Single module identification – local direct method

Paul Van den Hof

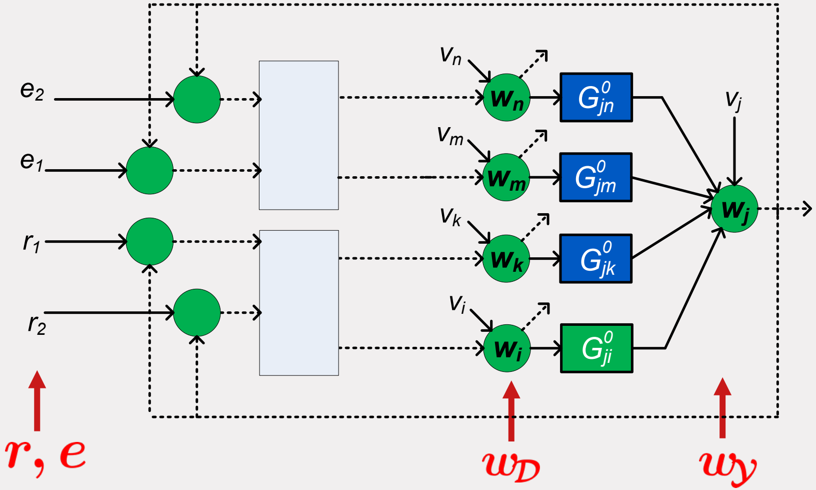
Doctoral School Lyon, France, 8 April 2021

www.sysdynet.eu
www.pvandenhof.nl
p.m.j.vandenhof@tue.nl

EINDHOVEN UNIVERSITY OF TECHNOLOGY

European Research Council

Local direct method



$$\varepsilon(t,\theta) = \bar{H}(q,\theta)^{-1}[w_{\mathcal{Y}}(t) - \bar{G}(q,\theta)w_{\mathcal{D}}(t)]$$

- Estimate transfer $w_{\mathcal{D}} \rightarrow w_{\mathcal{Y}}$ and model the disturbance process on the output.
- consistent estimate and ML properties

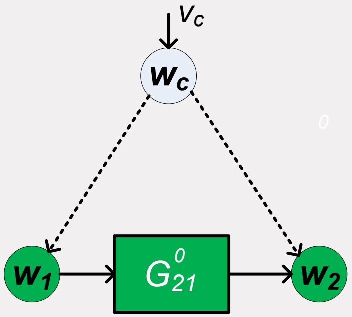
Additional problem:

If:
- v signals are correlated, i.e. $\Phi_v(\omega)$ non-diagonal, or
- some in-neighbors of $w_{\mathcal{Y}}$ are not included in $w_{\mathcal{D}}$

then confounding variables can occur, destroying the consistency results

[1] Dankers et al., IEEE-TAC, 2016; Dankers et al., IFAC 2017

Confounding variable



Confounding variable [1][2]**:**

Unmeasured signal that has (unmeasured paths) to both the input and output of an estimation problem.

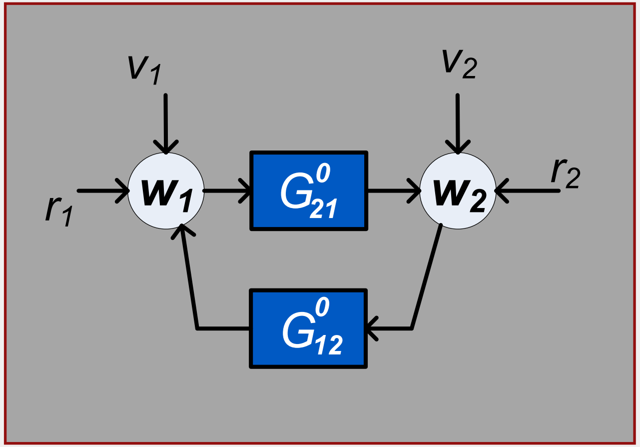
In networks they can appear in two different ways:

- If v disturbances on inputs and outputs are correlated
- If non-measured in-neighbors of w_j affect signals in $w_{\mathcal{D}}$

[1] J. Pearl, *Stat. Surveys, 3,* 96-146, 2009
[2] A.G. Dankers et al., *Proc. IFAC World Congress*, 2017.

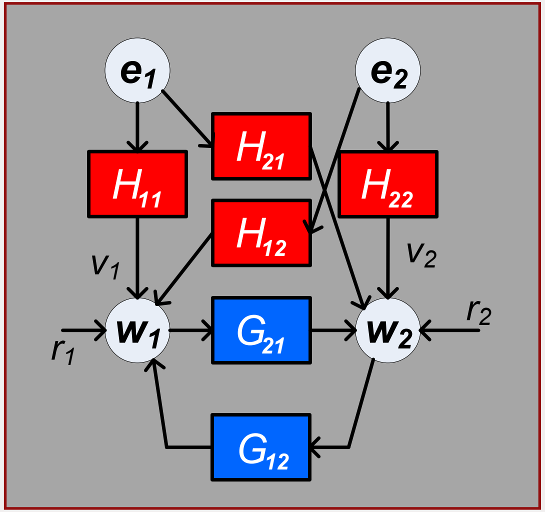
Confounding variables

- Direct confounding variable:



v_1, v_2
correlated

Multivariate
noise model



When estimating $w_1 \rightarrow w_2$ consistency is lost!

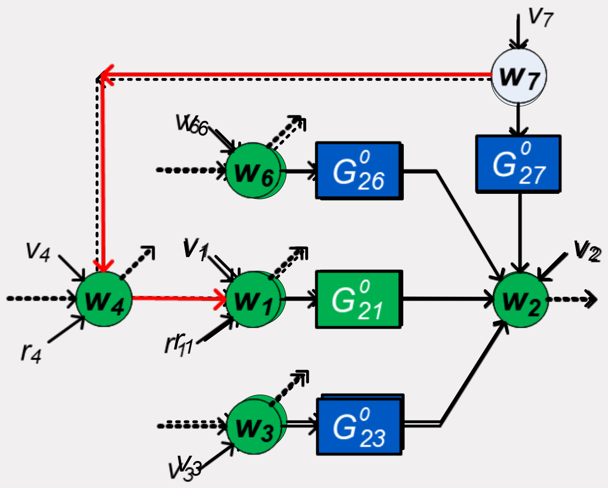
Predict both w_1 and w_2 ⟶ **Adding predicted outputs** [1],[2]

Becomes a multi output local identification problem.

