WWTP diagnosis based on robust principal component analysis

Y. Tharrault, G. Mourot, J. Ragot, M.-F. Harkat

Centre de Recherche en Automatique de Nancy
Nancy-université, CNRS
2, Avenue de la forêt de Haye
54516 Vandœuvre-lès-Nancy Cedex

7th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes
July 3, 2009, Barcelona, Spain

Outline

- Data matrix $X \in \Re^{N \times n}$ in a normal process operation

Principle of the Principal component analysis

- Data matrix $X \in \Re^{N \times n}$ in a normal process operation

PCA

Maximization of the variance projections $T = X\,P$

- $T \in \Re^{N \times n}$: principal component matrix
- $P \in \Re^{n \times n}$: projection matrix

- Data matrix $X \in \Re^{N\times n}$ in a normal process operation

PCA

Maximization of the variance projections $T = XP$

- $T \in \Re^{N\times n}$: principal component matrix
- $P \in \Re^{n\times n}$: projection matrix

Decomposition in eigenvalues/eigenvectors of the covariance matrix

$$\Sigma = \frac{1}{N-1} X^T X = P \Lambda P^T \qquad \text{with} \qquad PP^T = P^T P = I_n$$

$$\Sigma = \begin{bmatrix} P_\ell & P_{n-\ell} \end{bmatrix} \begin{bmatrix} \Lambda_\ell & 0 \\ 0 & \Lambda_{n-\ell} \end{bmatrix} \begin{bmatrix} P_\ell^T \\ P_{n-\ell}^T \end{bmatrix}$$

- Data matrix $X \in \Re^{N\times n}$ in a normal process operation

PCA

Maximization of the variance projections $T = X\,P$

- $T \in \Re^{N\times n}$: principal component matrix
- $P \in \Re^{n\times n}$: projection matrix

Decomposition in eigenvalues/eigenvectors of the covariance matrix

$$\Sigma = \frac{1}{N-1} X^T X = P \Lambda P^T \qquad \text{with} \qquad PP^T = P^T P = I_n$$

$$\Sigma = \begin{bmatrix} P_\ell & P_{n-\ell} \end{bmatrix} \begin{bmatrix} \Lambda_\ell & 0 \\ 0 & \Lambda_{n-\ell} \end{bmatrix} \begin{bmatrix} P_\ell^T \\ P_{n-\ell}^T \end{bmatrix}$$

- Data matrix $X \in \Re^{N \times n}$ in a normal process operation

PCA

Maximization of the variance projections $T = XP$

- $T \in \Re^{N \times n}$: principal component matrix
- $P \in \Re^{n \times n}$: projection matrix

Decomposition in eigenvalues/eigenvectors of the covariance matrix

$$\Sigma = \frac{1}{N-1} X^T X = P \Lambda P^T \qquad \text{with} \qquad P P^T = P^T P = I_n$$

$$\Sigma = \begin{bmatrix} P_\ell & P_{n-\ell} \end{bmatrix} \begin{bmatrix} \Lambda_\ell & 0 \\ 0 & \Lambda_{n-\ell} \end{bmatrix} \begin{bmatrix} P_\ell^T \\ P_{n-\ell}^T \end{bmatrix}$$

- Data matrix $X \in \Re^{N \times n}$ in a normal process operation

PCA

Maximization of the variance projections $T = X P$

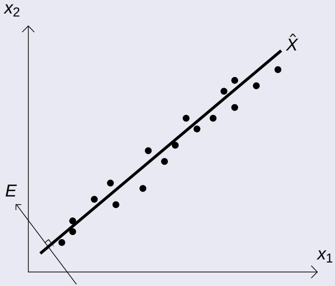
- $T \in \Re^{N \times n}$: principal component matrix
- $P \in \Re^{n \times n}$: projection matrix

