



# Article The Weighted Least-Squares Approach to State Estimation in Linear State Space Models: The Case of Correlated Noise Terms

Andreas Galka D

Bundeswehr Technical Centre for Ships and Naval Weapons, Maritime Technology and Research (WTD 71), 24340 Eckernförde, Germany; andreasgalka@bundeswehr.org

Abstract: In this article, a particular approach to deriving recursive state estimators for linear state space models is generalised, namely the weighted least-squares approach introduced by Duncan and Horn in 1972, for the case of the two noise processes arising in such models being cross-correlated; in this context, the fact that in the available literature two different non-equivalent recursive algorithms are presented for the task of state estimation in the aforementioned case is discussed. Although the origin of the difference between these two algorithms can easily be identified, the issue has only rarely been discussed so far. Then the situations in which each of the two original algorithms is proposed. While, strictly speaking, optimal state estimates can be obtained only through the non-recursive weighted least-squares approach, in examples of modelling simulated and real-world data, the recursive generalised Kalman filter shows almost as good performance as the optimal non-recursive filter.

Keywords: state space modelling; Kalman filtering; estimation; smoothing

# 1. Introduction

State space models represent an extremely well-established framework for the predictive modelling of time series, which has found countless applications in statistics, engineering, econometrics, neuroscience, and other fields [1–6]. If both the dynamical processes within the system to be studied and the observation process can be assumed to be linear, the optimal algorithm for recursively estimating the current state of the system is given by the widely used Kalman filter [7]; for estimating past states, smoothing algorithms are available, many of which themselves are based on the results of Kalman filtering. A well-known example of a recursive smoothing algorithm is given by the Rauch–Tung–Striebel (RTS) smoother [8].

There are numerous ways to derive the equations of the Kalman filter and the RTS smoother. This paper starts with an approach that was introduced by Duncan and Horn in 1972 [9]. Their approach is based on expressing the problem of state estimation as a weighted least-squares problem. The first purpose of the present paper is to generalise the weighted least-squares approach to the case of correlations being present between the two noise terms arising in state space models, i.e., the dynamical noise term and the observation noise term; in the paper by Duncan and Horn, it was assumed that these noise terms are uncorrelated.

Duncan and Horn's approach to state estimation corresponds with estimating the states for all time points simultaneously by inverting a single matrix, usually of large dimensions; Kalman filtering and RTS smoothing can be interpreted as a recursive approach to inverting this matrix. However, non-recursive approaches would also be possible, and some authors actually advocate for replacing Kalman filtering and RTS smoothing with other inversion algorithms. As an example, Chan and Jeliazkov [10] are mentioned, who have "rediscovered" the weighted least-squares approach to state estimation, based on



**Citation:** Galka, A . The Weighted Least-Squares Approach to State Estimation in Linear State Space Models: The Case of Correlated Noise Terms. *Algorithms* **2024**, *17*, 194. https://doi.org/10.3390/a17050194

Academic Editor: Alicia Cordero

Received: 22 January 2024 Revised: 27 April 2024 Accepted: 2 May 2024 Published: 4 May 2024



**Copyright:** © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). earlier work by Fahrmeir and Kaufmann [11], apparently without noticing the much earlier paper by Duncan and Horn.

Aravkin [12,13] has reformulated and extended the original work of Duncan and Horn in various directions, but excluding the case of correlated noise terms. He approaches the aforementioned task of inverting a large matrix through a blockwise LDL decomposition of that matrix; Chan and Jeliazkov [10], in contrast, employ Cholesky decomposition. Aravkin has shown that by applying a recursive inversion algorithm for one-block-banded matrices, originally formulated by Bell [14], to the LDL decomposition, the Kalman filter (in its "information filter" variant) and the RTS smoother can be retrieved. The derivation of square-root information filters via the Duncan–Horn approach has been discussed by Paige and Saunders [15] and by Hammarling [16].

The second purpose of this paper is to discuss the fact that for the task of state estimation in the case of correlated noise terms, two different generalised Kalman filter algorithms exist that are not equivalent. Most available textbooks on state space modelling and Kalman filtering present either one or the other of these algorithms; Harvey [2] is a rare example of an author presenting both algorithms. As also mentioned by Harvey, these two algorithms correspond to two slightly different forms of defining the underlying state space model. The difference between these two forms is given solely by the time index of the dynamical noise term, relative to the time index of the state vector; Harvey labels these two forms the "contemporaneous form" and the "future form" [17]. While in control engineering, the "future form" is used almost exclusively, in statistics, the "contemporaneous form" is also frequently used due to its relationship with autoregressive moving-average (ARMA) modelling, which represents an important tool for time series analysis [18].

While, as mentioned, generalisations of the Kalman filter for the case of correlated noise terms are available, generalisations of the RTS smoother for this case are rarely discussed in the literature. In the present paper, these generalisations will be derived using the weighted least-squares approach, both for contemporaneous-form and future-form state space models. It will also be pointed out that by switching between the contemporaneous form and future form, lag-zero cross-correlations between the two noise terms become lag-one cross-correlations, and vice versa; this will explain the fact that the corresponding generalised Kalman filter algorithms are not equivalent.

The next step will then be to define a further generalised state estimation problem that contains both lag-zero and lag-one cross-correlations; it will be demonstrated that for this case, through the weighted least-squares approach, no optimal recursive filtering and smoothing algorithms can be derived. Instead, an approximate recursive algorithm will be presented, which can be obtained by merging the available Kalman filter algorithms for contemporaneous-form and future-form state space models. Related work has been presented recently by Linares-Pérez and coauthors [19], although these authors arrive at a different recursive algorithm.

In practical applications, it is often assumed that, from prior knowledge, it would be known that the noise terms were uncorrelated. In many cases, this assumption may actually be justified, while in other cases, it has to be regarded as an unwarranted constraint. Sometimes, the same source of randomness affects both the dynamical process itself and the observation process; as an example, Chui and Chen [20] mention aircraft inertial navigation. There are also applications where the state space model is used mainly as a device for a quantitative description of the temporal correlation structure present in the original data and not as a model of an underlying physical reality, and, in such cases, it would be difficult to justify the constraint of uncorrelated noise terms. If it is decided to set the crosscorrelation to zero, this should be the result of a corresponding model comparison step, i.e., by proving that a model with vanishing cross-correlation displays the best performance in modelling the data.

The structure of the present paper is as follows. In Section 2, linear state space models are discussed. In Section 3, Kalman filters are discussed for the case of correlated noise

terms, including the corresponding information filters. In Section 4, a recursive approach to solving a certain class of linear equations is reviewed; later, this approach is employed to solve the weighted least-squares problem. Section 5 contains the main results of the paper, including the discussion and generalisation of the weighted least-squares approach to state estimation. In Section 6, the case of both lag-zero and lag-one cross-correlations being present simultaneously is discussed. In Section 7, an example of the application of the proposed algorithms to simulated data is presented, while in Section 8 an example of their application to real-world data follows. The paper closes in Section 9 with a final discussion and conclusions. The Appendix A contains a somewhat lengthy derivation, which establishes the equivalence of the two equations arising in Section 5.

### 2. State Space Models

Let  $\mathbf{y}_t$  denote a given time series of data vectors, where the subscript t = 1, ..., N denotes a discrete time; let *n* denote the dimensions of the data vectors. The given time series shall be modelled by a linear state space model given by a pair of equations, the first of which is known as the observation equation, given by

$$\mathbf{y}_t = \mathsf{C}_t \mathbf{x}_t + \boldsymbol{\epsilon}_t \tag{1}$$

where  $\mathbf{x}_t$  denotes a series of state vectors, with m denoting the dimensions of the state vectors,  $C_t$  denotes an  $(n \times m)$ -dimensional matrix, known as the observation matrix, and  $\epsilon_t$  denotes a series of noise vectors, known as observation noise, which are assumed to be drawn from a Gaussian distribution with a mean vector of zero and an  $(n \times n)$ -dimensional covariance matrix,  $R_t$ . As the subscript t indicates, the observation matrix and the covariance matrix of the observation noise may depend on time.

The second equation is known as the dynamical equation; it can be formulated in two different variants [17], the first, known as the "contemporaneous form", being given by

$$\mathbf{x}_t = \mathsf{A}_t \mathbf{x}_{t-1} + \boldsymbol{\eta}_t \tag{2}$$

while the second variant, known as the "future form", is given by

$$\mathbf{x}_t = \mathsf{A}_t \mathbf{x}_{t-1} + \boldsymbol{\eta}_{t-1} \tag{3}$$

where  $A_t$  denotes an  $(m \times m)$ -dimensional matrix, known as the state transition matrix, and  $\eta_t$  denotes a series of noise vectors, known as dynamical noise, which are assumed to be drawn from a Gaussian distribution with a mean vector of zero and an  $(m \times m)$ dimensional covariance matrix,  $Q_t$ . As the subscript *t* indicates, the state transition matrix and the covariance matrix of the dynamical noise may also depend on time. In the futureform dynamical equation, Equation (3), many authors choose to write the time dependence of the state transition matrix as  $A_{t-1}$  instead of  $A_t$ , but this is merely a matter of definition.

The two noise terms,  $\epsilon_t$  and  $\eta_t$ , may be correlated, i.e., the expectation of the product,  $\eta_t \epsilon_t^T$ , may not vanish, such that an  $(m \times n)$ -dimensional cross-covariance matrix needs to be defined by

$$\mathsf{S}_t = \mathbb{E}(\boldsymbol{\eta}_{\tau}\boldsymbol{\epsilon}_{\tau}^{\mathsf{T}}) \tag{4}$$

where the symbol  $\mathbb{E}$  denotes the expectation (over time  $\tau$ ); this cross-covariance matrix may also depend on time. Then, the covariance matrix of the stacked vector of both noise terms is given by

$$\mathbb{E}\left(\begin{bmatrix}\boldsymbol{\eta}_{\tau}\\\boldsymbol{\varepsilon}_{\tau}\end{bmatrix}\begin{bmatrix}\boldsymbol{\eta}_{\tau}\\\boldsymbol{\varepsilon}_{\tau}\end{bmatrix}^{\mathsf{I}}\right) = \begin{pmatrix}\mathsf{Q}_{t} & \mathsf{S}_{t}\\\mathsf{S}_{t}^{\mathsf{T}} & \mathsf{R}_{t}\end{pmatrix}$$
(5)

For later use, this cross-covariance matrix will now be inverted. First, new symbols for the inverses of the covariance matrices  $R_t$  and  $Q_t$  are introduced:

$$\mathbf{\Phi}_t := \mathsf{R}_t^{-1} , \quad \mathbf{\Theta}_t := \mathsf{Q}_t^{-1} \tag{6}$$

Note that  $\Phi_t$  and  $\Theta_t$  represent information matrices. The desired inverse matrix is formulated as

where

$$\begin{pmatrix} \mathbf{Q}_t & \mathbf{S}_t \\ \mathbf{S}_t^\mathsf{T} & \mathsf{R}_t \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{\Omega}_t & \mathbf{\Sigma}_t \\ \mathbf{\Sigma}_t^\mathsf{T} & \mathbf{\Psi}_t \end{pmatrix}$$
(7)

$$\mathbf{\Omega}_t := \left( \mathsf{Q}_t - \mathsf{S}_t \mathbf{\Phi}_t \mathsf{S}_t^\mathsf{T} \right)^{-1} \tag{8}$$

$$\mathbf{\Psi}_t := \left(\mathsf{R}_t - \mathsf{S}_t^\mathsf{T} \mathbf{\Theta}_t \mathsf{S}_t\right)^{-1} \tag{9}$$

$$\boldsymbol{\Sigma}_t := -\boldsymbol{\Theta}_t \boldsymbol{\mathsf{S}}_t \boldsymbol{\Psi}_t = -\boldsymbol{\Omega}_t \boldsymbol{\mathsf{S}}_t \boldsymbol{\Phi}_t \tag{10}$$

### 3. Kalman Filter Algorithms for the Case of Correlated Noise Terms

The task of state estimation consists of estimating the series of state vectors,  $\mathbf{x}_t$ , t = 1, ..., N, for given data,  $\mathbf{y}_t$ , and given model parameter matrices,  $A_t$ ,  $C_t$ ,  $R_t$ ,  $Q_t$ , and  $S_t$ ; an estimate of the initial state should also be available, e.g., at time t = 0, denoted as  $\mathbf{x}_0$ . The classical solution to this task is given by a twofold recursive algorithm consisting of a forward pass through the data by the Kalman filter, followed by a backward pass by a smoother, such as the Rauch–Tung–Striebel (RTS) smoother [3,21]. The estimates are given by a sequence of mean vectors and covariance matrices, corresponding to multivariate Gaussian distributions.

The Kalman filter consists of a recursion in the forward direction through time. At each time point, first the predicted states,  $\mathbf{x}_{t|t-1}$ , and the corresponding covariance matrices,  $\mathsf{P}_{t|t-1}$ , are computed. Then, second, the filtered states,  $\mathbf{x}_{t|t}$ , and the corresponding covariance matrices,  $\mathsf{P}_{t|t-1}$ , are computed. The notation  $x_{t_1|t_2}$  is defined as an "estimate of  $x_{t_1}$  based on the information available at time  $t_2$ ".

Instead of formulating the recursion for the states  $\mathbf{x}_{t|t-1}$  and  $\mathbf{x}_{t|t}$  and the corresponding covariance matrices,  $\mathsf{P}_{t|t-1}$  and  $\mathsf{P}_{t|t}$ , it may also be formulated for the information states,  $\mathsf{P}_{t|t-1}^{-1}\mathbf{x}_{t|t-1}$  and  $\mathsf{P}_{t|t}^{-1}\mathbf{x}_{t|t}$ , and the corresponding information matrices,  $\mathsf{P}_{t|t-1}^{-1}$  and  $\mathsf{P}_{t|t}^{-1}$ , leading to information filters. Again, new symbols shall be introduced:

$$\boldsymbol{\chi}_{t_1|t_2} := \mathsf{P}_{t_1|t_2}^{-1} \mathbf{x}_{t_1|t_2} , \quad \mathsf{F}_{t_1|t_2} := \mathsf{P}_{t_1|t_2}^{-1}$$
(11)

Depending on whether covariance matrices or information matrices are employed, there are covariance Kalman filters or information Kalman filters (the latter usually simply called information filters).

### 3.1. Contemporaneous Form and Future Form

In Equations (2) and (3), the "contemporaneous form" and the "future form" of the dynamical equation of a state space model are defined. The names of these forms refer to the question of whether the state vector,  $\mathbf{x}_t$ , on the left-hand side of the dynamical equation is contemporaneous to the dynamical noise term,  $\eta_t$ , on the right-hand side or whether it has advanced by one time step into the future relative to that noise term; this terminology was introduced by Harvey [17].

