Data-driven model learning in linear dynamic networks

Paul Van den Hof

39th Benelux Meeting on Systems and Control
Elspeet, the Netherlands, 11 March 2020

www.sysdynet.eu
www.pvandenhof.nl
p.m.j.vandenhof@tue.nl
Introduction – dynamic networks

Decentralized process control

- Model
  - Process control
    - Schedule plan
    - Dynamic on-line plant-wide optimisation
  - Historical data
    - MPC

- Static economics
  - MPC
  - Data-driven/Physical

- Autonomous driving
  - www.envidia.com

Smart power grid

- Smart Grid
  - Pierre et al. (2012)
  - www.envidia.com

Brain network

- LH
  - P. Hagmann et al. (2008)

Hydrocarbon reservoirs

- Mansoori (2014)
Introduction

Overall trend:

• (Large-scale) interconnected systems
• With hybrid dynamics
• Distributed / multi-agent type monitoring, control and optimization problems
• Data is “everywhere”, big data era
• Model-based operations require accurate/relevant models
• → Learning models from data (including physical insights when available)
Introduction

Distributed / multi-agent control:

With both physical and communication links between systems $G_i$ and controllers $C_i$

How to address data-driven modelling problems in such a setting?
Introduction

The classical (multivariable) identification problems\cite{Ljung1999}:

Identify a model of $G$ on the basis of measured signals $u$, $y$ (and possibly $r$), focusing on continuous LTI dynamics.

We have to move from a simple and fixed configuration to deal with structure in the problem.

\cite{Ljung1999, SoderstromStoica1989, PintelonSchoukens2012}
Early contributors

**Topology detection:** Materassi, Innocenti, Salapaka, Yuan, Stan, Warnick, Goncalves, Sanandaji, Vincent, Wakin, Chiuso, Pillonetto
exploring Granger causality, Bayesian networks, Wiener filters

Subspace algorithms for **spatially distributed systems** with
identical modules (Fraanje, Verhaegen, Werner), or
non-identical ones (Torres, van Wingerden, Verhaegen, Sarwar, Salapaka, Haber)

Here: focus on **structural aspects** in identification setups.
Contents

• Introduction and motivation
• How to model a dynamic network?
• Single module identification
• Global network identification
• Physical networks
• Extensions - Discussion
Dynamic networks for data-driven modeling
Network models

D. Materassi and M.V. Salapaka (2012)
R.N. Mantegna (1999)
J.C. Willems (2007)
D. Koller and N. Friedman (2009)
E.A. Carara and F.G. Moraes (2008)
P.E. Paré et al (2013)
X.Cheng (2019)
Network models

State space representation [1]

Module representation [2]

Dynamic network setup

$G_{76}$ module
$r_i$ external excitation
$v_i$ process noise
$w_i$ node signal
Dynamic network setup

- $r_i$: external excitation
- $v_i$: process noise
- $w_i$: node signal

Diagram:

- $G_{76}$: module
- $G_{61}$, $G_{26}$, $G_{27}$, $G_{37}$
- $G_{21}$, $G_{32}$, $G_{43}$, $G_{54}$, $G_{45}$
- $G_{12}$, $G_{23}$, $G_{34}$, $G_{18}$, $G_{84}$
Dynamic network setup

- $r_i$: external excitation
- $v_i$: process noise
- $w_i$: node signal
Dynamic network setup

- $v_i$: process noise
- $w_i$: node signal
- $r_i$: external excitation
Dynamic network setup

- **$v_i$** process noise
- **$w_i$** node signal
- **$r_i$** external excitation

Symbols:
- $G_{ij}$ module
- $V_i$ node signal
- $W_i$ external excitation
Dynamic network setup

