ELSEVIER Linear Algebra and its Applications 321 (2000) 209–232
www.elsevier.com/locate/laa

Construction of the exact Fisher information matrix of Gaussian time series models by means of matrix differential rules

André Klein [a,*], Guy Mélard [b,1,2], Toufik Zahaf [b]

[a] *Department of Actuarial Sciences, Econometrics and Quantitative Methods, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands*

[b] *Institut de Statistique, Université Libre de Bruxelles, Campus Plaine CP 210, Bd du Triomphe, B-1050 Bruxelles, Belgium*

Received 27 January 1998; accepted 12 February 1999

Submitted by S. Puntanen

Abstract

The Fisher information matrix is of fundamental importance for the analysis of parameter estimation of time series models. In this paper the exact information matrix of a multivariate Gaussian time series model expressed in state space form is derived. A computationally efficient procedure is used by applying matrix differential rules for the derivatives of a matrix function $J = J(\theta)$ with respect to its vector argument. An algorithm is given. It is sketched for the general state space structure without specifying a parametrization. It is then detailed for the vector autoregressive moving average (VARMA) model, with a given parametrization, where explicit recurrent relations are developed. © 2000 Elsevier Science Inc. All rights reserved.

Keywords: Matrix differentiation; Vector autoregressive moving average model; Fisher information matrix

[*] Corresponding author. Tel.: +31-20-5254245; fax: +31-20-5254349; e-mail: aklein@fee.uva.nl

[1] This paper has benefited from a grant “Action de Recherche Concertée” 96/01/205 of the “Communauté Francaise Wallonie–Bruxelles”.

[2] Tel.: +32-2-6505890; fax: +32-2-6505899; e-mail: gmelard@ulb.ac.be

PII: S 0 0 2 4 - 3 7 9 5 (9 9) 0 0 0 4 5 - 2

1. Introduction

This paper is concerned with the construction of the exact (finite sample) Fisher information matrix of a Gaussian time series model expressed in a state space form. A straightforward way of solving the problem is to explore the properties of matrix derivatives.

The interest among statisticians in the algebra of the Kronecker product of matrices and matrix vectorization has been growing over the last decade. Pollock [17] focuses on the connection between the algebra of Kronecker products and the abstract algebra of tensor products. Material on matrix derivatives and their applications in classical (i.e. with independent observations) statistical models depending on a vector θ of parameters, can be found in [23]. In both papers relevant references are provided.

Time series models introduce an additional dimension: time. Even in the Gaussian case, maximum likelihood estimation requires numerical optimization. The Fisher information matrix plays an important role in describing the covariance structure of the estimator. In a simpler case than that considered here, Klein and Spreij [8] have already pointed out the relationship between the information matrix of time series models and linear algebra concepts. We do not attempt to go in that direction here. We are more interested in the mere evaluation of the information matrix J, as a whole, of a general multivariate time series model, instead of element by element, J_{ij}, as it is generally done by considering the pair of scalar parameters indexed by i and j with respect to which derivatives are carried out.

Each approach has its relative advantages. One advantage of our approach is that all derivatives are explicitly stated in arrays (vectors and matrices). The explicitness of the full-array form can be useful for gaining mathematical insights and facilitating programming. Practically speaking, the key is having mathematical expressions which are easy to program as dense-matrix computations. If all the matrices in the Kronecker products in Section 4 were dense, then a program which duplicates what is written on the paper will be more or less efficient. But other routine things still need to be done, like making sure that only about one half of symmetric matrices are computed, etc. The point is that the full-array form is not always better. The scalar form involves much simpler algebraic structures because it does not require Kronecker products.

Since the purpose of this paper is to construct an algorithm of the exact Fisher information matrix for a Gaussian time series model expressed in a state space form it is worth mentioning that it is now more than three decades that statisticians and econometricians have the Kalman filter [5] at their disposal. Schweppe [19] had first noticed that the full information likelihood function of time series models expressed in a state space form can be evaluated by means of the Kalman filter. Often the state space system is time-invariant (i.e. the coefficients of the state space and observation equations are constant, not random or functions of time). For that case of the time invariant model, Morf et al. [13] have introduced the so-called Chandrasekhar equations, which are computationally more efficient by an order of magnitude. Fast procedures are often more complex but it is not the case here. On the contrary, the Chandrasekhar equations are also slightly simpler than the corresponding Kalman filter equations. Of course, the Chandrasekhar equations do not work in time-varying situations. Even with a time-invariant state space equation, the observation equation can be time-varying, for example when different series are observed at different frequencies [26].

The asymptotic Fisher information matrix of time series models has a long history. It was first introduced by Whittle [22]. It is based on an approximation of the Gaussian likelihood which is valid asymptotically. Subsequent developments are well covered in [16,2,24,12,6]. More recently, the sample or exact information matrix $J(\theta)$ has been studied. It is defined as minus the mathematical expectation of the Hessian of the exact likelihood function, evaluated at the final estimated value of the parameters. It can be written as $J(\theta) = \sum_{t=1}^{N} J_t(\theta)$, where N is the length of the time series (see (22)). Porat and Friedlander [18] have described an algorithm for a univariate ARMA model with a deterministic additive component. The method is both complex and computationally intensive. The number of scalar operations is indeed of order N^2. Independently, Zadrozny [24,25] and Teircero [21] have given a much more efficient algorithm, since the number of operations is proportional to N. The method is based on the Kalman filter, and has been applied to the vector autoregressive moving average (VARMA) model by the former, and to the general state space form by the latter. However, in all cases, the corresponding algorithm is given for one element (i,j) of $J_t(\theta)$. Mélard and Klein [11] have described a method for computing the exact information matrix of a univariate ARMA model. That method is based on the alternative expression of the Gaussian exact likelihood in terms of the Chandrasekhar equations instead of the Kalman filter equations. Note that in this case $J_t(\theta)$ is computed as a matrix. Klein et al. [7] have used that approach for a univariate single input single output (SISO) model.

In this paper a generalization to multivariate models is considered, starting from a general state space model with a detailed treatment for VARMA processes, including the initialization issues.

The main term of $J_t(\theta)$ is of the form

$$\mathbb{E}\left\{ \frac{\partial \tilde{z}_t}{\partial \theta^{\mathrm{T}}} \otimes \frac{\partial \tilde{z}_t}{\partial \theta^{\mathrm{T}}} \right\}^{\mathrm{T}},$$

where $\tilde{z}_t$ is the innovation (defined below in (10)). The purpose is to obtain a recurrence equation for the matrix as a whole. This is done by writing an expression for $(\partial\tilde{z}_t/\partial\theta^{\mathrm{T}})$ using a suitable vectorization of the Chandrasekhar recursion. An algorithm is set sketched for the general vector state space form without specifying a parametrization. It is then detailed for the VARMA model with a given parametrization. Since a parametrization is not given for the general vector state space model, the corresponding algorithm presented is incomplete because some terms are still unspecified, whereas for the VARMA case an explicit form of the recurrence relations can be established. A test program in the MATLAB environment is available from the authors for the VARMA model (at address http://ulb.ac.be/~gmelard/efimvarma.html).

This paper is organized as follows. In Section 2 some preliminaries are given. In Section 3 we formulate the closed form for the exact information matrix for general state space. The VARMA version of the exact information matrix, with the complete recursions, is studied in Section 4. The aspect of initialization is described in Section 5, which is followed by the conclusion.