Decomposition in eigenvalues/eigenvectors of the covariance matrix

$$\Sigma = \frac{1}{N-1} X^T X = P \Lambda P^T \qquad \text{with} \qquad P P^T = P^T P = I_n$$

$$\Sigma = \begin{bmatrix} P_\ell & P_{n-\ell} \end{bmatrix} \begin{bmatrix} \Lambda_\ell & 0 \\ 0 & \Lambda_{n-\ell} \end{bmatrix} \begin{bmatrix} P_\ell^T \\ P_{n-\ell}^T \end{bmatrix}$$

Decomposition

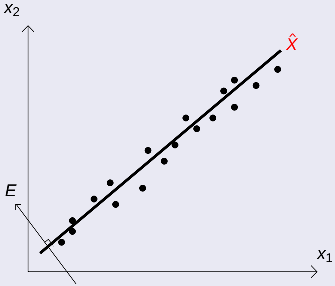


Decomposition

Principal part

$\hat{X} = X\, C_\ell$

with $C_\ell = P_\ell P_\ell^T$



Decomposition

Principal part

$\hat{X} = X\,C_\ell$
with $C_\ell = P_\ell P_\ell^T$

Residual part

$E = X - \hat{X}$
$\quad = X\,(I_n - C_\ell)$
$\quad = X\,P_{n-\ell}P_{n-\ell}^T$

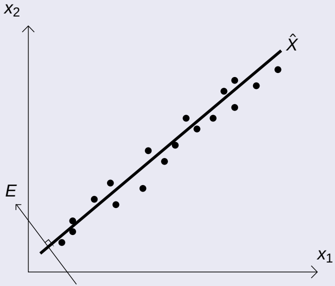
Decomposition

Principal part

$\hat{X} = X\,C_\ell$
with $C_\ell = P_\ell P_\ell^T$

Residual part

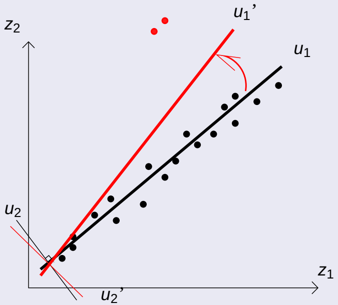
$E = X - \hat{X}$
$= X\,(I_n - C_\ell)$
$= X\,P_{n-\ell}P_{n-\ell}^T$



Determination of the number of principal components ℓ

Sensitive to outliers

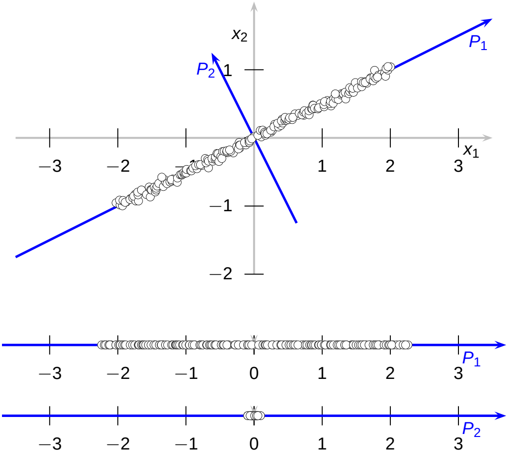
Outliers: Data different from the normal operating conditions (faulty data, data obtained during shutdown or stratup periods or data issued form different operating mode)



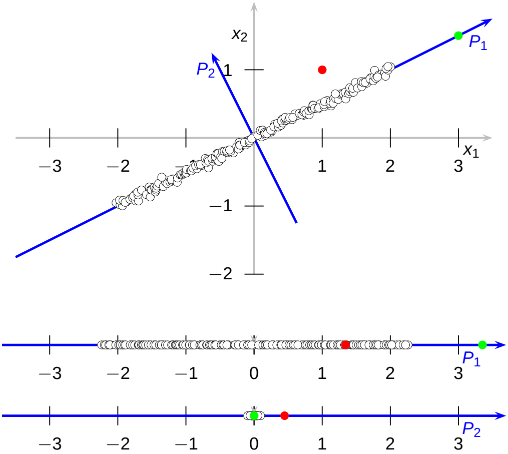
Robust PCA with respect to outliers

⟶ Outliers detection and isolation

- $n = 2$ variables (x_1, x_2)
- $X = [x_1\ x_2]$
- $\ell = 1$



- $n = 2$ variables (x_1, x_2)
- $X = [x_1\ x_2]$
- $\ell = 1$
- Outliers 1 : green observation
- Outliers 2 : red observation



