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]

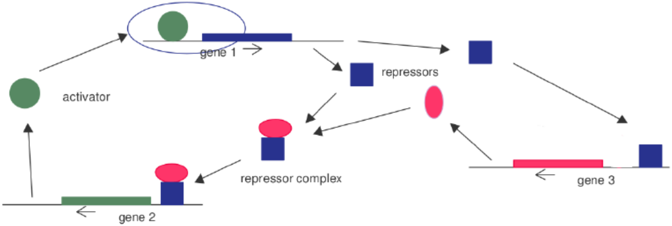
[1]CEA LIST, Gif-sur-Yvette, France
[2]IBISC EA 4526, Université d'Évry Val d'Essonne, Évry, France
[3]Department of Statistics, University of Michigan, USA
[4]INRIA-Saclay, LRI umr CNRS 86 23, Université Paris-Sud, Orsay, France

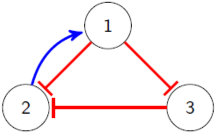
September 12, 2013

Outline

Outline

Example: Gene Regulatory Network





Signed directed graph
$G = (V, E)$

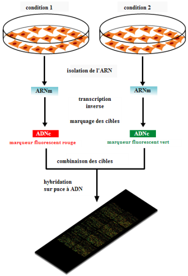
repression
induction

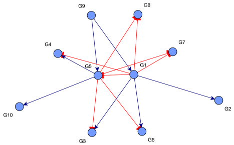
$$\begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & -1 \\ -1 & 0 & 0 \end{pmatrix}$$

Adjacency matrix

Reverse-engineering GRN

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



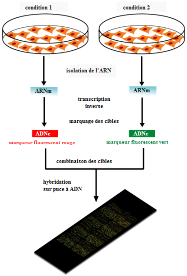


Signed directed graph

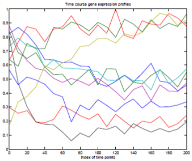
(*E. coli* subnetwork)

Reverse-engineering GRN

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

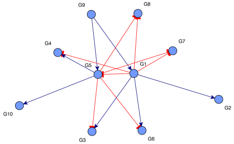


Processing →



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; Charbonnier *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]

Dynamical models and Network inference

Limitations

- Specific
- Undirected graph
- Linear
- Small systems

Dynamical models and Network inference

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

Modeling nonlinear dynamical systems

Model assumptions

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

$$\mathbf{x}_{t+1} = h(\mathbf{x}_t) + \mathbf{u}_t \tag{1}$$

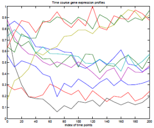
where

- $\mathbf{x}_0, \ldots, \mathbf{x}_{N-1} \in \mathbb{R}^d$: observed time series of a dynamical system comprising of d variables at time $t = 0, \ldots, N-1$
- $\mathbf{u}_t$ is a noise term

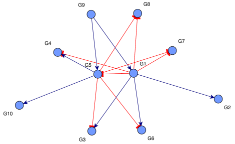
Network inference chart

Data

$\mathbf{x}_0, \ldots, \mathbf{x}_{N-1}$



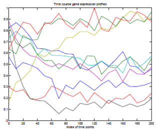
Network



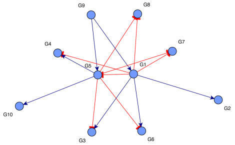
Network inference chart

Data

$\mathbf{x}_0, \ldots, \mathbf{x}_{N-1}$



Network

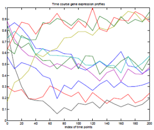


Choose h

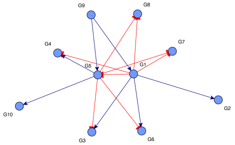
Network inference chart

Data

$\mathbf{x}_0, \ldots, \mathbf{x}_{N-1}$

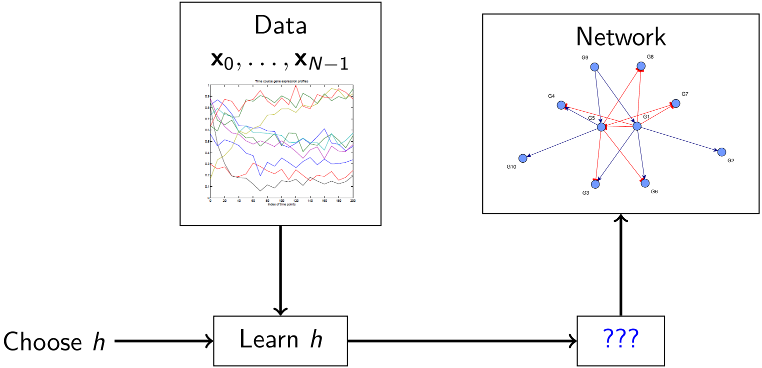


Network



Choose h → Learn h

Network inference chart



Network inference with linear models

1. $\mathbf{x}_{t+1} = B\mathbf{x}_t + \mathbf{u}_t$: learn B with a sparsity constraint
2. Threshold B to get an estimate of the adjacency matrix A

