Theoretical Foundations of Autoregressive Models for Time Series on Acyclic Directed Graphs

Ralph Högn, Claudia Czado
Zentrum Mathematik
Technische Universität München
Boltzmannstr. 3
D-85747 Garching bei München, Germany
hoegn@ma.tum.de, cczado@ma.tum.de

July 2, 2003

Abstract

Three classes of models for time series on acyclic directed graphs are considered. At first a review of tree-structured models constructed from a nested partitioning of the observation interval is given. This nested partitioning leads to several resolution scales. The concept of mass balance allowing to interpret the average over an interval as the sum of averages over the sub-intervals implies linear restrictions in the tree-structured model. Under a white noise assumption for transition and observation noise there is an change-of-resolution Kalman filter for linear least squares prediction of interval averages (Chou 1991). This class of models is generalized by modeling transition noise on the same scale in linear state space form. The third class deals with models on a more general class of directed acyclic graphs where nodes are allowed to have two parents. We show that these models have a linear state space representation with white system and coloured observation noise.

Key words: linear least squares prediction, tree-structured model, mass-balance, acyclic directed graph, linear state space model, linear Kalman filter, score vector.

Contents

1 Introduction

The main problem considered in this paper is the optimal prediction and smoothing of stochastic processes based on irregularly time-spaced observations for large or even huge data sets. Here optimality means minimal linear least square errors. It is assumed that the data is generated by a continuous time stochastic process. Distorted observations are made in discrete time points, where the distance between them may very short compared to the length of the observation interval. It is also allowed that the time points are irregularly spaced in time. Dacorogna et al. (2001) refer to such time series as irregular time series. As they point out, most methods in time series analysis rely on homogeneously time spaced time series. If the data is observed in irregularly spaced time points then a homogeneous time series has to be constructed from the raw data. For this they propose linear interpolation, previous-value interpolation or other operators such as the moving average operator. One drawback of these methods is that the length of the homogeneous intervals between two time points has to be chosen carefully and some information contained in the data may be lost. After making the time series homogeneous standard time series techniques can be applied.

We review and introduce models that are formulated by averages over different time horizons, where we do not need to construct homogeneous time series. The observation interval is divided on several resolution scales into sub-intervals where these sub-intervals become shorter from resolution scale to resolution scale. At the finest resolution scale therefore there are only short intervals that contain either no observation, one or just a few observations. The aim of these models is to allow for fast summaries for different time resolutions that are estimated by linear least squares prediction and smoothing. Another matter of interest may be the relationship between the averages for different time resolutions.

Three classes of models are discussed in this paper. At first a review of models of Huang et al. (2002) is given, which was formulated for spatial data. The spatial region is divided by nested partitioning into sub-regions on several resolution scales. Corresponding to the nested partitioning they define a tree-structured autoregressive stochastic process. Due to their application of nested partitioning they need and introduce the concept of mass balance. In their model they assume the transition and the observation noise to be independent, serially and mutually. In this case we speak of white noise otherwise the noise is coloured. They also develop an algorithm for linear least square prediction of the averages over the sub-regions which is connected to a tree-structure. Their algorithm is based on results of Chou (1991). Formulation of these models for time series is straightforward. Hence we consider these models as a starting point for modeling interval averages.

We first generalize these models by modeling transition noise on the same resolution scale in linear state space form. The observation noise is treated in the same manner. The concept of nested partitioning and mass balance was retained.

Due to restrictions implied by mass balance these models imply an artificial structure of the correlation matrix of the given data. So we formulate a third class of models where mass balance is omitted. As a consequence the concept of nested partitioning is no longer applicable and replaced by an overlapping interval arrangement. The tree-structure is no longer maintained when overlapping intervals are used so we formulate models on a more general class of acyclic directed graphs.

Autoregressive stochastic processes on acyclic directed graphs with white transition and observation noise and a corresponding algorithm are discussed by Huang and Cressie (2001). Again we generalize these models by allowing a linear state space model for the transition and observation noise.

Since the algorithm proposed by Chou (1991) and Huang and Cressie (2001) is no longer applicable for models on graphs with coloured noise we show how to find a linear state space representation with white system noise and coloured observation noise for these models. Therefore we give a brief review of these linear state space models. A Kalman filter (Chui and Chen 1999) is applicable and an existing Kalman fixed point smoothing algorithm (Durbin and Koopman 2001) was modified to allow for coloured transition noise in the state space domain. With regard to maximum likelihood estimation of the unknown model parameters we derive an analytical representation of the score vector.

2 Linear Least Squares Prediction of Interval Averages in an Additive Error Model

Huang et al. (2002) consider a continuous time stochastic process $X(s) = \mu(s) + \eta(s)$ over an spatial region G. Since we are interested in models for time series we replace the spatial region by the observation interval I. $\mu(s)$ is assumed a deterministic mean process and $\eta(s)$ a stochastic process with finite variance and zero mean. We assume that distorted observations $Z(t_i)$ of the process X are made at discrete time points t_i which are irregularly spaced in time, i.e.

$$Z(t_i) = X(t_i) + \epsilon(t_i) \ \ (t_i \in I),$$

where $\epsilon(t_i)$ is a random error variable with zero mean and finite variance.

As in Huang et al. (2002) we are interested in fast summaries of the process $X(s)$ over sub-intervals of I with different lengths. For this

$$Y := \frac{1}{|I|} \int_I \eta(s) ds$$

is considered as the average value of η over the interval I. Here $|I|$ denotes the length of the interval I. In particular I is divided on a first scale into several sub-intervals. These sub-intervals are divided into shorter sub-intervals on following scales. Let j denote the scale. On each scale the sub-intervals are numbered starting with $k = 1$ from the left to the right.

The goal is to find linear least square estimates in terms of the complete data vector $\boldsymbol{Z}$ of

$$Y_{j,k} := \frac{1}{|I_{j,k}|} \int_{I_{j,k}} \eta(s) ds. \tag{1}$$

Let $L(Y|\boldsymbol{Z})$ denote the linear least square prediction operator of a random variable Y given data $\boldsymbol{Z}$ and $\boldsymbol{A}^{-}$ the generalized inverse of the matrix $\boldsymbol{A}$. Then linear least square estimates are given by

$$L(Y_{j,k}|\boldsymbol{Z}) = E(Y) + cov(Y_{j,k}, \boldsymbol{Z}) \boldsymbol{\Sigma}_{\boldsymbol{Z}}^{-} (\boldsymbol{Z} - E(\boldsymbol{Z}), \tag{2}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{Z}} := var(\boldsymbol{Z})$. Formula (2) can be found e.g. in Hamilton (1994), Section 4.1.

For application of (2) one has to know the complete covariance-structure of $Y_{j,k}$ and $\boldsymbol{Z}$ and has to invert the matrix $\boldsymbol{\Sigma}_{\boldsymbol{Z}}$. For big or even massive data-sets $\boldsymbol{Z}$ this may be very consuming of computation time. Thus we introduce in the following sections algorithms for recursive computation.

3 Tree-structured Models with White Noise

Chou (1991) introduces tree-structured models with white system and observation noise. He derives a tree-structured Kalman filter and smoother for calculation of linear least squares prediction. In this section we review this model including the notion of mass balance, since we will consider this model as a first model for interval averages $Y_{j,k}$ defined in (1). Huang et al. (2002) partition the observation interval I into nested sub-intervals $I_{j,k}$. This means, they allow for no overlapping of intervals, and each sub-interval has just one parent. A precise formulation is given in the following definition

Definition 3.1. *(Nested Partitioning (Huang et al. 2002):)*
Let a real interval I with length $|I| > 0$, a natural number $J \in \mathbb{N}$ and a family of natural numbers $(N_j)_{j=1}^{J}$ with $N_{j+1} \geq N_j$ $(j = 1, \ldots, J)$ be given. A collection of sub-intervals $\{I_{j,k} \subset I : j =$

$1, \ldots, J,\ k = 1, \ldots, N_j\}$ *is called a nested partitioning on I, if the following conditions hold:*

(i) $|I_{j,k}| > 0 \ \ (j = 1, \ldots, J;\ k = 1, \ldots, N_j),$

(ii) $\bigcap_{k=1}^{N_j} I_{j,k} = \emptyset \ \ (j = 1, \ldots, J),$

(iii) $\bigcup_{k=1}^{N_j} I_{j,k} = I \ \ (j = 1, \ldots, J),$

(iv) $\forall I_{j,k}\ (j = 2, \ldots, J,\ k = 1, \ldots, N_j) \exists\, k' \in \{1, \ldots, N_{j-1}\}$
such that $I_{j,k} \subset I_{j-1,k'}$.

3.1 The Tree-structure

Huang et al. (2002) consider a (univariate or multivariate) random process indexed by the nodes of a tree (T,E). T denotes the set of the nodes and E the set of the directed edges. For the tree (T,E) we introduce the following notation:

J : finest scale.

j : scale, with $j = 0, \ldots, J$.

N_j : number of nodes on the scale j $(j = 0, \ldots, J)$.

(j, k) : k^{th} root on the scale j, counted from the left to the right.

$pa(j, k)$: parent node of (j, k).

$an(i, j, k)$: ancestor node on the scale i of the node (j, k).

n_0 : number of children of the root node.

$n_{j,k}$: number of children of the node (j, k).

$ch(j, k, l)$: l^{th} child of the node (j, k).

$de(i, j, k)$: the descendants on the scale i of the node (j, k).

Example 1. *As an example we show a tree with the finest scale* $J = 3$*, and* $n_{j,k} = 3$ $(j = 0, \ldots, J-1; k = 1, \ldots N_j)$ *in Figure 1. For example take the node* $(2, 1)$*. Then the following relations hold:*

$$\textit{parent node of } (2,1):\ pa(2,1) = (1,1),$$
$$\textit{ancestor node of } (2,1) \textit{ on the scale } j=1:\ an(1,2,1) = (1,1),$$
$$\textit{first child of } (2,1):\ ch(2,1,1) = (3,1),$$
$$\textit{descendants of } (2,1) \textit{ on the scale } j=3:\ de(3,2,1) = \big((3,1),(3,2),(3,3)\big).$$

Together with the nested partitioning from Definition 3.1 we immediately get a tree of intervals, where the original interval I *is assigned to the root. The interval* $I_{j,k}$ *from Definition 3.1(iv) could then be called a child of interval* $I_{j-1,k'}$ $(j = 1, 2, 3; k = 1, \ldots, N_j)$.

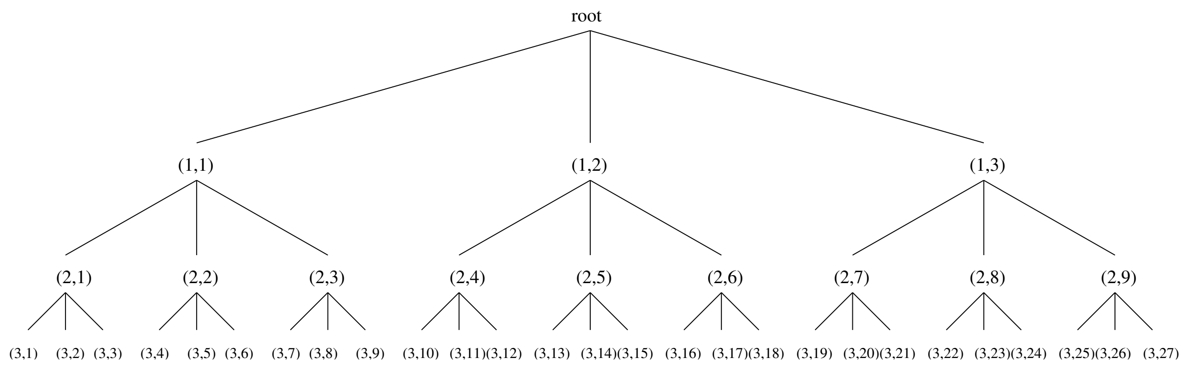

Figure 1: Tree-structure of Example 1

It is assumed that the process evolves from parents to children in an autoregressive manner. The process itself is considered as hidden. The corresponding observations are connected with the random variables on the finest scale by observation equations. Thus Chou (1991) considers the following model:

Definition 3.2. *(Univariate Tree-structured Model:)*
Let a tree (E, T) with the finest scale J and $n_{j,k}$ children $(j = 0, \ldots, J-1, k = 1, \ldots, N_j)$ be given. A tree-structured stochastic process can then be defined as follows:

	Root:	$Y_0,$
(3)	***Transition equation:***	$Y_{j,k} = Y_{pa(j,k)} + W_{j,k} \ \ (j = 1, \ldots, J;\ k = 1, \ldots, N_j),$
(4)	***Observation equation:***	$Z_k = Y_{J,k} + Q_k \ \ (k = 1, \ldots, N_J).$

We restrict ourselves to the special case where observations Z_k are available only at the finest scale J. Then we can index them with a single index k. $W_{j,k}$ is called transition noise and Q_k observation noise.

Let $X \perp Y$ denote that random variables X and Y are uncorrelated. If normal joint distribution of the random variables is assumed this is equivalent to independence. Chou (1991) makes the following white noise assumptions:

$$\begin{gathered} E(Y_0) = 0, E(Y_0^2) < \infty, \\ E(W_{j,k}) = 0, \sigma_{j,k}^2 := E(W_{j,k}^2) < \infty, \\ E(Q_k) = 0, \tau_k^2 := E(Q_k^2) < \infty, \\ W_{j,k} \perp Y_0,\ W_{j,k} \perp W_{i,l}\ (j \neq i), \\ W_{j,k} \perp Q_l,\ Q_k \perp Y_0. \end{gathered} \tag{5}$$

Further assumptions necessary for the transition and observation noise will be given in the following subsection.

3.2 The Concept of Mass Balance

The concept of mass balance was introduced by Huang et al. (2002). In (1) the average value of $\eta(s)$ was defined by

$$Y_{j,k} := \frac{1}{|I_{j,k}|} \int_{I_{j,k}} \eta(s) ds.$$

It is natural to assume that the average value of the process $\eta(s)$ over the interval $I_{j,k}$ is the sum of the average values of $\eta(s)$ over the sub-intervals $I_{ch(j,k,l)}$, $l = 1, \dots, n_{j,k}$. That is, it is assumed that

$$Y_{j,k} = \sum_{l=1}^{n_{j,k}} Y_{ch(j,k,l)} \tag{6}$$

holds for $j = 0, \dots, J-1$, $k = 1, \dots, N_j$. This assumption was denoted by Huang et al. (2002) as mass balance. This is equivalent to requiring

$$\int_{I_{j,k}} \eta(s) ds = \sum_{l=1}^{n_{j,k}} \int_{I_{ch(j,k,l)}} \eta(s) ds.$$

It can easily be shown that mass balance in (6) can be characterized by

$$\sum_{l=1}^{n_{j,k}} |I_{ch(j,k,l)}| W_{ch(j,k,l)} = 0, \text{ or} \tag{7}$$

$$var\left(\sum_{l=1}^{n_{j,k}} |I_{ch(j,k,l)}| W_{ch(j,k,l)} \right) = 0 \tag{8}$$

for $j = 0, \dots, J-1$, $k = 1, \dots, N_j$, $l = 1, \dots, n_{j,k}$. We can solve Equation (7) for a chosen $W_{ch(j,k,l)}$:

$$\begin{gathered} W_{ch(j,k,l)} = -\frac{1}{|I_{ch(j,k,l)}|} \sum_{\substack{\lambda=1 \\ \lambda \neq l}}^{n_{j,k}} |I_{ch(j,k,\lambda)}| W_{ch(j,k,\lambda)} \\ (j = 0, \dots, J-1,\ k = 1, \dots, N_j,\ l = 1, \dots, n_{j,k}). \end{gathered} \tag{9}$$

If $|I_{ch(j,k,l)}| = |I_{ch(j,k,1)}|$ $(l = 1, \dots, n_{j,k})$, i.e. for $(l = 1, \dots, n_{j,k})$ the sub-intervals have equal length Equation (9) simplifies to

$$W_{ch(j,k,l)} = -\sum_{\substack{\lambda=1 \\ \lambda \neq l}}^{n_{j,k}} W_{ch(j,k,\lambda)}. \tag{10}$$

3.3 Vectorized Tree-structured Models

Since the mass balance of a particular node involves conditions on all children of this node, it is convenient to combine these children in a vector. Together with Definition 3.2 this yields the following vectorization (see Huang et al. (2002), Subsection 3.2):

Definition 3.3. *Given a tree-structured model as in Definition 3.2. For $j = 0, \dots, J-1$, $k = 1, \dots, N_j$ Huang et al. (2002) define*

$$\boldsymbol{Y}_{ch(j,k)} := (Y_{ch(j,k,1)}, Y_{ch(j,k,2)}, \dots, Y_{ch(j,k,n_{j,k})})', \tag{11}$$

$$\boldsymbol{W}_{ch(j,k)} := (W_{ch(j,k,1)}, W_{ch(j,k,2)}, \dots, W_{ch(j,k,n_{j,k})})', \tag{12}$$

$$\boldsymbol{Z}_{k} := (Z_{ch(pa(J,k),1)}, Z_{ch(pa(J,k),2)}, \dots, Z_{ch(pa(J,k),r_{J,k})})', \tag{13}$$

$$\boldsymbol{Q}_{k} := (Q_{ch(pa(J,k),1)}, Q_{ch(pa(J,k),2)}, \dots, Q_{ch(pa(J,k),r_{J,k})})'. \tag{14}$$

$(j = 0, \dots, J-1; k = 1, \dots, N_j)$.

We introduce some additional notation:

K_j : Number of nodes on the scale j in the vectorized tree-structured model.

$r_{j,k}$: Number of elements in the random vector $\boldsymbol{Y}_{j,k}$ $(j = 1, \dots, J;\ k = 1, \dots, K_j)$.

With this notation the following relations hold

$$K_j = N_{j-1} \ \ (j = 1, \dots, J),$$
$$r_{j,k} = n_{pa(j,k)} \ \ (j = 1, \dots, J;\ k = 1, \dots, K_j),$$
$$\boldsymbol{W}_{j+1,k} = \boldsymbol{W}_{ch(j,k)} \ \ (j = 2, \dots, J-1;\ k = 1, \dots, N_j).$$

We illustrate this vectorization by the following Example:

Example 2. *For the tree in Figure 1 we define*

j=1: $$\boldsymbol{Y}_{1,1} := \begin{pmatrix} Y_{1,1} \\ Y_{1,2} \\ Y_{1,3} \end{pmatrix} \qquad \boldsymbol{W}_{1,1} := \begin{pmatrix} W_{1,1} \\ W_{1,2} \\ W_{1,3} \end{pmatrix}$$

j=2: $$\boldsymbol{Y}_{2,1} := \begin{pmatrix} Y_{2,1} \\ Y_{2,2} \\ Y_{2,3} \end{pmatrix} \qquad \boldsymbol{W}_{2,1} := \begin{pmatrix} W_{2,1} \\ W_{2,2} \\ W_{2,3} \end{pmatrix}$$

$$\boldsymbol{Y}_{2,2} := \begin{pmatrix} Y_{2,4} \\ Y_{2,5} \\ Y_{2,6} \end{pmatrix} \qquad \boldsymbol{W}_{2,2} := \begin{pmatrix} W_{2,4} \\ W_{2,5} \\ W_{2,6} \end{pmatrix}$$

$$\boldsymbol{Y}_{2,3} := \begin{pmatrix} Y_{2,7} \\ Y_{2,8} \\ Y_{2,9} \end{pmatrix} \qquad \boldsymbol{W}_{2,1} := \begin{pmatrix} W_{2,7} \\ W_{2,8} \\ W_{2,9} \end{pmatrix}$$

j=3: $$\boldsymbol{Y}_{3,1} := \begin{pmatrix} Y_{3,1} \\ Y_{3,2} \\ Y_{3,3} \end{pmatrix} \qquad \boldsymbol{W}_{3,1} := \begin{pmatrix} W_{3,1} \\ W_{3,2} \\ W_{3,3} \end{pmatrix}$$

$$\vdots \qquad\qquad \vdots$$

$$\boldsymbol{Y}_{3,9} := \begin{pmatrix} Y_{3,25} \\ Y_{3,26} \\ Y_{3,27} \end{pmatrix} \qquad \boldsymbol{W}_{3,9} := \begin{pmatrix} W_{3,25} \\ W_{3,26} \\ W_{3,27} \end{pmatrix}$$

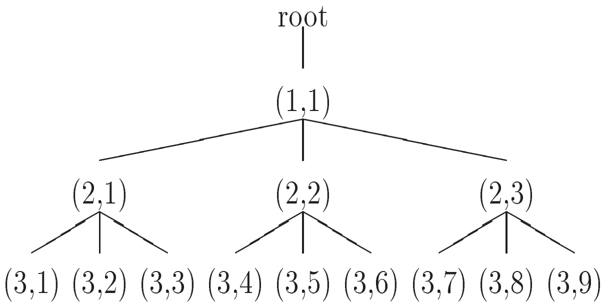

Figure 2: Tree for vectorized nodes in Example 2

This vectorized tree is shown in Figure 2.

