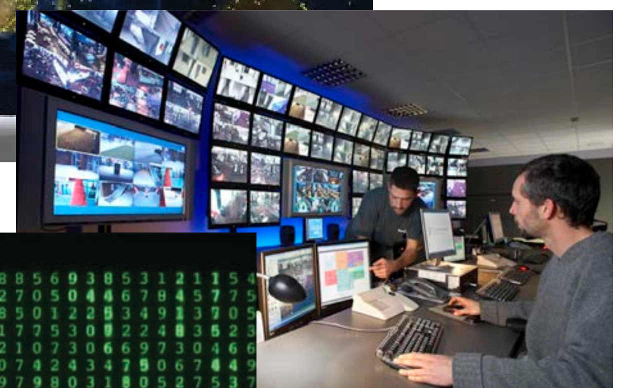
Northeastern University

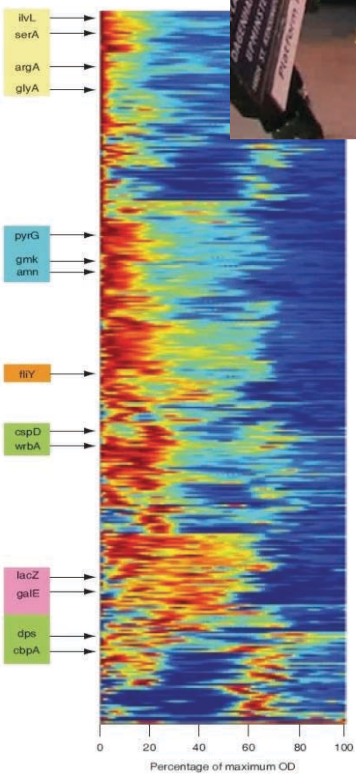
Moments Based Relaxations in Systems Identification and Machine Learning

M. Sznaier

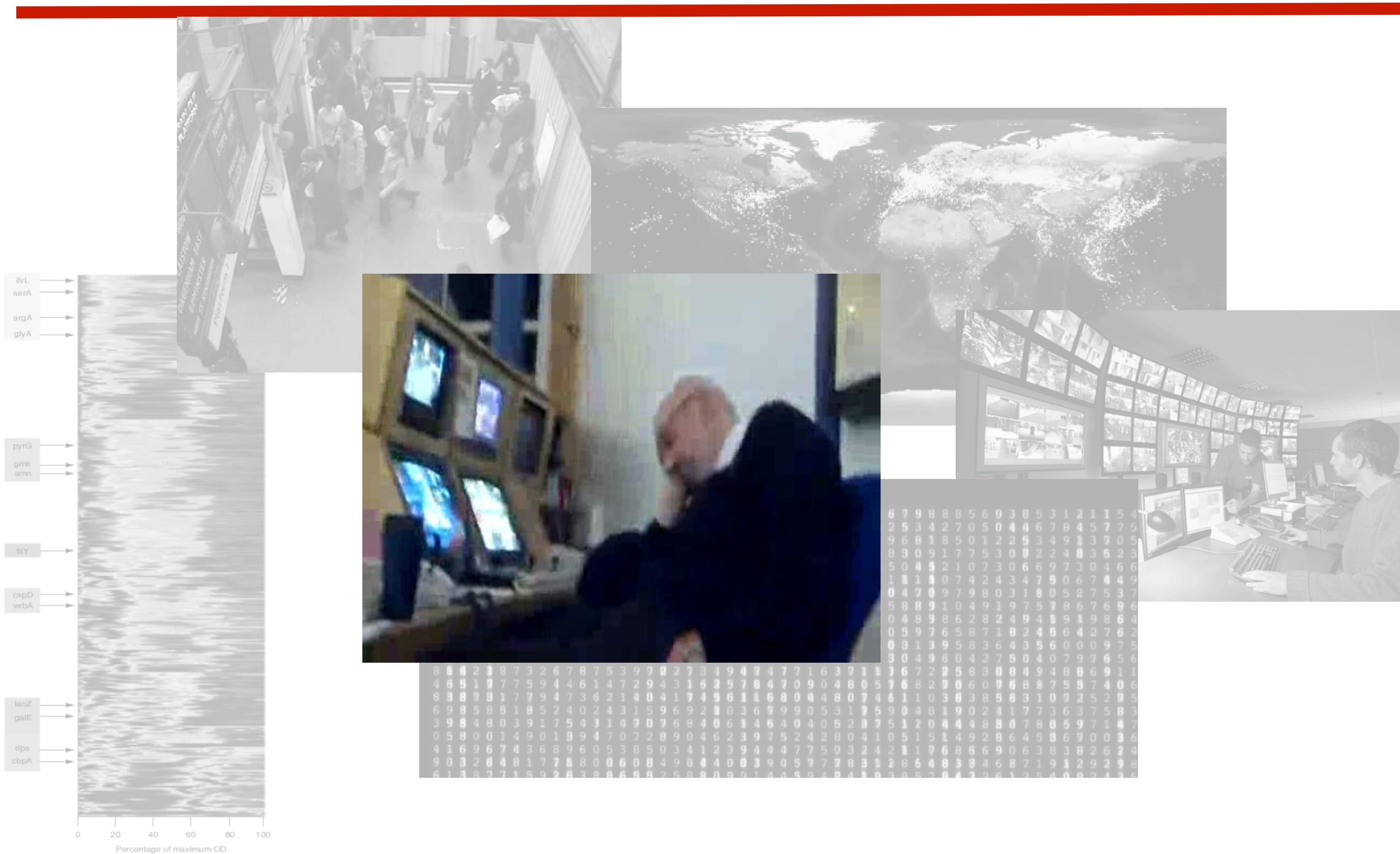
Robust Systems Lab.
Dept. of Electrical and Computer Eng.
Northeastern University

Motivation: data deluge

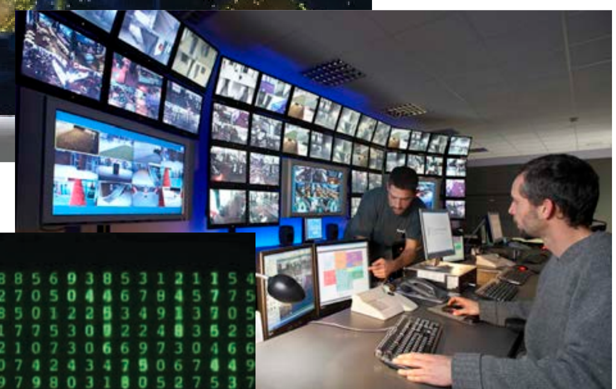


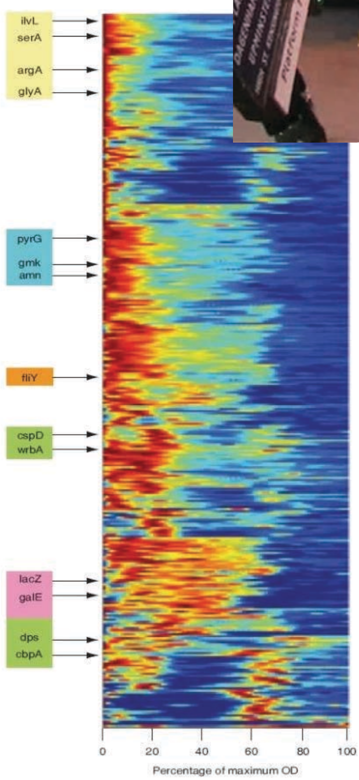


Motivation: data deluge



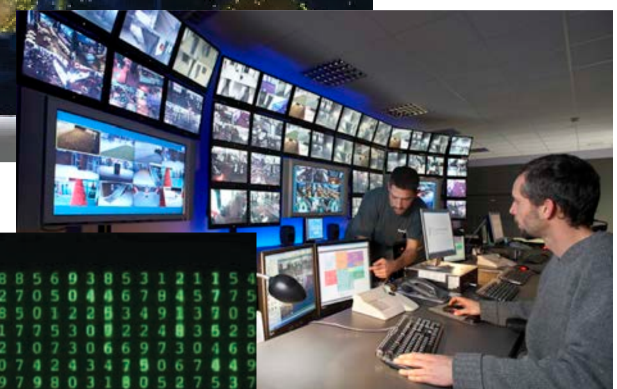
Motivation: data deluge

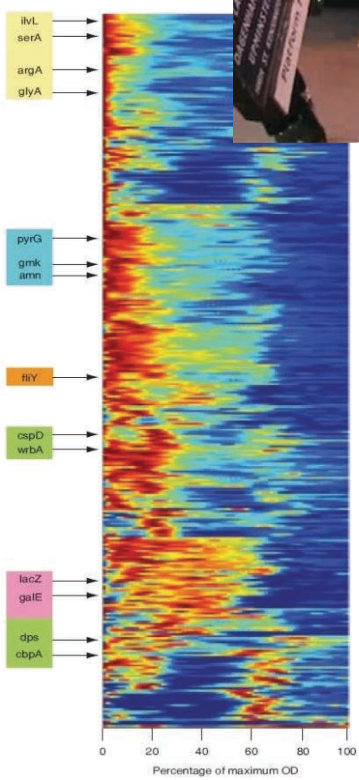




Claim: A hidden dynamical systems identification problem

Motivation: data deluge





Claim: A hidden "structured" robust PCA problem

Structured Robust PCA Problems

Prototype SRPCA problem:

- $$\begin{aligned}&\mathbf{min}\|\mathbf{S}(\mathbf{x})\|_{\mathbf{W},*} + \lambda_1 \|\mathbf{e}_1\|_1 + \tfrac{\lambda_2}{2}\|\mathbf{e}_2\|_2^2\\ &\text{subject to:}\\ &\mathbf{d} - \mathbf{F}\mathbf{x} = 0\\ &\mathbf{S}_1\mathbf{x} - \mathbf{e}_1 = 0\\ &\mathbf{S}_2\mathbf{x} - \mathbf{e}_2 = 0\\ &\mathbf{S}_\mathbf{Q}(\mathbf{x}) \succeq 0, \qquad \mathbf{S}, \mathbf{S}_\mathbf{Q} \text{ affine structural constraints}\end{aligned}$$

  - **Generalization of decompose H= H_L + H_E**

    H_L: low rank
    H_E: sparse

Prototype SRPCA problem:

- $$\begin{aligned}&\mathbf{min}\|\mathbf{S}(\mathbf{x})\|_{\mathbf{W},*}+\lambda_1\left\|\mathbf{e}_1\right\|_1+\tfrac{\lambda_2}{2}\left\|\mathbf{e}_2\right\|_2^2\\&\text{subject to:}\\&\mathbf{d}-\mathbf{F}\mathbf{x}=0\\&\mathbf{S}_1\mathbf{x}-\mathbf{e}_1=0\\&\mathbf{S}_2\mathbf{x}-\mathbf{e}_2=0\\&\mathbf{S}_{\mathbf{Q}}(\mathbf{x})\succeq 0,\qquad \mathbf{S},\mathbf{S}_{\mathbf{Q}}\text{ affine structural constraints}\end{aligned}$$

- **Solvable using interior point methods, but poor scaling properties (time: $O(n^3)$, memory $O(n^2)$)**

Prototype SRPCA problem:

- $$\begin{aligned}&\mathbf{min}\|\mathbf{S}(\mathbf{x})\|_{\mathbf{W},*} + \lambda_1 \|\mathbf{e}_1\|_1 + \tfrac{\lambda_2}{2} \|\mathbf{e}_2\|_2^2 \\ &\text{subject to:} \\ &\mathbf{d} - \mathbf{F}\mathbf{x} = 0 \\ &\mathbf{S}_1\mathbf{x} - \mathbf{e}_1 = 0 \\ &\mathbf{S}_2\mathbf{x} - \mathbf{e}_2 = 0 \\ &\mathbf{S}_{\mathbf{Q}}(\mathbf{x}) \succeq 0, \qquad \mathbf{S}, \mathbf{S}_{\mathbf{Q}} \text{ affine structural constraints}\end{aligned}$$

- **Solvable using interior point methods, but poor scaling properties (time: $O(n^3)$, memory $O(n^2)$)**
- **Alternative: ADMM methods**