[1] P.M.J. Van den Hof et al. , CDC 2019.
[2] K.R. Ramaswamy et al., IEEE-TAC, 2021.

Confounding variables

- Indirect confounding variable:



Non-measurable w_7 is a confounding variable

Two possible solutions:

1. Include w_4 ➡ add predictor input

$$w_{\mathcal{D}} = \{w_1, w_3, w_4, w_6\} \quad w_{\mathcal{Y}} = \{w_2\}$$

2. Predict w_1 too ➡ add predictor output

$$w_{\mathcal{D}} = \{w_1, w_3, w_6\} \quad w_{\mathcal{Y}} = \{w_1, w_2\}$$

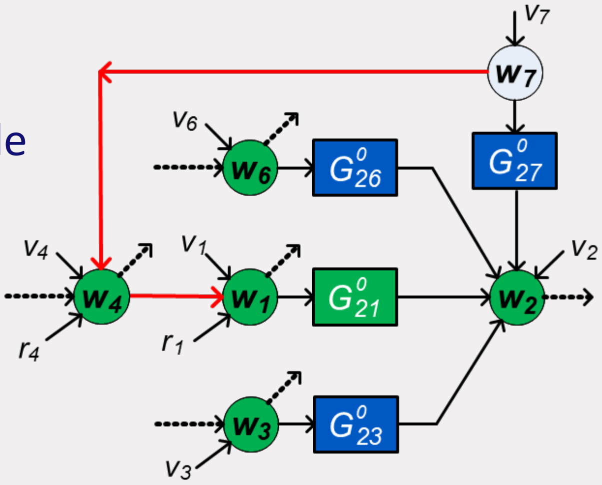
- There are degrees of freedom in choosing the predictor model

Handling confounding variables in local module identification

``Blocking'' confounding variables by adding predictor inputs

By adding w_4 as predictor input, new confounding variable for $w_4 \rightarrow w_2$.
Does this help?

Yes. Since we do not need an accurate model of G_{24}

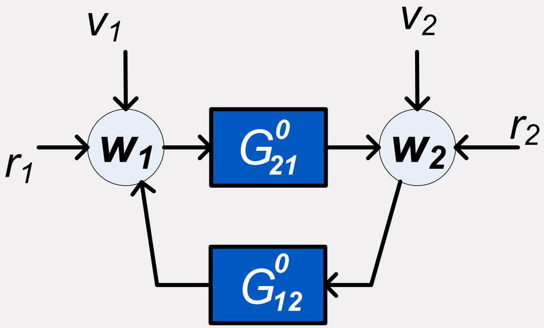


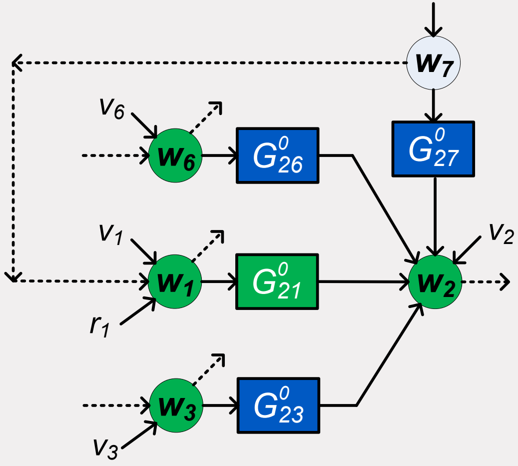
Handling confounding variables in local module identification

Confounding variables and closed-loop mechanisms

In closed-loop case
(when predicting only w_2):

- Correlation between w_1 and v_2 is no problem, as long as it passes through w_2.
- Correlation between v_1 and v_2 is a problem.





Algorithm for dealing with confounding variables

For estimating target module G_{ji}

1. Select input w_i and output w_j
2. Add inputs to satisfy the parallel path and loop condition
3. Check on direct confounding variables → add output and return to step 2
4. Check on indirect confounding variables
   a) Add output and return to step 2, **OR**
   b) Add input

Algorithm always reaches a convergence point where conditions are satisfied.

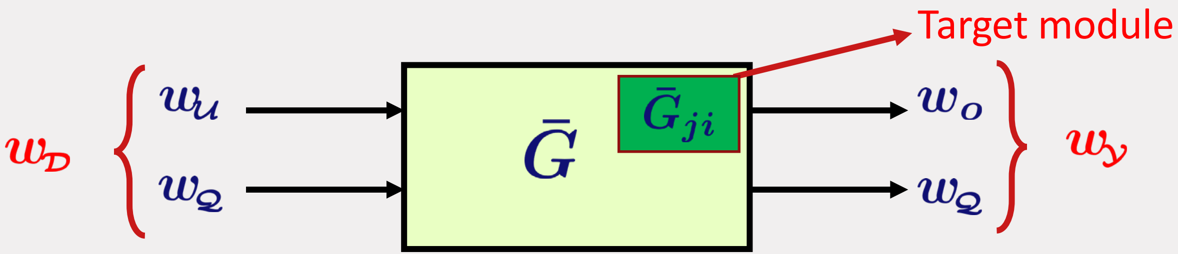
The choice options lead to different end-results for signals to be included

→ **different predictor models**

that all can reach consistency of $\hat{G}_{ji}$

Direct method

General setup:

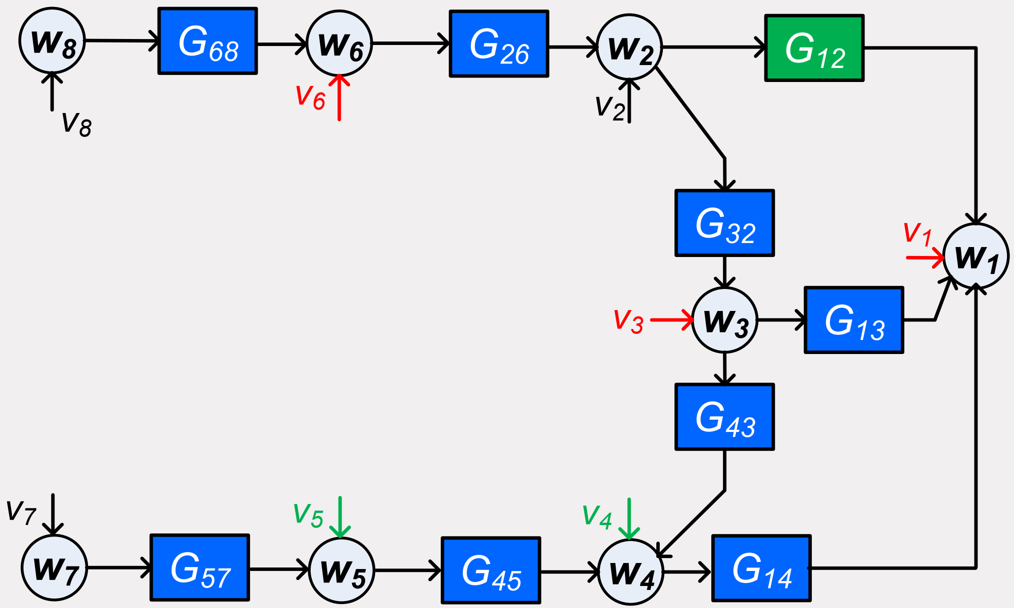


Different predictor models:

- Full input case: include all in-neighbors of $w_{\mathcal{Y}}$
- Minimum node signals case : maximize number of outputs
- User selection case : dedicated choice based on measurable nodes

Different strategies – direct method

- Full input case
- User selection case
- Minimum measurements case



Network with v_1 correlated with v_3 and v_6.
v_4 correlated with v_5.

Full input case

We include all in-neighbors of the predicted outputs as predictor inputs

Maximum use of information in signals

$$w_{\mathcal{D}} = \{2,3,4\} \quad w_{\mathcal{Y}} = \{1\}$$

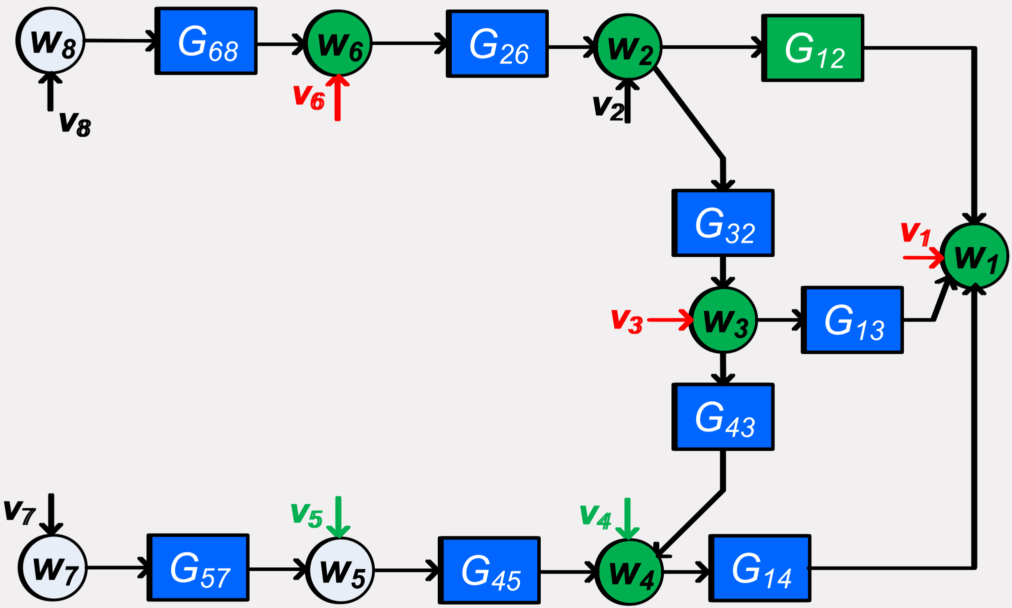
Handling direct confounding variable:

$$w_{\mathcal{D}} = \{2,3,4\} \quad w_{\mathcal{Y}} = \{1,3\}$$

Handling indirect confounding variable:

$$w_{\mathcal{D}} = \{2,3,4,6\} \quad w_{\mathcal{Y}} = \{1,3\}$$

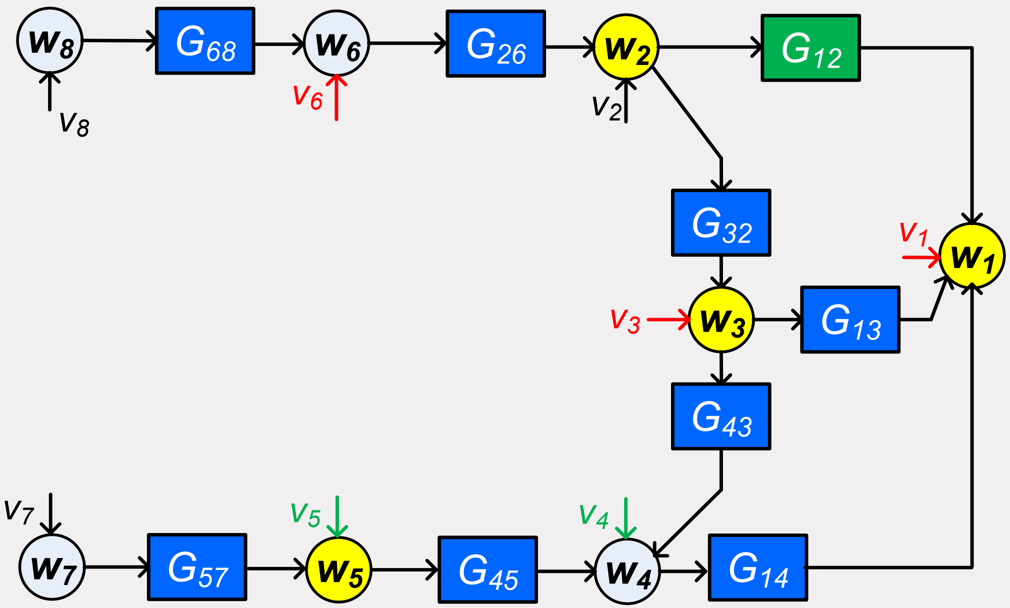
Direct identification $w_{\mathcal{D}} \rightarrow w_{\mathcal{Y}}$