2. Some preliminaries

2.1. The model

We first present the general state space system described by

$$x_{t+1} = \Phi x_t + \Gamma u_t + F w_t, \tag{1}$$

$$z_t = H x_t + D u_t + C v_t, \tag{2}$$

where $z_t \in \mathbb{R}^m$ is the vector of observations, $x_t \in \mathbb{R}^n$ is the vector of the state variables, $u_t \in \mathbb{R}^r$ the vector of exogenous variables, $t \in \mathbb{N}$, $w_t \in \mathbb{R}^q$ and $v_t \in \mathbb{R}^g$, and (v_t, w_t) is a Gaussian white noise process with $\mathbb{E}(w_t) = 0$, $\mathbb{E}(v_t) = 0$,

$$\mathbb{E}\begin{bmatrix} w_t \\ v_t \end{bmatrix}\begin{bmatrix} w_t^{\mathrm{T}} & v_t^{\mathrm{T}} \end{bmatrix} = \begin{pmatrix} Q & S \\ S^{\mathrm{T}} & R \end{pmatrix} \geqslant 0, \tag{3}$$

where $Q \geqslant 0$, $R > 0$, and $^{\mathrm{T}}$ denotes transposition. These are standard assumptions (e.g. [1]). Note that R, the covariance matrix of the innovation process, is supposed to be invertible to avoid degeneracy problems. The conditions for identification of state space models can be found in [3]. Finally, the usual Gaussian inferential process which uses the information matrix requires the data generating process to be stationary.

We suppose that the model (1)–(3) depends on ℓ parameters denoted by the vector $\theta = (\theta_1, \ldots, \theta_\ell)^{\mathrm{T}}$. Thus Φ, Γ, F, H, D, C, Q, S, and R are functions of θ and are supposed to be two times continuously differentiable.

2.2. The Chandrasekhar equations

There are several ways to express the exact likelihood function of a time series $\{z_1, \dots, z_N\}$ of length N. Except for the closed form expression of a normal multivariate density, the simplest representation is based on the Chandrasekhar recurrence equations, which are also the most computationally efficient, even with respect to the Kalman filter.

The Kalman filter consists in a system of several recurrences for computing $\hat{x}_{t|t-1}$, the prediction of the state vector, and $\hat{z}_{t|t-1}$, the prediction of the observation vector. Let $\tilde{z}_t$ be the (one-step-ahead) prediction error of the observations, and $\tilde{x}_t$ the prediction error of the state vector by

$$\tilde{x}_t = x_t - \hat{x}_{t|t-1}. \tag{4}$$

Let $P_{t|t-1} = \mathbb{E}\{\widetilde{x}_t \widetilde{x}_t^{\mathrm{T}}\}$, the covariance matrix of the prediction error of the state vector, an $n \times n$ matrix.

The Chandrasekhar equations make use of smaller matrices and are based on the assumption that $\operatorname{rank}(P_{t|t-1} - P_{t-1|t-2})$ (which is non-increasing in t) $\leqslant k = \operatorname{rank}(P_{1|0}) \leqslant n$. The following recurrence equations can then be written [13] to derive $\tilde{z}_t$, and $B_t = \mathbb{E}\left[\tilde{z}_t \tilde{z}_t^{\mathrm{T}}\right]$

$$B_t = B_{t-1} + H Y_{t-1} X_{t-1} Y_{t-1}^{\mathrm{T}} H^{\mathrm{T}}, \tag{5}$$

$$K_t = \left[K_{t-1} B_{t-1} + \Phi Y_{t-1} X_{t-1} Y_{t-1}^{\mathrm{T}} H^{\mathrm{T}}\right] B_t^{-1}, \tag{6}$$

$$Y_t = [\Phi - K_{t-1} H] Y_{t-1}, \tag{7}$$

$$X_t = X_{t-1} - X_{t-1} Y_{t-1}^{\mathrm{T}} H^{\mathrm{T}} B_t^{-1} H Y_{t-1} X_{t-1}, \tag{8}$$

$$\hat{z}_{t|t-1} = H \hat{x}_{t|t-1} + D u_t, \tag{9}$$

$$\tilde{z}_t = z_t - \hat{z}_{t|t-1} \tag{10}$$

$$\hat{x}_{t+1|t} = \Phi \hat{x}_{t|t-1} + \Gamma u_t + K_t \tilde{z}_t. \tag{11}$$

The auxiliary matrices K_t, X_t and Y_t have dimensions $n \times m$, $k \times k$ and $n \times k$, respectively. We also introduce the prediction error of the state vector based on (1) and (2)

$$\tilde{x}_{t+1} = \overline{\Phi}_t \tilde{x}_t - K_t C v_t + F w_t, \tag{12}$$

where we denote $\overline{\Phi}_t = \Phi - K_t H$. We can rewrite the innovation

$$\tilde{z}_t = H \tilde{x}_t + C v_t. \tag{13}$$

The initial conditions will be discussed in the VARMA case in Section 5, where $k = m$.

Given a time series of length N, the Chandrasekhar equations (5)–(11) are used to compute the negative logarithm of the likelihood of the system described by (1) and (2)

$$l(\theta) = -\log L(\theta) = \sum_{t=1}^{N} \left\{ \frac{m}{2}\log(2\pi) + \frac{1}{2}\log|B_t| + \frac{1}{2}\tilde{z}_t^{\mathrm{T}} B_t^{-1}\tilde{z}_t \right\}. \tag{14}$$

3. Closed form recurrences for the exact Fisher information matrix

In this section, we derive the matrix recurrences which are needed in order to obtain the exact information matrix. We start with an expression of the exact information matrix based on the derivatives with respect to one element of the vector θ from which we will deduce the corresponding derivatives with respect to the vector θ as a whole.

For the element (i, j) of the exact information matrix J we have

$$[J(\theta)]_{ij} = \mathbb{E}\left\{ \frac{\partial^2 l(\theta)}{\partial\theta_i \partial\theta_j} \right\}. \tag{15}$$

It can be shown that ([4,18,21])

$$[J(\theta)]_{ij} = \sum_{t=1}^{N} \left[\frac{1}{2}\mathrm{Tr}\left\{ B_t^{-1}\frac{\partial B_t}{\partial\theta_i} B_t^{-1}\frac{\partial B_t}{\partial\theta_j} \right\} + \mathrm{Tr}\left(B_t^{-1}\mathbb{E}\left\{ \frac{\partial\tilde{z}_t}{\partial\theta_i}\frac{\partial\tilde{z}_t^{\mathrm{T}}}{\partial\theta_j} \right\} \right) \right] \tag{16}$$

which is always given as a scalar in the literature. By applying the following rules [10]

$$\mathrm{Tr}(A^{\mathrm{T}}B) = (\mathrm{vec}\,A)^{\mathrm{T}}\mathrm{vec}\,B, \quad \text{where } A \in \mathbb{R}^{m\times n} \text{ and } B \in \mathbb{R}^{m\times n}, \tag{17}$$

$$\mathrm{vec}(ABC) = (C^{\mathrm{T}} \otimes A)\mathrm{vec}\,B, \quad \text{where } A \in \mathbb{R}^{m\times n}, \; B \in \mathbb{R}^{n\times p} \text{ and } C \in \mathbb{R}^{p\times s}, \tag{18}$$

$$\mathrm{Tr}(ABC) = \mathrm{Tr}(CAB), \quad \text{where } A \in \mathbb{R}^{m\times n}, \; B \in \mathbb{R}^{n\times p} \text{ and } C \in \mathbb{R}^{p\times m}, \tag{19}$$

$$(A \otimes B)(C \otimes D) = AC \otimes BD, \quad \text{where } A \in \mathbb{R}^{m\times n}, \; B \in \mathbb{R}^{p\times q}, \; C \in \mathbb{R}^{n\times k} \text{ and } D \in \mathbb{R}^{q\times l}, \tag{20}$$

where $\otimes$ is the Kronecker product (see also [9]), the first term of (16) is

$$\mathrm{Tr}\left\{B_t^{-1}\frac{\partial B_t}{\partial\theta_i}B_t^{-1}\frac{\partial B_t}{\partial\theta_j}\right\} = \left[\mathrm{vec}\left(\frac{\partial B_t^{\mathrm{T}}}{\partial\theta_i}B_t^{-\mathrm{T}}\right)\right]^{\mathrm{T}}\left[\mathrm{vec}\left(B_t^{-1}\frac{\partial B_t}{\partial\theta_j}\right)\right]$$

