Statistical Techniques for Monitoring Industrial Processes

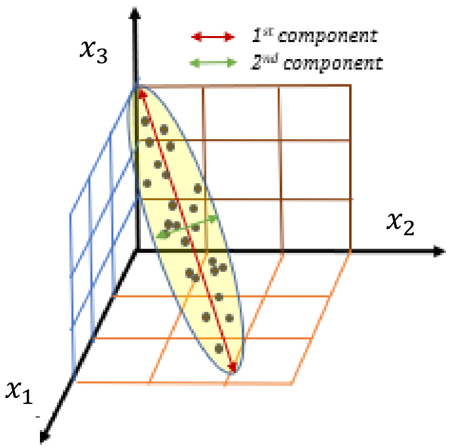
Lecture : PCA – Under the Hood

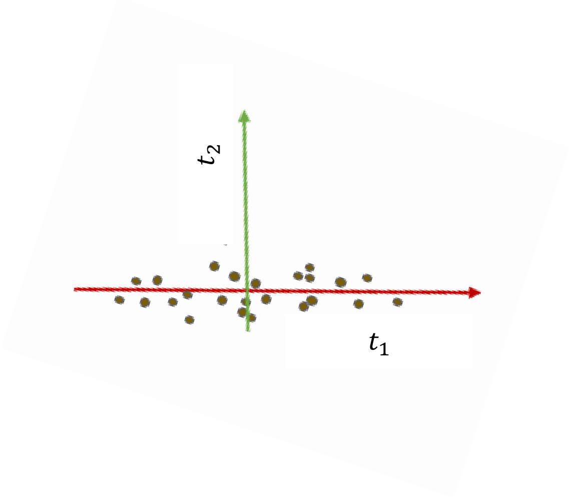
Module : PCA-based MSPM

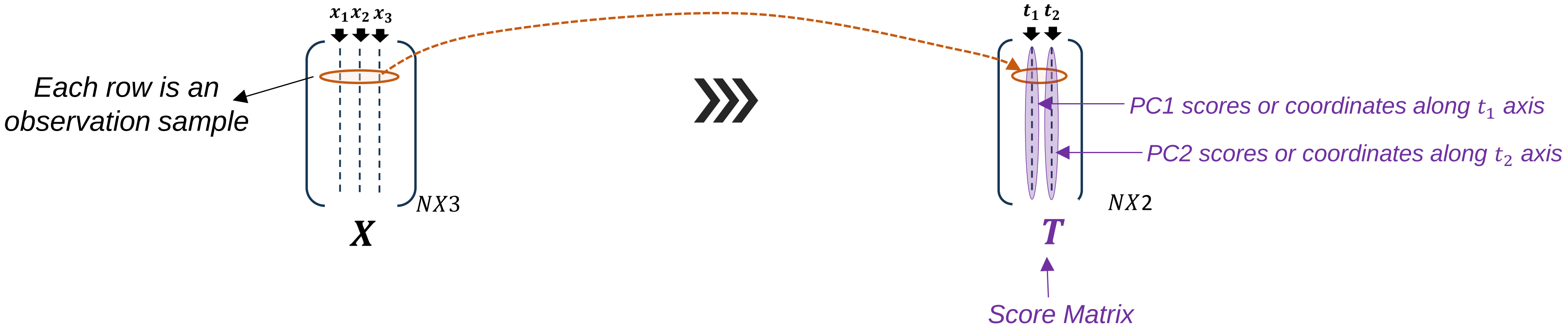
Course TOC

- 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
    - Fault detection & diagnosis (FDD) using PCA
    - Application to a Polymer Manufacturing process
  - Partial Least Squares (PLS) regression-based MSPM
  - Strategies for handling nonlinear, dynamic, multimode systems
- Deploying SPM solutions

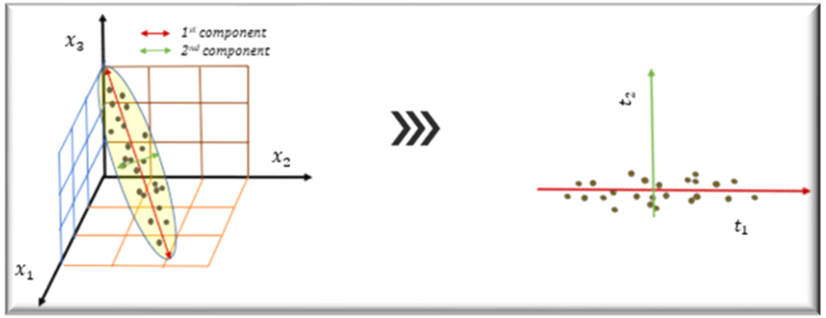
PCA Basic Terminology: Score Matrix







PCA Basic Terminology: Loading Vectors & Matrix



$$\underset{NX3}{X} \; \underset{3X1}{p_1} = \underset{NX2}{t_1}$$

PC1 scores are obtained by multiplying matrix X by a vector p_1, also called *loading vector*

$$\underset{NX3}{X} \; \underset{3X1}{p_2} = \underset{NX2}{t_2}$$

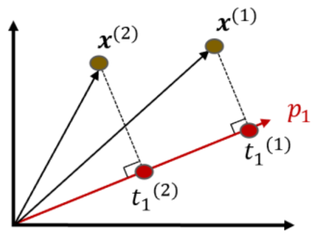
PC2 scores are obtained by multiplying matrix X by another loading vector p_2

$$X\,[p_1, p_2] = T$$

$$\Rightarrow X\underbrace{P}_{\text{Loading Matrix}} = T$$

PCA: Defining Loading Vectors

Finding p_1



$$\max_{\substack{\boldsymbol{p}_1 \\ \|\boldsymbol{p}_1\|=1}} \; variance\ of\ PC_1\ scores$$