The fact that two variants for the dynamical equation can be formulated is a consequence of using discrete-time dynamics; in continuous-time dynamics, there would be no such distinction. The relative choice of the time indices of these two terms may seem like an insignificant technical detail, and in many situations, this is indeed the case. If the two noise terms,  $\epsilon_t$  and  $\eta_t$ , are uncorrelated, exactly the same Kalman filter algorithm is valid for both the contemporaneous form and the future form.

However, if the two noise terms are correlated, this no longer holds true. In this case, two different Kalman filter algorithms exist that are not equivalent; these two algorithms will be reviewed below in Sections 3.2 and 3.3. Most papers and textbooks present only one of the two possible algorithms without discussing the issue further; the only exception known to this author is the textbook by Harvey [2].

### 3.2. Contemporaneous-Form Kalman Filter

Derivations of this version of the Kalman filter can be found in the textbook by Jazwinski [1], based on orthogonal projection, or the textbook by Brown and Hwang [22], based on a least-mean-squares approach; the algorithm is also given by several other authors, such as Harvey [2], Gibbs [23], and Grewal and Andrews [21]. Its equations are given as follows:

$$\mathbf{x}_{t|t-1} = \mathsf{A}_t \mathbf{x}_{t-1|t-1} \tag{12}$$

$$\mathsf{P}_{t|t-1} = \mathsf{A}_t \mathsf{P}_{t-1|t-1} \mathsf{A}_t^{\mathsf{T}} + \mathsf{Q}_t$$
(13)

$$\mathsf{V}_t = \mathsf{C}_t \mathsf{P}_{t|t-1} \mathsf{C}_t^\mathsf{T} + \mathsf{R}_t + \mathsf{C}_t \mathsf{S}_t + (\mathsf{C}_t \mathsf{S}_t)^\mathsf{T}$$
(14)

$$\mathsf{K}_t = \left(\mathsf{P}_{t|t-1}\mathsf{C}_t^\mathsf{T} + \mathsf{S}_t\right)\mathsf{V}_t^{-1} \tag{15}$$

$$\mathbf{x}_{t|t} = \mathbf{x}_{t|t-1} + \mathsf{K}_t(\mathbf{y}_t - \mathsf{C}_t \mathbf{x}_{t|t-1})$$
(16)

$$\mathsf{P}_{t|t} = \mathsf{P}_{t|t-1} - \mathsf{K}_t \mathsf{V}_t \mathsf{K}_t^\mathsf{T} \tag{17}$$

Compared to the standard Kalman filter with uncorrelated noise terms, only the expressions for the innovation covariance, Equation (14), and for the optimal Kalman gain, Equation (15), contain additional terms. It should be emphasised that for the computation of the innovation likelihood, Equation (14) is not to be used for the innovation covariance; rather, the standard expression,  $C_t P_{t|t-1} C_t^T + R_t$ , has to be used.

Also the contemporaneous form information filter is needed:

$$\mathsf{B}_t = \mathsf{F}_{t-1|t-1} + \mathsf{A}_t^{\mathsf{I}} \, \Theta_t \mathsf{A}_t \tag{18}$$

$$\boldsymbol{\chi}_{t|t-1} = \boldsymbol{\Theta}_t \mathsf{A}_t \mathsf{B}_t^{-1} \boldsymbol{\chi}_{t-1|t-1} \tag{19}$$

$$\mathsf{F}_{t|t-1} = \mathbf{\Theta}_t - \mathbf{\Theta}_t \mathsf{A}_t \mathsf{B}_t^{-1} \mathsf{A}_t^{\mathsf{T}} \mathbf{\Theta}_t \tag{20}$$

$$\boldsymbol{\Gamma}_t = (\mathsf{R}_t - \mathsf{S}_t^\mathsf{T} \mathsf{F}_{t|t-1} \mathsf{S}_t)^{-1} \tag{21}$$

$$\mathbf{G}_t = \mathbf{C}_t + \mathbf{S}_t^\mathsf{T} \mathbf{F}_{t|t-1} \tag{22}$$

$$\boldsymbol{\chi}_{t|t} = (\mathbf{I}^{(m)} + \mathbf{G}_t^{\mathsf{T}} \boldsymbol{\Gamma}_t \mathbf{S}_t^{\mathsf{T}}) \boldsymbol{\chi}_{t|t-1} + \mathbf{G}_t^{\mathsf{T}} \boldsymbol{\Gamma}_t \mathbf{y}_t$$
(23)

$$\mathsf{F}_{t|t} = \mathsf{F}_{t|t-1} + \mathsf{G}_t^{\mathsf{T}} \mathbf{\Gamma}_t \mathsf{G}_t \tag{24}$$

where  $I^{(m)}$  denotes the  $(m \times m)$ -dimensional unity matrix, and the quantities denoted by the symbols B<sub>t</sub>,  $\Gamma_t$ , and G<sub>t</sub> have been introduced in order to simplify the equations.

The information filter algorithm in Equations (18)–(24) can be condensed into two equations (plus the two equations defining  $\Gamma_t$  and  $G_t$ ), directly describing the update of the filtered information state estimate and its information matrix:

$$\boldsymbol{\chi}_{t|t} = (\mathbf{I}^{(m)} + \mathbf{G}_t^{\mathsf{T}} \boldsymbol{\Gamma}_t \mathbf{S}_t^{\mathsf{T}}) \boldsymbol{\Theta}_t \mathbf{A}_t (\mathbf{F}_{t-1|t-1} + \mathbf{A}_t^{\mathsf{T}} \boldsymbol{\Theta}_t \mathbf{A}_t)^{-1} \boldsymbol{\chi}_{t-1|t-1} + \mathbf{G}_t^{\mathsf{T}} \boldsymbol{\Gamma}_t \mathbf{y}_t$$
(25)

$$\mathsf{F}_{t|t} = \mathbf{\Theta}_t - \mathbf{\Theta}_t \mathsf{A}_t \big(\mathsf{F}_{t-1|t-1} + \mathsf{A}_t^\mathsf{T}\mathbf{\Theta}_t \mathsf{A}_t\big)^{-1} \mathsf{A}_t^\mathsf{T}\mathbf{\Theta}_t + \mathsf{G}_t^\mathsf{T}\mathbf{\Gamma}_t \mathsf{G}_t$$
(26)

# 3.3. Future-Form Kalman Filter

This version of the Kalman filter is given in numerous textbooks, such as those by Anderson and Moore [24], Kailath et al. [3], Chui and Chen [20], and Gómez [25]. It can be derived by decorrelating the noise terms, such that, formally, the standard Kalman filter for uncorrelated noise terms can be applied. Its equations are given as follows:

$$\tilde{\mathsf{A}}_t = \mathsf{A}_t - \mathsf{S}_{t-1} \mathbf{\Phi}_{t-1} \mathsf{C}_{t-1} \tag{27}$$

$$\tilde{\mathsf{Q}}_t = \mathsf{Q}_t - \mathsf{S}_t \mathbf{\Phi}_t \mathsf{S}_t^{\mathsf{T}} \tag{28}$$

$$\mathbf{x}_{t|t-1} = \tilde{A}_t \mathbf{x}_{t-1|t-1} + S_{t-1} \Phi_{t-1} \mathbf{y}_{t-1}$$
(29)

$$\mathsf{P}_{t|t-1} = \tilde{\mathsf{A}}_t \mathsf{P}_{t-1|t-1} \tilde{\mathsf{A}}_t^\mathsf{T} + \tilde{\mathsf{Q}}_{t-1} \tag{30}$$

$$\mathsf{V}_t = \mathsf{C}_t \mathsf{P}_{t|t-1} \mathsf{C}_t^\mathsf{T} + \mathsf{R}_t \tag{31}$$

$$\mathsf{K}_t = \mathsf{P}_{t|t-1}\mathsf{C}_t^\mathsf{T}\mathsf{V}_t^{-1} \tag{32}$$

$$\mathbf{x}_{t|t} = \mathbf{x}_{t|t-1} + \mathsf{K}_t(\mathbf{y}_t - \mathsf{C}_t \mathbf{x}_{t|t-1})$$
(33)

$$\mathsf{P}_{t|t} = \mathsf{P}_{t|t-1} - \mathsf{K}_t \mathsf{V}_t \mathsf{K}_t^\mathsf{T} \tag{34}$$

In this filter,  $\tilde{A}_t$ , as defined by Equation (27), represents a *modified state transition matrix*, and  $\tilde{Q}_t$ , as defined by Equation (28), represents a *modified dynamical noise covariance matrix*. Note that

$$\tilde{\mathsf{Q}}_t^{-1} = \mathbf{\Omega}_t \tag{35}$$

where  $\Omega_t$  is defined in Equation (8). As a further modification to the standard Kalman filter for uncorrelated noise terms, a new term,  $S_{t-1}\Phi_{t-1}\mathbf{y}_{t-1}$ , arises in the equation for the predicted state estimate, Equation (29); however, since this term is precisely known, it does not cause any further changes to the standard Kalman filter. As another subtle difference, in the case of the future-form Kalman filter, in Equation (30), the time index of the modified dynamical noise covariance matrix has to be chosen as t - 1 in contrast to Equation (13).

Also, the future-form information filter is needed:

$$\tilde{\mathsf{B}}_{t} = \mathsf{F}_{t-1|t-1} + \tilde{\mathsf{A}}_{t}^{\mathsf{T}} \mathbf{\Omega}_{t-1} \tilde{\mathsf{A}}_{t}$$
(36)

$$\mathsf{F}_{t|t-1} = \mathbf{\Omega}_{t-1} - \mathbf{\Omega}_{t-1} \tilde{\mathsf{A}}_t \tilde{\mathsf{B}}_t^{-1} \tilde{\mathsf{A}}_t^{\mathsf{T}} \mathbf{\Omega}_{t-1}$$
(37)

$$\boldsymbol{\chi}_{t|t-1} = \boldsymbol{\Omega}_{t-1} \tilde{\mathsf{A}}_t \tilde{\mathsf{B}}_t^{-1} \boldsymbol{\chi}_{t-1|t-1} + \mathsf{F}_{t|t-1} \mathsf{S}_{t-1} \boldsymbol{\Phi}_{t-1} \mathbf{y}_{t-1}$$
(38)

$$\boldsymbol{\chi}_{t|t} = \boldsymbol{\chi}_{t|t-1} + \mathsf{C}_t^\mathsf{T} \boldsymbol{\Phi}_t \mathbf{y}_t \tag{39}$$

$$\mathsf{F}_{t|t} = \mathsf{F}_{t|t-1} + \mathsf{C}_t^{\mathsf{T}} \, \boldsymbol{\Phi}_t \mathsf{C}_t \tag{40}$$

where the quantity denoted by the symbol  $\tilde{B}_t$  has been introduced in order to simplify the equations.

Again, the information filter algorithm in Equations (36)–(40) can be condensed into two equations:

$$\chi_{t|t} = \boldsymbol{\Omega}_{t-1} \tilde{\mathsf{A}}_t \big(\mathsf{F}_{t-1|t-1} + \tilde{\mathsf{A}}_t^{\mathsf{T}} \boldsymbol{\Omega}_{t-1} \tilde{\mathsf{A}}_t\big)^{-1} \chi_{t-1|t-1} + \Big(\boldsymbol{\Omega}_{t-1} - \boldsymbol{\Omega}_{t-1} \tilde{\mathsf{A}}_t \big(\mathsf{F}_{t-1|t-1} + \tilde{\mathsf{A}}_t^{\mathsf{T}} \boldsymbol{\Omega}_{t-1} \tilde{\mathsf{A}}_t\big)^{-1} \tilde{\mathsf{A}}_t^{\mathsf{T}} \boldsymbol{\Omega}_{t-1}\Big) \mathsf{S}_{t-1} \boldsymbol{\Phi}_{t-1} \mathbf{y}_{t-1} + \mathsf{C}_t^{\mathsf{T}} \boldsymbol{\Phi}_t \mathbf{y}_t$$
(41)

$$\mathsf{F}_{t|t} = \mathbf{\Omega}_{t-1} - \mathbf{\Omega}_{t-1} \tilde{\mathsf{A}}_t \big(\mathsf{F}_{t-1|t-1} + \tilde{\mathsf{A}}_t^\mathsf{T} \mathbf{\Omega}_{t-1} \tilde{\mathsf{A}}_t \big)^{-1} \tilde{\mathsf{A}}_t^\mathsf{T} \mathbf{\Omega}_{t-1} + \mathsf{C}_t^\mathsf{T} \mathbf{\Phi}_t \mathsf{C}_t \tag{42}$$

It has to be emphasised that the filter algorithm in Equations (12)–(17) is not equivalent to the filter algorithm in Equations (27)–(34), and this applies likewise to the information filter algorithms. This is obvious from a direct comparison of the equations and can readily be confirmed by numerical examples.

# 4. Recursive Solution of a Linear Equation with a Symmetric One-Block-Banded Matrix

Before embarking on the actual discussion of the weighted least-squares approach to state estimation, an algorithm for the recursive solution of a linear equation needs to be briefly reviewed. Let the linear equation be given by

$$\mathcal{V} = \mathcal{M}\mathcal{X},\tag{43}$$

where the square matrix, M, has a particular structure, known as symmetric one-blockbanded matrix structure. This structure is given by

$$\mathcal{M} = \begin{pmatrix} \mathsf{M}_{1} & \mathsf{W}_{2} & 0 & \dots & 0 & 0 \\ \mathsf{W}_{2}^{\mathsf{T}} & \mathsf{M}_{2} & \mathsf{W}_{3} & \dots & 0 & 0 \\ 0 & \mathsf{W}_{3}^{\mathsf{T}} & \mathsf{M}_{3} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \mathsf{M}_{N-1} & \mathsf{W}_{N} \\ 0 & 0 & 0 & \dots & \mathsf{W}_{N}^{\mathsf{T}} & \mathsf{M}_{N} \end{pmatrix},$$
(44)

where  $M_t$ , t = 1, ..., N, denotes a set of invertible symmetric positive definite  $(m \times m)$ dimensional square matrices, and  $W_t$ , t = 2, ..., N, denotes another set of  $(m \times m)$ dimensional square matrices. For the vectors  $\mathcal{V}$  and  $\mathcal{X}$ , a similar partitioning as for  $\mathcal{M}$  shall be defined:

$$\mathcal{V} = \begin{pmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_N \end{pmatrix} \qquad \qquad \mathcal{X} = \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_N \end{pmatrix}, \qquad (45)$$

where the dimensions of the subvectors  $\mathbf{v}_t$  and  $\mathbf{x}_t$ , t = 1, ..., N, shall be m.