Prototype ADMM:

$$\min f(\mathbf{x}_1, \ldots, \mathbf{x}_n) \text{ subject to: } h(\mathbf{x}_1, \ldots, \mathbf{x}_n) = 0$$

$$\min L(\mathbf{x}, \mathbf{Y}, \mu) \doteq f(\mathbf{x}_1, \ldots, \mathbf{x}_n) + \langle \mathbf{Y}, h(\mathbf{x}_1, \ldots, \mathbf{x}_n) \rangle + \frac{\mu}{2} \| h(\mathbf{x}_1, \ldots, \mathbf{x}_n) \|_F^2$$

while not converged **do**

1. **For i=1,...,n**

$\mathbf{x}_i^{k+1} = argmin_{\mathbf{x}_i} L(\mathbf{x}_1^{k+1}, \ldots, \mathbf{x}_{i-1}^{k+1}, \mathbf{x}_i, \mathbf{x}_{i+1}^k, \ldots, \mathbf{x}_n^k, \mathbf{Y}^k)$

end do

2. $\mathbf{Y}^{k+1} = \mathbf{Y}^k + \mu h(\mathbf{X}^{k+1})$

end while

See for instance tutorial by S. Boyd (2011)

Prototype ADMM:

$$\min f(\mathbf{x}_1,\ldots,\mathbf{x}_n) \text{ subject to: } h(\mathbf{x}_1,\ldots,\mathbf{x}_n) = 0$$

$$\min L(\mathbf{x},\mathbf{Y},\mu) \doteq f(\mathbf{x}_1,\ldots,\mathbf{x}_n) + \langle \mathbf{Y}, h(\mathbf{x}_1,\ldots,\mathbf{x}_n)\rangle + \frac{\mu}{2}\|h(\mathbf{x}_1,\ldots,\mathbf{x}_n)\|_F^2$$

while not converged **do**

1. **For i=1,...,n**

$\mathbf{x}_i^{k+1} = argmin_{\mathbf{x}_i} L(\mathbf{x}_1^{k+1},\ldots,\mathbf{x}_{i-1}^{k+1},\mathbf{x}_i,\ldots$

end do

2. $\mathbf{Y}^{k+1} = \mathbf{Y}^k + \mu h(\mathbf{X}^{k+1})$

end while

Only as efficient as this step

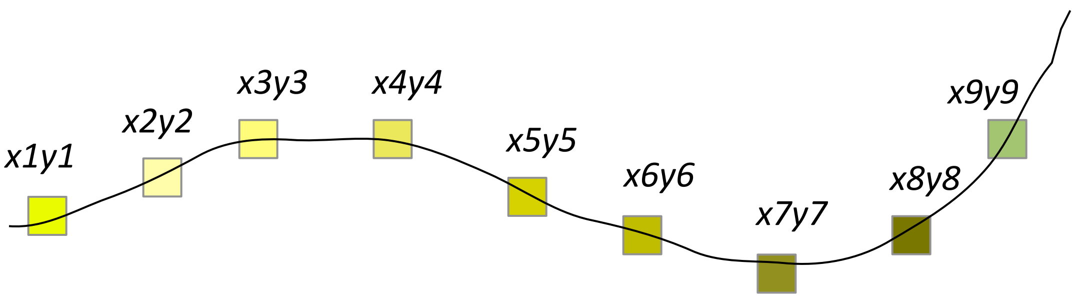
ADMM methods for SRPCA problems:

- **For problems of the form:**

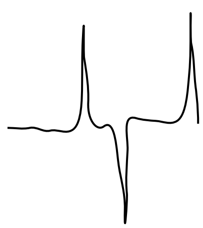
$$\begin{aligned}
&\min \|\mathbf{S}(\mathbf{x})\|_{\mathbf{W},*} + \lambda_1 \|\mathbf{e}_1\|_1 + \tfrac{\lambda_2}{2} \|\mathbf{e}_2\|_2^2 \\
&\text{subject to:} \\
&\mathbf{d} - \mathbf{F}\mathbf{x} = 0 \\
&\mathbf{S}_1\mathbf{x} - \mathbf{e}_1 = 0 \\
&\mathbf{S}_2\mathbf{x} - \mathbf{e}_2 = 0 \\
&\mathbf{S}_{\mathbf{Q}}(\mathbf{x}) \succeq 0, \qquad \mathbf{S}, \mathbf{S}_{\mathbf{Q}} \text{ affine structural constraints}
\end{aligned}$$

  - **Closed form solutions at each step**
  - **Many cheap iterations (cost of a partial SVD)**

Application: – Outlier Removal



+

outliers 

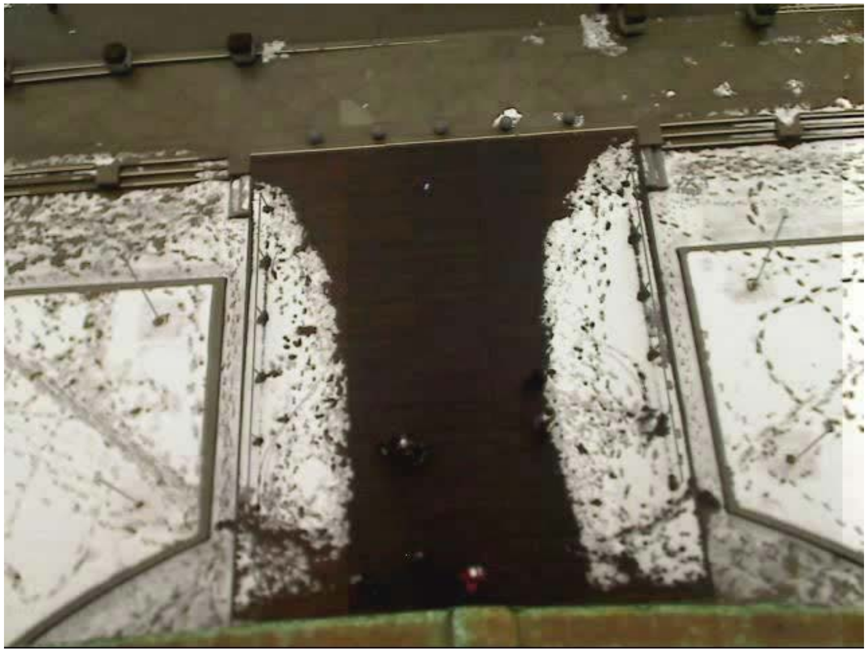
x1 y1	x2 y2	x3 y3	x4 y4	x5 y5
x2 y2	x3 y3	x4 y4	x5 y5	x6 y6
x3 y3	x4 y4	x5 y5	x6 y6	x7 y7
x4 y4	x5 y5	x6 y6	x7 y7	x8 y8
x5 y5	x6 y6	x7 y7	x8 y8	x9 y9

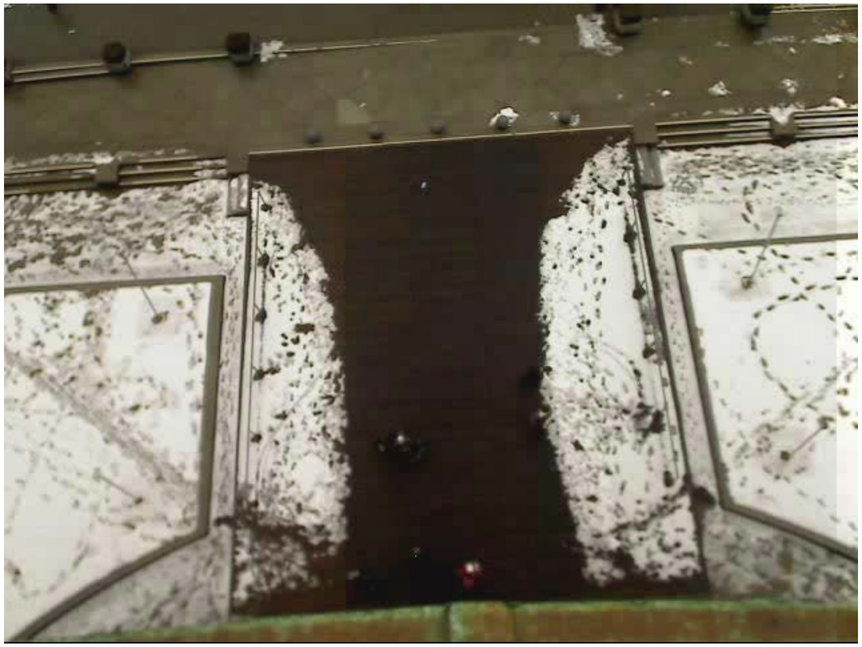
Find and remove outliers

Decompose $H = H_L + H_E$

H_L: low rank
H_E: sparse

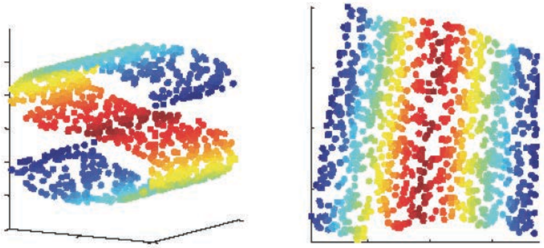
Application: – Outlier Removal





Speed: ADMM: 25 secs *Int. Point: Out of Mem.*

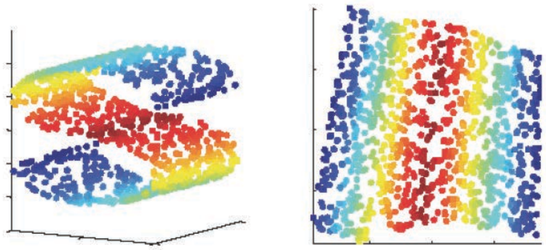
Nonlinear Dimensionality Reduction



Classical dimensionality reduction methods:

- **Use spatial correlations to project to a lower dimensional manifold**
  - Linear (PCA, SVD)
  - Non linear:
    - Locally Linear Embedddings
    - Hessian Eigenmaps
    - Maximum Variance Unfolding
    - Semi Definite Embeddings

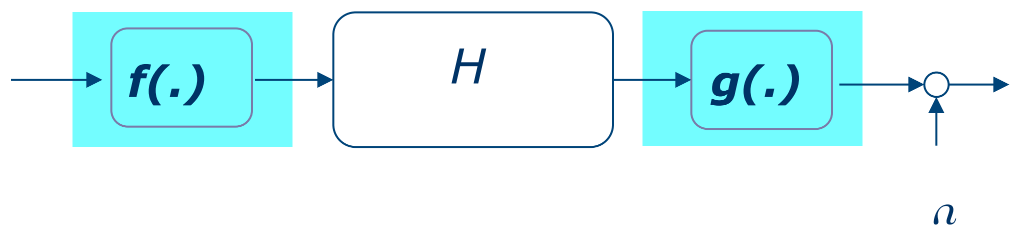
    ⋮



- **Typically these methods do not exploit temporal correlations**

Dimensionality reduction as an Id problem:

- **Map to/from manifold: a memoryless non-linearity**

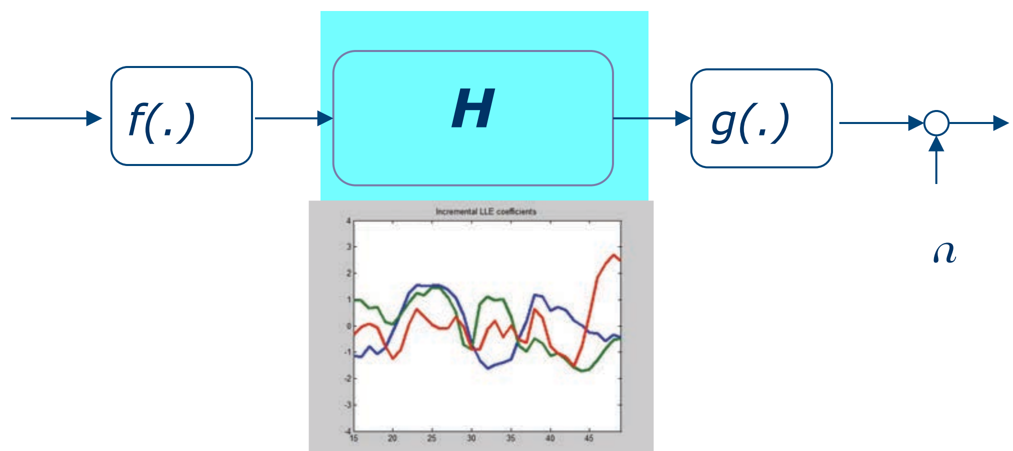


projections

Dimensionality reduction as an Id problem:

- **Map to/from manifold: a memoryless non-linearity**
- **Manifold dynamics: piece-wise linear:**

- **A switched Hammerstein/Wiener SysId problem:**



Dynamics on the manifold

A SRPCA Formulation:

$$\min \quad rank(\mathbf{G}) - \lambda trace(\mathbf{K})$$

$$\text{s.t.} \quad \mathbf{K}_{ii} + \mathbf{K}_{jj} + 2\mathbf{K}_{ij} = (1+\varepsilon) \| x_i - x_j \|_2^2 \text{ ,if } \boldsymbol{\eta}_{ij} = 1$$

$$\mathbf{K} \geq 0$$

$$\sum_{i,j} \mathbf{K}_{ij} = 0$$

where

$$\mathbf{G} = \mathbf{H}_Y^T \mathbf{H}_Y$$

$$\mathbf{K}_{i,n-1} = \begin{bmatrix} {y_i}^T y_i & {y_i}^T y_{i+1} & \cdots & {y_i}^T y_{i+n-1} \\ {y_{i+1}}^T y_i & {y_{i+1}}^T y_{i+1} & \cdots & {y_{i+1}}^T y_i \\ \vdots & \vdots & \ddots & \vdots \\ {y_{i+n-1}}^T y_i & {y_{i+n-1}}^T y_{i+1} & \cdots & {y_{i+n-1}}^T y_{i+n-1} \end{bmatrix}$$

A SRPCA Formulation:

$$\min \quad rank(\mathbf{G}) - \lambda trace(\mathbf{K})$$

$$\text{s.t.} \quad \mathbf{K}_{ii} + \mathbf{K}_{jj} + 2\mathbf{K}_{ij} = (1+\varepsilon)\,\|\boxed{x_i - x_j}\|_2^2 \text{ ,if } \boldsymbol{\eta}_{ij} = 1$$

$$\mathbf{K} \geq 0$$

$$\sum_{i,j} \mathbf{K}_{ij} = 0$$

High dimensional data (given)

where

$$\mathbf{G} = \mathbf{H}_Y^T \mathbf{H}_Y$$

Manifold data (unknown)

$$\mathbf{K}_{i,n-1} = \begin{bmatrix} \boxed{y_i{}^T y_i} & y_i{}^T y_{i+1} & \cdots & y_i{}^T y_{i+n-1} \\ y_{i+1}{}^T y_i & y_{i+1}{}^T y_{i+1} & \cdots & y_{i+1}{}^T y_i \\ \vdots & \vdots & \ddots & \vdots \\ y_{i+n-1}{}^T y_i & y_{i+n-1}{}^T y_{i+1} & \cdots & y_{i+n-1}{}^T y_{i+n-1} \end{bmatrix}$$

A SRPCA Formulation:

$$
\begin{aligned}
\min \quad & rank(\mathbf{G}) - \lambda \boxed{trace(\mathbf{K})} \\
\text{s.t.} \quad & \boxed{\mathbf{K}_{ii} + \mathbf{K}_{jj} + 2\mathbf{K}_{ij} = (1+\varepsilon)\| x_i - x_j \|_2^2}, \text{if } \boldsymbol{\eta}_{ij} = 1 \\
& \mathbf{K} \geq 0 \\
& \sum_{i,j} \mathbf{K}_{ij} = 0
\end{aligned}
$$

Spatial information

where

$$\mathbf{G} = \mathbf{H}_Y^T \mathbf{H}_Y$$

$$
\mathbf{K}_{i,n-1} = \begin{bmatrix}
{y_i}^T y_i & {y_i}^T y_{i+1} & \cdots & {y_i}^T y_{i+n-1} \\
{y_{i+1}}^T y_i & {y_{i+1}}^T y_{i+1} & \cdots & {y_{i+1}}^T y_i \\
\vdots & \vdots & \ddots & \vdots \\
{y_{i+n-1}}^T y_i & {y_{i+n-1}}^T y_{i+1} & \cdots & {y_{i+n-1}}^T y_{i+n-1}
\end{bmatrix}
$$

A SRPCA Formulation:

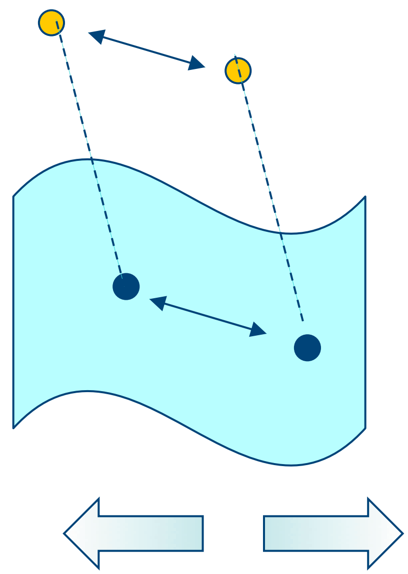
$$\min \quad rank(\mathbf{G}) - \lambda trace(\mathbf{K})$$

$$\text{s.t.} \quad \boxed{\mathbf{K}_{ii} + \mathbf{K}_{jj} + 2\mathbf{K}_{ij} = (1+\varepsilon)\| x_i - x_j \|_2^2} \text{, if } \boldsymbol{\eta}_{ij} = 1$$

$$\mathbf{K} \geq 0$$

$$\sum_{i,j} \mathbf{K}_{ij} = 0$$

Spatial information



A SRPCA Formulation:

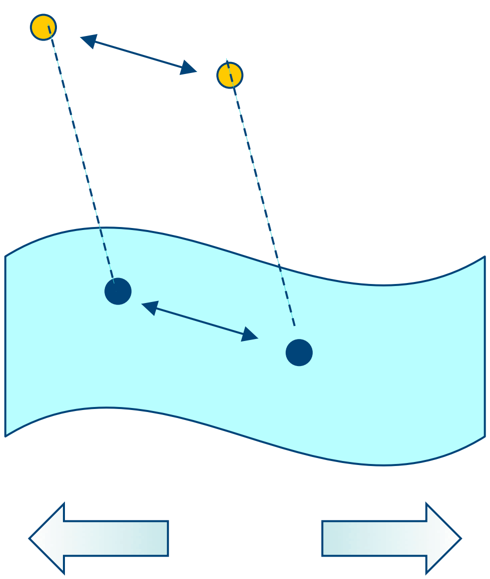
$$\min \quad rank(\mathbf{G}) - \lambda \boxed{trace(\mathbf{K})}$$

$$\text{s.t.} \quad \mathbf{K}_{ii} + \mathbf{K}_{jj} + 2\mathbf{K}_{ij} = (1+\varepsilon) \| x_i - x_j \|_2^2 \text{, if } \boldsymbol{\eta}_{ij} = 1$$

$$\mathbf{K} \geq 0$$

$$\sum_{i,j} \mathbf{K}_{ij} = 0$$

Spatial information



A SRPCA Formulation:

$$
\begin{aligned}
\min \quad & \boxed{rank(\mathbf{G})} - \lambda trace(\mathbf{K}) \\
\text{s.t.} \quad & \mathbf{K}_{ii} + \mathbf{K}_{jj} + 2\mathbf{K}_{ij} = (1+\varepsilon) \| x_i - x_j \|_2^2 \text{ ,if } \boldsymbol{\eta}_{ij} = 1 \\
& \mathbf{K} \geq 0 \\
& \sum_{i,j} \mathbf{K}_{ij} = 0
\end{aligned}
$$

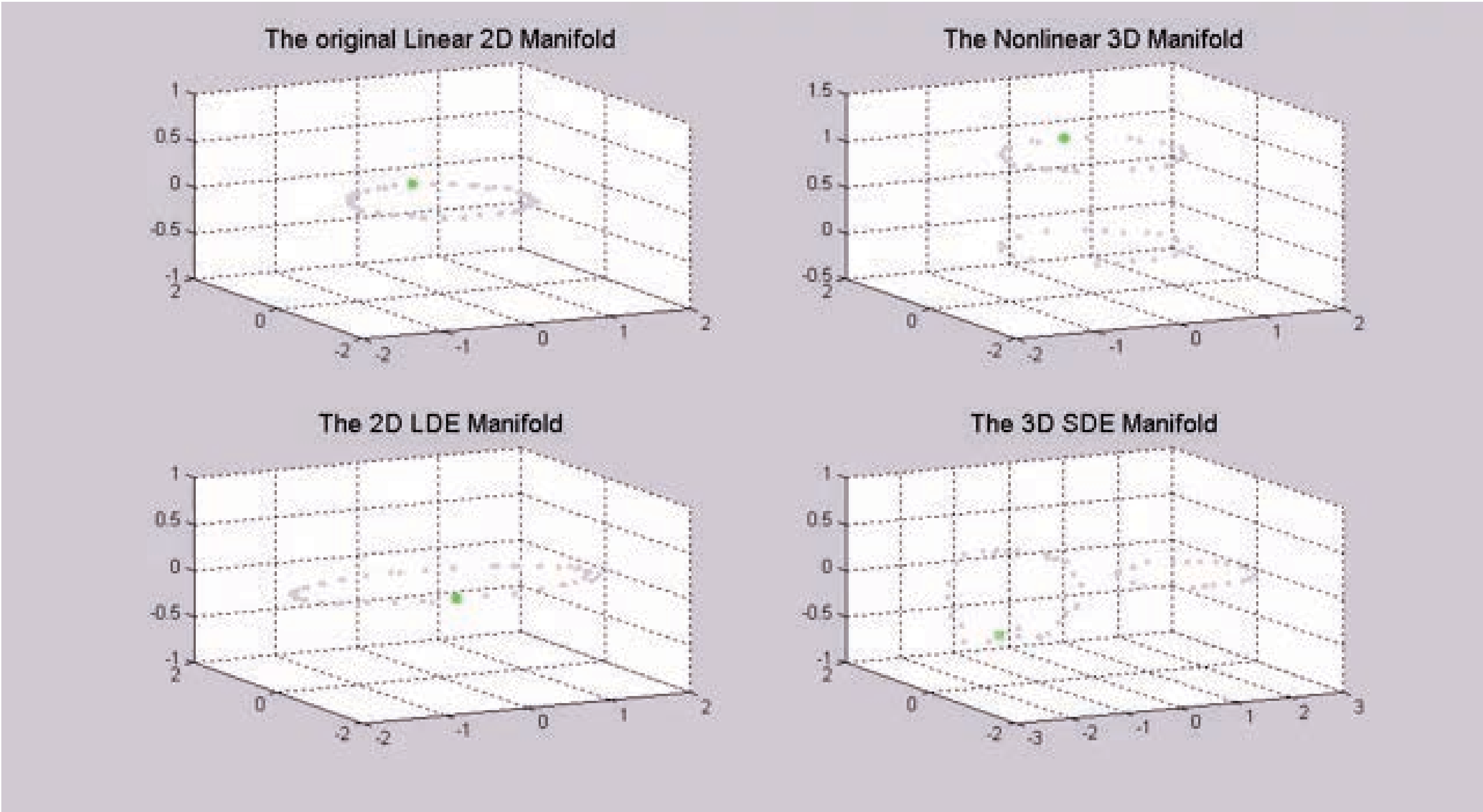
Low order dynamics

where

$$\boxed{\mathbf{G} = \mathbf{H}_Y^T \mathbf{H}_Y}$$

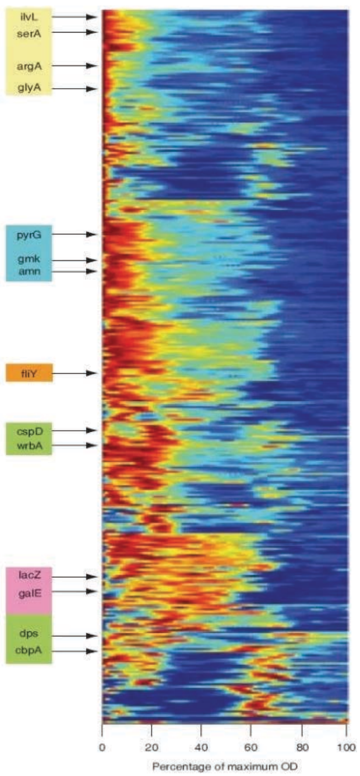
$$
\mathbf{K}_{i,n-1} = \begin{bmatrix}
y_i{}^T y_i & y_i{}^T y_{i+1} & \cdots & y_i{}^T y_{i+n-1} \\
y_{i+1}{}^T y_i & y_{i+1}{}^T y_{i+1} & \cdots & y_{i+1}{}^T y_i \\
\vdots & \vdots & \ddots & \vdots \\
y_{i+n-1}{}^T y_i & y_{i+n-1}{}^T y_{i+1} & \cdots & y_{i+n-1}{}^T y_{i+n-1}
\end{bmatrix}
$$