We now consider the variance matrix of $\boldsymbol{W}_{j,k}$, $j = 1, \ldots, J$, $k = 1, \ldots, N_j$, under mass balance. For this let

$$
\begin{aligned}
\boldsymbol{\Sigma}_{\boldsymbol{W}_{ch(j,k)}} &:= var(\boldsymbol{W}_{ch(j,k)}) \\
&= \begin{pmatrix}
E(W^2_{ch(j,k,1)}) & E(W_{ch(j,k,1)}W_{ch(j,k,2)}) & \ldots & E(W_{ch(j,k,l)}W_{ch(j,k,n_{j,k})}) \\
E(W_{ch(j,k,2)}W_{ch(j,k,1)}) & E(W^2_{ch(j,k,2)}) & \ldots & E(W_{ch(j,k,2)}W_{ch(j,k,n_{j,k})}) \\
\vdots & \vdots & \vdots & \vdots \\
E(W_{ch(j,k,n_{j,k})}W_{ch(j,k,1)}) & E(W_{ch(j,k,n_{j,k})}W_{ch(j,k,2)}) & \ldots & E(W^2_{ch(j,k,n_{j,k})})
\end{pmatrix} \\
&(j = 0, \ldots, J-1,\ k = 1, \ldots, K_j).
\end{aligned}
$$

Using Equation (9) the elements $l = 1, \ldots, n_{j,k} - 1$ of the last column of $\boldsymbol{\Sigma}_{\boldsymbol{W}_{ch(j,k)}}$ have to satisfy:

$$
\begin{aligned}
E(W_{ch(j,k,l)}W_{ch(j,k,n_{j,k})}) &= E\big(W_{ch(j,k,l)} \sum_{\lambda=1}^{n_{j,k}-1} -\frac{|I_{ch(j,k,\lambda)}|}{|I_{ch(j,k,n_{j,k})}|} W_{ch(j,k,\lambda)}\big) \\
&= -\frac{1}{|I_{ch(j,k,n_{j,k})}|} \sum_{\lambda=1}^{n_{j,k}-1} |I_{ch(j,k,\lambda)}| E\big(W_{ch(j,k,l)}W_{ch(j,k,\lambda)}\big)
\end{aligned} \tag{15}
$$

and the last element of the last column is $E(W^2_{ch(j,k,n_{j,k})})$. Since $\boldsymbol{\Sigma}_{\boldsymbol{W}_{ch(j,k)}}$ is symmetric, the last row of $\boldsymbol{\Sigma}_{\boldsymbol{W}_{ch(j,k)}}$ is simply the transponse of the last column of $\boldsymbol{\Sigma}_{\boldsymbol{W}_{ch(j,k)}}$. Huang et al. (2002) proposed a simple method for constructing positive semi-definite matrices satisfying mass balance which are used as variance matrices $\boldsymbol{\Sigma}_{\boldsymbol{W}_{j,k}}$ $(j = 1, \ldots, J, k = 1, \ldots, K_j)$. They assume that the condition

$$
\begin{aligned}
&min\big\{|I_{ch(j,k,1)}|^2\sigma^2_{ch(j,k,1)}, |I_{ch(j,k,2)}|^2\sigma^2_{ch(j,k,2)}, \ldots, |I_{ch(j,k,n_{j,k})}|^2\sigma^2_{ch(j,k,n_{j,k})}\big\} \geq \\
&\qquad \frac{1}{n_{j,k}(n_{j,k}-1)} \sum_{l=1}^{n_{j,k}} |I_{ch(j,k,l)}|^2\sigma^2_{ch(j,k,l)}
\end{aligned} \tag{16}
$$

is satisfied. Recall that $\sigma_{j,k}^2 := var(W_{j,k})$ for $j = 1, \ldots, J$, $k = 1, \ldots, N_j$. Further they define

$$\begin{aligned}
\boldsymbol{F}_{ch(j,k)} &:= \frac{1}{n_{j,k}-1}(n\boldsymbol{I}_{n_{j,k}} - \mathbf{1}_{n_{j,k}}\mathbf{1}'_{n_{j,k}}),\\
\boldsymbol{G}_{ch(j,k)} &:= \left(1 - \frac{1}{(n_{j,k}-1)^2}\right)\boldsymbol{I}_{n_{j,k}} + \frac{1}{(n_{j,k}-1)^2}\mathbf{1}_{n_{j,k}}\mathbf{1}'_{n_{j,k}},\\
\boldsymbol{r}_{ch(j,k)} &:= \left(|I_{ch(j,k,1)}|^2, \ldots, |I_{ch(j,k,n_{ch(j,k)})}|^2\right)',\\
\boldsymbol{a}_{ch(j,k)} &:= \left(|I_{ch(j,k,1)}|^2\sigma^2_{ch(j,k,1)}, \ldots, |I_{ch(j,k,n_{ch(j,k)})}|^2\sigma^2_{ch(j,k,n_{ch(j,k)})}\right)',\\
\boldsymbol{c}_{ch(j,k)} &:= \boldsymbol{G}^{-1}_{ch(j,k)}\boldsymbol{r}_{ch(j,k)},
\end{aligned}$$

then the matrix defined by

$$\boldsymbol{\Sigma}_{\boldsymbol{W}_{ch(j,k)}} := \left(diag(\boldsymbol{r}_{ch(j,k)})\right)^{-1}\boldsymbol{F}_{ch(j,k)}diag(\boldsymbol{c}_{ch(j,k)})\boldsymbol{F}_{ch(j,k)}\left(diag(\boldsymbol{r}_{ch(j,k)})\right)^{-1} \tag{17}$$

is semi positive-definite and can be used as variance matrix. If for a node (j,k) in a univariate tree-structured model $\sigma^2_{ch(j,k,l)} = \sigma^2_{ch(j,k,1)}$ $(l = 2, \ldots, n_{j,k})$ and $|I_{ch(j,k,l)}| = |I_{ch(j,k,1)}|$ holds, (17) reduces to

$$\boldsymbol{\Sigma}_{\boldsymbol{W}_{ch(j,k)}} := \sigma^2_{j+1}(\boldsymbol{I}_{n_j} + \frac{1}{n_j-1}\boldsymbol{I}_{n_j} - \frac{1}{n_j-1}\mathbf{1}_{n_j}\mathbf{1}'_{n_j}) \;\; (j = 1, \ldots, J-1).$$

Transition Equation (3) and Observation Equation (4) can be rewritten for the vectorized model given by Definition 3.3 as follows:

$$\textbf{Transition equation:} \qquad \boldsymbol{Y}_{j+1,k} = \mathbf{1}_{n_{j,k}}Y_{j,k} + \boldsymbol{W}_{j+1,k}, \tag{18}$$

$$\textbf{Observation equation:} \qquad \boldsymbol{Z}_k = \boldsymbol{Y}_{J,k} + \boldsymbol{Q}_k \;\; (k = 1, \ldots, K_J). \tag{19}$$

The white noise assumptions (5) for the univariate model imply for $j = 1, \ldots, J$, $k = 1, \ldots, K_j$

$$\boldsymbol{W}_{j,k} \perp Y_0, \; \boldsymbol{Q}_k \perp Y_0, \tag{20}$$

and $\left\{\boldsymbol{W}_{j,k}, j = 1, \ldots, J, k = 1 \ldots, K_j\right\}$ and $\left\{\boldsymbol{Q}_k, k = 1, \ldots, K_J\right\}$ are families of mutually and serially orthogonal random vectors.

Chou (1991) introduced an efficient tree-structured Kalman filter and smoother for these models. The algorithm has the advantage that, in case of a huge number of observations $\boldsymbol{Z}$, it can be easily performed in parallel problems of smaller dimensions. Estimation of unknown variance parameters of the transition noise can be done by an EM algorithm, as formulated in Huang et al. (2002). They assumed the variance of the observation noise to be known, for instance by information available for the measuring device or from independent experiments.

The most severe drawback of this setup in our opinion is the fact, that the model given by (18) and

(19) implies an artificial block structure of the correlation matrix of $\boldsymbol{Z}$. In Appendix A we give an Example that demonstrates this behavior. It is not possible to generate a stationary correlation function for $\boldsymbol{Z}$. As a possible solution Huang et al. (2002) suggested to compute the estimates as an average over a number of mass balanced, tree-structured models with different tree branches. They also pointed out, that the estimation variances and covariances will be considerably more complicated and the computational complexity will increase with the number of trees used.

4 Tree-structured Models with Coloured Transition and Observation Noise

In order to smooth the block structure of the implied correlation of $\boldsymbol{Z}$ we relax the white noise assumption while maintaining the orthogonality of transition noise on different scales and the orthogonality on the observation noise, i.e.

$$\boldsymbol{W}_{j,k} \perp \boldsymbol{W}_{i,l} \text{ for } j \neq i \text{ and } \boldsymbol{W}_{j,k} \perp \boldsymbol{Q}_l \text{ for } j = 1, \ldots, J,\ k, l = 1, \ldots, K_j.$$

But we now allow for correlation of the transition noise on the same scale. For this we note that there is only one node on the first scale of the vectorized tree. Further we assume for a given scale $j = 2, \ldots, J$ and $k = 1, \ldots, K_j - 1$ the vectorized transition noise to be a vector AR(1)-process (see for example Wei (1990), Section 14.3) given by

$$\boldsymbol{W}_{j,k+1} = \boldsymbol{B}_{j,k} \boldsymbol{W}_{j,k} + \boldsymbol{V}_{j,k}, \tag{21}$$

where $\boldsymbol{V}_{j,k}$ is a zero-mean random variable and all of its components have finite variance. Furthermore we assume $\boldsymbol{V}_{j,k} \perp \boldsymbol{W}_{j,1}$ and $\boldsymbol{V}_{j,k} \perp \boldsymbol{V}_{j,l}$ for $k \neq l$. This implies

$$L(\boldsymbol{W}_{j,k+1} | \boldsymbol{W}_{j,1}, \ldots, \boldsymbol{W}_{j,k}) = L(\boldsymbol{W}_{j,k+1} | \boldsymbol{W}_{j,k}) = cov(\boldsymbol{W}_{j,k+1}, \boldsymbol{W}_{j,k}) \boldsymbol{\Sigma}^{-}_{\boldsymbol{W}_{j,k}} \boldsymbol{W}_{j,k}.$$

Therefore, $\boldsymbol{B}_{j,k} = E(\boldsymbol{W}_{j,k+1} \boldsymbol{W}'_{j,k}) \boldsymbol{\Sigma}^{-}_{\boldsymbol{W}_{j,k}}$.

Let $\boldsymbol{w}_{j,k}$ be a realization of the random vector $\boldsymbol{W}_{j,k}$. The Mass Balance Equation (9), which is still assumed to hold, states that a single element in $\boldsymbol{w}_{j,k}$ is uniquely determined by the other elements. This implies a restriction on the covariance matrix $E(\boldsymbol{W}_{j,k+1} \boldsymbol{W}'_{j,k})$. More precisely, from Equation (8) it follows that for $\lambda = 1, \ldots, n_{j,k} - 1$

$$E(W_{ch(j,k+1,n_{j,k+1})} W_{ch(j,k,\lambda)}) = -\frac{1}{|I_{ch(j,k+1,n_{j,k+1})}|} \sum_{l=1}^{n_{j,k+1}-1} |I_{ch(j,k+1,l)}| E(W_{ch(j,k+1,l)} W_{ch(j,k,\lambda)}), \tag{22}$$

$$E(W_{ch(j,k+1,\lambda)} W_{ch(j,k,n_{j,k})}) = -\frac{1}{|I_{ch(j,k,n_{j,k})}|} \sum_{l=1}^{n_{j,k}-1} |I_{ch(j,k,l)}| E(W_{ch(j,k+1,\lambda)} W_{ch(j,k,n_{j,k})}), \tag{23}$$

$$E(W_{ch(j,k+1,n_{j,k+1})} W_{ch(j,k,n_{j,k})}) = -\frac{1}{|I_{ch(j,k+1,n_{j,k+1})}|} \sum_{\lambda=1}^{n_{j,k}-1} |I_{ch(j,k,\lambda)}| E(W_{ch(j,k+1,n_{j,k+1})} W_{ch(j,k,\lambda)}). \tag{24}$$

Note that $E(W_{ch(j,k+1,n_{j,k+1})} W_{ch(j,k,\lambda)})$ for $\lambda = 1, \ldots, n_{j,k}$ is computed in Equation (22).

Let $\boldsymbol{W}^{\dagger}_{ch(j,k)}$ denote the vector $\boldsymbol{W}_{ch(j,k)}$ without its last element $W_{ch(j,k,n_{j,k})}$. As seen from Equation (24) $E(W_{ch(j,k+1,n_{j,k+1})}W_{ch(j,k,n_{j,k})})$ is a weighted sum of all of the elements in the matrix $E(\boldsymbol{W}^{\dagger}_{j,k+1}\boldsymbol{W}^{\dagger\prime}_{j,k})$. Since $|cov(X_1,X_2)| \leq \sqrt{var(X_1)var(X_2)}$ holds for two random variables X_1, X_2,

(25)

$$\frac{1}{|I_{ch(j,k+1,n_{j,k+1})}|}\Big|\sum_{\lambda=1}^{n_{j,k}-1}|I_{ch(j,k,\lambda)}|E(W_{ch(j,k+1,n_{j,k+1})}W_{ch(j,k,\lambda)})\Big| \leq \sigma_{ch(j,k+1,n_{j,k+1})}\sigma_{ch(j,k,n_{j,k})}$$

has to hold. In simulations with several covariance matrices $E(\boldsymbol{W}_{j,k+1}\boldsymbol{W}'_{j,k})$ Condition (25) turned out to be just a necessary but not a sufficient condition to obtain a positive semidefinite matrix

$$var(\boldsymbol{V}_{j,k}) = var(\boldsymbol{W}_{j,k+1} - \boldsymbol{B}_{j,k}\boldsymbol{W}_{j,k}) = \boldsymbol{\Sigma}_{\boldsymbol{W}_{j,k+1}} - \boldsymbol{B}_{j,k}\boldsymbol{\Sigma}_{\boldsymbol{W}_{j,k}}\boldsymbol{B}'_{j,k}.$$

Since $\boldsymbol{B}_{j,k} = E(\boldsymbol{W}_{j,k+1}\boldsymbol{W}'_{j,k})\boldsymbol{\Sigma}^{-}_{\boldsymbol{W}_{j,k}}$ holds, the matrix $var(\boldsymbol{V}_{j,k})$ depends on $E(\boldsymbol{W}_{j,k+1}\boldsymbol{W}'_{j,k})$. Thus one problem of this model is to find sufficient conditions for the matrix $E(\boldsymbol{W}_{j,k+1}\boldsymbol{W}'_{j,k})$ to obtain a positive semidefinite matrix $var(\boldsymbol{V}_{j,k})$ such that mass balance is satisfied.

For the observation noise we proceed in a similar manner:

$$\boldsymbol{Q}_{k+1} = \boldsymbol{D}_k\boldsymbol{Q}_k + \boldsymbol{U}_k,$$

with $E(\boldsymbol{U}_k) = \boldsymbol{0}$, $\boldsymbol{U}_k \perp \boldsymbol{Q}_1$ and $\boldsymbol{U}_k \perp \boldsymbol{U}_l$. Since we do not require mass balance for the observations, the problems discussed for the transition noise don't occur here.

Since the derivation of the tree-structured Kalman filter and smoother mentioned in the previous section make distinct use of the orthogonality of the noise it doesn't apply to Model (18),(19) and (21). But this model has a state space representation with white system noise and coloured observation noise. This state space representation will be derived in Section 6 for more general models defined on acyclic directed graphs which allow several parents of a node. Chui and Chen (1999) derived a Kalman filter for such state space models. A Kalman smoothing algorithm, as in Durbin and Koopman (2001) can be modified for such state space models. Estimation of the unknown parameters can again be carried out by an EM-Algorithm.

For $cor(\boldsymbol{Z})$ the vector AR(1)-structure of the observation noise seems to have no great effect. The major effect is brought in by the vector AR(1)-structure of the transition noise. The structure of $cor(\boldsymbol{Z})$ depends on the specification of the covariance matrices $E(\boldsymbol{W}_{j,k+1}\boldsymbol{W}'_{j,k})$ for $j = 2,\ldots,J$ and $k = 1,\ldots,K_j-1$. But stationarity of the correlation function of $\boldsymbol{Z}$ was still not obtained. The correlation matrix of $\boldsymbol{Z}$ now depends also on specification of $E(\boldsymbol{W}_{j,k+1}\boldsymbol{W}'_{j,k})$ or $\boldsymbol{B}_{j,k}$, respectively. Example 6 given in Appendix A illustrates this. On the other side we have to specify $E(\boldsymbol{W}_{j,k+1}\boldsymbol{W}'_{j,k})$ carefully and therefore have more parameters to estimate than in the tree-structured model with white noise. Thus it is questionable, whether the model with coloured noise is really an improvement for the desired inference of time series data. It seems that these problems are the result from using the concept of mass balance. In the next section we therefore introduce a model, where no mass balance is assumed.

5 Extensions of White Noise Linear State Space Models

In the next section we will derive a linear state space representation for an autoregressive stochastic process indexed by the nodes of an acyclic directed graph. For this reason we give in this section a short review about extensions of linear state space models. For a detailed discussion see e.g. Chui and Chen (1999). This section is arranged into several subsections. At first, we discuss linear state space models with coloured observation noise which are needed in Section 6 for autoregressive models on acyclic directed graphs. Then we give a brief summary of the Kalman filter and Kalman one step predictor for these state space models. These two subsections refer to Chui and Chen (1999). In the next subsections we state smoothing algorithms, where we follow the approach in Durbin and Koopman (2001) for white noise linear state space models. Since we consider linear state space models with white system and coloured observation noise some modifications have to be made. Then we discuss the treatment of missing observations, which are necessary to consider for the models in Section 6. For maximum likelihood estimation of model parameters we then derive an analytical representation of the score statistic in the last subsection where we assume additionally normal distribution.

5.1 Linear State Space Model with White System Noise and Coloured Observation Noise

Chui and Chen (1999), Chapter 5, considered the following class of linear state space models with white system noise and coloured observation noise:

(26) **Transition equation:** $$\boldsymbol{X}_{k+1} = \bar{\boldsymbol{A}}_k \boldsymbol{X}_k + \boldsymbol{\Gamma}_k \boldsymbol{\xi}_k,$$

(27) **Observation equation:** $$\boldsymbol{Z}_k = \bar{\boldsymbol{C}}_k \boldsymbol{X}_k + \boldsymbol{\eta}_k,$$

where $\boldsymbol{X}_k \in \mathbb{R}^{m_k}$, $\boldsymbol{\xi}_k \in \mathbb{R}^{m_k^\xi}$, $\boldsymbol{Z} \in \mathbb{R}^{m_k^Z}$. Therefore $\boldsymbol{\eta}_k \in \mathbb{R}^{m_k^Z}$, $\bar{\boldsymbol{A}}_k \in \mathbb{R}^{m_{k+1} \times m_k}$, $\boldsymbol{\Gamma}_k \in \mathbb{R}^{m_{k+1} \times m_k^\xi}$ and $\bar{\boldsymbol{C}}_k \in \mathbb{R}^{m_k^Z \times m_k}$. $\boldsymbol{\Gamma}_k$ is assumed to be either a selection matrix, i.e. only some diagonal elements are equal to one, as all the other elements are equal to zero, or to be a matrix of the form

$$\boldsymbol{\Gamma}_k := \begin{pmatrix} block\ 1 \\ \vdots \\ block\ p \end{pmatrix},$$

where $p \in \mathbb{N}$ and $block\ m$ is a selection matrix, $m \in \{1, \ldots, p\}$, and $block\ l = \mathbf{0}$ for $l = 1, \ldots, m-1, m+1, \ldots, p$. In the latter case we say that $\boldsymbol{\Gamma}_k$ is a block selection matrix. For the transition noise

$\{\boldsymbol{\xi}_k, k \geq 1\}$ and the observation noise $\{\boldsymbol{\eta}_k, k \geq 1\}$ Chui and Chen (1999) assume

$$\begin{aligned} &E(\boldsymbol{\xi}_k) = \mathbf{0},\ var(\boldsymbol{\xi}_k) < \infty, E(\boldsymbol{\eta}_k) = \mathbf{0},\ var(\boldsymbol{\eta}_k) < \infty, \\ &\boldsymbol{\xi}_k \perp \boldsymbol{X}_k, \boldsymbol{\xi}_k \perp \boldsymbol{\xi}_l\ (k \neq l),\ \boldsymbol{\xi}_k \perp \boldsymbol{\eta}_l\ (k,l \geq 1), \boldsymbol{\eta}_k \perp \boldsymbol{X}_l\ (k,l \geq 1). \end{aligned} \tag{28}$$

For the observation noise $\boldsymbol{\eta}_k$ they additionally assume

$$\boldsymbol{\eta}_k = \bar{\boldsymbol{N}}_{k-1}\boldsymbol{\eta}_{k-1} + \boldsymbol{\gamma}_k\ \ (k \geq 2),\ \text{with}\ \boldsymbol{\gamma}_k \perp \boldsymbol{\eta}_{k-1},\ \boldsymbol{\gamma}_k \perp \boldsymbol{\gamma}_l\ (k \neq l). \tag{29}$$

$\bar{\boldsymbol{N}}_{k-1}$ is a $m_k^Z \times m_{k-1}^Z$ real matrix. If $\bar{\boldsymbol{N}}_{k-1} = \mathbf{0}$ Chui and Chen (1999) speak of white observation noise, and if on the other hand $\bar{\boldsymbol{N}}_{k-1} \neq \mathbf{0}$ they say that the observation noise is coloured. Of course all their derivations and our derivations given below hold for both cases. So we can regard the case of coloured observation noise as a more general case than white observation noise.

The state vectors $\boldsymbol{X}_k$ are assumed to be unobserved. Thus they have to be estimated from the data. For this we denote for $k \geq 1$

$$\begin{aligned} \boldsymbol{\Sigma}_{\boldsymbol{Z}_k} &:= var(\boldsymbol{Z}_k), \\ \boldsymbol{Z}^k &:= (\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_k)', \\ \hat{\boldsymbol{X}}_{k|k} &:= L(\boldsymbol{X}_k|\boldsymbol{Z}^k). \end{aligned}$$

The estimation error is defined by

$$\tilde{\boldsymbol{X}}_{k|k} := \boldsymbol{X}_k - \hat{\boldsymbol{X}}_{k|k}.$$

The Kalman filter is a recursion for calculating the linear least square estimates $\hat{\boldsymbol{X}}_{k|k}$ and the corresponding mean square error matrix $\tilde{\boldsymbol{\Sigma}}_{k|k} := var(\tilde{\boldsymbol{X}}_{k|k})$.

Kalman one step prediction deals with the computation of

$$\begin{aligned} \hat{\boldsymbol{X}}_{k+1|k} &:= L(\boldsymbol{X}_{k+1}|\boldsymbol{Z}^k), \\ \tilde{\boldsymbol{\Sigma}}_{k+1|k} &:= var(\tilde{\boldsymbol{X}}_{k+1|k}), \end{aligned}$$

where $\tilde{\boldsymbol{X}}_{k+1|k} := \boldsymbol{X}_{k+1} - \hat{\boldsymbol{X}}_{k+1|k}$ denotes the one step prediction error for $k \geq 1$.

Let N denote the number of state vectors $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N$. Further define $\boldsymbol{Z}^N := (\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_N)'$. The Kalman fixed point smoother is a recursion algorithm for the computation of

$$\begin{aligned} \hat{\boldsymbol{X}}_{k|N} &:= L(\boldsymbol{X}_k|\boldsymbol{Z}^N), \\ \tilde{\boldsymbol{\Sigma}}_{k|N} &:= var(\tilde{\boldsymbol{X}}_{k|N}), \end{aligned}$$

where $\tilde{\boldsymbol{X}}_{k|N} := \boldsymbol{X}_k - \hat{\boldsymbol{X}}_{k|N}$ denotes the estimation error of $\boldsymbol{X}_k$ in terms of $\boldsymbol{Z}^N$ for $k = 1, \ldots, N$.