$$= \left(\frac{\partial \mathrm{vec}\, B_t}{\partial\theta_i}\right)^{\mathrm{T}}\left(B_t^{-1}\otimes B_t^{-1}\right)\left(\frac{\partial \mathrm{vec}\, B_t}{\partial\theta_j}\right) \tag{21}$$

and the general expression of the exact Fisher information matrix can be written as the following $\ell\times\ell$ matrix

$$J(\theta) = \sum_{t=1}^{N}\left[\frac{1}{2}\left(\frac{\partial \mathrm{vec}\, B_t}{\partial\theta^{\mathrm{T}}}\right)^{\mathrm{T}}\left(B_t^{-1}\otimes B_t^{-1}\right)\left(\frac{\partial \mathrm{vec}\, B_t}{\partial\theta^{\mathrm{T}}}\right) + \mathbb{E}\left\{\left(\frac{\partial\tilde{z}_t}{\partial\theta^{\mathrm{T}}}\right)^{\mathrm{T}}B_t^{-1}\left(\frac{\partial\tilde{z}_t}{\partial\theta^{\mathrm{T}}}\right)\right\}\right]. \tag{22}$$

Note that $(\partial\mathrm{vec}\, B_t)/(\partial\theta^{\mathrm{T}})$ is an $m^2\times\ell$ matrix. Of course (22) can also be derived directly from (14) and (15), but the derivation is more complex than the one used in this paper.

For solving the first term of (22) the derivatives of the Chandrasekhar equations are used, whereas the situation is different for the second one, which involves the expected value of stochastic elements. In order to obtain an appropriate covariance structure, vectorization of $J(\theta)$ is recommended. Consequently we obtain

$$\mathrm{vec}\, J(\theta) = \sum_{t=1}^{N}\left\{\frac{1}{2}\left[\left(\frac{\partial \mathrm{vec}\, B_t}{\partial\theta^{\mathrm{T}}}\right)\otimes\left(\frac{\partial \mathrm{vec}\, B_t}{\partial\theta^{\mathrm{T}}}\right)\right]^{\mathrm{T}}\mathrm{vec}\left(B_t^{-1}\otimes B_t^{-1}\right) + \mathbb{E}\left\{\frac{\partial\tilde{z}_t}{\partial\theta^{\mathrm{T}}}\otimes\frac{\partial\tilde{z}_t}{\partial\theta^{\mathrm{T}}}\right\}^{\mathrm{T}}\mathrm{vec}\, B_t^{-1}\right\}. \tag{23}$$

For developing a recurrence for the second term of (23) new recurrences have to be constructed. To achieve this goal, a differential rule is used to derive an expression of the derivatives of $\tilde{x}_{t+1}$ in (12) and $\tilde{z}_t$ in (13). This will be given in the following theorem.

Theorem. *Let* $\tilde{x}_{t+1}$ *and* $\tilde{z}_t$ *be described by* (12) *and* (13), *respectively*, *then*

$$\frac{\partial\tilde{x}_{t+1}}{\partial\theta^{\mathrm{T}}} = \left(\tilde{x}_t^{\mathrm{T}}\otimes I_n\right)\frac{\partial\,\mathrm{vec}\,\overline{\Phi}_t}{\partial\theta^{\mathrm{T}}} + \overline{\Phi}_t\frac{\partial\tilde{x}_t}{\partial\theta^{\mathrm{T}}} - \left(v_t^{\mathrm{T}}C^{\mathrm{T}}\otimes I_n\right)\frac{\partial\,\mathrm{vec}\,K_t}{\partial\theta^{\mathrm{T}}} - \left(v_t^{\mathrm{T}}\otimes K_t\right)\frac{\partial\,\mathrm{vec}\,C}{\partial\theta^{\mathrm{T}}} + \left(w_t^{\mathrm{T}}\otimes I_n\right)\frac{\partial\,\mathrm{vec}\,F}{\partial\theta^{\mathrm{T}}} - K_tC\frac{\partial v_t}{\partial\theta^{\mathrm{T}}} + F\frac{\partial w_t}{\partial\theta^{\mathrm{T}}} \tag{24}$$

and

$$\frac{\partial \tilde{z}_t}{\partial \theta^{\mathrm{T}}} = (\tilde{x}_t^{\mathrm{T}} \otimes I_m)\frac{\partial \operatorname{vec} H}{\partial \theta^{\mathrm{T}}} + H\frac{\partial \tilde{x}_t}{\partial \theta^{\mathrm{T}}} + (v_t^{\mathrm{T}} \otimes I_m)\frac{\partial \operatorname{vec} C}{\partial \theta^{\mathrm{T}}} + C\frac{\partial v_t}{\partial \theta^{\mathrm{T}}}. \tag{25}$$

Proof. The following differential rule is used [10]. Consider a real, differentiable $(m \times n)$ matrix function $X(\theta)$ of real $(\ell \times 1)$ vector $\theta = (\theta_1, \ldots, \theta_\ell)^{\mathrm{T}}$, where m, n and ℓ are positive integers. Let $(m \times n)$ matrices $\partial_r X = (\partial X_{ij}/\partial \theta_r)$ with $r = 1, \ldots, \ell$ be the first order derivatives of $X(\theta)$ in partial derivative form with X_{ij} being the first element (i, j) of X. Write $\mathrm{d}X_{ij} = \sum_{r=1}^{\ell}(\partial X_{ij}/\partial \theta_r)\mathrm{d}\theta_r$, where $\mathrm{d}\theta_r$ is an arbitrary perturbation of θ_r. The $(m \times n)$ matrix $\mathrm{d}X = (\mathrm{d}X_{ij})$ is the differential form of the first order derivative $X(\theta)$. An expression in differential form can instantaneously be put into a partial derivative form by replacing d with ∂_r for $r = 1, \ldots, \ell$.

Let $X(\theta)$ and $Y(\theta)$ be real $(m \times n)$ and $(n \times p)$ differentiable matrix functions of the real vector $\theta(\ell \times 1)$, where m, n, p, and ℓ are positive integers. The usual scalar product rule of differentiation yields

$$\mathrm{d}(XY) = (\mathrm{d}X)Y + X(\mathrm{d}Y). \tag{26}$$

By applying this approach, first for $\tilde{x}_{t+1}$ in (12), we have

$$\begin{aligned}\mathrm{d}\tilde{x}_{t+1} = {} & (\mathrm{d}\overline{\Phi}_t)\tilde{x}_t + \overline{\Phi}_t \mathrm{d}\tilde{x}_t - (\mathrm{d}K_t)Cv_t - K_t(\mathrm{d}C)v_t \\ & - K_t C(\mathrm{d}v_t) + (\mathrm{d}F)w_t + F(\mathrm{d}w_t).\end{aligned} \tag{27}$$