Residual $r(k)$

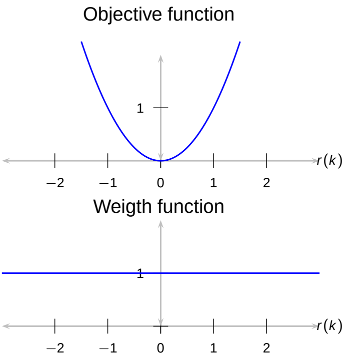
$$r(k) = ||P_{n-\ell}^T(x(k) - \mu)||^2$$

with $x(k)$ an observation
μ the mean of the data X
$P_{n-\ell}$ is the eigenvector matrix of the robust covariance matrix corresponding to its $n-\ell$ smallest eigenvalues

PCA minimise the following criterion:

$$\frac{1}{N}\sum_{k=1}^{N}(r(k))$$

with the constraint $P_{n-\ell}^T P_{n-\ell} = I_{n-\ell}$.



Residual $r(k)$

$$r(k) = ||P_{n-\ell}^T(x(k) - \mu)||^2$$

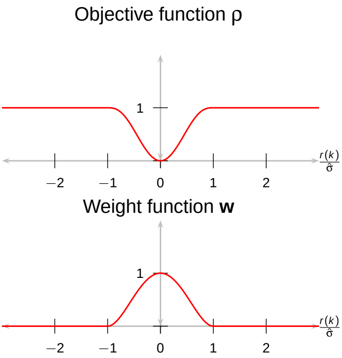
with $x(k)$ an observation
μ the mean of the data X
$P_{n-\ell}$ is the eigenvector matrix of the robust covariance matrix corresponding to its $n-\ell$ smallest eigenvalues

The general scale-M estimator minimizes the following objective function with the constraint $P_{n-\ell}^T P_{n-\ell} = I_{n-\ell}$:

$$\frac{1}{N}\sum_{k=1}^{N} \rho\left(\frac{r(k)}{\hat{\sigma}}\right)$$

with the function ρ as the objective function, $\hat{\sigma}$ the robust scale of the residual $r_(k)$ calculated by minimising the following criterion:

$$\frac{1}{N}\sum_{k=1}^{N} \rho\left(\frac{r(k)}{\hat{\sigma}}\right) = \delta$$



Initialization with a robust covariance matrix

Robust covariance matrix

$$V = \frac{\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} w(i,j)(x(i)-x(j))(x(i)-x(j))^T}{\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} w(i,j)}$$

where the weights $w(i,j)$ themselves are defined by:

$$w(i,j) = \exp\left(-\frac{\beta}{2}(x(i)-x(j))^T\Sigma^{-1}(x(i)-x(j))\right)$$

with β turning parameter

For $\beta = 0$, $V = 2\Sigma$

Only robust to outliers with a projection onto the residual space

sacle M-estimator

The scale M-estimator maximizes the following criterion with the constraint $P_\ell^T P_\ell = I_\ell$:

$$\frac{1}{N}\sum_{k=1}^{N} \rho\left(\frac{||P_\ell^T\,(x(k)-\mu)||^2}{\hat{\sigma}}\right)$$

with $\hat{\sigma}$ the robust scale of the residual r
and ρ the objective function.

