Operator-valued Kernel-based models for Gene Regulatory Network Inference

Néhémy Lim$^{1,2}$
Joint work with George Michailidis$^{3}$, Cédric Auliac$^1$ and Florence d’Alché-Buc$^{2,4}$

$^1$CEA LIST, Gif-sur-Yvette, France
$^2$IBISC EA 4526, Université d’Évry Val d’Essonne, Évry, France
$^3$Department of Statistics, University of Michigan, USA
$^4$INRIA-Saclay, LRI umr CNRS 86 23, Université Paris-Sud, Orsay, France

September 12, 2013
Outline

1. Reverse-engineering GRN
2. A new framework for network inference
3. A novel nonlinear vector autoregressive model
4. Learning the OKVAR model
5. Numerical studies
6. Conclusion
Outline

1. Reverse-engineering GRN
2. A new framework for network inference
3. A novel nonlinear vector autoregressive model
4. Learning the OKVAR model
5. Numerical studies
6. Conclusion
Example: Gene Regulatory Network

Signed directed graph:

\[ G = (V, E) \]

Adjacency matrix:

\[
\begin{pmatrix}
0 & 1 & 0 \\
-1 & 0 & -1 \\
-1 & 0 & 0
\end{pmatrix}
\]
Reverse-engineering GRN

Reconstruction of gene regulatory networks = Identify **direct** interactions from gene expression data

Signed directed graph

(E. coli subnetwork)
Reverse-engineering GRN

Reconstruction of gene regulatory networks = Identify direct interactions from gene expression data (time-series)

Processing → Time-course gene expression profiles → Reverse-modeling → Signed directed graph (E. coli subnetwork)
Dynamical models and Network inference : State of the art

- Correlations, Mutual Information [Butte et al., 2000; Basso et al., 2005; Faith et al., 2007]
Dynamical models and Network inference: State of the art

- Correlations, Mutual Information [Butte et al., 2000; Basso et al., 2005; Faith et al., 2007]

- **Linear models:**
  - Linear regression [D’Haeseleer, 1999], LASSO [van Someren et al., 2006], linear autoregressive models [Fujita et al., 2007; Shimamura et al., 2009], several-order autoregressive models [Lozano, 2009; Bolstad et al., 2011], Gaussian graphical models [Schäfer & Strimmer, 2005; Charbonnier et al., 2010]
  - Granger causality [Shojaie and Michailidis, 2010 and 2011]
  - State-space models [Perrin et al., 2003; Rangel et al., 2004]
Dynamical models and Network inference: State of the art

- Correlations, Mutual Information [Butte et al., 2000; Basso et al., 2005; Faith et al., 2007]

- **Linear models:**
  - Linear regression [D’Haeseleer, 1999], LASSO [van Someren et al., 2006], linear autoregressive models [Fujita et al., 2007; Shimamura et al., 2009], several-order autoregressive models [Lozano, 2009; Bolstad et al., 2011], Gaussian graphical models [Schäfer & Strimmer, 2005; Carbonnier et al., 2010]
  - Granger causality [Shojaie and Michailidis, 2010 and 2011]
  - State-space models [Perrin et al., 2003; Rangel et al., 2004]

- **Nonlinear models:**
  - Boolean logic [Liang et al., 1998]
  - Ordinary Differential Equations (ODEs) [Chen et al., 1999]
  - Bayesian networks [Murphy & Mian, 1999; Friedman et al., 2000; Perrin et al., 2003; Auliac et al., 2008]
  - Random forests [Huynh-Thu et al., 2010]
  - Non-parametric Gaussian processes [Äijö & Lähdesmäki, 2009]
  - Kernels and time-series [Ralaivola et d’Alché-Buc, 2005; Principe et al., 2011; Kallas et al., 2011]
Limitations

- Specific
- Undirected graph
- Linear
- Small systems
## Limitations

- Specific
- Undirected graph
- Linear
- Small systems

## Requirements

- Generic
- Causality
- Nonlinear
- Scalable

## Our approach

1. Introduce a general framework for nonlinear multivariate modeling and network inference
2. Extend the framework of linear modeling to sparse nonlinear modeling
Outline