Let us vectorize the matrix $X(\theta)$ defined above according to (18), then the $(mn \times \ell)$ matrix $\partial \operatorname{vec} X(\theta)/\partial \theta^{\mathrm{T}}$ is the gradient form of first order derivatives of $X(\theta)$ and can be defined as $\operatorname{vec} \mathrm{d}X(\theta) = (\partial(\operatorname{vec} X(\theta))/\partial \theta^{\mathrm{T}})\, \mathrm{d}\theta = \mathrm{d} \operatorname{vec} X(\theta)$ (see also [14]). Componentwise application of this rule gives

$$\begin{aligned}\mathrm{d}\tilde{x}_{t+1} = {} & (\tilde{x}_t^{\mathrm{T}} \otimes I_n)\mathrm{d} \operatorname{vec}\overline{\Phi}_t + \overline{\Phi}_t \mathrm{d}\tilde{x}_t - (v_t^{\mathrm{T}} C^{\mathrm{T}} \otimes I_n)\mathrm{d} \operatorname{vec} K_t - (K_t C)\mathrm{d}v_t \\ & - (v_t^{\mathrm{T}} \otimes K_t)\mathrm{d} \operatorname{vec} C + (w_t^{\mathrm{T}} \otimes I_n)\mathrm{d} \operatorname{vec} F + F\, \mathrm{d}w_t \\ = {} & (\tilde{x}_t^{\mathrm{T}} \otimes I_n)\frac{\partial \operatorname{vec} \overline{\Phi}_t}{\partial \theta^{\mathrm{T}}}\mathrm{d}\theta + \overline{\Phi}_t \frac{\partial \tilde{x}_t}{\partial \theta^{\mathrm{T}}}\mathrm{d}\theta - (v_t^{\mathrm{T}} C^{\mathrm{T}} \otimes I_n)\frac{\partial \operatorname{vec} K_t}{\partial \theta^{\mathrm{T}}}\mathrm{d}\theta \\ & + F\frac{\partial w_t}{\partial \theta^{\mathrm{T}}}\mathrm{d}\theta - (v_t^{\mathrm{T}} \otimes K_t)\frac{\partial \operatorname{vec} C}{\partial \theta^{\mathrm{T}}}\mathrm{d}\theta \\ & + (w_t^{\mathrm{T}} \otimes I_n)\frac{\partial \operatorname{vec} F}{\partial \theta^{\mathrm{T}}}\mathrm{d}\theta - K_t C\frac{\partial v_t}{\partial \theta^{\mathrm{T}}}\mathrm{d}\theta\end{aligned} \tag{28}$$

which yields (24). Analogously, the differential of $\tilde{z}_t$ in (13) is

$$\begin{aligned}
\mathrm{d}\tilde{z}_t &= (\mathrm{d}H)\tilde{x}_t + H\mathrm{d}\tilde{x}_t + (\mathrm{d}C)v_t + (C)\,\mathrm{d}v_t \\
&= (\tilde{x}_t^{\mathrm{T}} \otimes I_m)\mathrm{dvec}H + H\mathrm{d}\tilde{x}_t + (v_t^{\mathrm{T}} \otimes I_m)\ \mathrm{dvec}C + C\,\mathrm{d}v_t \\
&= (\tilde{x}_t^{\mathrm{T}} \otimes I_m)\frac{\partial \mathrm{vec}\, H}{\partial \theta^{\mathrm{T}}}\mathrm{d}\theta + H\frac{\partial \tilde{x}_t}{\partial \theta^{\mathrm{T}}}\mathrm{d}\theta + (v_t^{\mathrm{T}} \otimes I_m)\frac{\partial \mathrm{vec}\, C}{\partial \theta^{\mathrm{T}}}\mathrm{d}\theta \\
&\quad + C\frac{\partial v_t}{\partial \theta^{\mathrm{T}}}\mathrm{d}\theta
\end{aligned} \tag{29}$$

which yields (25). □

The following rules will be used where the orders of the matrices involved are supposed to be such that all the following operations are defined.
Rule 1. $(A + B) \otimes (C + D) = A \otimes C + A \otimes D + B \otimes C + B \otimes D.$
Rule 2. $A \otimes B \otimes C = (A \otimes B) \otimes C = A \otimes (B \otimes C).$
Rule 3. Let $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{p\times q}$, then $M_{p,m}(A \otimes B)M_{n,q} = B \otimes A$, where the commutation matrix $M_{m,n} \in \mathbb{R}^{mn\times mn}$ is defined by (see [15])

$$M_{m,n} = \sum_{i=1}^{m}\sum_{j=1}^{n}(H_{ij} \otimes H_{ij}^{\mathrm{T}}), \tag{30}$$

where $H_{ij}^{\mathrm{T}} = e_j^n(e_i^m)^{\mathrm{T}}$, and e_i^m is the unit column vector of order m. Note the additional properties which are of interest

$$\begin{aligned}
&M_{n,m}^{\mathrm{T}} = M_{m,n}, M_{1,n} = M_{n,1} = I_n, \\
&M_{n,m}M_{m,n} = I_{mn}\ \text{(orthogonality property)}.
\end{aligned} \tag{31}$$

Rule 4. Let the random vectors $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ be jointly distributed with $\mathbb{E}(x) = \mu_1$, $\mathbb{E}(y) = \mu_2$ and $\mathbb{E}\{(y - \mu_2)(x - \mu_1)^{\mathrm{T}}\} = \Omega$. Then (see [15])

$$\mathbb{E}(x \otimes y) = \mathrm{vec}\,\Omega + \mu_1 \otimes \mu_2$$

In order to simplify the presentation of the results, we introduce the following notations in the recurrences, $(\partial \mathrm{vec}\, A)/(\partial\theta^{\mathrm{T}}) = A^\theta$ and $(\partial b)/(\partial\theta^{\mathrm{T}}) = b^\theta$, for any matrix A and any vector b. Be careful, however, that A^θ and b^θ are matrices.

Note that $\tilde{x}_t$ and $\tilde{x}_t^\theta$ are uncorrelated with the vector $\left(w_s^{\mathrm{T}}, v_s^{\mathrm{T}}\right)$ for $s \geqslant t$. Indeed, from (12), $\tilde{x}_{t+1} - \overline{\Phi}_t\tilde{x}_t = K_tCv_t + Fw_t$ is uncorrelated with v_{t+1} and w_{t+1} since the noises are white noise processes. To express $\mathbb{E}\{\tilde{z}_t^\theta \otimes \tilde{z}_t^\theta\}^{\mathrm{T}}$ in an appropriate form, it is straightforward to combine the theorem given above, rules (1–4) and (20):

$$\begin{aligned}
\mathbb{E}\left\{\tilde{z}_t^\theta \otimes \tilde{z}_t^\theta\right\}^{\mathrm{T}} = {} & \left\{H^\theta \otimes H^\theta\right\}^{\mathrm{T}}\left\{\left(M_{nm,n}\left[\mathbb{E}\left\{\tilde{x}_t \otimes \tilde{x}_t\right\} \otimes I_m\right]\right) \otimes I_m\right\} \\
& + \left\{\left(H^\theta\right)^{\mathrm{T}} \otimes I_\ell\right\}\left(M_{nm,\ell}\left[\mathbb{E}\left\{\left(\tilde{x}_t^\theta\right)^{\mathrm{T}} \otimes \tilde{x}_t\right\} \otimes I_m\right] M_{n,m}\right)\left(I_m \otimes H^{\mathrm{T}}\right) \\
& + \left\{I_\ell \otimes \left(H^\theta\right)^{\mathrm{T}}\right\}\left\{\left[\mathbb{E}\left\{\left(\tilde{x}_t^\theta\right)^{\mathrm{T}} \otimes \tilde{x}_t\right\} \otimes I_m\right]\left(H^{\mathrm{T}} \otimes I_m\right)\right\} \\
& + \mathbb{E}\left\{\tilde{x}_t^\theta \otimes \tilde{x}_t^\theta\right\}^{\mathrm{T}}(H \otimes H)^{\mathrm{T}} + \mathbb{E}\left\{v_t^\theta \otimes \tilde{x}_t^\theta\right\}^{\mathrm{T}}(C \otimes H)^{\mathrm{T}} \\
& + \left\{C^\theta \otimes C^\theta\right\}^{\mathrm{T}}\left\{\left(M_{gm,g}[\operatorname{vec} R \otimes I_m]\right) \otimes I_m\right\} \\
& + \left\{\left(H^\theta\right)^{\mathrm{T}} \otimes I_\ell\right\}\left(M_{mn,\ell}\left[\mathbb{E}\left\{\left(v_t^\theta\right)^{\mathrm{T}} \otimes \tilde{x}_t\right\} C^{\mathrm{T}} \otimes I_m\right] M_{m,m}\right) \\
& + \left\{\left(C^\theta\right)^{\mathrm{T}} \otimes I_\ell\right\}\left(M_{gm,\ell}\left[\mathbb{E}\left\{\left(v_t^\theta\right)^{\mathrm{T}} \otimes v_t\right\} C^{\mathrm{T}} \otimes I_m\right] M_{m,m}\right) \\
& + \mathbb{E}\left\{\tilde{x}_t^\theta \otimes v_t^\theta\right\}^{\mathrm{T}}(H \otimes C)^{\mathrm{T}} + \mathbb{E}\left\{v_t^\theta \otimes v_t^\theta\right\}^{\mathrm{T}}(C \otimes C)^{\mathrm{T}} \\
& + \left\{I_\ell \otimes \left(C^\theta\right)^{\mathrm{T}}\right\}\left[\mathbb{E}\left\{\left(v_t^\theta\right)^{\mathrm{T}} \otimes v_t\right\} C^{\mathrm{T}} \otimes I_m\right] \\
& + \left\{I_\ell \otimes \left(H^\theta\right)^{\mathrm{T}}\right\}\left[\mathbb{E}\left\{\left(v_t^\theta\right)^{\mathrm{T}} \otimes \tilde{x}_t\right\} C^{\mathrm{T}} \otimes I_m\right] \qquad (32)
\end{aligned}$$

