Identification of linear dynamic networks with reduced-rank noise

Paul M.J. Van den Hof

co-authors: Arne Dankers and Harm Weerts

CARMA Workshop, University of Newcastle, Australia, 7 December 2017

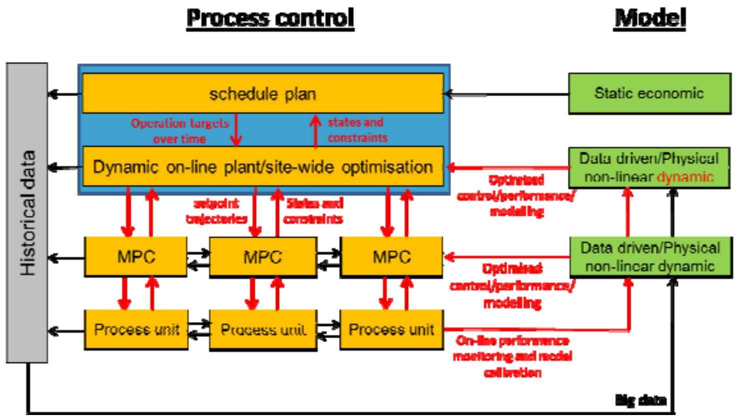
European Research Council

TU/e Technische Universiteit Eindhoven University of Technology

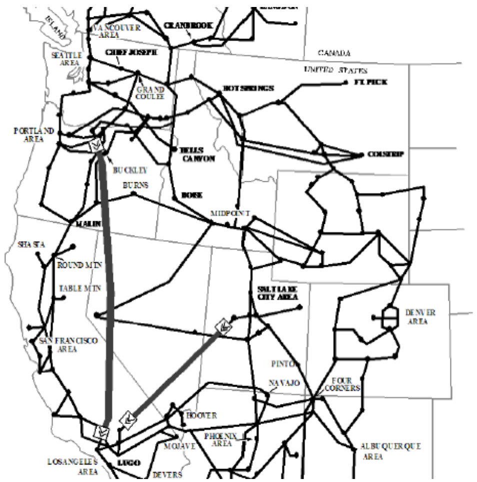
Where innovation starts

Introduction – dynamic networks

Decentralized process control

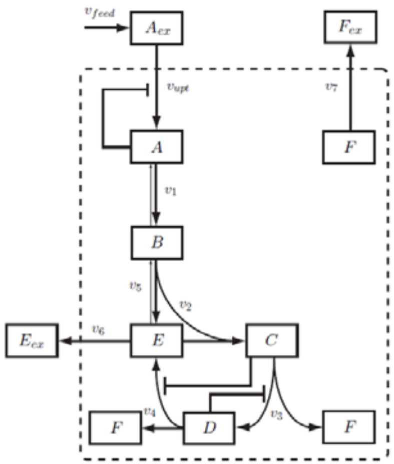


Power grid



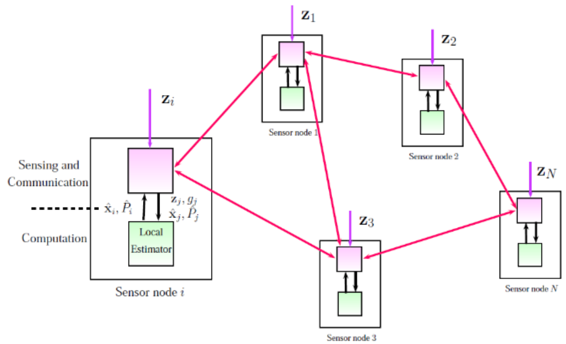
Pierre et al. (2012)

Metabolic network



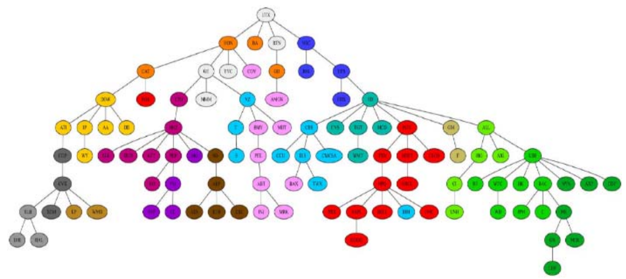
Hillen (2012)

Distributed control (robotic networks)



Simonetto (2012)