Kalman filtering, prediction and fixed point smoothing rely on decomposition of the data vector $\boldsymbol{Z}$ into so called innovations $\boldsymbol{\theta}$ which are defined for $k \geq 2$ as linear least squares prediction errors

$$\boldsymbol{\theta}_k := \boldsymbol{Z}_k - L(\boldsymbol{Z}_k|\boldsymbol{Z}^{k-1}), \tag{30}$$

where $\boldsymbol{Z}^{k-1}$ denotes the vector $(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_{k-1})'$. With Schneider (1986) (Satz (2-15)) the relation $\boldsymbol{\theta}_k \perp \boldsymbol{Z}^{k-1}$ holds and thus $\{\boldsymbol{\theta}_k, k \geq 2\}$ is serially orthogonal. Similarly to Chui and Chen (1999) we can derive two representations for $\boldsymbol{\theta}_k$. Using the Observation Equation (27) and the Transition Equation (26) we get for $k = 1, \ldots, N$

$$\begin{aligned}
\boldsymbol{Z}_k &= \bar{\boldsymbol{C}}_k\boldsymbol{X}_k + \boldsymbol{\eta}_k = \bar{\boldsymbol{C}}_k\boldsymbol{X}_k + \bar{\boldsymbol{N}}_{k-1}\boldsymbol{\eta}_{k-1} + \boldsymbol{\gamma}_k \\
&= \bar{\boldsymbol{C}}_k(\bar{\boldsymbol{A}}_{k-1}\boldsymbol{X}_{k-1} + \boldsymbol{\Gamma}_{k-1}\boldsymbol{\xi}_{k-1}) + \bar{\boldsymbol{N}}_{k-1}(\boldsymbol{Z}_{k-1} - \bar{\boldsymbol{C}}_{k-1}\boldsymbol{X}_{k-1}) + \boldsymbol{\gamma}_k \\
&= (\bar{\boldsymbol{C}}_k\bar{\boldsymbol{A}}_{k-1} - \bar{\boldsymbol{N}}_{k-1}\bar{\boldsymbol{C}}_{k-1})\boldsymbol{X}_{k-1} + \bar{\boldsymbol{N}}_{k-1}\boldsymbol{Z}_{k-1} + \bar{\boldsymbol{C}}_k\boldsymbol{\Gamma}_{k-1}\boldsymbol{\xi}_{k-1} + \boldsymbol{\gamma}_k \\
&= \boldsymbol{H}_{k-1}\boldsymbol{X}_{k-1} + \bar{\boldsymbol{N}}_{k-1}\boldsymbol{Z}_{k-1} + \bar{\boldsymbol{C}}_k\boldsymbol{\Gamma}_{k-1}\boldsymbol{\xi}_{k-1} + \boldsymbol{\gamma}_k, \text{ with} \qquad (31)
\end{aligned}$$

$$\boldsymbol{H}_{k-1} := \bar{\boldsymbol{C}}_k\bar{\boldsymbol{A}}_{k-1} - \bar{\boldsymbol{N}}_{k-1}\bar{\boldsymbol{C}}_{k-1}. \tag{32}$$

Since $\boldsymbol{\xi}_k \perp \boldsymbol{X}_k$ and $\boldsymbol{\xi}_k \perp \boldsymbol{\gamma}_k$ we conclude that $E(\boldsymbol{\xi}_k\boldsymbol{Z}_k') = \mathbf{0}$ and thus $L(\boldsymbol{\xi}_k|\boldsymbol{Z}^k) = \mathbf{0}$. From Assumption (29) it follows that $L(\boldsymbol{\gamma}_k|\boldsymbol{Z}^{k-1}) = 0$. Substitution of (31) into (30) yields

$$\begin{aligned}
\boldsymbol{\theta}_k &= \boldsymbol{Z}_k - L(\boldsymbol{H}_{k-1}\boldsymbol{X}_{k-1} + \bar{\boldsymbol{N}}_{k-1}\boldsymbol{Z}_{k-1} + \bar{\boldsymbol{C}}_k\boldsymbol{\Gamma}_{k-1}\boldsymbol{\xi}_{k-1} + \boldsymbol{\gamma}_k|\boldsymbol{Z}^{k-1}) \\
&= \boldsymbol{Z}_k - \boldsymbol{H}_{k-1}\hat{\boldsymbol{X}}_{k-1|k-1} - \bar{\boldsymbol{N}}_{k-1}\boldsymbol{Z}_{k-1}.
\end{aligned} \tag{33}$$

Further substitution leads to

$$\begin{aligned}
\boldsymbol{\theta}_k &= \boldsymbol{H}_{k-1}\boldsymbol{X}_{k-1} + \bar{\boldsymbol{N}}_{k-1}\boldsymbol{Z}_{k-1} + \bar{\boldsymbol{C}}_k\boldsymbol{\Gamma}_{\boldsymbol{k-1}}\boldsymbol{\xi}_{k-1} + \boldsymbol{\gamma}_k - \boldsymbol{H}_{k-1}\hat{\boldsymbol{X}}_{k-1|k-1} - \bar{\boldsymbol{N}}_{k-1}\boldsymbol{Z}_{k-1} \\
&= \boldsymbol{H}_{k-1}(\boldsymbol{X}_{k-1} - \hat{\boldsymbol{X}}_{k-1|k-1}) + \bar{\boldsymbol{C}}_k\boldsymbol{\Gamma}_{\boldsymbol{k-1}}\boldsymbol{\xi}_{k-1} + \boldsymbol{\gamma}_k.
\end{aligned} \tag{34}$$

Note that $\tilde{\boldsymbol{X}}_{k-1|k-1}$ is a linear function of $\boldsymbol{X}_{k-1}$ and $\boldsymbol{Z}^{k-1}$. Therefore $E(\tilde{\boldsymbol{X}}_{k-1|k-1}\boldsymbol{\xi}_{k-1}') = \mathbf{0}$ and $E(\tilde{\boldsymbol{X}}_{k-1|k-1}\boldsymbol{\gamma}_k') = \mathbf{0}$ holds for $k = 2, \ldots, N$. Thus

$$\boldsymbol{\Delta}_k := var(\boldsymbol{\theta}_k) = \boldsymbol{H}_{k-1}\tilde{\boldsymbol{\Sigma}}_{k-1|k-1}\boldsymbol{H}_{k-1}' + \bar{\boldsymbol{C}}_k\boldsymbol{\Gamma}_{k-1}var(\boldsymbol{\xi}_{k-1})\boldsymbol{\Gamma}_{k-1}'\bar{\boldsymbol{C}}_k' + var(\boldsymbol{\gamma}_k). \tag{35}$$

5.2 Kalman Filter and One Step Predictor for State Space Models with White System and Coloured Observation Noise

In Chui and Chen (1999) the Kalman filter is initialized by

$$\begin{aligned}
\hat{\boldsymbol{X}}_{1|1} &:= E(\boldsymbol{X}_1) - var(\boldsymbol{X}_1)\bar{\boldsymbol{C}}_1'\boldsymbol{\Sigma}_{\boldsymbol{Z}_1}^{-1}\bar{\boldsymbol{C}}_1\big(E(\boldsymbol{X}_1) - \boldsymbol{Z}_1\big), \\
\tilde{\boldsymbol{\Sigma}}_{1|1} &:= var(\boldsymbol{X}_1) - var(\boldsymbol{X}_1)\bar{\boldsymbol{C}}_1'\boldsymbol{\Sigma}_{\boldsymbol{Z}_1}^{-1}\bar{\boldsymbol{C}}_1 var(\boldsymbol{X}_1).
\end{aligned} \tag{36}$$

Then the following recursion formulas hold for the Model (26), (27) with $k \geq 2$ (Chui and Chen (1999), (5.21), page 73):

$$G_k := (\bar{A}_{k-1}\tilde{\Sigma}_{k-1|k-1}H'_{k-1} + \Gamma_{k-1}var(\xi_{k-1})\Gamma'_{k-1}\bar{C}'_k) \cdot \Delta_k^{-1}, \tag{37}$$

$$F_k := \bar{A}_{k-1} - G_k H_{k-1}, \tag{38}$$

$$\begin{aligned} \tilde{\Sigma}_{k|k} &= F_k\tilde{\Sigma}_{k-1|k-1}\bar{A}'_{k-1} + (I_{m_k} - G_k\bar{C}_k)\Gamma_{k-1}var(\xi_{k-1})\Gamma'_{k-1}, \\ \hat{X}_{k|k} &= \bar{A}_{k-1}\hat{X}_{k-1|k-1} + G_k\theta_k. \end{aligned} \tag{39}$$

In Equation (39) the Representation (33) is used. Note that the Representation (34) contains unobserved components. These are the filtering equations. With Transition Equation (26) we get for the one step predictions for $k \geq 1$

$$\begin{aligned} X_{k+1|k} &= \bar{A}_k\hat{X}_{k|k}, \\ \tilde{\Sigma}_{k+1|k} &= \bar{A}_k\tilde{\Sigma}_{k+1|k}\bar{A}'_k, \end{aligned} \tag{40}$$

since $L(\xi_k|Z^k) = 0$.

To obtain $\hat{X}_{k|N}$ and $\tilde{\Sigma}_{k|N}$ for $k = 1, \ldots, N$ we have to apply a Kalman fixed point smoothing algorithm. Since the algorithm makes use of the innovations θ_k, rather than of the data vectors Z_k, we can replace Z_k by θ_k successively to save memory space. The matrices Δ_k^{-1} and G_k have to be stored. It may be the case, that these matrices are not different for all k but $\Delta_k^{-1} = \Delta_l^{-1}$ holds for some $l = 2, \ldots, N$ and $G_k = G_m$ for some $m = 2, \ldots, N$. Then only the different matrices have to be stored, together with the information to which indices they correspond. The matrices $\{F_k, k = 2, \ldots, N\}$ need not to be stored, if sufficient memory space is a problem. It may be the case, that there are as well only a relative small number of different matrices F_k. On the other hand, they could be computed in the smoothing step again. In our applications $\bar{A}_k$ and $\bar{C}_k$, $k = 1, \ldots, N-1$ happen to be sparse matrices of simple structure, where matrices $\bar{N}_k$ have relatively small dimensions and may be only a small number of different $\bar{N}_k$ for $k = 1, \ldots, N-1$. Thus $\bar{A}_k, \bar{C}_k$, and $\bar{N}_k$ need not much memory space and computation of F_k and H_k can be done without much effort. Matrices $\bar{C}_k$ are needed anyway for the computation of $\hat{\xi}_{k|N}$ and $var(\tilde{\xi}_{k|N})$. Matrices $\{\Gamma_k, k = 1, \ldots, N-1\}$ are also needed in the smoothing step. $\{\Gamma_k, k = 1, \ldots, N-1\}$ are sparse selection matrices. For smoothing we need also the matrices $\{var(\gamma_k), k = 2, \ldots, N\}$ and $\{var(\xi_k), k = 1, \ldots, N-1\}$. Again, there may be only a relatively small number of different $var(\gamma_k)$ and $var(\xi_k)$.

5.3 Kalman Fixed Point Smoother for Linear State Space Models with White System and Coloured Observation Noise

Similarly to Durbin and Koopman (2001) (4.25) and (4.30) we define recursively for $k = N, \dots, 2$

$$\boldsymbol{P}_{k-1} := \boldsymbol{H}'_{k-1}\boldsymbol{\Delta}_k^{-1}\boldsymbol{H}_{k-1} + \boldsymbol{F}'_k\boldsymbol{P}_k\boldsymbol{F}_k, \tag{41}$$

$$\boldsymbol{R}_{k-1} := \boldsymbol{H}'_{k-1}\boldsymbol{\Delta}_k^{-1}\boldsymbol{\theta}_k + \boldsymbol{F}'_k\boldsymbol{R}_k, \tag{42}$$

where $\boldsymbol{P}_N = \mathbf{0}$ and $\boldsymbol{R}_N = \mathbf{0}$. If index $k = N-1, \dots, 1$ is interpreted as a time index then $\boldsymbol{P}_k$ is a linear function of the inverse variance matrices of innovations occurring after time k, and $\boldsymbol{R}_k$ is a linear function of innovations occurring after time k. Then we compute for $k = N-1, \dots, 1$ using $\boldsymbol{X}_{k|k}$ and $\tilde{\boldsymbol{\Sigma}}_{k|k}$ from the Kalman filtering (36) - (39)

$$\begin{aligned}\hat{\boldsymbol{X}}_{k|N} &= \hat{\boldsymbol{X}}_{k|k} + \tilde{\boldsymbol{\Sigma}}_{k|k}\boldsymbol{R}_k,\\ \tilde{\boldsymbol{\Sigma}}_{k|N} &= \tilde{\boldsymbol{\Sigma}}_{k|k} - \tilde{\boldsymbol{\Sigma}}_{k|k}\boldsymbol{P}_k\tilde{\boldsymbol{\Sigma}}_{k|k}.\end{aligned} \tag{43}$$

Note that $\hat{\boldsymbol{X}}_{N|N}$ and $\tilde{\boldsymbol{\Sigma}}_{N|N}$ were already computed by the filtering step (36) - (39). The derivation of (43) is given in Appendix B.

For the derivation of the score vector we need the smoothed disturbances $\hat{\boldsymbol{\gamma}}_{k|N} := L(\boldsymbol{\gamma}_k|\boldsymbol{Z}^N)$, $k = N, \dots, 2$ and $\hat{\boldsymbol{\xi}}_{m|N} := L(\boldsymbol{\xi}_{m|N}|\boldsymbol{Z}^N)$, $m = N-1, \dots, 1$. The corresponding smoothed estimation errors are denoted by

$$\begin{aligned}\tilde{\boldsymbol{\gamma}}_{k|N} &:= \boldsymbol{\gamma}_k - L(\boldsymbol{\gamma}_k|\boldsymbol{Z}^N),\\ \tilde{\boldsymbol{\xi}}_{m|N} &:= \boldsymbol{\xi}_m - L(\boldsymbol{\xi}_{m|N}|\boldsymbol{Z}^N).\end{aligned}$$

The mean squared error matrices $var(\tilde{\boldsymbol{\gamma}}_{k|N})$ and $var(\tilde{\boldsymbol{\xi}}_{m|N})$ are also needed for the derivation of the score vector. Computation of the smoothed disturbances and the corresponding mean squared error matrices can be done using the following recursion formulas:

$$\begin{aligned}\hat{\boldsymbol{\gamma}}_{k|N} &= var(\boldsymbol{\gamma}_k)\big(\boldsymbol{\Delta}_k^{-1}\boldsymbol{\theta}_k - \boldsymbol{G}'_k\boldsymbol{R}_k\big),\\ var(\tilde{\boldsymbol{\gamma}}_{k|N}) &= var(\boldsymbol{\gamma}_k) - var(\boldsymbol{\gamma}_k)\big(\boldsymbol{\Delta}_k^{-1} - \boldsymbol{G}'_k\boldsymbol{P}_k\boldsymbol{G}_k\big)var(\boldsymbol{\gamma}_k),\end{aligned} \tag{44}$$

$$\begin{aligned}\hat{\boldsymbol{\xi}}_{k|N} &= var(\boldsymbol{\xi}_k)\boldsymbol{\Gamma}'_k\bar{\boldsymbol{C}}'_{k+1}\boldsymbol{\Delta}_{k+1}^{-1}\boldsymbol{\theta}_{k+1} + var(\boldsymbol{\xi}_k)(\boldsymbol{\Gamma}_k - \boldsymbol{G}_{k+1}\bar{\boldsymbol{C}}_{k+1})'\boldsymbol{R}_{k+1},\\ var(\tilde{\boldsymbol{\xi}}_{k|N}) &= var(\boldsymbol{\xi}_k) - var(\boldsymbol{\xi}_k)\boldsymbol{\Gamma}'_k\bar{\boldsymbol{C}}'_{k+1}\boldsymbol{\Delta}_{k+1}^{-1}\bar{\boldsymbol{C}}_{k+1}\boldsymbol{\Gamma}_k var(\boldsymbol{\xi}_k)\\ &\quad - var(\boldsymbol{\xi}_k)(\boldsymbol{\Gamma}_k - \boldsymbol{G}_{k+1}\bar{\boldsymbol{C}}_{k+1})'\boldsymbol{P}_{k+1}(\boldsymbol{\Gamma}_k - \boldsymbol{G}_{k+1}\bar{\boldsymbol{C}}_{k+1})var(\boldsymbol{\xi}_k).\end{aligned} \tag{45}$$

In Appendix C the derivation of the equations for $\hat{\boldsymbol{\gamma}}_{k|N}$ and $\hat{\boldsymbol{\xi}}_{k|N}$ and in Appendix D the derivation of the equations for $var(\tilde{\boldsymbol{\gamma}}_{k|N})$ and $var(\tilde{\boldsymbol{\xi}}_{k|N})$ are given.

5.4 Missing Observations

In our applications of Kalman filter and smoother algorithms for linear state space models with white system and coloured observation noise we might have to allow for missing observations, that is there is no observation connected with the state vector $\boldsymbol{X}_k$ for some $k = 1, \ldots, N$. In this case we define $\bar{\boldsymbol{C}}_k$ to be a zero matrix of appropriate dimension and we re-define (27), (29) and (32) more generally as:

$$
\begin{aligned}
\boldsymbol{Z}_k &:= \begin{cases} \bar{\boldsymbol{C}}_k \boldsymbol{X}_k + \boldsymbol{\eta}_k & \text{, for } \bar{\boldsymbol{C}}_k \neq \boldsymbol{0} \\ \boldsymbol{Z}_{k-1} & \text{, for } \bar{\boldsymbol{C}}_k = \boldsymbol{0}, \end{cases} \\
\boldsymbol{\eta}_k &:= \begin{cases} \bar{\boldsymbol{N}}_{k-1} \boldsymbol{\eta}_{k-1} + \boldsymbol{\gamma}_k & \text{, for } \bar{\boldsymbol{C}}_k \neq \boldsymbol{0} \\ \boldsymbol{\eta}_{k-1} & \text{, for } \bar{\boldsymbol{C}}_k = \boldsymbol{0}, \end{cases} \\
\boldsymbol{H}_{k-1} &:= \begin{cases} \bar{\boldsymbol{C}}_k \bar{\boldsymbol{A}}_{k-1} - \bar{\boldsymbol{N}}_{k-1} \bar{\boldsymbol{C}}_{k-1} & \text{, for } \bar{\boldsymbol{C}}_k \neq \boldsymbol{0} \\ \boldsymbol{0} & \text{, for } \bar{\boldsymbol{C}}_k = \boldsymbol{0}. \end{cases}
\end{aligned}
\tag{46}
$$

For $\bar{\boldsymbol{C}}_k = \boldsymbol{0}$ (46) implies $\bar{\boldsymbol{N}}_{k-1} = \boldsymbol{I}$ and $\boldsymbol{\gamma}_k = \boldsymbol{0}$. There is no new information related to $\boldsymbol{Z}_k$ and the innovation $\boldsymbol{\theta}_k = \boldsymbol{Z}_k - L(\boldsymbol{Z}_k | \boldsymbol{Z}^{k-1})$ is a zero vector with variance matrix $\boldsymbol{\Delta}_k = \boldsymbol{0}$. But derivations of Kalman filter and smoother algorithms hold also for the case, when the inverse variance matrices that occur are replaced by their generalized inverse (see e.g. Hamilton (1994), Section 4.1). For $\boldsymbol{\Delta}_k = \boldsymbol{0}$ the generalized inverse matrix is $\boldsymbol{\Delta}_k^- = \boldsymbol{0}$, implying for (37) and (38) $\boldsymbol{G}_k = \boldsymbol{0}$ and $\boldsymbol{F}_k = \bar{\boldsymbol{A}}_{k-1}$. The filtering and the smoothing equations can then be applied in both cases when observations connected to a specific state vector are observed or not.

5.5 Derivation of the Score Statistic

Since often model parameters are unknown in practice they have to be estimated. For linear state space models maximum likelihood estimation is commonly used, see e.g. Durbin and Koopman (2001) and Harvey (1987). Here the score statistic becomes important for the application of an EM algorithm or for numerical maximization. For this reason we give in this section a derivation of an analytic representation of the score statistic for linear state space models with white transition noise and coloured observation noise.

Let $\boldsymbol{\psi}$ denote the vector of the unknown parameters in a parameter space $\boldsymbol{\Omega}$. The parameter vector $\boldsymbol{\psi}$ might consist of unknown variance and covariance parameters and some nonnegative weights. Therefore we assume that $\boldsymbol{\Omega} \subset \mathbb{R}_+^q \times \mathbb{R}^r$ with $q, r \in \mathbb{N}$. The likelihood function of $\boldsymbol{\psi}$ formed from the observed data is given by

$$L(\boldsymbol{\psi}; \boldsymbol{Z}^N) := p(\boldsymbol{Z}^N; \boldsymbol{\psi}),$$

where $p(\boldsymbol{Z}^N; \boldsymbol{\psi})$ denotes the probability density function of $\boldsymbol{Z}^N$ in terms of the parameters $\boldsymbol{\psi}$. Similarly the likelihood function of $\boldsymbol{\psi}$ formed from the complete set of the unobserved state vectors $\boldsymbol{X}^N :=$

$(\boldsymbol{X}_1,\dots,\boldsymbol{X}_N)'$ and the observed data $\boldsymbol{Z}^N$ is given by

$$L(\boldsymbol{\psi};\boldsymbol{X}^N,\boldsymbol{Z}^N) := p(\boldsymbol{X}^N,\boldsymbol{Z}^N;\boldsymbol{\psi}),$$

where $p(\boldsymbol{X}^N,\boldsymbol{Z}^N;\boldsymbol{\psi})$ is the joint density of $\boldsymbol{X}^N$ and $\boldsymbol{Z}^N$ in terms of the parameters $\boldsymbol{\psi}$.

As in McLachlan and Krishnan (1997) we define the score statistic based on the observed data $\boldsymbol{Z}^N$ as

$$\begin{aligned} S(\boldsymbol{Z}^N;\boldsymbol{\psi}) &:= \frac{\partial log\ L(\boldsymbol{\psi};\boldsymbol{Z}^N)}{\partial\boldsymbol{\psi}}, \text{ and corresponding to the complete data } (\boldsymbol{X}^N,\boldsymbol{Z}^N) \text{ by} \\ S(\boldsymbol{X}^N,\boldsymbol{Z}^N;\boldsymbol{\psi}) &:= \frac{\partial log\ L(\boldsymbol{\psi};\boldsymbol{X}^N,\boldsymbol{Z}^N)}{\partial\boldsymbol{\psi}}. \end{aligned} \tag{47}$$

Under regularity conditions like continuity, the interchange of integration and differentiation is valid. For this case McLachlan and Krishnan (1997) show that for a fixed value $\bar{\boldsymbol{\psi}} \in \boldsymbol{\Omega}$ $S(\boldsymbol{Z}^N;\bar{\boldsymbol{\psi}})$ can be written as

$$S(\boldsymbol{Z}^N;\bar{\boldsymbol{\psi}}) = \frac{\partial}{\partial\boldsymbol{\psi}} E_{\bar{\boldsymbol{\psi}}}[log\ L(\boldsymbol{\psi};\boldsymbol{X}^N,\boldsymbol{Z}^N)]\,|_{\boldsymbol{\psi}=\bar{\boldsymbol{\psi}}} \tag{48}$$

We use (48) for the derivation of the analytic representation. For this we assume that $\boldsymbol{X}^N$ and $\boldsymbol{Z}^N$ are jointly normally distributed in addition to assumption (28). Further, $var(\boldsymbol{\xi}_k)$ and $var(\boldsymbol{\gamma}_m)$ are either non-singular or zero matrices, $k = 0,\dots,N-1$, $m = 1,\dots,N$. In the latter case the terms $var(\boldsymbol{\xi}_k)^{-1}$, $var(\boldsymbol{\gamma}_l)^{-1}$ in equations below have to be replaced by $\boldsymbol{0}$.