1. Reverse-engineering GRN
2. A new framework for network inference
3. A novel nonlinear vector autoregressive model
4. Learning the OKVAR model
5. Numerical studies
6. Conclusion
Modeling nonlinear dynamical systems

Model assumptions

The temporal evolution of the system is ruled by a first-order stationary nonlinear model $h : \mathbb{R}^d \to \mathbb{R}^d$:

$$x_{t+1} = h(x_t) + u_t$$  \hspace{1cm} (1)

where

- $x_0, \ldots, x_{N-1} \in \mathbb{R}^d$: observed time series of a dynamical system comprising of $d$ variables at time $t = 0, \ldots, N - 1$
- $u_t$ is a noise term
Network inference chart

Data
\[ x_0, \ldots, x_{N-1} \]

Network

OKVAR models
September 12, 2013
Network inference chart

Choose $h$

Data
$x_0, \ldots, x_{N-1}$

Network

Néhémy Lim  (CEA LIST/IBISC)  OKVAR models  September 12, 2013  10 / 30
Network inference chart

Data
\[ x_0, \ldots, x_{N-1} \]

Choose \( h \) → Learn \( h \)

Network

Néhémy Lim (CEA LIST/IBISC)
Network inference chart

Data
\[ x_0, \ldots, x_{N-1} \]

Choose \( h \) → Learn \( h \) → ???

Network
Network inference with linear models

1. \[ x_{t+1} = Bx_t + u_t \] : learn \( B \) with a sparsity constraint
2. Threshold \( B \) to get an estimate of the adjacency matrix \( A \)
Network inference with linear models

1. \( x_{t+1} = Bx_t + u_t \): learn \( B \) with a sparsity constraint
2. Threshold \( B \) to get an estimate of the adjacency matrix \( A \)

Network inference with nonlinear models

1. Learn a nonlinear model \( h : \mathbb{R}^d \rightarrow \mathbb{R}^d : x_{t+1} = h(x_t) + u_t \)
Network inference with linear models

1. \( x_{t+1} = Bx_t + u_t \): learn \( B \) with a sparsity constraint
2. Threshold \( B \) to get an estimate of the adjacency matrix \( A \)

Network inference with nonlinear models

1. Learn a nonlinear model \( h : \mathbb{R}^d \rightarrow \mathbb{R}^d : x_{t+1} = h(x_t) + u_t \)
2. Compute the average empirical Jacobian matrix of \( h \):

\[
J(h)_{ij} = \frac{1}{N-1} \sum_{t=0}^{N-2} \frac{\partial h(x_t)_i}{\partial (x_t)_j}
\]

3. Threshold \( J(h) \) to get an estimate of the adjacency matrix \( A \)
Outline

1. Reverse-engineering GRN
2. A new framework for network inference
3. A novel nonlinear vector autoregressive model
4. Learning the OKVAR model
5. Numerical studies
6. Conclusion
Network inference chart

Choose $h$ \rightarrow \text{Learn } h \rightarrow \text{Jacobian}
From single-valued functions . . .

- Binary classification
- Real-valued regression
From single-valued functions . . .

- Binary classification SVM
- Real-valued regression SVR
- Kernels: popular nonparametric nonlinear methods

![Diagram of input and feature space with kernel function \( \phi \)]
From single-valued functions . . .

- Binary classification SVM
- Real-valued regression SVR
- **Scalar** kernels: popular nonparametric nonlinear methods

. . . to vector-valued functions

Recent interest in **operator-valued** kernels [Senkene and Tempel’man, 1973; Michelli & Pontil, 2005; Caponnetto *et al.*, 2008]

Development of new learning tasks:

- Multi-task learning [Evgeniou *et al.*, 2005]
- Functional regression [Kadri *et al.*, 2010]
- Structured output prediction [Brouard *et al.*, 2011]
RKHS theory for vector-valued functions

Notations

- Input set: $\mathcal{X}$
- Output Hilbert space: $\mathcal{F}_y$
- We consider functions $h : \mathcal{X} \to \mathcal{F}_y$
RKHS theory for vector-valued functions

Notations

- Input set: $\mathcal{X}$
- Output Hilbert space: $\mathcal{F}_y$
- We consider functions $h : \mathcal{X} \rightarrow \mathcal{F}_y$


$K : \mathcal{X} \times \mathcal{X} \rightarrow L(\mathcal{F}_y)$ is an operator-valued kernel if:

- $\forall (x, z) \in \mathcal{X} \times \mathcal{X}$, $K(x, z) = K(z, x)^*$
- $\forall m \in \mathbb{N}$, $\forall \{(x_i, y_i)\}_{i=1}^m \subseteq \mathcal{X} \times \mathcal{F}_y$, $\sum_{i,j=1}^m \langle y_i, K(x_i, x_j)y_j \rangle_{\mathcal{F}_y} \geq 0$
Representer theorem [Michelli and Pontil (2005)]
Let \( \lambda > 0 \), \( S_n = \{(x_1, y_1), \ldots, (x_n, y_n)\} \subset \mathbb{R}^d \times \mathbb{R}^d \). Then, the following optimization problem:

\[
\arg\min_{h \in \mathcal{H}} \mathcal{L}(h) = \sum_{i=1}^{n} \| h(x_i) - y_i \|^2 + \lambda \| h \|^2_{\mathcal{H}}
\]

admits a solution of the form:

\[
\hat{h}(\cdot; S_n) = \sum_{\ell=1}^{n} K(x_{\ell}, \cdot) c_{\ell}
\]

(2)

where \( c_{\ell} \in \mathbb{R}^d \), \( \ell = \{1, \cdots, n\} \) are to be learned.
**Representer theorem** [Michelli and Pontil (2005)]

Let $\lambda > 0$, $S_n = \{(x_1, y_1), \ldots, (x_n, y_n)\} \subset \mathbb{R}^d \times \mathbb{R}^d$. Then, the following optimization problem:

$$\arg\min_{h \in \mathcal{H}} \mathcal{L}(h) = \sum_{i=1}^{n} \|h(x_i) - y_i\|^2 + \lambda \|h\|^2_{\mathcal{H}}$$

admits a solution of the form:

$$\hat{h}(\cdot; S_n) = \sum_{\ell=1}^{n} K(x_{\ell}, \cdot) c_{\ell} \quad (2)$$

where $c_{\ell} \in \mathbb{R}^d$, $\ell = \{1, \cdots, n\}$ are to be learned.

---

**Operator-valued Kernel-based Vector AutoRegressive (OKVAR) model**

Given the observed time series $S_N = \{(x_0, x_1), \ldots, (x_{N-2}, x_{N-1})\} \subset \mathbb{R}^d \times \mathbb{R}^d$, the OKVAR model $h$ is defined as

$$h(x_t; S_N) = \sum_{\ell=0}^{N-2} K(x_{\ell}, x_t) c_{\ell} \quad (3)$$
The OKVAR model family

Examples of matrix-valued kernels

1. \( K_1(x, z) = k_1(x, z)B \) with \( k_1(x, z) = \exp(-\gamma_1 \| x - z \|^2) \) and \( B \in S^+_d(\mathbb{R}) \)

2. \( \forall (p, q) \in \{1, \ldots, d\}^2, K_2(x, z)_{pq} = \exp(-\gamma_2(x^p - z^q)^2) \)

3. \( K(x, z) = k_1(x, z)B \circ K_2(x, z) \)
The OKVAR model family

Examples of matrix-valued kernels

1. \( K_1(x, z) = k_1(x, z)B \) with \( k_1(x, z) = \exp(-\gamma_1 \| x - z \|^2) \) and \( B \in S^+_d(\mathbb{R}) \)

\[
J(h_1)_{ij}(t) = \sum_{\ell=0}^{N-2} \sum_{q=1}^{d} b_{iq} c_{q} \frac{\partial k_1(x_t, x_\ell)}{\partial x^j_t}
\]

2. \( \forall (p, q) \in \{1, \ldots, d\}^2, K_2(x, z)_{pq} = \exp(-\gamma_2 (x^p - z^q)^2) \)

\[
J(h_2)_{ij}(t) = 2\gamma_2 (x_t^i - x_t^j) \exp \left(-\gamma_2 (x_t^i - x_t^j)^2 \right) c_t^j
\]