User selection case

- The user does not have access to all node signals
- Four node signals can be measured
- Parallel path and loop condition is satisfied
- Start with:

$$\mathcal{W}_{\mathcal{D}} = \{2,3\} \quad \mathcal{W}_{\mathcal{Y}} = \{1\}$$



User selection case

$$w_{\mathcal{D}} = \{2,3\} \quad w_{\mathcal{Y}} = \{1\}$$

Handling direct confounding variable:

$$w_{\mathcal{D}} = \{2,3\} \quad w_{\mathcal{Y}} = \{1,3\}$$

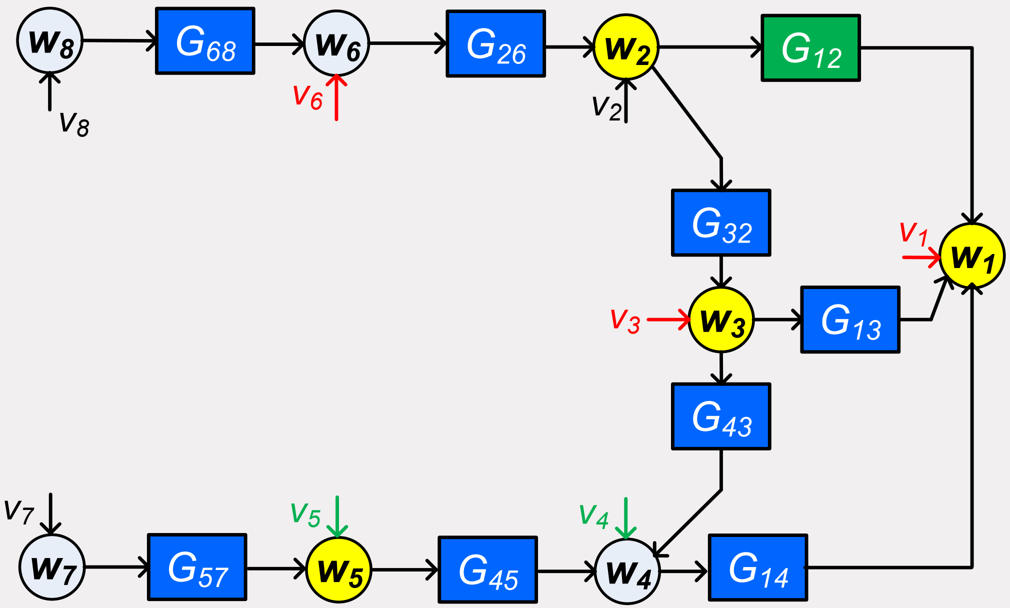
Indirect confounding variables:

(v_4, v_5):

$$w_{\mathcal{D}} = \{2,3\} \quad w_{\mathcal{Y}} = \{1,3,5\}$$

v_6:

$$w_{\mathcal{D}} = \{2,3\} \quad w_{\mathcal{Y}} = \{1,2,3,5\}$$



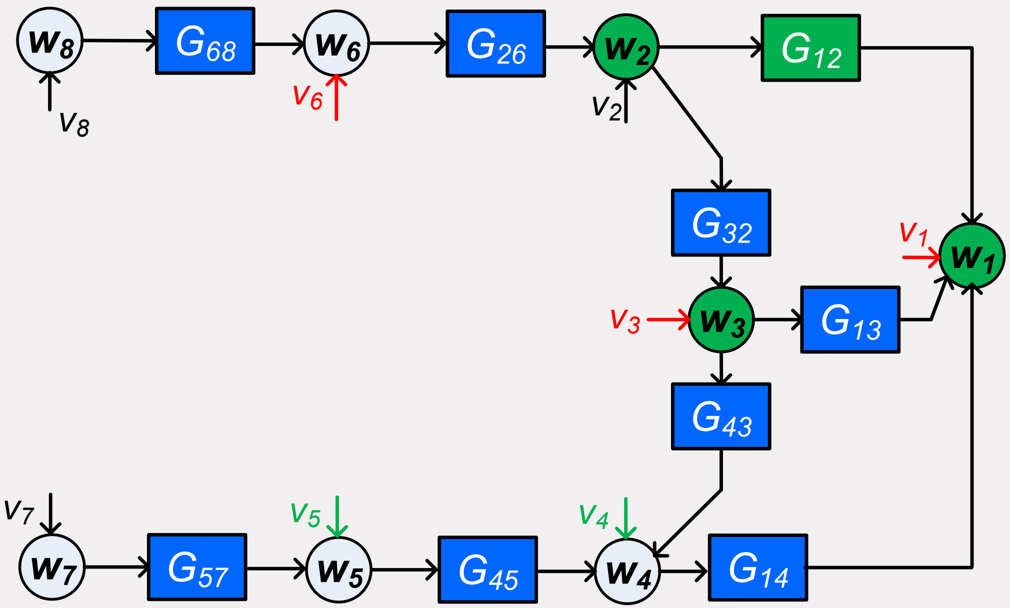
Direct identification $w_{\mathcal{D}} \rightarrow w_{\mathcal{Y}}$

TU/e

Minimum measurements case

- Select signals to satisfy the parallel path and loop condition
- Handle all confounding variables by including signals in output

$$w_{\mathcal{D}} = \{2, 3\} \quad w_{\mathcal{Y}} = \{1, 2, 3\}$$



Direct identification $w_{\mathcal{D}} \rightarrow w_{\mathcal{Y}}$

TU/e

Different strategies for same network and target module

Same network with different identification setups that lead to **consistent estimate of the target module** with **Maximum likelihood properties** based on the strategy used.

Full input case	User selection case	Minimum measurements case
$\begin{bmatrix} w_2 \\ w_3 \\ w_4 \\ w_6 \end{bmatrix} \rightarrow \begin{bmatrix} w_1 \\ w_3 \end{bmatrix}$	$\begin{bmatrix} w_2 \\ w_3 \\ w_5 \end{bmatrix} \rightarrow \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_5 \end{bmatrix}$	$\begin{bmatrix} w_2 \\ w_3 \end{bmatrix} \rightarrow \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}$

Data informativity conditions: $dim(r) \geq dim(w_y)$ (see later)

Theory for local direct method (MIMO)

Theory for single module direct method (MIMO)