We will need the following result (see e.g. Kailath et al. (2000), Appendix 3.C): Let $\boldsymbol{U},\boldsymbol{V}$ be two jointly normally distributed random vectors. Then

$$\begin{aligned} L(\boldsymbol{U}|\boldsymbol{V}) &= E(\boldsymbol{U}|\boldsymbol{V}), \\ var[\boldsymbol{U} - L(\boldsymbol{U}|\boldsymbol{V})] &= var(\boldsymbol{U}|\boldsymbol{V}), \end{aligned} \tag{49}$$

where $E(\boldsymbol{U}|\boldsymbol{V})$ denotes the conditional expectation and $var(\boldsymbol{U}|\boldsymbol{V})$ the conditional variance of $\boldsymbol{U}$ given $\boldsymbol{V}$. Furthermore, the random vector $\boldsymbol{U}|\boldsymbol{V}$ conditioned on $\boldsymbol{V}$ is normally distributed with mean $E(\boldsymbol{U}|\boldsymbol{V})$ and variance matrix $var(\boldsymbol{U}|\boldsymbol{V})$.

Applying Bayes Theorem yields

$$p(\boldsymbol{X}^N,\boldsymbol{Z}^N;\boldsymbol{\psi}) = p(\boldsymbol{X}^N;\boldsymbol{\psi})p(\boldsymbol{Z}^N|\boldsymbol{X}^N;\boldsymbol{\psi}) = p(\boldsymbol{X}_1;\boldsymbol{\psi})\prod_{k=2}^N p(\boldsymbol{X}_k|\boldsymbol{X}_1,\dots\boldsymbol{X}_{k-1};\boldsymbol{\psi}). \tag{50}$$

Using (26), (28) and (49) we get

$$\begin{aligned} L(\boldsymbol{X}_k|\boldsymbol{X}_1,\dots,\boldsymbol{X}_{k+1}) = L(\boldsymbol{X}_k|\boldsymbol{X}_{k-1}) &= E(\boldsymbol{X}_k|\boldsymbol{X}_{k-1}) = \bar{\boldsymbol{A}}_{k-1}\boldsymbol{X}_{k-1}, \\ var(\boldsymbol{X}_k|\boldsymbol{X}_{k-1}) = var[\boldsymbol{X}_k - E(\boldsymbol{X}_k|\boldsymbol{X}_{k-1})] &= \boldsymbol{\Gamma}_{k-1} var(\boldsymbol{\xi}_{k-1})\boldsymbol{\Gamma}'_{k-1}. \end{aligned} \tag{51}$$

In addition we have

$$p(\boldsymbol{X}^N;\boldsymbol{\psi}) = p(\boldsymbol{X}_1;\boldsymbol{\psi})\prod_{k=2}^N p(\boldsymbol{X}_k|\boldsymbol{X}_1,\dots\boldsymbol{X}_{k-1};\boldsymbol{\psi}) = p(\boldsymbol{X}_1;\boldsymbol{\psi})\prod_{k=2}^N p(\boldsymbol{X}_k|\boldsymbol{X}_{k-1};\boldsymbol{\psi}), \tag{52}$$

where $p(\boldsymbol{X}_k|\boldsymbol{X}_{k-1};\boldsymbol{\psi})$ is a multivariate normal density with mean $\bar{\boldsymbol{A}}_{k-1}\boldsymbol{X}_{k-1}$ and variance matrix $\boldsymbol{\Gamma}_{k-1}var(\boldsymbol{\xi}_{k-1})\boldsymbol{\Gamma}'_{k-1}$. Using again Bayes Theorem we have

$$p(\boldsymbol{Z}^N|\boldsymbol{X}^N;\boldsymbol{\psi}) = p(\boldsymbol{Z}_1|\boldsymbol{X}^N;\boldsymbol{\psi})\prod_{k=2}^{N} p(\boldsymbol{Z}_k|\boldsymbol{Z}^{k-1},\boldsymbol{X}^N;\boldsymbol{\psi}).$$

Applying (27) and (28) together with (49) it follows

$$\begin{aligned}
E(\boldsymbol{Z}_1|\boldsymbol{X}_1) &= \bar{\boldsymbol{C}}_1 E(\boldsymbol{X}_1|\boldsymbol{X}_1) + E(\boldsymbol{\eta}_1|\boldsymbol{X}_1) = \bar{\boldsymbol{C}}_1\boldsymbol{X}_1,\\
var(\boldsymbol{Z}_1|\boldsymbol{X}_1) &= var[\boldsymbol{Z}_1 - E(\boldsymbol{Z}_1|\boldsymbol{X}_1)] = var(\boldsymbol{\eta}_1),\\
E(\boldsymbol{Z}_k|\boldsymbol{Z}^{k-1},\boldsymbol{X}^N) &= \bar{\boldsymbol{C}}_k E(\boldsymbol{X}_k|\boldsymbol{Z}^{k-1},\boldsymbol{X}^N) + E(\boldsymbol{\eta}_k|\boldsymbol{Z}^{k-1},\boldsymbol{X})\\
&= \bar{\boldsymbol{C}}_k E(\boldsymbol{X}_k|\boldsymbol{X}_k) + \boldsymbol{N}_{k-1}E(\boldsymbol{\eta}_{k-1}|\boldsymbol{Z}^{k-1},\boldsymbol{X}^N) + E(\boldsymbol{\gamma}_k|\boldsymbol{Z}^{k-1},\boldsymbol{X}^N)\\
&= \bar{\boldsymbol{C}}_k E(\boldsymbol{X}_k|\boldsymbol{X}_k) + \boldsymbol{N}_{k-1}E(\boldsymbol{\eta}_{k-1}|\boldsymbol{Z}_{k-1},\boldsymbol{X}_{k-1}) + \boldsymbol{0}\\
&= E(\boldsymbol{Z}_k|\boldsymbol{Z}_{k-1},\boldsymbol{X}_{k-1},\boldsymbol{X}_k),
\end{aligned} \tag{53}$$

since $\boldsymbol{\eta}_{k-1} = \boldsymbol{Z}_{k-1} - \bar{\boldsymbol{C}}_{k-1}\boldsymbol{X}_{k-1}$ is a linear function of $\boldsymbol{Z}_{k-1}$ and $\boldsymbol{X}_{k-1}$, and $\boldsymbol{\gamma}_k \perp \boldsymbol{Z}^{k-1}, \boldsymbol{X}^N$. This yields

$$\begin{aligned}
E(\boldsymbol{Z}_k|\boldsymbol{Z}^{k-1},\boldsymbol{X}^N) &= \bar{\boldsymbol{C}}_k\boldsymbol{X}_k + \boldsymbol{N}_{k-1}\boldsymbol{\eta}_{k-1},\\
var(\boldsymbol{Z}_k|\boldsymbol{Z}^{k-1},\boldsymbol{X}^N) &= var[\boldsymbol{Z}_k - E(\boldsymbol{Z}_k|\boldsymbol{Z}^{k-1},\boldsymbol{X}^N)] = var(\boldsymbol{\gamma}_k).
\end{aligned} \tag{54}$$

We define

$$\boldsymbol{\xi}_0 := \boldsymbol{X}_1 - E\boldsymbol{X}_1 \text{ and } \boldsymbol{\gamma}_1 := \boldsymbol{\eta}_1.$$

Since $\boldsymbol{\Gamma}_k$ is a selection matrix or a block selection matrix the relation $\boldsymbol{\Gamma}'_k\boldsymbol{\Gamma}_k = \boldsymbol{I}$ holds and it follows that

$$\boldsymbol{\xi}_k = \boldsymbol{\Gamma}'_k\boldsymbol{\Gamma}_k\boldsymbol{\xi}_k = \boldsymbol{\Gamma}'_k(\boldsymbol{X}_{k+1} - \boldsymbol{\Gamma}_k\boldsymbol{X}_k).$$

The complete log likelihood can now be calculated as follows

(55)

$$\begin{aligned}
log\ L(\boldsymbol{\psi};\boldsymbol{X}^N,\boldsymbol{Z}^N) &= log\ p(\boldsymbol{X}^N,\boldsymbol{Z}^N;\boldsymbol{\psi}) = log\ p(\boldsymbol{X}^n;\boldsymbol{\psi}) + log\ p(\boldsymbol{Z}^N|\boldsymbol{X}^N;\boldsymbol{\psi})\\
&= log\ p(\boldsymbol{X}_1;\boldsymbol{\psi}) + \textstyle\sum_{k=2}^N log\ p(\boldsymbol{X}_k|\boldsymbol{X}_{k-1};\boldsymbol{\psi}) + log\ p(\boldsymbol{Z}_1|\boldsymbol{X}_1;\boldsymbol{\psi}) + \sum_{k=2}^N log\ p(\boldsymbol{Z}_k|\boldsymbol{Z}_{k-1},\boldsymbol{X}_{k-1},\boldsymbol{X}_k;\boldsymbol{\psi})\\
&= \text{const.} - \tfrac{1}{2}\Bigg[log(|var(\boldsymbol{\xi}_0)|) + \boldsymbol{\xi}'_0 var(\boldsymbol{\xi}_0)^{-1}\boldsymbol{\xi}_0 + log(|var(\boldsymbol{\gamma}_1)|) + \boldsymbol{\gamma}'_1 var(\boldsymbol{\gamma}_1)^{-1}\boldsymbol{\gamma}_1\\
&\quad + \textstyle\sum_{k=2}^N \left[log(|var(\boldsymbol{\xi}_{k-1})|) + \boldsymbol{\xi}'_{k-1}var(\boldsymbol{\xi}_{k-1})^{-1}\boldsymbol{\xi}_{k-1} + log(|var(\boldsymbol{\gamma}_k)|) + \boldsymbol{\gamma}'_k var(\boldsymbol{\gamma}_k)^{-1}\boldsymbol{\gamma}_k\right]\Bigg]\\
&= \text{const.} - \tfrac{1}{2}\textstyle\sum_{k=1}^N \left[log(|var(\boldsymbol{\xi}_{k-1})|) + log(|var(\boldsymbol{\gamma}_k)|) + \boldsymbol{\xi}'_{k-1}var(\boldsymbol{\xi}_{k-1})^{-1}\boldsymbol{\xi}_{k-1} + \boldsymbol{\gamma}'_k var(\boldsymbol{\gamma}_k)^{-1}\boldsymbol{\gamma}_k.\right.
\end{aligned}$$

Since for a random vector $\boldsymbol{V}$ of size n and $\boldsymbol{M} \in \mathbb{R}^{n\times n}$ symmetric

$$E(\boldsymbol{V}'\boldsymbol{M}\boldsymbol{V}) = tr[\boldsymbol{M}var(\boldsymbol{V}) + E(\boldsymbol{V})E(\boldsymbol{V})'] \tag{56}$$

(Seber (1977), Theorem 1.7) holds we have for $\boldsymbol{\psi} = \bar{\boldsymbol{\psi}}$

(57)

$$S(\boldsymbol{Z}^N;\bar{\boldsymbol{\psi}}) = \frac{\partial}{\partial \boldsymbol{\psi}} E_{\bar{\boldsymbol{\psi}}}\left[log\ L(\boldsymbol{\psi};\boldsymbol{X}^N,\boldsymbol{Z}^N)\right]\Big|_{\boldsymbol{\psi}=\bar{\boldsymbol{\psi}}}$$

$$= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\psi}} \sum_{k=1}^{N}\left\{log(|var(\boldsymbol{\xi}_{k-1})|) + log(|var(\boldsymbol{\gamma}_k)|) + E_{\bar{\boldsymbol{\psi}}}\left[\boldsymbol{\xi}'_{k-1}var(\boldsymbol{\xi}_{k-1})^{-1}\boldsymbol{\xi}_{k-1}|\boldsymbol{Z}^N\right] + E_{\bar{\boldsymbol{\psi}}}\left[\boldsymbol{\gamma}'_k var(\boldsymbol{\gamma}_k)^{-1}\boldsymbol{\gamma}_k|\boldsymbol{Z}^N\right]\right\}\Bigg|_{\boldsymbol{\psi}=\bar{\boldsymbol{\psi}}}$$

$$= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\psi}} \sum_{k=1}^{N}\Big[log(|var(\boldsymbol{\xi}_{k-1})|) + log(|var(\boldsymbol{\gamma}_k)|) + tr\left(\hat{\boldsymbol{\xi}}_{k-1|N}\hat{\boldsymbol{\xi}}'_{k-1|N} + var(\boldsymbol{\xi}_{k-1})^{-1}var(\boldsymbol{\xi}_{k-1}|\boldsymbol{Z}^N)\right)$$

$$+ tr\left(\hat{\boldsymbol{\gamma}}_{k|N}\hat{\boldsymbol{\gamma}}'_{k|N} + var(\boldsymbol{\gamma}_k)^{-1}var(\boldsymbol{\gamma}_k|\boldsymbol{Z}^N)\right)\Big]\Bigg|_{\boldsymbol{\psi}=\bar{\boldsymbol{\psi}}}$$

Note that $E(\boldsymbol{\xi}_{k-1}|\boldsymbol{Z}^N) = \hat{\boldsymbol{\xi}}_{k-1|N}$ and $E(\boldsymbol{\gamma}_k|\boldsymbol{Z}^N) = \hat{\boldsymbol{\gamma}}_{k|N}$. Since $\hat{\boldsymbol{\xi}}_{k-1|N}$, $var(\boldsymbol{\xi}_{k-1}|\boldsymbol{Z}^N) = var(\tilde{\boldsymbol{\xi}}_{k-1|N})$ and $\hat{\boldsymbol{\gamma}}_{k|N}$, $var(\boldsymbol{\gamma}_k|\boldsymbol{Z}^N) = var(\tilde{\boldsymbol{\gamma}}_{k|N})$ were computed in (44) and (45), respectively, under the assumption $\boldsymbol{\psi} = \bar{\boldsymbol{\psi}}$ these terms do not vary with $\boldsymbol{\psi}$. Thus in (57) only the terms in $var(\boldsymbol{\xi})$ and $var(\boldsymbol{\gamma})$ require differentiation with respect to $\boldsymbol{\psi}$.

In the case of missing observations in the state space model the number of state vectors $N_{\boldsymbol{X}}$ is greater than the number of observations $\boldsymbol{N_Z}$. We have augmented the original data vector $\boldsymbol{Z}^{N_{\boldsymbol{Z}}}$ by some vectors $\boldsymbol{Z}_k = \boldsymbol{Z}_{k-1}$ when there was no observation connected with the state vector $\boldsymbol{X}_k$, $k = 1,\ldots,N_{\boldsymbol{X}}$. Let $\boldsymbol{Z}^{N_{\boldsymbol{X}}}$ denote the augmented data vector $\boldsymbol{Z}^{N_{\boldsymbol{Z}}}$. Then for the likelihood $L(\boldsymbol{\psi};\boldsymbol{X}^{N_{\boldsymbol{X}}},\boldsymbol{Z}^{N_{\boldsymbol{X}}}) = L(\boldsymbol{\psi};\boldsymbol{X}^{N_{\boldsymbol{X}}},\boldsymbol{Z}^{N_{\boldsymbol{Z}}})$ holds. Thus we can skip the corresponding terms in Equation (57) when $\boldsymbol{Z}_k = \boldsymbol{Z}_{k-1}$. Therefore we write Equation (57) as

(58)

$$S(\boldsymbol{Z}^{N_{\boldsymbol{Z}}};\bar{\boldsymbol{\psi}}) = -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\psi}}\sum_{k=1}^{N_{\boldsymbol{X}}}\left[log(|var(\boldsymbol{\xi}_{k-1})|) + tr\left(\hat{\boldsymbol{\xi}}_{k-1|N_{\boldsymbol{Z}}}\hat{\boldsymbol{\xi}}'_{k-1|N_{\boldsymbol{Z}}} + var(\boldsymbol{\xi}_{k-1})^{-1}var(\boldsymbol{\xi}_{k-1}|\boldsymbol{Z}^{N_{\boldsymbol{Z}}})\right)\right]$$

$$+\sum_{k=1}^{N_{\boldsymbol{Z}}}\left[log(|var(\boldsymbol{\gamma}_k)|N_{\boldsymbol{Z}}) + tr\left(\hat{\boldsymbol{\gamma}}_{k|N_{\boldsymbol{Z}}}\hat{\boldsymbol{\gamma}}'_{k|N_{\boldsymbol{Z}}} + var(\boldsymbol{\gamma}_k)^{-1}var(\boldsymbol{\gamma}_k|\boldsymbol{Z}^{N_{\boldsymbol{Z}}})\right)\right]\Bigg|_{\boldsymbol{\psi}=\bar{\boldsymbol{\psi}}}$$

6 Autoregressive Models on Acyclic Directed Graphs

6.1 Introduction

Huang and Cressie (2001) relaxed the tree-structure and allowed for structures on more general acyclic directed graphs. As Huang et al. (2002) they assume white system and observation noise. For these models they derive so-called junction trees. The tree-structured Kalman filter and smoother work now on these junction trees. We took these models as a starting point to formulate a model for time series, omitting mass balance. For this we drop the assumption of nested partitioning and use an overlapping arrangement of sub-intervals instead. We will now define directed acyclic graphs and required additional notation. Finally we define stochastic processes, indexed by nodes of a specific class of acyclic directed graphs. System noise and observation noise are then modeled in linear state space form where we do not require white noise but can allow for coloured noise.

6.2 Overlapping Interval Arrangement and Acyclic Directed Graphs

Definition 6.1. *Overlapping Interval Arrangement*

Let interval $I \subset \mathbb{R}$ with length $|I| > 0$, a natural number $J \in \mathbb{N}$ and a family of natural numbers $(N_j)_{j=1}^{J}$ with $N_{j+1} \geq N_j$ $(j = 1, \dots, J)$ be given. A collection of sub-intervals $\{I_{j,k} \subset I : j = 1, \dots, J,\ k = 1, \dots, N_j\}$ is called an overlapping interval arrangement on I, if the following conditions hold:

(i) $$|I_{j,k}| > 0 \ \ (j = 1, \dots, J;\ k = 1, \dots, N_j),$$

(ii) $$I_{j,k} \cap I_{j,k+1} \neq \emptyset \ \ (j = 1, \dots, J;\ k = 1, \dots, N_j - 1),$$

(iii) $$\bigcup_{k=1}^{N_j} I_{j,k} = I \ \ (j = 1, \dots, J),$$

We consider the indices of these sub-intervals as nodes in a directed acyclic graph, which is defined in graph theory as follows:

Definition 6.2. *Let a finite set of nodes T and a set of edges $E \in V \times V$ be given. If for all $(v, v') \in E$ it follows that $(v', v) \notin E$, then the pair (T, V) is called an acyclic graph. A path of length $k \geq 1$ from v_0 to $v_k \in T$ is a sequence of nodes $v_0, v_1, \dots, v_k$ such that (v_i, v_{i+1}) is an edge for each $i = 0, \dots, k-1$. A cycle of length $k \geq 1$ is a path $v_0, v_1, \dots, v_k$ such that $v_0 = v_k$. An acyclic directed graph is a directed graph that has no cycles in it. For a directed edge (v, v'), v is said to be a parent of v', and v' is said to be a child of v. A node v of a directed graph is said to be a root, if it has no parent, and it is called a terminal node, if it has no children.*

These definitions were used by Huang and Cressie (2001). Note that with this definition a graph can have more than one root. For the models we consider we make additional definitions:
We say that a node v is on scale $j = 1$ if a root is the parent of v. Roots are then nodes on scale $j = 0$. The further scales are defined recursively: We say that a node is on scale $j + 1$ if its parent or parents are on scale j. We call the scale J with only terminal nodes on it the finest scale. The number of nodes on a scale $j = 0, \dots, J$ is denoted by N_j. The numbering of nodes on a scale j starts with 1 at the left and proceeds to the right up to N_j. Thus the kth node on the scale $j = 0, \dots, J$ can be denoted by the pair (j, k).

The nodes are allowed to have up to two parents:
A node (j, k) is called a left parent of the node $(j + 1, k')$ if (j, k) is the only parent or if (j, k) is a parent of $(j + 1, k')$ and if there is a node $(j, k + 1)$ that is also a parent of $(j + 1, k')$. We denote the left parent of $(j + 1, k')$ by $lpa(j + 1, k')$. A node (j, k) is called a right parent of $(j + 1, k')$, if (j, k) is the only parent or if (j, k) is a parent of $(j + 1, k')$ and if there is a node $(j, k - 1)$ that is also a parent of

$(j+1,k')$. We denote the right parent of $(j+1,k')$ by $rpa(j+1,k')$.

A path from the root to the terminal node (J,k) is called the left path from the root to (J,k) if only (J,k) and left parents of nodes of the path are on the path. A path from the root to the terminal node (J,k) is called the right path from the root to (J,k) if only (J,k) and right parents of nodes of the path are on the path.

Let P_j denote the number of nodes on scale j that are left parents for $j=0,\dots,J-1$ and N_j the number of nodes on scale j, $j=0,\dots,J$. For $j=1,\dots,J$ we can specify $K_j\in\mathbb{N}$, $P_{j-1}\le K_j\le N_j$ and $l_{j,k},r_{j,k}\in\mathbb{N}$ with $l_{j,k}\le r_{j,k}$ for $k=1,\dots,K_j$ such that the nodes $(j,l_{j,k}),\dots,(j,r_{j,k})$ have the same left parent. We set $\nu_{j,k}:=\big\{(j,l_{j,k}),\dots,(j,r_{j,k})\big\}$ and denote the cardinal number of $\nu_{j,k}$ by $n_{j,k}$.

For the models discussed below we will consider only acyclic directed graphs with one root and where all the terminal nodes are on the finest scale J only. Furthermore we assume that the nodes of scale $j=2,\dots,J$ have up to two parents. For easier reference we call such graphs two-parent-terminal graphs.

The whole observation interval and the sub-intervals obtained by overlapping interval arrangement can now be indexed by the nodes of an acyclic directed graph.

Example 3. *To illustrate these notations we give an example with $J=2$ and three sub-intervals on the scale $j=1$ and six sub-intervals on the scale $j=2$, see Figure 3.*

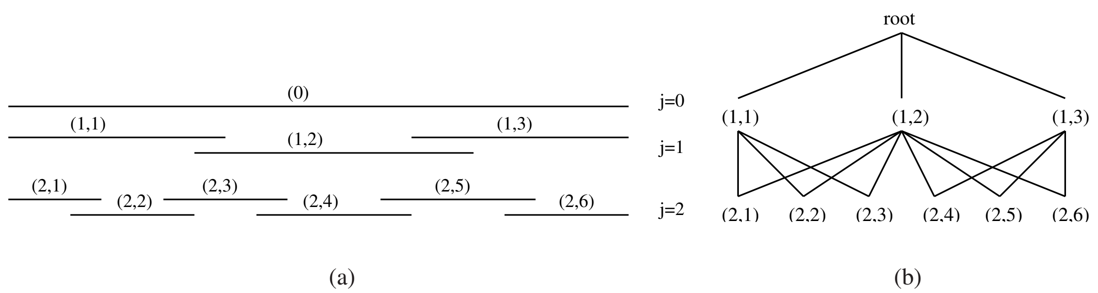

Figure 3: Overlapping Interval Arrangement (a) and Corresponding Acyclic Directed Graph (b) for Example 3

Since we have no nested partitioning as in Definition 3.1 but overlapping sub-intervals we don't have to pay attention to the linearity of the integral and thus no mass balance is needed. One possible acyclic directed graph, corresponding to this overlapping interval arrangement is given in Figure 3(b). The nodes on $j=1$ have only one parent, the root, while each node on $j=2,\dots,J$ has a left parent lp and a right parent rp.