The solution to Equation (43) can be based on the blockwise LDL decomposition of  ${\cal M}$  according to

$$\mathcal{M} = \mathcal{L}\mathcal{D}\mathcal{L}^{\mathsf{T}},\tag{46}$$

where

$$\mathcal{L} = \begin{pmatrix} \mathsf{I}^{(m)} & 0 & 0 & \dots & 0 & 0 \\ \mathsf{W}_2^{\mathsf{T}}\mathsf{Z}_1^{-1} & \mathsf{I}^{(m)} & 0 & \dots & 0 & 0 \\ 0 & \mathsf{W}_3^{\mathsf{T}}\mathsf{Z}_2^{-1} & \mathsf{I}^{(m)} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \mathsf{I}^{(m)} & 0 \\ 0 & 0 & 0 & \dots & \mathsf{W}_N^{\mathsf{T}}\mathsf{Z}_{N-1}^{-1} & \mathsf{I}^{(m)} \end{pmatrix}$$
(47)

and

$$\mathcal{D} = \begin{pmatrix} Z_1 & 0 & 0 & \dots & 0 & 0 \\ 0 & Z_2 & 0 & \dots & 0 & 0 \\ 0 & 0 & Z_3 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & Z_{N-1} & 0 \\ 0 & 0 & 0 & \dots & 0 & Z_N \end{pmatrix}$$
(48)

Then, in the first step, the equation

$$\mathcal{V} = \mathcal{M}\mathcal{X} = \mathcal{L}\mathcal{D}\mathcal{L}^{\mathsf{T}}\mathcal{X}$$

is recursively solved for  $\mathcal{Z}$ , which is defined by

$$\mathcal{Z} = \mathcal{D}\mathcal{L}^{\mathsf{T}}\mathcal{X},\tag{49}$$

(54)

while in the second step, Equation (49) is recursively solved for  $\mathcal{X}$ . The corresponding recursions were given by Bell [14], as follows:

$$1. \quad \mathsf{Z}_1 = \mathsf{M}_1 \tag{50}$$

 $z_1 = v_1$ 2. "forward" recursion over t from t = 2 until t = N: (51)

$$Z_t = M_t - W_t^T Z_{t-1}^{-1} W_t$$
(52)

$$\mathbf{z}_t = \mathbf{v}_t - \mathbf{W}_t^{\mathsf{T}} \mathbf{Z}_{t-1}^{-1} \mathbf{z}_{t-1}$$
(53)

- 3.  $\mathbf{x}_N = \mathsf{Z}_N^{-1} \mathbf{z}_N$
- 4. "backward" recursion over t from t = N 1 until t = 1:  $\mathbf{x}_t = \mathbf{Z}_t^{-1}(\mathbf{z}_t - \mathbf{W}_{t+1}\mathbf{x}_{t+1})$ (55)

It is not hard to see that the "forward" recursion leads to a Kalman filter, while the "backward" recursion leads to a smoother.

In addition to the equations given by Bell, another equation needs to be added to the "backward" recursion. For this purpose, the task of computing the inverse of the matrix  $\mathcal{M}$  explicitly is studied, which can be carried out by inverting the matrices  $\mathcal{L}$  and  $\mathcal{D}$  separately.

 $\mathcal{D}$  has a block-diagonal structure, with blocks  $[\mathcal{D}]_{tt} = Z_t$ , where the index *t* refers to blocks. Its inverse also has a block-diagonal structure, with blocks  $[\mathcal{D}^{-1}]_{tt} = Z_t^{-1}$ .

 $\mathcal{L}$  has a block-bidiagonal structure, with blocks  $[\mathcal{L}]_{tt} = \mathsf{I}^{(m)}$  and  $[\mathcal{L}]_{t+1,t} = \mathsf{W}_{t+1}^{\mathsf{T}}\mathsf{Z}_{t}^{-1}$ . Its inverse has a lower block-triangular structure, with blocks

$$[\mathcal{L}^{-1}]_{st} = \begin{cases} \mathsf{I}^{(m)} & \text{if } s = t \\ (-1)^{s+t} \prod_{\tau=s}^{t-1} [\mathcal{L}]_{\tau+1,\tau} & \text{if } s > t \end{cases}$$

where the indices *s* and *t* refer to blocks.

Then, if the product  $\mathcal{L}^{-T}\mathcal{D}^{-1}\mathcal{L}^{-1}$  is formed, it is found that the blocks on the diagonal of  $\mathcal{M}^{-1}$  can be computed by a "backward" recursion over *t* from t = N - 1 to t = 1, according to

$$[\mathcal{M}^{-1}]_{NN} = \mathsf{Z}_{N}^{-1} \tag{56}$$

$$[\mathcal{M}^{-1}]_{tt} = \mathsf{Z}_t^{-1} + \mathsf{Z}_t^{-1} \mathsf{W}_{t+1} [\mathcal{M}^{-1}]_{t+1,t+1} \mathsf{W}_{t+1}^{\mathsf{T}} \mathsf{Z}_t^{-1}$$
(57)

The idea of inverting the matrix  $\mathcal{M}$  via blockwise LDL decomposition, in the context of state estimation, has also been proposed by Fahrmeir and Kaufmann [11], while Chan and Jeliazkov [10] prefer to employ Cholesky decomposition for this purpose.

# 5. Weighted Least-Squares Approach to State Estimation

Various authors have noticed that the state estimation problem can be cast as a regression problem; a good review of the earlier literature has been provided by Solo [26]. A particularly influential paper was published in 1972 by Duncan and Horn [9], who showed that the linear state estimation problem can be solved by weighted least squares. Later, the same approach was "rediscovered" by Chan and Jeliazkov [10]; see also the related work by Fahrmeir and Kaufmann [11]. The derivation given in the paper by Duncan and Horn will now briefly be reproduced, but it will be generalised for the case  $S_t \neq 0$ .

# 5.1. Contemporaneous Form

First, the contemporaneous form of the state space model is considered, which is also the form employed by Duncan and Horn. The starting point is given by writing down the equations of the state space model, Equations (1) and (2), for all time points, t = 1, ..., N:

$$\mathbf{x}_{1} = \mathsf{A}_{1}\mathbf{x}_{0} + \boldsymbol{\eta}_{1}$$

$$\mathbf{y}_{1} = \mathsf{C}_{1}\mathbf{x}_{1} + \boldsymbol{\epsilon}_{1}$$

$$\mathbf{x}_{2} = \mathsf{A}_{2}\mathbf{x}_{1} + \boldsymbol{\eta}_{2}$$

$$\mathbf{y}_{2} = \mathsf{C}_{2}\mathbf{x}_{2} + \boldsymbol{\epsilon}_{2}$$

$$\vdots$$

$$\mathbf{x}_{N} = \mathsf{A}_{N}\mathbf{x}_{N-1} + \boldsymbol{\eta}_{N}$$

$$\mathbf{y}_{N} = \mathsf{C}_{N}\mathbf{x}_{N} + \boldsymbol{\epsilon}_{N}$$
(58)

Note that in the first of these equations, the term  $A_1x_0$  can be interpreted as the prediction of the state at time t = 1, denoted as  $\mathbf{x}_{1|0} = A_1 \mathbf{x}_0$ . Then, the set of Equation (58) may be rearranged as

$$-\mathbf{x}_{1|0} = -\mathbf{x}_{1} + \boldsymbol{\eta}_{1}$$
  

$$\mathbf{y}_{1} = \mathsf{C}_{1}\mathbf{x}_{1} + \boldsymbol{\epsilon}_{1}$$
  

$$\mathbf{0} = \mathsf{A}_{2}\mathbf{x}_{1} - \mathbf{x}_{2} + \boldsymbol{\eta}_{2}$$
  

$$\mathbf{y}_{2} = \mathsf{C}_{2}\mathbf{x}_{2} + \boldsymbol{\epsilon}_{2}$$
  

$$\vdots$$
  

$$\mathbf{0} = \mathsf{A}_{N}\mathbf{x}_{N-1} - \mathbf{x}_{N} + \boldsymbol{\eta}_{N}$$
  

$$\mathbf{y}_{N} = \mathsf{C}_{N}\mathbf{x}_{N} + \boldsymbol{\epsilon}_{N}$$
  
(59)

By defining

$$\mathcal{Y} = \begin{pmatrix} -\mathbf{x}_{1|0} \\ \mathbf{y}_1 \\ \mathbf{0} \\ \mathbf{y}_2 \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{y}_N \end{pmatrix}, \quad \mathcal{X} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \\ \vdots \\ \mathbf{x}_{N-1} \\ \mathbf{x}_N \end{pmatrix}, \quad \mathcal{E} = \begin{pmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\epsilon}_1 \\ \boldsymbol{\eta}_2 \\ \boldsymbol{\epsilon}_2 \\ \boldsymbol{\eta}_3 \\ \vdots \\ \boldsymbol{\eta}_N \\ \boldsymbol{\epsilon}_N \end{pmatrix}$$
(60)

and

$$\mathcal{A} = \begin{pmatrix} -\mathbf{I}^{(m)} & 0 & 0 & \dots & 0 & 0 \\ C_1 & 0 & 0 & \dots & 0 & 0 \\ A_2 & -\mathbf{I}^{(m)} & 0 & \dots & 0 & 0 \\ 0 & C_2 & 0 & \dots & 0 & 0 \\ 0 & A_3 & -\mathbf{I}^{(m)} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & A_N & -\mathbf{I}^{(m)} \\ 0 & 0 & 0 & \dots & 0 & C_N \end{pmatrix},$$
(61)

~

the set of Equation (59) can be rewritten as

$$\mathcal{Y} = \mathcal{A}\mathcal{X} + \mathcal{E} \tag{62}$$

`

The vector  $\mathcal{Y}$  contains the known data, including the prediction of the state at t = 1. The vector  $\mathcal{E}$  contains the unknown noise terms, for which the covariance matrix is given by

$$\mathcal{R} = \begin{pmatrix} Q_1 & S_1 & 0 & 0 & \dots & 0 & 0 \\ S_1^{\mathsf{T}} & \mathsf{R}_1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \mathsf{Q}_2 & \mathsf{S}_2 & \dots & 0 & 0 \\ 0 & 0 & \mathsf{S}_2^{\mathsf{T}} & \mathsf{R}_2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & \mathsf{Q}_N & \mathsf{S}_N \\ 0 & 0 & 0 & 0 & \dots & \mathsf{S}_N^{\mathsf{T}} & \mathsf{R}_N \end{pmatrix}$$
(63)

The inverse of  $\mathcal{R}$  is given by

$$\mathcal{R}^{-1} = \begin{pmatrix} \Omega_1 & \Sigma_1 & 0 & 0 & \dots & 0 & 0 \\ \Sigma_1^{\mathsf{T}} & \Psi_1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \Omega_2 & \Sigma_2 & \dots & 0 & 0 \\ 0 & 0 & \Sigma_2^{\mathsf{T}} & \Psi_2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & \Omega_N & \Sigma_N \\ 0 & 0 & 0 & 0 & \dots & \Sigma_N^{\mathsf{T}} & \Psi_N \end{pmatrix}$$
(64)

where the matrices  $\Omega_t$ ,  $\Psi_t$ , and  $\Sigma_t$  have been defined in Equations (8)–(10).

Equations (60)–(63) represent a weighted least-squares problem to be solved for  $\mathcal{X}$ , i.e., the sequence of the states. The solution,  $\hat{\mathcal{X}}$ , to this problem has to fulfil the equation

$$\mathcal{A}^{\mathsf{T}}\mathcal{R}^{-1}\mathcal{Y} = \mathcal{A}^{\mathsf{T}}\mathcal{R}^{-1}\mathcal{A}\,\hat{\mathcal{X}} \tag{65}$$

If the matrix product  $\mathcal{A}^{\mathsf{T}}\mathcal{R}^{-1}\mathcal{A}$  has full rank, the estimated solution is given by

$$\hat{\mathcal{X}} = \left(\mathcal{A}^{\mathsf{T}} \mathcal{R}^{-1} \mathcal{A}\right)^{-1} \mathcal{A}^{\mathsf{T}} \mathcal{R}^{-1} \mathcal{Y}$$
(66)

The covariance matrix of the estimated solution  $\hat{\mathcal{X}}$  is given by

$$\mathcal{P}(\hat{\mathcal{X}}) = \left(\mathcal{A}^{\mathsf{T}} \mathcal{R}^{-1} \mathcal{A}\right)^{-1} \tag{67}$$

By comparing with Equation (43), the following correspondences can be found:

$$\mathcal{V} = \mathcal{A}^{\mathsf{T}} \mathcal{R}^{-1} \mathcal{Y}$$
 and  $\mathcal{M} = \mathcal{A}^{\mathsf{T}} \mathcal{R}^{-1} \mathcal{A}$ 

By evaluating the matrix product  $\mathcal{A}^{\mathsf{T}}\mathcal{R}^{-1}\mathcal{A}$ , using Equation (7), it can easily be confirmed that a symmetric one-block-banded matrix is produced, with a shape given by Equation (44) and blocks given by

$$\mathsf{M}_{t} = \begin{cases} \mathbf{\Omega}_{t} - \mathbf{\Sigma}_{t}\mathsf{C}_{t} - (\mathbf{\Sigma}_{t}\mathsf{C}_{t})^{\mathsf{T}} + \mathsf{C}_{t}^{\mathsf{T}}\Psi_{t}\mathsf{C}_{t} + \mathsf{A}_{t+1}^{\mathsf{T}}\mathbf{\Omega}_{t+1}\mathsf{A}_{t+1} & \text{if } t = 1, \dots, N-1\\ \mathbf{\Omega}_{t} - \mathbf{\Sigma}_{t}\mathsf{C}_{t} - (\mathbf{\Sigma}_{t}\mathsf{C}_{t})^{\mathsf{T}} + \mathsf{C}_{t}^{\mathsf{T}}\Psi_{t}\mathsf{C}_{t} & \text{if } t = N \end{cases}$$
(68)

$$W_t = A_t^{\mathsf{T}} \Sigma_t C_t - A_t^{\mathsf{T}} \Omega_t \quad , \quad t = 2, \dots, N$$
(69)

such that the algorithm in Equations (50)–(55) can be applied to solve Equation (65).

