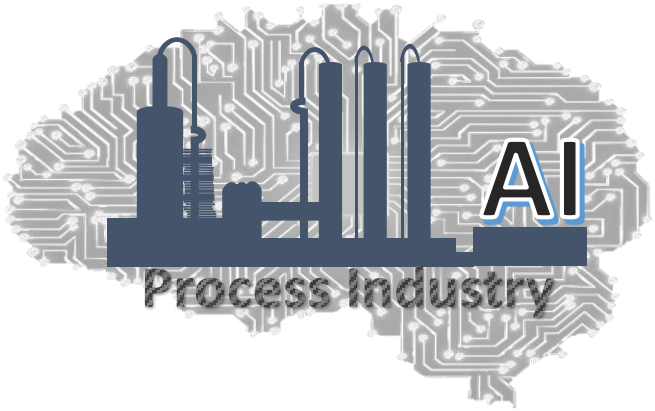
Statistical Techniques for Monitoring Industrial Processes



Lecture : Introduction to PLS

Module : PLS-based MSPM

Course TOC

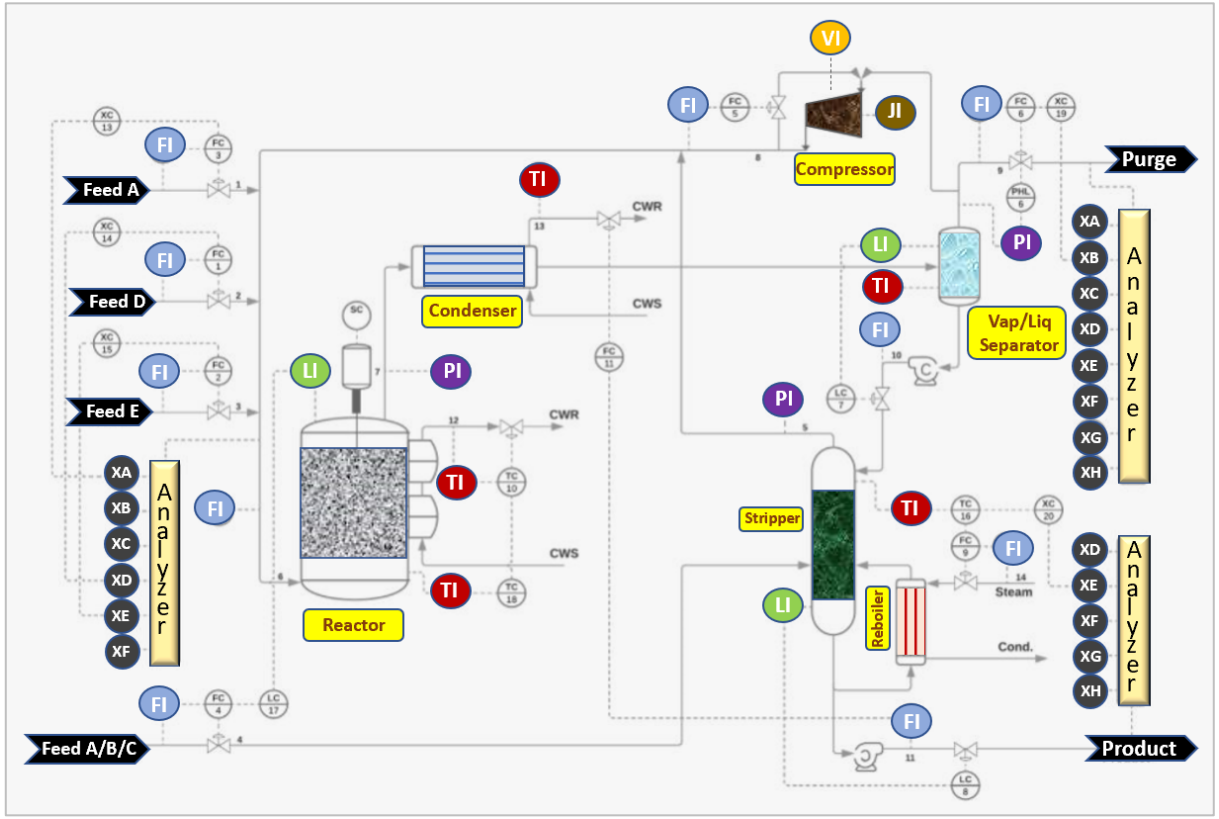
- ❑ Introduction to Statistical Process Monitoring (SPM)
- ❑ Python Installation and basics (optional)
- ❑ Univariate SPM & Control Charts
  - ➢ Shewhart Charts
  - ➢ CUSUM Charts
  - ➢ EWMA Charts
- ❑ Multivariate SPM
  - ➢ Principal Component Analysis (PCA)-based MSPM
  - ➢ Partial Least Squares (PLS) regression-based MSPM
    - ➢ Fault detection & diagnosis (FDD) using PLS
    - ➢ Application to a LDPE reactor monitoring
  - ➢ Strategies for handling nonlinear, dynamic, multimode systems
- ❑ Deploying SPM solutions