with $\mathbb{E}\{v_t \otimes v_t\} = \operatorname{vec} R$ and $\mathbb{E}\{\widetilde{x}_t \otimes \widetilde{x}_t\} = \operatorname{vec} \mathbb{E}\{\widetilde{x}_t\widetilde{x}_t^{\mathrm{T}}\} = \operatorname{vec} P_t$ (see rule 4 above) where P_t is the solution of the Riccati equation

$$P_{t+1|t} = \overline{\Phi}_t P_{t|t-1}\overline{\Phi}_t^{\mathrm{T}} + FQF^{\mathrm{T}} - K_t B_t K_t^{\mathrm{T}}. \qquad (33)$$

In all the recurrences, the notation $M_{ab,c}$ will be used as in rule 3 with $m = ab$ and $n = c$, where a, b and c are integers.

We shall not go on because, in the general case considered here, it is not possible to obtain the derivatives of the state space and observation disturbances. [24] encountered the same difficulty and solved it in [25] along the lines of [21]. This problem disappears in Section 4, where the general recurrences are applied to the VARMA model, because the process disturbance, the observation disturbance and the innovation are identical.

4. The vector ARMA model

In this part of the paper, the complete form of the algorithm described in Section 3 will be developed for the VARMA model.

We set forth a special case of (1–2) which is the VARMA model of order (p, s). The parametrization of this structure is given by

$$z_t = \alpha_1 z_{t-1} + \alpha_2 z_{t-2} + \cdots + \alpha_p z_{t-p} + w_t - \beta_1 w_{t-1} - \beta_2 w_{t-2} - \cdots - \beta_s w_{t-s}, \tag{34}$$

where w_t is the innovation process (a sequence of independently and normally distributed random vectors with mean 0 and an invertible covariance matrix Q). We assume that the stationarity and invertibility conditions are fulfilled,

$$\det\,(I_m - \alpha_1 z - \alpha_2 z^2 - \cdots - \alpha_p z^p) \neq 0 \text{ for } |z| \leqslant 1, \tag{35}$$

$$\det\,(I_m - \beta_1 z - \beta_2 z^2 - \cdots - \beta_s z^s) \neq 0 \text{ for } |z| \leqslant 1 \tag{36}$$

and that the model is uniquely specified which implies that the autoregressive and moving average matrix polynomials are left coprime.

The state space form can be written (for example, [20])

$$x_{t+1} = \Phi x_t + F w_t, \tag{37}$$

$$z_t = H x_t + w_t, \tag{38}$$

where

$$\Phi = \begin{pmatrix} \alpha_1 & I_m & 0_m & \cdots & 0_m \\ \alpha_2 & 0_m & I_m & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0_m \\ \vdots & \vdots & \ddots & 0_m & I_m \\ \alpha_h & 0_m & \cdots & \cdots & 0_m \end{pmatrix}, \quad F = \begin{pmatrix} \alpha_1 - \beta_1 \\ \alpha_2 - \beta_2 \\ \vdots \\ \alpha_h - \beta_h \end{pmatrix} \quad \text{and}$$

$$H^{\mathrm{T}} = \begin{pmatrix} I_m \\ 0_m \\ \vdots \\ 0_m \end{pmatrix}, \tag{39}$$

$h = \max(p, s)$, $\alpha_i = 0_m, i > p$, $\beta_i = 0_m, i > s$, and consequently $n = hm$. Note that $H^\theta = 0$.

We start from an equivalent representation of (12–13), which is

$$\tilde{z}_t = H\tilde{x}_t + w_t \tag{40}$$

and

$$\tilde{x}_{t+1} = \overline{\Phi}_t \tilde{x}_t + \bar{F}_t w_t \tag{41}$$

with $\overline{\Phi}_t = \Phi - K_t H$, and $\bar{F}_t = (F - K_t)$. Instead of writing the recurrences in function of $t-1$ we write them in function of t for typographical brevity, like in (33) above.

Expression (24) becomes

$$\frac{\partial \tilde{x}_{t+1}}{\partial \theta^{\mathrm{T}}} = (\tilde{x}_t^{\mathrm{T}} \otimes I_n)\frac{\partial \mathrm{vec}\, \overline{\Phi}_t}{\partial \theta^{\mathrm{T}}} + \overline{\Phi}_t \frac{\partial \tilde{x}_t}{\partial \theta^{\mathrm{T}}} + (w_t^{\mathrm{T}} \otimes I_n)\frac{\partial \mathrm{vec}\, \bar{F}_t}{\partial \theta^{\mathrm{T}}} + \bar{F}_t \frac{\partial w_t}{\partial \theta^{\mathrm{T}}} \tag{42}$$

and the equivalent of (25) is

$$\frac{\partial \tilde{z}_t}{\partial \theta^{\mathrm{T}}} = H \frac{\partial \tilde{x}_t}{\partial \theta^{\mathrm{T}}} + \frac{\partial w_t}{\partial \theta^{\mathrm{T}}}. \tag{43}$$

We additionally need the following recurrence equations which can be derived from (37) and (38)

$$\frac{\partial x_{t+1}}{\partial \theta^{\mathrm{T}}} = \left(x_t^{\mathrm{T}} \otimes I_n\right)\frac{\partial \mathrm{vec}\, \Phi}{\partial \theta^{\mathrm{T}}} + \Phi \frac{\partial x_t}{\partial \theta^{\mathrm{T}}} + \left(w_t^{\mathrm{T}} \otimes I_n\right)\frac{\partial \mathrm{vec}\, F}{\partial \theta^{\mathrm{T}}} + F \frac{\partial w_t}{\partial \theta^{\mathrm{T}}} \tag{44}$$

and

$$\frac{\partial w_{t+1}}{\partial \theta^{\mathrm{T}}} = -\left\{ \left(x_t^{\mathrm{T}} \otimes H\right)\frac{\partial \mathrm{vec}\, \Phi}{\partial \theta^{\mathrm{T}}} + H\Phi \frac{\partial x_t}{\partial \theta^{\mathrm{T}}} + \left(w_t^{\mathrm{T}} \otimes H\right)\frac{\partial \mathrm{vec}\, F}{\partial \theta^{\mathrm{T}}} + HF \frac{\partial w_t}{\partial \theta^{\mathrm{T}}} \right\}, \tag{45}$$

because the derivatives of the observations z_t with respect to θ are equal to 0. Notice that (45) is fully implementable because w_t is now the innovation, but the corresponding recurrences in Section 3 are not implementable. By taking the equations (37), (38), (40), (41), (42), (44) and (45) into account it can be seen that $\mathbb{E}\left\{\left(w_t^{\theta}\right)^{\mathrm{T}} \otimes w_t\right\} = 0$, $\mathbb{E}\left\{x_t^{\theta} \otimes w_t^{\mathrm{T}}\right\} = 0$, and $\mathbb{E}\left\{\widetilde{x}_t^{\theta} \otimes w_t^{\mathrm{T}}\right\} = 0$. By using the differential rule applied in the theorem, we formulate the derivatives of the Chandrasekhar equations which are necessary for solving the obtained recurrences.

