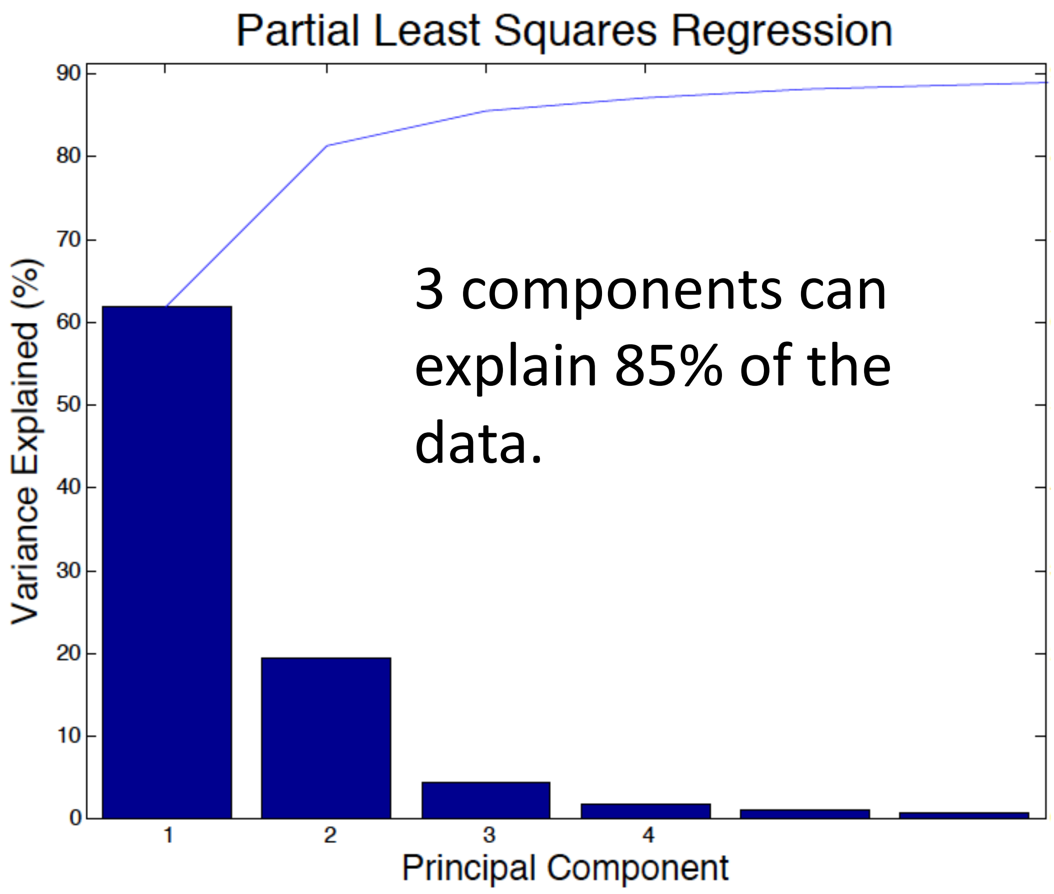
Output joint displacements may be collinear and correlated to each other.

$$\mathbf{Y} = \begin{pmatrix} \mathbf{y}^1 \\ \vdots \\ \mathbf{y}^N \end{pmatrix} = \boxed{100 \times 8}$$