PLS (Partial Least Squares) : *X* and *Y* Blocks of Variables

Input or *X* variables

- Incoming feed flow
- Intermediate temperatures
- Intermediate pressures
- Intermediate flows

⋮



*Typical Chemical Plant (Tennessee Eastman Process)**

Output or *Y* variables

- Analyzer measurements
- Process efficiency
- Product flows

⋮

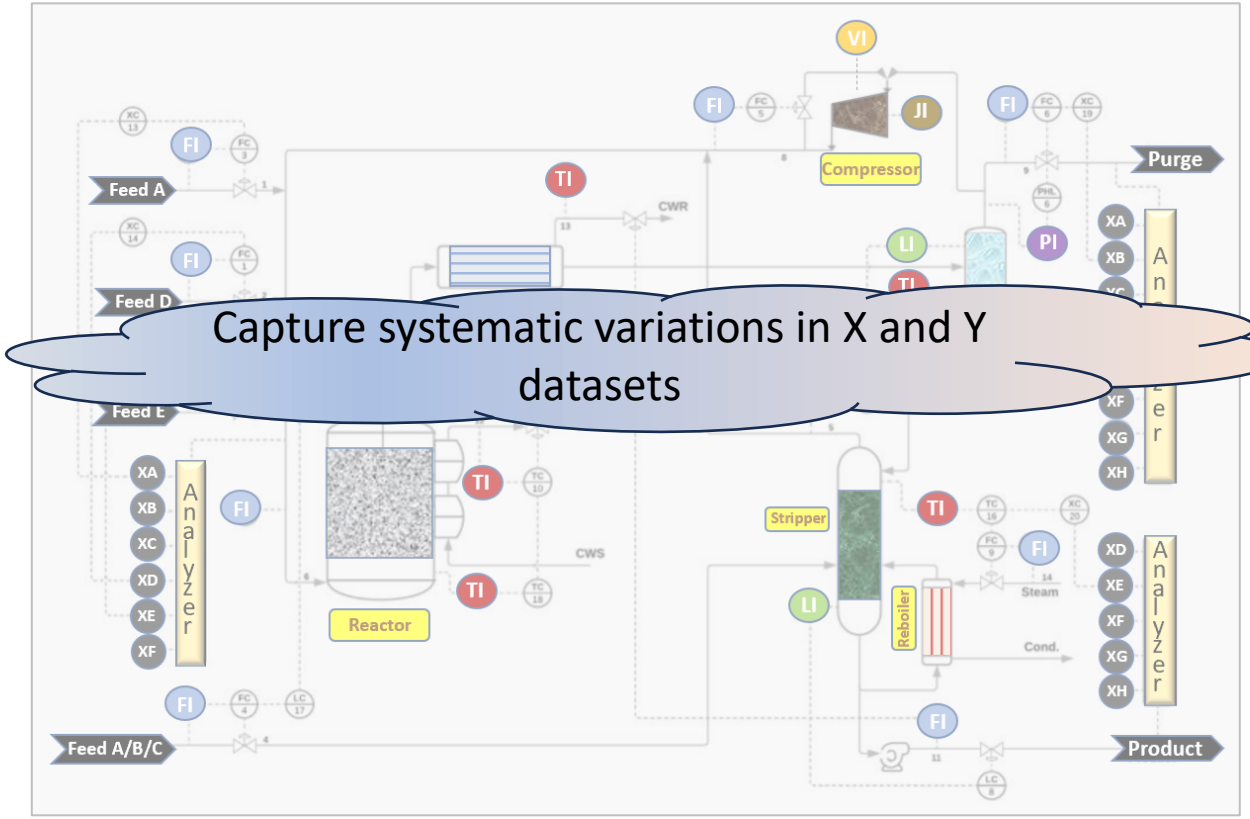
* Adapted from the original flowsheet by Gilberto Xavier (https://github.com/gmxavier/TEP-meets-LSTM) provided under Creative-Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/).

