MODEL IDENTIFICATION USING ITERATIVE PCA

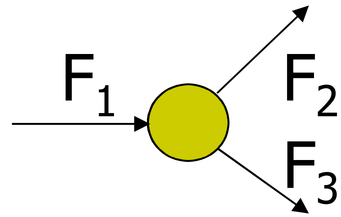
Shankar Narasimhan
Dept. of Chemical Engineering,
IIT Madras, Chennai, INDIA

MODEL IDENTIFICATION USING PCA

- Model $Az^*(i) = 0$
- Measurements : $z(i) = z^*(i) + e(i)$
- Errors $e(i)$ identically distributed with mean zero and covariance matrix Σ_e
- Model used in chemometrics
- Can be extended to dynamic process model identification

MOTIVATING EXAMPLE

- Simple flow process



- At steady state : $F_1 - F_2 - F_3 = 0$
- Constraint model : $\begin{bmatrix} 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \\ F_3 \end{bmatrix} = 0$

$$A \quad z \quad = \quad 0$$

Constraint matrix true values

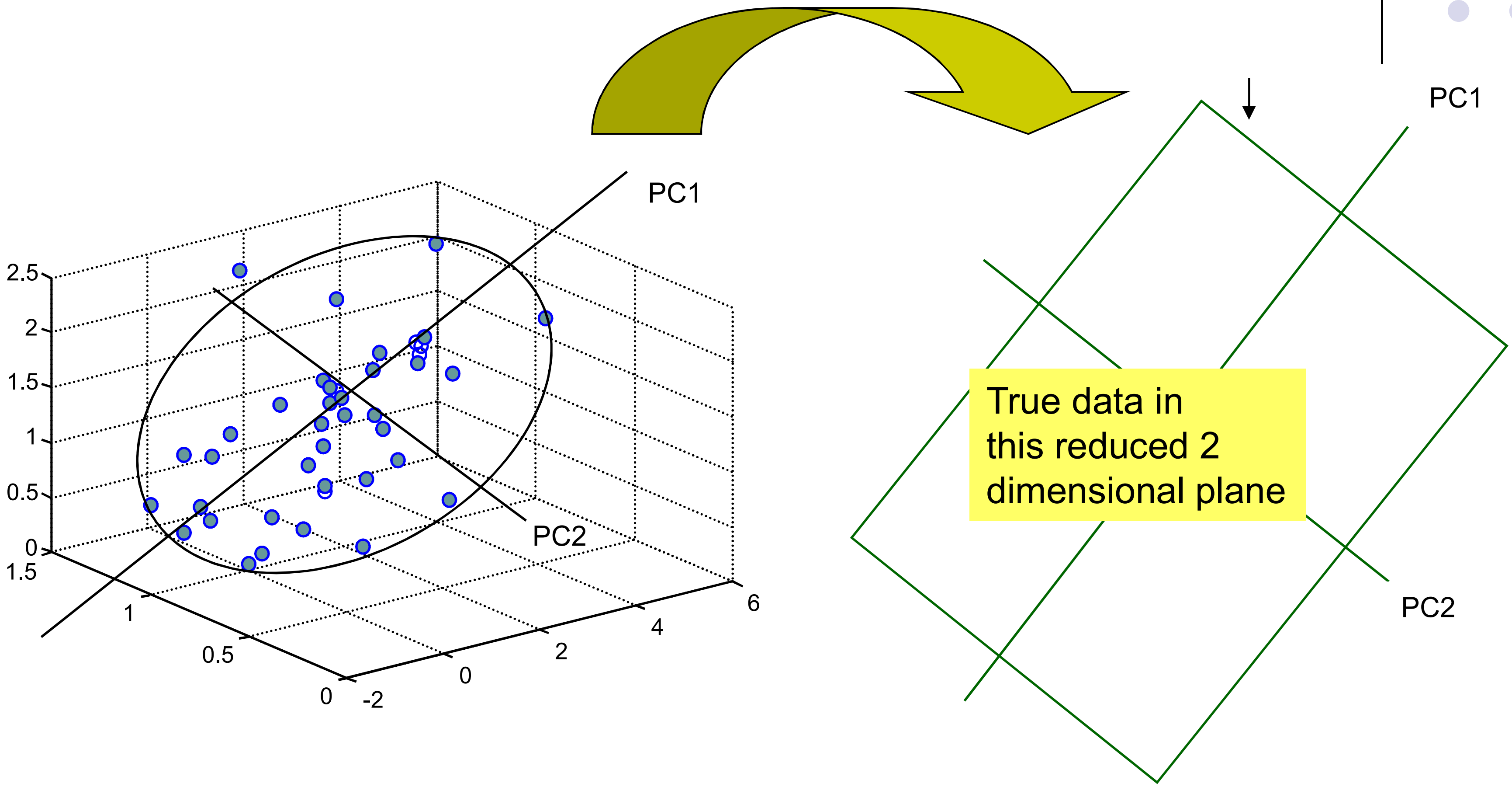
- *Objective: estimate A from a sample of measurements of z*