Network inference with linear models

1. $\mathbf{x}_{t+1} = B\mathbf{x}_t + \mathbf{u}_t$: 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 $h : \mathbb{R}^d \to \mathbb{R}^d$: $\mathbf{x}_{t+1} = h(\mathbf{x}_t) + \mathbf{u}_t$

Network inference with linear models

1. $\mathbf{x}_{t+1} = B\mathbf{x}_t + \mathbf{u}_t$: learn B with a sparsity constraint
2. Threshold B to get an estimate of the adjacency matrix A

Network inference with nonlinear models

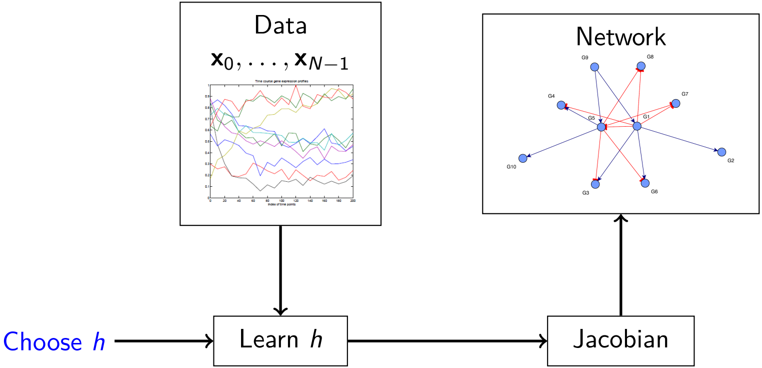
1. Learn a nonlinear model $h : \mathbb{R}^d \to \mathbb{R}^d$: $\mathbf{x}_{t+1} = h(\mathbf{x}_t) + \mathbf{u}_t$
2. Compute the average empirical Jacobian matrix of h :

$$J(h)_{ij} = \frac{1}{N-1} \sum_{t=0}^{N-2} \frac{\partial h(\mathbf{x}_t)_i}{\partial (\mathbf{x}_t)_j}$$

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

Outline

Network inference chart

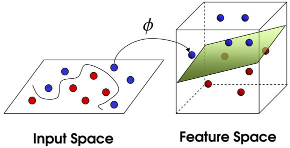


From single-valued functions . . .

- Binary classification
- Real-valued regression

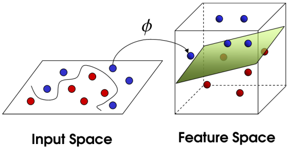
From single-valued functions ...

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



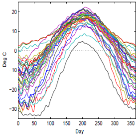
From single-valued functions . . .

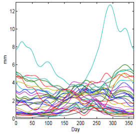
- Binary classification SVM
- Real-valued regression SVR
- **Scalar** kernels : popular nonparametric nonlinear methods



. . . *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 $h : \mathcal{X} \to \mathcal{F}_y$

Operator-valued kernel [Senkene and Tempel'man (1973) , Caponnetto *et al* (2008)]

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

- $\forall (x, z) \in \mathcal{X} \times \mathcal{X}, \quad K(x, z) = K(z, x)^*$
- $\forall m \in \mathbb{N}, \forall \{(x_i, \mathbf{y}_i)\}_{i=1}^m \subseteq \mathcal{X} \times \mathcal{F}_y, \quad \sum_{i,j=1}^m \langle \mathbf{y}_i, K(x_i, x_j)\mathbf{y}_j \rangle_{\mathcal{F}_y} \geq 0$

Representer theorem [Michelli and Pontil (2005)]
Let $\lambda > 0$, $\mathcal{S}_n = \{(x_1, \mathbf{y}_1), \ldots, (x_n, \mathbf{y}_n)\} \subset \mathbb{R}^d \times \mathbb{R}^d$. Then, the following optimization problem:

$$\operatorname*{argmin}_{h \in \mathcal{H}} \mathcal{L}(h) = \sum_{i=1}^{n} \|h(x_i) - \mathbf{y}_i\|^2 + \lambda \|h\|_{\mathcal{H}}^2$$

admits a solution of the form:

$$\hat{h}(\cdot; \mathcal{S}_n) = \sum_{\ell=1}^{n} K(x_\ell, \cdot) \mathbf{c}_\ell \quad (2)$$