Separate the node variables of the network into

$$w = \begin{bmatrix} w_{\mathcal{Q}} \\ w_{\mathcal{o}} \\ w_{\mathcal{U}} \\ w_{\mathcal{Z}} \end{bmatrix} = \begin{bmatrix} \textbf{nodes that appear in input and output} \\ \textbf{output of target module, if not present in } w_{\mathcal{Q}} \\ \textbf{nodes that appear only in the input} \\ \textbf{unmeasured nodes} \end{bmatrix}$$

and write the network equations:

$$\begin{bmatrix} w_{\mathcal{Q}} \\ w_{o} \\ w_{\mathcal{U}} \\ w_{\mathcal{Z}} \end{bmatrix} = \begin{bmatrix} G_{\mathcal{Q}\mathcal{Q}} & G_{\mathcal{Q}o} & G_{\mathcal{Q}\mathcal{U}} & G_{\mathcal{Q}\mathcal{Z}} \\ G_{o\mathcal{Q}} & G_{oo} & G_{o\mathcal{U}} & G_{o\mathcal{Z}} \\ G_{\mathcal{U}\mathcal{Q}} & G_{\mathcal{U}o} & G_{\mathcal{U}\mathcal{U}} & G_{\mathcal{U}\mathcal{Z}} \\ G_{\mathcal{Z}\mathcal{Q}} & G_{\mathcal{Z}o} & G_{\mathcal{Z}\mathcal{U}} & G_{\mathcal{Z}\mathcal{Z}} \end{bmatrix} \begin{bmatrix} w_{\mathcal{Q}} \\ w_{o} \\ w_{\mathcal{U}} \\ w_{\mathcal{Z}} \end{bmatrix} + R(q)r + \begin{bmatrix} H_{\mathcal{Q}\mathcal{Q}} & H_{\mathcal{Q}o} & H_{\mathcal{Q}\mathcal{U}} & H_{\mathcal{Q}\mathcal{Z}} \\ H_{o\mathcal{Q}} & H_{oo} & H_{o\mathcal{U}} & H_{o\mathcal{Z}} \\ H_{\mathcal{U}\mathcal{Q}} & H_{\mathcal{U}o} & H_{\mathcal{U}\mathcal{U}} & H_{\mathcal{U}\mathcal{Z}} \\ H_{\mathcal{Z}\mathcal{Q}} & H_{\mathcal{Z}o} & H_{\mathcal{Z}\mathcal{U}} & H_{\mathcal{Z}\mathcal{Z}} \end{bmatrix} \begin{bmatrix} e_{\mathcal{Q}} \\ e_{o} \\ e_{\mathcal{U}} \\ e_{\mathcal{Z}} \end{bmatrix}$$

Then remove node variables $w_{\mathcal{Z}}$ from the equations through immersion

Theory for single module direct method (MIMO)

Upon immersing node variables $w_{\mathcal{Z}}$ there exists a system transform into the equivalent network representation

$$\underbrace{\left[\begin{array}{c} w_{\mathcal{Q}} \\ w_{o} \\ \hline w_{\mathcal{U}} \end{array}\right]}_{w_m} = \underbrace{\left[\begin{array}{c|c} \bar{G} & 0 \\ \hline \bar{G}_{\mathcal{U}\mathcal{D}} & \bar{G}_{\mathcal{U}o} \end{array}\right]}_{\bar{G}_m} \left[\begin{array}{c} w_{\mathcal{Q}} \\ w_{\mathcal{U}} \\ \hline w_{o} \end{array}\right] + \underbrace{\left[\begin{array}{c|c} \bar{H} & 0 \\ \hline 0 & \bar{H}_{\mathcal{U}} \end{array}\right]}_{\bar{H}_m} \underbrace{\left[\begin{array}{c} \xi_{\mathcal{Q}} \\ \xi_{o} \\ \hline \xi_{\mathcal{U}} \end{array}\right]}_{\xi_m}$$

with ξ_m a white noise process, and $\bar{H}_m$ monic, stable and stably invertible.

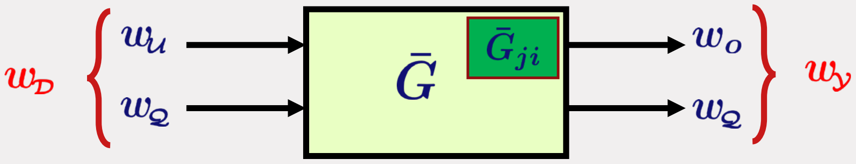
Showing that disturbances on inputs and outputs can be decoupled.

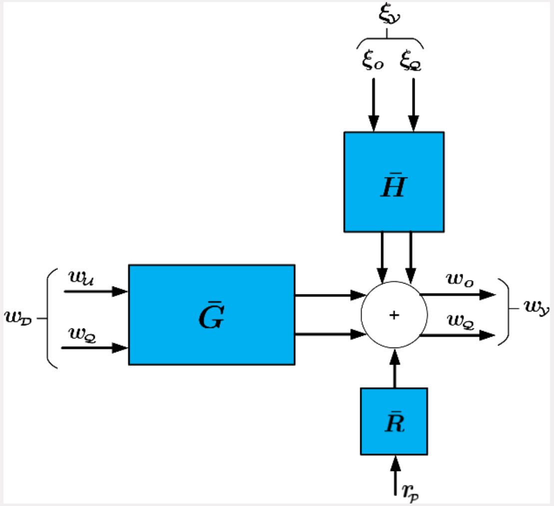
Upper part of the equation leads to:

$$\underbrace{\begin{bmatrix} w_{\mathcal{Q}} \\ w_{o} \end{bmatrix}}_{w_{\mathcal{Y}}} = \underbrace{\begin{bmatrix} \bar{G}_{\mathcal{Q}\mathcal{Q}} & \bar{G}_{\mathcal{Q}\mathcal{U}} \\ \bar{G}_{o\mathcal{Q}} & \bar{G}_{o\mathcal{U}} \end{bmatrix}}_{\bar{G}} \underbrace{\begin{bmatrix} w_{\mathcal{Q}} \\ w_{\mathcal{U}} \end{bmatrix}}_{w_{\mathcal{D}}} + \underbrace{\begin{bmatrix} \bar{H}_{\mathcal{Q}\mathcal{Q}} & \bar{H}_{\mathcal{Q}o} \\ \bar{H}_{o\mathcal{Q}} & \bar{H}_{oo} \end{bmatrix}}_{\bar{H}} \underbrace{\begin{bmatrix} \xi_{\mathcal{Q}} \\ \xi_{o} \end{bmatrix}}_{\xi_{\mathcal{Y}}}$$

to be used for identification

Local direct method





Target module $\bar{G}_{ji}$ is embedded in (possible) MIMO system

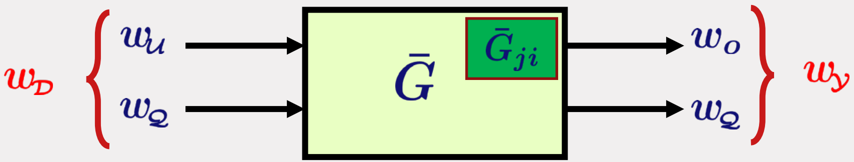
Prediction error: $\varepsilon(t,\theta) = \bar{H}(q,\theta)^{-1}[w_{\mathcal{Y}}(t) - \bar{G}(q,\theta)w_{\mathcal{D}}(t) - \bar{R}r_{\mathcal{P}}(t)]$