3. \( K(x, z) = k_1(x, z)B \circ K_2(x, z) \)
The OKVAR model family

Examples of matrix-valued kernels

1. \( K_1(x, z) = k_1(x, z)B \) with \( k_1(x, z) = \exp(-\gamma_1 \| x - z \|^2) \) and \( B \in S_d^+(\mathbb{R}) \)

\[
J(h_1)_{ij}(t) = \sum_{\ell=0}^{N-2} \sum_{q=1}^{d} b_{iq} c_q \frac{\partial k_1(x_t, x_\ell)}{\partial x_t^j}
\]

2. \( \forall (p, q) \in \{1, \ldots, d\}^2, K_2(x, z)_{pq} = \exp(-\gamma_2(x^p - z^q)^2) \)

\[
J(h_2)_{ij}(t) = 2\gamma_2(x_t^i - x_t^j) \exp\left(-\gamma_2(x_t^i - x_t^j)^2\right) c_t^j
\]

3. \( K(x, z) = k_1(x, z)B \circ K_2(x, z) \)

\[
J_{ij}(t) = 2\gamma_2 b_{ij} (x_t^i - x_t^j) \exp\left(-\gamma_2(x_t^i - x_t^j)^2\right) c_t^j
- 2\gamma_1 \sum_{\ell \neq t} k_1(x_t, x_\ell) (x_t^i - x_\ell^i) \sum_{p=1}^{d} b_{ip} \exp\left(-\gamma_2(x_t^i - x_\ell^p)^2\right) c_\ell^p
\]
Outline

1. Reverse-engineering GRN
2. A new framework for network inference
3. A novel nonlinear vector autoregressive model
4. Learning the OKVAR model
5. Numerical studies
6. Conclusion
Network inference chart

Data $x_0, \ldots, x_{N-1}$

Choose $h$ → Learn $h$ → Jacobian

Network
Learning the OKVAR model

We aim to solve the following optimization problem:

\[
\begin{align*}
\text{minimize} & \quad \mathcal{L}(B, C) = \sum_{t=0}^{N-2} \|h(x_t; B, C) - x_{t+1}\|^2 + \Omega(B, C) \\
\text{s.t.} & \quad B \in S^+_{d_d}(\mathbb{R}) \\
\end{align*}
\]

with \(\Omega(B, C) = \lambda_h \|h_{B,C}\|_{\mathcal{H}}^2 + \lambda_C \|C\|_{\ell_1} + \lambda_B \|B\|_{\ell_1}\)
Learning the OKVAR model

For fixed $B$ and for $c_\ell$ the loss function to be minimized becomes:

$$
\mathcal{L}(\hat{B}, C, \ell) = \sum_{t=0}^{N-2} \| h(x_t; \hat{B}, C) - x_{t+1} \|^2 + \lambda_h \| h_{\hat{B}, C} \|^2_\mathcal{H} + \lambda_C \| C \|_\ell_1
$$  \hspace{1cm} (5)

For given $\hat{C}$, the loss function to be minimized is the following:

$$
\mathcal{L}(B, \hat{C}) = \sum_{t=0}^{N-2} \| h(x_t; B, \hat{C}) - x_{t+1} \|^2 + \lambda_h \| h_{B, \hat{C}} \|^2_\mathcal{H} + \lambda_B \| B \|_\ell_1
$$  \hspace{1cm} (6)
Learning the OKVAR model

- For fixed $B$ and for $c_\ell$ the loss function to be minimized becomes:

$$\mathcal{L}(\hat{B}, C, \ell) = \sum_{t=0}^{N-2} \| h(x_t; \hat{B}, C) - x_{t+1} \|^2 + \lambda_h \| h_{\hat{B}, C} \|^2_H + \lambda C \| C \|_1 (5)$$

- Proximal gradient algorithms [Martinet (1970); Beck and Teboulle (2010)]

- For given $\hat{C}$, the loss function to be minimized is the following:

$$\mathcal{L}(B, \hat{C}) = \sum_{t=0}^{N-2} \| h(x_t; B, \hat{C}) - x_{t+1} \|^2 + \lambda_h \| h_{B, \hat{C}} \|^2_H + \lambda B \| B \|_1 (6)$$

- Matrix exponentiated gradient updates [Tsuda et al (2005)]
Outline

1. Reverse-engineering GRN
2. A new framework for network inference
3. A novel nonlinear vector autoregressive model
4. Learning the OKVAR model
5. Numerical studies
6. Conclusion
DREAM3 data set