Assumptions:
- Total of $L$ nodes
- Network is well-posed and stable
- Modules are dynamic LTI, may be unstable
- Disturbances are stationary stochastic and can be correlated

\[
\begin{bmatrix}
  w_1 \\
  w_2 \\
  \vdots \\
  w_L
\end{bmatrix}
= \begin{bmatrix}
  0 & G_{12}^0 & \cdots & G_{1L}^0 \\
  G_{21}^0 & 0 & \cdots & G_{2L}^0 \\
  \vdots & \vdots & \ddots & \vdots \\
  G_{L1}^0 & G_{L2}^0 & \cdots & 0
\end{bmatrix}
\begin{bmatrix}
  w_1 \\
  w_2 \\
  \vdots \\
  w_L
\end{bmatrix}
+ R^0
\begin{bmatrix}
  r_1 \\
  r_2 \\
  \vdots \\
  r_K
\end{bmatrix}
+ \begin{bmatrix}
  v_1 \\
  v_2 \\
  \vdots \\
  v_L
\end{bmatrix}
\]

\[
 w(t) = G^0(q)w(t) + R^0(q)r(t) + v(t)
\]
Dynamic network setup

Setup covers the situation of bilaterally coupled (physical) systems:
Dynamic network setup

Many new data-driven modeling questions can be formulated

Measured time series:
\[ \{w_i(t)\}_{i=1,...,L}; \quad \{r_j(t)\}_{j=1,...,K} \]
Under which conditions can we estimate the topology and/or dynamics of the full network?
Model learning problems

How/when can we learn a local module from data (with known/unknown network topology)? Which signals to measure?
Model learning problems

Where to optimally locate sensors and actuators?
Model learning problems

Same questions for a subnetwork
Model learning problems

How can we benefit from known modules?
Model learning problems

Fault detection and diagnosis; detect/handle nonlinear elements
Model learning problems

Can we distribute the computations?
Dynamic network setup

- Identification of a local module (known topology)
- Identification of the full network
- Topology estimation
- Identifiability
- Sensor and excitation allocation
- Fault detection
- User prior knowledge of modules
- Distributed identification
- Scalable algorithms

Measured time series:
\[ \{w_i(t)\}_{i=1,...,L}; \quad \{r_j(t)\}_{j=1,...,K} \]

Many new data-driven modeling questions can be formulated
Dynamic network setup - graph

Nodes are vertices; modules/links are edges

Extended graph: including the external signals and disturbance correlations
Application: Networks of (damped) oscillators

- Power systems, vehicle platoons, thermal building dynamics, ...
- Spatially distributed
- Bilaterally coupled
- No central coordination \(\Rightarrow\) local identification problems
Single module identification - Example

Decentralized MPC
2 interconnected MPC loops

Target:
Identify interaction dynamics $G_{21}, G_{12}$

Addressed by Gudi & Rawlings (2006) for the situation $G_{12} = 0$ (no cycles)

Contents

• Introduction and motivation
• How to model a dynamic network?
• Single module identification
• Global network identification
• Physical networks
• Extensions - Discussion
Single module identification
Single module identification

For a network with known topology:

- Identify $G_{21}^0$ on the basis of measured signals
- Which signals to measure? Preference for local measurements
- When is there enough excitation / data informativity?
Naïve approach: identify based on $w_1$ and $w_2$ : in general does not work.
Single module identification

Identifying $G_{21}^0$ is part of a 4-input, 1-output problem

If noises $v_k$ are correlated it may even be part of a MIMO problem
Single module identification

Input signals will be correlated:
similar as in a closed-loop situation

What is required for
identifiability / data informativity?

Ability to distinguish between models
independent of id-method

Information content of signals
dependent on id-method

Identifying $G_{21}^0$ is part of a
4-input, 1-output problem
Single module identification

Identifiying $G_{21}^0$ is part of a 4-input, 1-output problem

All parallel paths, and loops around the output, plus input $w_1$ should have an independent external signal $r$ or $v$

[1] Weerts et al., Automatica 2018, CDC 2018
Single module identification

Which node signals to measure?

