Prediction error identification with rank-reduced output noise

Paul M.J. Van den Hof

with Harm Weerts (TU/e) and Arne Dankers (Calgary)

2017 American Control Conference, Seattle, WA, 24 May 2017

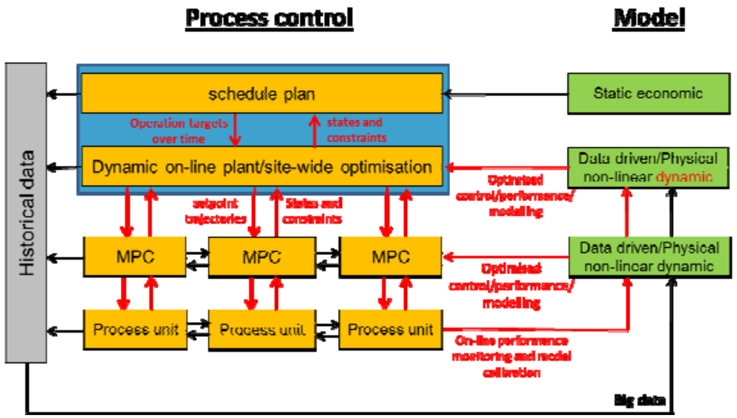
European Research Council

Technische Universiteit Eindhoven University of Technology

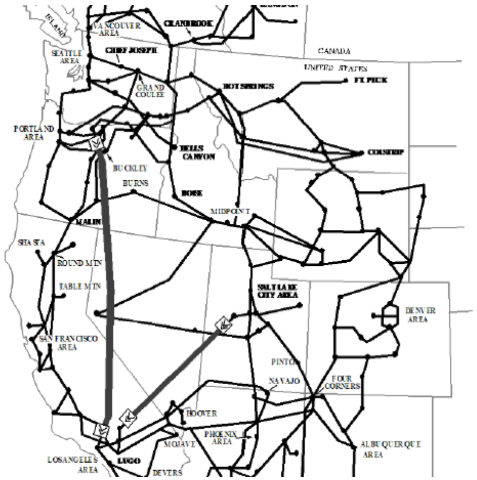
Where innovation starts

Introduction – dynamic networks

Decentralized process control

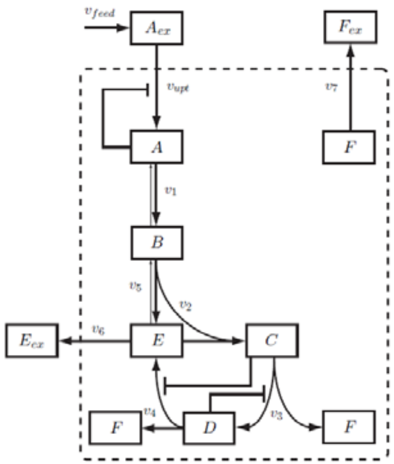


Power grid



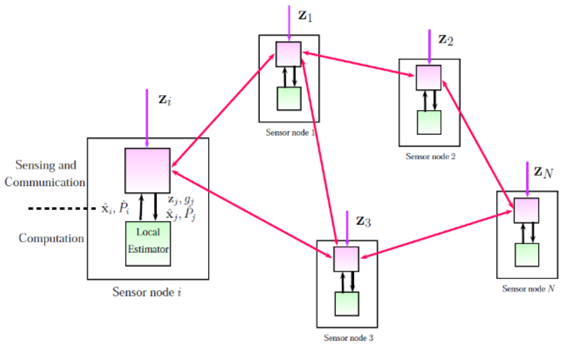
Pierre et al. (2012)

Metabolic network



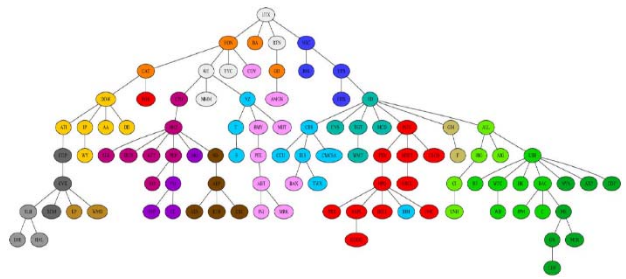
Hillen (2012)