PLS (Partial Least Squares) : *X* and *Y* Blocks of Variables

Input or *X* variables

- Incoming feed flow
- Intermediate temperatures
- Intermediate pressures
- Intermediate flows

Quantify noise in input variables



Capture systematic variations in X and Y datasets

*Typical Chemical Plant (Tennessee Eastman Process)**

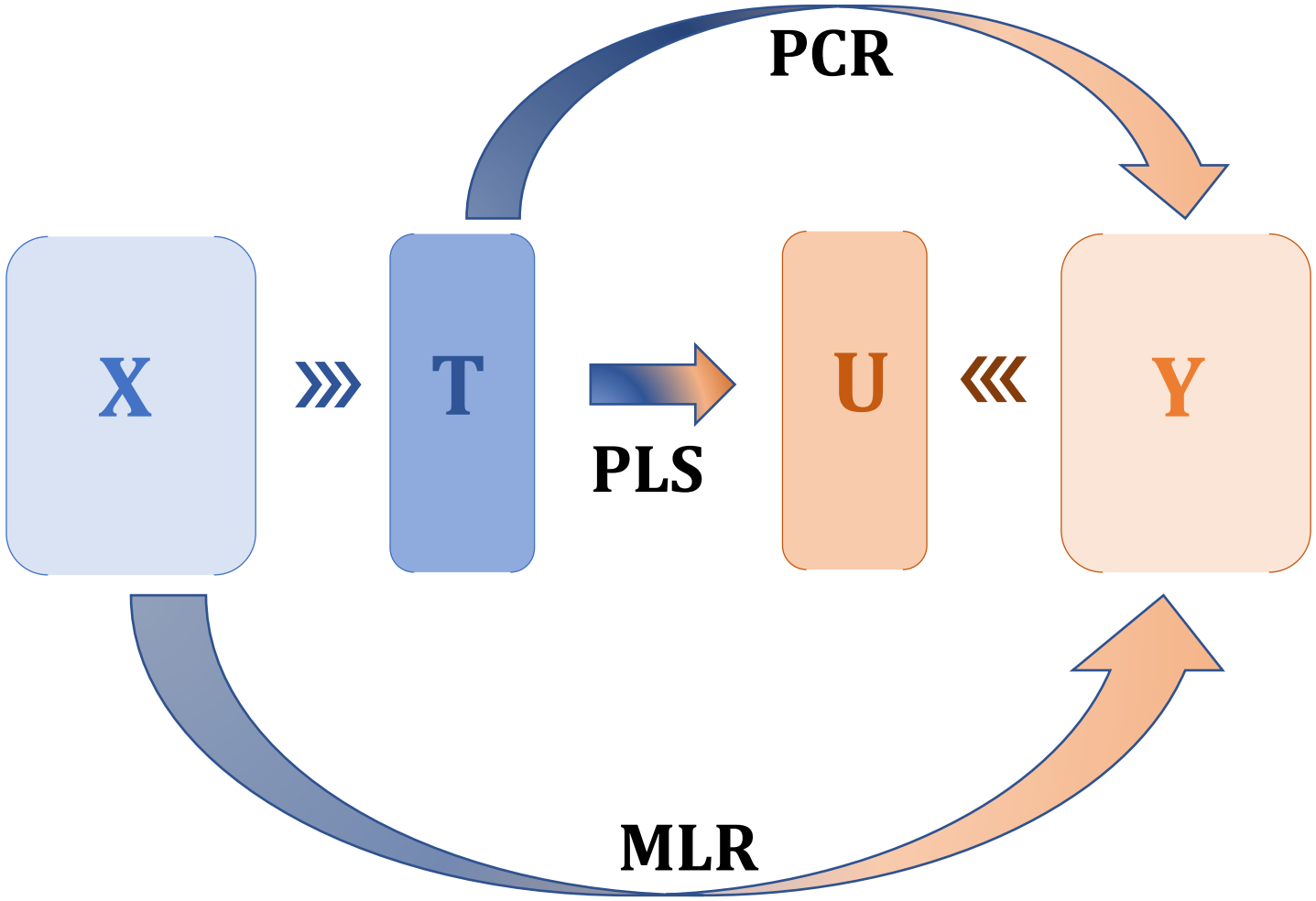
Output or *Y* variables

- Analyzer measurements
- Process efficiency