The vector  $\mathcal{A}^{\mathsf{T}}\mathcal{R}^{-1}\mathcal{Y}$ , corresponding to the "given-data" vector,  $\mathcal{V}$ , in Equation (43), consists of stacked *m*-dimensional subvectors, given by

$$\mathbf{v}_{t} = \begin{cases} (\mathbf{C}_{1}^{\mathsf{T}} \mathbf{\Psi}_{1} - \mathbf{\Sigma}_{1}) \mathbf{y}_{1} + \mathbf{A}_{2}^{\mathsf{T}} \mathbf{\Sigma}_{2} \mathbf{y}_{2} - (\mathbf{C}_{1}^{\mathsf{T}} \mathbf{\Sigma}_{1}^{\mathsf{T}} - \mathbf{\Omega}_{1}) \mathbf{x}_{1|0} & \text{if } t = 1 \\ (\mathbf{C}_{t}^{\mathsf{T}} \mathbf{\Psi}_{t} - \mathbf{\Sigma}_{t}) \mathbf{y}_{t} + \mathbf{A}_{t+1}^{\mathsf{T}} \mathbf{\Sigma}_{t+1} \mathbf{y}_{t+1} & \text{if } t = 2, \dots, N-1 \\ (\mathbf{C}_{N}^{\mathsf{T}} \mathbf{\Psi}_{N} - \mathbf{\Sigma}_{N}) \mathbf{y}_{N} & \text{if } t = N \end{cases}$$
(70)

By inserting Equations (68) and (69) into Equation (52), it follows, for t = 1, ..., N - 1, that

$$Z_{t} = \boldsymbol{\Omega}_{t} - \boldsymbol{\Sigma}_{t} C_{t} - (\boldsymbol{\Sigma}_{t} C_{t})^{\mathsf{T}} + C_{t}^{\mathsf{T}} \boldsymbol{\Psi}_{t} C_{t} + A_{t+1}^{\mathsf{T}} \boldsymbol{\Omega}_{t+1} A_{t+1} - (A_{t}^{\mathsf{T}} \boldsymbol{\Sigma}_{t} C_{t} - A_{t}^{\mathsf{T}} \boldsymbol{\Omega}_{t})^{\mathsf{T}} Z_{t-1}^{-1} (A_{t}^{\mathsf{T}} \boldsymbol{\Sigma}_{t} C_{t} - A_{t}^{\mathsf{T}} \boldsymbol{\Omega}_{t})$$
(71)

By inserting Equations (69) and (70) into Equation (53), it follows, for t = 2, ..., N - 1, that

$$\mathbf{z}_{t} = \left(\mathsf{C}_{t}^{\mathsf{T}} \mathbf{\Psi}_{t} - \boldsymbol{\Sigma}_{t}\right) \mathbf{y}_{t} + \mathsf{A}_{t+1}^{\mathsf{T}} \boldsymbol{\Sigma}_{t+1} \mathbf{y}_{t+1} - \left(\mathsf{A}_{t}^{\mathsf{T}} \boldsymbol{\Sigma}_{t} \mathsf{C}_{t} - \mathsf{A}_{t}^{\mathsf{T}} \boldsymbol{\Omega}_{t}\right)^{\mathsf{T}} \mathsf{Z}_{t-1}^{-1} \mathbf{z}_{t-1}$$
(72)

By inserting Equation (69) into Equations (55) and (57), it follows, for t = N - 1, ..., 1, that

$$\mathbf{x}_{t} = \mathsf{Z}_{t}^{-1} \Big( \mathbf{z}_{t} - \big(\mathsf{A}_{t+1}^{\mathsf{T}} \boldsymbol{\Sigma}_{t+1} \mathsf{C}_{t+1} - \mathsf{A}_{t+1}^{\mathsf{T}} \boldsymbol{\Omega}_{t+1} \big) \mathbf{x}_{t+1} \Big)$$
(73)

$$\mathsf{P}_{t|N} = \mathsf{Z}_{t}^{-1} + \mathsf{Z}_{t}^{-1} \big(\mathsf{A}_{t+1}^{\mathsf{T}} \boldsymbol{\Sigma}_{t+1} \mathsf{C}_{t+1} - \mathsf{A}_{t+1}^{\mathsf{T}} \boldsymbol{\Omega}_{t+1} \big) \times \mathsf{P}_{t+1|N} \big(\mathsf{A}_{t+1}^{\mathsf{T}} \boldsymbol{\Sigma}_{t+1} \mathsf{C}_{t+1} - \mathsf{A}_{t+1}^{\mathsf{T}} \boldsymbol{\Omega}_{t+1} \big)^{\mathsf{T}} \mathsf{Z}_{t}^{-1} \quad (74)$$

Equations (71) and (72) represent an algorithm for forward recursive filtering, in the case of contemporaneous-form models, to be applied for t = 2, ..., N - 1; the modifications for the limit cases t = 1 and t = N follow accordingly from Equations (50), (51), (69), and (70).

Equations (73) and (74) represent an algorithm for backward recursive smoothing, in the case of contemporaneous-form models, to be applied for t = N - 1, ..., 1; the modifications for the limit case t = N follow accordingly from Equations (54) and (56), using  $Z_N^{-1}$ , as provided by the forward filter. Note that replacing  $[\mathcal{M}^{-1}]_{tt}$  in Equation (57) with  $P_{t|N}$  is justified due to Equation (67).

# 5.2. Future Form

Now, it shall be assumed that the state space model is formulated in the future form; see Equations (1) and (3). The corresponding temporal shift in the dynamical noise term, as compared to the state space model in the contemporaneous form, has the effect that, within the matrix  $\mathcal{R}$ , the cross-covariance matrix S<sub>t</sub> moves to different block positions:

$$\mathcal{R} = \begin{pmatrix} Q_0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & \mathsf{R}_1 & \mathsf{S}_1^\mathsf{T} & \dots & 0 & 0 & 0 \\ 0 & \mathsf{S}_1 & \mathsf{Q}_1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \mathsf{R}_{N-1} & \mathsf{S}_{N-1}^\mathsf{T} & 0 \\ 0 & 0 & 0 & \dots & \mathsf{S}_{N-1} & \mathsf{Q}_{N-1} & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & \mathsf{R}_N \end{pmatrix}$$
(75)

The inverse of  $\mathcal{R}$  is given by

$$\mathcal{R}^{-1} = \begin{pmatrix} \Theta_0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & \Psi_1 & \Sigma_1^{\mathsf{T}} & \dots & 0 & 0 & 0 \\ 0 & \Sigma_1 & \Omega_1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \Psi_{N-1} & \Sigma_{N-1}^{\mathsf{T}} & 0 \\ 0 & 0 & 0 & \dots & \Sigma_{N-1} & \Omega_{N-1} & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & \Phi_N \end{pmatrix}$$
(76)

Also, in this case, the matrix product  $\mathcal{A}^{\mathsf{T}}\mathcal{R}^{-1}\mathcal{A}$  yields a one-block-banded matrix, according to Equation (44), with blocks given by

$$M_{t} = \begin{cases} \Theta_{0} + A_{2}^{\mathsf{T}} \Sigma_{1} C_{1} + (A_{2}^{\mathsf{T}} \Sigma_{1} C_{1})^{\mathsf{T}} + C_{1}^{\mathsf{T}} \Psi_{1} C_{1} + A_{2}^{\mathsf{T}} \Omega_{1} A_{2} & \text{if } t = 1 \\ \Omega_{t-1} + A_{t+1}^{\mathsf{T}} \Sigma_{t} C_{t} + (A_{t+1}^{\mathsf{T}} \Sigma_{t} C_{t})^{\mathsf{T}} + C_{t}^{\mathsf{T}} \Psi_{t} C_{t} \\ + A_{t+1}^{\mathsf{T}} \Omega_{t} A_{t+1} & \text{if } t = 2, \dots, N-1 \\ \Omega_{N-1} + C_{N}^{\mathsf{T}} \Phi_{N} C_{N} & \text{if } t = N \end{cases}$$

$$W_{t} = -C_{t-1}^{\mathsf{T}} \Sigma_{t-1}^{\mathsf{T}} - A_{t}^{\mathsf{T}} \Omega_{t-1}$$
(77)

The vector  $\mathcal{A}^{\mathsf{T}}\mathcal{R}^{-1}\mathcal{Y}$  consists of stacked *m*-dimensional subvectors, given by

$$\mathbf{v}_{t} = \begin{cases} (\mathsf{C}_{1}^{\mathsf{T}} \mathbf{\Psi}_{1} + \mathsf{A}_{2}^{\mathsf{T}} \boldsymbol{\Sigma}_{1}) \mathbf{y}_{1} + \boldsymbol{\Theta}_{0} \mathbf{x}_{1|0} & \text{if } t = 1\\ (\mathsf{C}_{t}^{\mathsf{T}} \mathbf{\Psi}_{t} + \mathsf{A}_{t+1}^{\mathsf{T}} \boldsymbol{\Sigma}_{t}) \mathbf{y}_{t} - \boldsymbol{\Sigma}_{t-1} \mathbf{y}_{t-1} & \text{if } t = 2, \dots, N-1\\ \mathsf{C}_{N}^{\mathsf{T}} \boldsymbol{\Phi}_{N} \mathbf{y}_{N} - \boldsymbol{\Sigma}_{N-1} \mathbf{y}_{N-1} & \text{if } t = N \end{cases}$$
(79)

By inserting Equations (77) and (78) into Equation (52), it follows, for t = 2, ..., N - 1, that

$$Z_{t} = \boldsymbol{\Omega}_{t-1} + \mathsf{A}_{t+1}^{\mathsf{T}}\boldsymbol{\Sigma}_{t}\mathsf{C}_{t} + (\mathsf{A}_{t+1}^{\mathsf{T}}\boldsymbol{\Sigma}_{t}\mathsf{C}_{t})^{\mathsf{T}} + \mathsf{C}_{t}^{\mathsf{T}}\boldsymbol{\Psi}_{t}\mathsf{C}_{t} + \mathsf{A}_{t+1}^{\mathsf{T}}\boldsymbol{\Omega}_{t}\mathsf{A}_{t+1} - \left(\mathsf{C}_{t-1}^{\mathsf{T}}\boldsymbol{\Sigma}_{t-1}^{\mathsf{T}} + \mathsf{A}_{t}^{\mathsf{T}}\boldsymbol{\Omega}_{t-1}\right)^{\mathsf{T}}\mathsf{Z}_{t-1}^{-1}\left(\mathsf{C}_{t-1}^{\mathsf{T}}\boldsymbol{\Sigma}_{t-1}^{\mathsf{T}} + \mathsf{A}_{t}^{\mathsf{T}}\boldsymbol{\Omega}_{t-1}\right)$$
(80)

By inserting Equations (78) and (79) into Equation (53), it follows, for t = 2, ..., N - 1, that

$$\mathbf{z}_{t} = \left(\mathsf{C}_{t}^{\mathsf{T}} \boldsymbol{\Psi}_{t} + \mathsf{A}_{t+1}^{\mathsf{T}} \boldsymbol{\Sigma}_{t}\right) \mathbf{y}_{t} - \boldsymbol{\Sigma}_{t-1} \mathbf{y}_{t-1} + \left(\mathsf{C}_{t-1}^{\mathsf{T}} \boldsymbol{\Sigma}_{t-1}^{\mathsf{T}} + \mathsf{A}_{t}^{\mathsf{T}} \boldsymbol{\Omega}_{t-1}\right)^{\mathsf{T}} \mathsf{Z}_{t-1}^{-1} \mathbf{z}_{t-1}$$
(81)

By inserting Equation (78) into Equations (55) and (57), it follows, for t = N - 1, ..., 1, that

$$\mathbf{x}_{t} = \mathsf{Z}_{t}^{-1} \Big( \mathbf{z}_{t} + \big( \mathsf{C}_{t}^{\mathsf{T}} \boldsymbol{\Sigma}_{t}^{\mathsf{T}} + \mathsf{A}_{t+1}^{\mathsf{T}} \boldsymbol{\Omega}_{t} \big) \mathbf{x}_{t+1} \Big)$$
(82)

$$\mathsf{P}_{t|N} = \mathsf{Z}_t^{-1} + \mathsf{Z}_t^{-1} \big( \mathsf{C}_t^\mathsf{T} \boldsymbol{\Sigma}_t^\mathsf{T} + \mathsf{A}_{t+1}^\mathsf{T} \boldsymbol{\Omega}_t \big) \mathsf{P}_{t+1|N} \big( \mathsf{C}_t^\mathsf{T} \boldsymbol{\Sigma}_t^\mathsf{T} + \mathsf{A}_{t+1}^\mathsf{T} \boldsymbol{\Omega}_t \big)^\mathsf{T} \mathsf{Z}_t^{-1}$$
(83)

Equations (80) and (81) represent an algorithm for forward recursive filtering, in the case of future-form models, to be applied for t = 2, ..., N - 1; the modifications for the limit cases t = 1 and t = N follow accordingly from Equations (50), (51), (77), and (79).

Equations (82) and (83) represent an algorithm for backward recursive smoothing, in the case of future-form models, to be applied for t = N - 1, ..., 1; the modifications for the limit case t = N follow accordingly from Equations (54) and (56), using  $Z_N^{-1}$ , as provided by the forward filter.

# 5.3. Interpretation of the Filters and Smoothers

The next required task is interpreting the filters and smoothers that were derived in the previous section. By construction, the output of the backward smoothing recursions, simply denoted by  $\mathbf{x}_t$  in Equations (73) and (82), represents the smoothed state estimates,  $\mathbf{x}_{t|N}$ , and from Equation (67), it is also clear that the recursion of Equations (74) and (83) describes the covariance matrix of the smoothed state estimates. However, the meaning of the intermediate quantities  $Z_t$  and  $\mathbf{z}_t$ , to be computed by Equations (71), (72), (80), and (81), is not immediately obvious. In the case of non-correlated noise terms,  $\mathbf{z}_t$  would directly correspond to the filtered information state,  $\chi_{t|t}$ , but in the general case, this is no longer true.

The condition for interpreting the quantities  $Z_t$  and  $z_t$  is that the known information filters, as reviewed above in Section 3, be reproduced. This condition leads to the following interpretations: for the contemporaneous form,

$$Z_{t} = \begin{cases} \mathsf{F}_{t|t} + \mathsf{A}_{t+1}^{\mathsf{T}} \mathbf{\Omega}_{t+1} \mathsf{A}_{t+1} & \text{if } t = 1, \dots, N-1 \\ \mathsf{F}_{t|t} & \text{if } t = N \end{cases}$$
(84)

$$\mathbf{z}_{t} = \begin{cases} \boldsymbol{\chi}_{t|t} + \mathsf{A}_{t+1}^{\mathsf{T}} \boldsymbol{\Sigma}_{t+1} \mathbf{y}_{t+1} & \text{if } t = 1, \dots, N-1 \\ \boldsymbol{\chi}_{t|t} & \text{if } t = N \end{cases}$$
(85)

and for the future form,

$$\mathsf{Z}_{t} = \begin{cases} \mathsf{F}_{t|t} + \tilde{\mathsf{A}}_{t+1}^{\mathsf{T}} \mathbf{\Omega}_{t} \tilde{\mathsf{A}}_{t+1} & \text{if } t = 1, \dots, N-1 \\ \mathsf{F}_{t|t} & \text{if } t = N \end{cases}$$
(86)

$$\mathbf{z}_{t} = \begin{cases} \boldsymbol{\chi}_{t|t} + \tilde{\mathsf{A}}_{t+1}^{\mathsf{T}} \boldsymbol{\Sigma}_{t} \mathbf{y}_{t} & \text{if } t = 1, \dots, N-1 \\ \boldsymbol{\chi}_{t|t} & \text{if } t = N \end{cases}$$
(87)

Equation (84) has already been given by Fahrmeir and Kaufmann [11], although only for the non-correlated case,  $S_t = 0$ .