RELATION TO REGRESSION

- Since one relation exists between three flow variables, can choose one flow variable as dependent (say F_1) and the remaining two as independent variables
- Can relate all flows to independent flow variables

$$\underbrace{\begin{bmatrix} F_1 \\ F_2 \\ F_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\text{Regression matrix B}} \underbrace{\begin{bmatrix} F_2 \\ F_3 \end{bmatrix}}_{\text{Independent variables}}$$

- *Objective: Estimate B from a sample of measurements of z*



3D Euclidean basis space

Principal subspace

PCA – Method

- Data matrix

$$\mathbf{Z} = \begin{bmatrix} z_1(1) & z_2(2) & \ldots & z_1(N) \\ z_2(1) & z_2(2) & \ldots & z_2(N) \\ \vdots & \vdots & \vdots & \vdots \\ z_n(1) & z_n(2) & \cdots & z_n(N) \end{bmatrix}$$

- Find orthonormal eigenvectors of $\mathbf{S_z}$

$$\mathbf{S_z} = \frac{1}{N}\mathbf{Z}\mathbf{Z}^T$$

- Eigenvectors corresponding to largest n–m eigenvalues is basis for the data subspace
- Transpose of eigenvectors corresponding to last *m* smallest eigenvalues (which are rejected) give a basis for the row space of constraints between variables **Az** = **0**

PCA - INTERPRETATION

- Given a sample of data in R^n
  - PCA decomposes R^n into two orthogonal subspaces V_z and V_c such that V_z is the space in which data lies and V_c is the row space of constraints
  - The orthonormal eigenvectors form orthonormal bases for V_z and V_c
  - Eigenvector corresponding to largest eigenvalue is also direction of maximum variability in data (Is it useful?)

CASE – Perfect measurements

- Data variance matrix $\mathbf{S}_{\mathbf{z}}$ ($\mathbf{S}_{\mathbf{z}^*}$) will be singular
- Number of zero eigenvalues of $\mathbf{S}_{\mathbf{z}}$ is equal to number of relationships
- Eigenvectors corresponding to non-zero eigenvalues is a basis for V_z
- Eigenvectors corresponding to zero eigenvalues is a basis for V_c

CASE – Perfect Measurements Effect of Scaling

- If PCA applied to transformed data $\mathbf{z}_s = \mathbf{D}\mathbf{z}$
  - Number of zero eigenvalues still equal to number of relations
  - Eigenvectors corresponding to zero eigenvalues gives a basis $\mathbf{A}_s$ for the space orthogonal to $\mathbf{z}_s$ space
  - Basis (not orthonormal) for V_c can be obtained as $\mathbf{A}_s\mathbf{D}$

PROBLEMS WITH PCA

- How to deal with noisy data in a statistically rigorous manner?
- What happens if we scale the data before applying PCA?
  - PCA is scale dependent
  - Optimal scaling method unknown
- How to determine the number of independent variables (number of relations or model order)?
  - Heuristic criteria used for selecting model order

PCA - Optimization Problem

- The solution for A obtained using PCA is the solution of the following optimization problem

$$\underset{\mathbf{A},\,\mathbf{z}^*(i)}{Min} \quad \sum_{i=1}^{N} [\mathbf{z}(i) - \mathbf{z}^*(i)]^T [\mathbf{z}(i) - \mathbf{z}^*(i)]$$

$$\mathbf{A}\mathbf{z}^*(i) = \mathbf{0} \quad \forall \; i = 1 \ldots N$$

$$\mathbf{A}\mathbf{A}^T = \mathbf{I}$$

- Note all variables are given same weight in objective function
- PCA is statistically optimal if errors in all variables have identical variance