The reconstruction $\hat{x}_R$

Minimizing the influence of fault

$$\hat{x}_R(k) = x(k) - \Xi_R f_R$$

with $x(k)$: an observation

f_R : the fault magnitude (unknown)

Ξ_R : matrix of reconstruction directions

The reconstruction $\hat{x}_R$

Minimizing the influence of fault

$$\hat{x}_R(k) = x(k) - \Xi_R f_R$$

with $x(k)$: an observation
f_R : the fault magnitude (unknown)
Ξ_R : matrix of reconstruction directions

For example, to reconstruct 2 variables ($R = 2, 4$) among 5 variables

$$\Xi_R = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}^T$$

The reconstruction $\hat{x}_R$

Minimizing the influence of fault

$$\hat{x}_R(k) = x(k) - \Xi_R f_R$$

with $x(k)$: an observation
f_R : the fault magnitude (unknown)
Ξ_R : matrix of reconstruction directions

For example, to reconstruct 2 variables ($R = 2, 4$) among 5 variables

$$\Xi_R = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}^T$$

Estimation of the fault magnitude $\hat{f}_R$:

$$\hat{f}_R = \arg\min_{f_R} \{D_R(k)\}$$

with $D_R(k) = \hat{x}_R(k)^T P \Lambda^{-1} P^T \hat{x}_R(k)$

The reconstruction vector $\hat{x}_R(k)$ of the vector $x(k)$ is given by:

$$\hat{x}_R(k) = \left(I - \Xi_R(\Xi_R^T P \Lambda^{-1} P^T \Xi_R)^{-1} \Xi_R^T P \Lambda^{-1} P^T\right) x(k)$$

The reconstruction vector $\hat{x}_R(k)$ of the vector $x(k)$ is given by:

$$\hat{x}_R(k) = \left(I - \Xi_R(\Xi_R^T P \Lambda^{-1} P^T \Xi_R)^{-1} \Xi_R^T P \Lambda^{-1} P^T\right) x(k)$$

Condition of reconstruction:

Existence of $(\Xi_R^T P \Lambda^{-1} P^T \Xi_R)^{-1} \Rightarrow$ matrix $\Xi_R^T P \Lambda^{-1} P^T \Xi_R$ of full rank

The reconstruction vector $\hat{x}_R(k)$ of the vector $x(k)$ is given by:

$$\hat{x}_R(k) = \left(I - \Xi_R(\Xi_R^T P \Lambda^{-1} P^T \Xi_R)^{-1} \Xi_R^T P \Lambda^{-1} P^T\right) x(k)$$

Condition of reconstruction:

Existence of $(\Xi_R^T P \Lambda^{-1} P^T \Xi_R)^{-1} \Rightarrow$ matrix $\Xi_R^T P \Lambda^{-1} P^T \Xi_R$ of full rank

To reconstruct a fault, it must be at least projected onto the principal space ($r \leq \ell$) or onto the principal space ($r \leq n - \ell$). This implies that the number of reconstructed variables r must respect the following inequality:

$$r \leq \max(n - \ell, \ell)$$

with r : Number of reconstructed variables

ℓ : Number of principal components

n : Number of variables

The reconstruction vector $\hat{x}_R(k)$ of the vector $x(k)$ is given by:

$$\hat{x}_R(k) = \left(I - \Xi_R(\Xi_R^T P \Lambda^{-1} P^T \Xi_R)^{-1} \Xi_R^T P \Lambda^{-1} P^T\right) x(k)$$

Condition of reconstruction:

Existence of $(\Xi_R^T P \Lambda^{-1} P^T \Xi_R)^{-1} \Rightarrow$ matrix $\Xi_R^T P \Lambda^{-1} P^T \Xi_R$ of full rank

To reconstruct a fault, it must be at least projected onto the principal space ($r \leq \ell$) or onto the principal space ($r \leq n - \ell$). This implies that the number of reconstructed variables r must respect the following inequality:

$$r \leq \max(n - \ell, \ell)$$

with r : Number of reconstructed variables
ℓ : Number of principal components
n : Number of variables

The number of maximum reconstruction:

$$\sum_{r=1}^{\max(n-\ell,\ell)-1} \mathbb{C}_n^r$$

with $\mathbb{C}_n^r$ denotes the combination of r from n.