We additionally apply the differential rule $\mathrm{d}(A^{-1}) = -A^{-1}(\mathrm{d}A)A^{-1}$.

$$\begin{aligned} B_t^{\theta} = {} & B_{t-1}^{\theta} + \left[\left(HY_{t-1}X_{t-1}^{\mathrm{T}}\right) \otimes H\right] Y_{t-1}^{\theta} + [(HY_{t-1}) \otimes (HY_{t-1})]X_{t-1}^{\theta} \\ & + [H \otimes (HY_{t-1}X_{t-1})]\left(Y_{t-1}^{\mathrm{T}}\right)^{\theta}, \end{aligned} \tag{46}$$

$$\begin{aligned} K_t^{\theta} = {} & \left[\left(B_t^{-1}B_{t-1}\right) \otimes I_n\right]K_{t-1}^{\theta} + \left[\left(B_t^{-1}HY_{t-1}X_{t-1}^{\mathrm{T}}Y_{t-1}^{\mathrm{T}}\right) \otimes I_n\right]\Phi^{\theta} \\ & + \left[B_t^{-1} \otimes K_{t-1}\right]B_{t-1}^{\theta} + \left[\left(B_t^{-1}HY_{t-1}X_{t-1}^{\mathrm{T}}\right) \otimes \Phi\right] Y_{t-1}^{\theta} \\ & - \left[B_t^{-1} \otimes \left(K_{t-1}B_{t-1}B_t^{-1}\right)\right]B_t^{\theta} + \left[\left(B_t^{-1}HY_{t-1}\right) \otimes \Phi Y_{t-1}\right]X_{t-1}^{\theta} \\ & + \left[\left(B_t^{-1}H\right) \otimes (\Phi Y_{t-1}X_{t-1})\right](Y_{t-1}^{\mathrm{T}})^{\theta} \\ & - \left[B_t^{-1} \otimes \left(\Phi Y_{t-1}X_{t-1}Y_{t-1}^{\mathrm{T}}H^{\mathrm{T}}B_t^{-1}\right)\right]B_t^{\theta}, \end{aligned} \tag{47}$$

$$\begin{aligned} Y_t^{\theta} = {} & \left[Y_{t-1}^{\mathrm{T}} \otimes I_n\right]\Phi^{\theta} + [I_m \otimes \Phi]Y_{t-1}^{\theta} - \left[\left(Y_{t-1}^{\mathrm{T}}H^{\mathrm{T}}\right) \otimes I_n\right]K_{t-1}^{\theta} \\ & - [I_n \otimes (K_{t-1}H)]Y_{t-1}^{\theta}, \end{aligned} \tag{48}$$

$$
\begin{aligned}
X_t^\theta = X_{t-1}^\theta &- \left[\left(X_{t-1}^{\mathrm{T}} Y_{t-1}^{\mathrm{T}} H^{\mathrm{T}} B_t^{-1} H Y_{t-1}\right) \otimes I_m\right] X_{t-1}^\theta \\
&- \left[\left(X_{t-1}^{\mathrm{T}} Y_{t-1}^{\mathrm{T}} H^{\mathrm{T}} B_t^{-1} H\right) \otimes X_{t-1}\right] \left(Y_{t-1}^{\mathrm{T}}\right)^\theta \\
&+ \left[\left(X_{t-1}^{\mathrm{T}} Y_{t-1}^{\mathrm{T}} H^{\mathrm{T}} B_t^{-1}\right) \otimes \left(X_{t-1} Y_{t-1}^{\mathrm{T}} H^{\mathrm{T}} B_t^{-1}\right)\right] B_t^\theta \\
&- \left[X_{t-1}^{\mathrm{T}} \otimes \left(X_{t-1} Y_{t-1}^{\mathrm{T}} H^{\mathrm{T}} B_t^{-1} H\right)\right] Y_{t-1}^\theta \\
&- \left[I_m \otimes \left(X_{t-1} Y_{t-1}^{\mathrm{T}} H^{\mathrm{T}} B_t^{-1} H Y_{t-1}\right)\right] X_{t-1}^\theta. \qquad (49)
\end{aligned}
$$

We then have the VARMA version of (32)

$$
\begin{aligned}
\mathbb{E}\left\{\tilde{z}_t^\theta \otimes \tilde{z}_t^\theta\right\}^{\mathrm{T}} = &\left[\mathbb{E}\left\{\tilde{x}_t^\theta \otimes \tilde{x}_t^\theta\right\}^{\mathrm{T}}\right] (H \otimes H)^{\mathrm{T}} + \mathbb{E}\left\{w_t^\theta \otimes w_t^\theta\right\}^{\mathrm{T}} \\
&+ \left[\mathbb{E}\left\{\tilde{x}_t^\theta \otimes w_t^\theta\right\}^{\mathrm{T}}\right] \left(H^{\mathrm{T}} \otimes I_m\right) \\
&+ \left[\mathbb{E}\left\{w_t^\theta \otimes \tilde{x}_t^\theta\right\}^{\mathrm{T}}\right] \left(I_m \otimes H^{\mathrm{T}}\right). \qquad (50)
\end{aligned}
$$

We can now give the recurrences, in the order they will be used in the computer program which is available for the interested reader

$$
\begin{aligned}
\mathbb{E}\left\{\tilde{x}_{t+1}^\theta \otimes \tilde{x}_{t+1}^\theta\right\}^{\mathrm{T}} = &\left\{\overline{\Phi}_t^\theta \otimes \overline{\Phi}_t^\theta\right\}^{\mathrm{T}} \left\{\left(M_{n^2,n}\left[\mathbb{E}\left\{\tilde{x}_t \otimes \tilde{x}_t\right\} \otimes I_n\right]\right) \otimes I_n\right\} \\
&+ \left\{\left(\overline{\Phi}_t^\theta\right)^{\mathrm{T}} \otimes I_\ell\right\} \left\{M_{n^2,\ell}\left(\left[\mathbb{E}\left\{\left(\tilde{x}_t^\theta\right)^{\mathrm{T}} \otimes \tilde{x}_t\right\} \overline{\Phi}_t^{\mathrm{T}}\right] \otimes I_n\right) M_{n,n}\right\} \\
&+ \left\{I_\ell \otimes \left(\overline{\Phi}_t^\theta\right)^{\mathrm{T}}\right\} \left\{\left[\mathbb{E}\left\{\left(w_t^\theta\right)^{\mathrm{T}} \otimes \tilde{x}_t\right\} \bar{F}_t^{\mathrm{T}}\right] \otimes I_n\right\} \\
&+ \left\{\left(\overline{\Phi}_t^\theta\right)^{\mathrm{T}} \otimes I_\ell\right\} \left\{M_{n^2,\ell}\left(\left[\mathbb{E}\left\{\left(w_t^\theta\right)^{\mathrm{T}} \otimes \tilde{x}_t\right\} \bar{F}_t^{\mathrm{T}}\right] \otimes I_n\right) M_{n,n}\right\} \\
&+ \left\{I_\ell \otimes \left(\overline{\Phi}_t^\theta\right)^{\mathrm{T}}\right\} \left(\left[\mathbb{E}\left\{\left(\tilde{x}_t^\theta\right)^{\mathrm{T}} \otimes \tilde{x}_t\right\} \overline{\Phi}_t^{\mathrm{T}}\right] \otimes I_n\right) \\
&+ \left\{\bar{F}_t^\theta \otimes \bar{F}_t^\theta\right\}^{\mathrm{T}} \left\{\left(M_{mn,m}\left[\operatorname{vec} Q \otimes I_n\right]\right) \otimes I_n\right\} \\
&+ \left[\mathbb{E}\left\{w_t^\theta \otimes w_t^\theta\right\}^{\mathrm{T}}\right] \left(\bar{F}_t \otimes \bar{F}_t\right)^{\mathrm{T}} + \left[\mathbb{E}\left\{\tilde{x}_t^\theta \otimes w_t^\theta\right\}^{\mathrm{T}}\right] \left(\overline{\Phi}_t \otimes \bar{F}_t\right)^{\mathrm{T}} \\
&+ \left[\mathbb{E}\left\{w_t^\theta \otimes \tilde{x}_t^\theta\right\}^{\mathrm{T}}\right] \left(\bar{F}_t \otimes \overline{\Phi}_t\right)^{\mathrm{T}} + E\left\{\tilde{x}_t^\theta \otimes \tilde{x}_t^\theta\right\}^{\mathrm{T}} \left(\overline{\Phi}_t \otimes \overline{\Phi}_t\right)^{\mathrm{T}},
\end{aligned}
\qquad (51)
$$