- Product flows

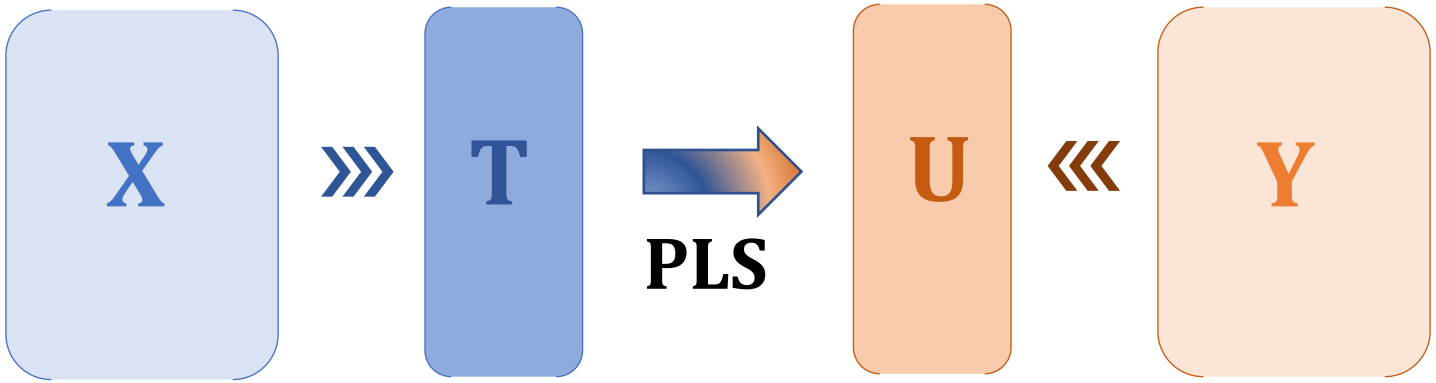
Quantify noise in output variables

* Adapted from the original flowsheet by Gilberto Xavier (https://github.com/gmxavier/TEP-meets-LSTM) provided under Creative-Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/).

PLS: Data Transformation



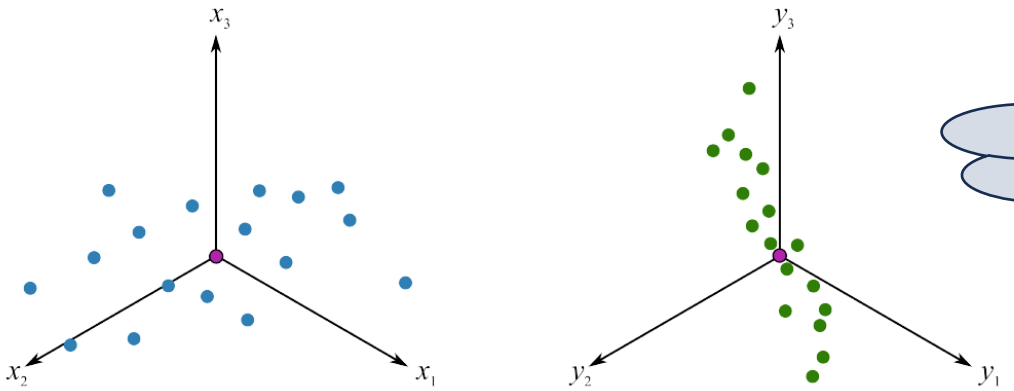
PLS: Data Transformation



PLS finds $\boldsymbol{T}$ and $\boldsymbol{U}$ matrices such that:

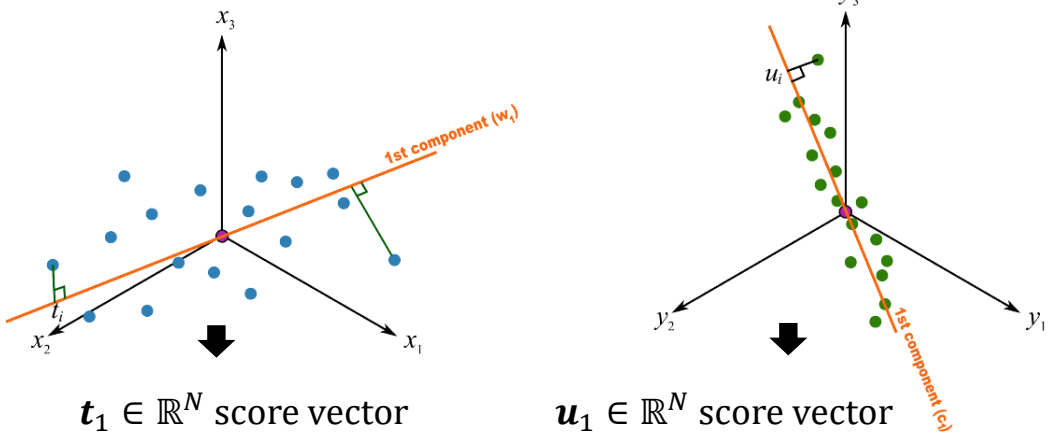
- explain as much variance as possible in ***X*** dataset
- explain as much variance as possible in ***Y*** dataset
- relationship between ***X*** and ***Y*** datasets is maximally captured

PLS: Geometric Interpretation*



Here, # of input variables (m) = # of output variables (p) just for convenience. Usually, p < m

First component



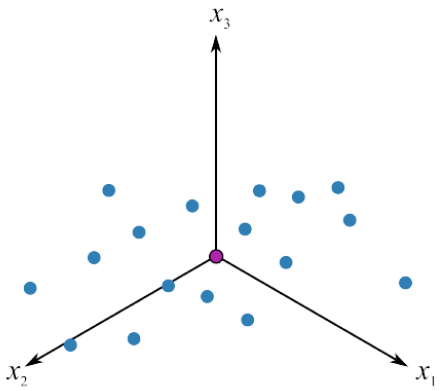
$\boldsymbol{t}_1 \in \mathbb{R}^N$ score vector

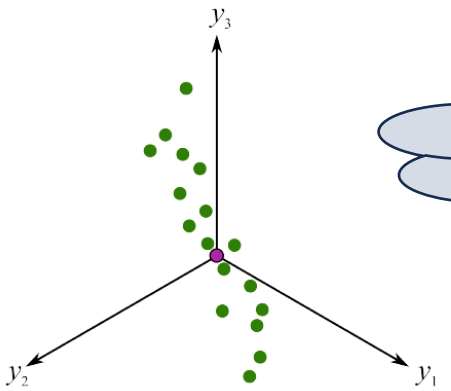
$\boldsymbol{u}_1 \in \mathbb{R}^N$ score vector

PLS finds $\boldsymbol{w}_1$ and $\boldsymbol{c}_1$ such that Cov($\boldsymbol{t}_1$, $\boldsymbol{u}_1$) is maximal!

* Visualizations are copyrighted work of Kevin Dunn (https://learnche.org/pid/) and shared under CC BY-SA 4.0 license

