Maximum likelihood estimation in dynamic networks with rank-reduced noise

TU/e Technische Universiteit Eindhoven University of Technology

Maximum likelihood estimation in dynamic networks with rank-reduced noise

Harm H.M. Weerts (h.h.m.weerts@tue.nl), Paul M.J. Van den Hof, Arne Dankers

European Research Council
Established by the European Commission

Introduction

Dynamic networks are rising in popularity as tool to model systems.

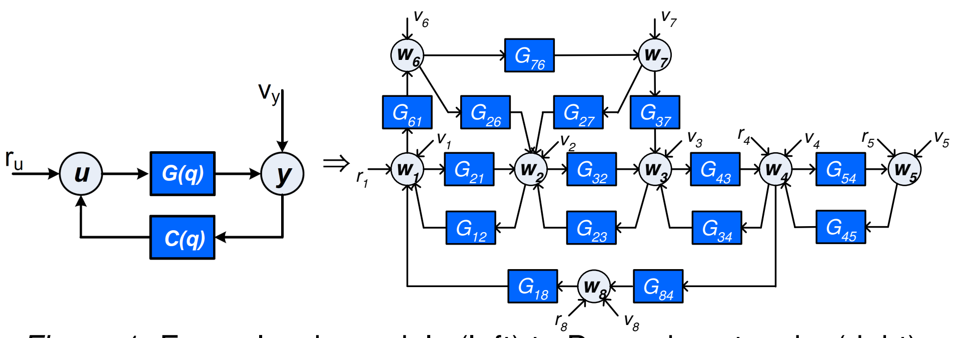
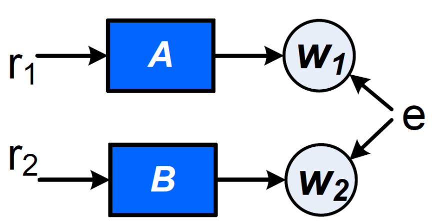


Figure 1: From classic models (left) to Dynamic networks (right).

With many measurements, the noise can be generated by fewer sources than nodes, i.e. noise is *rank-reduced*.

Goal: Obtain maximum likelihood estimates of the network, and provide a lower bound on the variance.

A special structure can result in a variance-free estimate, which is a special case of the maximum likelihood estimate.



$$\varepsilon_A = w_1 - \theta_A r_1, \quad \varepsilon_B = w_2 - \theta_B r_2$$

Should have: $\varepsilon_A = \varepsilon_B$

$$e + (A - \theta_A) r_1 = e + (B - \theta_B) r_2$$

Which eliminates noise

Figure 2: Example of rank-reduced identification. Parameters $\theta_A = A$ and $\theta_B = B$ can be determined variance-free.

The model

We assume that the first p nodes are affected by a full rank noise, and the other noise is fully dependent on this first part.

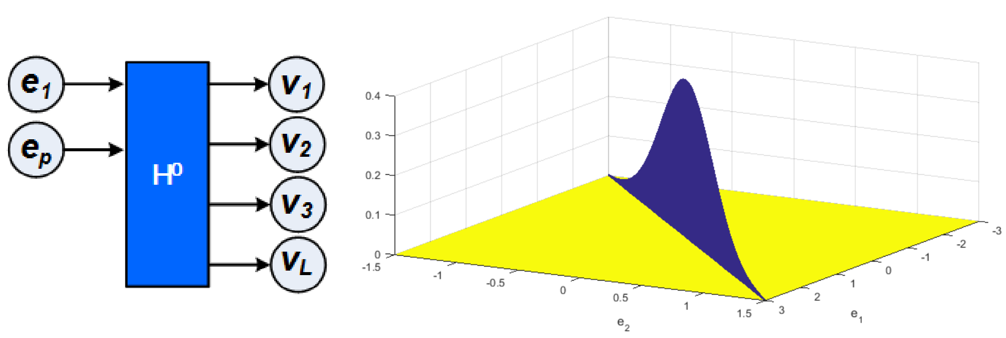


Figure 3: Rank-reduced noise model.

A prediction error is defined, with a full rank and a dependent part

$$\varepsilon(t, \theta_0) = \begin{bmatrix} \varepsilon_a(t, \theta_0) \\ \varepsilon_b(t, \theta_0) \end{bmatrix} = \begin{bmatrix} I \\ \Gamma \end{bmatrix} e(t) =: \check{e}(t).$$

The innovation is white noise $e(t)$ which is mapped to higher dimensional space.

Maximum likelihood

Assuming initial conditions are 0, and $e(t)$ is Gaussian, the maximum likelihood estimate can be obtained.

Result: Constrained optimization results in the Maximum Likelihood estimate[1]

$$\hat{\theta}_{ML} = \arg\min_{\theta} \det \left(\frac{1}{N} \sum_{t=1}^{N} \varepsilon_a(t,\theta) \varepsilon_a^T(t,\theta) \right)$$

$$\textbf{subject to } \frac{1}{N} \sum_{t=1}^{N} \varepsilon^T(t,\theta) \begin{bmatrix} \Gamma^T(\theta) \\ -I \end{bmatrix} \begin{bmatrix} \Gamma(\theta) & -I \end{bmatrix} \varepsilon(t,\theta) = 0.$$

A consistent (but not minimum variance) estimate can be obtained by[2]

$$\hat{\theta}_{WLS} = \arg\min_{\theta} \frac{1}{N} \sum_{t=1}^{N} \varepsilon^T(t,\theta) Q \varepsilon(t,\theta).$$

Variance

Typical variance expressions are invalid for constrained criteria. For analysis use a new parameter ρ that always satisfies the constraint.

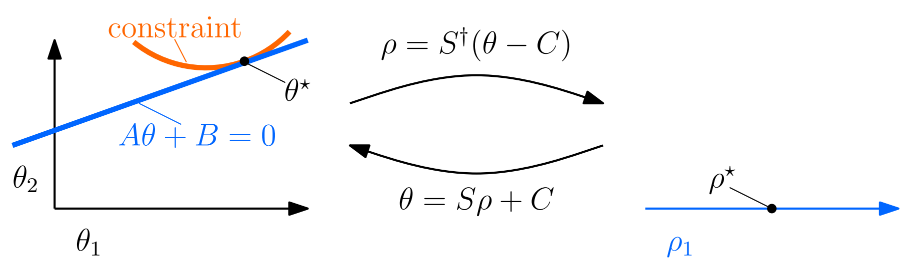


Figure 4: Original parameter space with constraint (orange), approximation of constraint (blue), and a mapping (defined by $A(S\rho + C) + B = 0\ \forall \rho$) to an unconstrained lower dimensional space.

An unconstrained criterion where θ is substituted for ρ can be solved. The ρ has distribution

$$\sqrt{N}(\hat{\rho} - \rho^*) \sim \mathcal{N}(0, P_\rho),$$

where the variance has some lower bound $P_\rho \geq P_\rho^0$. Via the reverse mapping above we obtain the variance and lower bound on θ.

Result: The variance of $\hat{\theta}$ is[1] $P_\theta \geq P_\theta^0 = S\, P_\rho^0\, S^T$.

References

[1] H.H.M. Weerts, P.M.J. Van den Hof, A. Dankers. Identification of linear dynamic networks with rank-reduced noise, *In preparation*.

[2] H.H.M. Weerts, P.M.J. Van den Hof, A. Dankers. Identification of dynamic networks with rank-reduced process noise, *IFAC WC*, 2017.