Stock market

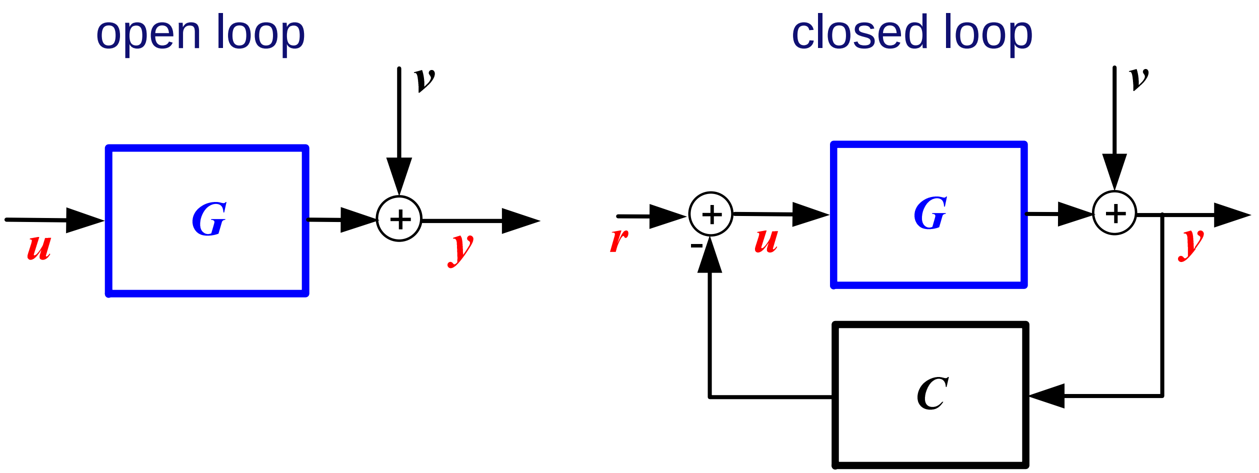


Materassi et al. (2010)

Technische Universiteit Eindhoven University of Technology

Introduction – identification

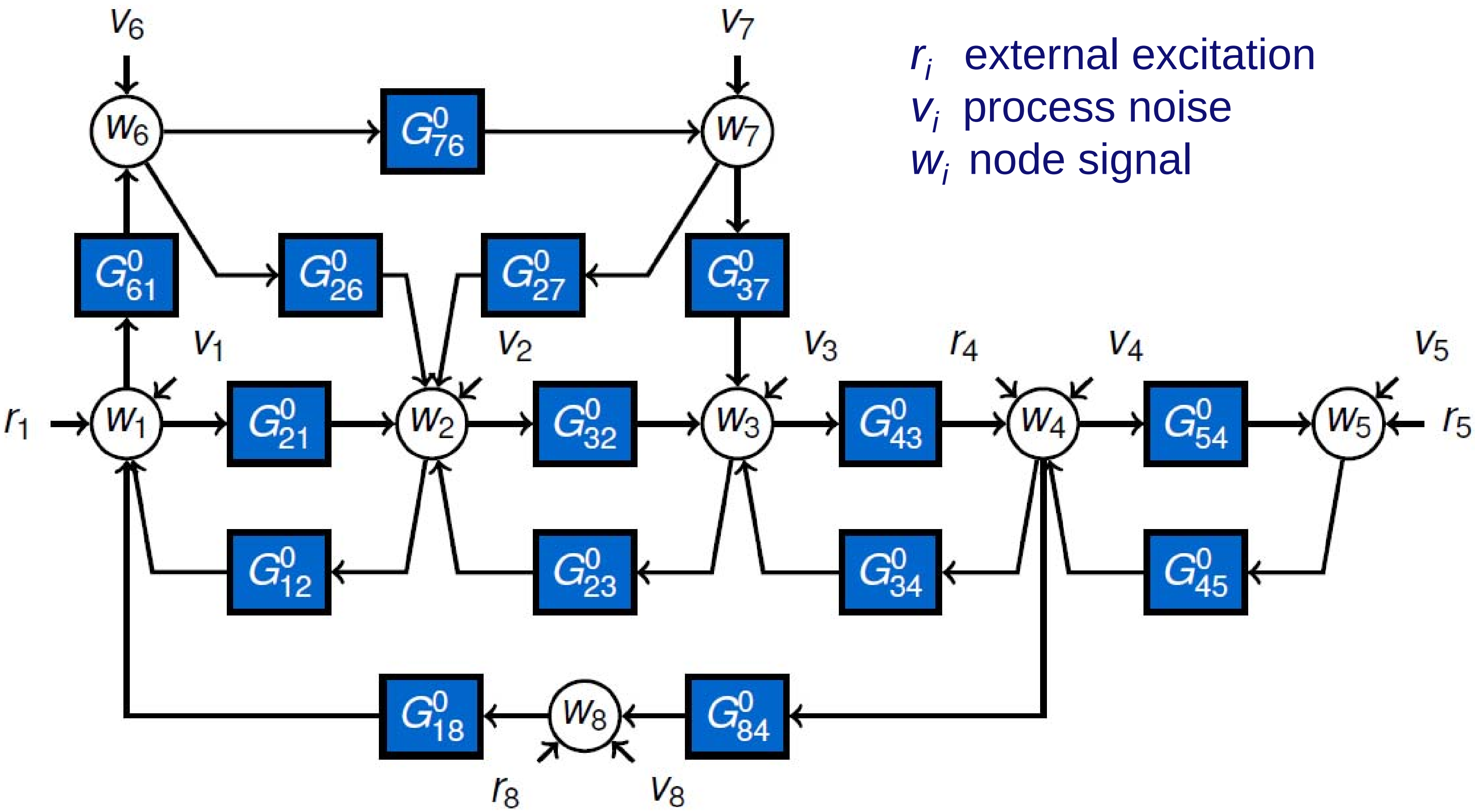
The classical (multivariable) identification problems: [Ljung (1999)]