Fault detection indicator D_R

$$D_R(k) = \hat{x}_R(k)^T P \Lambda^{-1} P^T \hat{x}_R(k)$$

Fault detection indicator D_R

$$D_R(k) = \hat{x}_R(k)^T P \Lambda^{-1} P^T \hat{x}_R(k)$$

Fault detection

A fault is detected, if:

$$D_R(k) > \gamma_\alpha^2$$

with γ_α^2 the detection threshold of indicator D_R,

Fault detection indicator D_R

$$D_R(k) = \hat{x}_R(k)^T P \Lambda^{-1} P^T \hat{x}_R(k)$$

Fault detection

A fault is detected, if:

$$D_R(k) > \gamma_\alpha^2$$

with γ_α^2 the detection threshold of indicator D_R,

Fault isolation

For the faulty observations, the faulty variables $\hat{R}$ are determined as follows:

$$\hat{R} = \arg_{R \in \Im} D_R(k) < \gamma_\alpha^2$$

with γ_α^2 the detection threshold of indicator D_R
and $\Im$ all combinations of possible reconstruction directions.

Reduction of the number of reconstruction

Determination of the colinear direction projection

a global indicator K is built:

$$K(R_1,R_2) = \max\{(d(R_1,R_2),\tilde{d}(R_1,R_2)\}$$

with R_1 and R_2 correspond to sets of variable reconstruction
and $d(R_1,R_2)$ distance between two sub-spaces onto principal space
and $\tilde{d}(R_1,R_2)$ distance between two sub-spaces onto residual space

$$d(R_1,R_2) = ||\hat{\Xi}_{R_1}(\hat{\Xi}_{R_1}^T\hat{\Xi}_{R_1})^{-1}\hat{\Xi}_{R_1}^T - \hat{\Xi}_{R_2}(\hat{\Xi}_{R_2}^T\hat{\Xi}_{R_2})^{-1}\hat{\Xi}_{R_2}^T||_2$$
$$\tilde{d}(R_1,R_2) = ||\tilde{\Xi}_{R_1}(\tilde{\Xi}_{R_1}^T\tilde{\Xi}_{R_1})^{-1}\tilde{\Xi}_{R_1}^T - \tilde{\Xi}_{R_2}(\tilde{\Xi}_{R_2}^T\tilde{\Xi}_{R_2})^{-1}\tilde{\Xi}_{R_2}^T||_2$$

with $\hat{\Xi}_{R_1} = \Lambda_\ell^{-1/2} P_\ell^T \Xi_{R_1}$, $\tilde{\Xi}_{R_1} = \Lambda_{n-\ell}^{-1/2} P_{n-\ell}^T \Xi_{R_1}$.

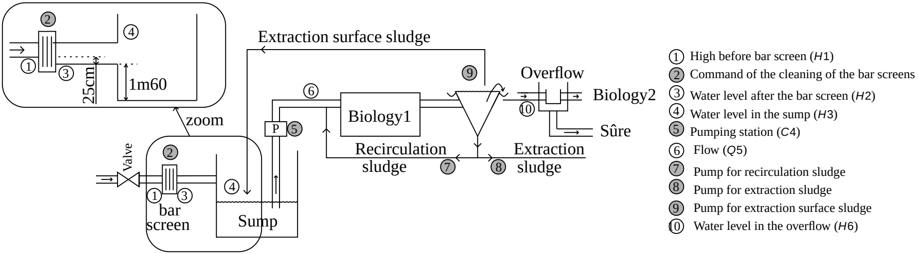
For example, to reconstruct 2 variables ($R_1 = 2, 4$) among 5 variables

$$\Xi_{R_1} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}^T$$

1. $r = 1$
2. Calculate for all available directions ($R_1 \in \Im$ and $R_2 \in \Im$) the indicator $K(R_1, R_2)$.
   If $K(R_1, R_2)$ is equal to zero:
   - only a set of variables potentially faulty may be determined, i.e. the faulty variables are associated to the indices R_1 or R_2 or R_1 and R_2. Thus, it is only required to determine one direction, for example R_1.