Distributed control (robotic networks)



Simonetto (2012)

Stock market

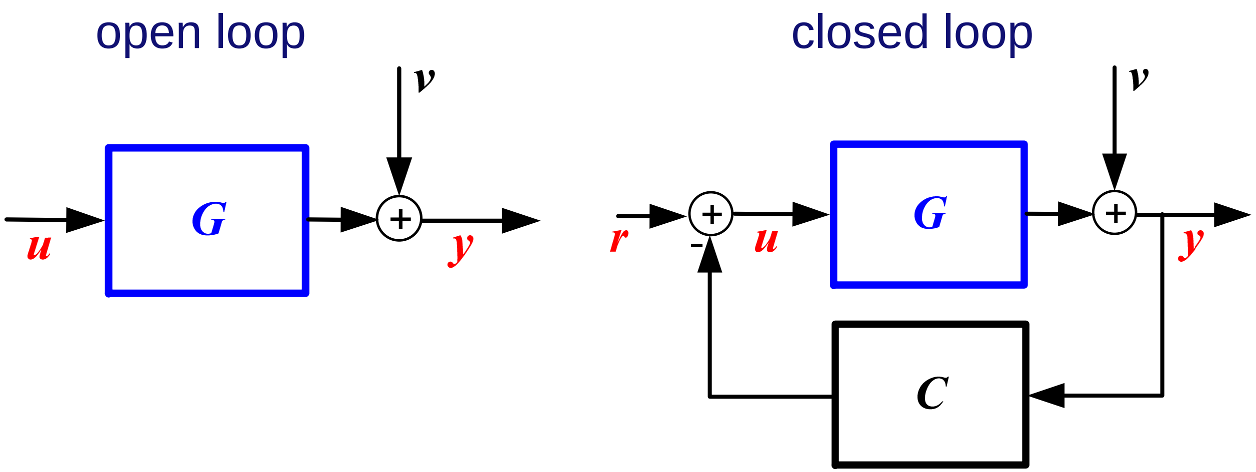


Materassi et al. (2010)