Identify a plant model $\hat{G}$ on the basis of measured signals *u, y* *(*and possibly *r)*

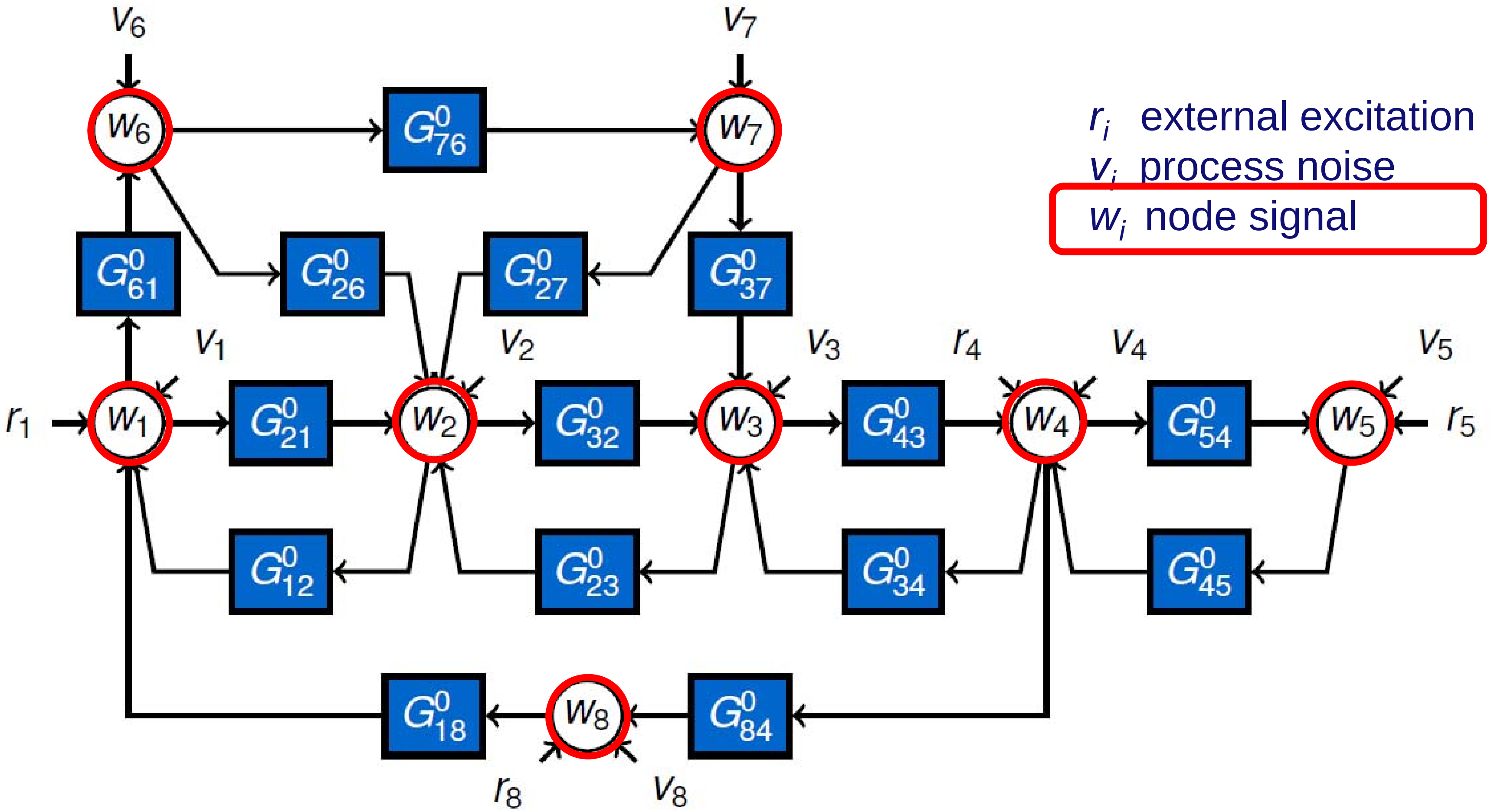
We have to move from a fixed and known configuration to deal with and exploit ***structure*** in the problem.