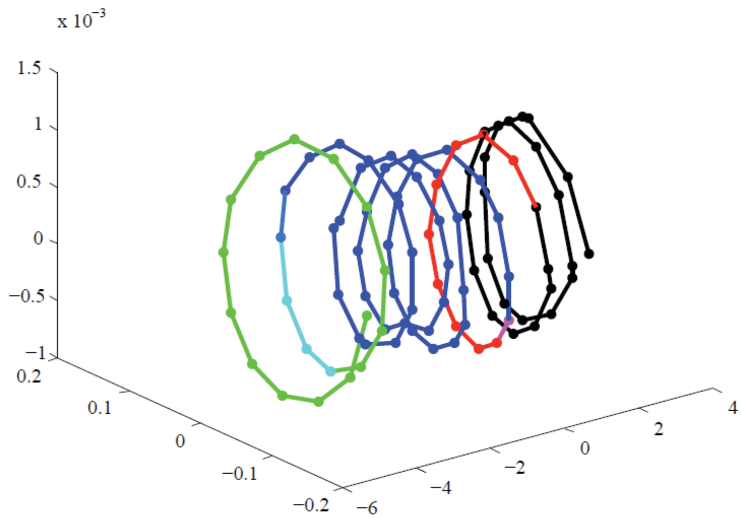
Academic Example:



Dynamic **Static**

Application: diauxic shift analysis

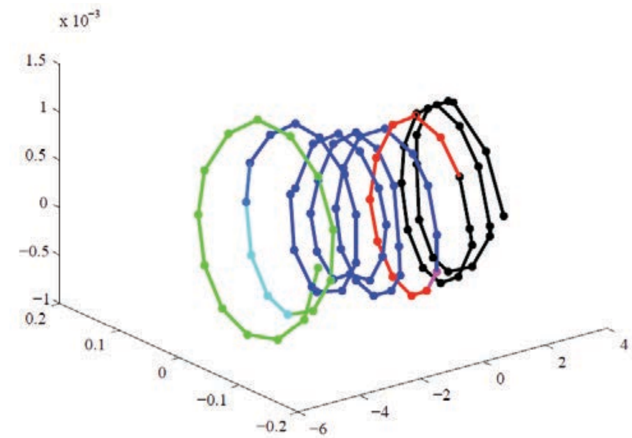




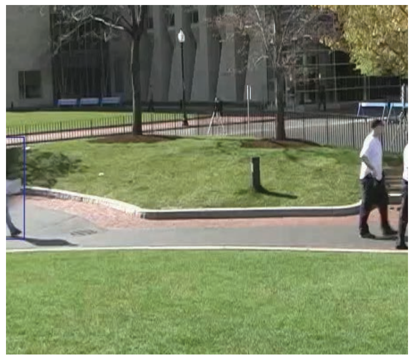
Manifold evolution

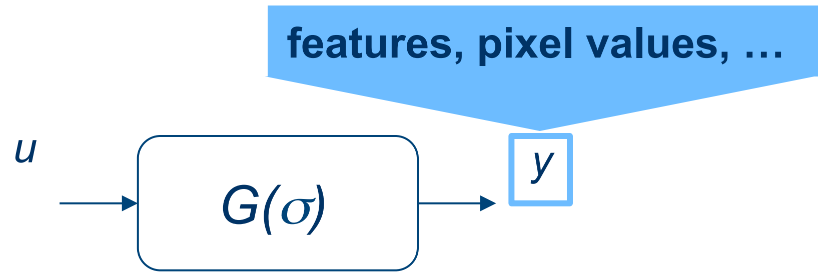
Original data: 2000 promoters

Information Extraction

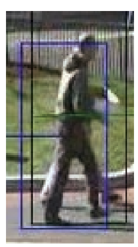


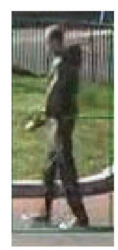
Information extraction as an Id problem:



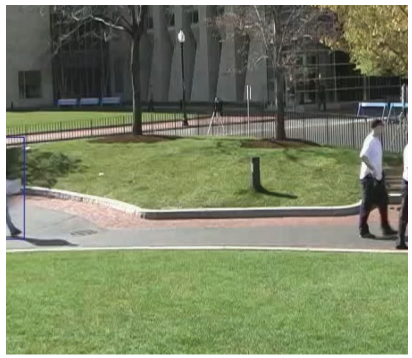


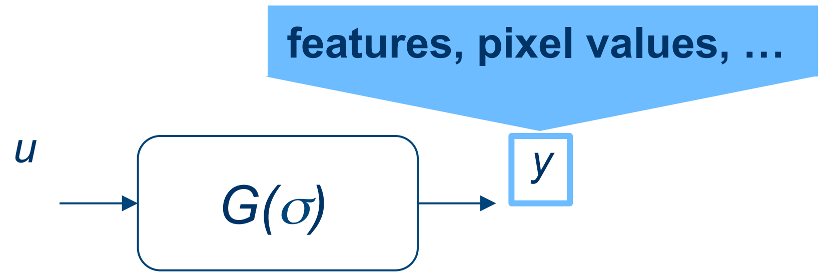
- **Model data streams as outputs of switched systems**
- **"Interesting" events ⇔ Model invariant(s) changes**
- **"Homogeneous" segments ⇔ output of a single sub-system**



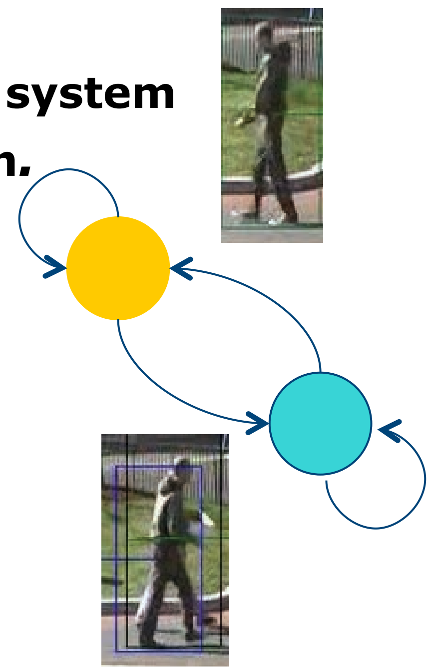


Information extraction as an Id problem:

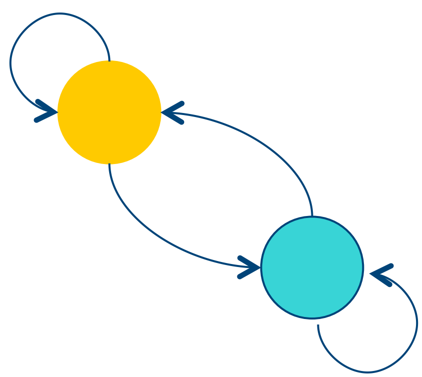




- **Equivalent to detecting changes in a switched system**
- **An identification/model (in)validation problem.**

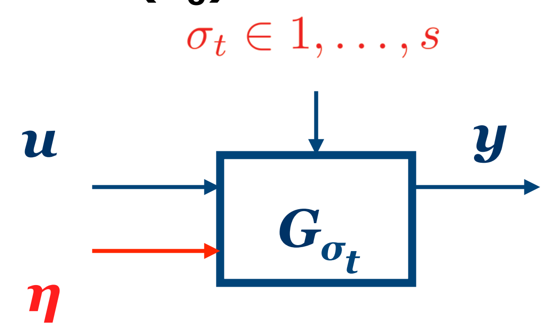


Identifying Switched ARX Models



SARX Id problem:

- **Given:**
  - **Bounds on noise ($\|\eta\|_\infty \leq \varepsilon$), sub-system order ($n_o$)**
  - **Input/output data (u,y)**
  - **Number of sub-models**

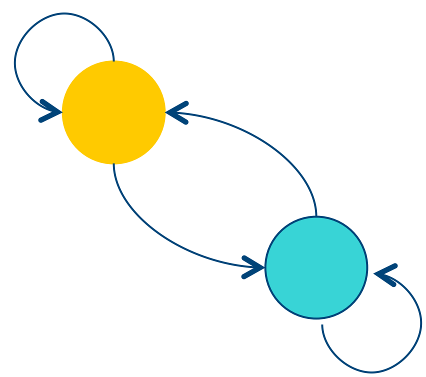


- **Find:**
  - **A piecewise affine model such that:**

$$y_t = \sum_{i=1}^{n_a} a_i(\sigma_t) y_{t-i} + \sum_{i=1}^{n_c} c_i(\sigma_t) u_{t-i} + f(\sigma_t) + \eta_t$$

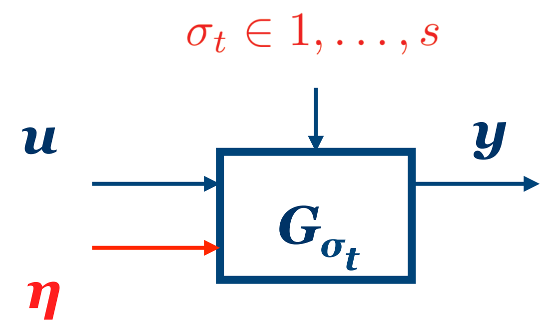
$$\Updownarrow$$

$$0 = \mathbf{b}(\sigma_t)^T \mathbf{r}_t + \eta_t$$



SARX Id problem:

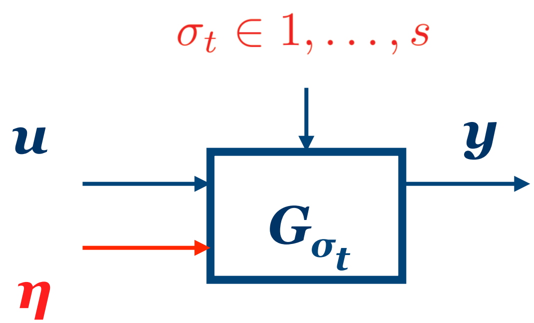
- **Problem is (generically) NP hard:**
- **Solutions based on:**
  - **Heuristics:**
  - **Optimization**
  - **Probabilistic priors**
  - **Convex Relaxations**



Vidal, Chiuso, Roll, Bemporad, Paoletti, Garulli, Vicino, Juloski, Ferrari-Trecate, Ozay, Bako, and many others

SARX Id problem:

- **Problem is (generically) NP hard:**
- **Solutions based on:**
  - **Heuristics:**
  - **Optimization**
  - **Probabilistic priors**
  - **Convex Relaxations**

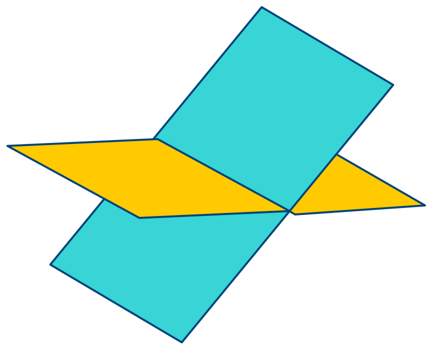


Vidal, Chiuso, Roll, Bemporad, Paoletti, Garulli, Vicino, Juloski, Ferrari-Trecate, Ozay, Bako, and many others