Then, in the case of the contemporaneous form, the information filter becomes, for t = 2, ..., N,

$$\boldsymbol{\chi}_{t|t} = \left(\mathsf{C}_{t}^{\mathsf{T}}\boldsymbol{\Psi}_{t} - \boldsymbol{\Sigma}_{t}\right)\mathbf{y}_{t} - \left(\mathsf{A}_{t}^{\mathsf{T}}\boldsymbol{\Sigma}_{t}\mathsf{C}_{t} - \mathsf{A}_{t}^{\mathsf{T}}\boldsymbol{\Omega}_{t}\right)^{\mathsf{T}}\left(\mathsf{F}_{t-1|t-1} + \mathsf{A}_{t}^{\mathsf{T}}\boldsymbol{\Omega}_{t}\mathsf{A}_{t}\right)^{-1}\left(\boldsymbol{\chi}_{t-1|t-1} + \mathsf{A}_{t}^{\mathsf{T}}\boldsymbol{\Sigma}_{t}\mathbf{y}_{t}\right) \quad (88)$$

$$\mathsf{F}_{t|t} = \mathbf{\Omega}_t - \mathbf{\Sigma}_t \mathsf{C}_t - (\mathbf{\Sigma}_t \mathsf{C}_t)^{\mathsf{T}} + \mathsf{C}_t^{\mathsf{T}} \mathbf{\Psi}_t \mathsf{C}_t - \left(\mathsf{A}_t^{\mathsf{T}} \mathbf{\Sigma}_t \mathsf{C}_t - \mathsf{A}_t^{\mathsf{T}} \mathbf{\Omega}_t\right)^{\mathsf{T}} \left(\mathsf{F}_{t-1|t-1} + \mathsf{A}_t^{\mathsf{T}} \mathbf{\Omega}_t \mathsf{A}_t\right)^{-1} \left(\mathsf{A}_t^{\mathsf{T}} \mathbf{\Sigma}_t \mathsf{C}_t - \mathsf{A}_t^{\mathsf{T}} \mathbf{\Omega}_t\right)$$
(89)

and the smoother becomes, for t = N - 1, ..., 1,

$$\mathbf{x}_{t|N} = (\mathbf{F}_{t|t} + \mathbf{A}_{t+1}^{\mathsf{T}} \mathbf{\Omega}_{t+1} \mathbf{A}_{t+1})^{-1} \times \left( \chi_{t|t} + \mathbf{A}_{t+1}^{\mathsf{T}} \mathbf{\Sigma}_{t+1} \mathbf{y}_{t+1} - (\mathbf{A}_{t+1}^{\mathsf{T}} \mathbf{\Sigma}_{t+1} \mathbf{C}_{t+1} - \mathbf{A}_{t+1}^{\mathsf{T}} \mathbf{\Omega}_{t+1}) \mathbf{x}_{t+1|N} \right)$$
(90)  

$$\mathbf{P}_{t|N} = (\mathbf{F}_{t|t} + \mathbf{A}_{t+1}^{\mathsf{T}} \mathbf{\Omega}_{t+1} \mathbf{A}_{t+1})^{-1} + (\mathbf{F}_{t|t} + \mathbf{A}_{t+1}^{\mathsf{T}} \mathbf{\Omega}_{t+1} \mathbf{A}_{t+1})^{-1} \times \left( \mathbf{A}_{t+1}^{\mathsf{T}} \mathbf{\Sigma}_{t+1} \mathbf{C}_{t+1} - \mathbf{A}_{t+1}^{\mathsf{T}} \mathbf{\Omega}_{t+1}) \mathbf{P}_{t+1|N} (\mathbf{A}_{t+1}^{\mathsf{T}} \mathbf{\Sigma}_{t+1} \mathbf{C}_{t+1} - \mathbf{A}_{t+1}^{\mathsf{T}} \mathbf{\Omega}_{t+1})^{\mathsf{T}} \times \left( \mathbf{F}_{t|t} + \mathbf{A}_{t+1}^{\mathsf{T}} \mathbf{\Omega}_{t+1} \mathbf{A}_{t+1} \right)^{-1}$$
(91)

In the case of future form, the information filter becomes, for t = 2, ..., N,

$$\chi_{t|t} = C_t^{\mathsf{T}} \boldsymbol{\Phi}_t \mathbf{y}_t - \boldsymbol{\Sigma}_{t-1} \mathbf{y}_{t-1} + \left( C_{t-1}^{\mathsf{T}} \boldsymbol{\Sigma}_{t-1}^{\mathsf{T}} + A_t^{\mathsf{T}} \boldsymbol{\Omega}_{t-1} \right)^{\mathsf{T}} \left( \mathsf{F}_{t-1|t-1} + \tilde{A}_t^{\mathsf{T}} \boldsymbol{\Omega}_{t-1} \tilde{A}_t \right)^{-1} \left( \chi_{t-1|t-1} + \tilde{A}_t^{\mathsf{T}} \boldsymbol{\Sigma}_{t-1} \mathbf{y}_{t-1} \right)$$
(92)  
$$\mathsf{F}_{t|t} = \boldsymbol{\Omega}_{t-1} + C_t^{\mathsf{T}} \boldsymbol{\Phi}_t C_t$$

$$-\left(\mathsf{C}_{t-1}^{\mathsf{T}}\boldsymbol{\Sigma}_{t-1}^{\mathsf{T}} + \mathsf{A}_{t}^{\mathsf{T}}\boldsymbol{\Omega}_{t-1}\right)^{\mathsf{T}}\left(\mathsf{F}_{t-1|t-1} + \tilde{\mathsf{A}}_{t}^{\mathsf{T}}\boldsymbol{\Omega}_{t-1}\tilde{\mathsf{A}}_{t}\right)^{-1}\left(\mathsf{C}_{t-1}^{\mathsf{T}}\boldsymbol{\Sigma}_{t-1}^{\mathsf{T}} + \mathsf{A}_{t}^{\mathsf{T}}\boldsymbol{\Omega}_{t-1}\right) \quad (93)$$

and the smoother becomes, for t = N - 1, ..., 1,

$$\mathbf{x}_{t|N} = \left(\mathsf{F}_{t|t} + \tilde{\mathsf{A}}_{t+1}^{\mathsf{T}} \mathbf{\Omega}_{t} \tilde{\mathsf{A}}_{t+1}\right)^{-1} \left(\boldsymbol{\chi}_{t|t} + \tilde{\mathsf{A}}_{t+1}^{\mathsf{T}} \boldsymbol{\Sigma}_{t} \mathbf{y}_{t} - \left(\mathsf{C}_{t}^{\mathsf{T}} \boldsymbol{\Sigma}_{t}^{\mathsf{T}} - \tilde{\mathsf{A}}_{t+1}^{\mathsf{T}} \mathbf{\Omega}_{t}\right) \mathbf{x}_{t+1|N}\right)$$
(94)

$$\mathsf{P}_{t|N} = \left(\mathsf{F}_{t|t} + \mathsf{A}_{t+1}^{\mathsf{T}} \mathbf{\Omega}_{t} \mathsf{A}_{t+1}\right)^{-1} + \left(\mathsf{F}_{t|t} + \mathsf{A}_{t+1}^{\mathsf{T}} \mathbf{\Omega}_{t} \mathsf{A}_{t+1}\right)^{-1} \times \left(\mathsf{C}_{t}^{\mathsf{T}} \boldsymbol{\Sigma}_{t}^{\mathsf{T}} + \mathsf{A}_{t+1}^{\mathsf{T}} \mathbf{\Omega}_{t}\right) \mathsf{P}_{t+1|N} \left(\mathsf{C}_{t}^{\mathsf{T}} \boldsymbol{\Sigma}_{t}^{\mathsf{T}} + \mathsf{A}_{t+1}^{\mathsf{T}} \mathbf{\Omega}_{t}\right)^{\mathsf{T}} \left(\mathsf{F}_{t|t} + \tilde{\mathsf{A}}_{t+1}^{\mathsf{T}} \mathbf{\Omega}_{t} \tilde{\mathsf{A}}_{t+1}\right)^{-1}$$
(95)

The modifications for the limit case t = 1 follow accordingly from Equations (50), (51), (77), and (79); the case t = N requires no modifications to the forward filter pass.

It therefore follows that the contemporaneous-form filter in Equations (88) and (89) is equivalent to the filter in Equations (25) and (26) and that the future-form filter in Equations (92) and (93) is equivalent to the filter in Equations (41) and (42), although these equivalences are far from obvious; however, they can be verified either through their application to numerical examples, or—preferably—by explicit algebraic manipulation. As an example of the latter approach, it is shown in the Appendix A how Equation (88) can be transformed into Equation (25). The transformation for the corresponding expressions for the covariance matrix  $P_{t|t}$  could probably be performed in a similar way; the same remark also applies to the expressions for the conditional mean vector and covariance matrix of the future-form filter.

Finally, both the contemporaneous-form smoother algorithm in Equations (90) and (91) and the future-form smoother algorithm in Equations (94) and (95) represent generalisations of the standard RTS smoother algorithm.

### 5.4. The Origin of the Non-Equivalence of the two Kalman Filters

In Section 3, the covariance Kalman and information filter algorithms were reviewed for the case of correlated noise terms, both for the contemporaneous-form and future-form state space models, and it was stated that they are not equivalent. The reason for this lack of equivalence is not hard to find: depending on whether the contemporaneous form or future form is used, the statement that the cross-correlation matrix,  $S_t$ , does not vanish has different meaning. If in the contemporaneous form,  $S_t$  is defined by

$$\mathsf{S}_t = \mathbb{E}(\boldsymbol{\eta}_{\tau}\boldsymbol{\epsilon}_{\tau}^\mathsf{T}),$$

then in future form, it would have to be defined by

$$\mathsf{S}_t = \mathbb{E}(\boldsymbol{\eta}_{\tau-1}\boldsymbol{\epsilon}_{\tau}^\mathsf{T}),$$

and this is precisely the definition for  $S_t$  chosen by Brown and Hwang [22] since these authors use the future form but at the same time present the algorithm in Equations (12)–(17), which in the present paper, is listed as a contemporaneous-form algorithm.

Then, it becomes obvious that actually two cross-correlation matrices should be defined, one for lag-zero (instantaneous) correlation and one for lag-one correlation:

$$\mathsf{S}_t^{(0)} = \mathbb{E}(\pmb{\eta}_{ au} \pmb{\epsilon}_{ au}^\mathsf{T}) , \quad \mathsf{S}_t^{(1)} = \mathbb{E}(\pmb{\eta}_{ au-1} \pmb{\epsilon}_{ au}^\mathsf{T})$$

Swapping the two noise terms in the definition of  $S_t^{(0)}$  would give

$$(\mathsf{S}_1^{(0)})^\mathsf{T} = \mathbb{E}(\boldsymbol{\epsilon}_\tau \boldsymbol{\eta}_\tau^\mathsf{T})$$

However, in the definition of  $S_t^{(1)}$ , the two noise terms may not be swapped since

$$\mathbb{E}(\boldsymbol{\eta}_{\tau-1}\boldsymbol{\epsilon}_{\tau}^{\mathsf{T}}) \neq \mathbb{E}(\boldsymbol{\epsilon}_{\tau-1}\boldsymbol{\eta}_{\tau}^{\mathsf{T}})$$

From the viewpoint of the contemporaneous form, the Kalman filter in Equations (12)-(17) is valid for the case

$$\mathsf{S}_t^{(0)} \neq \mathsf{0} \text{ , } \quad \mathsf{S}_t^{(1)} = \mathsf{0}$$

(with S<sub>t</sub> to be replaced by  $S_t^{(0)}$ ), while the Kalman filter in Equations (27)–(34) is valid for the case S<sup>(0)</sup> (1)

$$S_t^{(0)} = 0$$
 ,  $S_t^{(1)} \neq 0$ 

(with S<sub>t</sub> to be replaced by  $S_t^{(1)}$ ). Naturally, both filters are valid for the case

$$S_t^{(0)} = 0$$
,  $S_t^{(1)} = 0$ .

# 6. Merging the Contemporaneous Form and Future Form

In the discussion of the previous section, one case has not yet been addressed:

$$S_t^{(0)} \neq 0$$
,  $S_t^{(1)} \neq 0$ .

Is there a Kalman filter for this case?

In order to investigate this case, the covariance matrices in Equations (63) and (75) have to be merged as follows:

$$\mathcal{R} = \begin{pmatrix} Q_0 & S_1^{(0)} & 0 & \dots & 0 & 0 & 0\\ (S_1^{(0)})^{\mathsf{T}} & \mathsf{R}_1 & (S_1^{(1)})^{\mathsf{T}} & \dots & 0 & 0 & 0\\ 0 & S_1^{(1)} & Q_1 & \dots & 0 & 0 & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots\\ 0 & 0 & 0 & \dots & \mathsf{R}_{N-1} & (\mathsf{S}_{N-1}^{(1)})^{\mathsf{T}} & 0\\ 0 & 0 & 0 & \dots & \mathsf{S}_{N-1}^{(1)} & \mathsf{Q}_{N-1} & \mathsf{S}_N^{(0)}\\ 0 & 0 & 0 & \dots & 0 & (\mathsf{S}_N^{(0)})^{\mathsf{T}} & \mathsf{R}_N \end{pmatrix}.$$
(96)

This covariance matrix is no longer block-diagonal, and as a consequence, the matrix  $\mathcal{A}^{\mathsf{T}}\mathcal{R}^{-1}\mathcal{A}$  is not a one-block-banded matrix; rather, it will generically be a full matrix, such that its inverse will also be a full matrix. Consequently, no recursive algorithm for optimally solving this state estimation problem can be derived through the approach presented above in Sections 5.1 and 5.2.