Dynamic network: what is it?



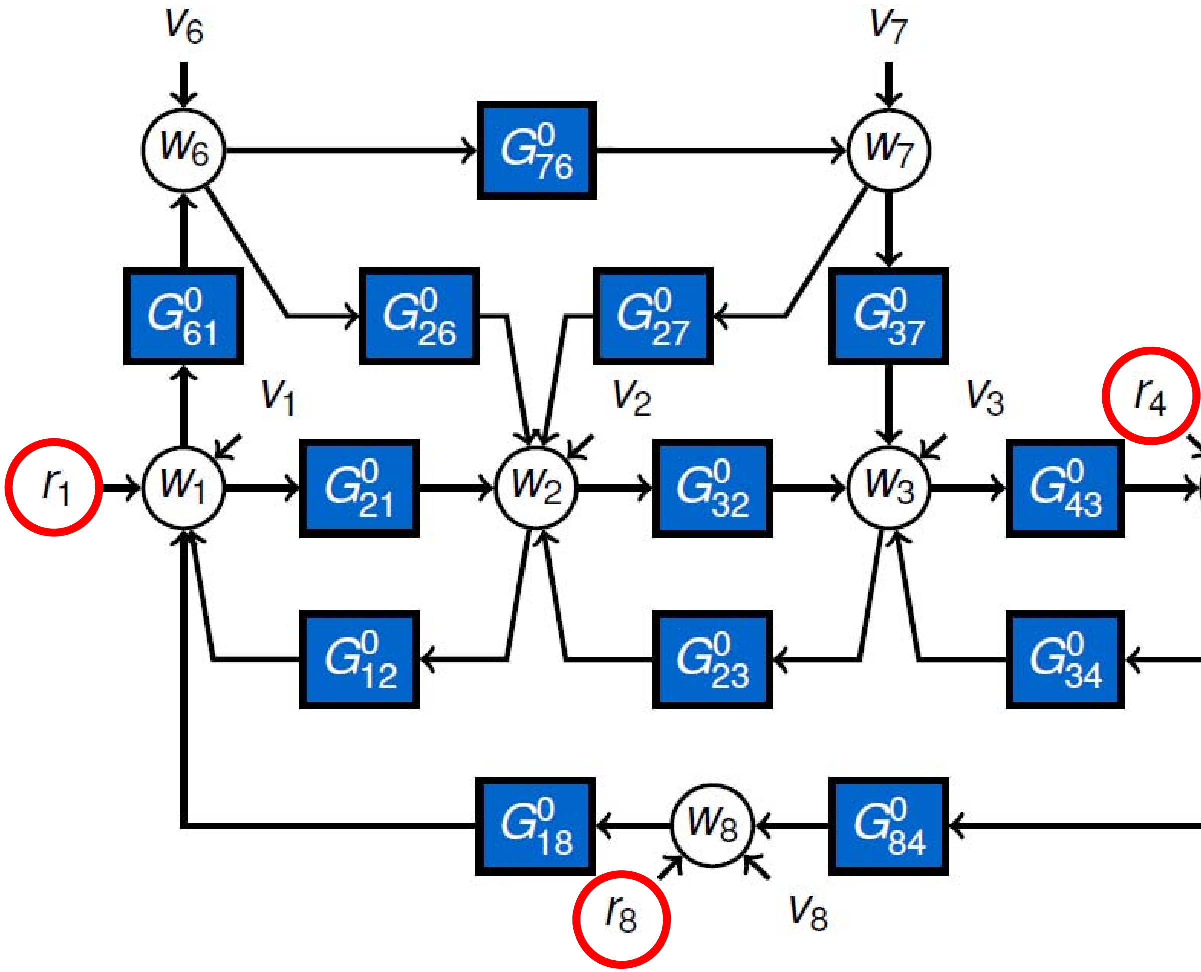
r_i external excitation
v_i process noise
w_i node signal

Dynamic network: what is it?