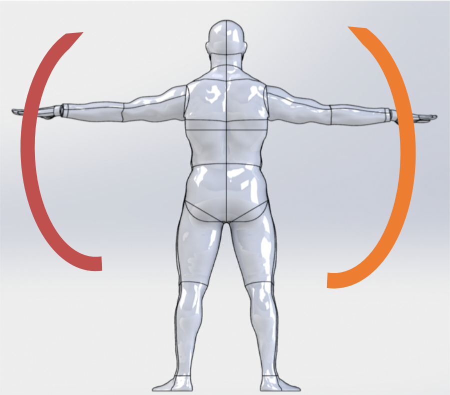
Four Arm Coordination



PLS Analysis

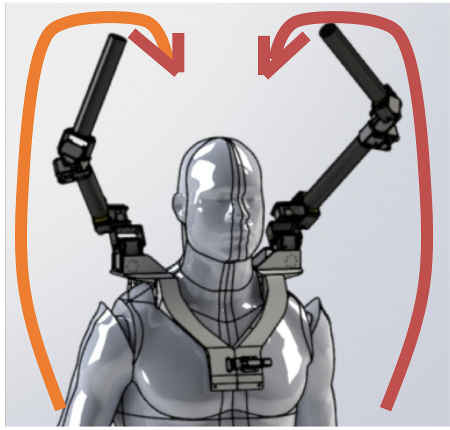


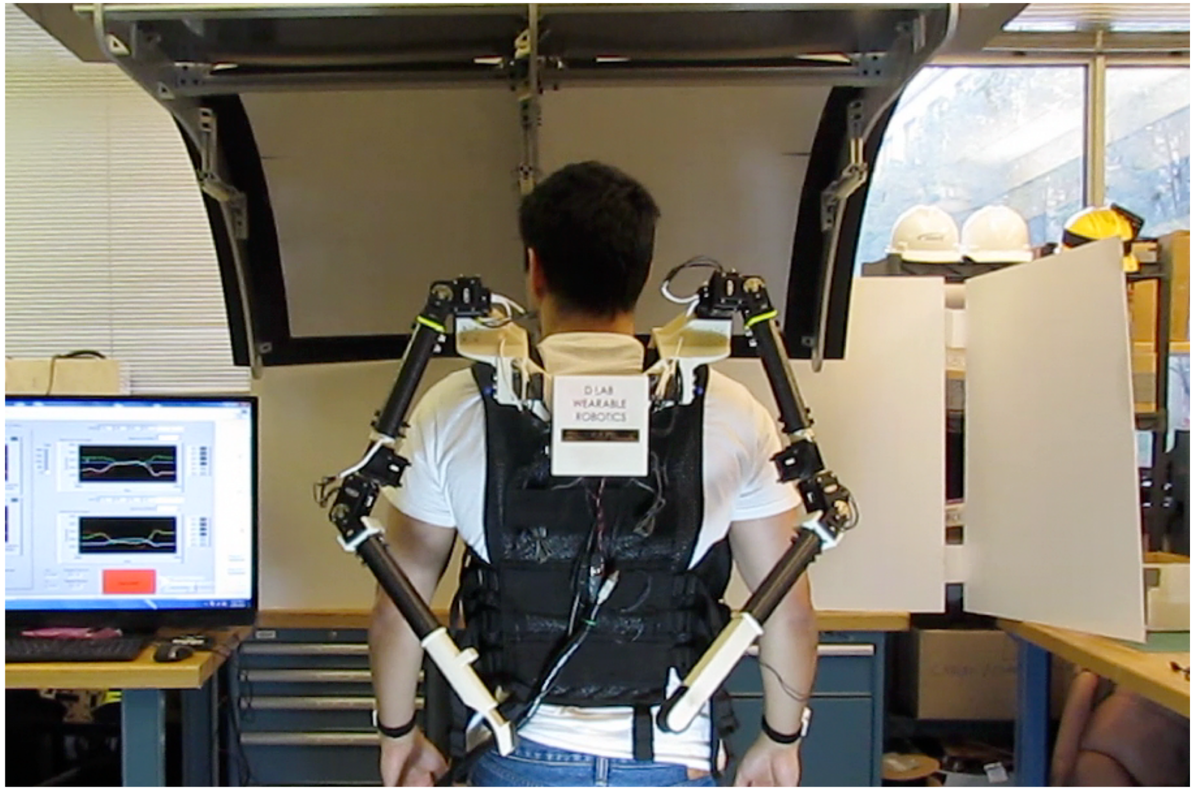
Partial Least Squares Regression: Mode 1