Technische Universiteit Eindhoven University of Technology

Introduction - identification

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

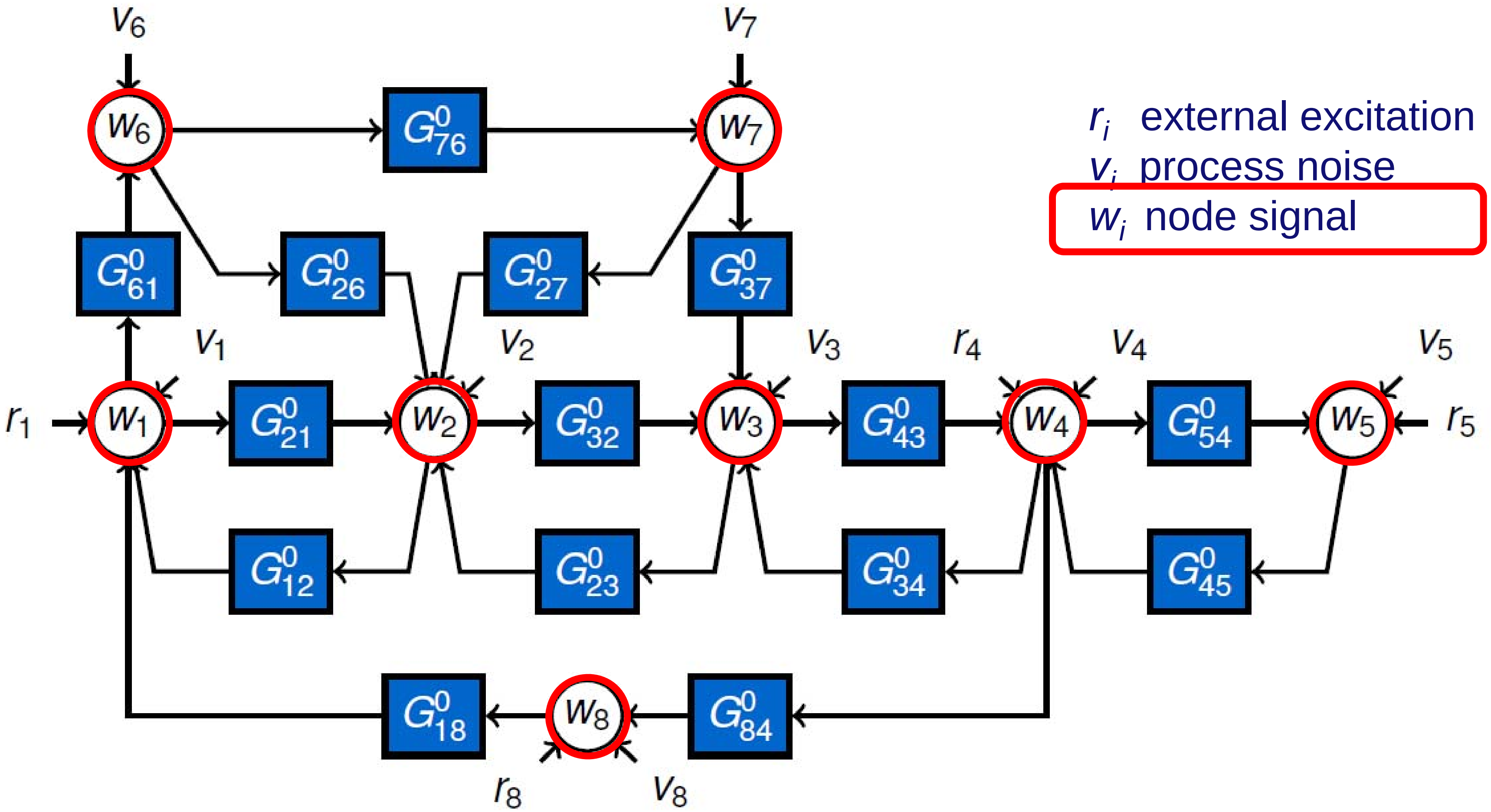


Identify a plant model $\hat{G}$ on the basis of measured signals *u*, *y* (and possibly *r*)

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

→ Dynamic networks

Introduction – Dynamic network identification



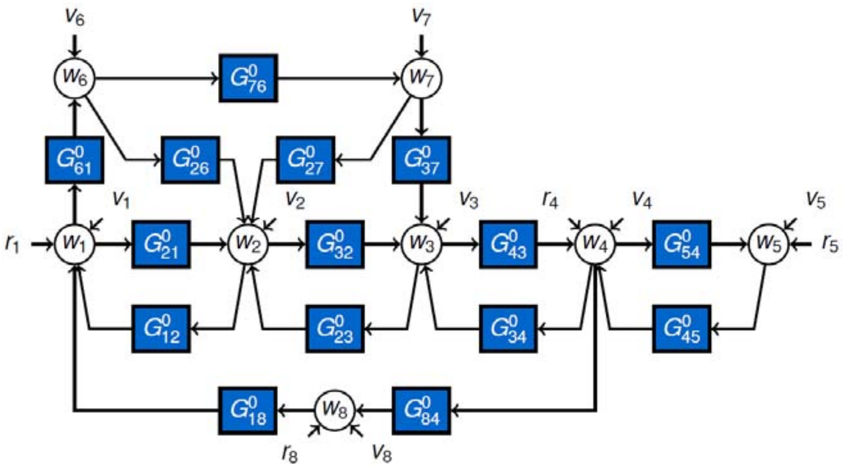
r_i external excitation

v_i process noise

w_i node signal

[P.M.J. Van den Hof, A. Dankers, P.S.C. Heuberger and X. Bombois. *Automatica*, October 2013]

Introduction – Dynamic network identification



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?