Dependent on

- $\nu$ signals uncorrelated or not
- Excitation conditions satisfied through $r$- and/or $\nu$-signals

Typical solution:

- One additional measured signal for each parallel path/loop
- Additional signals if excitation is through $\nu$ signals
- Variation in available algorithms / options

[4] Bazanella et al., CDC2019
[5] PVdH, Ramaswamy, CDC2019
**Single module identification**

one signal per parallel path/loop:

With a 3-input, 1 output model we can consistently identify $G_{21}^0$

When excitation is through disturbance signals $\nu$:

- dealing with **confounding variables**, [1][2] i.e. correlated disturbances on inputs and outputs
- can be addressed by adding inputs/outputs to the estimation problem [3]

---

[3] PVdH et al, CDC 2019
Single module identification

Typical solution:

- MISO (sometimes MIMO) estimation problem
- to be solved by any (closed-loop) identification algorithm, e.g. direct/indirect method
Machine learning in local module identification

• MISO identification with all modules parameterized
• Brings in two major problems:
  ▶ Large number of parameters to estimate
  ▶ Model order selection step for each module (CV, AIC, BIC)

• For 5 modules, combinations = 244,140,625

  Increases variance
  Computationally challenging

• We need only the target module. No NUISANCE!
Machine learning in local module identification

Maximize marginal likelihood of output data: $\hat{\eta} = \arg\max p(w_j; \eta)$

$$\eta := [\theta \ \lambda_j \ \lambda_{k_1} \ldots \ \lambda_{k_p} \ \beta_j \ \beta_{k_1} \ldots \ \beta_{k_p} \ \sigma_j^2]^T$$

Numerical simulation

- Identify $G_{31}$ given data
- 50 independent MC simulation
- Data = 500
Summary single module identification

• Path-based conditions for **network identifiability** (where to excite?)

• Graph tools for checking conditions

• Degrees of freedom in selection of measured signals – sensor selection

• Methods for **consistent** and **minimum variance** module estimation, and effective (scalable) algorithms

• A priori known modules can be accounted for
Contents

• Introduction and motivation
• How to model a dynamic network?
• Single module identification
• Global network identification
• Diffusively coupled physical networks
• Extensions - Discussion
Full network identification

Under which conditions can we estimate the topology and/or dynamics of the full network?
Network identifiability

Question: Can different dynamic networks be distinguished from each other from measured signals $w$, $r$?
Network identifiability

The identifiability problem:

The network model:

\[ w(t) = G(q)w(t) + R(q)r(t) + \frac{H(q)e(t)}{v(t)} \]

can be transformed with any rational \( P(q) \) :

\[ P(q)w(t) = P(q)\{G(q)w(t) + R(q)r(t) + H(q)e(t)\} \]

to an equivalent model:

\[ w(t) = \tilde{G}(q)w(t) + \tilde{R}(q)r(t) + \tilde{H}(q)e(t) \]

Nonuniqueness, unless there are structural constraints on \( G, R, H \).

---

[2] Bottegal et al., SYSID 2017
Network identifiability

Consider a **network model set**:

\[ \mathcal{M} = \{ (G(\theta), R(\theta), H(\theta)) \} \theta \in \Theta \]

representing structural constraints on the considered models:

- modules that are fixed and/or zero (topology)
- locations of excitation signals
- disturbance correlation

**Generic identifiability of** \( \mathcal{M} \) :

- There do not exist distinct equivalent models
- for almost all models in the set.

---

[1] Weerts et al., SYSID2015; Weerts et al., Automatica, March 2018;
Example 5-node network

Conditions for identifiability: rank conditions on transfer function

For the **generic case**, the rank can be calculated by a graph-based condition\(^1\),\(^2\),\(^3\):

\[
\text{Generic rank} = \text{number of vertex-disjoint paths}
\]

\(2\) vertex-disjoint paths \(\rightarrow\) full row rank 2

The rank condition has to be checked for all nodes.

\[^1\] Van der Woude, 1991
\[^2\] Hendrickx, Gevers & Bazanella, CDC 2017, TAC 2019
\[^3\] van Waarde et al., ArXiv, 2018.
Synthesis solution for network identifiability

Allocating external signals for generic identifiability:

1. Cover the graph of the network model set by a set of disjoint pseudo-trees
   Pseudo-trees:
   - Tree with root in green
   - Cycle with outgoing trees; Any node in cycle is root

   Edges are disjoint and all out-neighbours of a node are in the same pseudo-tree

2. Assign an independent external signal (\( r \) or \( e \)) at a root of each pseudo-tree.
   This guarantees generic identifiability of the model set.