$$\begin{aligned}\mathbb{E}\left\{\left(\tilde{x}^{\theta}_{t+1}\right)^{\mathrm{T}}\otimes\tilde{x}_{t+1}\right\} &= \left\{\left(\overline{\Phi}^{\theta}_{t}\right)^{\mathrm{T}}\otimes\overline{\Phi}_{t}\right\}\left(M_{n^2,n}\left[\mathbb{E}\left\{\tilde{x}_t\otimes\tilde{x}_t\right\}\otimes I_n\right]\right)\\ &\quad+(I_\ell\otimes\overline{\Phi}_t)\mathbb{E}\left\{\left(\tilde{x}^{\theta}_{t}\right)^{\mathrm{T}}\otimes\tilde{x}_t\right\}\overline{\Phi}^{\mathrm{T}}_t+(I_\ell\otimes\overline{\Phi}_t)\mathbb{E}\left\{\left(w^{\theta}_{t}\right)^{\mathrm{T}}\otimes\tilde{x}_t\right\}\bar{F}^{\mathrm{T}}_t\\ &\quad+\left\{\left(\bar{F}^{\theta}_{t}\right)^{\mathrm{T}}\otimes\bar{F}_t\right\}(M_{mn,m}[\mathrm{vec}\,Q\otimes I_n]),\end{aligned}\tag{52}$$

$$\begin{aligned}\mathbb{E}\left\{x^{\theta}_{t+1}\otimes\tilde{x}^{\theta}_{t+1}\right\}^{\mathrm{T}} &= \left\{\Phi^{\theta}\otimes\overline{\Phi}^{\theta}_{t}\right\}^{\mathrm{T}}\left\{\left(M_{n^2,n}\left[\mathbb{E}\left\{\tilde{x}_t\otimes x_t\right\}\otimes I_n\right]\right)\otimes I_n\right\}\\ &\quad+\left\{\left(\Phi^{\theta}\right)^{\mathrm{T}}\otimes I_\ell\right\}\left(M_{n^2,\ell}\left[\mathbb{E}\left\{\left(\tilde{x}^{\theta}_{t}\right)^{\mathrm{T}}\otimes x_t\right\}\otimes I_n\right]M_{n,n}\right)\left(I_n\otimes\overline{\Phi}^{\mathrm{T}}_t\right)\\ &\quad+\left(\left\{I_\ell\otimes\left(\overline{\Phi}^{\theta}_{t}\right)^{\mathrm{T}}\right\}\left[\mathbb{E}\left\{\left(x^{\theta}_{t}\right)^{\mathrm{T}}\otimes\tilde{x}_t\right\}\otimes I_n\right]\right)\left(\Phi^{\mathrm{T}}\otimes I_n\right)\\ &\quad+\left[\mathbb{E}\left\{x^{\theta}_{t}\otimes\tilde{x}^{\theta}_{t}\right\}^{\mathrm{T}}\right]\left(\Phi\otimes\overline{\Phi}_t\right)^{\mathrm{T}}+\left[\mathbb{E}\left\{x^{\theta}_{t}\otimes w^{\theta}_{t}\right\}^{\mathrm{T}}\right]\left(\Phi\otimes\bar{F}_t\right)^{\mathrm{T}}\\ &\quad+\left\{F^{\theta}\otimes\bar{F}^{\theta}_{t}\right\}^{\mathrm{T}}\left\{(M_{nm,m}[\mathrm{vec}\,Q\otimes I_n])\otimes I_n\right\}\\ &\quad+\left[\mathbb{E}\left\{w^{\theta}_{t}\otimes\tilde{x}^{\theta}_{t}\right\}^{\mathrm{T}}\right]\left(F\otimes\overline{\Phi}_t\right)^{\mathrm{T}}+\left[\mathbb{E}\left\{w^{\theta}_{t}\otimes w^{\theta}_{t}\right\}^{\mathrm{T}}\right]\left(F\otimes\bar{F}_t\right)^{\mathrm{T}}\\ &\quad+\left\{\left(\Phi^{\theta}\right)^{\mathrm{T}}\otimes I_\ell\right\}\left\{M_{n^2,\ell}\left[\left(\mathbb{E}\left\{\left(w^{\theta}_{t}\right)^{\mathrm{T}}\otimes x_t\right\}\overline{F}^{\mathrm{T}}_t\right)\otimes I_n\right]M_{n,n}\right\}\\ &\quad+\left(\left\{I_\ell\otimes\left(\overline{\Phi}^{\theta}_{t}\right)^{\mathrm{T}}\right\}\left[\mathbb{E}\left\{\left(w^{\theta}_{t}\right)^{\mathrm{T}}\otimes\tilde{x}_t\right\}\otimes I_n\right]\right)\left(F^{\mathrm{T}}\otimes I_n\right),\end{aligned}\tag{53}$$

$$\begin{aligned}-\mathbb{E}\left\{w^{\theta}_{t+1}\otimes\tilde{x}^{\theta}_{t+1}\right\}^{\mathrm{T}} &= \left\{\Phi^{\theta}\otimes\overline{\Phi}^{\theta}_{t}\right\}^{\mathrm{T}}\left(\left[M_{n^2,n}\left\{\mathbb{E}\left(\tilde{x}_t\otimes x_t\right)\otimes H^{\mathrm{T}}\right\}\right]\otimes I_n\right)\\ &\quad+\left(M_{\ell,\ell}\left\{\left[I_\ell\otimes\left(\Phi^{\theta}\right)^{\mathrm{T}}\right]\left[\mathbb{E}\left\{\left(\tilde{x}^{\theta}_{t}\right)^{\mathrm{T}}\otimes x_t\right\}\otimes H^{\mathrm{T}}\right]\right\}M_{n,m}\right)\left(I_m\otimes\overline{\Phi}^{\mathrm{T}}_t\right)\\ &\quad+\left(M_{\ell,\ell}\left\{\left[I_\ell\otimes\left(\Phi^{\theta}\right)^{\mathrm{T}}\right]\left[\mathbb{E}\left\{\left(w^{\theta}_{t}\right)^{\mathrm{T}}\otimes x_t\right\}\otimes H^{\mathrm{T}}\right]\right\}M_{m,m}\right)\left(I_m\otimes F^{\mathrm{T}}\right)\\ &\quad+\left[I_\ell\otimes\left(\overline{\Phi}^{\theta}_{t}\right)^{\mathrm{T}}\right]\left(\left[\mathbb{E}\left\{\left(w^{\theta}_{t}\right)^{\mathrm{T}}\otimes\tilde{x}_t\right\}(HF)^{\mathrm{T}}\right]\otimes I_n\right)\\ &\quad+\left\{\left[I_\ell\otimes\left(\overline{\Phi}^{\theta}_{t}\right)^{\mathrm{T}}\right]\left[\mathbb{E}\left\{\left(x^{\theta}_{t}\right)^{\mathrm{T}}\otimes\tilde{x}_t\right\}\otimes I_n\right]\right\}\left(H\Phi\otimes I_n\right)^{\mathrm{T}}\\ &\quad+\left[\mathbb{E}\left\{x^{\theta}_{t}\otimes\tilde{x}^{\theta}_{t}\right\}^{\mathrm{T}}\right]\left(H\Phi\otimes\overline{\Phi}_t\right)^{\mathrm{T}}+\left[\mathbb{E}\left\{x^{\theta}_{t}\otimes w^{\theta}_{t}\right\}^{\mathrm{T}}\right]\left(H\Phi\otimes\bar{F}_t\right)^{\mathrm{T}}\\ &\quad+\left[\mathbb{E}\left\{w^{\theta}_{t}\otimes\tilde{x}^{\theta}_{t}\right\}^{\mathrm{T}}\right]\left(HF\otimes\overline{\Phi}_t\right)^{\mathrm{T}}+\left[\mathbb{E}\left\{w^{\theta}_{t}\otimes w^{\theta}_{t}\right\}^{\mathrm{T}}\right]\left(HF\otimes\bar{F}_t\right)^{\mathrm{T}}\\ &\quad+\left\{F^{\theta}\otimes\bar{F}^{\theta}_{t}\right\}^{\mathrm{T}}\left\{\left(M_{mn,n}\left[\mathrm{vec}\,Q\otimes H^{\mathrm{T}}\right]\right)\otimes I_n\right\},\end{aligned}\tag{54}$$