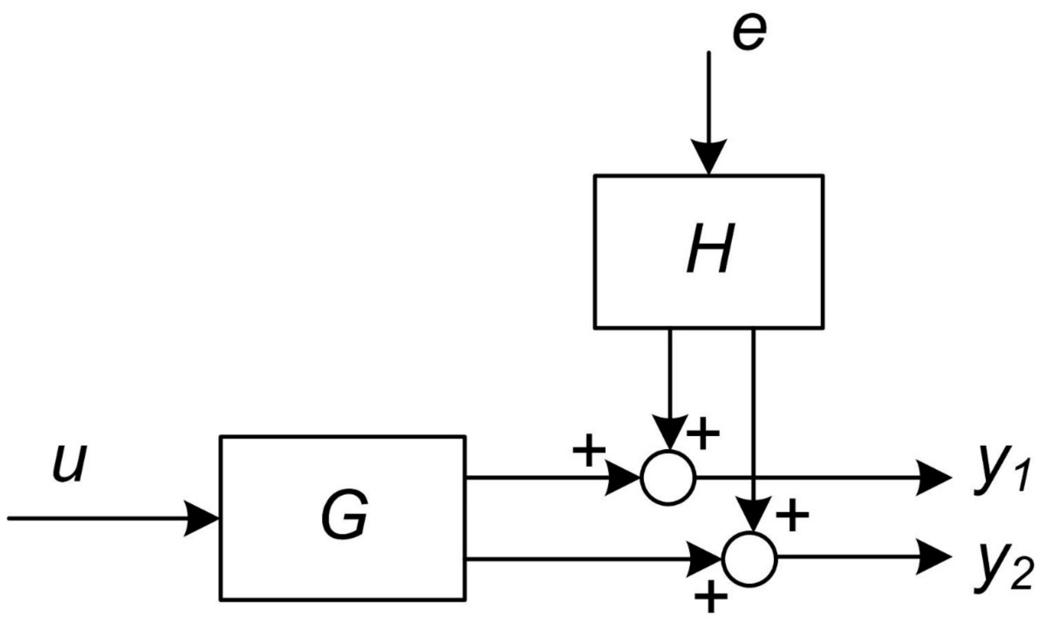
- Independent white noise processes
- Vector stochastic process with full rank spectrum, $rank\ \Phi_v(\omega) = L \ \ a.e.$ leading to a square noise model:

$$v(t) = H(q)e(t)$$

- If $dim(e) < L$ then we have “singular” or “reduced-rank” noise

Identification with reduced-rank noise

The simplest case to consider:



$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = G(q)u(t) + \begin{bmatrix} H_1(q) \\ H_2(q) \end{bmatrix} e(t)$$

Identification with reduced-rank noise

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = G(q)u(t) + \begin{bmatrix} H_1(q) \\ H_2(q) \end{bmatrix} e(t) \qquad \lim_{z\to\infty} H(z) := H^\infty = \begin{bmatrix} 1 \\ \eta \end{bmatrix}$$

The innovation form:

One-step-ahead predictor: $\hat{y}(t|t-1) := \bar{\mathbb{E}}\{y(t) \mid y^{t-1}, u^t\}$

$$y(t) - \hat{y}(t|t-1) = H^\infty e(t) = \tilde{H}(q)[y(t) - G(q)u(t)]$$

$$\tilde{H}(q) := [I - [H(q) - H^\infty]H^+(q)] \qquad H^+(z)H(z) = I$$

In full-rank case: $H^\infty = I; \quad \tilde{H}(q) := H^{-1}(q)$

In our case: $\hat{y}(t|t-1) = [I - \tilde{H}(q)]y(t) + \tilde{H}(q)G(q)u(t)$

The predictor filters are non-unique

Identification with reduced-rank noise

If we restrict the predictor inputs to the full-rank output only:

$$\begin{bmatrix} \hat{y}_1(t|t-1) \\ \hat{y}_2(t|t-1) \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} y_1(t) \\ u(t) \end{bmatrix}$$

$$\begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} = \begin{bmatrix} 1 - H_1^{-1} & [H_1^{-1} \;\; 0]G \\ H_1^{-1}(H_2 - \eta) & [-H_1^{-1}(H_2 - \eta) \;\; 1]G \end{bmatrix}$$

then predictor filters become unique (and suitable for parametrization)

$$\begin{bmatrix} P_{11}(\theta) & P_{12}(\theta) \\ P_{21}(\theta) & P_{22}(\theta) \end{bmatrix}$$

prediction error: $\varepsilon(t,\theta) := y(t) - \hat{y}(t|t-1;\theta)$

$$\begin{bmatrix} \varepsilon_1(t,\theta) \\ \varepsilon_2(t,\theta) \end{bmatrix} := \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} - \begin{bmatrix} \hat{y}_1(t|t-1;\theta) \\ \hat{y}_2(t|t-1;\theta) \end{bmatrix}$$