Where to allocate external excitations for network identifiability?

All indicated modules are parametrized

Two disjoint pseudo-trees
Where to allocate external excitations for network identifiability?

Two independent excitations guarantee generic network identifiability.

Where to allocate external excitations for network identifiability?

- Nodes are signals $w$ and external signals $(r, e)$ that are input to parametrized link.
- Known (nonparametrized) links do not need to be covered.

Summary identifiability of full network

Identifiability of network model sets is determined by

- Presence and location of external signals, and
- Correlation of disturbances
- Topology of parametrized modules

- Graphic-based tool for synthesizing allocation of external signals

Extensions:
- Situations where not all node signals are measured [1]

Algorithms for identification of full network

(Prediction error) identification methods will typically lead to large-scale non-convex optimization problems.

Convex relaxation algorithms are being developed\cite{1} as well as machine learning tools.

[1] Weerts, Galrinho et al., SYSID 2018
Topology identification

- Topology resulting from full dynamic model

- Alternative: non-parametric models (Wiener filters [1]) or kernel-based approaches [2][3]

- Modeling module dynamics by Gaussian processes,
  
  kernel with 2 parameters for each dynamic module

- Optimizing likelihood of the data as function of parameters and topology:

  $$p(\{w(t)\}_{t=1}^{N} | \theta, \mathcal{G})$$

- Forward-backward search over topologies + empirical Bayes (EM) for parameters

[3] Shi, Bottegal, PVdH, ECC 2019
Topology identification

50 MC realizations of network with 6 nodes.

[1] Shi, Bottegal, PVdH, ECC 2019
Neurodynamic effect of listening to Mozart music

Identifying changes in network connections in the brain, after intensely listening for one week

Figure 3: Spatial maps of the 30 active brain networks found through the ICA decomposition. Each image consists of 3 relevant horizontal slices of the brain, where the spatial map is indicated by the red color scale.

Algorithms for identification of full network

Particular feature for larger networks:

Modeling disturbances as a **reduced rank process**: (cf dynamic factor analysis\[^1\])

\[ \text{dim}(e) < \text{dim}(v) \]

Consequences for **estimation**\[^3\]:

- Optimization becomes a **constrained quadratic problem** with ML properties for Gaussian noise
- Reworked Cramer Rao lower bound
- Some parameters can be estimated variance free $\rightarrow$ regularization effect

---

\[^1\] Deistler et al., EJC, 2010.
\[^2\] Zorzi and Chiuso, Automatica 2017.
\[^3\] Weerts et al., Automatica dec 2018.
Contents

• Introduction and motivation
• How to model a dynamic network?
• Single module identification
• Global network identification
• Physical networks
• Extensions - Discussion
Physical networks
Back to the basics of physical interconnections

In connecting physical systems, there is often no predetermined direction of information [1]

Example: resistor / spring connection in electrical / mechanical system:

Difference of node signals drives the interaction: **diffusive coupling**

Equation for node $j$:

$$M_j \ddot{w}_j(t) + D_{j0} \dot{w}_j(t) + \sum_{k \neq j} D_{jk}(\dot{w}_j(t) - \dot{w}_k(t)) + K_{j0} w_j(t) + \sum_{k \neq j} K_{jk}(w_j(t) - w_k(t)) = u_j(t),$$
Mass-spring-damper system