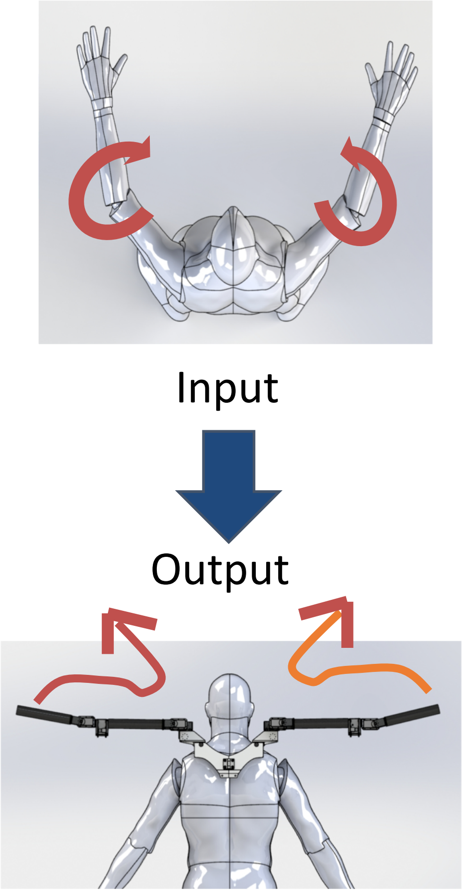
Input

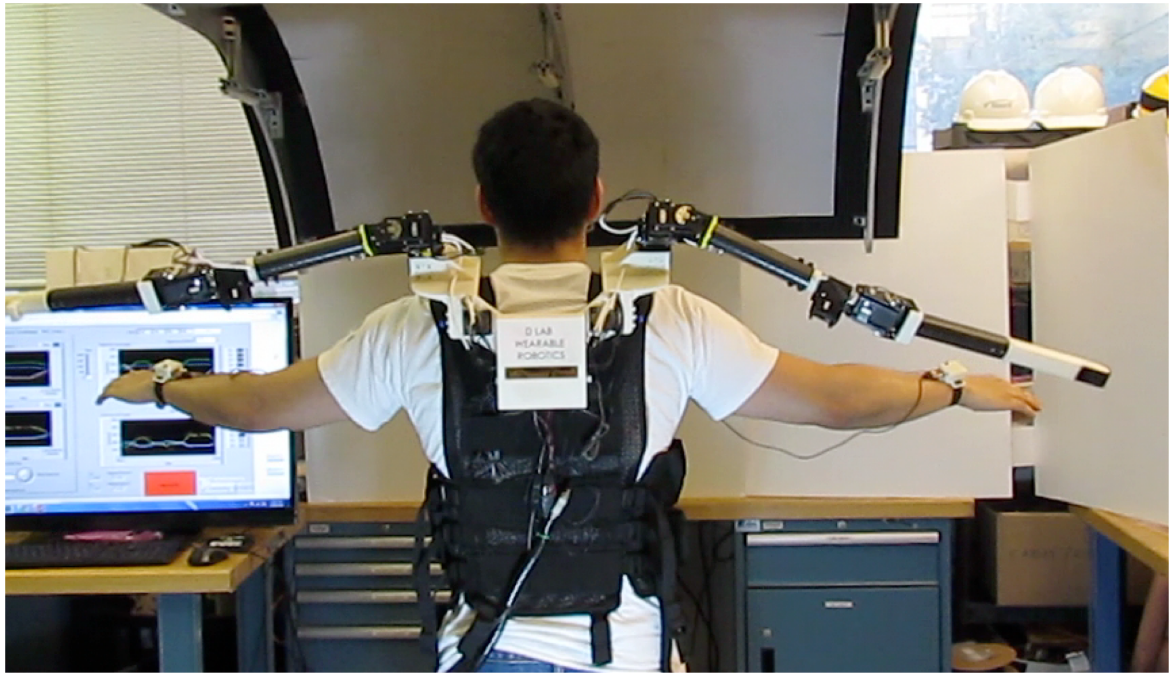
Output

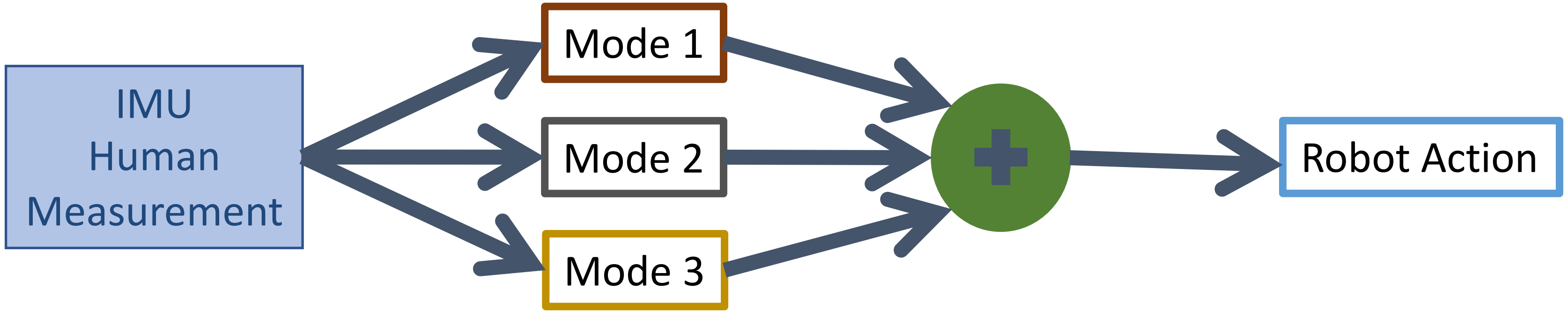


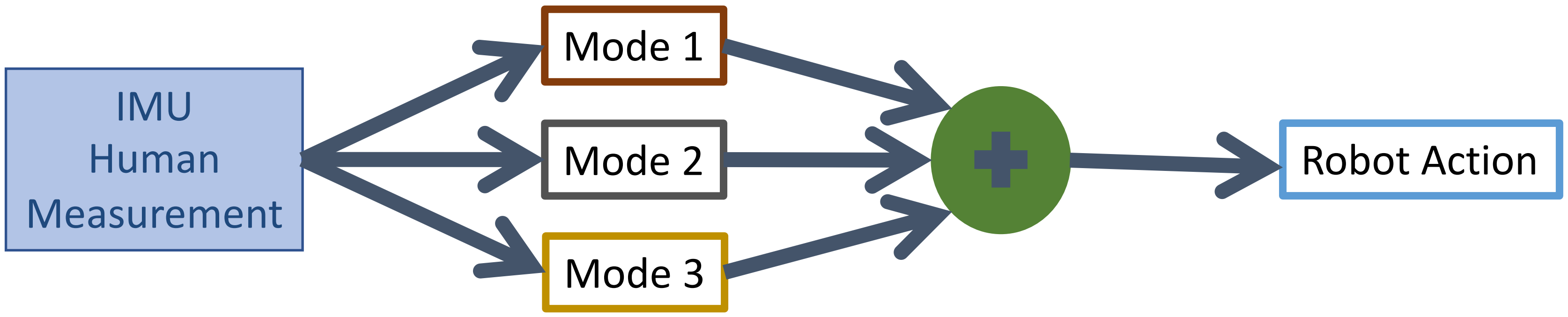


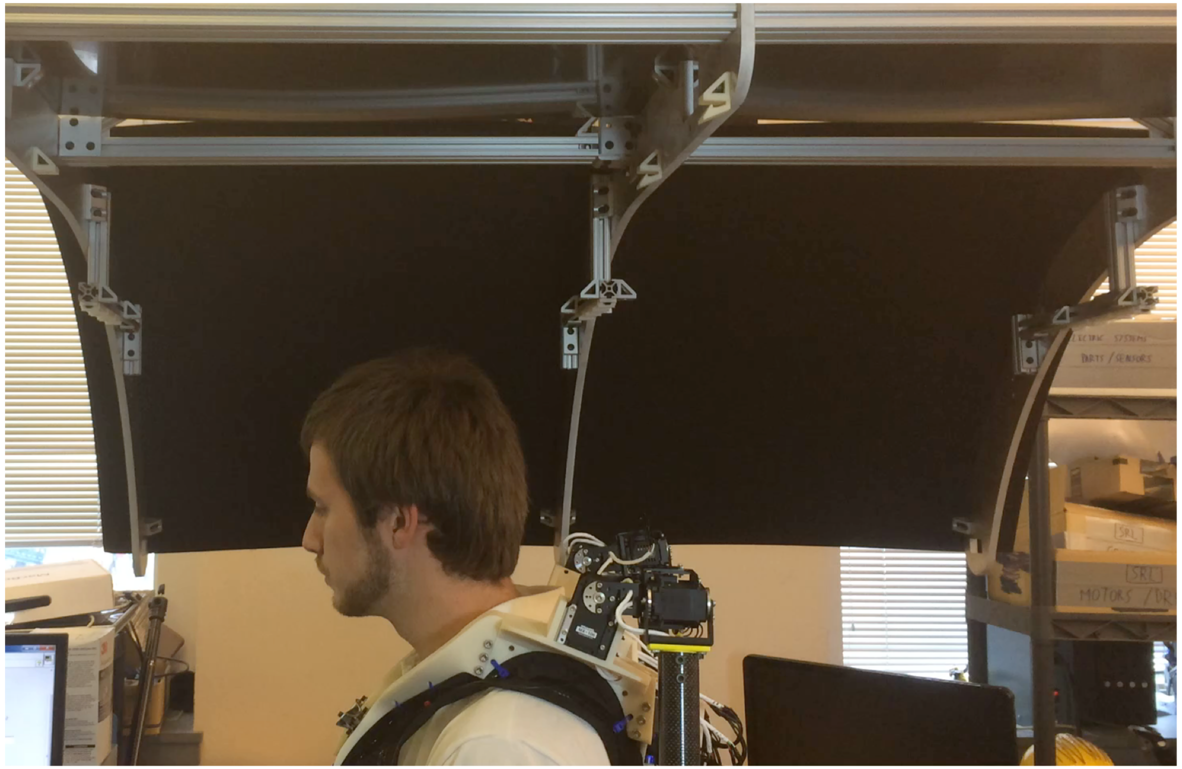
Partial Least Squares Regression: Mode 3






IMU
Human
Measurement
Mode 1
Mode 2
Mode 3
Robot Action


IMU
Human
Measurement
Mode 1
Mode 2
Mode 3
Robot Action



Concluding Discussion

- Dealing with high-dimensional input and output spaces is an important challenge with many practical applications of today's interest.
- PCR and PLSR are linear predictors, but as we use more input and output variables, some nonlinearities can be well captured with the high-dimensional linear regression.
- We will revisit high-dimensional spaces in Part 4. Specifically, it is closely related to extended feature space, kernel trick, and lifting linearization based on Koopman Operator and Dual-Faceted Linearization of nonlinear dynamical systems.