Identification criterion

The innovation process:

$$\begin{bmatrix}1\\ \eta\end{bmatrix} e(t) = y(t) - \hat{y}(t|t-1)$$

Implication for identification based on prediction errors:

- $\min_{\theta\in\Theta} \bar{\mathbb{E}}\, \varepsilon^T(t,\theta)\varepsilon(t,\theta)$ **will not satisfy the rank condition of the innovation process in** $\theta = \theta_0$

Constrained identification criterion:

$$\min_{\theta\in\Theta} \bar{\mathbb{E}}\, \varepsilon_1(t,\theta)^2 \;\; \text{subject to} \;\; \varepsilon_2(t,\theta) = \eta\varepsilon_1(t,\theta) \;\; \text{for all } t.$$

The constraint accounts for the fact that $y_1(t)$ and $y_2(t)$ are driven by the same (white) noise

Identification criterion

Constrained identification criterion:

$$\theta^* := \min_{\theta \in \Theta} \bar{\mathbb{E}}\, \varepsilon_1(t,\theta)^2 \;\; \text{subject to} \;\; \varepsilon_2(t,\theta) = \eta\varepsilon_1(t,\theta) \;\; \text{for all } t.$$

"Consistency" result:

Let data be generated by $(G(\theta_0), H(\theta_0))$

Then $(G(\theta^*), H(\theta^*)) = (G(\theta_0), H(\theta_0))$ provided that

- The system is in the model set, and
- The input signal is sufficiently exciting

Constrained criterion may not be very practical

Constraint relaxation

After constraint relaxation:

$$\min_{\theta\in\Theta} \bar{\mathbb{E}} \left[\varepsilon_1(t,\theta)^2 \; + \lambda(\varepsilon_2(t,\theta) - \eta\varepsilon_1(t,\theta))^2\right]$$

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

For $\lambda > 0$ the consistency result remains true.

For $\lambda \to \infty$ constraint satisfaction

The above criterion is equivalent to:

$$\min_{\theta\in\Theta} \bar{\mathbb{E}}\, \varepsilon^T(t,\theta) Q \varepsilon(t,\theta))$$

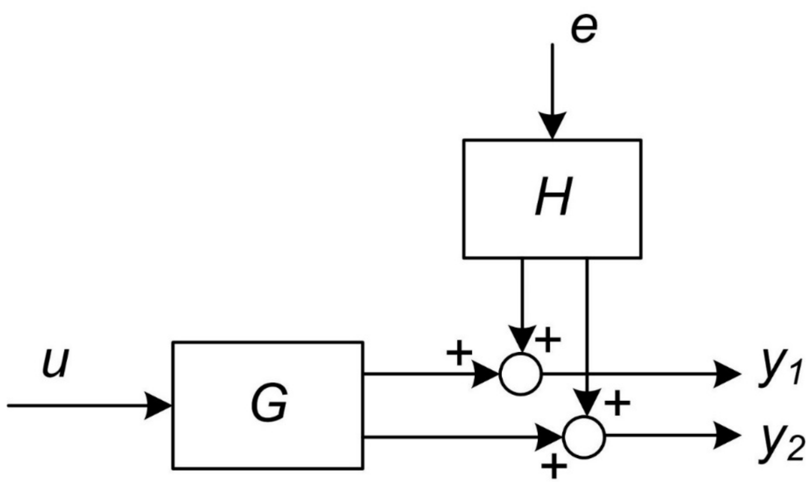
with weighting matrix $Q = \begin{bmatrix} 1+\lambda\eta^2 & -\eta\lambda \\ -\eta\lambda & \lambda \end{bmatrix}$

allowing a variance analysis.