[*] Only those r-signals that lead to a constant, non-dynamic, transfer $\bar{\boldsymbol{R}}$ can be handled by a direct method. Other r-signals occur in the disturbance terms.

Quadratic identification criterion: $\hat{\theta}_N := \arg\min_{\theta} \frac{1}{N}\sum_{t=0}^{N-1} \varepsilon(t,\theta)^T Q \varepsilon(t,\theta) \qquad Q > 0$

[1] K.R. Ramaswamy et al., IEEE-TAC, 2021.
[2] VdH et al., CDC-2020.

Consistency result



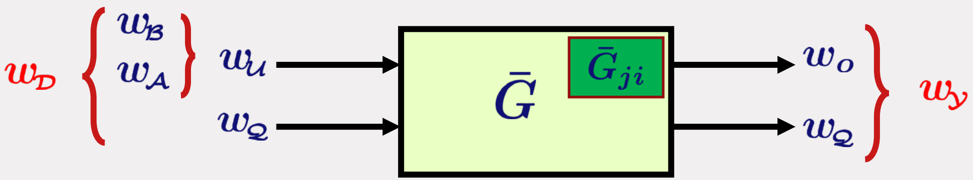
$G_{ji}(q, \hat{\theta}_N)$ is a consistent estimate of G^0_{ji}, if

- $\mathcal{S} \in \mathcal{M}$
- The parallel path and loop condition is satisfied
- A technical condition on presence of delays is satisfied
- Confounding variable conditions are satisfied
- Data set is informative with respect to $\mathcal{M}$

According to PEM/ML theory, the estimator can achieve the CRLB

[1] K.R. Ramaswamy et al., IEEE-TAC, 2021.
[2] VdH et al., CDC-2020.

Consistency result



Confounding variable conditions:

- $i \in \mathcal{Q} \cup \mathcal{A}$
- No confouding variables between $w_{\mathcal{A}}$ and $w_{\mathcal{Y}}$
- No confouding variables between $w_{\mathcal{A}}$ and $w_{\mathcal{B}}$
- No unmeasured paths from $\{i, j\}$ to $w_{\mathcal{B}}$

These conditons can always be satisfied by appropriate choices of $w_{\mathcal{A}}, w_{\mathcal{B}}, w_{\mathcal{Q}}$ and influence the selection of the predictor model

Data informativity (classical definition)

Predictor model: $\hat{w}_{\mathcal{Y}}(t,\theta) = W(q,\theta)z(t)$

for a model set $\mathcal{M} := (\bar{G}(q,\theta), \bar{H}(q,\theta), \bar{R})_{\theta\in\Theta}$ with $z(t) := \begin{bmatrix} w_{\mathcal{Y}}(t) \\ w_{\mathcal{D}}(t) \\ r_{\mathcal{P}}(t) \end{bmatrix}$

Then a quasi-stationary data sequence $\{z(t)\}_{t=0,\dots}$ is informative with respect to $\mathcal{M}$ if for any two models in $\mathcal{M}$:

$$\bar{\mathbb{E}}[(W_1(q) - W_2(q))z(t)]^2 = 0 \implies W_1(e^{i\omega}) \equiv W_2(e^{i\omega})$$

A sufficient condition for this is that z is persistently exciting:

$\Phi_z(\omega) > 0$ for almost all ω

[1] L. Ljung, Englewood Cliffs, NJ: Prentice-Hall, 1999

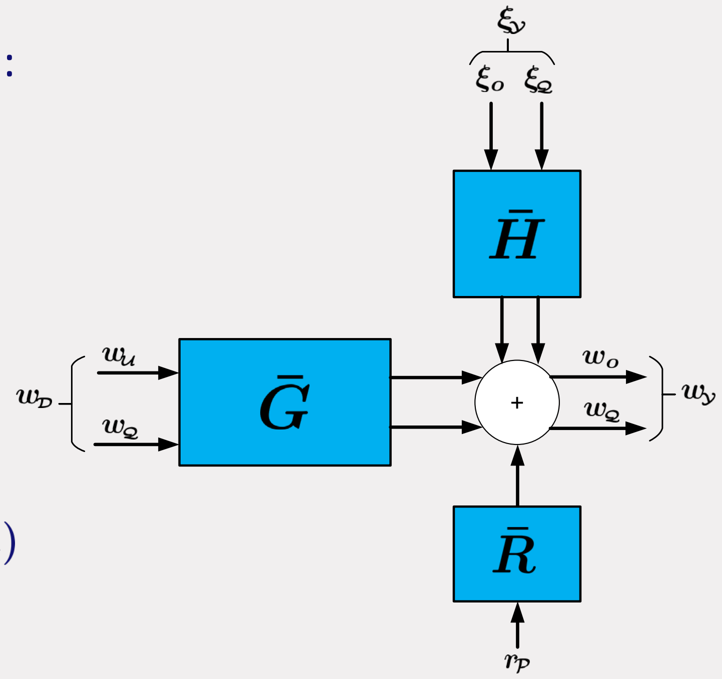
Data informativity - network case

In our situation of specific predictor model:

$\boxed{\Phi_\kappa(\omega) > 0 \text{ for almost all } \omega}$

$$\kappa(t) := \begin{bmatrix} w_{\mathcal{D}}(t) \\ \xi_{\mathcal{Y}}(t) \end{bmatrix}$$

- Note that κ is a filtered version of (r, e)
- with (r, e) persistently exciting

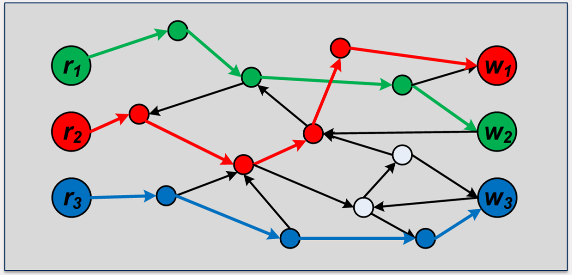


TU/e

Data informativity (path-based condition)

A signal $y(t) = F(q)x(t)$ with x persistently exciting, is persistently exciting iff F has **full row rank**.