SARX Id problem in the noise free case:

- **GPCA: an algebraic geometric method due to Vidal *et al.***
- **Main Idea:**

$$\mathrm{b}(\sigma_t)^T \mathbf{r}_t = 0, \quad \sigma_t \in \{1, \ldots, s\}$$



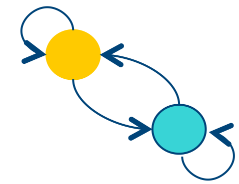
arrangement of subspaces

$$p_s(\mathbf{r}) = \prod_{i=1}^{s} (\mathrm{b_i}^T \mathbf{r}_t) = 0$$

vanishing ideal

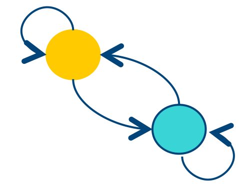
Toy example: 2 first order systems:

$$y_t = a(\sigma_t)y_{t-1} + b(\sigma_t)u_{t-1},\ \sigma_t \in \{1,2\}$$



Toy example: 2 first order systems:

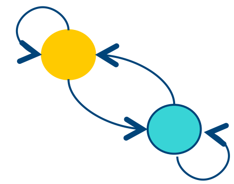
$$y_t = a(\sigma_t)y_{t-1} + b(\sigma_t)u_{t-1},\ \sigma_t \in \{1,2\}$$



$$[y_t - a(\sigma_1)y_{t-1} + b(\sigma_1)u_{t-1}]\,[y_t - a(\sigma_2)y_{t-1} + b(\sigma_2)u_{t-1}] = 0$$

Toy example: 2 first order systems:

$$y_t = a(\sigma_t)y_{t-1} + b(\sigma_t)u_{t-1},\ \sigma_t \in \{1,2\}$$



$$[y_t - a(\sigma_1)y_{t-1} + b(\sigma_1)u_{t-1}]\,[y_t - a(\sigma_2)y_{t-1} + b(\sigma_2)u_{t-1}] = 0$$

$$\underbrace{\begin{bmatrix} y_t^2 & -y_t y_{t-1} & -y_t u_{t-1} & y_{t-1}^2 & y_{t-1}u_{t-1} \end{bmatrix}}_{\nu_s(\mathbf{r}_t)} \!\!\begin{matrix} u_{t-1}^2 \,] \end{matrix} \underbrace{\begin{bmatrix} 1 \\ -(a_1+a_2) \\ -(b_1+b_2) \\ a_1a_2 \\ a_1b_2+a_2b_1 \\ b_1b_2 \end{bmatrix}}_{c_s} = 0$$

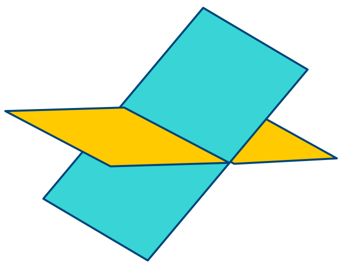
Function of the data only

System parameters
Independent of the data

One such equation per data point

SARX Id problem in the noise free case:

- **GPCA: an algebraic geometric method due to Vidal *et al.***
- **Main Idea:**



$$\mathrm{b}(\sigma_t)^T \mathbf{r}_t = 0, \quad \sigma_t \in \{1, \ldots, s\}$$

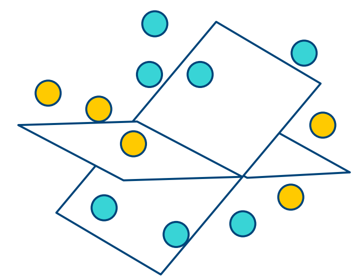
$$p_s(\mathbf{r}) = \mathbf{c_s}^T \nu_s(\mathbf{r}_t) = 0$$

$$\mathbf{V_s c_s} \doteq \begin{bmatrix} \nu_s(\mathbf{r}_{t_0})^T \\ \vdots \\ \nu_s(\mathbf{r}_T)^T \end{bmatrix} \mathbf{c_s} = \mathbf{0}$$

- **Solve for c_s from the null space of the embedded data matrix.**
- **Get b_i from c_s via polynomial differentiation**

Details in Vidal et al., 2003

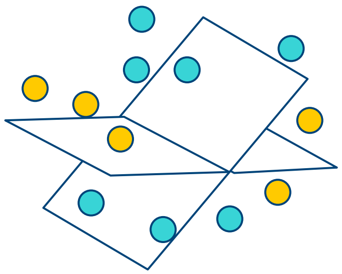
What happens with noisy measurements?



$$\mathrm{b}(\sigma_t)^T \mathbf{r}_t = \boldsymbol{\eta}_t, \quad \sigma_t \in \{1, \ldots, s\}$$

$$p_s(\mathbf{r}) = \mathbf{c_s}^T \nu_s(\mathbf{r}_t, \boldsymbol{\eta}_t) = 0 \quad \rightarrow \quad \mathbf{V_s c_s} \doteq \begin{bmatrix} \nu_s(\mathbf{r}_{t_0}, \boldsymbol{\eta}_{t_o})^T \\ \vdots \\ \nu_s(\mathbf{r}_T, \boldsymbol{\eta}_T)^T \end{bmatrix} \mathbf{c_s} = \mathbf{0}$$

What happens with noisy measurements?



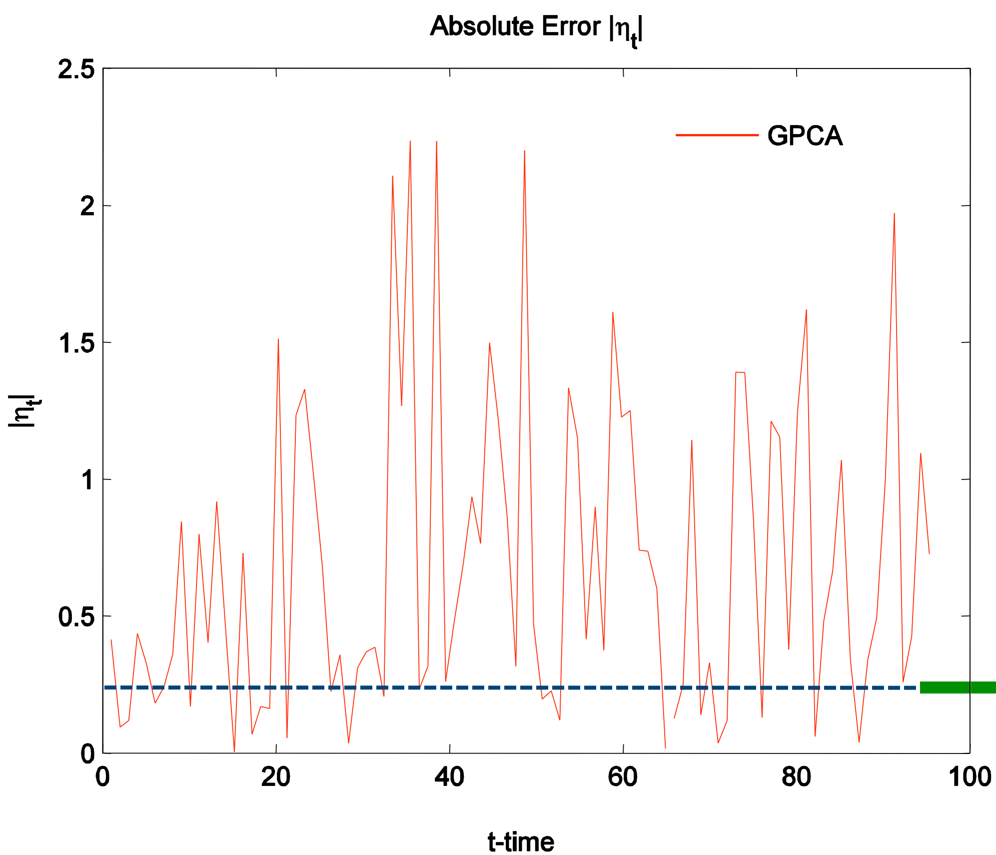
$$\mathrm{b}(\sigma_t)^T \mathbf{r}_t = \boldsymbol{\eta_t}, \quad \sigma_t \in \{1, \ldots, s\}$$

$$p_s(\mathbf{r}) = \mathbf{c_s}^T \nu_s(\mathbf{r}_t, \boldsymbol{\eta_t}) = 0$$

$$\mathbf{V_s c_s} \doteq \begin{bmatrix} \nu_s(\mathbf{r}_{t_0}, \boldsymbol{\eta_{t_o}})^T \\ \vdots \\ \nu_s(\mathbf{r}_T, \boldsymbol{\eta_T})^T \end{bmatrix} \mathbf{c_s} = \mathbf{0}$$

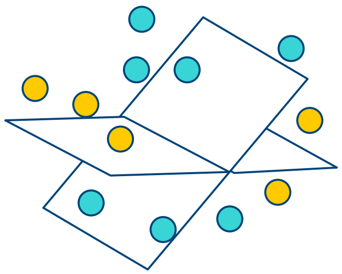
Need to find the null space of a noisy matrix: Obvious approach: SVD

Academic Example



Noise bound: 0.25

What happens with noisy measurements?



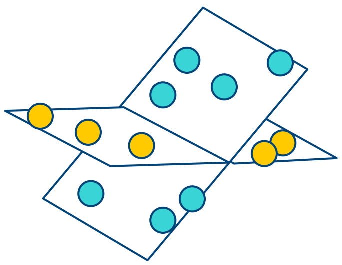
$$\mathrm{b}(\sigma_t)^T \mathbf{r}_t = \boldsymbol{\eta}_t, \quad \sigma_t \in \{1, \ldots, s\}$$

$$p_s(\mathbf{r}) = \mathbf{c_s}^T \nu_s(\mathbf{r}_t, \boldsymbol{\eta}_t) = 0 \quad \rightarrow \quad \mathbf{V_s c_s} \doteq \begin{bmatrix} \nu_s(\mathbf{r}_{t_0}, \boldsymbol{\eta}_{t_o})^T \\ \vdots \\ \nu_s(\mathbf{r}_T, \boldsymbol{\eta}_T)^T \end{bmatrix} \mathbf{c_s} = \mathbf{0}$$

Need to find the null space of a noisy matrix: Minimize rank V_s w.r.t η_t

What happens with noisy measurements?

$$\mathrm{b}(\sigma_t)^T \mathbf{r}_t = \boldsymbol{\eta}_t, \quad \sigma_t \in \{1, \ldots, s\}$$



$$p_s(\mathbf{r}) = \mathbf{c_s}^T \nu_s(\mathbf{r}_t, \boldsymbol{\eta}_t) = 0 \quad \rightarrow \quad \mathbf{V_s c_s} \doteq \begin{bmatrix} \nu_s(\mathbf{r}_{t_0}, \boldsymbol{\eta}_{t_o})^T \\ \vdots \\ \nu_s(\mathbf{r}_T, \boldsymbol{\eta}_T)^T \end{bmatrix} \mathbf{c_s} = \mathbf{0}$$

Need to find the null space of a noisy matrix: Minimize rank V_s w.r.t η_t

Detour: Polynomial Optimization

From Lasserre 01:

$$
\begin{aligned}
p_1^* &= \min_{x \in K} p(x) = \min_{x \in K} \sum a_i x^i \\
p_2^* &= \min_{\mu \text{ supp in} K} E_\mu(p) = \min_{\mu} \int_\mu (\sum a_i x^i) d\mu
\end{aligned}
$$

Theorem: **$p_1^* = p_2^*$**

Detour: Polynomial Optimization

From Lasserre 01:

$$p_1^* \quad = \quad \min_{x \in K} p(x) = \min_{x \in K} \sum a_i x^i$$

$$p_2^* \quad = \quad \min_{\mu \text{ supp in} K} E_\mu(p) = \min_\mu \int_\mu (\sum a_i x^i) d\mu$$

$$= \min \sum a_i m_i$$

subject to:

$$m_i = E_\mu(x^i)$$

Detour: Polynomial Optimization

From Lasserre 01:

$$p_1^* \quad = \quad \min_{x \in K} p(x) = \min_{x \in K} \sum a_i x^i$$

$$p_2^* \quad = \quad \min_{\mu \text{ supp in} K} E_\mu(p) = \min_\mu \int_\mu$$

Affine in m_i

$$= \min \sum a_i m_i$$

subject to:

$$m_i = E_\mu(x^i)$$

Detour: Polynomial Optimization

From Lasserre 01:

$$p_1^* = \min_{x \in K} p(x) = \min_{x \in K} \sum a_i x^i$$

$$p_2^* = \min_{\mu \text{ supp in} K} E_\mu(p) = \min_{\mu} \int_\mu \left(\sum a_i x^i\right) d\mu$$

$$= \min \sum a_i m_i$$

subject to:

$$\boxed{m_i = E_\mu(x^i)}$$

Hausdorff, Hamburger moments problem.