Simulation example

Data-generating system:

$$G_1(q) = 0.3 + 0.7q^{-1} + 0.3q^{-2},$$
$$G_2(q) = 0.15 + 0.9q^{-1} - 0.5q^{-2},$$
$$H_1(q) = \frac{1}{1+0.3q^{-1}+0.4q^{-2}},$$
$$H_2(q) = \frac{1}{2-0.4q^{-1}+0.2q^{-2}},$$



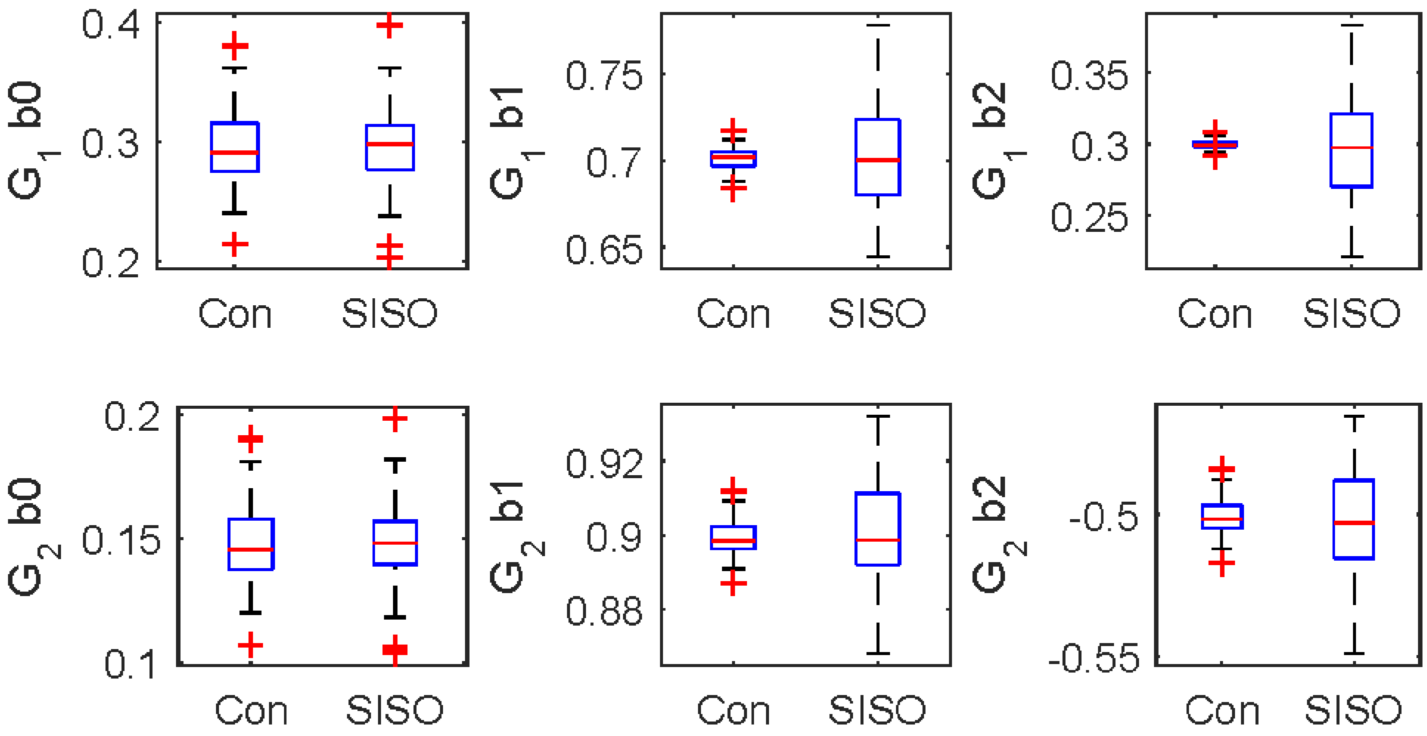
Parametrized model:

$$\begin{aligned} G_i(q,\theta) &= b_{0,i} + b_{1,i}q^{-1} + b_{2,i}q^{-2}, \quad i = \{1,2\} \\ H_j(q,\theta)^{-1} &= d_{0,j} + d_{1,j}q^{-1} + d_{2,j}q^{-2}, \quad j = \{1,2\}. \end{aligned}$$