This condition can be verified in a generic sense, by considering the **generic rank** of F [1],[2]



$b_{\mathcal{R}\to\mathcal{W}} = 3$

linking to the maximum number of **vertex disjoint paths** between inputs and outputs

κ is persistently exciting holds **generically** if there are $|\mathcal{D}| + |\mathcal{Y}|$ **vertex disjoint paths** between external signals (r, e) and $\kappa = \begin{bmatrix} w_{\mathcal{D}} \\ \xi_{\mathcal{Y}} \end{bmatrix}$

Rework the conditions, since $\xi_{\mathcal{Y}}$ is also a (filtered) external signal (white noise)

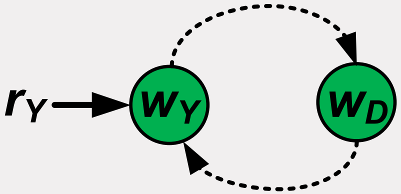
[1] Van der Woude, 1991
[2] Hendrickx, Gevers & Bazanella, CDC 2017, TAC 2019.

External signals on original network

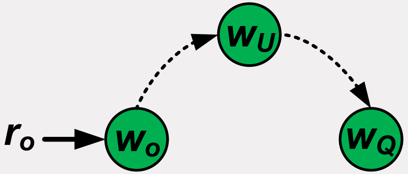
Final result:

Generic data informativity check becomes:
$|\mathcal{D}|$ vertex disjoint paths between external signals $(r_{\mathcal{P}}, x_{\mathcal{U}})$ and $w_{\mathcal{D}}$

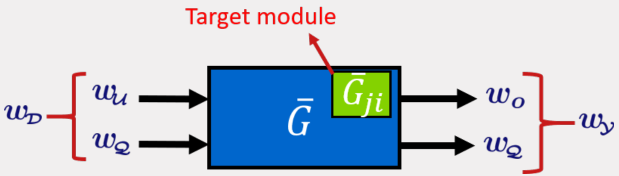
Signals in $r_{\mathcal{P}} \in r_{\mathcal{Y}}$



Loops through $w_{\mathcal{Y}}$
pass through $w_{\mathcal{D}}$



Path from w_o to $w_{\mathcal{Q}}$
passes through $w_{\mathcal{U}}$



Signals in $x_{\mathcal{U}}$: All external signals (r, e) that have a direct or unmeasured path to $w_{\mathcal{U}}$

Example

Target: identify G_{21}

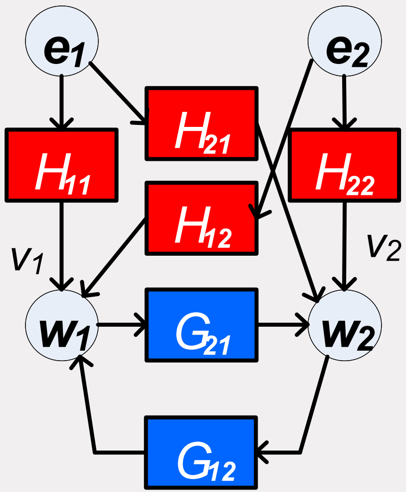
Predictor model: $\underbrace{\{w_1\}}_{w_{\mathcal{D}}} \rightarrow \underbrace{\{w_1, w_2\}}_{w_{\mathcal{Y}}}$

2 x 2 noise model accounts for confounding variable

$w_{\mathcal{Q}} = \{w_1\} \quad w_{\mathcal{U}} = \emptyset \quad x_{\mathcal{U}} = \emptyset$

Situation 1: $r_1 = r_2 = 0 \qquad r_{\mathcal{P}} = \emptyset$

There are no external signals available for exciting $w_{\mathcal{D}} = \{w_1\}$



Data-informativity condition NOT satisfied

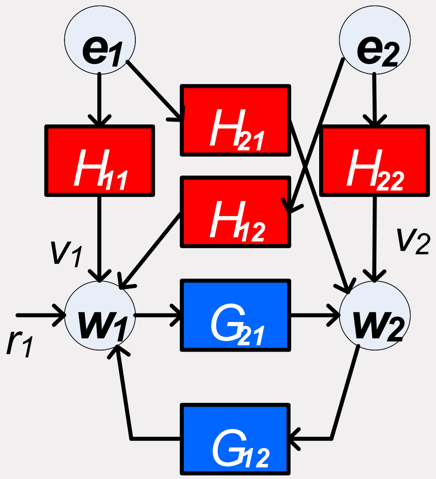
Example

Target: identify G_{21}

Predictor model: $\underbrace{\{w_1\}}_{w_{\mathcal{D}}} \rightarrow \underbrace{\{w_1, w_2\}}_{w_{\mathcal{Y}}}$

2 x 2 noise model accounts for confounding variable

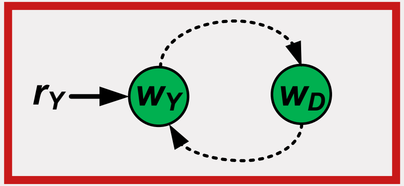
$w_{\mathcal{Q}} = \{w_1\} \quad w_{\mathcal{U}} = \emptyset \quad x_{\mathcal{U}} = \emptyset$



Situation 2: $r_2 = 0; r_1$ **present**

r_1 runs through a loop that does not pass through an input signal $\Rightarrow$ $r_{\mathcal{P}} = \emptyset$