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

Representer theorem [Michelli and Pontil (2005)]
Let $\lambda > 0$, $\mathcal{S}_n = \{(x_1, \mathbf{y}_1), \ldots, (x_n, \mathbf{y}_n)\} \subset \mathbb{R}^d \times \mathbb{R}^d$. Then, the following optimization problem:

$$\underset{h \in \mathcal{H}}{\operatorname{argmin}}\, \mathcal{L}(h) = \sum_{i=1}^{n} \|h(x_i) - \mathbf{y}_i\|^2 + \lambda \|h\|_{\mathcal{H}}^2$$

admits a solution of the form:

$$\hat{h}(\cdot; \mathcal{S}_n) = \sum_{\ell=1}^{n} K(x_\ell, \cdot)\mathbf{c}_\ell \qquad (2)$$

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

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

Given the observed time series $\mathcal{S}_N = \{(\mathbf{x}_0, \mathbf{x}_1), \ldots, (\mathbf{x}_{N-2}, \mathbf{x}_{N-1})\} \subset \mathbb{R}^d \times \mathbb{R}^d$, the OKVAR model h is defined as

$$h(\mathbf{x}_t; \mathcal{S}_N) = \sum_{\ell=0}^{N-2} K(\mathbf{x}_\ell, \mathbf{x}_t)\mathbf{c}_\ell \qquad (3)$$

The OKVAR model family

Examples of matrix-valued kernels

1. $K_1(\mathbf{x}, \mathbf{z}) = k_1(\mathbf{x}, \mathbf{z})B$ with $k_1(\mathbf{x}, \mathbf{z}) = \exp(-\gamma_1 \parallel \mathbf{x} - \mathbf{z} \parallel^2)$ and $B \in S_d^+(\mathbb{R})$
2. $\forall (p, q) \in \{1, \dots, d\}^2, K_2(\mathbf{x}, \mathbf{z})_{pq} = \exp(-\gamma_2(x^p - z^q)^2)$
3. $K(\mathbf{x}, \mathbf{z}) = k_1(\mathbf{x}, \mathbf{z})B \circ K_2(\mathbf{x}, \mathbf{z})$

The OKVAR model family

Examples of matrix-valued kernels

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

$$J(h_1)_{ij}(t) = \sum_{\ell=0}^{N-2}\sum_{q=1}^{d} b_{iq} c_\ell^q \frac{\partial k_1(\mathbf{x}_t,\mathbf{x}_\ell)}{\partial x_t^j}$$

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

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

3. $K(\mathbf{x},\mathbf{z}) = k_1(\mathbf{x},\mathbf{z})B \circ K_2(\mathbf{x},\mathbf{z})$

The OKVAR model family

Examples of matrix-valued kernels

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

$$J(h_1)_{ij}(t) = \sum_{\ell=0}^{N-2} \sum_{q=1}^{d} b_{iq} c_\ell^q \frac{\partial k_1(\mathbf{x}_t, \mathbf{x}_\ell)}{\partial x_t^j}$$

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

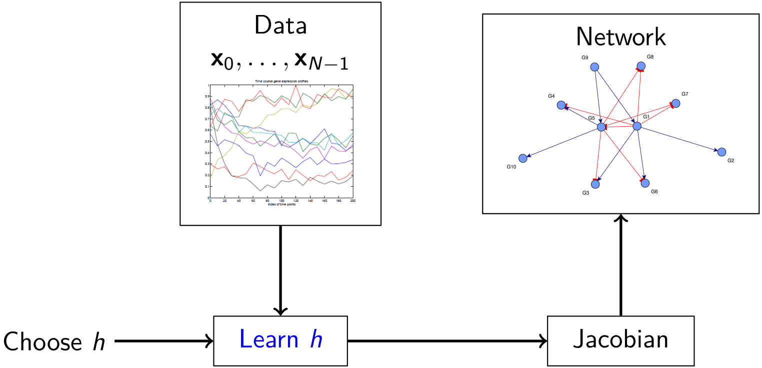
$$J(h_2)_{ij}(t) = 2\gamma_2 (x_t^i - x_t^j) \exp\left(-\gamma_2 (x_t^i - x_t^j)^2\right) c_t^j$$