PCA using known Σ_e

- Square root of $\boldsymbol{\Sigma}_e$ $(\boldsymbol{\Sigma}_e = \mathbf{L}\mathbf{L}^T)$
- Transform data $\mathbf{z}_s = \mathbf{L}^{-1}\mathbf{z}$ $\Rightarrow$ $\boldsymbol{\Sigma}_{\mathbf{z}_s} = \boldsymbol{\Sigma}_{\mathbf{z}_s^*} + \mathbf{I}$
- Eigenvectors of $\boldsymbol{\Sigma}_{\mathbf{z}_s}$ same as $\boldsymbol{\Sigma}_{\mathbf{z}_s^*}$
- Eigenvalues of $\boldsymbol{\Sigma}_{\mathbf{z}_s}$ equal to eigenvalues of $\boldsymbol{\Sigma}_{\mathbf{z}_s^*}$ shifted by unity

  $\Rightarrow$ Number of unity eigenvalues of $\boldsymbol{\Sigma}_{\mathbf{z}_s}$ is equal to number of relationships
- Eigenvectors corresponding to unity eigenvalues of $\boldsymbol{\Sigma}_{\mathbf{z}_s}$ gives basis for V_c

Case – Noisy measurements Unknown Σ_e

- Simultaneously obtain estimates for $\boldsymbol{\Sigma}_e$ and **A**
- Estimate $\boldsymbol{\Sigma}_e$ given an initial estimate $\hat{\mathbf{A}}^0$ using likelihood function of constraint residuals

$$\mathbf{r}(j) = \hat{\mathbf{A}}^0\mathbf{z}(j) \sim N(0, \boldsymbol{\Sigma}_r) \quad \boldsymbol{\Sigma}_r = \hat{\mathbf{A}}^0\boldsymbol{\Sigma}_e(\hat{\mathbf{A}}^0)^T$$

$$\underset{\boldsymbol{\Sigma}_e}{Min} \;\; N\log\left|\hat{\mathbf{A}}^0\boldsymbol{\Sigma}_e\hat{\mathbf{A}}^{0^T}\right| + \sum_{j=1}^{N}\mathbf{r}^T(j)(\hat{\mathbf{A}}^0\boldsymbol{\Sigma}_e\hat{\mathbf{A}}^{0^T})^{-1}\mathbf{r}(j)$$

  - Above procedure is similar to estimating $\boldsymbol{\Sigma}_e$ given covariance matrix of constraint residuals $\boldsymbol{\Sigma}_\mathbf{r}$

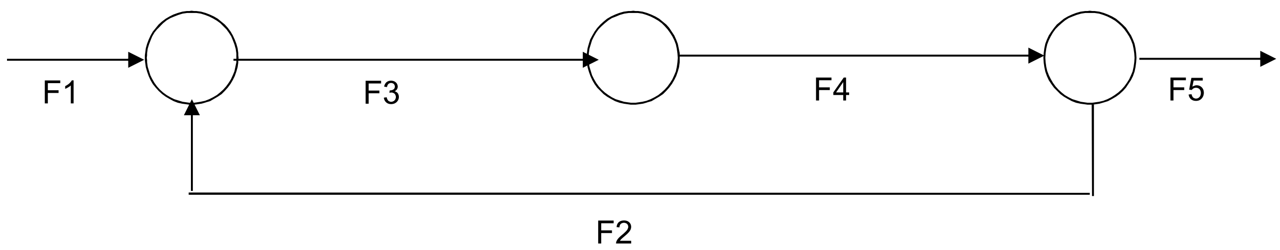
ALGORITHM – ITERATIVE PCA

- Use PCA on original data to get initial estimate of A say A^0
- Obtain estimate of Σ_e by maximizing likelihood function of constraint residuals
- Use estimated Σ_e to scale data and apply PCA to get new estimate of A.
- Repeat steps 2 and 3 until convergence

Estimating Σ_e - Limitations

- Estimating Σ_e from Σ_r has limitations
  - Estimating full Σ_e from $\mathbf{\Sigma_r}$ does not give unique solution since $\mathbf{\Sigma_r}$ is m x m and $\mathbf{\Sigma_e}$ is n x n (n usually greater than m)
  - Maximum # of elements of Σ_e that can be estimated is m(m+1)/2 (exploiting symmetry)
  - If Σ_e is assumed to be diagonal, then it can be estimated provided $m(m+1)/2 \geq n$

