Learning parsimonious representations: A Systems and Controls Perspective

M. Sznaier

Robust Systems Lab
Electrical and Computer Engineering
Northeastern University

Verifiable, Control-Oriented Learning On The Fly MURI Kickoff Meeting
Representing Dynamic Data
Rank of the data matrix encapsulates spatial correlation information
Background: Hankel matrices.

Rank of the data matrix encapsulates spatial correlation information
BUT: it is invariant to temporal reordering
Background: Hankel matrices.

\[ H = \begin{bmatrix}
  y_k & y_{k+1} & y_{k+2} & \cdots & y_{k+n} \\
  y_{k+1} & y_{k+2} & \cdots & \cdots & y_{k+n+1} \\
  y_{k+2} & \cdots & \cdots & \cdots & y_{k+n+2} \\
  \vdots & \vdots & \vdots & \ddots & \vdots \\
  y_{k+n} & y_{k+n+1} & y_{k+n+2} & \cdots & y_{k+2n}
\end{bmatrix} \]
Background: Hankel matrices.

\[ H = \begin{bmatrix}
  y_k & y_{k+1} & y_{k+2} & \cdots & y_{k+n} \\
  y_{k+1} & y_{k+2} & \cdots & \cdots & y_{k+n+1} \\
  y_{k+2} & \cdots & \cdots & \cdots & y_{k+n+2} \\
  \vdots & \vdots & \vdots & \vdots & \vdots \\
  y_{k+n} & y_{k+n+1} & y_{k+n+2} & \cdots & y_{k+2n}
\end{bmatrix} \]
Background: Hankel matrices.

Model:

\[ y_k = a_1 y_{k-1} + a_2 y_{k-2} + \ldots + a_n y_{k-n} \]

\[ H = \begin{bmatrix}
    y_k & y_{k+1} & y_{k+2} & \cdots \\
    y_{k+1} & y_{k+2} & \cdots & \cdots \\
    y_{k+2} & \cdots & \cdots & \cdots \\
    \vdots & \vdots & \cdots & \cdots \\
    y_{k+n} & y_{k+n+1} & y_{k+n+2} & \cdots
\end{bmatrix} \]
Background: Hankel matrices.

model:

\[ y_k = a_1 y_{k-1} + a_2 y_{k-2} + \ldots + a_n y_{k-n} \]

\[
\begin{bmatrix}
Y_k & Y_{k+1} & Y_{k+2} & \cdots & Y_{k+n} \\
Y_{k+1} & Y_{k+2} & \cdots & \cdots & Y_{k+n+1} \\
Y_{k+2} & \cdots & \cdots & \cdots & Y_{k+n+2} \\
\vdots & \vdots & \cdots & \cdots & \vdots \\
Y_{k+n} & Y_{k+n+1} & Y_{k+n+2} & \cdots & Y_{k+2n}
\end{bmatrix}
\begin{bmatrix}
a_n \\
\vdots \\
a_2 \\
a_1 \\
-1
\end{bmatrix} = 0
\]

Rank of the Hankel matrix measures the complexity of the dynamics.
Learning parsimonious representations
Classical dimensionality reduction methods:

- Use spatial correlations to project to a lower dimensional manifold

  - Linear (PCA, SVD)
  - Non linear:
    - Locally Linear Embeddings
    - Hessian Eigenmaps
    - Maximum Variance Unfolding
    - Semi Definite Embeddings

- Typically these methods do not fully exploit temporal correlations
Dimensionality reduction of dynamic data:

- Project to a manifold where the dynamics are low order:

\[ f(.) \rightarrow H \rightarrow g(.) \]

**Dynamics on the manifold**

*A hidden Hammerstein-Wiener identification problem*
Dimensionality reduction of dynamic data:

- Project to a manifold where the dynamics are low order:

  Functional interpolation  
  C. Fefferman, A. Israel
A Kernel Formulation:

\[
\min_{K \succeq 0} \ rank(G) - \lambda \text{trace}(K)
\]

subject to:

\[K_{ii} - 2K_{ij} + K_{jj} = \|x_i - x_j\|^2_2 \ \forall \eta_{ij} = 1\]

where

\[G = H_y^T H_y = \sum_{i,j} K_{i,j}\]

\[K_{i,n-1} = \begin{bmatrix}
y_i^T & y_i & y_i^T & y_{i+1} & \cdots & y_i^T & y_{i+n-1} \\
y_{i+1}^T & y_i & y_{i+1}^T & y_{i+1} & \cdots & y_{i+1}^T & y_i \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
y_{i+n-1}^T & y_i & y_{i+n-1}^T & y_{i+1} & \cdots & y_{i+n-1}^T & y_{i+n-1}
\end{bmatrix}\]
A Kernel Formulation:

\[
\min_{K \succeq 0} \text{rank}(G) - \lambda \text{trace}(K)
\]

subject to:

\[
K_{ii} - 2K_{ij} + K_{jj} = \|x_i - x_j\|^2_2 \ \forall \eta_{ij} = 1
\]

where

\[
G = H_y^TH_y = \sum_{i,j} K_{i,j}
\]

\[
K_{i,n-1} = \\
\begin{bmatrix}
Y_i^T & Y_i & Y_i^T & Y_{i+1} & \ldots & Y_i^T & Y_{i+n-1} \\
Y_{i+1}^T & Y_i & Y_{i+1}^T & Y_{i+1} & \ldots & Y_{i+1}^T & Y_i \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
Y_{i+n-1}^T & Y_i & Y_{i+n-1}^T & Y_{i+1} & \ldots & Y_{i+n-1}^T & Y_{i+n-1}
\end{bmatrix}
\]
A Kernel Formulation:

\[
\min_{K \succeq 0} \text{rank}(G) - \lambda \text{trace}(K)
\]

subject to:

\[
K_{ii} - 2K_{ij} + K_{jj} = \|x_i - x_j\|_2^2 \quad \forall \eta_{ij} = 1
\]

where

\[
G = H_y^T H_y = \sum_{i,j} K_{i,j}
\]

\[
K_{i,n-1} = \begin{bmatrix}
y_i^T & y_i & y_i^T & y_{i+1} & \cdots & y_i^T & y_{i+n-1} \\
y_{i+1}^T & y_i & y_{i+1}^T & y_{i+1} & \cdots & y_{i+1}^T & y_i \\
\vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\
y_{i+n-1}^T & y_i & y_{i+n-1}^T & y_{i+1} & \cdots & y_{i+n-1}^T & y_{i+n-1}
\end{bmatrix}
\]

Spatial information
A Kernel Formulation:

\[
\min_{K \succeq 0} \text{rank}(G) - \lambda \text{trace}(K)
\]

subject to:

\[
K_{ii} - 2K_{ij} + K_{jj} = \|x_i - x_j\|_2^2 \quad \forall \eta_{i,j} = 1
\]

*Spatial information*
A Kernel Formulation:

\[
\begin{align*}
\min & \quad rank(G) - \lambda \text{trace}(K) \\
\text{s.t.} & \quad K_{ii} + K_{jj} + 2K_{ij} = (1 + \varepsilon) \| x_i - x_j \|_2^2, \text{if } \eta_{ij} = 1 \\
& \quad K \geq 0 \\
& \quad \sum_{i,j} K_{ij} = 0
\end{align*}
\]

Note: Captures local spatial properties
Note: Captures temporal information

\text{Spatial information}
A Kernel Formulation:

\[
\min_{K \succeq 0} \begin{bmatrix} \text{rank}(G) - \lambda \text{trace}(K) \\
\end{bmatrix}
\]

subject to:

\[
K_{ii} - 2K_{ij} + K_{jj} = \|x_i - x_j\|^2_2 \ \forall i,j = 1
\]

where

\[
G = H_Y^T H_Y
\]

\[
K_{i,n-1} = \begin{bmatrix}
 y_i^T & y_i & y_i^T & y_{i+1} & \cdots & y_i^T & y_{i+n-1} \\
 y_{i+1} & y_i & y_{i+1}^T & y_{i+1} & \cdots & y_{i+1} & y_i \\
 \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
 y_{i+n-1} & y_i & y_{i+n-1}^T & y_{i+n-1} & \cdots & y_{i+n-1} & y_{i+n-1} \\
\end{bmatrix}
\]

Low order dynamics
\begin{align*}
\min_{K \succeq 0} \quad & \text{trace}(G) - \lambda \text{trace}(K) \\
\text{subject to:} \quad & K_{ii} - 2K_{ij} + K_{jj} = \|x_i - x_j\|_2^2 \forall \eta_{ij} = 1
\end{align*}

where

\[ G = H_Y^T H_Y \]

\[ K_{i,n-1} = \begin{bmatrix}
    y_i^T y_i & y_i^T y_{i+1} & \cdots & y_i^T y_{i+n-1} \\
    y_{i+1}^T y_i & y_{i+1}^T y_{i+1} & \cdots & y_{i+1}^T y_i \\
    \vdots & \vdots & \ddots & \vdots \\
    y_{i+n-1}^T y_i & y_{i+n-1}^T y_{i+1} & \cdots & y_{i+n-1}^T y_{i+n-1}
\end{bmatrix} \]
SDP in primal form:
\[
\min_{X \succeq 0} \langle X, A_o \rangle \text{ subject to: }
\]
\[
\langle X, A_i \rangle = c_i, \ i = 1, \ldots, m
\]

- **Correlative Sparsity Graph:**
  - \( n \) vertices
  - There is an edge \((i,j)\) if \(X_{ij}\) appears in the objective or a constraint
Exploting Sparsity in SDPs

SDP in primal form:

$$\min_{X \succeq 0} \langle X, A_o \rangle \text{ subject to:}$$

$$\langle X, A_i \rangle = c_i, \ i = 1, \ldots, m$$

- **Correlative Sparsity Graph:**
  - $n$ vertices
  - There is an edge $(i,j)$ if $X_{ij}$ appears in the objective or a constraint

- If the graph is sparse, the constraint $X \succeq 0$ can be replaced by smaller constraints. Each corresponding to a clique.
Exploting Sparsity in Learning

- Correlative Sparsity Graph:
  - \( n^2 \) vertices, each corresponds to \( <y_i, y_j> \)

```
Complexity dominated by the memory of the system, not the number of points
```

(via Grover’s Theorem on PSD matrix completions)
Academic Example:

- **Ground Truth**
- **3-D lifted data**

**DGCM Manifold**

**SDE Manifold**

**LDE Manifold**
Example: using manifold dynamics for analysis

Original data: 2000 promoters
Model aware Data Driven methods:

- Dynamic data as manifestation of hidden, “sparse” dynamic structures
- Learning from high volume data streams: finding these structures
- Key ingredients: Structure, Randomness, Natural data ordering
- An interesting connection between several communities:
  - Control, sparse optimization, machine learning,....
Future work:

- Incorporation of side information and physics based priors (connection with A. A. Ahmadi, C. Rowley)

- Extension to piece-wise linear dynamics

- Extension to polynomial and rational dynamics (connection with R. Ward)

- All of these are expected to be computationally demanding
  - Need to develop tailored methods that exploit sparsity
  - Explore the use of randomized linear algebra