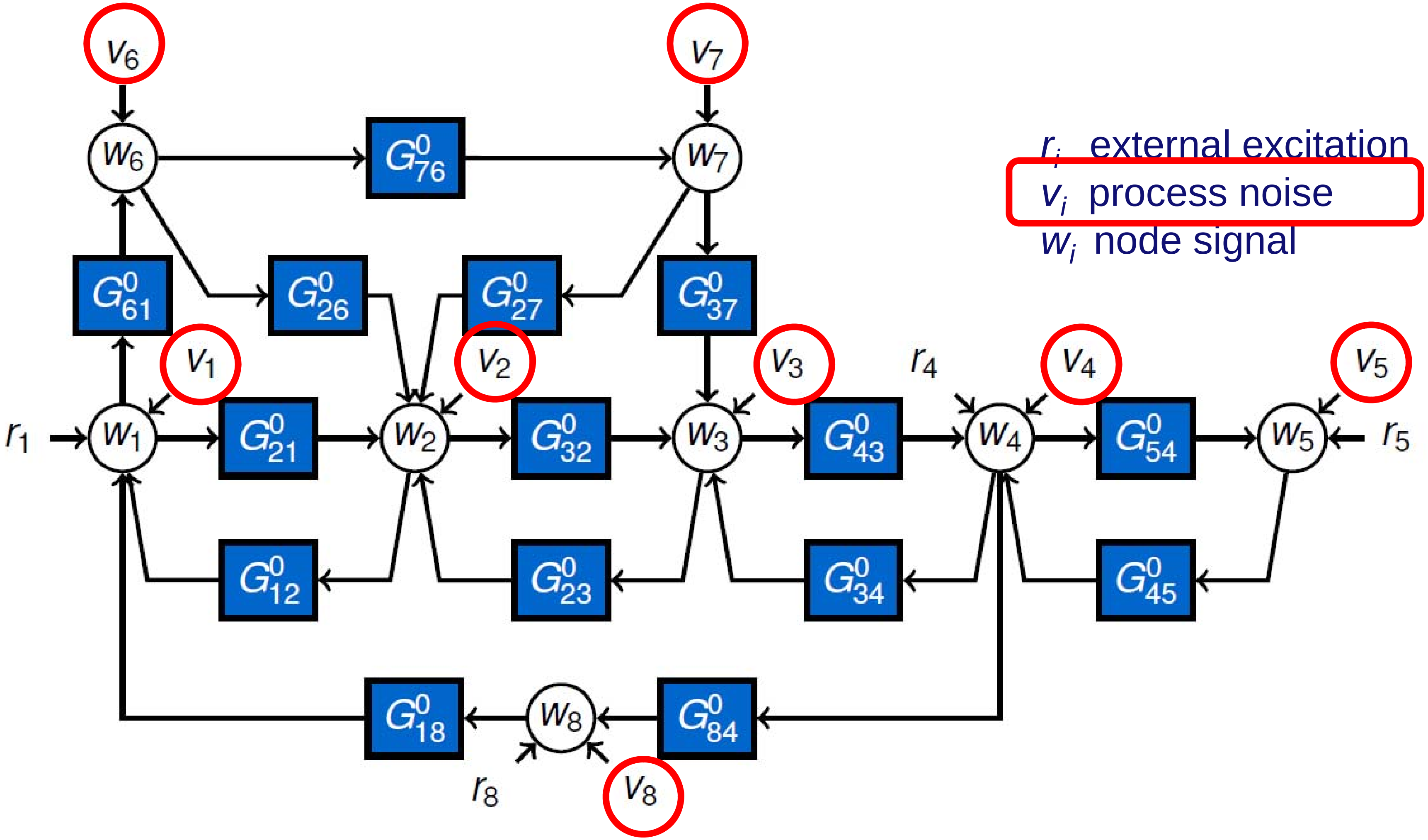
r_i external excitation
v_i process noise
w_i node signal

Dynamic network: what is it?



r_i external excitation
v_i process noise
w_i node signal

Dynamic network: what is it?