There are no external signals available for exciting $w_{\mathcal{D}} = \{w_1\}$



Data-informativity condition NOT satisfied

TU/e

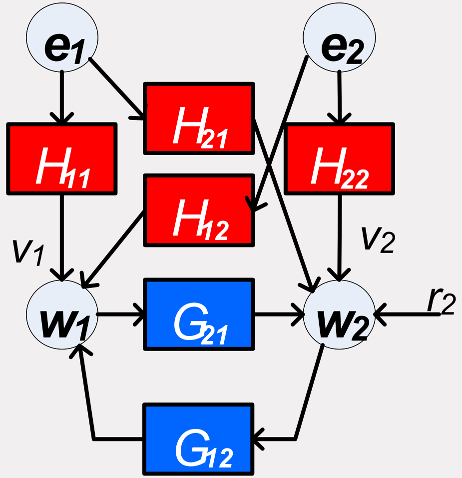
Example

Target: identify G_{21}

Predictor model: $\underbrace{\{w_1\}}_{w_{\mathcal{D}}} \rightarrow \underbrace{\{w_1, w_2\}}_{w_{\mathcal{Y}}}$

2 x 2 noise model accounts for confounding variable

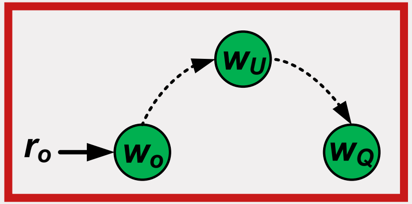
$w_{\mathcal{Q}} = \{w_1\} \quad w_{\mathcal{U}} = \emptyset \quad x_{\mathcal{U}} = \emptyset$



Situation 3: $r_1 = 0; r_2$ **present**

r_2 has a path to $w_{\mathcal{Q}} = \{w_1\}$ that does not pass through $w_{\mathcal{U}}$ ➡ $r_{\mathcal{P}} = \emptyset$

There are no external signals available for exciting $w_{\mathcal{D}} = \{w_1\}$



Data-informativity condition NOT satisfied

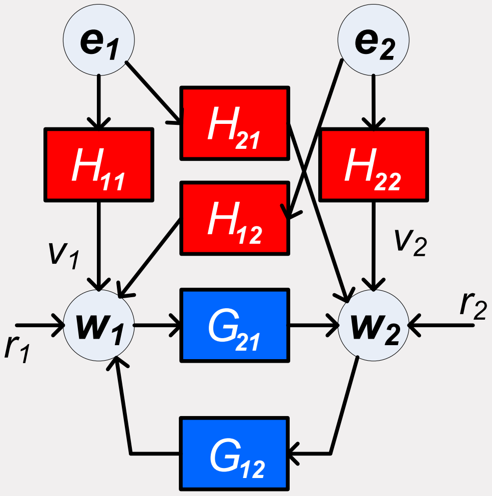
Example

Target: identify G_{21}

Predictor model: $\boxed{\underbrace{\{w_1, w_2\}}_{w_{\mathcal{D}}} \rightarrow \underbrace{\{w_1, w_2\}}_{w_{\mathcal{Y}}}}$

2 x 2 noise model accounts for confounding variable

$w_{\mathcal{Q}} = \{w_1, w_2\} \quad w_{\mathcal{U}} = \emptyset \quad x_{\mathcal{U}} = \emptyset$



Situation 3: r_1, r_2 **present**

r_1, r_2 satisfies the conditions for $r_{\mathcal{P}}$ ➡ $r_{\mathcal{P}} = \{r_1, r_2\}$

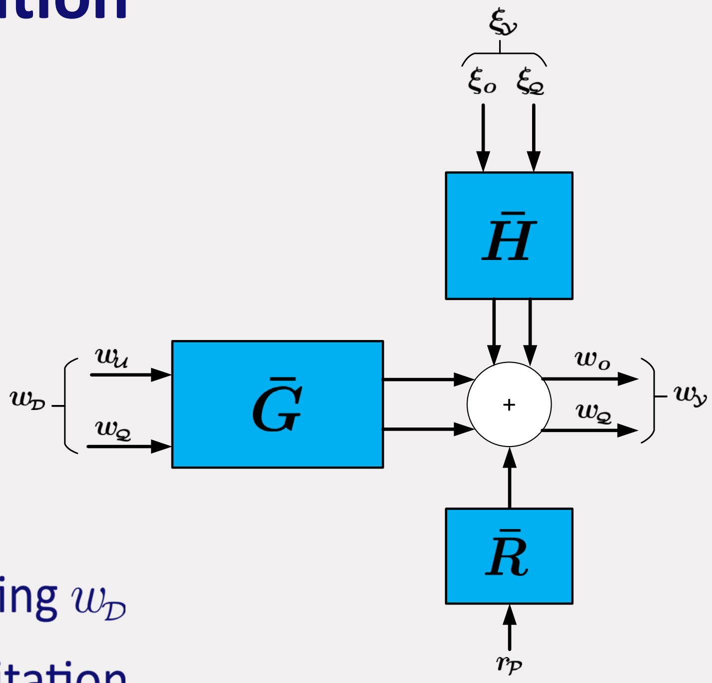
Since there are two vertex-disjoint paths from r_1, r_2 to $w_{\mathcal{D}} = \{w_1, w_2\}$

Data-informativity satisfied !

Data informativity - interpretation

$\boxed{\Phi_\kappa(\omega) > 0 \text{ for almost all } \omega}$

$$\kappa(t) := \begin{bmatrix} w_{\mathcal{D}}(t) \\ \xi_{\mathcal{Y}}(t) \end{bmatrix}$$



- Disturbances $\xi_{\mathcal{Y}}$ can not be used for exciting $w_{\mathcal{D}}$
- For every signal in $w_{\mathcal{Q}}$ we need an r-excitation
- More “expensive” experiments with growing # outputs

TU/e

Summary local direct method for single module ID

- Flexible algorithm for selecting measured signals in a predictor model
- that leads to consistent (and minimum variance) module estimates
- Verifiable conditions on the network topology (assumed a priori known)
- Path-based conditions also for (generic) data informativity
- For the actual identification algorithm: preferably regularized techniques
- Extensions:
  - effective use of r-signals can further relax the conditions for signal selection[1]
  - include topology estimation as a first step[2]

[1] Ramaswamy, VdH, Dankers, CDC 2019.
[2] Rajagopal, MSc, November 2020; CDC 2021 submitted.

Identifiability and data informativity

- For a **particular** identification method:
  Consistency conditions include aspects of data-informativity
  and underlying conditions of identifiability (implicitly)

- Current consistency conditions can be split in (a) identifiability conditions and (b) data informativity conditions

- Network identifiability is identification method - independent
  Reflects choice of predictor model:

  – presence and location of excitation and disturbance signals
  – parametrized model set (fixed modules and disturbance correlations)