- DREAM = Dialogue for Reverse Engineering Assessments and Methods
- 5 size-10 and 5 size-100 networks (subgraphs of *E. coli* and *S. cerevisiae*) have been generated
  - *Ecoli1, Ecoli2, Yeast1, Yeast2, Yeast3*
- An example of gene regulatory network: *S. cerevisiae* subnetwork

  ![Gene regulatory network diagram](image)

  → activation
  → inhibition

- Challenge: Reconstruct the networks from time-series data of $N = 21$ points
## DREAM3 size-10 data sets: Results

### Table 1: AUROC

<table>
<thead>
<tr>
<th></th>
<th>Ecoli1</th>
<th>Ecoli2</th>
<th>Yeast1</th>
<th>Yeast2</th>
<th>Yeast3</th>
</tr>
</thead>
<tbody>
<tr>
<td>OKVAR + True B</td>
<td>0.956</td>
<td>0.918</td>
<td>0.806</td>
<td>0.781</td>
<td>0.780</td>
</tr>
<tr>
<td>OKVAR</td>
<td>0.717</td>
<td>0.724</td>
<td>0.644</td>
<td>0.740</td>
<td>0.705</td>
</tr>
<tr>
<td>LASSO</td>
<td>0.500</td>
<td>0.547</td>
<td>0.528</td>
<td>0.627</td>
<td>0.582</td>
</tr>
<tr>
<td>GPODE</td>
<td>0.607</td>
<td>0.516</td>
<td>0.494</td>
<td>0.613</td>
<td>0.571</td>
</tr>
<tr>
<td>G1DBN</td>
<td>0.604</td>
<td>0.573</td>
<td>0.494</td>
<td>0.540</td>
<td>0.601</td>
</tr>
<tr>
<td>Team 236</td>
<td>0.621</td>
<td>0.650</td>
<td><strong>0.646</strong></td>
<td>0.438</td>
<td>0.488</td>
</tr>
<tr>
<td>Team 190</td>
<td>0.573</td>
<td>0.515</td>
<td>0.631</td>
<td>0.577</td>
<td>0.603</td>
</tr>
</tbody>
</table>

### Table 2: AUPR

<table>
<thead>
<tr>
<th></th>
<th>Ecoli1</th>
<th>Ecoli2</th>
<th>Yeast1</th>
<th>Yeast2</th>
<th>Yeast3</th>
</tr>
</thead>
<tbody>
<tr>
<td>OKVAR + True B</td>
<td>0.752</td>
<td>0.677</td>
<td>0.473</td>
<td>0.523</td>
<td>0.586</td>
</tr>
<tr>
<td>OKVAR</td>
<td>0.385</td>
<td>0.678</td>
<td>0.430</td>
<td>0.480</td>
<td>0.447</td>
</tr>
<tr>
<td>LASSO</td>
<td>0.119</td>
<td>0.531</td>
<td>0.244</td>
<td>0.305</td>
<td>0.255</td>
</tr>
<tr>
<td>GPODE</td>
<td>0.180</td>
<td>0.146</td>
<td>0.089</td>
<td>0.377</td>
<td>0.341</td>
</tr>
<tr>
<td>G1DBN</td>
<td>0.159</td>
<td>0.534</td>
<td>0.192</td>
<td>0.226</td>
<td>0.248</td>
</tr>
<tr>
<td>Team 236</td>
<td>0.197</td>
<td>0.378</td>
<td>0.194</td>
<td>0.236</td>
<td>0.239</td>
</tr>
<tr>
<td>Team 190</td>
<td>0.152</td>
<td>0.181</td>
<td>0.167</td>
<td>0.371</td>
<td>0.373</td>
</tr>
</tbody>
</table>
### Table 1: AUROC