However, recursive algorithms for non-optimally solving this problem may exist. As an example, the Kalman filter algorithms in Equations (12)–(17) and Equations (27)–(34) are considered again. Since the generalisations with respect to the standard Kalman filter without correlated noise terms occur in different equations, these algorithms may, in a purely heuristical way, be merged as follows:

$$\tilde{\mathsf{A}}_{t} = \mathsf{A}_{t} - \mathsf{S}_{t-1}^{(1)} \mathbf{\Phi}_{t-1} \mathsf{C}_{t-1}$$
(97)

$$\tilde{\mathsf{Q}}_t = \mathsf{Q}_t - \mathsf{S}_t^{(1)} \boldsymbol{\Phi}_t \big(\mathsf{S}_t^{(1)}\big)^\mathsf{T}$$
(98)

$$\mathbf{x}_{t|t-1} = \tilde{A}_t \mathbf{x}_{t-1|t-1} + S_{t-1}^{(1)} \mathbf{\Phi}_{t-1} \mathbf{y}_{t-1}$$
(99)

$$\mathsf{P}_{t|t-1} = \tilde{\mathsf{A}}_t \mathsf{P}_{t-1|t-1} \tilde{\mathsf{A}}_t^\mathsf{T} + \tilde{\mathsf{Q}}_{t-1}$$
(100)

$$\mathbf{V}_{t} = \mathbf{C}_{t} \mathbf{P}_{t|t-1} \mathbf{C}_{t}^{\mathsf{T}} + \mathbf{R}_{t} + \mathbf{C}_{t} \mathbf{S}_{t}^{(0)} + \left(\mathbf{C}_{t} \mathbf{S}_{t}^{(0)}\right)^{\mathsf{T}}$$
(101)

$$\mathsf{K}_{t} = \left(\mathsf{P}_{t|t-1}\mathsf{C}_{t}^{\mathsf{T}} + \mathsf{S}_{t}^{(0)}\right)\mathsf{V}_{t}^{-1} \tag{102}$$

$$\mathbf{x}_{t|t} = \mathbf{x}_{t|t-1} + \mathsf{K}_t(\mathbf{y}_t - \mathsf{C}_t \mathbf{x}_{t|t-1})$$
(103)

$$\mathsf{P}_{t|t} = \mathsf{P}_{t|t-1} - \mathsf{K}_t \mathsf{V}_t \mathsf{K}_t^{\mathsf{I}} \tag{104}$$

It is less obvious how the smoother algorithms for the contemporaneous form and for the future form may be merged, and in this paper, this problem will not be considered further.

Here, it should be mentioned that recently Linares-Pérez and coauthors [19] also presented a generalised recursive filter algorithm for state estimation, which contains the Kalman filters in Equations (12)–(17) and Equations (27)–(34) as special cases. However, for the case  $S_t^{(0)} \neq 0$ ,  $S_t^{(1)} \neq 0$ , their recursive filter differs from the generalised Kalman filter in Equations (97)–(104). The exact relationship between the two filter algorithms still needs to be investigated.

# 7. Application Example: Simulated Data

Now, an example for the application of the generalised Kalman filter in Equations (97)–(104) to simulated data will be presented. The state and data dimensions are chosen as m = n = 1, and the model parameters are chosen as

$$\begin{aligned} \mathsf{A}_t &\equiv a = 0.95 , & \mathsf{C}_t &\equiv c = 1.0 , \\ \mathsf{Q}_t &\equiv q = 1.0 , & \mathsf{R}_t &\equiv r = 1.0 , \\ \mathsf{S}_t^{(0)} &\equiv s^{(0)} = 0.75 , & \mathsf{S}_t^{(1)} &\equiv s^{(1)} &= -0.25 , \end{aligned}$$

i.e., the parameters do not depend on time. From a state space model employing these parameters, 1000 simulated time series are created, each of length N = 1024 points, using different sets of dynamical and observation noise values. The noise values are created such that they approximately fulfil the following constraints:

$$\mathbb{E}\left(\begin{bmatrix}\eta_{\tau}\\\epsilon_{\tau}\end{bmatrix}\begin{bmatrix}\eta_{\tau}\\\epsilon_{\tau}\end{bmatrix}^{\mathsf{T}}\right) = \begin{pmatrix}q & s^{(0)}\\s^{(0)} & r\end{pmatrix} \quad \text{and} \quad \mathbb{E}\left(\begin{bmatrix}\eta_{\tau-1}\\\epsilon_{\tau-1}\end{bmatrix}\begin{bmatrix}\eta_{\tau}\\\epsilon_{\tau}\end{bmatrix}^{\mathsf{T}}\right) = \begin{pmatrix}0 & s^{(1)}\\0 & 0\end{pmatrix}$$

using the chosen values for q, r,  $s^{(0)}$ , and  $s^{(1)}$ , as given above.

Since this is a simulation, the true states are known. Then, the same state space model that was used to create the simulated time series is used to estimate the states, either using the weighted least-squares approach, i.e., solving Equation (65) non-recursively, or the generalised Kalman filter algorithm in Equations (97)–(104). While the former algorithm provides smoothed state estimates, the latter provides filtered state estimates, but nevertheless, these two sets of estimates shall be compared.

In the particular setting of this simulation, state estimation is equivalent to noise reduction with respect to the observational noise term. Therefore, the performance of the two algorithms for state estimation may be compared by quantifying the achieved level of noise reduction; this is possible since the true states are known. The result is shown via histograms in Figure 1. Note that a value of 0 dB on the horizontal axis corresponds to failure in noise reduction.

In the figure, it can be seen that for the 1000 time series created, both algorithms achieve values for noise reduction between 5 dB and 7 dB; if the distributions are fitted by Gaussians, for the weighted least-squares algorithm, a mean of 6.3234 and a standard deviation of 0.1624 are obtained, while for the generalised Kalman filter algorithm, a mean of 5.8242 and a standard deviation of 0.2079 are obtained. Thus, at least in this simulation example, the generalised Kalman filter achieves almost as good performance as the weighted least-squares algorithm, despite representing a non-optimal algorithm for state estimation.



**Figure 1.** Histogram of reduction in variance in observational noise for 1000 time series that were created from a simulated state space model and analysed by estimating the states using the non-recursive weighted least-squares approach in Equation (65) (smoothed state estimates: red colour) and by estimating the states using the generalised Kalman filter in Equations (97)–(104) (filtered state estimates: green colour)

# 8. Application Example: Real-World Data

Next, an example of the application of the generalised Kalman filter in Equations (97)–(104) to a real-world dataset is presented. A hydroacoustic recording is chosen that was obtained in the Baltic Sea by a floating hydrophone while a ship was passing at a distance of a few hundred metres. The original sampling rate was 20 kHz, but the data have been subsampled to 5 kHz. From a much longer time series, a short window of 0.4096 s in length is selected, corresponding to 2048 samples. The selected data are transformed into a mean of zero and a variance of one. Within the selected window, the underwater sound emission of the passing ship is dominated by a single line at a frequency of approx. 498.5 Hz. The aim of the analysis is to extract this line from background noise and other signal components.

The data dimension is n = 1, while the state dimension is chosen as m = 2. Since it is intended to model a single stochastic oscillation, the structure of the state space model is chosen such that it corresponds to an autoregressive moving-average (ARMA) model with model orders p = 2 and q = 1:

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + b_1 \eta_{t-1} + \eta_t.$$

From this model, the state transition matrix and the dynamical noise covariance matrix of the state space model, in *observer canonical form*, are as follows:

$$\mathsf{A} = \begin{pmatrix} a_1 & 1 \\ a_2 & 0 \end{pmatrix} \quad , \quad \mathsf{Q} = \begin{pmatrix} 1 & b_1 \\ b_1 & b_1^2 + \beta \end{pmatrix}.$$

In a pure ARMA model, the additional parameter  $\beta$  should be zero; however, in order to avoid a singular covariance matrix, Q, it is necessary to allow this parameter to be non-zero; it is constrained to be positive.

All the model parameters are fitted to the selected data through numerical maximisation of the innovation likelihood. The innovation likelihood is computed with the generalised Kalman filter in Equations (97)–(104). The resulting estimates of the model parameters are

$a_1 = 1.6180$ ,	$a_2 = -0.9974$ ,
C = (-0.0099, 0),	$b_1 = -0.0013$ ,
eta=0.0020 ,	r = 0.9329 ,
$S^{(0)} = (0.7056, 0)^{T}$ ,	$S^{(1)} = (-0.2599, 0)$

The estimates of  $a_1$  and  $a_2$  correspond to a frequency of 498.5751 Hz. Also, an estimate of the initial state,  $\mathbf{x}_{1|0}$ , is obtained by maximising the innovation likelihood. It should be

noted that these estimates of the model parameters may represent only a local maximum of the innovation likelihood.

Then, the optimised state space model is employed to estimate the states that correspond to the selected data, using either the non-recursive weighted least-squares approach of solving Equation (65) directly or the generalised Kalman filter in Equations (97)–(104). While the first approach yields smoothed estimates, the latter yields filtered estimates. Both time series of state estimates are displayed in Figure 2, either versus time or as a perpendicular scatter plot. At each time point, the state is a two-dimensional vector, but only the first component is displayed.



**Figure 2.** Left panel: state estimates vs. time for the first 100 samples of the hydroacoustic time series: smoothed state estimates obtained by the non-recursive weighted least-squares approach in Equation (65) (lines and symbols in red colour) and filtered state estimates obtained by the generalised Kalman filter in Equations (97)–(104) (lines and symbols in green colour). Right panel: perpendicular scatter plot of the smoothed state estimates vs. the filtered state estimates for all 2048 samples of the time series (symbols in blue colour).

In the figure, it can be seen that both time series of state estimates are very similar. The smoothed state estimates obtained by the non-recursive approach (denoted by lines and symbols in red colour) represent the correct estimates by definition. Differences between the two time series of state estimates arise for two reasons: first, smoothed estimates will always differ somewhat from filtered estimates, and second, as has been shown above, the estimates obtained by the non-recursive approach represent only an non-optimal approximation of the correct estimates. The numerical results shown in the figure illustrate the good quality achieved by this approximation.

### 9. Discussion and Conclusions

This paper examines the problem of state estimation in linear state space models when non-zero cross-correlations exist between the dynamical noise and observation noise terms. Lag-zero and lag-one cross-correlations are distinguished, and it is highlighted that it depends on which form is chosen for the state space model (the contemporaneous form or the future form) whether a given cross-correlation matrix refers to lag-zero or lag-one cross-correlations. As previously noted by Harvey [2], the existence of these two forms leads to two distinct, non-equivalent Kalman filtering algorithms for the case of correlated noise.

The main contributions of this paper are as follows: First, the weighted least-squares framework, as developed by Duncan and Horn [9] for the problem of state estimation in linear state space models, has been generalised to the case of correlated noise terms, thereby obtaining Kalman filter and smoother algorithms, both for the contemporaneous form and the future form. While the resulting generalised Kalman filter algorithms coincide with already known information filter algorithms, the resulting generalised smoother algorithms may be of wider interest, as in the available literature, smoothing in the case of correlated noise terms is rarely discussed. In order to obtain a recursion for the covariance of the

smoothed state estimates, the original recursive inversion algorithm by Bell was augmented by two additional expressions, given by Equations (56) and (57).

Second, it has been demonstrated that Bell's recursive algorithm for inverting oneblock-banded matrices via blockwise LDL decomposition, as reviewed in Section 4, provides a general canonical structure for information filters and RTS smoothers. This structure differs considerably from the structure that is usually chosen; see, e.g., the difference between Equations (25) and (26) and Equations (88) and (89). As shown in the Appendix A, by suitable algebraic transformations Equation (88) can be transformed into Equation (25), and without doubt, the same holds true for the other pairs of corresponding equations for filtered state estimates and their covariance matrices. Actually, in order to prove the equivalence of these pairs of equations, such explicit algebraic transformations are not necessary since the equivalence is already proven by the fact that each pair of equations has been derived from the same underlying state space model. Therefore there is no need to develop a formal proof. It is possible that the equivalence of Equation (26) and Equation (89) has already been established in the theory of Riccati equations, but the author has not verified this.

Third, a generalisation to the case where both lag-zero and lag-one cross-correlations are present simultaneously has been proposed, building upon the previously developed theory for lag-zero and lag-one cross-correlations of the noise terms in state space models. This generalisation offers the potential to merge the filter algorithms for the contemporaneous form and the future form. Strictly speaking, as it has been shown by using the weighted least-squares approach, in this case, no recursive algorithm for optimally solving the corresponding state estimation problem can be derived. Nevertheless, an optimal solution to the state estimation problem can still be obtained through the non-recursive approach of directly solving the weighted least-squares problem. This approach is applicable to any lagged cross-correlation structure for the noise terms that might be chosen. It is important to note, however, that the computational cost of this non-recursive algorithm, in terms of time and memory, scales significantly with the length of the time series being modelled.

For this reason, an approximative recursive algorithm has been proposed that was obtained by merging the recursive Kalman filter algorithms for the contemporaneous-form and future-form state space models; no smoother backward pass was included in this algorithm. The resulting generalised Kalman filter algorithm was obtained using a purely heuristical approach. Nevertheless, it has been demonstrated that, both in a simulation study and in an example of the analysis of real-world data, this approximative recursive algorithm performs almost as well as the optimal non-recursive algorithm.

A further potential benefit of formulating state estimation problems as weighted least-squares problems is given by the possibility of imposing constraints and applying regularisation. If one is willing to perform state estimation using non-recursive algorithms, as suggested by Chan and Jeliazkov [10], the considerable repertoire of methods and experience that have been accumulated in the least-squares field can be utilised.

Future research directions in this field include the design of approximate recursive smoothers for the case of both lag-zero and lag-one cross-correlations being present simultaneously. Additionally, the development of square-root variants of the information filter and RTS smoother algorithms presented in this paper could be explored. Preliminary experience with modelling real-world data sets using state space models with correlated noise terms indicates that the covariance matrix of the filtered state estimate, given by Equation (104), exhibits a propensity to lose the property of positive definiteness, potentially hindering parameter estimation via maximum likelihood methods; this is a well-known potential effect of forming differences in matrices [3]. The same problem may arise with the definition of the modified dynamical noise covariance matrix for the future-form Kalman filter; see Equation (28). Square-root filtering algorithms would provide a convenient solution to these kinds of numerical problems.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The simulated data will be provided by the author upon request.

**Acknowledgments:** This work has benefitted significantly from discussions with Mohinder Grewal. The comments of two anonymous reviewers have contributed to improving the paper.

Conflicts of Interest: The author declares no conflicts of interest.