Input: white noise process

Simulation example

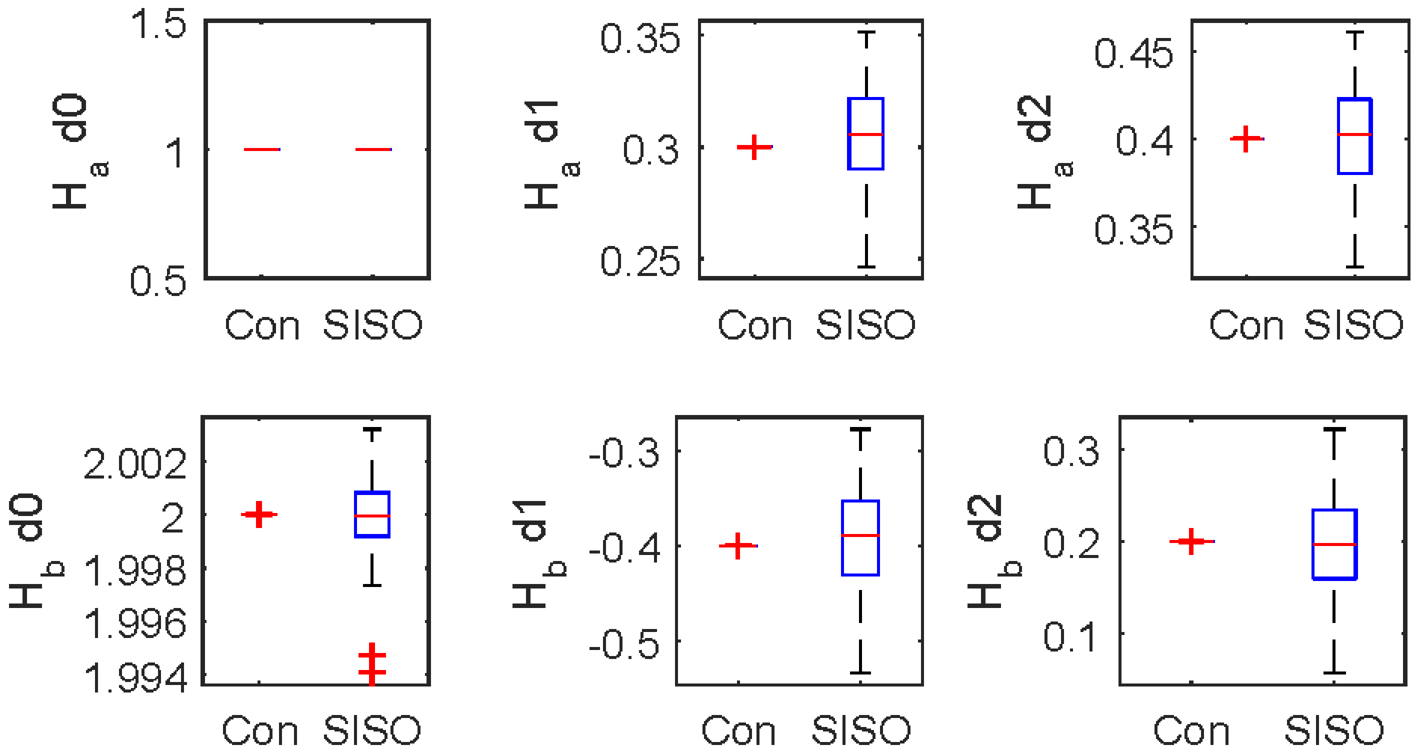
Result of 100 Monte Carlo simulation runns (*N=1000*):



"SISO"= two independent SISO identifications of (G_1, H_1) and (G_2, H_2)