EXAMPLE - FLOW PROCESS



Constraints m = 3 ➔ m(m+1)/2 = 6

Variables n = 5

SIMULATION PARAMETERS

Flow variable	True values		Std of measurement error
	Base value	Std of fluctuation	
F1	10	1.0	0.1
F2	10	2.0	0.08
F3	F1 + F2		0.15
F4	F3		0.2
F5	F4 – F2		0.18

RESULTS – High S/N ratio

Case	Scale	$\alpha \times 10^3$	θ (deg)
PCA	None	5.86	0.17
PCA	σ_y	10.22	0.24
PCA	σ_ϵ	1.62	0.028
IPCA		1.2	0.03

α is distance between the rows of true A and estimated subspace A

θ Is the angle (subspace angle) between estimated A and true A

RESULTS - Eigenvalues

- Eigenvalues

  PCA Case 1: [33.32 1.9 0.18 0.16 0.11]
  Case 2 : [238.9 18.8 1.06 0.99 0.97]
  Case 3 : [19.84 1.35 0.15 0.08 0.06]

  IPCA : [236.5 17.7 1.01 1.0 0.99].

- Estimated Σ_e using IPCA

  [0.1121 0.0837 0.1406 0.2031 0.1775] - Estimated
  [0.1 0.08 0.15 0.20 0.18] - True

RESULTS – Low S/N ratio

Signal fluctuations standard deviations reduced to 0.2 each

Case	Scale	$\alpha \times 10^3$	θ (deg)
PCA	None	447.0	12.73
PCA	σ_y	252.1	7.61
PCA	σ_ϵ	21.1	0.49
IPCA		32.5	1.39

Converged eigenvalues for IPCA are [233.5 2.6 1.01 1.0 0.98]

RESULTS

- m incorrectly assumed as 4
  - Eigenvalues obtained using IPCA :

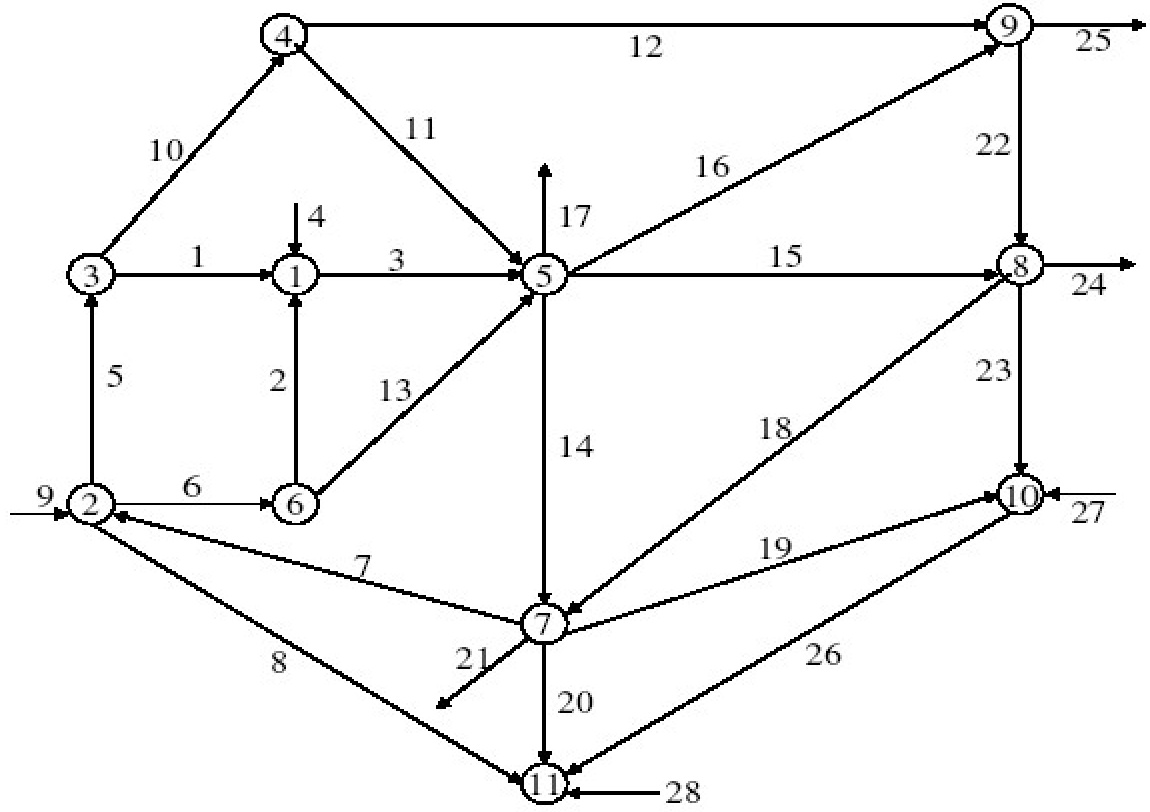
    [434.37 1.72 1.00 0.15 0.11]