- Masses $M_j$
- Springs $K_{jk}$
- Dampers $D_{jk}$
- Input $u_j$

\[
\begin{bmatrix}
M_1 & M_2 \\
M_2 & M_3
\end{bmatrix}
\begin{bmatrix}
\dot{w}_1 \\
\dot{w}_2 \\
\dot{w}_3
\end{bmatrix}
+ \begin{bmatrix}
0 \\
D_{20}
\end{bmatrix}
\begin{bmatrix}
\dot{w}_1 \\
\dot{w}_2 \\
\dot{w}_3
\end{bmatrix}
+ \begin{bmatrix}
K_{10} & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
w_1 \\
w_2 \\
w_3
\end{bmatrix}
+ \begin{bmatrix}
D_{13} & 0 & -D_{13} \\
0 & D_{23} & -D_{23} \\
-D_{13} & -D_{23} & D_{13} + D_{23}
\end{bmatrix}
\begin{bmatrix}
\dot{w}_1 \\
\dot{w}_2 \\
\dot{w}_3
\end{bmatrix}
+ \begin{bmatrix}
K_{12} + K_{13} & -K_{12} & -K_{13} \\
-K_{12} & K_{12} & 0 \\
-K_{13} & 0 & K_{13}
\end{bmatrix}
\begin{bmatrix}
w_1 \\
w_2 \\
w_3
\end{bmatrix}
= \begin{bmatrix}
0 \\
u_2
\end{bmatrix}
\]

\[
\begin{bmatrix}
A(p) \\
B(p)
\end{bmatrix}
\text{ diagonal Laplacian}
\]

\[
\begin{bmatrix}
A(p) \\
B(p)
\end{bmatrix}
w(t) = u(t) \quad A(p), B(p) \quad \text{polynomial} \quad p = \frac{d}{dt}
\]
Mass-spring-damper system

\[
\begin{bmatrix}
\underbrace{A(p)}_{\text{diagonal}} + \underbrace{B(p)}_{\text{Laplacian}}
\end{bmatrix} w(t) = u(t) \quad A(p), B(p) \text{ polynomial}
\]

\[
\begin{bmatrix}
\underbrace{Q(p)}_{\text{diagonal}} - \underbrace{P(p)}_{\text{hollow\&symmetric}}
\end{bmatrix} w(t) = u(t)
\]

This fully fits in the earlier module representation:

\[w(t) = Gw(t) + \underbrace{Rr(t) + He(t)}_{Q^{-1}(p)u(t)}\]

with the additional condition that:

\[G(p) = Q(p)^{-1}P(p) \quad Q(p), P(p) \text{ polynomial}\]

\[P(p) \text{ symmetric, } Q(p) \text{ diagonal}\]
Module representation

Consequences for node interactions:

- Node interactions come in pairs of modules
- Where numerators are the same

Framework for network identification remains the same

- Symmetry can simply be incorporated in identification
Local network identification

Identification of one physical interconnection

Identification of two modules $G_{jk}$ and $G_{kj}$
Immersion conditions

For simultaneously identifying two modules in one interconnection:

The parallel path and loops-around-the-output condition, now simplifies to:

Measuring/exciting all neighbouring nodes of $w_2$ and $w_3$ leads to a solution
Summary physical networks

• Physical networks fit within the module framework (special case)
  - no restriction to second order equations
• Earlier identification framework can be utilized
• Local identification is well-addressed (and stays really local)
• Framework is fit for representing cyber-physical systems
  (combining physical bi-directional links, and cyber uni-directional links).
Extensions - Discussion
Extensions - Discussion

• Including sensor noise \[1\]
  - Errors-in-variables problems can be more easily handled in a network setting

• Distributed estimation (MISO models) \[2\]
  - Communication constraints between different agents
  - Recursive (distributed) estimator converges to global optimizer (more slowly)

• Experiment design \[3,4\]
  - Design of least costly experiments

Summary

• **Dynamic network modeling:**
  intriguing research topic with many open questions

• The (centralized) LTI framework is only just the beginning

• Further move towards data-aspects related to distributed control

• and large-scale aspects

• and bring it to real-life applications
Acknowledgements

Lizan Kivits, Shengling Shi, Karthik Ramaswamy, Tom Steentjes, Mircea Lazar, Jobert Ludlage, Mannes Dreef, Tijs Donkers, Giulio Bottegal, Maarten Schoukens, Xiaodong Cheng

Co-authors, contributors and discussion partners:

Arne Dankers
Harm Weerts

Xavier Bombois
Peter Heuberger
Donatelllo Materassi
Manfred Deistler
Michel Gevers
Jonas Linder
Sean Warnick
Alessandro Chiuso
Hakan Hjalmarsson
Miguel Galrinho
Martin Enqvist
Further reading


Papers available at www.pvandenhof.nl
The end