Maximum likelihood identification in dynamic networks with rank-reduced noise

Maximum likelihood identification in dynamic networks with rank-reduced noise

H.H.M. Weerts
Eindhoven University of Technology[1]

P.M.J. Van den Hof
Eindhoven University of Technology

A.G. Dankers
University of Calgary

1 Introduction

A typical assumption in multivariable and dynamic network system identification is that process noise is full rank. When a network becomes large-scale, e.g. when many sensors are placed physically close to each other, we have to consider the possibility that noise is rank-reduced[2]. Many of the typical steps in a prediction error identification framework can not immediately be taken in case noise is rank-reduced. Recently in a prediction error setting we have derived conditions under which consistent estimates of the network can be obtained in case noise is rank-reduced [1]. Our objective is to extend those results towards maximum likelihood estimates of the network dynamics, even when noise is of a reduced rank.

2 Approach

In order to estimate the network a prediction error approach is taken, leading to the multivariable prediction error $\varepsilon(t,\theta)$ of dimension L, the number of measured node variables. The typical identification criterion is the weighted least squares (WLS)

$$\theta^* := \arg\min_{\theta} \bar{\mathbb{E}}\, \varepsilon^T(t,\theta) Q \varepsilon(t,\theta). \qquad (1)$$

If Q is chosen as the identity matrix this criterion leads to consistent estimates under rather standard conditions, even when noise is rank-reduced [1]. A question that needs to be answered is whether this criterion leads to maximum likelihood estimates. Typically, taking Q as the inverse of the covariance matrix of the innovation process leads to a maximum likelihood estimate. When noise is rank-reduced the covariance matrix is singular, and therefore can not be inverted.

Due to the rank-reduced situation the covariance of the innovation can be written as $B\Lambda B^T$, with B an $L \times p$ dimensional matrix and Λ the $p \times p$ dimensional covariance matrix of $e(t)$. We use the left-inverse $B^\dagger$ of B and the nullspace of B spanned by the rows of matrix N to construct a generalized inverse of $B\Lambda B^T$ as weighting matrix

$$Q_\lambda = (B^\dagger)^T \Lambda^{-1} B^\dagger + \lambda N^T N. \qquad (2)$$

The B and N terms of the weight actually represent two orthogonal components of the innovation process. λ governs the balance between the two orthogonal components. When $\lambda \to \infty$ the variance of the estimate reduces to some lower bound, possibly 0.

3 Maximum likelihood

When the innovation ϵ is of rank $p < L$ then it does not have a probability density function (pdf) in the usual sense. However a pdf for the innovation exists on a p dimensional subspace [2], which is then mapped to the L dimensional space by matrix B. This means that $B^\dagger\epsilon(t) = e(t)$ does have a pdf, and moreover $N\epsilon(t) = 0$ with probability 1 for all t. The distribution of ϵ is specified by a pdf on the hyperplane $N\epsilon(t) = 0$.

When weight Q_λ is used in (1) with $\lambda \to \infty$ then the second term of Q_λ ensures that the estimate is taken on the hyperplane $N\epsilon(t) = 0$. The first term of Q_λ maps the prediction error to dimension p via $B^\dagger$, and then applies the appropriately weight for estimation. When the appropriate B, N and Λ are chosen then the estimate (1) with weight $Q = Q_\lambda$ is a maximum likelihood estimate.

References

[1] H.H.M. Weerts, P.M.J. Van den Hof, A. Dankers, "Identification of dynamic networks with rank-reduced process noise," IFAC World Congress, 2017, *to appear.*

[2] C.R. Rao, "Linear Statistical Inference and Its Applications," John Wiley & Sons, 1973.

[1]Contact details: h.h.m.weerts@tue.nl or P.O. Box 513, 5600 MB Eindhoven, The Netherlands.

[2]A noise process $v(t)$ of dimension L is called rank-reduced if it can be generated by filtering a white noise process $e(t)$ of dimension $p < L$.