3. $K(\mathbf{x}, \mathbf{z}) = k_1(\mathbf{x}, \mathbf{z})B \circ K_2(\mathbf{x}, \mathbf{z})$

$$\begin{aligned} J_{ij}(t) &= 2\gamma_2 b_{ij} (x_t^i - x_t^j) \exp\left(-\gamma_2 (x_t^i - x_t^j)^2\right) c_t^j \\ &\quad -2\gamma_1 \sum_{\ell \neq t} k_1(\mathbf{x}_t, \mathbf{x}_\ell)(x_t^j - x_\ell^j) \sum_{p=1}^{d} b_{ip} \exp\left(-\gamma_2 (x_t^i - x_\ell^p)^2\right) c_\ell^p \end{aligned}$$

Outline

Network inference chart



Learning the OKVAR model

We aim to solve the following optimization problem :

$$\begin{aligned} \underset{B\in M_d(\mathbb{R}),C\in M_{N-1,d}(\mathbb{R})}{\text{minimize}} \quad & \mathcal{L}(B,C) = \sum_{t=0}^{N-2} ||h(\mathbf{x}_t;B,C) - \mathbf{x}_{t+1}||^2 + \Omega(B,C) \\ \text{s.t.} \quad & B \in \mathcal{S}_d^+(\mathbb{R}) \end{aligned} \tag{4}$$

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

Learning the OKVAR model

- For fixed B and for $\mathbf{c}_\ell$ the loss function to be minimized becomes:

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

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

$$\mathcal{L}(B, \hat{C}) = \sum_{t=0}^{N-2} ||h(\mathbf{x}_t; B, \hat{C}) - \mathbf{x}_{t+1}||^2 + \lambda_h ||h_{B,\hat{C}}||^2_{\mathcal{H}} + \lambda_B ||B||_{\ell_1} \quad (6)$$

Learning the OKVAR model

- For fixed B and for $\mathbf{c}_\ell$ the loss function to be minimized becomes:

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

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

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

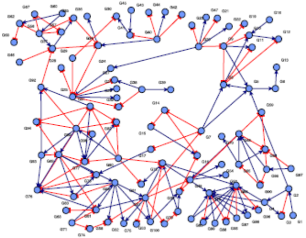
$$\mathcal{L}(B, \hat{C}) = \sum_{t=0}^{N-2} ||h(\mathbf{x}_t; B, \hat{C}) - \mathbf{x}_{t+1}||^2 + \lambda_h ||h_{B,\hat{C}}||^2_{\mathcal{H}} + \lambda_B ||B||_{\ell_1} \quad (6)$$

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

Outline

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



$\rightarrow$ activation
$\rightarrow$ inhibition

- Challenge : Reconstruct the networks from time-series data of $N = 21$ points

DREAM3 size-10 data sets : Results

Table 1: AUROC

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

Table 2: AUPR

	Ecoli1	Ecoli2	Yeast1	Yeast2	Yeast3
OKVAR + *True B*	0.752	0.677	0.473	0.523	0.586
OKVAR	**0.385**	**0.678**	**0.430**	**0.480**	**0.447**
LASSO	0.119	0.531	0.244	0.305	0.255
GPODE	0.180	0.146	0.089	0.377	0.341
G1DBN	0.159	0.534	0.192	0.226	0.248
Team 236	0.197	0.378	0.194	0.236	0.239
Team 190	0.152	0.181	0.167	0.371	0.373

DREAM3 size-100 data sets : Results

Table 1: AUROC

	Ecoli1	Ecoli2	Yeast1	Yeast2	Yeast3
OKVAR + *True B*	0.962	0.971	0.958	0.906	0.897
OKVAR	**0.618**	**0.620**	**0.537**	**0.553**	**0.522**
LASSO	0.519	0.512	0.507	0.530	0.506
G1DBN	0.553	0.548	0.510	0.509	0.506
Team 236	0.527	0.546	0.532	0.508	0.508

Table 2: AUPR

	Ecoli1	Ecoli2	Yeast1	Yeast2	Yeast3
OKVAR + *True B*	0.432	0.516	0.279	0.407	0.364
OKVAR	**0.029**	**0.093**	0.024	**0.052**	0.053
LASSO	0.016	0.057	0.016	0.044	0.044
G1DBN	0.018	0.052	0.022	0.043	0.049
Team 236	0.019	0.042	**0.035**	0.046	**0.065**

Outline

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

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** 2013, pages 1–8
  doi:10.1093/bioinformatics/btt167

  System biology Advance Access publication April 10, 2013

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

- N. Lim*, F. d'Alché-Buc*, C. Auliac, G. Michailidis, Operator-valued Kernel based Vector Autoregressive Models for Network Inference, submitted to *Machine Learning Journal* (under revision)

Thank you for your attention !