   If $K(R_1, R_2)$ is closed to zero:
   - magnitude of the fault has to be important to ensure fault isolation.

   Else the fault are isolable
3. $r = r + 1$
4. While $r < \max(\ell, n - \ell)$ do to the step 2



Construction of the data matrix

Dynamic process : Temporal lag

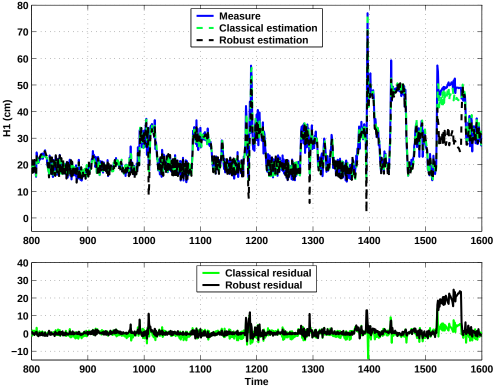
Non linear process : Transformed variables

$$x(k) = [\ H1(k) \quad H2(k) \quad H3(k) \quad Q5(k) \quad H6(k) \quad \tanh((Q5(k-1) - 550)./150) \quad H1(k-1) \quad H6(k-1) \quad C4(k)\ ]^T \tag{1}$$

The data matrix X is constituted of N observations of the vector $x(k)$.

Construction of the model

4 principal components are determined



Measure and estimation of $H1$

Analysis of the reconstruction directions

The maximum number of reconstructions is then equal to 255 ($max(n-\ell,\ell)-1=4$)

K		R_1								
		1	2	3	4	5	6	7	8	9
R_2	1	0	0.94	0.99	1.00	0.99	0.98	0.97	0.99	1.00
	2		0	0.52	1.00	0.98	1.00	1.00	1.00	1.00
	3			0	0.99	1.00	1.00	1.00	1.00	0.99
	4				0	1.00	0.99	1.00	1.00	0.33
	5					0	0.82	0.96	0.90	0.99
	6						0	0.80	1.00	1.00
	7							0	1.00	0.97
	8								0	0.97

Indicator K for $r=1$

The number of useful reconstruction can be reduced to 202

Fault detection

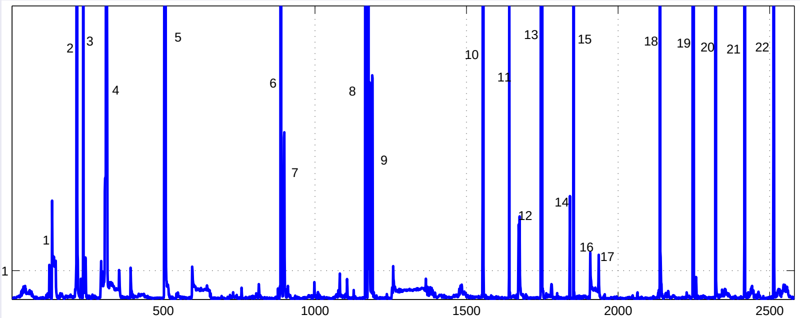


Figure: Fault detection with Mahalanobis distance

Fault isolation

Fault index	Reconstruction direction under the detection threshold
16, 17	D_1
3, 10, 11, 14, 15, 19, 20, 21	D_3
7, 12	$D_{1,6}$
6, 8, 13, 18, 22	$D_{3,9}$
1, 2, 5	$D_{1,3,4}$
4	$D_{3,7,9}$
9	$D_{1,2,3,7}$

Table: Summary of fault isolation

Faults 16 and 17 are under the threshold when the first variable ($H1(k)$) is reconstructed.
$\Rightarrow$ variable $H1(k)$ is faulty

Conclusion

- Robust PCA with respect to outliers
  –> directly applicable on data containing potential faults
- use of the principle of reconstruction and projection of the reconstructed data together
  –> outliers detection and isolation
- Reduction of the computational load
  –> determination of the dectable and isolable faults
- Application to hydraulic part of a wastewater treatment plant