Following Huang and Cressie (2001) we define a stochastic process $\boldsymbol{Y} := \{Y_0, Y_{j,k}, \{j,k\} \in T\}$ indexed by the nodes T of a two-parent-terminal graph, starting with Y_0 at the root:

(59) Transition Eq. $j = 1$: $\quad Y_{1,k} = Y_0 + W_{1,k} \quad (k = 1, \ldots, N_1).$

(60) Transition Eq. $j = 2, \ldots, J$: $\quad Y_{j,k} = \alpha_{j,k} Y_{lp(j,k)} + \beta_{j,k} Y_{rp(j,k)} + W_{j,k} \quad (k = 1, \ldots, N_j).$

(61) Observation Eq.: $\quad Z_k = c_k Y_{J,K} + Q_k,$

where $c_k$is a real number and Q_k a random variable for $k = 1, \ldots, N_J$. We also say that $Y_{j,k}$ is a child of $Y_{lpa(j,k)}$ and $Y_{rpa(j,k)}$ and that $Y_{lpa(j,k)}$, $Y_{rpa(j,k)}$ are parents of $Y_{j,k}$ for $j = 2, \ldots, J$, $k = 1, \ldots, K_j$. Note that if we set $\alpha_{j,k} = 0 \; \forall \; j,k$ and consider e. g. right parents as the only parents in the graph, we are back to trees discussed in Chapter 4.

In this model observations again are only associated with the finest scale J. Further assumptions are: $Y_0, W_{j,k}, Q_k$ are zero mean random variables with finite variance, $\alpha_{j,k}, \beta_{j,k} \in \mathbb{R}_+ := \{x \in \mathbb{R} : x \geq 0\}$ with $\alpha_{j,k} + \beta_{j,k} = 1$, $W_{j,k} \perp Y_0$, $W_{j,k} \perp \boldsymbol{Q}_l$, $\boldsymbol{Q}_l \perp Y_0$ and $W_{j,k} \perp W_{i,l}$, $(j \neq i)$.

For j fixed, the transition noise $\{W_{j,k}, k = 1, \ldots n_j\}$ is modeled in linear state space form:

$$\boldsymbol{W}_{j,k+1} = \boldsymbol{B}_{j,k} \boldsymbol{W}_{j,k} + \boldsymbol{V}_{j,k}, \;\; k = 1, \ldots, K_j - 1 \tag{62}$$

where the elements of the random vector $\boldsymbol{W}_{j,k}$ are the elements of the set $\{W_{j,\lambda} : (j,\lambda) \in \nu_{j,k}\}$, and $\boldsymbol{B}_{j,k} \in \mathbb{R}^{n_{j,k+1} \times n_{j,k}}$ and $\boldsymbol{V}_{j,k}$ is a zero mean random vector with $n_{j,k+1}$ elements. Further we assume $\{\boldsymbol{V}_{j,k}, k = 1, \ldots, K_j\}$ to be a family of uncorrelated random vectors, also uncorrelated to $\boldsymbol{W}_{j,1}$. The state space representation (62) implies

$$L(\boldsymbol{W}_{j,k+1} | \boldsymbol{W}_{j,1}, \ldots, \boldsymbol{W}_{j,k}) = L(\boldsymbol{W}_{j,k+1} | \boldsymbol{W}_{j,k}). \tag{63}$$

Kailath et al. (2000) call this weak Markov property, since in general the Markov property is defined by conditional independence rather then by covariance.

The observation noise $\{Q_m, m = 1, \ldots, N_J\}$ is modeled in a similar manner with

$$\boldsymbol{Q}_{k+1} = \boldsymbol{D}_k \boldsymbol{Q}_k + \boldsymbol{U}_k, \;\; k = 1, \ldots, K_J - 1. \tag{64}$$

where the elements of the random vector $\boldsymbol{Q}_k$ are the elements of the set $\{Q_\lambda : (J,\lambda) \in \nu_{J,k}\}$, $\boldsymbol{D}_k \in \mathbb{R}^{n_{J,k+1} \times n_{J,k}}$ and $\boldsymbol{U}_k$ a zero mean random vector with $\boldsymbol{U}_k \perp \boldsymbol{Q}_k$ and $\boldsymbol{U}_k \perp \boldsymbol{U}_l$ for $k \neq l$.

6.3 Linear State Space Representation for Models on Acyclic Directed Graphs

We explain how to find a linear state space representation with white system and coloured observation noise by a simple example.

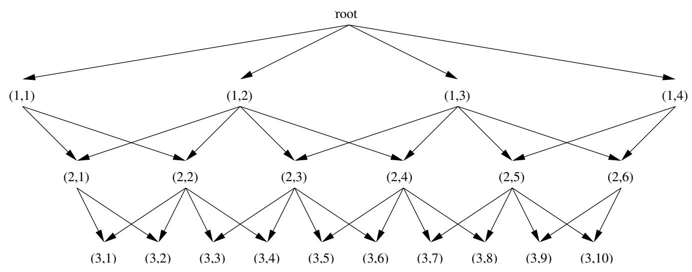

Figure 4: Two-parent-terminal Graph for Example 4

Example 4. *Consider the graph shown in Figure 4. On the finest scale $J = 3$ the nodes $(3, 1)$ and $(3, 2)$ have both parents in common. Thus we set $\nu_{3,1} := \{(3, 1), (3, 2)\}$, $\boldsymbol{Y}_{3,1} := (Y_{3,1}, Y_{3,2})'$ and $\boldsymbol{W}_{3,1} := (W_{3,1}, W_{3,2})'$. Proceeding in the same manner on scale 3 we get random vectors $\boldsymbol{Y}_{3,1}, \ldots, \boldsymbol{Y}_{3,5}$ and $\boldsymbol{W}_{3,1}, \ldots, \boldsymbol{W}_{3,5}$. Observations Z_m, $m = 1, \ldots, 10$, are similarly compounded into vectors:*

$$\boldsymbol{Z}_1 := (Z_1, Z_2)', \ldots, \boldsymbol{Z}_5 := (Z_9, Z_{10})',$$
$$\boldsymbol{Q}_1 := (Q_1, Q_2)', \ldots, \boldsymbol{Q}_5 := (Q_9, Q_{10})'.$$

On scale $j = 2$ we get the random vectors $\boldsymbol{W}_{2,1} := (W_{2,1}, W_{2,2})'$, $\boldsymbol{W}_{2,2} := (W_{2,3}, W_{2,4})'$, $\boldsymbol{W}_{2,3} := (W_{2,5}, W_{2,6})'$ and $\boldsymbol{Y}_{2,1} := (Y_{2,1}, Y_{2,2})'$, $\boldsymbol{Y}_{2,2} := (Y_{2,3}, Y_{2,4})'$, $\boldsymbol{Y}_{2,3} := (Y_{2,5}, Y_{2,6})'$. The nodes on scale $j = 1$ have the root as parent. Therefore we define $\boldsymbol{W}_{1,1} := (W_{1,1}, W_{1,2})'$, $\boldsymbol{W}_{1,2} := (W_{1,3}, W_{1,4})'$ and $\boldsymbol{Y}_{1,1} := (Y_{1,1}, Y_{1,2})'$, $\boldsymbol{Y}_{1,2} := (Y_{1,3}, Y_{1,4})'$. Using Transition Equations (59) and (60) yields the *transition equations for the vectorized model:*

$$
\begin{aligned}
\boldsymbol{Y}_{1,1} &= \begin{pmatrix} Y_{1,1} \\ Y_{1,2} \end{pmatrix} = \begin{pmatrix} Y_0+W_{1,1} \\ Y_0+W_{1,2} \end{pmatrix} = Y_0 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} W_{1,1} \\ W_{1,2} \end{pmatrix}, \\
\boldsymbol{Y}_{1,2} &= \begin{pmatrix} Y_{1,3} \\ Y_{1,4} \end{pmatrix} = \begin{pmatrix} Y_0+W_{1,3} \\ Y_0+W_{1,4} \end{pmatrix} = Y_0 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} W_{1,3} \\ W_{1,4} \end{pmatrix}, \\
\boldsymbol{Y}_{2,1} &= \begin{pmatrix} Y_{2,1} \\ Y_{2,2} \end{pmatrix} = \begin{pmatrix} \alpha_{2,1}Y_{1,1}+\beta_{2,1}Y_{1,2}+W_{2,1} \\ \alpha_{2,2}Y_{1,1}+\beta_{2,2}Y_{1,2}+W_{2,2} \end{pmatrix} = \begin{pmatrix} \alpha_{2,1} & \beta_{2,1} \\ \alpha_{2,2} & \beta_{2,2} \end{pmatrix} \boldsymbol{Y}_{1,1}+\boldsymbol{W}_{2,1} =: \boldsymbol{A}_{2,1}\boldsymbol{Y}_{1,1}+\boldsymbol{W}_{2,1}, \\
\boldsymbol{Y}_{2,2} &= \begin{pmatrix} Y_{2,3} \\ Y_{2,4} \end{pmatrix} = \begin{pmatrix} \alpha_{2,3}Y_{1,2}+\beta_{2,3}Y_{1,3}+W_{2,3} \\ \alpha_{2,4}Y_{1,2}+\beta_{2,4}Y_{1,3}+W_{2,4} \end{pmatrix} = \begin{pmatrix} 0 & \alpha_{2,3} \\ 0 & \alpha_{2,4} \end{pmatrix} \boldsymbol{Y}_{1,1} + \begin{pmatrix} \beta_{2,3} & 0 \\ \beta_{2,4} & 0 \end{pmatrix} \boldsymbol{Y}_{1,2}+\boldsymbol{W}_{2,2} \\
&=: \boldsymbol{A}^l_{2,2}\boldsymbol{Y}_{1,1}+\boldsymbol{A}^r_{2,2}\boldsymbol{Y}_{1,2}+\boldsymbol{W}_{2,2}, \\
\boldsymbol{Y}_{2,3} &= \begin{pmatrix} Y_{2,5} \\ Y_{2,6} \end{pmatrix} = \begin{pmatrix} \alpha_{2,5}Y_{1,3}+\beta_{2,5}Y_{1,4}+W_{2,5} \\ \alpha_{2,6}Y_{1,3}+\beta_{2,6}Y_{1,4}+W_{2,6} \end{pmatrix} = \begin{pmatrix} \alpha_{2,5} & \beta_{2,5} \\ \alpha_{2,6} & \beta_{2,6} \end{pmatrix} \boldsymbol{Y}_{1,2}+\boldsymbol{W}_{2,3} =: \boldsymbol{A}_{2,3}\boldsymbol{Y}_{1,2}+\boldsymbol{W}_{2,3}, \\
\boldsymbol{Y}_{3,1} &= \begin{pmatrix} Y_{3,1} \\ Y_{3,2} \end{pmatrix} = \begin{pmatrix} \alpha_{3,1}Y_{2,1}+\beta_{3,1}Y_{2,2}+W_{3,1} \\ \alpha_{3,2}Y_{2,1}+\beta_{3,2}Y_{2,2}+W_{3,2} \end{pmatrix} = \begin{pmatrix} \alpha_{3,1} & \beta_{3,1} \\ \alpha_{3,2} & \beta_{3,2} \end{pmatrix} \boldsymbol{Y}_{2,1}+\boldsymbol{W}_{3,1} =: \boldsymbol{A}_{3,1}\boldsymbol{Y}_{2,1}+\boldsymbol{W}_{3,1}, \\
\boldsymbol{Y}_{3,2} &= \begin{pmatrix} Y_{3,3} \\ Y_{3,4} \end{pmatrix} = \begin{pmatrix} 0 & \alpha_{3,3} \\ 0 & \alpha_{3,4} \end{pmatrix} \boldsymbol{Y}_{2,1} + \begin{pmatrix} \beta_{3,3} & 0 \\ \beta_{3,4} & 0 \end{pmatrix} \boldsymbol{Y}_{2,2}+\boldsymbol{W}_{3,2} =: \boldsymbol{A}^l_{3,2}\boldsymbol{Y}_{2,1}+\boldsymbol{A}^r_{3,2}\boldsymbol{Y}_{2,2}+\boldsymbol{W}_{3,2}, \\
\boldsymbol{Y}_{3,3} &= \begin{pmatrix} Y_{3,5} \\ Y_{3,6} \end{pmatrix} = \begin{pmatrix} \alpha_{3,5} & \beta_{3,5} \\ \alpha_{3,6} & \beta_{3,6} \end{pmatrix} \boldsymbol{Y}_{2,2}+\boldsymbol{W}_{3,3} =: \boldsymbol{A}_{3,3}\boldsymbol{Y}_{2,2}+\boldsymbol{W}_{3,3}, \\
\boldsymbol{Y}_{3,4} &= \begin{pmatrix} Y_{3,7} \\ Y_{3,8} \end{pmatrix} = \begin{pmatrix} 0 & \alpha_{3,7} \\ 0 & \alpha_{3,8} \end{pmatrix} \boldsymbol{Y}_{2,2} + \begin{pmatrix} \beta_{3,7} & 0 \\ \beta_{3,7} & 0 \end{pmatrix} \boldsymbol{Y}_{2,3}+\boldsymbol{W}_{3,4} =: \boldsymbol{A}^l_{3,4}\boldsymbol{Y}_{2,2}+\boldsymbol{A}^r_{3,4}\boldsymbol{Y}_{2,3}+\boldsymbol{W}_{3,4}, \\
\boldsymbol{Y}_{3,5} &= \begin{pmatrix} Y_{3,9} \\ Y_{3,10} \end{pmatrix} = \begin{pmatrix} \alpha_{3,9} & \beta_{3,9} \\ \alpha_{3,10} & \beta_{3,10} \end{pmatrix} \boldsymbol{Y}_{2,3}+\boldsymbol{W}_{3,5} =: \boldsymbol{A}_{3,5}\boldsymbol{Y}_{2,3}+\boldsymbol{W}_{3,5}.
\end{aligned} \tag{65}
$$

For the observation equations in the vectorized model we get with Equation (61):

$$
\begin{aligned}
\boldsymbol{Z}_1 &= \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} = \begin{pmatrix} c_1 & c_2 \end{pmatrix} \begin{pmatrix} Y_{3,1}+Q_1 \\ Y_{3,2}+Q_2 \end{pmatrix} =: \boldsymbol{C}_1\boldsymbol{Y}_{3,1}+\boldsymbol{Q}_1, \\
&\vdots \\
\boldsymbol{Z}_5 &= \begin{pmatrix} Z_9 \\ Z_{10} \end{pmatrix} = \begin{pmatrix} c_9 & c_{10} \end{pmatrix} \begin{pmatrix} Y_{3,9}+Q_9 \\ Y_{3,10}+Q_{10} \end{pmatrix} =: \boldsymbol{C}_5\boldsymbol{Y}_{3,5}+\boldsymbol{Q}_5.
\end{aligned} \tag{66}
$$

This is a model on a new directed acyclic graph shown in Figure 5.

Transition noise vectors on the same scale are modeled as a vector AR(1) model, i.e.

$$
\begin{aligned}
\boldsymbol{W}_{1,2} &= \boldsymbol{B}_{1,1}\boldsymbol{W}_{1,1}+\boldsymbol{V}_{1,1}, \\
\boldsymbol{W}_{2,2} &= \boldsymbol{B}_{2,1}\boldsymbol{W}_{2,1}+\boldsymbol{V}_{2,1}, \\
\boldsymbol{W}_{2,3} &= \boldsymbol{B}_{2,2}\boldsymbol{W}_{2,2}+\boldsymbol{V}_{2,2}, \\
\boldsymbol{W}_{3,2} &= \boldsymbol{B}_{3,1}\boldsymbol{W}_{3,1}+\boldsymbol{V}_{3,1}, \\
&\vdots \\
\boldsymbol{W}_{3,5} &= \boldsymbol{B}_{3,4}\boldsymbol{W}_{3,4}+\boldsymbol{V}_{3,4}.
\end{aligned} \tag{67}
$$

Modeling the observation noise vectors as vector AR(1) model yields

$$
\begin{aligned}
\boldsymbol{Q}_2 &= \boldsymbol{D}_1\boldsymbol{Q}_1+\boldsymbol{U}_1, \\
&\vdots \\
\boldsymbol{Q}_5 &= \boldsymbol{D}_4\boldsymbol{Q}_4+\boldsymbol{U}_4.
\end{aligned} \tag{68}
$$

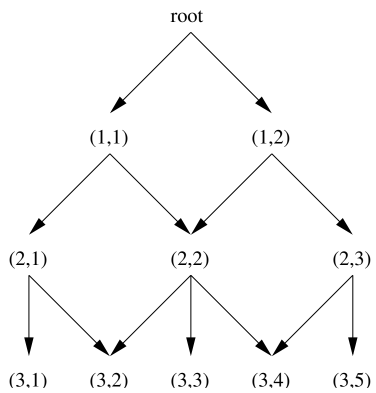

Figure 5: Graph for Vectorized Nodes for Example 4

To put the vectorized model (65)-(68) into linear state space form we specify the initial state vector such that Y_0 *and all* $\boldsymbol{Y}_{j,1}$ *and* $\boldsymbol{W}_{j,1}$ *indexed by nodes* $(j,1)$, $j = 1,\dots,J$, *on the left path from the root to* $(J,1)$ *of the vectorized graph are elements of* $\boldsymbol{X}_1$. *In our example*

$$\boldsymbol{X}_1 := (Y_0, \boldsymbol{Y}_{1,1}, \boldsymbol{Y}_{2,1}, \boldsymbol{Y}_{3,1}, \boldsymbol{W}_{1,1}, \boldsymbol{W}_{2,1}, \boldsymbol{W}_{3,1})'.$$

We will denote the observations in (26) and (27) by $\bar{\boldsymbol{Z}}_k$ *to distinguish them from the observations in the acyclic graph model. The first observation equation can therefore be written as*

$$\bar{\boldsymbol{Z}}_1 = \bar{\boldsymbol{C}}_1 \boldsymbol{X}_1 + \boldsymbol{\eta}_1,$$

with $\boldsymbol{\eta}_1 := \boldsymbol{Q}_1$ *and* $\bar{\boldsymbol{C}}_1 := \begin{pmatrix} \mathbf{0}_{2,5} & \boldsymbol{C}_1 & \mathbf{0}_{2,6} \end{pmatrix}$, *where* $\mathbf{0}_{p,q}$ *denotes a zero* $(p \times q)$ *matrix and* $\boldsymbol{I}_p$ *the identity matrix of dimension* p. *Similarly, let* $\mathbf{0}_p$ *denote the zero matrix of dimension* $p \times p$.

In model (65)-(68) the data vector $\boldsymbol{Z}_2$ *is connected with* $\boldsymbol{Y}_{J,2}$ *by (66). Since* $\boldsymbol{Y}_{J,2} \notin \boldsymbol{X}_1$ *the* $\boldsymbol{Y}_{j,k}$ *and* $\boldsymbol{W}_{j,k}$ *indexed by the nodes on the right path from the root to* $(J,2)$ *have to be successively integrated into state vectors. Since* $\boldsymbol{Y}_{1,2} = Y_0 \mathbf{1}_{n_0} + \boldsymbol{W}_{1,2}$, *where* $\mathbf{1}_p$ *denotes a column vector of* p *ones, the first step is to update from* $\boldsymbol{W}_{1,1}$ *to* $\boldsymbol{W}_{1,2}$ *in* $\boldsymbol{X}_2$ *using (67). Generally, before integrating* $\boldsymbol{Y}_{i,m}$, $i = 1,\dots,J$, $m = 1,\dots,K_j$, *into a state vector we have to integrate* $\boldsymbol{W}_{i,m}$ *into the state vector. All* $\boldsymbol{W}_{j,k}$, $j = 1,\dots,J$, $k = 1,\dots,K_j$, *in the actual state vector that are not needed for the integration of* $\boldsymbol{W}_{i,m}$ *into the next state vector are retained in the next state vector as long as* $k < K_j$, $j = 1,\dots,J$. *The* $\boldsymbol{Y}_{j,k}$ *in the actual state vector are retained in the next state vector as long as not all their children are either in the actual state vector, have been in previous state vectors or will be in the next state vector.* $\boldsymbol{Y}_{J,1}$ *has no child and therefore "all children of* $\boldsymbol{Y}_{J,1}$*" are in* $\boldsymbol{X}_1$, *implying* $\boldsymbol{Y}_{J,1}$ *is not needed in* $\boldsymbol{X}_2$. *The transition matrix*

$$\bar{\boldsymbol{A}}_m = \begin{pmatrix} block(1,1) & \dots & block(1,e_m) \\ \vdots & \vdots & \vdots \\ block(e_{m+1},1) & \dots & block(e_{m+1},e_m) \end{pmatrix}$$

consists of block matrices denoted by $block(r,c)$, $r = 1,\dots,e_{m+1}$, $c = 1,\dots,e_m$, *where* e_m *denotes the number of sub-vectors in* $\boldsymbol{X}_m$, $m \geq 1$. *In our example* $e_1 = 7$. $\boldsymbol{W}_{1,1}$ *is the fifth sub-vector of* $\boldsymbol{X}_1$ *in our example and* $\boldsymbol{W}_{1,2}$ *will be the fourth sub-vector of* $\boldsymbol{X}_2$. *Thus in* $\bar{\boldsymbol{A}}_1$ *we have* $block(4,5)$ *defined to be* $\boldsymbol{B}_{1,1}$. *If the* n_1*th sub-vector of* $\boldsymbol{X}_1$ *will be the* n_2*th sub-vector of* $\boldsymbol{X}_2$, *we define* $block(n_2,n_1)$ *to be the identity matrix of appropriate dimension. In our example we get*

$$\boldsymbol{X}_2 := \begin{pmatrix} Y_0 \\ \boldsymbol{Y}_{1,1} \\ \boldsymbol{Y}_{2,1} \\ \boldsymbol{W}_{1,2} \\ \boldsymbol{W}_{2,1} \\ \boldsymbol{W}_{3,1} \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{0}_{1,2} & \mathbf{0}_{1,2} & \mathbf{0}_{1,2} & \mathbf{0}_{1,2} & \mathbf{0}_{1,2} & \mathbf{0}_{1,2} \\ \mathbf{0}_{2,1} & \boldsymbol{I}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 \\ \mathbf{0}_{2,1} & \mathbf{0}_2 & \boldsymbol{I}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 \\ \mathbf{0}_{2,1} & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \boldsymbol{B}_{1,1} & \mathbf{0}_2 & \mathbf{0}_2 \\ \mathbf{0}_{2,1} & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \boldsymbol{I}_2 & \mathbf{0}_2 \\ \mathbf{0}_{2,1} & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \boldsymbol{I}_2 \end{pmatrix} \boldsymbol{X}_1 + \begin{pmatrix} \mathbf{0}_{5,2} \\ \boldsymbol{I}_2 \\ \mathbf{0}_{4,2} \end{pmatrix} \boldsymbol{V}_{1,1} =: \bar{\boldsymbol{A}}_1 \boldsymbol{X}_1 + \boldsymbol{\Gamma}_1 \boldsymbol{\xi}_1.$$

No observation is connected with state vector $\boldsymbol{X}_2$ *since there is no terminal node element of* $\boldsymbol{X}_2$. *We interprete this as a case of missing observation and define* $\bar{\boldsymbol{C}}_2 := \mathbf{0}_{2,11}$, $\bar{\boldsymbol{Z}}_2 := \bar{\boldsymbol{Z}}_1 = \boldsymbol{Z}_1$ *and* $\boldsymbol{\eta}_2 := \boldsymbol{\eta}_1 = \boldsymbol{Q}_1$. *Like in (46) this implies* $\bar{\boldsymbol{N}}_1 = \boldsymbol{I}_2$ *and* $\boldsymbol{\gamma}_2 = \mathbf{0}_{2,1}$.