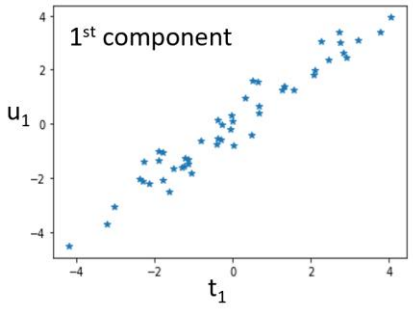
PLS: Geometric Interpretation*

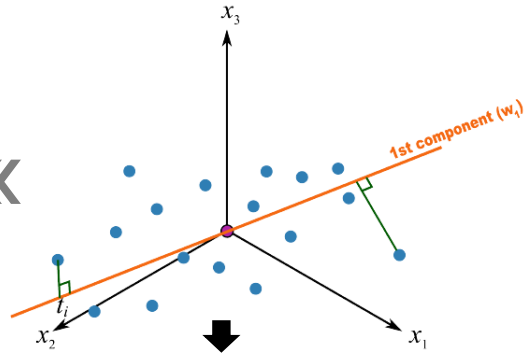




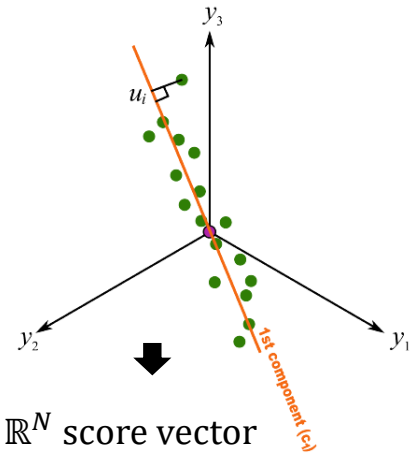
Here, # of input variables (*m*) = # of output variables (*p*) just for convenience. Usually, p < m

First component





$\boldsymbol{t_1} \in \mathbb{R}^N$ score vector



$\boldsymbol{u_1} \in \mathbb{R}^N$ score vector

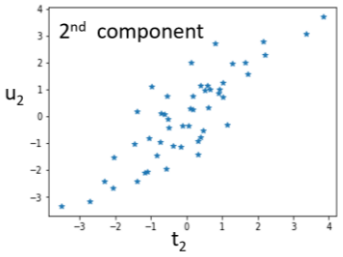
PLS finds $\boldsymbol{w_1}$ and $\boldsymbol{c_1}$ such that Cov($\boldsymbol{t_1}$, $\boldsymbol{u_1}$) is maximal!

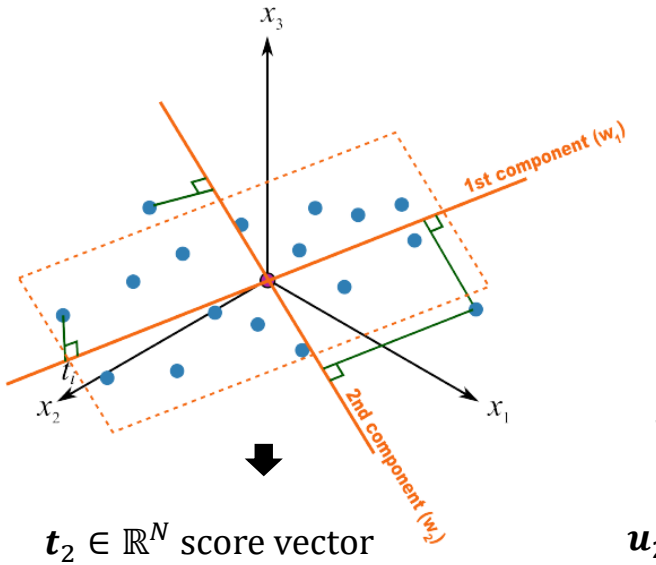
$$Cov(\boldsymbol{t_1}, \boldsymbol{u_1}) = Correlation(\boldsymbol{t_1}, \boldsymbol{u_1}) * \sqrt{Var(\boldsymbol{t_1})} * \sqrt{Var(\boldsymbol{u_1})}$$

* Visualizations are copyrighted work of Kevin Dunn (https://learnche.org/pid/) and shared under CC BY-SA 4.0 license

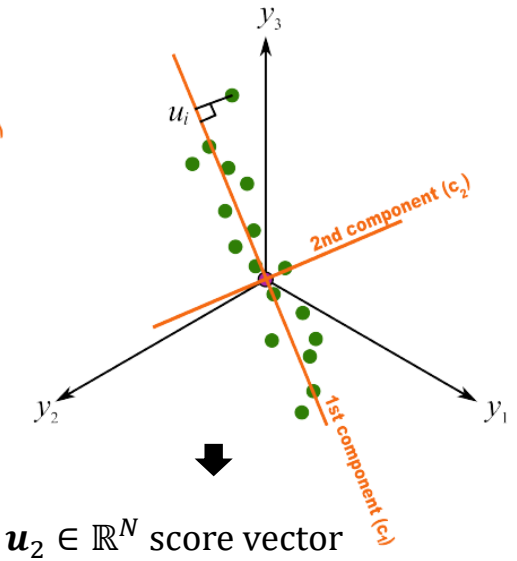
PLS: Geometric Interpretation*

Second component





$t_2 \in \mathbb{R}^N$ score vector



$u_2 \in \mathbb{R}^N$ score vector

PLS finds w_2 and c_2 such that Cov(t_2, u_2) is maximal!