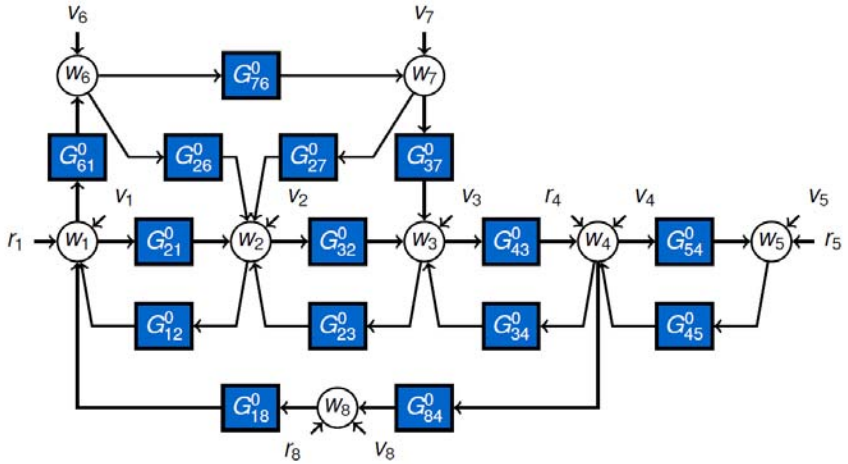


r_i external excitation

v_i process noise

w_i node signal

Introduction



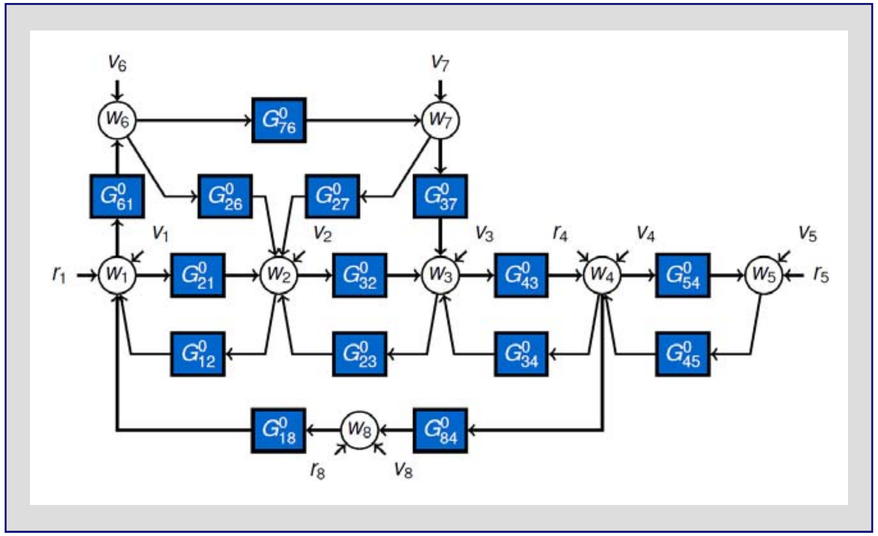
r_i external excitation
v_i process noise
w_i node signal

$$v(t) = \begin{bmatrix} v_1(t) \\ \vdots \\ v_L(t) \end{bmatrix}$$

What are assumptions on process noises when identifying (parts of) a network?

- Independent white noise processes
- Vector stochastic process with full rank spectrum, $rank\ \Phi_v(\omega) = L\ \ a.e.$ leading to a square noise model: $v(t) = H(q)e(t)$
- If $dim(e) < L$ then we have “singular” or “reduced-rank” noise

Network Setup



Assumptions:

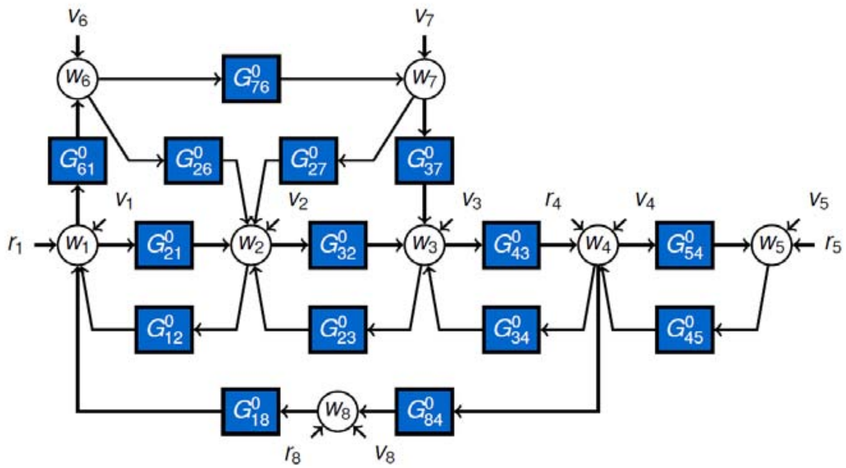
- Total of *L* nodes
- Network is well-posed and stable
- Modules may be unstable
- Node signals and excitation signals can be measured

$$\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_L \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & G^0_{12} & \cdots & G^0_{1L} \\ G^0_{21} & 0 & \ddots & G^0_{2L} \\ \vdots & \ddots & \ddots & \vdots \\ G^0_{L1} & G^0_{L2} & \cdots & 0 \end{bmatrix}}_{G^0} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_L \end{bmatrix} + R^0(q) \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_K \end{bmatrix} + \underbrace{H^0(q) \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_p \end{bmatrix}}_{v}$$

$$w = G^0 w + R^0 r + H^0 e$$

- P.M.J. Van den Hof, A.G. Dankers, P.S.H. Heuberger and X. Bombois. *Automatica, 49*, 2994-3006, 2013.