Now we can integrate $\boldsymbol{Y}_{1,2}$ *into* $\boldsymbol{X}_3$ *using (65). For this we note that* Y_0 *is the first sub-vector and* $\boldsymbol{W}_{1,2}$ *is the fourth sub-vector in* $\boldsymbol{X}_2$. $\boldsymbol{Y}_{1,2}$ *will be the third sub-vector in* $\boldsymbol{X}_3$. *Thus in* $\bar{\boldsymbol{A}}_2$ *we set* $block(3,1) := \mathbf{1}_{2,1}$ *and* $block(3,4) := \boldsymbol{I}_2$, *where* $\mathbf{1}_{p,q}$ *is a* $(p \times q)$ *matrix of ones. We further update from* $\boldsymbol{W}_{2,1}$ *to* $\boldsymbol{W}_{2,2}$. *Since there is no* $\boldsymbol{W}_{1,3}$ *we can skip* $\boldsymbol{W}_{1,2}$ *in* $\boldsymbol{X}_3$. *The other sub-vectors of* $\boldsymbol{X}_2$ *are retained in* $\boldsymbol{X}_3$.

$$\boldsymbol{X}_3 := \begin{pmatrix} Y_0 \\ \boldsymbol{Y}_{1,1} \\ \boldsymbol{Y}_{1,2} \\ \boldsymbol{Y}_{2,1} \\ \boldsymbol{W}_{2,2} \\ \boldsymbol{W}_{3,1} \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{0}_{1,2} & \mathbf{0}_{1,2} & \mathbf{0}_{1,2} & \mathbf{0}_{1,2} & \mathbf{0}_{1,2} \\ \mathbf{0}_{2,1} & \boldsymbol{I}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 \\ \mathbf{1}_{2,1} & \mathbf{0}_2 & \mathbf{0}_2 & \boldsymbol{I}_2 & \mathbf{0}_2 & \mathbf{0}_2 \\ \mathbf{0}_{2,1} & \mathbf{0}_2 & \boldsymbol{I}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 \\ \mathbf{0}_{2,1} & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \boldsymbol{B}_{2,1} & \mathbf{0}_2 \\ \mathbf{0}_{2,1} & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \boldsymbol{I}_2 \end{pmatrix} \boldsymbol{X}_2 + \begin{pmatrix} \mathbf{0}_{7,2} \\ \boldsymbol{I}_2 \\ \mathbf{0}_2 \end{pmatrix} \boldsymbol{V}_{2,1} =: \bar{\boldsymbol{A}}_2 \boldsymbol{X}_2 + \boldsymbol{\Gamma}_2 \boldsymbol{\xi}_2,$$

Again there is no observation connected with $\boldsymbol{X}_3$ *and thus we set* $\bar{\boldsymbol{C}}_3 := \mathbf{0}_{2,11}$, $\bar{\boldsymbol{Z}}_3 := \bar{\boldsymbol{Z}}_2 = \boldsymbol{Z}_1$ *and* $\boldsymbol{\eta}_3 := \boldsymbol{\eta}_2 = \boldsymbol{Q}_1$, *implying that* $\bar{\boldsymbol{N}}_2 = \boldsymbol{I}_2$ *and* $\boldsymbol{\gamma}_3 = \mathbf{0}_{2,1}$.

Going down the right path from the root to $(J,2)$ *the next step is done by integrating* $\boldsymbol{Y}_{2,2}$ *into* $\boldsymbol{X}_4$. *Using (65) we note that* $\boldsymbol{Y}_{1,1}$ *is the second sub-vector,* $\boldsymbol{Y}_{1,2}$ *is the third sub-vector and* $\boldsymbol{W}_{2,2}$ *is the fifth sub-vector of* $\boldsymbol{X}_3$. $\boldsymbol{Y}_{2,2}$ *will be the third sub-vector in* $\boldsymbol{X}_4$. *Thus in* $\bar{\boldsymbol{A}}_4$ *we define* $block(3,2) := \boldsymbol{A}^l_{2,2}$, $block(3,3) := \boldsymbol{A}^r_{2,2}$ *and* $block(3,5) := \boldsymbol{I}_2$. *We further update from* $\boldsymbol{W}_{3,1}$ *to* $\boldsymbol{W}_{3,2}$. Y_0 *and* $\boldsymbol{Y}_{1,1}$ *are no longer needed in the state vector since all their children are, already have been or will be in the state vectors* $\boldsymbol{X}_1$ *to* $\boldsymbol{X}_4$.

$$\boldsymbol{X}_4 := \begin{pmatrix} \boldsymbol{Y}_{1,2} \\ \boldsymbol{Y}_{2,1} \\ \boldsymbol{Y}_{2,2} \\ \boldsymbol{W}_{2,2} \\ \boldsymbol{W}_{3,2} \end{pmatrix} = \begin{pmatrix} \mathbf{0}_{2,1} & \mathbf{0}_2 & \boldsymbol{I}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 \\ \mathbf{0}_{2,1} & \mathbf{0}_2 & \mathbf{0}_2 & \boldsymbol{I}_2 & \mathbf{0}_2 & \mathbf{0}_2 \\ \mathbf{0}_{2,1} & \boldsymbol{A}^l_{2,2} & \boldsymbol{A}^r_{2,2} & \mathbf{0}_2 & \boldsymbol{I}_2 & \mathbf{0}_2 \\ \mathbf{0}_{2,1} & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \boldsymbol{I}_2 & \mathbf{0}_2 \\ \mathbf{0}_{2,1} & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \boldsymbol{B}_{3,1} \end{pmatrix} \boldsymbol{X}_3 + \begin{pmatrix} \mathbf{0}_{8,2} \\ \boldsymbol{I}_2 \end{pmatrix} \boldsymbol{V}_{3,1} =: \bar{\boldsymbol{A}}_3 \boldsymbol{X}_3 + \boldsymbol{\Gamma}_2 \boldsymbol{\xi}_3.$$

Again this is treated as a case of missing observations, setting $\bar{\boldsymbol{C}}_4 := \mathbf{0}_{2,8}$, $\bar{\boldsymbol{Z}}_4 := \bar{\boldsymbol{Z}}_3 = \boldsymbol{Z}_1$ *and* $\boldsymbol{\eta}_4 := \boldsymbol{\eta}_3 = \boldsymbol{Q}_1$. *This implies* $\bar{\boldsymbol{N}}_3 = \boldsymbol{I}_2$ *and* $\boldsymbol{\gamma}_4 = \mathbf{0}_{2,1}$.

Now we can easily put $\boldsymbol{Y}_{3,2}$ *into* $\boldsymbol{X}_5$*, noting that* $\boldsymbol{Y}_{3,2}$ *has the parents* $\boldsymbol{Y}_{2,1}$ *and* $\boldsymbol{Y}_{2,2}$ *which are subvectors of* $\boldsymbol{X}_4$*. Then all the children of* $\boldsymbol{Y}_{2,1}$ *are, already have been or will be in the state vectors* $\boldsymbol{X}_1$ *to* $\boldsymbol{X}_5$ *and thus we can omit* $\boldsymbol{Y}_{2,1}$ *in* $\boldsymbol{X}_5$*. We also update from* $\boldsymbol{W}_{3,2}$ *to* $\boldsymbol{W}_{3,3}$*. In Equation (66) observation* $\boldsymbol{Z}_2$ *is connected with* $\boldsymbol{Y}_{3,2}$ *and thus connected with* $\boldsymbol{X}_5$*. In particular we have*

$$\boldsymbol{X}_5 := \begin{pmatrix} \boldsymbol{Y}_{1,2} \\ \boldsymbol{Y}_{2,2} \\ \boldsymbol{Y}_{3,2} \\ \boldsymbol{W}_{2,2} \\ \boldsymbol{W}_{3,3} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 \\ \mathbf{0}_2 & \mathbf{0}_2 & \boldsymbol{I}_2 & \mathbf{0}_2 & \mathbf{0}_2 \\ \mathbf{0}_2 & \boldsymbol{A}^l_{3,2} & \boldsymbol{A}^r_{3,2} & \mathbf{0}_2 & \boldsymbol{I}_2 \\ \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \boldsymbol{I}_2 & \mathbf{0}_2 \\ \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \boldsymbol{B}_{3,2} \end{pmatrix} \boldsymbol{X}_3 + \begin{pmatrix} \mathbf{0}_{8,2} \\ \boldsymbol{I}_2 \end{pmatrix} \boldsymbol{V}_{3,2} =: \bar{\boldsymbol{A}}_4 \boldsymbol{X}_4 + \boldsymbol{\Gamma}_4 \boldsymbol{\xi}_4.$$

The corresponding observation equation is given by

$$\bar{\boldsymbol{Z}}_5 := \boldsymbol{Z}_2 = \begin{pmatrix} \mathbf{0}_{2,4} & \boldsymbol{C}_2 & \mathbf{0}_{2,4} \end{pmatrix} \boldsymbol{X}_5 + \boldsymbol{Q}_5 =: \bar{\boldsymbol{C}}_5 \boldsymbol{X}_5 + \boldsymbol{\eta}_5,$$

and we set for the observation noise $\bar{\boldsymbol{N}}_4 := \boldsymbol{D}_1$ *and* $\boldsymbol{\gamma}_5 := \boldsymbol{U}_1$.

In the next state vector we integrate $\boldsymbol{Y}_{3,3}$ *and update from* $\boldsymbol{W}_{2,2}$ *to* $\boldsymbol{W}_{2,3}$. $\boldsymbol{Y}_{3,2}$ *is omitted in* $\boldsymbol{X}_6$ *since it has no child. Therefore we define*

$$\boldsymbol{X}_6 := \begin{pmatrix} \boldsymbol{Y}_{1,2} \\ \boldsymbol{Y}_{2,2} \\ \boldsymbol{Y}_{3,3} \\ \boldsymbol{W}_{2,3} \\ \boldsymbol{W}_{3,3} \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 \\ \mathbf{0}_2 & \boldsymbol{I}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 \\ \mathbf{0}_2 & \boldsymbol{A}_{3,3} & \mathbf{0}_2 & \mathbf{0}_2 & \boldsymbol{I}_2 \\ \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \boldsymbol{B}_{2,2} & \mathbf{0}_2 \\ \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \boldsymbol{I}_2 \end{pmatrix} \boldsymbol{X}_5 + \begin{pmatrix} \mathbf{0}_{6,2} \\ \boldsymbol{I}_2 \\ \mathbf{0}_2 \end{pmatrix} \boldsymbol{V}_{2,2} =: \bar{\boldsymbol{A}}_5 \boldsymbol{X}_5 + \boldsymbol{\Gamma}_5 \boldsymbol{\xi}_5,$$

together with the observation equation

$$\bar{\boldsymbol{Z}}_6 := \boldsymbol{Z}_3 = \begin{pmatrix} \mathbf{0}_{2,4} & \boldsymbol{C}_3 & \mathbf{0}_{2,4} \end{pmatrix} \boldsymbol{X}_6 + \boldsymbol{Q}_3 =: \bar{\boldsymbol{C}}_6 \boldsymbol{X}_6 + \boldsymbol{\eta}_6.$$

For the observation noise we set $\bar{\boldsymbol{N}}_5 := \boldsymbol{D}_2$ *and* $\boldsymbol{\gamma}_6 := \boldsymbol{U}_2$.

Now we have to integrate $\boldsymbol{Y}_{2,3}$ *into the state vector since* $\boldsymbol{Y}_{2,3}$ *is the right parent of* $\boldsymbol{Y}_{3,4}$*. We omit* $\boldsymbol{Y}_{1,2}$ *because all of its children are or already have been or will be in the state vector. We update from* $\boldsymbol{W}_{3,3}$ *to* $\boldsymbol{W}_{3,4}$*. Since there is no* $\boldsymbol{W}_{2,4}$ *we do not need* $\boldsymbol{W}_{2,3}$ *in the state vector any more. This gives*

$$\boldsymbol{X}_7 := \begin{pmatrix} \boldsymbol{Y}_{2,2} \\ \boldsymbol{Y}_{2,3} \\ \boldsymbol{W}_{3,4} \end{pmatrix} = \begin{pmatrix} \mathbf{0}_2 & \boldsymbol{I}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 \\ \boldsymbol{A}_{2,3} & \mathbf{0}_2 & \mathbf{0}_2 & \boldsymbol{I}_2 & \mathbf{0}_2 \\ \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \mathbf{0}_2 & \boldsymbol{B}_{3,3} \end{pmatrix} \boldsymbol{X}_6 + \begin{pmatrix} \mathbf{0}_{4,2} \\ \boldsymbol{I}_2 \end{pmatrix} \boldsymbol{V}_{3,3} =: \bar{\boldsymbol{A}}_6 \boldsymbol{X}_6 + \boldsymbol{\Gamma}_6 \boldsymbol{\xi}_6.$$

Again we have a case of missing observation and set $\bar{\boldsymbol{C}}_7 := \boldsymbol{0}_{2,6}$, $\bar{\boldsymbol{Z}}_7 := \bar{\boldsymbol{Z}}_6 = \boldsymbol{Z}_3$ *and* $\boldsymbol{\eta}_7 := \boldsymbol{\eta}_6 = \boldsymbol{Q}_3$. *It follows that* $\bar{\boldsymbol{N}}_6 = \boldsymbol{I}_2$ *and* $\boldsymbol{\gamma}_7 = \boldsymbol{0}_{2,1}$.

Now we put $\boldsymbol{Y}_{3,4}$ *into the next state vector* $\boldsymbol{X}_8$. $\boldsymbol{Y}_{2,2}$ *is not needed any longer. Therefore we have*

$$\boldsymbol{X}_8 := \begin{pmatrix} \boldsymbol{Y}_{2,3} \\ \boldsymbol{Y}_{3,4} \\ \boldsymbol{W}_{3,5} \end{pmatrix} = \begin{pmatrix} \boldsymbol{0}_2 & \boldsymbol{I}_2 & \boldsymbol{0}_2 \\ \boldsymbol{A}^l_{3,4} & \boldsymbol{A}^r_{3,4} & \boldsymbol{I}_2 \\ \boldsymbol{0}_2 & \boldsymbol{0}_2 & \boldsymbol{B}_{3,4} \end{pmatrix} \boldsymbol{X}_7 + \begin{pmatrix} \boldsymbol{0}_{4,2} \\ \boldsymbol{I}_2 \end{pmatrix} \boldsymbol{V}_{3,4} =: \bar{\boldsymbol{A}}_7 \boldsymbol{X}_7 + \boldsymbol{\Gamma}_7 \boldsymbol{\xi}_7.$$
$$\bar{\boldsymbol{Z}}_8 := \boldsymbol{Z}_4 = \begin{pmatrix} \boldsymbol{0}_2 & \boldsymbol{C}_4 & \boldsymbol{0}_2 \end{pmatrix} \boldsymbol{X}_8 + \boldsymbol{\eta}_8,$$

where $\boldsymbol{\eta}_8 := \boldsymbol{Q}_4$, $\bar{\boldsymbol{N}}_7 := \boldsymbol{D}_3$ *and* $\boldsymbol{\gamma}_8 := \boldsymbol{U}_3$.

The last state vector $\boldsymbol{X}_9$ *is simply given by*

$$\boldsymbol{X}_9 := \boldsymbol{Y}_{3,5} = \begin{pmatrix} \boldsymbol{A}_{3,5} & \boldsymbol{0}_2 & \boldsymbol{I}_2 \end{pmatrix} \boldsymbol{X}_8 =: \bar{\boldsymbol{A}}_8 \boldsymbol{X}_8.$$
$$\bar{\boldsymbol{Z}}_9 := \boldsymbol{Z}_5 = \bar{\boldsymbol{C}}_9 \boldsymbol{X}_9 + \boldsymbol{\eta}_9,$$

where $\bar{\boldsymbol{C}}_9 := \boldsymbol{C}_5$, $\boldsymbol{\eta}_9 := \boldsymbol{Q}_5$, $\bar{\boldsymbol{N}}_8 := \boldsymbol{D}_3$ *and* $\boldsymbol{\gamma}_9 := \boldsymbol{U}_4$.

Let $N^{\boldsymbol{X}}$ denote the number of state vectors $\boldsymbol{X}_k$. For $m = 1, \ldots, N^{\boldsymbol{X}}$ we have $\boldsymbol{\xi}_m = \boldsymbol{0}$ or there is a $j \in \{1, \ldots, J\}$ and a $k \in \{1, \ldots, K_j\}$ such that $\boldsymbol{\xi}_m = V_{j,k}$. From $\boldsymbol{\xi}_m = \boldsymbol{V}_{j,k}$ it follows that $\boldsymbol{W}_{j,k} \in \boldsymbol{X}_m$. Since $\boldsymbol{V}_{j,k} \perp \boldsymbol{W}_{j,k}$ we get $\boldsymbol{\xi}_m \perp \boldsymbol{X}_m$. Furthermore, since $\{\boldsymbol{V}_{j,k}, j = 1, \ldots, J,\ k = 1, \ldots, K_j\}$ is a family of uncorrelated random vectors $\{\boldsymbol{\xi}_m, m = 1 \ldots, N^{\boldsymbol{X}}\}$ is also a family of uncorrelated random vectors. Similarly, for $m = 1 \ldots, N^{\boldsymbol{X}}$ we have set $\boldsymbol{\eta}_m = \boldsymbol{Q}_k$ with $k \in \{1, \ldots, K_J\}$ and $k \leq m$. $\bar{\boldsymbol{N}}_{m-1}$ was then defined either by $\bar{\boldsymbol{N}}_{m-1} := \boldsymbol{I}$ or $\bar{\boldsymbol{N}}_{m-1} := \boldsymbol{D}_{k-1}$ and $\boldsymbol{\gamma}_m$ was defined by $\boldsymbol{\gamma}_m := \boldsymbol{0}$ or $\boldsymbol{\gamma}_m := \boldsymbol{U}_{k-1}$. Since $\{\boldsymbol{U}_k, k = 1, \ldots, K_J\}$ is a family of uncorrelated random vectors $\{\boldsymbol{\gamma}_m, m = 1 \ldots, N^{\boldsymbol{X}}\}$ is also a family of uncorrelated random vectors. With $\boldsymbol{U}_k \perp \boldsymbol{Q}_k$, $k = 1, \ldots, K_J$ it also follows that $\boldsymbol{\gamma}_m \perp \boldsymbol{\eta}_{m-1}$. Therefore we have a state space representation such that model equations (26), (28) and (29) are satisfied. Though the state vectors have relatively big dimension the system matrices $\bar{\boldsymbol{A}}_k, \boldsymbol{\Gamma}_k$ and have a block structure such that calculation can be done efficiently. For the initialization of the Kalman filter we have to derive $var(\boldsymbol{X}_1)$. This will be done element wise. Recall that $W_{j,k} \perp W_{i,l}$, $(j \neq i)$. At first we note $cov(Y_0, W_{j,k}) = 0$ and $cov(Y_{j,k}, W_{i,l}) = 0$ for $j = 1, \ldots, J-1$, $i = j+1, \ldots, J$, $k, l = 1, 2$. Using (59) and (60) yields for $j = 1, 2, 3$:

$$cov(Y_{j,1}, W_{j,1}) = var(W_{j,1}), \qquad cov(Y_{j,2}, W_{j,1}) = cov(W_{j,2}, W_{j,1}), \tag{69}$$
$$cov(Y_{j,1}, W_{j,2}) = cov(W_{j,1}, W_{j,2}), \qquad cov(Y_{j,2}, W_{j,2}) = var(W_{j,2}). \tag{70}$$

For $j = 1, 2$ we get

$$\begin{aligned}
cov(Y_{j+1,1}, W_{j,1}) &= cov[(\alpha_{j+1,1} Y_{j,1} + \beta_{j+1,1} Y_{j,2} + W_{j+1,1}), W_{j,1}] \\
&= \alpha_{j+1,1} var(W_{j,1}) + \beta_{j+1,1} cov(W_{j,2}, W_{j,1}), \\
cov(Y_{j+1,1}, W_{j,2}) &= \alpha_{j+1,1} cov(W_{j,1}, W_{j,2}) + \beta_{j+1,1} var(W_{j,2}), \\
cov(Y_{j+1,2}, W_{j,1}) &= cov[(\alpha_{j+1,2} Y_{j,1} + \beta_{j+1,1} Y_{j,2} + W_{j+1,2}), W_{j,2}] \\
&= \alpha_{j+1,2} var(W_{j,1}) + \beta_{j+1,2} cov(W_{j,2}, W_{j,1}), \\
cov(Y_{j+1,2}, W_{j,2}) &= \alpha_{j+1,2} cov(W_{j,1}, W_{j,2}) + \beta_{j+1,2} var(W_{j,2}).
\end{aligned}$$

For $j = 1$ we get recursively

$$\begin{aligned}
cov(Y_{3,1}, W_{1,1}) &= cov[(\alpha_{3,1} Y_{2,1} + \beta_{3,1} Y_{2,2} + W_{3,1}), W_{1,1}] \\
&= \alpha_{3,1} cov(Y_{2,1}, W_{1,1}) + \beta_{3,1} cov(Y_{2,2}, W_{1,1}), \\
cov(Y_{3,1}, W_{1,2}) &= \alpha_{3,1} cov(Y_{2,1}, W_{1,2}) + \beta_{3,1} cov(Y_{2,2}, W_{1,2}), \\
cov(Y_{3,2}, W_{1,1}) &= cov[(\alpha_{3,2} Y_{2,1} + \beta_{3,2} Y_{2,2} + W_{3,2}), W_{1,1}] \\
&= \alpha_{3,2} cov(Y_{2,1}, W_{1,1}) + \beta_{3,2} cov(Y_{1,2}, W_{1,1}), \\
cov(Y_{j+2,1}, W_{j,2}) &= \alpha_{j+2,2} cov(Y_{j+1,1}, W_{j,2}) + \beta_{j+2,2} cov(Y_{j+1,2}, W_{j,2}).
\end{aligned}$$

Finally the quantities $var(Y_{j,k})$ and $cov(Y_{j,1}, Y_{j,2})$ can be computed recursively using (59) and (60) starting at the root or at the first scale, respectively. In particular we obtain at the first scale for $k = 1, 2$

$$\begin{aligned}
var(Y_{1,k}) &= var(Y_0) + var(W_{1,k}), \\
cov(Y_{1,1}, Y_{1,2}) &= cov[(Y_0 + W_{1,1}), (Y_0 + W_{1,2})] = var(Y_0) + cov(W_{1,1}, W_{1,2}).
\end{aligned}$$

For $j = 2, 3$ and $k = 1, 2$ we get

$$var(Y_{j,k}) = \alpha_{j,k}^2 var(Y_{j-1,1}) + \beta_{j,k}^2 var(Y_{j-1,2}) + 2\alpha_{j,k}\beta_{j,k} cov(Y_{j-1,1}, Y_{j-1,2}) + var(W_{j,k}),$$

$$\begin{aligned}
cov(Y_{j,1}, Y_{j,2}) &= cov([\alpha_{j,1} Y_{j-1,1} + \beta_{j,1} Y_{j-1,1} + W_{j,1}], [\alpha_{j,2} Y_{j-1,2} + \beta_{j,2} Y_{j-1,2} + W_{j,2}]) \\
&= \alpha_{j,1}\alpha_{j,2} var(Y_{j-1,1}) + \beta_{j,1}\beta_{j,2} var(Y_{j-1,2}) + (\alpha_{j,1}\beta_{j,2} + \alpha_{j,2}\beta_{j,1}) cov(Y_{j-1,1}, Y_{j-1,2}) \\
&\quad + cov(W_{j,1}, W_{j,2}).
\end{aligned}$$

7 Discussion

We reviewed and presented three classes of models on acyclic directed graphs. The first class, introduced by Chou (1991) and Huang et al. (2002) applies nested partitioning that makes the assumption of mass balance necessary. The restrictions implied by mass balance together with the white noise assumption for transition and observation noise imply an artificial block structure of the correlation matrix of the data

$\boldsymbol{Z}$ and thus it is not possible to obtain a stationary autocorrelation function for Z. Huang et al. (2002) suggest as a possible solution of this problem to compute the predictions as an average over a number of mass balanced, tree-structured models with different tree branches that represent children shifted to have different parents. As they point out, the prediction variances and covariances will be more complicated and the computational complexity will increase with the number of trees used. The models imply that all average values $Y_{j,k}$ have the same mean. The advantage of their models is that computation can be done efficiently by a change-of-resolution Kalman filter in such a way that computations can be easily performed in parallel problems of smaller dimensions. Furthermore, the number of parameters to be estimated is smaller then in the models defined in Sections 4 and 6.