  - Estimated error covariance matrix

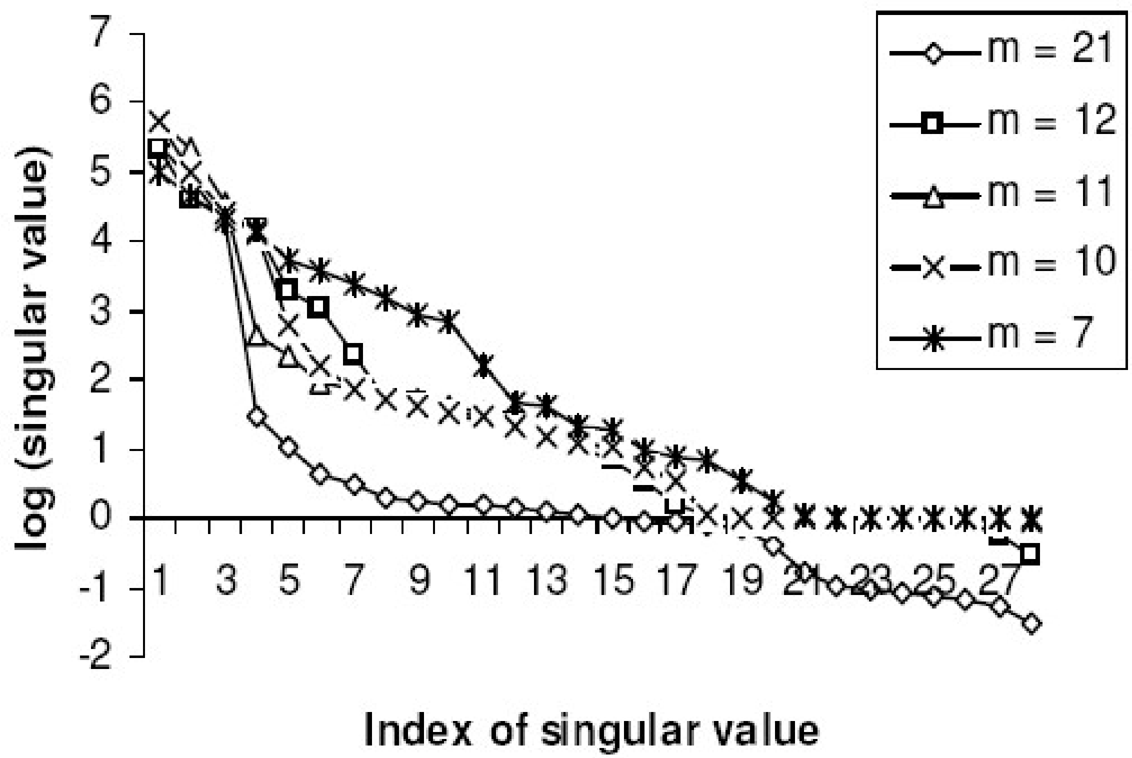
    [0.1121 0.0838 0.1406 0.2031 0.1774]

LARGER EXAMPLE

n =28, m = 11



RESULTS – MODEL ORDER DETERMINATION



If model order is overestimated, the last m eigenvalues are not equal to unity

CONCLUSIONS

- An improved PCA method developed for heteroscedastic noise – Noise variances estimated simultaneously along with model
- Optimal scaling issue resolved
- Rigorous method for model order determination based on analysis of eigenvalues

ACKNOWLEDGEMENTS

- Prof. Sirish Shah, University of Alberta, Canada
- Dr. Nirav Bhatt, DST Inspire faculty, Chemical Engg. Dept. IIT Madras