Introduction



r_i external excitation
v_i process noise
w_i node signal

$$\begin{bmatrix} v_1(t) \\ \vdots \\ v_L(t) \end{bmatrix} = H^0(q) \begin{bmatrix} e_1(t) \\ \vdots \\ e_p(t) \end{bmatrix}$$

Main question:
How to identify (parts of) a dynamic network,
when the process noise is of reduced rank *(p < L)*?

Contents

- Modelling a reduced-rank stochastic process
- Multi-output identification in a dynamic network
  the joint-direct method with weighted LS
- Constrained LS and maximum likelihood estimation
- Variance-free estimation, minimum variance and the CRLB
- Simulation example

Modelling reduced rank noise

Assumption

The node signals w_j are ordered in such a way that the first *p* noise components $v_j, \;\; j = 1, \cdots p$ constitute a full rank process.

Modelling reduced rank noise

A reduced-rank stochastic process $\boldsymbol{v}$ with dimension *L* and rank *p* can equivalently be described in two ways:

a) $\boldsymbol{v(t) = \check{H}^0(q)\check{e}(t)}$

With $\boldsymbol{\check{H}^0} \in \mathbb{R}^{\boldsymbol{L\times L}}\boldsymbol{(z)},\ \boldsymbol{\check{e}(t)} \in \mathbb{R}^{\boldsymbol{L}}$ a white noise process,
$\boldsymbol{\check{H}^0}$ stable, stably invertible, and monic, and
$\boldsymbol{cov(\check{e}) = \check{\Lambda}^0}$ having rank *p*

b) $\boldsymbol{v(t) = H^0(q)e(t)}$

With $\boldsymbol{H^0} \in \mathbb{R}^{\boldsymbol{L\times p}}\boldsymbol{(z)},\ \boldsymbol{e(t)} \in \mathbb{R}^{\boldsymbol{p}}$ a white noise process,
$\boldsymbol{H^0} = \begin{bmatrix} \boldsymbol{H_a^0} \\ \boldsymbol{H_b^0} \end{bmatrix}$ with $\boldsymbol{H_a^0}$ square, stable, stably invertible, and monic,
$\boldsymbol{cov(e) = \Lambda^0}$ having full rank *p*

Modelling reduced rank noise

Relations between descriptions:

$$v(t) = \check{H}^0(q)\check{e}(t) = \begin{bmatrix} H_a^0(q) & 0 \\ H_b^0(q) - \Gamma^0 & I \end{bmatrix} \begin{bmatrix} e \\ \Gamma^0 e \end{bmatrix}$$

with $\Gamma^0 = \lim_{z \to \infty} H_b^0(z)$

while $\check{\Lambda}^0 = \begin{bmatrix} I \\ \Gamma^0 \end{bmatrix} \Lambda^0 \begin{bmatrix} I \\ \Gamma^0 \end{bmatrix}^T$ and $\begin{bmatrix} \Gamma^0 & -I \end{bmatrix} \check{e}(t) = 0$

Both noise models $\check{H}^0$ and $H^0 = \begin{bmatrix} H_a^0 \\ H_b^0 \end{bmatrix}$ will be used.

Joint-direct identification method

We follow a prediction error approach, by predicting **all** node variables:

$$\hat{w}(t|t-1) := \mathbb{E}\left\{w(t) \mid w^{t-1},\ r^{t}\right\}$$

Then: $$\hat{w}(t|t-1) = W_w^0(q)w(t) + W_r^0(q)r(t)$$

with:
$$\begin{aligned} W_w^0(q) &= I - (\check{H}^0(q))^{-1}(I - G^0(q)), \\ W_r^0(q) &= (\check{H}^0(q))^{-1}R^0(q). \end{aligned}$$

being the unique predictor filters.

Joint-direct identification method

The **network** is defined by: $(G^0, R^0, H^0, \Lambda^0)$

a network model is denoted by: $M = (G, R, H, \Lambda)$

and a **network model set** by:

$$\mathcal{M} = \{M(\theta) = (G(\theta), R(\theta), H(\theta), \Lambda(\theta)), \theta \in \Theta\}$$

Then the parametrized predictor:

$$\hat{w}(t|t-1) = W_w(q,\theta)w(t) + W_r(q,\theta)r(t)$$

leads to the prediction error: $\varepsilon(t,\theta) = w(t) - \hat{w}(t|t-1;\theta)$

Weighted LS criterion:

$$\hat{\theta}_N^{WLS} = \arg\min_{\theta\in\Theta} \frac{1}{N}\sum_{t=1}^{N} \varepsilon^T(t,\theta)\ Q\ \varepsilon(t,\theta) \qquad Q > 0$$