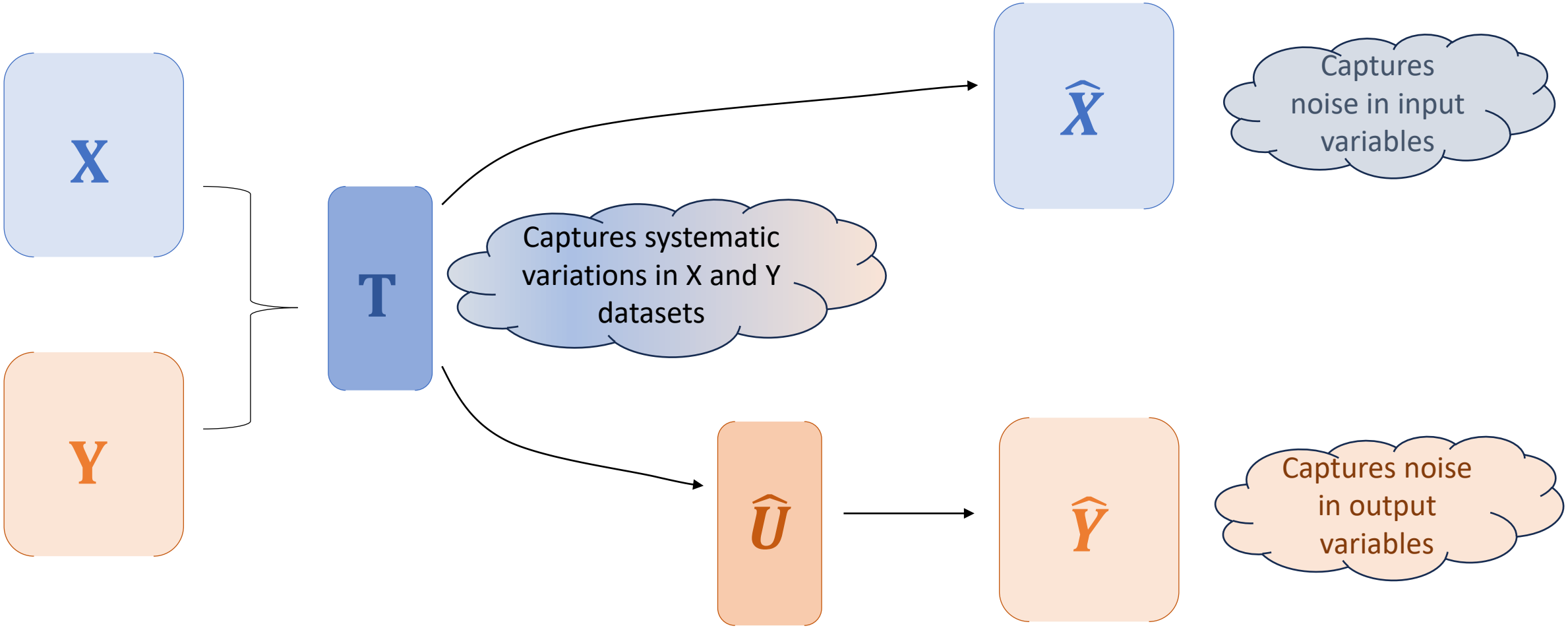
- w_2 is orthogonal to w_1

Note: t_2 (or u_2) is not direct projection of X (or Y) onto w_2 (or c_2)

Maximum number of components extractable equals the number of input variables (*m*)

* Visualizations are copyrighted work of Kevin Dunn (https://learnche.org/pid/) and shared under CC BY-SA 4.0 license

PLS: Data Transformation



Statistical Techniques for Monitoring Industrial Processes

Next Lecture : PLS – Under the Hood

Module : PLS-based MSPM