### Abbreviations

The following abbreviations are used in this manuscript:

- RTS Rauch-Tung-Striebel
- LDL (not actually an abbreviation but a type of matrix decomposition)

### Appendix A

In this appendix, it is shown that Equation (25) can be derived from Equation (88). For this purpose, first, two relationships need to be stated that follow from the following well-known *matrix inversion lemma*: Let A denote an invertible square matrix of dimensions  $(m \times m)$ , let B denote a matrix of dimensions  $(m \times n)$ , and let C denote an invertible square matrix of dimensions  $(n \times n)$ ; then, it holds true that

$$(A \pm BCB^{\mathsf{T}})^{-1} = A^{-1} \mp A^{-1}B(C^{-1} \pm B^{\mathsf{T}}A^{-1}B)^{-1}B^{\mathsf{T}}A^{-1}$$
(A1)

and

$$(A \pm BCB^{\mathsf{T}})^{-1}BC = A^{-1}B(C^{-1} \pm B^{\mathsf{T}}A^{-1}B)^{-1}.$$
 (A2)

After rearranging, Equation (88) consists of two terms, one containing  $\chi_{t-1|t-1}$  and the other containing  $y_t$ :

$$\chi_{t|t} = (\boldsymbol{\Omega}_{t}\boldsymbol{A}_{t} - \boldsymbol{C}_{t}^{\mathsf{T}}\boldsymbol{\Sigma}_{t}^{\mathsf{T}}\boldsymbol{A}_{t})(\boldsymbol{F}_{t-1|t-1} + \boldsymbol{A}_{t}^{\mathsf{T}}\boldsymbol{\Omega}_{t}\boldsymbol{A}_{t})^{-1}\boldsymbol{\chi}_{t-1|t-1} + \left((\boldsymbol{C}_{t}^{\mathsf{T}}\boldsymbol{\Psi}_{t} - \boldsymbol{\Sigma}_{t}) + (\boldsymbol{\Omega}_{t}\boldsymbol{A}_{t} - \boldsymbol{C}_{t}^{\mathsf{T}}\boldsymbol{\Sigma}_{t}^{\mathsf{T}}\boldsymbol{A}_{t})(\boldsymbol{F}_{t-1|t-1} + \boldsymbol{A}_{t}^{\mathsf{T}}\boldsymbol{\Omega}_{t}\boldsymbol{A}_{t})^{-1}\boldsymbol{A}_{t}^{\mathsf{T}}\boldsymbol{\Sigma}_{t}\right)\boldsymbol{y}_{t}.$$
(A3)

Equation (25) has the same structure:

$$\boldsymbol{\chi}_{t|t} = (\mathbf{I}^{(m)} + \mathbf{G}_t^{\mathsf{T}} \mathbf{\Gamma}_t \mathbf{S}_t^{\mathsf{T}}) \boldsymbol{\Theta}_t \mathbf{A}_t (\mathbf{F}_{t-1|t-1} + \mathbf{A}_t^{\mathsf{T}} \boldsymbol{\Theta}_t \mathbf{A}_t)^{-1} \boldsymbol{\chi}_{t-1|t-1} + \mathbf{G}_t^{\mathsf{T}} \mathbf{\Gamma}_t \mathbf{y}_t,$$
(A4)

so the coefficient matrices of these terms can be studied separately. Note that

$$\mathbf{\Omega}_t = \mathbf{\Theta}_t + \mathbf{\Theta}_t \mathbf{S}_t \mathbf{\Psi}_t \mathbf{S}_t^{\mathsf{T}} \mathbf{\Theta}_t, \tag{A5}$$

where Equations (8), (9), and (A1) have been used. It follows that

$$\left(\mathsf{F}_{t-1|t-1} + \mathsf{A}_t^{\mathsf{T}} \mathbf{\Omega}_t \mathsf{A}_t\right)^{-1} = \left(\mathsf{F}_{t-1|t-1} + \mathsf{A}_t^{\mathsf{T}} \mathbf{\Theta}_t \mathsf{A}_t + \mathsf{A}_t^{\mathsf{T}} \mathbf{\Theta}_t \mathsf{S}_t^{\mathsf{T}} \mathbf{\Theta}_t \mathsf{A}_t\right)^{-1} \tag{A6}$$

$$= \left(\mathsf{B}_t + \mathsf{A}_t^{\mathsf{T}} \mathbf{\Theta}_t \mathsf{S}_t \Psi_t \mathsf{S}_t^{\mathsf{T}} \mathbf{\Theta}_t \mathsf{A}_t\right)^{-1} \tag{A7}$$

$$=\mathsf{B}_{t}^{-1}-\mathsf{B}_{t}^{-1}\mathsf{A}_{t}^{\mathsf{T}}\boldsymbol{\Theta}_{t}\mathsf{S}_{t}(\boldsymbol{\Psi}_{t}^{-1}+\mathsf{S}_{t}^{\mathsf{T}}\boldsymbol{\Theta}_{t}\mathsf{A}_{t}\mathsf{B}_{t}^{-1}\mathsf{A}_{t}^{\mathsf{T}}\boldsymbol{\Theta}_{t}\mathsf{S}_{t})^{-1}\times$$

 $\mathsf{S}_t^{\mathsf{T}} \boldsymbol{\Theta}_t \mathsf{A}_t \mathsf{B}_t^{-1}, \qquad (\mathrm{A8})$ 

where Equations (18) and (A1) have been used.

Then, the coefficient matrix of  $\chi_{t-1|t-1}$  in Equation (A3) can be transformed as follows:

$$(\mathbf{\Omega}_{t}\mathsf{A}_{t} - \mathsf{C}_{t}^{\mathsf{T}}\mathbf{\Sigma}_{t}^{\mathsf{T}}\mathsf{A}_{t}) (\mathsf{F}_{t-1|t-1} + \mathsf{A}_{t}^{\mathsf{T}}\mathbf{\Omega}_{t}\mathsf{A}_{t})^{-1}$$

$$= (\mathbf{\Theta}_{t}\mathsf{A}_{t} + \mathbf{\Theta}_{t}\mathsf{S}_{t}\mathbf{\Psi}_{t}\mathsf{S}_{t}^{\mathsf{T}}\mathbf{\Theta}_{t}\mathsf{A}_{t} - \mathsf{C}_{t}^{\mathsf{T}}\mathbf{\Sigma}_{t}^{\mathsf{T}}\mathsf{A}_{t}) (\mathsf{B}_{t} + \mathsf{A}_{t}^{\mathsf{T}}\mathbf{\Theta}_{t}\mathsf{S}_{t}\mathbf{\Psi}_{t}\mathsf{S}_{t}^{\mathsf{T}}\mathbf{\Theta}_{t}\mathsf{A}_{t})^{-1}$$
(A9)

by using Equations (A5) and (A7),

$$= (\boldsymbol{\Theta}_{t}\boldsymbol{A}_{t} + \boldsymbol{\Theta}_{t}\boldsymbol{S}_{t}\boldsymbol{\Psi}_{t}\boldsymbol{S}_{t}^{\mathsf{T}}\boldsymbol{\Theta}_{t}\boldsymbol{A}_{t} + \boldsymbol{C}_{t}^{\mathsf{T}}\boldsymbol{\Psi}_{t}\boldsymbol{S}_{t}^{\mathsf{T}}\boldsymbol{\Theta}_{t}\boldsymbol{A}_{t})(\boldsymbol{B}_{t} + \boldsymbol{A}_{t}^{\mathsf{T}}\boldsymbol{\Theta}_{t}\boldsymbol{S}_{t}\boldsymbol{\Psi}_{t}\boldsymbol{S}_{t}^{\mathsf{T}}\boldsymbol{\Theta}_{t}\boldsymbol{A}_{t})^{-1}$$
(A10)

by using Equation (10),

$$= (C_t^{\mathsf{T}} + \Theta_t S_t) \Psi_t S_t^{\mathsf{T}} \Theta_t A_t (B_t + A_t^{\mathsf{T}} \Theta_t S_t \Psi_t S_t^{\mathsf{T}} \Theta_t A_t)^{-1} + \Theta_t A_t (B_t + A_t^{\mathsf{T}} \Theta_t S_t \Psi_t S_t^{\mathsf{T}} \Theta_t A_t)^{-1}$$
(A11)  
$$= (C_t^{\mathsf{T}} + \Theta_t S_t) (\Psi_t^{-1} + S_t^{\mathsf{T}} \Theta_t A_t B_t^{-1} A_t^{\mathsf{T}} \Theta_t S_t)^{-1} S_t^{\mathsf{T}} \Theta_t A_t B_t^{-1} + \Theta_t A_t (B_t + A_t^{\mathsf{T}} \Theta_t S_t \Psi_t S_t^{\mathsf{T}} \Theta_t A_t)^{-1}$$
(A12)

by using Equation (A2),

$$= (\mathsf{C}_{t}^{\mathsf{T}} + \boldsymbol{\Theta}_{t}\mathsf{S}_{t})(\boldsymbol{\Psi}_{t}^{-1} + \mathsf{S}_{t}^{\mathsf{T}}\boldsymbol{\Theta}_{t}\mathsf{A}_{t}\mathsf{B}_{t}^{-1}\mathsf{A}_{t}^{\mathsf{T}}\boldsymbol{\Theta}_{t}\mathsf{S}_{t})^{-1}\mathsf{S}_{t}^{\mathsf{T}}\boldsymbol{\Theta}_{t}\mathsf{A}_{t}\mathsf{B}_{t}^{-1} + \boldsymbol{\Theta}_{t}\mathsf{A}_{t}\mathsf{B}_{t}^{-1} - \boldsymbol{\Theta}_{t}\mathsf{A}_{t}\mathsf{B}_{t}^{-1}\mathsf{A}_{t}^{\mathsf{T}}\boldsymbol{\Theta}_{t}\mathsf{S}_{t}(\boldsymbol{\Psi}_{t}^{-1} + \mathsf{S}_{t}^{\mathsf{T}}\boldsymbol{\Theta}_{t}\mathsf{A}_{t}\mathsf{B}_{t}^{-1}\mathsf{A}_{t}^{\mathsf{T}}\boldsymbol{\Theta}_{t}\mathsf{S}_{t})^{-1}\mathsf{S}_{t}^{\mathsf{T}}\boldsymbol{\Theta}_{t}\mathsf{A}_{t}\mathsf{B}_{t}^{-1}$$
(A13)

by using Equation (A8),

$$= \boldsymbol{\Theta}_{t} \mathsf{A}_{t} \mathsf{B}_{t}^{-1} + (\mathsf{C}_{t}^{\mathsf{T}} + \boldsymbol{\Theta}_{t} \mathsf{S}_{t} - \boldsymbol{\Theta}_{t} \mathsf{A}_{t} \mathsf{B}_{t}^{-1} \mathsf{A}_{t}^{\mathsf{T}} \boldsymbol{\Theta}_{t} \mathsf{S}_{t}) \times (\boldsymbol{\Psi}_{t}^{-1} + \mathsf{S}_{t}^{\mathsf{T}} \boldsymbol{\Theta}_{t} \mathsf{A}_{t} \mathsf{B}_{t}^{-1} \mathsf{A}_{t}^{\mathsf{T}} \boldsymbol{\Theta}_{t} \mathsf{S}_{t})^{-1} \mathsf{S}_{t}^{\mathsf{T}} \boldsymbol{\Theta}_{t} \mathsf{A}_{t} \mathsf{B}_{t}^{-1}$$
(A14)

$$= \boldsymbol{\Theta}_{t} \boldsymbol{A}_{t} \boldsymbol{B}_{t}^{-1} + \left(\boldsymbol{C}_{t} + \boldsymbol{S}_{t}^{\mathsf{T}} \boldsymbol{\Theta}_{t} - \boldsymbol{S}_{t}^{\mathsf{T}} \boldsymbol{\Theta}_{t} \boldsymbol{A}_{t} \boldsymbol{B}_{t}^{-1} \boldsymbol{A}_{t}^{\mathsf{T}} \boldsymbol{\Theta}_{t}\right)^{\mathsf{T}} \times \left(\boldsymbol{\Psi}_{t}^{-1} + \boldsymbol{S}_{t}^{\mathsf{T}} \boldsymbol{\Theta}_{t} \boldsymbol{A}_{t} \boldsymbol{B}_{t}^{-1} \boldsymbol{A}_{t}^{\mathsf{T}} \boldsymbol{\Theta}_{t} \boldsymbol{S}_{t}\right)^{-1} \boldsymbol{S}_{t}^{\mathsf{T}} \boldsymbol{\Theta}_{t} \boldsymbol{A}_{t} \boldsymbol{B}_{t}^{-1} \qquad (A15)$$

$$= \boldsymbol{\Theta}_{t} \mathsf{A}_{t} \mathsf{B}_{t}^{-1} + \left(\mathsf{C}_{t} + \mathsf{S}_{t}^{\mathsf{T}} \boldsymbol{\Theta}_{t} - \mathsf{S}_{t}^{\mathsf{T}} \boldsymbol{\Theta}_{t} \mathsf{A}_{t} \mathsf{B}_{t}^{-1} \mathsf{A}_{t}^{\mathsf{T}} \boldsymbol{\Theta}_{t}\right)^{\mathsf{T}} \times \left(\mathsf{R}_{t} - \mathsf{S}_{t}^{\mathsf{T}} \boldsymbol{\Theta}_{t} \mathsf{S}_{t} + \mathsf{S}_{t}^{\mathsf{T}} \boldsymbol{\Theta}_{t} \mathsf{A}_{t} \mathsf{B}_{t}^{-1} \mathsf{A}_{t}^{\mathsf{T}} \boldsymbol{\Theta}_{t} \mathsf{S}_{t}\right)^{-1} \mathsf{S}_{t}^{\mathsf{T}} \boldsymbol{\Theta}_{t} \mathsf{A}_{t} \mathsf{B}_{t}^{-1} \qquad (A16)$$

by using Equation (9),

$$= \boldsymbol{\Theta}_{t} \mathsf{A}_{t} \mathsf{B}_{t}^{-1} + \left(\mathsf{C}_{t} + \mathsf{S}_{t}^{\mathsf{T}} \mathsf{F}_{t|t-1}\right)^{\mathsf{T}} \left(\mathsf{R}_{t} - \mathsf{S}_{t}^{\mathsf{T}} \mathsf{F}_{t|t-1} \mathsf{S}_{t}\right)^{-1} \mathsf{S}_{t}^{\mathsf{T}} \boldsymbol{\Theta}_{t} \mathsf{A}_{t} \mathsf{B}_{t}^{-1}$$
(A17)

by using Equation (20),

$$= \mathbf{\Theta}_t \mathsf{A}_t \mathsf{B}_t^{-1} + \mathsf{G}_t^\mathsf{T} \mathbf{\Gamma}_t \mathsf{S}_t^\mathsf{T} \mathbf{\Theta}_t \mathsf{A}_t \mathsf{B}_t^{-1}$$
(A18)

by using Equations (21) and (22), and

$$= \left(\mathsf{I}^{(m)} + \mathsf{G}_{t}^{\mathsf{T}} \boldsymbol{\Gamma}_{t} \mathsf{S}_{t}^{\mathsf{T}}\right) \boldsymbol{\Theta}_{t} \mathsf{A}_{t} \left(\mathsf{F}_{t-1|t-1} + \mathsf{A}_{t}^{\mathsf{T}} \boldsymbol{\Theta}_{t} \mathsf{A}_{t}\right)^{-1}$$
(A19)

by using Equation (18). Equation (A19) is finally identical to the coefficient matrix of  $\chi_{t-1|t-1}$  in Equation (A4), as desired.