Joint-direct identification method

Weighted LS criterion:

$$\hat{\theta}_N^{WLS} = \arg\min_{\theta\in\Theta} \frac{1}{N}\sum_{t=1}^{N} \varepsilon^T(t,\theta)\; Q\; \varepsilon(t,\theta) \qquad\qquad \boldsymbol{Q > 0}$$

Properties:

- Consistent estimate under regularity conditions,
- Provided model set large enough, appropriate excitation, global network identifiability,
- But for minimum variance an optimal $\boldsymbol{Q}$ has to be chosen

Typical choice, leading to minimum variance estimatorm for $\boldsymbol{Q} \in \mathbb{R}^{\boldsymbol{L\times L}}$

$$\boldsymbol{Q = [cov(\breve{e})]^{-1} = (\breve{\Lambda}^0)^{-1}}$$

but in our situation $\boldsymbol{\breve{\Lambda}^0}$ is singular

Constrained LS and Maximum Likelihood

The WLS estimator does not take account of the dependencies in the innovation:

$$\begin{bmatrix}\boldsymbol{\Gamma}^0 & -\boldsymbol{I}\end{bmatrix} \check{e}(t) = 0$$

or differently formulated:

$$\begin{bmatrix}\boldsymbol{\Gamma}^0 & -\boldsymbol{I}\end{bmatrix} \begin{bmatrix}\varepsilon_a(t,\theta_0) \\ \varepsilon_b(t,\theta_0)\end{bmatrix} = 0$$

This can be imposed, by restricting the parametrized model to satisfy:

$$\underbrace{\Gamma(\theta)\varepsilon_a(t,\theta) - \varepsilon_b(t,\theta)}_{:= Z(t,\theta)} = 0$$

We denote: $:= Z(t,\theta)$

Constrained LS and Maximum Likelihood

Constrained LS criterion:

$$\hat{\theta}_N^{CLS} = \arg\min_{\theta \in \Theta} \frac{1}{N} \sum_{t=1}^{N} \varepsilon_a^T(t,\theta)\, Q_a\, \varepsilon_a(t,\theta) \qquad Q_a > 0$$

$$\text{subject to } \frac{1}{N} \sum_{t=1}^{N} Z^T(t,\theta) Z(t,\theta) = 0$$

Properties:

- Consistent estimate under similar conditions as WLS
- The choice
  $$Q_a = (\Lambda^0)^{-1}$$
  leads to minimum variance, and ML properties in case of Gaussian noise.
- For indendently parametrized $\Lambda(\theta)$, the cost function turns into a determinant function

Constrained LS and Maximum Likelihood

Implementation:

In practice, constraints could be unfeasible, e.g. in case $\mathcal{S} \notin \mathcal{M}$

Constraint relaxation:

$$\hat{\theta}_N^{rel} = \arg\min_{\theta} \frac{1}{N}\sum_{t=1}^{N}\Big(\varepsilon_a^T(t,\theta)Q_a\varepsilon_a(t,\theta) + \lambda Z^T(t,\theta)Z(t,\theta)\Big), \quad \lambda \in \mathbb{R}$$

with tuning parameter $\lambda \in \mathbb{R}$

For $\lambda > 0$ the consistency result remains true.
For $\lambda \to \infty$ constraint satisfaction

The criterion is equivalent to WLS with

$$Q(\theta) = \begin{bmatrix} Q_a + \lambda\Gamma^T(\theta)\Gamma(\theta) & -\lambda\Gamma^T(\theta) \\ -\lambda\Gamma(\theta) & \lambda I \end{bmatrix}$$

Minimum variance and CRLB

Asymptotic criterion:

$$\theta^\star = \arg\min_{\theta\in\Theta} \ \bar{\mathbb{E}}\ \varepsilon^T(t,\theta)\ Q_a\ \varepsilon(t,\theta)$$ subject to

$$\bar{\mathbb{E}} Z(t,\theta) Z^T(t,\theta) = 0$$

When linearizing $Z(t,\theta)$ in the neighbourhood of the optimum:

$$Z(t,\theta) \approx Z(t,\theta^*) + A(t)(\theta-\theta^*)$$

the constrained parameter space can be characterized by

$$\theta = S\rho + C \qquad \rho \in \mathbb{R}^{n_\rho}$$ of reduced dimension

with S, C determined by:

$$\begin{cases} \Pi S &= 0 \\ C &= -\Pi^\dagger \Pi \theta^* \end{cases}$$

and S full rank, where $\bar{\mathbb{E}} A^T(t) A(t) = \Pi^T \Pi$

$\Pi^\dagger$ right inverse

P. Stoica and B.C. Ng, *IEEE Signal Processing Letters, 5* (7), 177-179, 1998.