We generalized their models by modeling the transition noise on a same scale in linear state space form. Mass-balance implies restrictions to the covariance matrix $E(\boldsymbol{W}_{j,k+1}\boldsymbol{W}'_{j,k})$ that are, at least so far, not easy to deal with. They also imply an artificial structure of the correlation matrix of $\boldsymbol{Z}$ such that a stationary autocorrelation of $\boldsymbol{Z}$ cannot be obtained. Here there are not only variance parameters but also covariance parameters to estimate. It is questionable if this way of modeling is an improvement to the models by Chou (1991) and Huang et al. (2002).

The third class of models requires no mass balance. The observation interval is divided into sub-intervals by overlapping interval arrangement. The stochastic process is indexed by the nodes of a directed acyclic graph which allows the nodes to have up to two parents. Thus weighting parameters are additionally needed, which also have to be estimated. These models on acyclic directed graphs have a linear state space representation with white system and coloured observation noise. The linear least squares prediction of interval averages was done by a Kalman filter and Kalman fixed point smoother. The advantage of this model is that now dynamic structures in transition and in observation noise can be modeled, that may be a matter of interest in themselves. Thus the zero mean assumption for the $Y_{j,k}$ can be relaxed.

Further topics for future research are incorporation of explanatory variables and terms to capture seasonality and trend in the data. The score vector needed for parameter estimation depends on the value of the initial state vector $\boldsymbol{X}_1$ which will often be unknown in practice. The approach in Koopman and Durbin (2001) using so called diffuse initial state vector for linear state space models with white system and observation noise should also work for state space models with coloured observation noise. A further important topic is to find concepts to reduce the number of variance, covariance and weighting parameters. For some applications one could model the observation noise process for itself and then integrate this model into a model on an acyclic directed graph. Then the variance and covariance parameters of the observation noise are not required to be estimated within the model on acyclic directed graphs. Finally, we aim on applying these models to high frequency financial data.

A Examples for Correlation Matrix of Z

Here we give examples of possible structures of $cor(\boldsymbol{Z})$ implied by tree structured models described in Section 4.

Example 5. *Figure 6 shows the correlation of the observed data vector $\boldsymbol{Z}$ arising from the following tree-structured model: The finest scale is $J = 3$. The root has $n_0 = 4$ children. Each node on $j = 1$ has $n_{1,k} = 6$ children, $k = 1, \ldots, 4$ and each node on the second scale has also $n_{2,k} = 6$ children, $k = 1, \ldots, 24$. The variance of the univariate transition noise is $\sigma_0^2 = \sigma_{j,k}^2 = 1$, $E(\boldsymbol{W}_{j,k+1}\boldsymbol{W}_{j,k}) = \mathbf{0}$, $j = 1, 2, 3$, $k = 1, \ldots, N_j$. The variance of the observation noise is given by $\tau_m^2 = 1$ with $E(Q_l Q_m) = 0$, $m, l = 1, \ldots, 144$, $m \neq l$.*

Example 6. *Figure 7 shows the correlation of the observed data vector $\boldsymbol{Z}$ arising from the tree-structured model with finest scale $J = 3$. The root has $n_0 = 4$ children. Each node on the the first scale has $n_{1,k} = 6$ children, $k = 1, \ldots, 4$ and each node on the second scale has also $n_{2,k} = 6$ children, $k = 1, \ldots, 24$. For $j = 1, 2, 3$ and $k = 1, \ldots, N_j$ we have chosen $var(W_{j,k}) := \sigma_{j,k}^2 = 1$ and*

$$E(\boldsymbol{W}_{j,k+1}\boldsymbol{W}'_{j,k}) = \begin{pmatrix} -0.39 & \frac{1}{10} & -\frac{1}{8} & \frac{1}{6} & -\frac{1}{4} & \frac{1}{2} \\ 0.19 & -\frac{1}{12} & \frac{1}{10} & -\frac{1}{8} & \frac{1}{6} & -\frac{1}{4} \\ -0.13 & \frac{1}{14} & -\frac{1}{12} & \frac{1}{10} & -\frac{1}{8} & \frac{1}{6} \\ 0.09 & -\frac{1}{16} & \frac{1}{14} & -\frac{1}{12} & \frac{1}{10} & -\frac{1}{8} \\ -0.08 & \frac{1}{18} & -\frac{1}{16} & \frac{1}{14} & -\frac{1}{12} & \frac{1}{10} \\ 0.31 & -0.08 & 0.10 & -0.13 & 0.19 & -0.39 \end{pmatrix}.$$

Furthermore, $var(Q_m) := \tau_m^2 = 1$ and $Q_{m+1} = 0.5Q_m + U_m$, $m = 1, \ldots, 144$.

B Kalman Fixed Point Smoothing for Linear State Space Models with White System and Coloured Observation Noise

Considering the model defined by (26), (28) and (29) we can modify the algorithm and the derivation given in Durbin and Koopman (2001) for linear state space models with white system and white observation noise, using Lemma 2.13 in Durbin and Koopman (2001) where it is assumed that $\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}$ are jointly distributed random vectors of arbitrary order with $E(\boldsymbol{\theta}) = \mathbf{0}$ and $E(\boldsymbol{Z}\boldsymbol{\theta}') = \mathbf{0}$. Defining

$$\tilde{\boldsymbol{X}}_{\boldsymbol{Z\theta}} := \boldsymbol{X} - L(\boldsymbol{X}|\boldsymbol{Z}, \boldsymbol{\theta}),$$
$$\tilde{\boldsymbol{X}}_{\boldsymbol{Z}} := \boldsymbol{X} - L(\boldsymbol{X}|\boldsymbol{Z})$$

then their Lemma 2.13 states that the following equations hold:

$$\begin{aligned} L(\boldsymbol{X}|\boldsymbol{Z}, \boldsymbol{\theta}) &= L(\boldsymbol{X}|\boldsymbol{Z}) + E(\boldsymbol{X}, \boldsymbol{\theta}) var(\boldsymbol{\theta})^{-1}\boldsymbol{\theta}, \\ var(\tilde{\boldsymbol{X}}_{\boldsymbol{Z\theta}}) &= var(\tilde{\boldsymbol{X}}_{\boldsymbol{Z}}) - E(\boldsymbol{X}, \boldsymbol{\theta}) var(\boldsymbol{\theta})^{-1} E(\boldsymbol{\theta}, \boldsymbol{X}). \end{aligned} \tag{71}$$

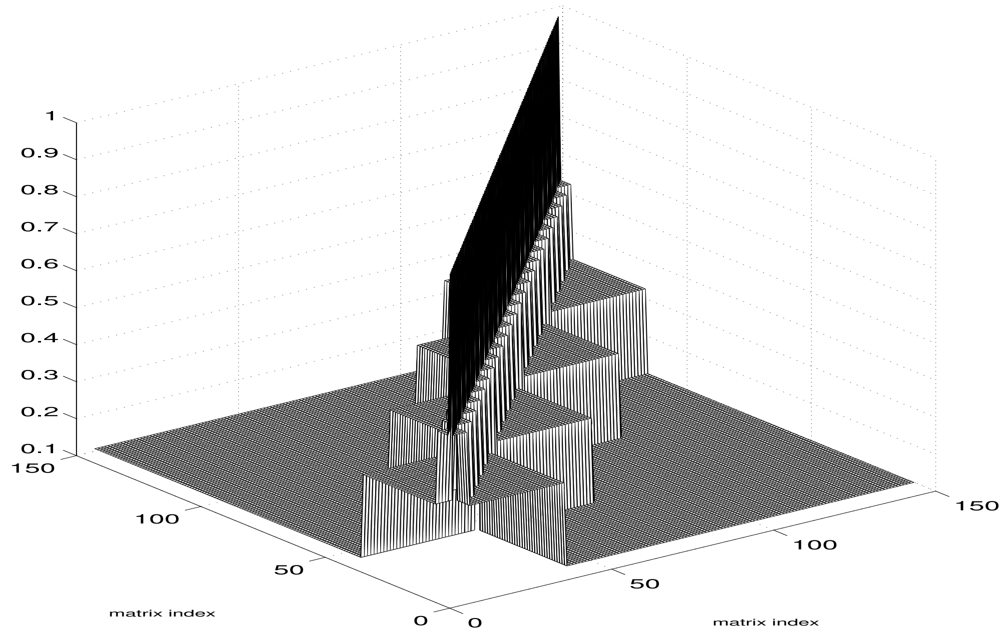

Figure 6: Cor($\boldsymbol{Z}$) of Example 5

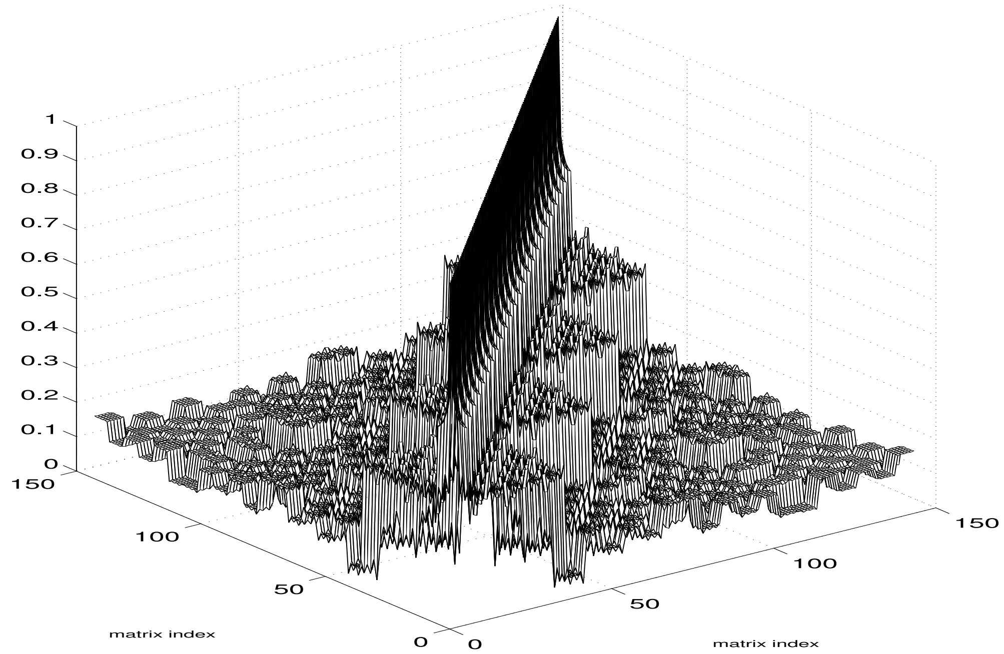

Figure 7: Cor($\boldsymbol{Z}$) of Example 6

For $N \in \mathbb{N}$ fixed and with $\boldsymbol{Z}^N := (\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}^N)'$ we can use Equation (71) together with (30) and (35) to get for $k = 1, \ldots, N$

$$(72) \qquad \hat{\boldsymbol{X}}_{k|N} := L(\boldsymbol{X}_k|\boldsymbol{Z}^N) = L(\boldsymbol{X}_k|\boldsymbol{Z}^k, \boldsymbol{\theta}_{k+1}, \ldots, \boldsymbol{\theta}_N) = \hat{\boldsymbol{X}}_{k|k} + \sum_{l=k+1}^{N} E(\boldsymbol{X}_k \boldsymbol{\theta}_l') \boldsymbol{\Delta}_l^{-1} \boldsymbol{\theta}_l.$$

Note that $E(\boldsymbol{\theta}_k) = \mathbf{0}$ for $k = 2, \ldots, N$. Further we compute using (34)

(73)
$$\begin{aligned} E(\boldsymbol{X}_k \boldsymbol{\theta}_l') &= E[\boldsymbol{X}_k (\boldsymbol{H}_{l-1} \tilde{\boldsymbol{X}}_{l-1|l-1} + \bar{\boldsymbol{C}}_l \boldsymbol{\Gamma}_{l-1} \boldsymbol{\xi}_{l-1} + \boldsymbol{\gamma}_l)'] \\ &= E(\boldsymbol{X}_k \tilde{\boldsymbol{X}}_{l-1|l-1}') \boldsymbol{H}_{l-1}'. \end{aligned}$$

For $l = k+1, \ldots, N$ $E(\boldsymbol{X}_k \tilde{\boldsymbol{X}}_{l-1|l-1}')$ can be obtained recursively, starting with $l = k+1$:

(74)
$$\begin{aligned} E[\boldsymbol{X}_k (\boldsymbol{X}_k - \hat{\boldsymbol{X}}_{k|k})'] &= E(\boldsymbol{X}_k \boldsymbol{X}_k') - E[\boldsymbol{X}_k (E(\boldsymbol{X}_k \boldsymbol{Z}^{k'}) \boldsymbol{\Sigma}_{\boldsymbol{Z}^k}^{-1} \boldsymbol{Z}^k)'] \\ &= var(\boldsymbol{X}_k) - E(\boldsymbol{X}_k \boldsymbol{Z}^{k'}) \boldsymbol{\Sigma}_{\boldsymbol{Z}^k}^{-1} E(\boldsymbol{Z}^k \boldsymbol{X}_k') = \tilde{\boldsymbol{\Sigma}}_{k|k}. \end{aligned}$$

Using Equation (74) together with (26), (39) and (34) we get

$$\begin{aligned} E[\boldsymbol{X}_k (\tilde{\boldsymbol{X}}_{k+1|k+1})'] &= E[\boldsymbol{X}_k (\bar{\boldsymbol{A}}_k \boldsymbol{X}_k + \boldsymbol{\Gamma}_k \boldsymbol{\xi}_k - \bar{\boldsymbol{A}}_k \hat{\boldsymbol{X}}_{k|k} - \boldsymbol{G}_{k+1} \boldsymbol{\theta}_{k+1})'] \\ &= E[\boldsymbol{X}_k (\bar{\boldsymbol{A}}_k \tilde{\boldsymbol{X}}_{k|k} - \boldsymbol{G}_{k+1} [\boldsymbol{H}_k \tilde{\boldsymbol{X}}_{k|k} + \bar{\boldsymbol{C}}_{k+1} \boldsymbol{\Gamma}_k \boldsymbol{\xi}_k + \boldsymbol{\gamma}_{k+1}])'] \\ &= E(\boldsymbol{X}_k \tilde{\boldsymbol{X}}_{k|k}') \bar{\boldsymbol{A}}_k' - E(\boldsymbol{X}_k \tilde{\boldsymbol{X}}_{k|k}') \boldsymbol{H}_k' \boldsymbol{G}_{k+1}' \\ &= E(\boldsymbol{X}_k \tilde{\boldsymbol{X}}_{k|k}') (\bar{\boldsymbol{A}}_k - \boldsymbol{G}_{k+1} \boldsymbol{H}_k)', \end{aligned}$$

since $E(\boldsymbol{X}_k \boldsymbol{\xi}_k') = \mathbf{0}$ and $E(\boldsymbol{X}_k \boldsymbol{\gamma}_{k+1}') = \mathbf{0}$. With (38) we can write

(75)
$$\begin{aligned} E[\boldsymbol{X}_k (\tilde{\boldsymbol{X}}_{k+1|k+1})'] &= E(\boldsymbol{X}_k \tilde{\boldsymbol{X}}_{k|k}') \boldsymbol{F}_{k+1}' = \tilde{\boldsymbol{\Sigma}}_{k|k} \boldsymbol{F}_{k+1}', \\ E[\boldsymbol{X}_k (\tilde{\boldsymbol{X}}_{k+2|k+2})'] &= \tilde{\boldsymbol{\Sigma}}_{k|k} \boldsymbol{F}_{k+1}' \boldsymbol{F}_{k+2}', \\ &\vdots \\ E[\boldsymbol{X}_k (\tilde{\boldsymbol{X}}_{N-1|N-1})'] &= \tilde{\boldsymbol{\Sigma}}_{k|k} \boldsymbol{F}_{k+1}' \ldots \boldsymbol{F}_{N-1}'. \end{aligned}$$

For the computation of the smoothed state space vector $\hat{\boldsymbol{X}}_{k|N}$ we will now substitute (74) and (75) into (72), using (42):

(76)
$$\hat{\boldsymbol{X}}_{k|N} = \hat{\boldsymbol{X}}_{k|k} + \tilde{\boldsymbol{\Sigma}}_{k|k} \boldsymbol{R}_k.$$

For computation of the mean squared error matrices $\tilde{\boldsymbol{\Sigma}}_{k|N}$ we proceed in a similar way. The starting point is again Equation (72). Since the innovations are serially orthogonal we get using (71)

(77)
$$\tilde{\boldsymbol{\Sigma}}_{k|N} = \tilde{\boldsymbol{\Sigma}}_{k|k} - \sum_{l=k+1}^{N} E(\boldsymbol{X}_k \boldsymbol{\theta}_l') \boldsymbol{\Delta}_l^{-1} E(\boldsymbol{\theta}_l \boldsymbol{X}_k').$$

Using (74), (75) and (41) Equation (77) can be written as

(78)
$$\tilde{\boldsymbol{\Sigma}}_{k|N} = \tilde{\boldsymbol{\Sigma}}_{k|k} - \tilde{\boldsymbol{\Sigma}}_{k|k} \boldsymbol{P}_k \tilde{\boldsymbol{\Sigma}}_{k|k}.$$

C Disturbance Smoother for Linear State Space Models with Coloured Observation Noise

The so called disturbance smoother (Durbin and Koopman (2001)) computes $\hat{\boldsymbol{\gamma}}_{k|N} := L(\boldsymbol{\gamma}_k | \boldsymbol{Z}^N)$ and $\hat{\boldsymbol{\xi}}_{k|N} := L(\boldsymbol{\xi}_k | \boldsymbol{Z}^N)$. We follow their approach for linear state space models with white system and coloured observation noise and modify it where needed. At first we need a recursive representation for the state estimation errors. Using Equations (39) and (34) for the innovations gives:

$$
\begin{aligned}
\tilde{\boldsymbol{X}}_{k+1|k+1} &:= \boldsymbol{X}_{k+1} - \hat{\boldsymbol{X}}_{k+1|k+1} \\
&= \bar{\boldsymbol{A}}_k \boldsymbol{X}_k + \boldsymbol{\Gamma}_k \boldsymbol{\xi}_k - \bar{\boldsymbol{A}}_k \hat{\boldsymbol{X}}_{k|k} - \boldsymbol{G}_{k+1} \boldsymbol{\theta}_{k+1} \\
&= \bar{\boldsymbol{A}}_k \tilde{\boldsymbol{X}}_{k|k} + \boldsymbol{\Gamma}_k \boldsymbol{\xi}_k - \boldsymbol{G}_{k+1} \boldsymbol{H}_k \tilde{\boldsymbol{X}}_{k|k} - \boldsymbol{G}_{k+1} \bar{\boldsymbol{C}}_{k+1} \boldsymbol{\Gamma}_k \boldsymbol{\xi}_k - \boldsymbol{G}_{k+1} \boldsymbol{\gamma}_{k+1} \\
&= \boldsymbol{F}_{k+1} \tilde{\boldsymbol{X}}_{k|k} + (\boldsymbol{\Gamma}_k - \boldsymbol{G}_{k+1} \bar{\boldsymbol{C}}_{k+1}) \boldsymbol{\xi}_k - \boldsymbol{G}_{k+1} \boldsymbol{\gamma}_{k+1},
\end{aligned}
\tag{79}
$$

where $\boldsymbol{F}_k$ was defined in (38). With (71) we have

$$
\hat{\boldsymbol{\gamma}}_{k|N} := L(\boldsymbol{\gamma}_k | \boldsymbol{Z}^N) = L(\boldsymbol{\gamma}_k | \boldsymbol{Z}^{k-1}, \boldsymbol{\theta}_k, \ldots, \boldsymbol{\theta}_N) = L(\boldsymbol{\gamma}_k | \boldsymbol{Z}^{k-1}) + \sum_{l=k}^{N} E(\boldsymbol{\gamma}_k \boldsymbol{\theta}_l') \boldsymbol{\Delta}_l^{-1} \boldsymbol{\theta}_l.
$$

Since $\boldsymbol{\gamma}_k \perp \boldsymbol{Z}^{k-1}$ and $E(\boldsymbol{\gamma}_k) = \mathbf{0}$ we have $L(\boldsymbol{\gamma}_k | \boldsymbol{Z}^{k-1}) = \mathbf{0}$. Thus

$$
\hat{\boldsymbol{\gamma}}_{k|N} = \sum_{l=k}^{N} E(\boldsymbol{\gamma}_k \boldsymbol{\theta}_l') \boldsymbol{\Delta}_l^{-1} \boldsymbol{\theta}_l. \tag{80}
$$

Using Equation (34) we yield

$$
E(\boldsymbol{\gamma}_k \boldsymbol{\theta}_k') = E[\boldsymbol{\gamma}_k (\boldsymbol{H}_{k-1} \tilde{\boldsymbol{X}}_{k-1|k-1} + \bar{\boldsymbol{C}}_k \boldsymbol{\Gamma}_{k-1} \boldsymbol{\xi}_{k-1} + \boldsymbol{\gamma}_k)'].
$$