Next, consider the coefficient matrix of  $y_t$  in Equation (A3):

$$(\mathsf{C}_{t}^{\mathsf{T}} \boldsymbol{\Psi}_{t} - \boldsymbol{\Sigma}_{t}) + (\boldsymbol{\Omega}_{t} \mathsf{A}_{t} - \mathsf{C}_{t}^{\mathsf{T}} \boldsymbol{\Sigma}_{t}^{\mathsf{T}} \mathsf{A}_{t}) (\mathsf{F}_{t-1|t-1} + \mathsf{A}_{t}^{\mathsf{T}} \boldsymbol{\Omega}_{t} \mathsf{A}_{t})^{-1} \mathsf{A}_{t}^{\mathsf{T}} \boldsymbol{\Sigma}_{t}$$

$$= (\mathsf{C}_{t}^{\mathsf{T}} \boldsymbol{\Psi}_{t} - \boldsymbol{\Sigma}_{t}) + (\mathsf{C}_{t}^{\mathsf{T}} + \boldsymbol{\Theta}_{t} \mathsf{S}_{t}) \boldsymbol{\Psi}_{t} \mathsf{S}_{t}^{\mathsf{T}} \boldsymbol{\Theta}_{t} \mathsf{A}_{t} (\mathsf{B}_{t} + \mathsf{A}_{t}^{\mathsf{T}} \boldsymbol{\Theta}_{t} \mathsf{S}_{t} \boldsymbol{\Psi}_{t} \mathsf{S}_{t}^{\mathsf{T}} \boldsymbol{\Theta}_{t} \mathsf{A}_{t})^{-1} \mathsf{A}_{t}^{\mathsf{T}} \boldsymbol{\Sigma}_{t}$$

$$+ \boldsymbol{\Theta}_{t} \mathsf{A}_{t} (\mathsf{B}_{t} + \mathsf{A}_{t}^{\mathsf{T}} \boldsymbol{\Theta}_{t} \mathsf{S}_{t} \boldsymbol{\Psi}_{t} \mathsf{S}_{t}^{\mathsf{T}} \boldsymbol{\Theta}_{t} \mathsf{A}_{t})^{-1} \mathsf{A}_{t}^{\mathsf{T}} \boldsymbol{\Sigma}_{t}$$

$$+ \boldsymbol{\Theta}_{t} \mathsf{A}_{t} (\mathsf{B}_{t} + \mathsf{A}_{t}^{\mathsf{T}} \boldsymbol{\Theta}_{t} \mathsf{S}_{t} \boldsymbol{\Psi}_{t} \mathsf{S}_{t}^{\mathsf{T}} \boldsymbol{\Theta}_{t} \mathsf{A}_{t})^{-1} \mathsf{A}_{t}^{\mathsf{T}} \boldsymbol{\Sigma}_{t}$$

$$(A20)$$

in analogy to Equations (A9)–(A11),

$$= (\mathsf{C}_{t}^{\mathsf{T}} \boldsymbol{\Psi}_{t} - \boldsymbol{\Sigma}_{t}) + (\mathsf{C}_{t}^{\mathsf{T}} + \boldsymbol{\Theta}_{t}\mathsf{S}_{t}) \boldsymbol{\Psi}_{t}\mathsf{S}_{t}^{\mathsf{T}} \boldsymbol{\Theta}_{t}\mathsf{A}_{t} (\mathsf{B}_{t} + \mathsf{A}_{t}^{\mathsf{T}} \boldsymbol{\Theta}_{t}\mathsf{S}_{t} \boldsymbol{\Psi}_{t}\mathsf{S}_{t}^{\mathsf{T}} \boldsymbol{\Theta}_{t}\mathsf{A}_{t})^{-1} \mathsf{A}_{t}^{\mathsf{T}} \boldsymbol{\Sigma}_{t} + \boldsymbol{\Theta}_{t}\mathsf{A}_{t}\mathsf{B}_{t}^{-1}\mathsf{A}_{t}^{\mathsf{T}} \boldsymbol{\Sigma}_{t} - \boldsymbol{\Theta}_{t}\mathsf{A}_{t}\mathsf{B}_{t}^{-1}\mathsf{A}_{t}^{\mathsf{T}} \boldsymbol{\Theta}_{t}\mathsf{S}_{t} (\boldsymbol{\Psi}_{t}^{-1} + \mathsf{S}_{t}^{\mathsf{T}} \boldsymbol{\Theta}_{t}\mathsf{A}_{t}\mathsf{B}_{t}^{-1}\mathsf{A}_{t}^{\mathsf{T}} \boldsymbol{\Theta}_{t}\mathsf{S}_{t})^{-1} \times \mathbf{S}_{t}^{\mathsf{T}} \boldsymbol{\Theta}_{t}\mathsf{A}_{t}\mathsf{B}_{t}^{-1}\mathsf{A}_{t}^{\mathsf{T}} \boldsymbol{\Sigma}_{t} \qquad (A21)$$

by using Equation (A1),

$$= (\mathsf{C}_{t}^{\mathsf{T}} \Psi_{t} - \Sigma_{t}) + (\mathsf{C}_{t}^{\mathsf{T}} + \Theta_{t} \mathsf{S}_{t}) \Psi_{t} \mathsf{S}_{t}^{\mathsf{T}} \Theta_{t} \mathsf{A}_{t} (\mathsf{B}_{t} + \mathsf{A}_{t}^{\mathsf{T}} \Theta_{t} \mathsf{S}_{t} \Psi_{t} \mathsf{S}_{t}^{\mathsf{T}} \Theta_{t} \mathsf{A}_{t})^{-1} \mathsf{A}_{t}^{\mathsf{T}} \Sigma_{t} + \Theta_{t} \mathsf{A}_{t} \mathsf{B}_{t}^{-1} \mathsf{A}_{t}^{\mathsf{T}} \Sigma_{t} - \Theta_{t} \mathsf{A}_{t} \mathsf{B}_{t}^{-1} \mathsf{A}_{t}^{\mathsf{T}} \Theta_{t} \mathsf{S}_{t} \Psi_{t} \mathsf{S}_{t}^{\mathsf{T}} \Theta_{t} \mathsf{A}_{t} \times (\mathsf{B}_{t} + \mathsf{A}_{t}^{\mathsf{T}} \Theta_{t} \mathsf{S}_{t} \Psi_{t} \mathsf{S}_{t}^{\mathsf{T}} \Theta_{t} \mathsf{A}_{t})^{-1} \mathsf{A}_{t}^{\mathsf{T}} \Sigma_{t}$$
(A22)

by using Equation (A2),

$$= (\mathsf{C}_{t}^{\mathsf{T}} + \boldsymbol{\Theta}_{t}\mathsf{S}_{t})\boldsymbol{\Psi}_{t} - \boldsymbol{\Theta}_{t}\mathsf{A}_{t}\mathsf{B}_{t}^{-1}\mathsf{A}_{t}^{\mathsf{T}}\boldsymbol{\Theta}_{t}\mathsf{S}_{t}\boldsymbol{\Psi}_{t} - (\mathsf{C}_{t}^{\mathsf{T}} + \boldsymbol{\Theta}_{t}\mathsf{S}_{t} - \boldsymbol{\Theta}_{t}\mathsf{A}_{t}\mathsf{B}_{t}^{-1}\mathsf{A}_{t}^{\mathsf{T}}\boldsymbol{\Theta}_{t}\mathsf{S}_{t}) \times$$
$$\boldsymbol{\Psi}_{t}\mathsf{S}_{t}^{\mathsf{T}}\boldsymbol{\Theta}_{t}\mathsf{A}_{t}(\mathsf{B}_{t} + \mathsf{A}_{t}^{\mathsf{T}}\boldsymbol{\Theta}_{t}\mathsf{S}_{t}\boldsymbol{\Psi}_{t}\mathsf{S}_{t}^{\mathsf{T}}\boldsymbol{\Theta}_{t}\mathsf{A}_{t})^{-1}\mathsf{A}_{t}^{\mathsf{T}}\boldsymbol{\Theta}_{t}\mathsf{S}_{t}\boldsymbol{\Psi}_{t}$$
(A23)

by using Equation (10) and rearraging,

$$= \left(\mathsf{C}_{t}^{\mathsf{T}} + \boldsymbol{\Theta}_{t}\mathsf{S}_{t} - \boldsymbol{\Theta}_{t}\mathsf{A}_{t}\mathsf{B}_{t}^{-1}\mathsf{A}_{t}^{\mathsf{T}}\boldsymbol{\Theta}_{t}\mathsf{S}_{t}\right) \times \left(\boldsymbol{\Psi}_{t} - \boldsymbol{\Psi}_{t}\mathsf{S}_{t}^{\mathsf{T}}\boldsymbol{\Theta}_{t}\mathsf{A}_{t}\left(\mathsf{B}_{t} + \mathsf{A}_{t}^{\mathsf{T}}\boldsymbol{\Theta}_{t}\mathsf{S}_{t}\boldsymbol{\Psi}_{t}\mathsf{S}_{t}^{\mathsf{T}}\boldsymbol{\Theta}_{t}\mathsf{A}_{t}\right)^{-1}\mathsf{A}_{t}^{\mathsf{T}}\boldsymbol{\Theta}_{t}\mathsf{S}_{t}\boldsymbol{\Psi}_{t}\right)$$
(A24)

$$= \left(\mathsf{C}_{t}^{\mathsf{T}} + \boldsymbol{\Theta}_{t}\mathsf{S}_{t} - \boldsymbol{\Theta}_{t}\mathsf{A}_{t}\mathsf{B}_{t}^{-1}\mathsf{A}_{t}^{\mathsf{T}}\boldsymbol{\Theta}_{t}\mathsf{S}_{t}\right)\left(\boldsymbol{\Psi}_{t}^{-1} + \mathsf{S}_{t}^{\mathsf{T}}\boldsymbol{\Theta}_{t}\mathsf{A}_{t}\mathsf{B}_{t}^{-1}\mathsf{A}_{t}^{\mathsf{T}}\boldsymbol{\Theta}_{t}\mathsf{S}_{t}\right)^{-1}$$
(A25)

and by using Equation (A1),

$$= \mathsf{G}_t^\mathsf{T} \mathbf{\Gamma}_t \tag{A26}$$

in analogy to Equations (A15)–(A18). Equation (A26) is finally identical to the coefficient matrix of  $y_t$  in Equation (A4), as desired.

#### References

- 1. Jazwinski, A. H. Stochastic Processes and Filtering Theory; Academic Press: San Diego, CA, USA, 1970.
- Harvey, A.C. Forecasting, Structural Time Series Models and the Kalman Filter; Cambridge University Press: Cambridge, MA, USA, 1989.
- 3. Kailath, T.; Sayed, A.; Hassibi, B. Linear Estimation; Prentice-Hall: Englewood Cliffs, NJ, USA, 2000.
- 4. Durbin, J.; Koopman, S.J. Time Series Analysis by State Space Methods, 2nd ed.; Oxford University Press: Oxford, UK, 2012.
- Casals, J.; Garcia-Hiernaux, A.; Jerez, M.; Sotoca, S.; Trindade, A.A. State-Space Methods for Time Series Analysis; CRC Press: Boca Raton, FL, USA, 2016.
- 6. Shumway, R.H.; Stoffer, D.S. Time Series Analysis and its Applications; 4th ed.; Springer: Berlin/Heidelberg, Germany, 2017.
- 7. Kalman, R.E. A new approach to linear filtering and prediction problems. J. Basic Eng. 1960, 82, 35–45. [CrossRef]
- 8. Rauch, H.E.; Tung, G.; Striebel, C.T. Maximum likelihood estimates of linear dynamic systems. *Am. Inst. Aeronaut. Astronaut. (Aiaa) J.* **1965**, *3*, 1445–1450. [CrossRef]
- 9. Duncan, D.B.; Horn, S.D. Linear dynamic recursive estimation from the viewpoint of regression analysis. *J. Am. Stat. Assoc.* **1972**, 67, 815–821. [CrossRef]

- 10. Chan, J.C.C.; Jeliazkov, I. Efficient simulation and integrated likelihood estimation in state space models. *Int. J. Math. Model. Numer. Optim.* **2009**, *1*, 101–120. [CrossRef]
- 11. Fahrmeir, L.; Kaufmann, H. On Kalman filtering, posterior mode estimation and Fisher scoring in dynamic exponential family regression. *Metrika* **1991**, *38*, 37–60. [CrossRef]
- 12. Aravkin, A.Y. Robust Methods for Kalman Filtering/Smoothing and Bundle Adjustment. Ph.D. Thesis, Department of Mathematics, University of Washington, Seattle, WA, USA, 2010.
- 13. Aravkin, A.Y.; Burke, J.V.; Pillonetto, G. Optimization viewpoint on Kalman smoothing, with applications to robust and sparse estimation. *arXiv*, **2013**, arXiv:1303.1993.
- 14. Bell, B.M. The marginal likelihood for parameters in a discrete Gauss-Markov process. *IEEE Trans. Signal Process* 2000, *48*, 870–873. [CrossRef] [PubMed]
- 15. Paige, C.C.; Saunders, M.A. Least squares estimation of discrete linear dynamic systems using orthogonal transformations. *Siam J. Numer. Anal.* **1977**, *14*, 180–193. [CrossRef]
- 16. Hammarling, S. The numerical solution of the Kalman filtering problem. In *Computational and Combinatorial Methods in System Theory*; Byrnes, C.I., Lindquist, A., Eds.; Elsevier, Amsterdam, The Netherlands, 1986; pp. 23–36.
- 17. Harvey, A.C.; Proietti, T. (Eds.) Readings in Unobserved Component Models; Oxford University Press: Oxford, UK, 2005.
- 18. Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis, Forecasting and Control*, 5th ed.; Wiley: Hoboken, NJ, USA, 2015.
- 19. Linares-Pérez, J.; Caballero-Águila, R.; García-Garrido, I. Optimal linear filter design for systems with correlation in the measurement matrices and noises: Recursive algorithm and applications. *Int. J. Syst. Sci.* **2014**, *45*, 1548–1562. [CrossRef]
- 20. Chui, C.K.; Chen, G. Kalman Filtering with Real-time Applications, 5th ed.; Springer: Berlin/Heidelberg, Germany, 2017.
- 21. Grewal, M.S.; Andrews, A.P. Kalman Filtering: Theory and Practice Using MATLAB, 4th ed.; Wiley: Hoboken, NJ, USA, 2015.
- 22. Brown, R.G.; Hwang, P.Y.C. Introduction to Random Signals and Applied Kalman Filtering, 4th ed.; Wiley: Hoboken, NJ, USA, 2012.
- 23. Gibbs, B.P. Advanced Kalman Filtering, Least-Squares and Modeling; Wiley: Hoboken, NJ, USA, 2011.
- 24. Anderson, B.D.O.; Moore, J.B. Optimal Filtering; Prentice-Hall: Englewood Cliffs, NJ, USA, 1979.
- 25. Gómez, V. Multivariate Time Series with Linear State Space Structure; Springer: Berlin/Heidelberg, Germany, 2016.
- 26. Solo, V. Topics in advanced time series analysis. In *Lectures in Probability and Statistics*; del Pino, G., Rebolledo, R., Eds.; Lecture Notes in Mathematics; Springer: Berlin/Heidelberg, Germany, 1986; Volume 1215, pp. 165–328.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.