<table>
<thead>
<tr>
<th></th>
<th>Ecoli1</th>
<th>Ecoli2</th>
<th>Yeast1</th>
<th>Yeast2</th>
<th>Yeast3</th>
</tr>
</thead>
<tbody>
<tr>
<td>OKVAR + True B</td>
<td>0.962</td>
<td>0.971</td>
<td>0.958</td>
<td>0.906</td>
<td>0.897</td>
</tr>
<tr>
<td>OKVAR</td>
<td>0.618</td>
<td>0.620</td>
<td>0.537</td>
<td>0.553</td>
<td>0.522</td>
</tr>
<tr>
<td>LASSO</td>
<td>0.519</td>
<td>0.512</td>
<td>0.507</td>
<td>0.530</td>
<td>0.506</td>
</tr>
<tr>
<td>G1DBN</td>
<td>0.553</td>
<td>0.548</td>
<td>0.510</td>
<td>0.509</td>
<td>0.506</td>
</tr>
<tr>
<td>Team 236</td>
<td>0.527</td>
<td>0.546</td>
<td>0.532</td>
<td>0.508</td>
<td>0.508</td>
</tr>
</tbody>
</table>

### Table 2: AUPR

<table>
<thead>
<tr>
<th></th>
<th>Ecoli1</th>
<th>Ecoli2</th>
<th>Yeast1</th>
<th>Yeast2</th>
<th>Yeast3</th>
</tr>
</thead>
<tbody>
<tr>
<td>OKVAR + True B</td>
<td>0.432</td>
<td>0.516</td>
<td>0.279</td>
<td>0.407</td>
<td>0.364</td>
</tr>
<tr>
<td>OKVAR</td>
<td>0.029</td>
<td>0.093</td>
<td>0.024</td>
<td>0.052</td>
<td>0.053</td>
</tr>
<tr>
<td>LASSO</td>
<td>0.016</td>
<td>0.057</td>
<td>0.016</td>
<td>0.044</td>
<td>0.044</td>
</tr>
<tr>
<td>G1DBN</td>
<td>0.018</td>
<td>0.052</td>
<td>0.022</td>
<td>0.043</td>
<td>0.049</td>
</tr>
<tr>
<td>Team 236</td>
<td>0.019</td>
<td>0.042</td>
<td><strong>0.035</strong></td>
<td>0.046</td>
<td><strong>0.065</strong></td>
</tr>
</tbody>
</table>
Outline

1. Reverse-engineering GRN
2. A new framework for network inference
3. A novel nonlinear vector autoregressive model
4. Learning the OKVAR model
5. Numerical studies
6. Conclusion
Conclusion

- **A key problem**: GRN inference from multivariate time-series data
Conclusion

- **A key problem:** GRN inference from multivariate time-series data

- **Requirements:**
  - Generic
  - Causality
  - Nonlinear
  - Scalable

Results:
- Very good performance of the OKVAR model on simulated benchmark data sets
Conclusion

- **A key problem:** GRN inference from multivariate time-series data
- **Requirements:**
  - Generic ✓
  - Causality ☐
    - Network inference method via the Jacobian
  - Nonlinear
  - Scalable
Conclusion

- **A key problem:** GRN inference from multivariate time-series data

- **Requirements:**
  - Generic ✓
  - Causality □
  - Network inference method via the Jacobian
  - Nonlinear □
    - A novel operator-valued kernel based vector autoregressive model
  - Scalable
Conclusion

- **A key problem:** GRN inference from multivariate time-series data

- **Requirements:**
  - Generic ✓
  - Causality ☐
  - Network inference method via the Jacobian
  - Nonlinear ☐
    - A novel operator-valued kernel based vector autoregressive model
  - Scalable ☐
    - Learning the OKVAR model’s parameters \( \sim \) minutes for size-100 data sets

- **Results:**
  - Very good performance of the OKVAR model on simulated benchmark data sets
Perspectives

- **Theoretical results:**
  - Universality of kernels, consistency of the Jacobian estimator [Fouchet]
  - Generalization error
Perspectives

- **Theoretical results:**
  - Universality of kernels, consistency of the Jacobian estimator [Fouchet]
  - Generalization error

- Probabilistic framework with informative priors
Perspectives

- **Theoretical results:**
  - Universality of kernels, consistency of the Jacobian estimator [Fouchet]
  - Generalization error

- Probabilistic framework with informative priors

- Exploit the OKVAR model for prediction
List of recent papers

- N. Lim*, Y. Senbabaoglu*, G. Michailidis, F. d’Alché-Buc

**BIOINFORMATICS**  **ORIGINAL PAPER**

System biology

OKVAR-Boost: a novel boosting algorithm to infer nonlinear dynamics and interactions in gene regulatory networks

Thank you for your attention!