Since $\tilde{\boldsymbol{X}}_{k-1|k-1}$ is a linear function of $\boldsymbol{Z}^{k-1}$ and $\boldsymbol{X}_{k-1}$ it follows that $E(\boldsymbol{\gamma}_k \tilde{\boldsymbol{X}}'_{k-1|k-1}) = \mathbf{0}$. With $E(\boldsymbol{\gamma}_k \boldsymbol{\xi}_k') = \mathbf{0}$ we get

$$
E(\boldsymbol{\gamma}_k \boldsymbol{\theta}_k') = var(\boldsymbol{\gamma}_k). \tag{81}
$$

Noting that $E(\boldsymbol{\gamma}_k \boldsymbol{\gamma}_l') = \mathbf{0}$ holds we get for $l = k+1, \ldots, N$:

$$
E(\boldsymbol{\gamma}_k \boldsymbol{\theta}_l') = E[\boldsymbol{\gamma}_k (\boldsymbol{H}_{l-1} \tilde{\boldsymbol{X}}_{l-1|l-1} + \bar{\boldsymbol{C}}_l \boldsymbol{\Gamma}_{l-1} \boldsymbol{\xi}_{l-1} + \boldsymbol{\gamma}_l)'] = E(\boldsymbol{\gamma}_k \tilde{\boldsymbol{X}}'_{l-1|l-1}) \boldsymbol{H}'_{l-1}. \tag{82}
$$

Here we have used Equation (33). The recursion for the state estimation errors (79) yields

$$
\tilde{\boldsymbol{X}}_{k|k} = \boldsymbol{F}_k \tilde{\boldsymbol{X}}_{k-1|k-1} + (\boldsymbol{\Gamma}_{k-1} - \boldsymbol{G}_k \bar{\boldsymbol{C}}_k) \boldsymbol{\xi}_{k-1} - \boldsymbol{G}_k \boldsymbol{\gamma}_k.
$$

Since $\tilde{\boldsymbol{X}}_{k-1|k-1} \perp \boldsymbol{\gamma}_k$ and $\boldsymbol{\xi}_{k-1} \perp \boldsymbol{\gamma}_k$ we get

$$
E(\boldsymbol{\gamma}_k \boldsymbol{\theta}'_{k+1}) = E(\boldsymbol{\gamma}_k \tilde{\boldsymbol{X}}'_{k|k}) \boldsymbol{H}'_k = E[\boldsymbol{\gamma}_k (-\boldsymbol{G}_k \boldsymbol{\gamma}_k)'] \boldsymbol{H}'_k = -var(\boldsymbol{\gamma}_k) \boldsymbol{G}'_k \boldsymbol{H}'_k. \tag{83}
$$

Further lags are now computed recursively:

$$
\begin{aligned}
E(\boldsymbol{\gamma}_k \boldsymbol{\theta}'_{k+2}) &= E(\boldsymbol{\gamma}_k \tilde{\boldsymbol{X}}'_{k+1|k+1}) \boldsymbol{H}'_{k+1} = E(\boldsymbol{\gamma}_k \tilde{\boldsymbol{X}}'_{k|k}) \boldsymbol{F}'_{k+1} \boldsymbol{H}'_{k+1} \\
&= -var(\boldsymbol{\gamma}_k) \boldsymbol{G}'_k \boldsymbol{F}'_{k+1} \boldsymbol{H}'_{k+1}, \\
E(\boldsymbol{\gamma}_k \boldsymbol{\theta}'_{k+3}) &= E(\boldsymbol{\gamma}_k \tilde{\boldsymbol{X}}'_{k+2|k+2}) \boldsymbol{H}'_{k+2} = E(\boldsymbol{\gamma}_k \tilde{\boldsymbol{X}}'_{k+1|k+1}) \boldsymbol{F}'_{k+2} \boldsymbol{H}'_{k+2} \\
&= -var(\boldsymbol{\gamma}_k) \boldsymbol{G}'_k \boldsymbol{F}'_{k+1} \boldsymbol{F}'_{k+2} \boldsymbol{H}'_{k+2}, \\
&\vdots \\
E(\boldsymbol{\gamma}_k \boldsymbol{\theta}'_N) &= E(\boldsymbol{\gamma}_k \tilde{\boldsymbol{X}}'_{N-1|N-1}) \boldsymbol{H}'_{N-1} = E(\boldsymbol{\gamma}_k \tilde{\boldsymbol{X}}'_{N-2|N-2}) \boldsymbol{F}'_{N-1} \boldsymbol{H}'_{N-1} \\
&= -var(\boldsymbol{\gamma}_k) \boldsymbol{G}'_k \boldsymbol{F}'_{k+1} \ldots \boldsymbol{F}'_{N-1} \boldsymbol{H}'_{N-1}.
\end{aligned}
$$

Substitution into (80) yields

$$\hat{\gamma}_{k|N} = var(\boldsymbol{\gamma}_k)\boldsymbol{\Delta}_k^{-1}\boldsymbol{\theta}_k - var(\boldsymbol{\gamma}_k)\boldsymbol{G}'_k\big(\boldsymbol{H}'_k\boldsymbol{\Delta}_{k+1}^{-1}\boldsymbol{\theta}_{k+1} + \boldsymbol{F}'_{k+1}\boldsymbol{H}'_{k+1}\boldsymbol{\Delta}_{k+2}^{-1}\boldsymbol{\theta}_{k+2} + \dots + \boldsymbol{F}'_{k+1}\dots\boldsymbol{F}'_{N-1}\boldsymbol{H}'_{N-1}\boldsymbol{\Delta}_N^{-1}\boldsymbol{\theta}_N\big).$$

Using (42) we can write

$$\hat{\boldsymbol{\gamma}}_{k|N} = var(\boldsymbol{\gamma}_k)\boldsymbol{\Delta}_k^{-1}\boldsymbol{\theta}_k - var(\boldsymbol{\gamma}_k)\boldsymbol{G}'_k\boldsymbol{R}_k = var(\boldsymbol{\gamma}_k)\big(\boldsymbol{\Delta}_k^{-1}\boldsymbol{\theta}_k - \boldsymbol{G}'_k\boldsymbol{R}_k\big) \tag{84}$$

Similarly it follows for $\hat{\boldsymbol{\xi}}_{k|N}$ that

$$\hat{\boldsymbol{\xi}}_{k|N} := L(\boldsymbol{\xi}_k|\boldsymbol{Z}^N) = L(\boldsymbol{\xi}_k|\boldsymbol{Z}^k, \boldsymbol{\theta}_{k+1}, \dots, \boldsymbol{\theta}_N) = L(\boldsymbol{\xi}_k|\boldsymbol{Z}^k) + \sum_{l=k+1}^{N} E(\boldsymbol{\xi}_k\boldsymbol{\theta}'_l)\boldsymbol{\Delta}_l^{-1}\boldsymbol{\theta}_l$$

$$= \sum_{l=k+1}^{N} E(\boldsymbol{\xi}_k\boldsymbol{\theta}'_l)\boldsymbol{\Delta}_l^{-1}\boldsymbol{\theta}_l. \tag{85}$$

since $\boldsymbol{\xi}_k \perp \boldsymbol{Z}^k$ and $E(\boldsymbol{\xi}_k) = 0$ for $k = 1, \dots, N-1$, which is a consequence of the assumptions (28). Using Equation (34) we can write

$$E(\boldsymbol{\xi}_k\boldsymbol{\theta}'_{k+1}) = E[\boldsymbol{\xi}_k(\boldsymbol{H}_k\tilde{\boldsymbol{X}}_{k|k} + \bar{\boldsymbol{C}}_{k+1}\boldsymbol{\Gamma}_k\boldsymbol{\xi}_k + \boldsymbol{\gamma}_k)'].$$

$\tilde{\boldsymbol{X}}_{k|k}$ is a linear function of $\boldsymbol{X}_k$ and $\boldsymbol{Z}^k$ and thus $E(\boldsymbol{\xi}_k\tilde{\boldsymbol{X}}'_{k|k}) = \boldsymbol{0}$. Since $E(\boldsymbol{\xi}_k\gamma'_l) = \boldsymbol{0}$ for $k = 1, \dots, N-1$, $l = 2, \dots, N$ this gives

$$E(\boldsymbol{\xi}_k\boldsymbol{\theta}'_{k+1}) = E[\boldsymbol{\xi}_k(\bar{\boldsymbol{C}}_{k+1}\boldsymbol{\Gamma}_k\boldsymbol{\xi}_k)'] = var(\boldsymbol{\xi}_k)\boldsymbol{\Gamma}'_k\bar{\boldsymbol{C}}'_{k+1}. \tag{86}$$

(34) yields

$$E(\boldsymbol{\xi}_k\boldsymbol{\theta}'_{k+2}) = E[\boldsymbol{\xi}_k(\boldsymbol{H}_{k+1}\tilde{\boldsymbol{X}}_{k+1|k+1} + \bar{\boldsymbol{C}}_{k+2}\boldsymbol{\Gamma}_{k+1}\boldsymbol{\xi}_{k+1} + \boldsymbol{\gamma}_{k+1})'].$$

Since $\boldsymbol{\xi}_k \perp \boldsymbol{\xi}_l$ for $k \neq l$ and using (79) this leads to

$$\begin{aligned} E(\boldsymbol{\xi}_k\boldsymbol{\theta}'_{k+2}) &= E[\boldsymbol{\xi}_k(\boldsymbol{H}_{k+1}\tilde{\boldsymbol{X}}_{k+1|k+1})'] = E[\boldsymbol{\xi}_k(\boldsymbol{F}_{k+1}\tilde{\boldsymbol{X}}_{k|k} + (\boldsymbol{\Gamma}_k - \boldsymbol{G}_{k+1}\bar{\boldsymbol{C}}_{k+1})\boldsymbol{\xi}_k)'] \\ &= var(\boldsymbol{\xi}_k)(\boldsymbol{\Gamma}_k - \boldsymbol{G}_{k+1}\bar{\boldsymbol{C}}_{k+1})'\boldsymbol{H}'_{k+1}. \end{aligned} \tag{87}$$

Proceeding recursively we get

$$\begin{aligned} E(\boldsymbol{\xi}_k\boldsymbol{\theta}'_{k+3}) &= E[\boldsymbol{\xi}_k(\boldsymbol{H}_{k+2}\tilde{\boldsymbol{X}}_{k+2|k+2})'] = E[\boldsymbol{\xi}_k(\boldsymbol{F}_{k+2}\tilde{\boldsymbol{X}}_{k+1|k+1})']\boldsymbol{H}'_{k+2} \\ &= var(\boldsymbol{\xi}_k)(\boldsymbol{\Gamma}_k - \boldsymbol{G}_{k+1}\bar{\boldsymbol{C}}_{k+1})'\boldsymbol{F}'_{k+2}\boldsymbol{H}'_{k+2}. \\ &\vdots \\ E(\boldsymbol{\xi}_k\boldsymbol{\theta}'_N) &= E[\boldsymbol{\xi}_k(\boldsymbol{H}_{N-1}\tilde{\boldsymbol{X}}_{N-1|N-1})'] = E[\boldsymbol{\xi}_k(\boldsymbol{F}_{N-1}\tilde{\boldsymbol{X}}_{N-2|N-2})']\boldsymbol{H}'_{N-1} \\ &= var(\boldsymbol{\xi}_k)(\boldsymbol{\Gamma}_k - \boldsymbol{G}_{k+1}\bar{\boldsymbol{C}}_{k+1})'\boldsymbol{F}'_{k+2}\dots\boldsymbol{F}'_{N-1}\boldsymbol{H}'_{N-1}. \end{aligned} \tag{88}$$

Substitution into (85) yields

$$\hat{\xi}_{k|N} = var(\boldsymbol{\xi}_k)\boldsymbol{\Gamma}'_k\bar{\boldsymbol{C}}'_{k+1}\boldsymbol{\Delta}_{k+1}^{-1}\boldsymbol{\theta}_{k+1} + var(\boldsymbol{\xi}_k)(\boldsymbol{\Gamma}_k - \boldsymbol{G}_{k+1}\bar{\boldsymbol{C}}_{k+1})'\big(\boldsymbol{H}'_{k+1}\boldsymbol{\Delta}_{k+2}^{-1}\boldsymbol{\theta}_{k+2} + \boldsymbol{F}'_{k+2}\boldsymbol{H}'_{k+2}\boldsymbol{\Delta}_{k+3}^{-1}\boldsymbol{\theta}_{k+3} + \boldsymbol{F}'_{k+2}\dots\boldsymbol{F}'_{N-1}\boldsymbol{H}'_{N-1}\boldsymbol{\Delta}_N^{-1}\boldsymbol{\theta}_N\big).$$

Using Equation (42) we get

$$\hat{\boldsymbol{\xi}}_{k|N} = var(\boldsymbol{\xi}_k)\boldsymbol{\Gamma}'_k\bar{\boldsymbol{C}}'_{k+1}\boldsymbol{\Delta}_{k+1}^{-1}\boldsymbol{\theta}_{k+1} + var(\boldsymbol{\xi}_k)(\boldsymbol{\Gamma}_k - \boldsymbol{G}_{k+1}\bar{\boldsymbol{C}}_{k+1})'\boldsymbol{R}_{k+1}. \tag{89}$$

D Smoothed Disturbance Variance Matrices for Linear State Space Models with Coloured Observation Noise

Defining

$$\tilde{\boldsymbol{\gamma}}_{k|N} := \boldsymbol{\gamma}_k - \hat{\boldsymbol{\gamma}}_{k|N} \text{ and } \tilde{\boldsymbol{\xi}}_{k|N} := \boldsymbol{\xi}_k - \hat{\boldsymbol{\xi}}_{k|N},$$

we want to derive an algorithm for recursive computation of $var(\tilde{\boldsymbol{\gamma}}_{k|N})$ and $var(\tilde{\boldsymbol{\xi}}_{k|N})$ for $k \geq 1$. These are also needed for maximum likelihood parameter estimation. For the derivation of this algorithm we can again follow Durbin and Koopman (2001), where the derivation for a linear state space model with white system is shown. Using (2) we consider at first $var(\tilde{\boldsymbol{\gamma}}_{k|k-1}) = var(\boldsymbol{\gamma}_k) - E(\boldsymbol{\gamma}_k \boldsymbol{Z}^{k-1'}) var(\boldsymbol{Z}^{k-1})^{-1} E(\boldsymbol{Z}^{k-1}\boldsymbol{\gamma}_k')$, where $\tilde{\boldsymbol{\gamma}}_{k|k-1} := \boldsymbol{\gamma}_k - L(\boldsymbol{\gamma}_k|\boldsymbol{Z}^{k-1})$ for $k = 2, \ldots, N$. Since $\boldsymbol{\gamma}_k \perp \boldsymbol{Z}^{k-1}$ it follows that $var(\tilde{\boldsymbol{\gamma}}_{k|k-1}) = var(\boldsymbol{\gamma}_k)$. Thus Equation (71) yields

$$var(\tilde{\boldsymbol{\gamma}}_{k|N}) = var(\boldsymbol{\gamma}_k) - \sum_{l=k}^{N} E(\boldsymbol{\gamma}_k \boldsymbol{\theta}_l') \boldsymbol{\Delta}_l^{-1} E(\boldsymbol{\theta}_l \boldsymbol{\gamma}_k') \tag{90}$$

By (81)-(84) we get

$$\begin{aligned}
var(\tilde{\boldsymbol{\gamma}}_{k|N}) = {} & var(\boldsymbol{\gamma}_k) - \big(var(\boldsymbol{\gamma}_k)\boldsymbol{\Delta}_k^{-1} var(\boldsymbol{\gamma}_k) \\
& - var(\boldsymbol{\gamma}_k)\boldsymbol{G}_k' \big[\boldsymbol{H}_k' \boldsymbol{\Delta}_{k+1}^{-1} \boldsymbol{H}_k \boldsymbol{G}_k var(\boldsymbol{\gamma}_k) + \boldsymbol{F}_{k+1}' \boldsymbol{H}_{k+1}' \boldsymbol{\Delta}_{k+2}^{-1} \boldsymbol{H}_{k+1} F_{k+1} \boldsymbol{G}_k var(\boldsymbol{\gamma}_k) + \ldots \\
& + \boldsymbol{F}_{k+1}' \ldots \boldsymbol{F}_{N-1}' \boldsymbol{H}_{N-1}' \boldsymbol{\Delta}_N^{-1} \boldsymbol{H}_{N-1} F_{N-1} \ldots \boldsymbol{F}_{k+1} \boldsymbol{G}_k var(\boldsymbol{\gamma}_k)\big]\big) \\
= {} & var(\boldsymbol{\gamma}_k) - var(\boldsymbol{\gamma}_k)\boldsymbol{\Delta}_k^{-1} var(\boldsymbol{\gamma}_k) + var(\boldsymbol{\gamma}_k)\boldsymbol{G}_k'\big[\boldsymbol{H}_k'\boldsymbol{\Delta}_{k+1}^{-1}\boldsymbol{H}_k \\
& + \boldsymbol{F}_{k+1}'\boldsymbol{H}_{k+1}'\boldsymbol{\Delta}_{k+2}^{-1}\boldsymbol{H}_{k+1}\boldsymbol{F}_{k+1} + \ldots + \boldsymbol{F}_{k+1}' \ldots \boldsymbol{F}_{N-1}'\boldsymbol{H}_{N-1}'\boldsymbol{\Delta}_N^{-1}\boldsymbol{H}_{N-1}\boldsymbol{F}_{N-1} \ldots \boldsymbol{F}_{k+1}\big] \\
& \boldsymbol{G}_k var(\boldsymbol{\gamma}_k).
\end{aligned}$$

Using (41) we can write

$$\begin{aligned}
var(\tilde{\boldsymbol{\gamma}}_{k|N}) &= var(\boldsymbol{\gamma}_k) - var(\boldsymbol{\gamma}_k)\boldsymbol{\Delta}_k^{-1} var(\boldsymbol{\gamma}_k) + var(\boldsymbol{\gamma}_k)\boldsymbol{G}_k'\boldsymbol{P}_k\boldsymbol{G}_k var(\boldsymbol{\gamma}_k) \\
&= var(\boldsymbol{\gamma}_k) - var(\boldsymbol{\gamma}_k)\big(\boldsymbol{\Delta}_k^{-1} - \boldsymbol{G}_k'\boldsymbol{P}_k\boldsymbol{G}_k\big) var(\boldsymbol{\gamma}_k).
\end{aligned} \tag{91}$$

We derive $var(\tilde{\boldsymbol{\xi}}_{k|N})$ analogously, starting by using again (71), recalling that $\boldsymbol{\xi}_k \perp \boldsymbol{Z}^k$ and thus $var(\tilde{\boldsymbol{\xi}}_{k|k}) = var(\boldsymbol{\xi}_k)$:

$$var(\tilde{\boldsymbol{\xi}}_{k|N}) = var(\boldsymbol{\xi}_k) - \sum_{l=k+1}^{n} E(\boldsymbol{\xi}_k \boldsymbol{\theta}_l') \boldsymbol{\Delta}_l^{-1} E(\boldsymbol{\theta}_l \boldsymbol{\xi}_k'). \tag{92}$$

Substitution of (86)-(88) into (92) yields

$$\begin{aligned}
var(\tilde{\boldsymbol{\xi}}_{k|N}) = {} & var(\boldsymbol{\xi}_k) - var(\boldsymbol{\xi}_k)\boldsymbol{\Gamma}_k'\bar{\boldsymbol{C}}_{k+1}'\boldsymbol{\Delta}_{k+1}^{-1}\bar{\boldsymbol{C}}_{k+1}\boldsymbol{\Gamma}_k var(\boldsymbol{\xi}_k) - var(\boldsymbol{\xi}_k)(\boldsymbol{\Gamma}_k - \boldsymbol{G}_{k+1}\bar{\boldsymbol{C}}_{k+1})' \cdot \\
& \cdot \big(\boldsymbol{H}_{k+1}'\boldsymbol{\Delta}_{k+2}^{-1}\boldsymbol{H}_{k+1} + \boldsymbol{F}_{k+2}'\boldsymbol{H}_{k+2}'\boldsymbol{\Delta}_{k+3}^{-1}\boldsymbol{H}_{k+2}\boldsymbol{F}_{k+2} + \ldots \\
& + \boldsymbol{F}_{k+2}' \ldots \boldsymbol{F}_{N-1}'\boldsymbol{H}_{N-1}'\boldsymbol{\Delta}_N^{-1}\boldsymbol{H}_{N-1}\boldsymbol{F}_{N-1} \ldots \boldsymbol{F}_{k+2}\big)(\boldsymbol{\Gamma}_k - \boldsymbol{G}_{k+1}\bar{\boldsymbol{C}}_{k+1}) var(\boldsymbol{\xi}_k).
\end{aligned}$$

Using Definition (41) we get

$$
\begin{aligned}
var(\tilde{\boldsymbol{\xi}}_{k|N}) &= var(\boldsymbol{\xi}_k) - var(\boldsymbol{\xi}_k)\boldsymbol{\Gamma}_k'\bar{\boldsymbol{C}}_{k+1}'\boldsymbol{\Delta}_{k+1}^{-1}\bar{\boldsymbol{C}}_{k+1}\boldsymbol{\Gamma}_k var(\boldsymbol{\xi}_k) \\
&\quad - var(\boldsymbol{\xi}_k)(\boldsymbol{\Gamma}_k - \boldsymbol{G}_{k+1}\bar{\boldsymbol{C}}_{k+1})'\boldsymbol{P}_{k+1}(\boldsymbol{\Gamma}_k - \boldsymbol{G}_{k+1}\bar{\boldsymbol{C}}_{k+1})var(\boldsymbol{\xi}_k).
\end{aligned} \tag{93}
$$

References

Chou, K. (1991, May). *A Stochastic Modeling Approach to Multiscale Signal Processing*. Ph. D. thesis, Massachusetts Institute of Technology.

Chui, C. and G. Chen (1999). *Kalman Filtering With Real Time Applications* (3 ed.). Springer.

Dacorogna, M., R. Gencay, U. A. Müller, R. B. Olsen, and O. V. Pictet (2001). *An Introduction to High Frequency Finance*. Academic Press.

Durbin, J. and S. Koopman (2001). *Time Series Analysis by State Space Methods*. Oxford University Press.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

Harvey, A. C. (1987). *Forecasting,Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Huang, H. and N. Cressie (2001). Multiscale graphical modeling in space: Applications to command and control. In M. Moore (Ed.), *Spatial Statistics: Methological Aspects and Applications*, pp. 83–113. Springer.

Huang, H., N. Cressie, and J. Gabrosek (2002). Fast; resolution-consistent spatial prediction of global processes from satellite data. *Journal of Computational and Graphical Statistics 11*(1), 63–88.

Kailath, T., A. H. Sayed, and B. Hassibi (2000). *Linear Estimation*. Prentice Hall.

Koopman, S. J. and J. Durbin (2001). Filtering and smoothing of state vector for diffuse state space models. Mimeo. Free University Amsterdam, http://www.ssfpack.com/dkbook/.

McLachlan, G. and T. Krishnan (1997). *The EM Algorithm and Extensions*. Wiley.

Schneider, W. (1986). *Der Kalmanfilter als Instrument zur Diagnose und Schätzung variabler Parameter in ökonometrischen Modellen*. Physica-Verlag.

Seber, G. A. F. (1977). *Linear Regression Analysis*. Wiley.

Wei, W. (1990). *Time Series Analysis: Univariate and Multivariate Methods*. Addison Wesley.