Set of LMIs.

What happens with noisy measurements?

Optimization Problem 1:

$$\begin{array}{ll} \text{minimize}_{\eta_t} & \text{rank}\mathbf{V}_{\mathbf{s}}(\mathbf{r}_t, \eta_t) \\ \text{subject to} & \|\eta_t\|_\infty \leq \epsilon \end{array}$$

- **Rank is not a polynomial function. Can we use ideas from polynomial optimization?**

What happens with noisy measurements?

Optimization Problem 1:

$$\begin{array}{ll} \text{minimize}_{\eta_t} & \text{rank}\mathbf{V}_{\mathbf{s}}(\mathbf{r}_t, \eta_t) \\ \text{subject to} & \|\eta_t\|_\infty \leq \epsilon \end{array}$$

$$m_i^{(t)} = \int_\mu \eta_t^i d\mu$$

Optimization Problem 2:

$$\begin{array}{ll} \text{minimize}_{\mathbf{m}^{(t)}} & \text{rank}\tilde{\mathbf{V}}_{\mathbf{s}}(\mathbf{r}_t, \mathbf{m}^{(t)}) \\ \text{subject to} & \text{each } \mathbf{m}^{(t)} \text{ is a} \\ & \text{moment sequence} \end{array}$$

What happens with noisy measurements?

Optimization Problem 1:

$$\begin{array}{ll} \text{minimize}_{\eta_t} & \text{rank}\,\mathbf{V}_{\mathbf{s}}(\mathbf{r}_t, \eta_t) \\ \text{subject to} & \|\eta_t\|_\infty \leq \epsilon \end{array}$$

$$m_i^{(t)} = \int_\mu \eta_t^i d\mu$$

Optimization Problem 2:

$$\begin{array}{ll} \text{minimize}_{\mathbf{m}^{(t)}} & \text{rank}\,\tilde{\mathbf{V}}_{\mathbf{s}}(\mathbf{r}_t, \mathbf{m}^{(t)}) \\ \text{subject to} & \boxed{\text{each } \mathbf{m}^{(t)} \text{ is a moment sequence}} \end{array}$$

Convex constraint set!!

What happens with noisy measurements?

Optimization Problem 1:

$$\begin{array}{ll} \text{minimize}_{\eta_t} & \text{rank}\,\mathbf{V_s}(\mathbf{r}_t, \eta_t) \\ \text{subject to} & \|\eta_t\|_\infty \leq \epsilon \end{array}$$

Optimization Problem 2:

$$\begin{array}{ll} \text{minimize}_{\mathbf{m}^{(t)}} & \text{rank}\,\tilde{\mathbf{V}}_\mathbf{s}(\mathbf{r}_t, \mathbf{m}^{(t)}) \\ \text{subject to} & \boxed{\begin{array}{l}\text{each } \mathbf{m}^{(t)} \text{ is a} \\ \text{moment sequence}\end{array}} \end{array}$$

Convex constraint set!!

- **Matrix rank minimization**
- **Subject to LMI constraints**

- **(approx) solvable using a convex relaxation (e.g. log-det heuristic of Fazel *et al.*).**

What happens with noisy measurements?

Optimization Problem 1:

$$\begin{array}{ll} \text{minimize}_{\eta_t} & \text{rank}\,\mathbf{V_s}(\mathbf{r}_t, \eta_t) \\ \text{subject to} & \|\eta_t\|_\infty \leq \epsilon \end{array}$$

Optimization Problem 2:

$$\begin{array}{ll} \text{minimize}_{\mathbf{m}^{(t)}} & \text{rank}\,\tilde{\mathbf{V}}_\mathbf{s}(\mathbf{r}_t, \mathbf{m}^{(t)}) \\ \text{subject to} & \boxed{\begin{array}{l}\text{each } \mathbf{m}^{(t)} \text{ is a} \\ \text{moment sequence}\end{array}} \end{array}$$

Convex constraint set!!

- **Fact:**
  - **There exists a rank deficient solution for Problem 2 if and only if there exists a rank deficient solution for Problem 1.**
  - **If c belongs to the nullspace of the solution of Problem 2, there exists a noise value η_t with $\|\eta_t\|_\infty \leq \epsilon$ such that c belongs to the nullspace of $\mathbf{V_s}(\mathbf{r_t}, \mathbf{m^{(t)}})$**

What happens with noisy measurements?

Optimization Problem 1:

$$\begin{array}{ll} \text{minimize}_{\eta_t} & \text{rank}\,\mathbf{V}_\mathbf{s}(\mathbf{r}_t, \eta_t) \\ \text{subject to} & \|\eta_t\|_\infty \leq \epsilon \end{array}$$

Optimization Problem 2:

$$\begin{array}{ll} \text{minimize}_{\mathbf{m}^{(t)}} & \text{rank}\,\tilde{\mathbf{V}}_\mathbf{s}(\mathbf{r}_t, \mathbf{m}^{(t)}) \\ \text{subject to} & \text{each } \mathbf{m}^{(t)} \text{ is a} \\ & \text{moment sequence} \end{array}$$

- **Rank is not a polynomial function. Can we use ideas from polynomial optimization?**
  - YES

- **Use a convex relaxation (e.g. log-det heuristic of Fazel *et al.*) to solve Problem 2**
- **Find a vector c in the nullspace**
- **Estimate noise by root finding ($\mathbf{V}_\mathbf{s}\mathbf{c} = 0$ polynomials of one variable)**
- **Proceed as in noise-free case**

What happens with noisy measurements?

Optimization Problem 1:

$$\begin{array}{ll} \text{minimize}_{\eta_t} & \text{rank}\mathbf{V}_\mathbf{s}(\mathbf{r}_t, \eta_t) \\ \text{subject to} & \|\eta_t\|_\infty \leq \epsilon \end{array}$$

Optimization Problem 2:

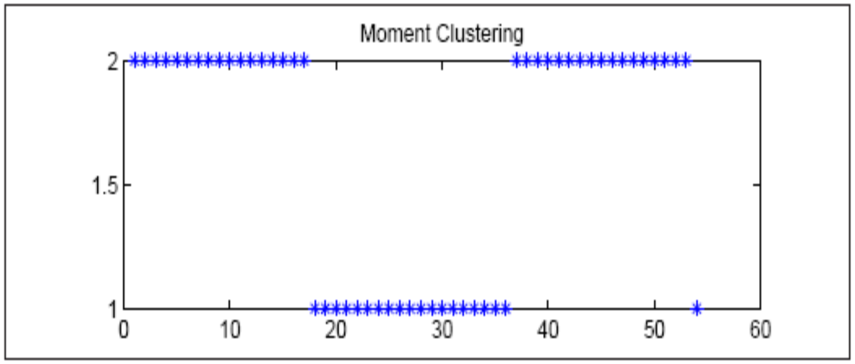
$$\begin{array}{ll} \text{minimize}_{\mathbf{m}^{(t)}} & \text{rank}\tilde{\mathbf{V}}_\mathbf{s}(\mathbf{r}_t, \mathbf{m}^{(t)}) \\ \text{subject to} & \text{each } \mathbf{m}^{(t)} \text{ is a} \\ & \text{moment sequence} \end{array}$$

- **Rank is not a polynomial function. Can we use ideas from polynomial optimization?**
  - **YES**

- **Use a convex relaxation (e.g. log-det heuristic of Fazel *et al.*) to solve Problem 2**
- **Find a vector c in the nullspace**
- **Estimate noise by root finding ($V_s c = 0$ polynomials of one variable)**
- **Proceed as in noise-free case**

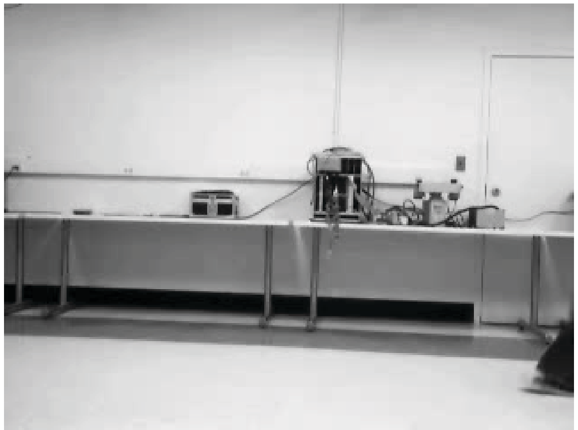
Provably convergent as the information is completed

Example: Human Activity Analysis



WALK ***BEND*** ***WALK***

Handling outliers

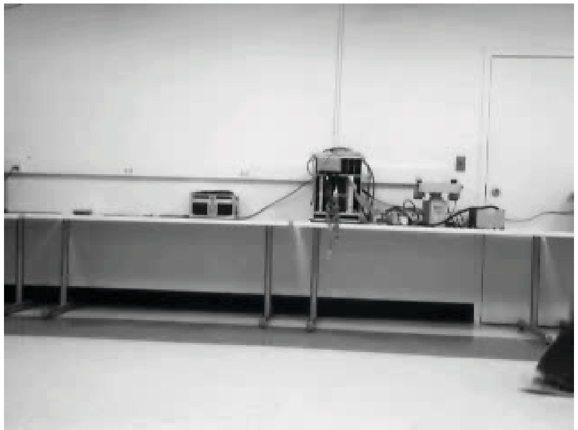


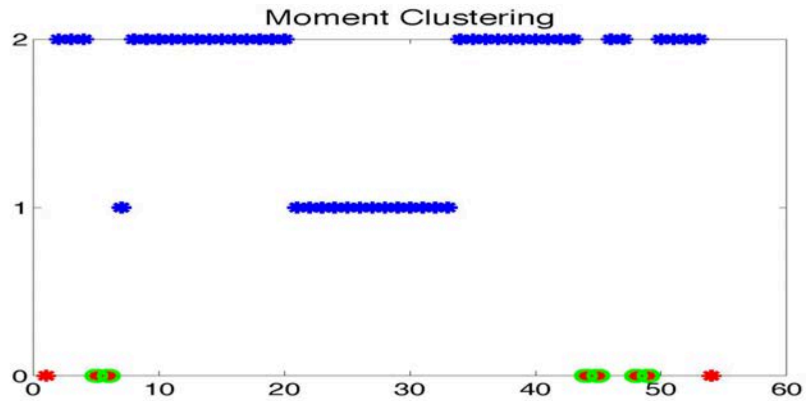
$$\min \left\{ \mathbf{rank} \left[\tilde{\mathbf{V}}(\mathbf{r},\mathbf{m}) + \mathbf{E} \right] + \lambda_1 \|\mathbf{E}\|_{0,row} \right\}$$

subject to

$$\mathbf{M}(\mathbf{m}) \succeq \mathbf{0},\ \mathbf{L}(\mathbf{m}) \succeq \mathbf{0}$$

Handling outliers

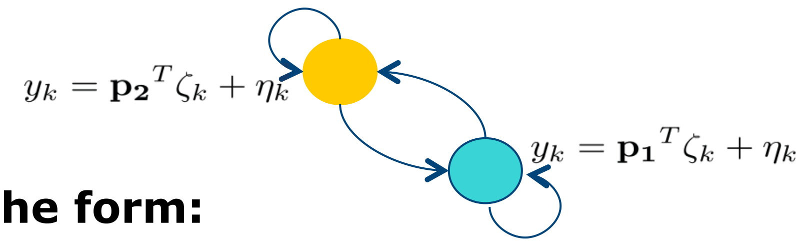




WALK ***BEND*** ***WALK***

(In)Validating SARX Models

Model (In)validation of SARX Systems



- **Given:**
  - **A nominal switched model of the form:**

$$\begin{aligned} \mathbf{y}_t &= \textstyle\sum_{k=1}^{n_a} \mathbf{A}_k(\sigma_t)\mathbf{y}_{t-k} + \sum_{k=1}^{n_c} \mathbf{C}_k(\sigma_t)\mathbf{u}_{t-k} + \mathbf{f}(\sigma_t) \\ \tilde{\mathbf{y}}_t &= \mathbf{y}_t + \boldsymbol{\eta}_t \end{aligned}$$