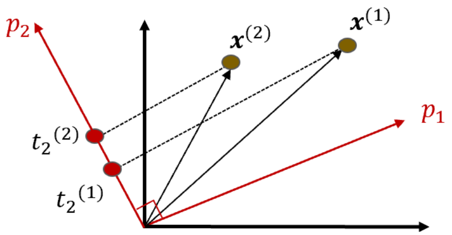
⬇

$$\max_{\substack{\boldsymbol{p}_1 \\ \|\boldsymbol{p}_1\|=1}} \; \boldsymbol{t}_1^T \boldsymbol{t}_1$$

⬇

$$\max_{\substack{\boldsymbol{p}_1 \\ \|\boldsymbol{p}_1\|=1}} \; (\boldsymbol{X}\boldsymbol{p}_1)^T \boldsymbol{X}\boldsymbol{p}_1$$

Finding p_2



$$\max_{\substack{\boldsymbol{p}_2 \\ \|\boldsymbol{p}_2\|=1 \\ \boldsymbol{p}_1^T\boldsymbol{p}_2=0}} \; variance\ of\ PC_2\ scores$$

⬇

$$\max_{\substack{\boldsymbol{p}_2 \\ \|\boldsymbol{p}_2\|=1 \\ \boldsymbol{p}_1^T\boldsymbol{p}_2=0}} \; \boldsymbol{t}_2^T \boldsymbol{t}_2$$

⬇

$$\max_{\substack{\boldsymbol{p}_2 \\ \|\boldsymbol{p}_2\|=1 \\ \boldsymbol{p}_1^T\boldsymbol{p}_2=0}} \; (\boldsymbol{X}\boldsymbol{p}_2)^T \boldsymbol{X}\boldsymbol{p}_2$$

…

Finding p_m

…

PCA: Calculating Loading Vectors via Eigenvalue Decomposition

Loading vectors can be efficiently computed via eigenvalue decomposition of sample covariance matrix

$$\frac{1}{N-1}\boldsymbol{X}^T\mathbf{X} = \boldsymbol{S} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T$$

- Software routines exist to perform the above decomposition
- $\boldsymbol{V} = [\boldsymbol{p}_1, \boldsymbol{p}_2, \cdots, \boldsymbol{p}_m]$
- $\boldsymbol{\Lambda} = diag(\lambda_1, \lambda_2, \cdots, \lambda_m)$

λ_2 *is variance of PC2 scores* ($\boldsymbol{t}_2^T\boldsymbol{t}_2$)

λ_1 *is variance of PC1 scores* ($\boldsymbol{t}_1^T\boldsymbol{t}_1$)

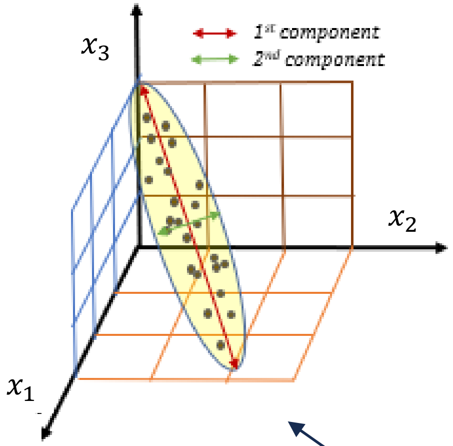
PCA: Variance Captured by Scores

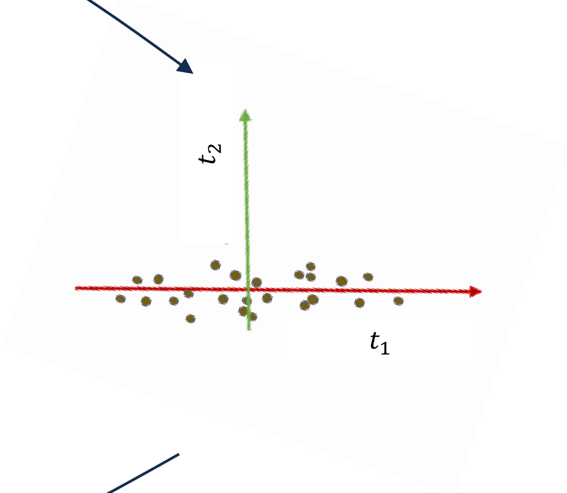
For our 3D numerical example

- Compute variances of PC1, PC2, and PC3 scores
- Discuss strategies for deciding the number of principal components to retain

PCA: Reconstruction

$$T = XP$$





$$\hat{X} = TP^T$$

If number of retained PCs equals $m \Rightarrow \hat{X} = X$

Statistical Techniques for Monitoring Industrial Processes

Next Lecture : PCA – An Industrial Case Study

Module : PCA-based MSPM