Simulation example

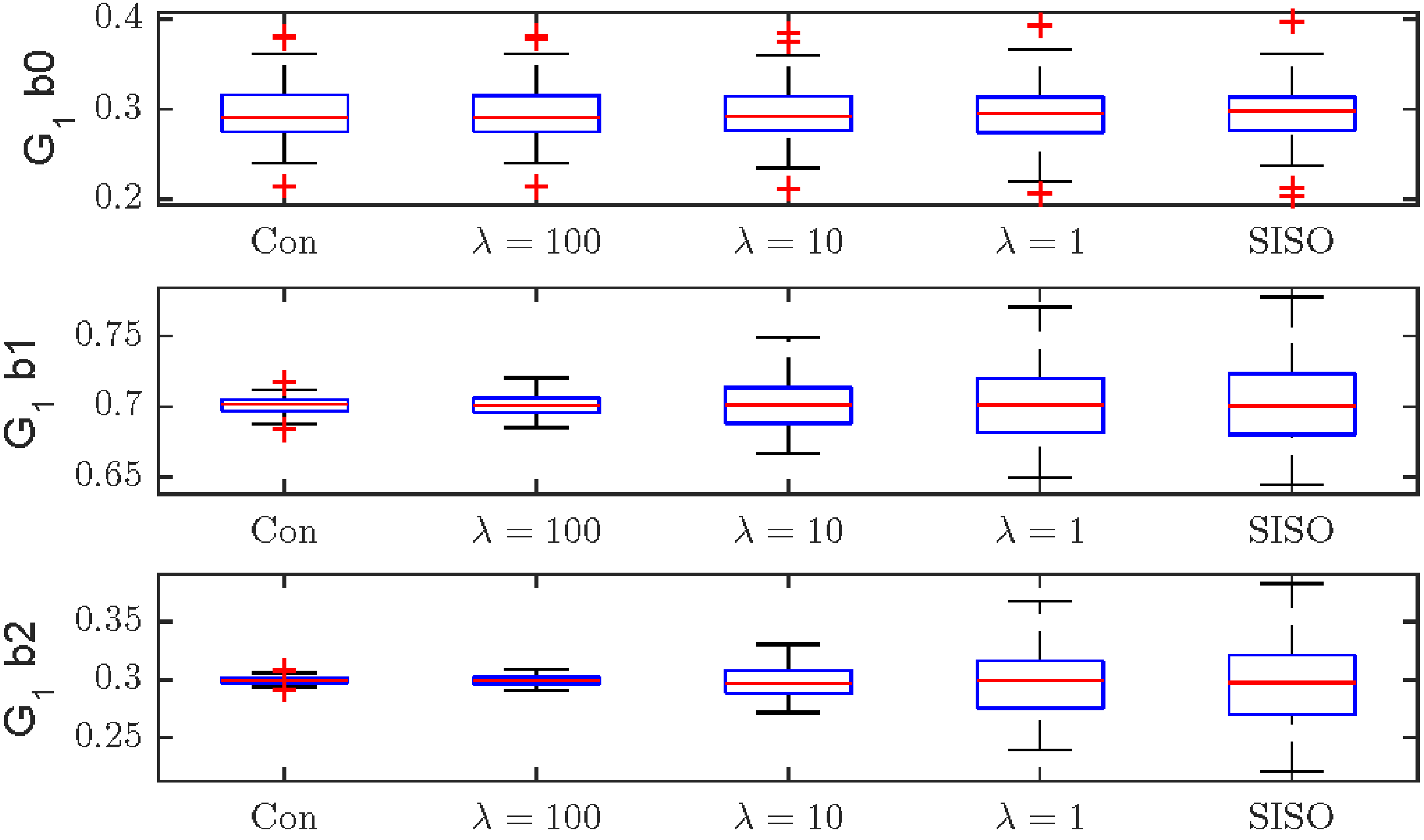
Estimated noise model parameters:



Some parameters are estimated variance-free,
see also the SIMO results in Everitt et al., Automatica 2017.

Simulation example

Criterion with constraint relaxation:



Conclusions

- The situation of reduced-rank (singular) noise in prediction error identification is addressed.
- PE / ML criterion becomes a quadratic cost function with an additional constaint
- Constraint criterion can be relaxed to arrive at minimizing a (weighted) quadratic cost function
- Simulation shows: taking account of the noise dependencies reduces the parameter variance
- Results are extended to cyclic dynamic networks in
  H.H.M. Weerts, P.M.J. Van den Hof and A.G. Dankers, IFAC World Congress, 2017

Technische Universiteit Eindhoven University of Technology