  - **A bound on the noise ($\|\eta\|_\infty \leq \varepsilon$)**
  - **Experimental Input/Output Data** $\{\mathbf{u}_t, \tilde{\mathbf{y}}_t\}_{t=t_0}^{T}$

- **Determine:**
  - **whether there exist noise and switching sequences**
  - **consistent with a priori information and experimental data**

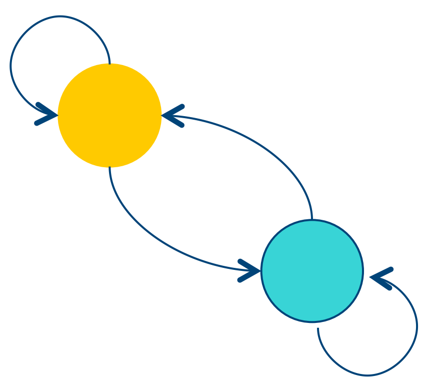
Equivalent to checking emptyness of a semialgebraic set

Reduces to SPD via moments and duality

$$\mathbf{y}_t + \boldsymbol{\eta}_t - \sum_{i=1}^{n_a} \mathbf{A}_i(\;)(\mathbf{y}_{t-i} + \boldsymbol{\eta}_{t-i}) - \sum_{i=1}^{n_c} \mathbf{C}_i(\;)\mathbf{u}_{t-i} = 0$$

or

$$\mathbf{y}_t + \boldsymbol{\eta}_t - \sum_{i=1}^{n_a} \mathbf{A}_i(\;)(\mathbf{y}_{t-i} + \boldsymbol{\eta}_{t-i}) - \sum_{i=1}^{n_c} \mathbf{C}_i(\;)\mathbf{u}_{t-i} = 0$$



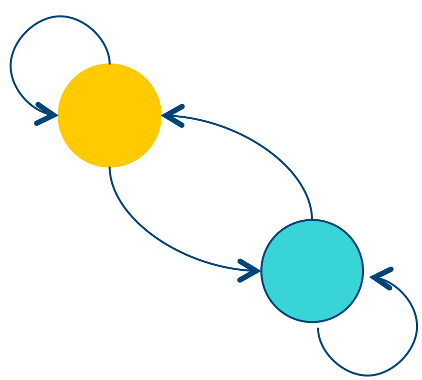
$$\mathbf{s_{1,t}}\left(\mathbf{y}_t + \boldsymbol{\eta}_t - \sum_{i=1}^{n_a} \mathbf{A}_i(\sigma_1)(\mathbf{y}_{t-i} + \boldsymbol{\eta}_{t-i}) - \sum_{i=1}^{n_c} \mathbf{C}_i(\sigma_1)\mathbf{u}_{t-i}\right) = 0$$

and

$$\mathbf{s_{2,t}}\left(\mathbf{y}_t + \boldsymbol{\eta}_t - \sum_{i=1}^{n_a} \mathbf{A}_i(\sigma_2)(\mathbf{y}_{t-i} + \boldsymbol{\eta}_{t-i}) - \sum_{i=1}^{n_c} \mathbf{C}_i(\sigma_2)\mathbf{u}_{t-i}\right) = 0$$

Subject to:

$$s_{i,t} \in \{0, 1\}, \; \sum_i s_{i,t} = 1$$



(In)validation Certificates:

- **The model is invalid if and only if**

$$\mathcal{T}(\mathbf{y}) \doteq \left\{ (s_{i,t}, \eta) : \begin{array}{l} \mathbf{s_{i,t}}(\mathbf{g}_{i,t} + \mathbf{h}_{i,t}\boldsymbol{\eta}_{t-n_a:t}) = 0 \\ \sum_i s_{i,t} = 1 \\ s_{i,t}^2 = 1 \\ \|\boldsymbol{\eta}\|_\infty \leq \epsilon \end{array} \right\} = \emptyset$$

(In)validation Certificates:

- **The model is invalid if and only if**

$$d^* \doteq \left\{ \begin{array}{l} \min_{\mathbf{s},\boldsymbol{\eta}} \sum_{t=1}^{T} \mathbf{s_{i,t}^2}(\|\mathbf{g}_{i,t} + \mathbf{h}_{i,t}\boldsymbol{\eta}_{t-n_a:t})\|_2^2 \\ \text{subject to:} \\ \sum_i s_{i,t} = 1 \\ s_{i,t}^2 = 1 \\ \|\boldsymbol{\eta}\|_\infty \leq \epsilon \end{array} \right\} > 0$$

(In)validation Certificates:

- **The model is invalid if and only if there exists N such that $d_N^* > 0$, where:**

$$
\begin{aligned}
d_N^* = \quad & \mathrm{min}_{\mathbf{m}} \textstyle\sum_{t=n_a}^{T} l_t(\mathbf{m}_{t-n_a:t}) \\
& \text{s.t.} \\
& \mathbf{M}_N(\mathbf{m}_{t-n_a:t}) \succeq 0 \; \forall t \in [n_a, T] \\
& \mathbf{L}_N(f_{t,j}\mathbf{m}_{t-n_a:t}) \succeq 0 \;\; \forall t \in [n_a+1, T], j \in \mathrm{N}_{n_y} \\
& \mathbf{L}_N(f_{t,j}\mathbf{m}_{0:n_a}) \succeq 0 \;\; \forall t \in [0, n_a], j \in \mathrm{N}_{n_y}
\end{aligned}
$$

(In)validation Certificates:

- **The model is invalid if and only if there exists N such that $d_N^* > 0$, where:**

$$
\begin{aligned}
d_N^* = \quad & \min_{\mathbf{m}} \textstyle\sum_{t=n_a}^{T} l_t(\mathbf{m}_{t-n_a:t}) \\
& \text{s.t.} \\
& \mathbf{M}_N(\mathbf{m}_{t-n_a:t}) \succeq 0 \; \forall t \in [n_a, T] \\
& \mathbf{L}_N(f_{t,j}\mathbf{m}_{t-n_a:t}) \succeq 0 \;\; \forall t \in [n_a+1, T], j \in \mathrm{N}_{n_y} \\
& \mathbf{L}_N(f_{t,j}\mathbf{m}_{0:n_a}) \succeq 0 \;\; \forall t \in [0, n_a], j \in \mathrm{N}_{n_y}
\end{aligned}
$$

- **Fact: N<T+2**

$$
\mathbf{d} - \mathbf{d}^* = \mathbf{u}_0 + \sum_{j=1}^{m} \mathbf{u}_j \mathbf{g_j}, \qquad \begin{array}{l} \mathbf{u}_j \in \Sigma \\ \mathbf{deg}(\mathbf{u}_0), \mathbf{deg}(\mathbf{u}_j \mathbf{g}_j) \leq 2(T+1) \end{array}
$$

(In)validation Certificates:

- **The model is invalid if and only if there exists N such that $d_N^* > 0$, where:**

$$
\begin{aligned}
d_N^* = \quad & \min_{\mathbf{m}} \textstyle\sum_{t=n_a}^{T} l_t(\mathbf{m}_{t-n_a:t}) \\
& \text{s.t.} \\
& \mathbf{M}_N(\mathbf{m}_{t-n_a:t}) \succeq 0 \ \forall t \in [n_a, T] \\
& \mathbf{L}_N(f_{t,j}\mathbf{m}_{t-n_a:t}) \succeq 0 \ \ \forall t \in [n_a+1, T], j \in \mathsf{N}_{n_y} \\
& \mathbf{L}_N(f_{t,j}\mathbf{m}_{0:n_a}) \succeq 0 \ \ \forall t \in [0, n_a], j \in \mathsf{N}_{n_y}
\end{aligned}
$$

- **Fact: N<T+2**

- **Conjecture: N =2**

Example: Activity Monitoring

- **A priori switched model: walking and waiting, 4% noise**
- **Test sequences of hybrid behavior:**

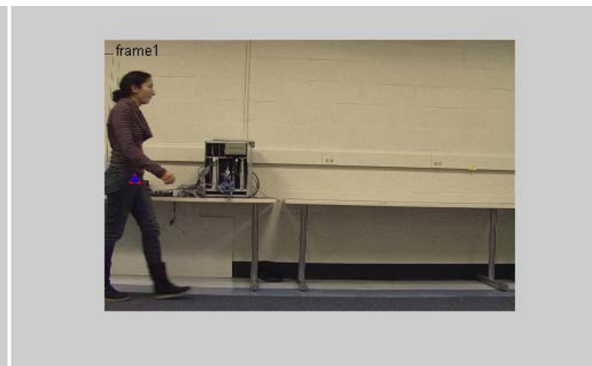
WALK, WAIT

Not Invalidated

RUN

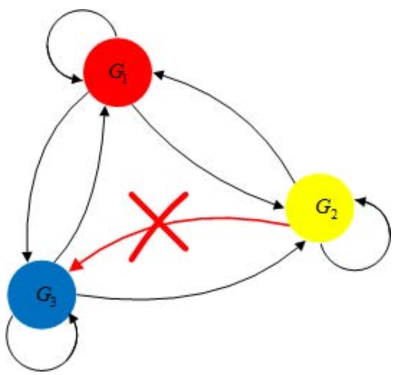
Invalidated

WALK, JUMP



Invalidated

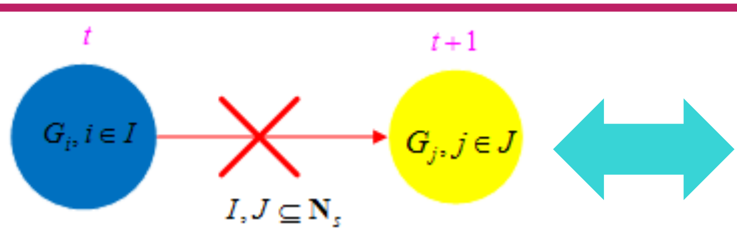
Adding topological constraints:



- **The model is invalid if and only if**

$$d^* \doteq \left\{ \begin{array}{l} \min_{\mathbf{s},\boldsymbol{\eta}} \sum_{t=1}^{T} \mathbf{s}^{\mathbf{2}}_{\mathbf{i,t}}(\|\mathbf{g}_{i,t} + \mathbf{h}_{i,t}\boldsymbol{\eta}_{t-n_a:t})\|_2^2 \\ \text{subject to:} \\ \sum_i s_{i,t} = 1 \\ s_{i,t}^2 = 1 \\ \|\boldsymbol{\eta}\|_\infty \leq \epsilon \end{array} \right\} > 0$$

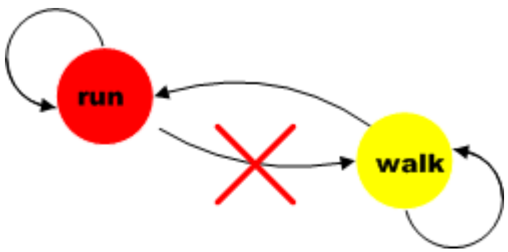
plus additional linear constraints:

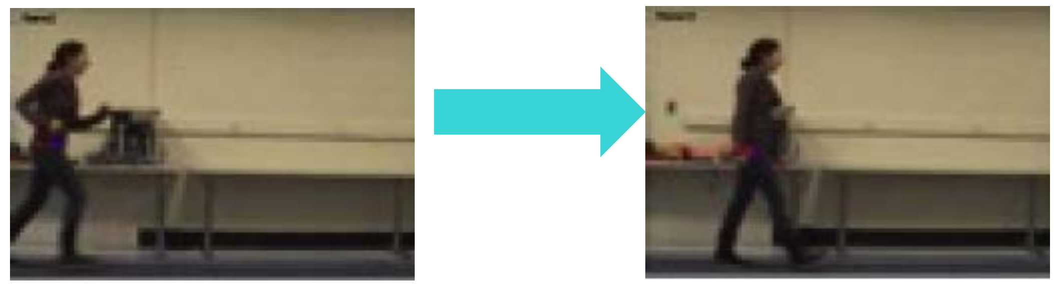


$$s_{i,t} + s_{j,t+1} \leq 1, \forall i \in I, \forall j \in J$$

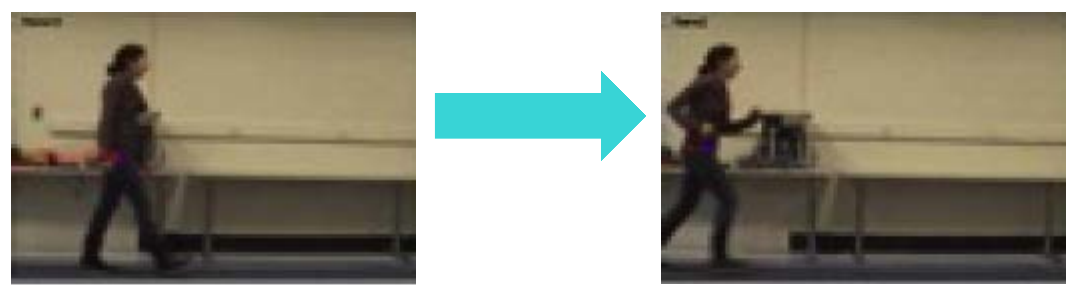
Example: Activity Monitoring

A Priori information



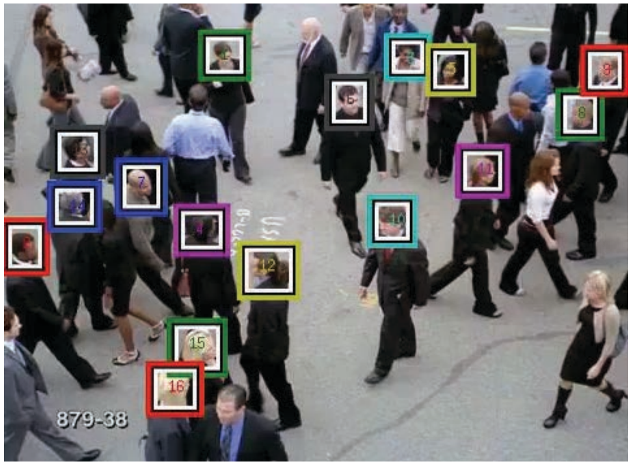


Invalidated (d=0.175)

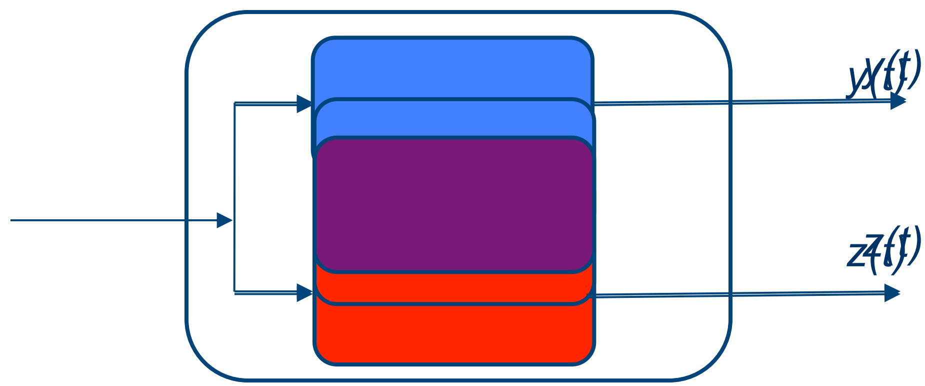


Not Invalidated (d=-3e-8)

Learning: Dynamic data association

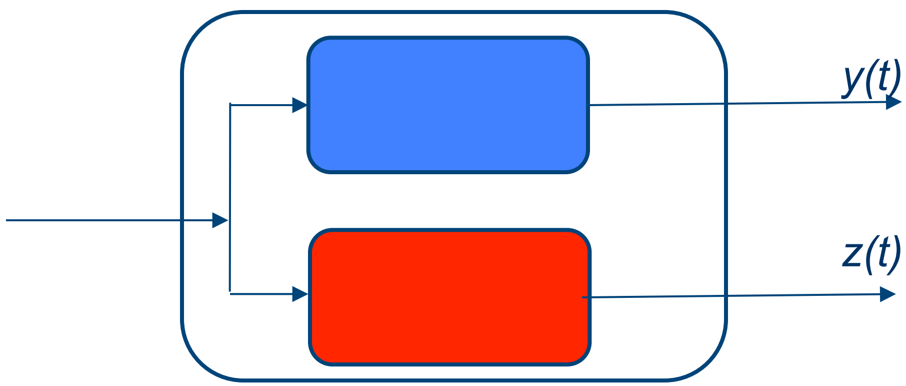


Fast Dynamic Data Association:



Look for simplest joint models

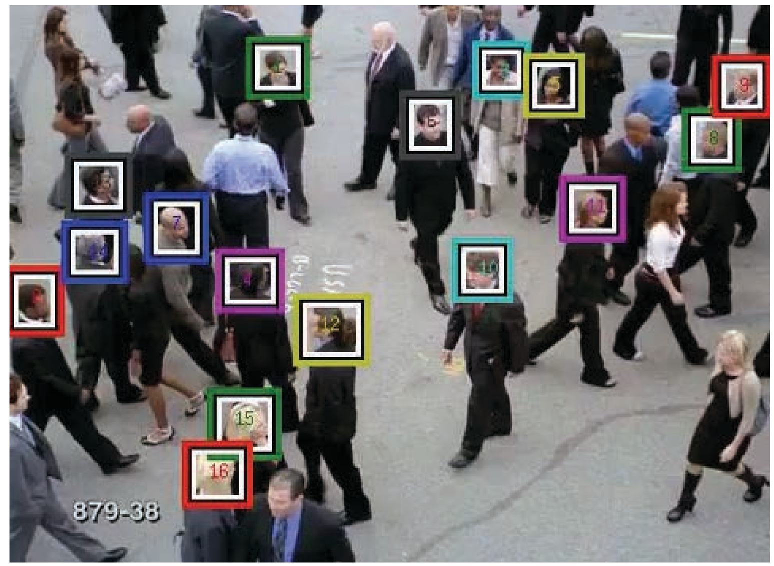
Fast Dynamic Data Association:



Group according to the "similarity score":

$$\frac{\mathrm{rank}(\mathbf{H}_i)+\mathrm{rank}(\mathbf{H}_j)}{\mathrm{rank}([\mathbf{H}_i \ \mathbf{H}_j])} - 1$$

Application: Tracking by detection



Reduces to a min-cut problem with "dynamics- induced" weights

$$\frac{\mathrm{rank}(\mathbf{H}_i)+\mathrm{rank}(\mathbf{H}_j)}{\mathrm{rank}([\mathbf{H}_i\ \mathbf{H}_j])} - 1$$

Fast Dynamic Data Classification:

- **Sequences from the same "class" live in the same subspace:**

$$\mathbf{a}^T\mathbf{y} = 0$$

- **Use a "SVM-like" classifier: find w such that**
  - $\|\mathbf{H}_y\mathbf{w}\| \approx 0$ **in class**
  - $\|\mathbf{H}_y\mathbf{w}\| \gg 0$ **out of class**

Fast Dynamic Data Classification:

- **Sequences for the same "class" live in the same subspace:**

$$\mathbf{a}^T\mathbf{y} = 0$$

- **Use a "SVM-like" classifier:**

$$\min_{W,\gamma,\xi,\zeta_i,z_i} \quad \frac{1}{2}\text{trace}(W) + c\sum \xi_i + c\sum |\zeta_j|$$

s.t.

$$l_i(\gamma - \text{trace}(H_{\mathbf{y}_i}^T H_{\mathbf{y}_i} W)) + \xi_i \geq 1 \quad ;\xi_i \geq 0;\ \text{known labels}$$

$$\gamma - \text{trace}(H_{\mathbf{y}_i}^T H_{\mathbf{y}_i} W) + \zeta_i + z_i = 0 \quad ;|z_i| \geq 1;\ \text{unknown labels}$$

$$\gamma < s||\eta||^2\text{trace}(W) \quad ;W \geq 0$$

Fast Dynamic Data Classification:

- **Sequences for the same "class" live in the same subspace:**

$$\mathbf{a}^T\mathbf{y} = 0$$

- **Use a "SVM-like" classifier:**

$$\min_{W,\gamma,\xi,\zeta_i,z_i} \quad \frac{1}{2}\text{trace}(W) + c\sum \xi_i + c\sum |\zeta_j|$$

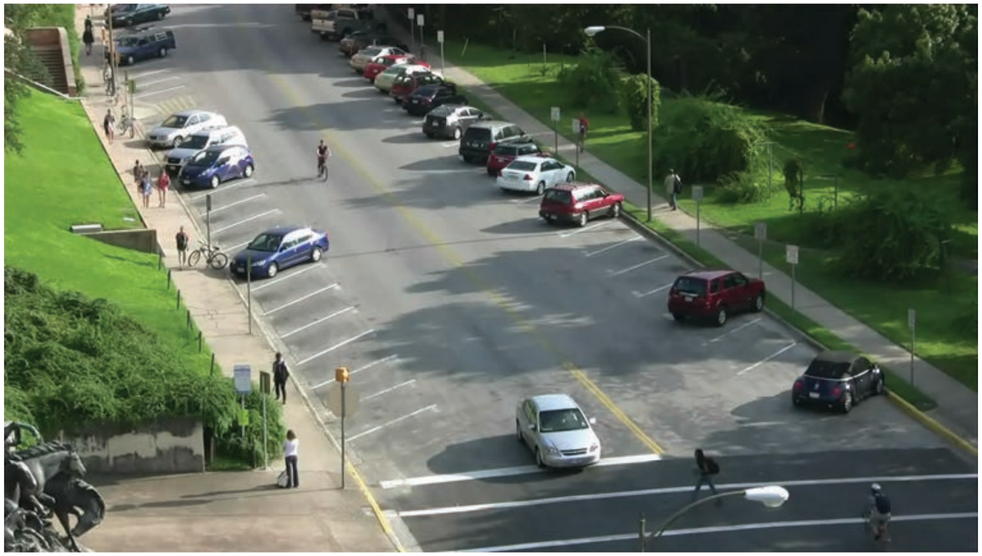
s.t.

$$l_i(\gamma - \text{trace}(H_{\mathbf{y}_i}^T H_{\mathbf{y}_i} W)) + \xi_i \geq 1 \quad ; \xi_i \geq 0; \text{ known labels}$$

$$\gamma - \text{trace}(H_{\mathbf{y}_i}^T H_{\mathbf{y}_i} W) + \zeta_i + z_i = 0 \quad ; |z_i| \geq 1; \text{ unknown labels}$$

$$\gamma < s||\eta||^2\text{trace}(W) \quad ; W \geq 0$$

Fast Dynamic Data Classification:



From sensing to “information”:

Dynamic models as the key to encapsulate
and
analyze (extremely) high dimensional data

- **Data as manifestation of hidden, “sparse” dynamic structures**
- **Extracting information from high volume data streams: finding changes in dynamic invariants (often no need to find the models)**
- **Dynamic models as very compact, robust data surrogates**
- **An interesting connection between several communities:**
  - **Control, computer vision, systems biology, compressive sensing, machine learning,….**

Acknowledgements:

- **Many thanks to:**
  - **IPAM, workshop organizers and audience**
  - **Students**
    - **Dr. M. Ayazoglu, Prof. N. Ozay, Dr. T. Ding, Y. Cheng, C. Dicle, Y. Wang, B. Yilmaz.**
  - **Colleagues:**
    - **Prof. O. Camps, Prof. C. Lagoa**
  - **Funding agencies (AFOSR, DHS, NSF)**

More information as http://robustsystems.ece.neu.edu