$$
\begin{aligned}
\mathbb{E}\left\{x_{t+1}^{\theta} \otimes x_{t+1}^{\theta}\right\}^{\mathrm{T}} = {} & \left\{\Phi^{\theta} \otimes \Phi^{\theta}\right\}^{\mathrm{T}}\left\{\left(M_{n^2,n}[\mathbb{E}\{x_t \otimes x_t\} \otimes I_n]\right) \otimes I_n\right\} \\
& + \left\{\left(\Phi^{\theta}\right)^{\mathrm{T}} \otimes I_{\ell}\right\}\left(M_{n^2,\ell}\left[\mathbb{E}\left\{\left(x_t^{\theta}\right)^{\mathrm{T}} \otimes x_t\right\} \otimes I_n\right]M_{n,n}\right)\left(I_n \otimes \Phi^{\mathrm{T}}\right) \\
& + \left\{\left(\Phi^{\theta}\right)^{\mathrm{T}} \otimes I_{\ell}\right\}\left(M_{n^2,\ell}\left[\mathbb{E}\left\{\left(w_t^{\theta}\right)^{\mathrm{T}} \otimes x_t\right\} \otimes I_n\right]M_{m,n}\right)\left(I_n \otimes F^{\mathrm{T}}\right) \\
& + \left[\mathbb{E}\left\{x_t^{\theta} \otimes x_t^{\theta}\right\}^{\mathrm{T}}\right](\Phi \otimes \Phi)^{\mathrm{T}} + \left[\mathbb{E}\left\{x_t^{\theta} \otimes w_t^{\theta}\right\}^{\mathrm{T}}\right](\Phi \otimes F)^{\mathrm{T}} \\
& + \left\{I_{\ell} \otimes \left(\Phi^{\theta}\right)^{\mathrm{T}}\right\}\left(\left[\mathbb{E}\left\{\left(x_t^{\theta}\right)^{\mathrm{T}} \otimes x_t\right\}\Phi^{\mathrm{T}}\right] \otimes I_n\right) \\
& + \left\{F^{\theta} \otimes F^{\theta}\right\}^{\mathrm{T}}\{(M_{nm,m}[\operatorname{vec} Q \otimes I_n]) \otimes I_n\} \\
& + \left\{I_{\ell} \otimes \left(\Phi^{\theta}\right)^{\mathrm{T}}\right\}\left[\mathbb{E}\left\{\left(w_t^{\theta}\right)^{\mathrm{T}} \otimes x_t\right\} \otimes I_n\right]\left(F^{\mathrm{T}} \otimes I_n\right) \\
& + \mathbb{E}\left\{w_t^{\theta} \otimes x_t^{\theta}\right\}^{\mathrm{T}}(F \otimes \Phi)^{\mathrm{T}} + \mathbb{E}\left\{w_t^{\theta} \otimes w_t^{\theta}\right\}^{\mathrm{T}}(F \otimes F)^{\mathrm{T}}, \qquad (55)
\end{aligned}
$$

$$
\begin{aligned}
-\mathbb{E}\left\{x_{t+1}^{\theta} \otimes w_{t+1}^{\theta}\right\}^{\mathrm{T}} = {} & \left\{\Phi^{\theta} \otimes \Phi^{\theta}\right\}^{\mathrm{T}}\left\{\left(M_{n^2,n}[\mathbb{E}\{x_t \otimes x_t\} \otimes I_n]\right) \otimes H^{\mathrm{T}}\right\} \\
& + \left\{\left(\Phi^{\theta}\right)^{\mathrm{T}} \otimes I_{\ell}\right\}\left(M_{n^2,\ell}\left[\mathbb{E}\left\{\left(x_t^{\theta}\right)^{\mathrm{T}} \otimes x_t\right\} \otimes I_n\right]M_{n,n}\right)(I_n \otimes H\Phi)^{\mathrm{T}} \\
& + \left\{\left(\Phi^{\theta}\right)^{\mathrm{T}} \otimes I_{\ell}\right\}\left(M_{n^2,\ell}\left[\mathbb{E}\left\{\left(w_t^{\theta}\right)^{\mathrm{T}} \otimes x_t\right\} \otimes I_n\right]M_{m,n}\right)(I_n \otimes HF)^{\mathrm{T}} \\
& + \left[\mathbb{E}\left\{x_t^{\theta} \otimes x_t^{\theta}\right\}^{\mathrm{T}}\right](\Phi \otimes H\Phi)^{\mathrm{T}} + \left[\mathbb{E}\left\{x_t^{\theta} \otimes w_t^{\theta}\right\}^{\mathrm{T}}\right](\Phi \otimes HF)^{\mathrm{T}} \\
& + \left\{I_{\ell} \otimes \left(\Phi^{\theta}\right)^{\mathrm{T}}\right\}\left(\left[\mathbb{E}\left\{\left(x_t^{\theta}\right)^{\mathrm{T}} \otimes x_t\right\}\Phi^{\mathrm{T}}\right] \otimes H^{\mathrm{T}}\right) \\
& + \left\{F^{\theta} \otimes F^{\theta}\right\}^{\mathrm{T}}\left\{(M_{mn,m}[\operatorname{vec} Q \otimes I_n]) \otimes H^{\mathrm{T}}\right\} \\
& + \left\{I_{\ell} \otimes \left(\Phi^{\theta}\right)^{\mathrm{T}}\right\}\left[\mathbb{E}\left\{\left(w_t^{\theta}\right)^{\mathrm{T}} \otimes x_t\right\} \otimes H^{\mathrm{T}}\right]\left(F^{\mathrm{T}} \otimes I_m\right) \\
& + \mathbb{E}\left\{w_t^{\theta} \otimes x_t^{\theta}\right\}^{\mathrm{T}}(F \otimes H\Phi)^{\mathrm{T}} + \mathbb{E}\left\{w_t^{\theta} \otimes w_t^{\theta}\right\}^{\mathrm{T}}(F \otimes HF)^{\mathrm{T}},
\end{aligned}
\qquad (56)
$$

$$
\begin{aligned}
\mathbb{E}\left\{w_{t+1}^{\theta} \otimes w_{t+1}^{\theta}\right\}^{\mathrm{T}} = {} & \left\{\Phi^{\theta} \otimes \Phi^{\theta}\right\}^{\mathrm{T}}\left(\left[M_{n^2,n}\left[\mathbb{E}\{x_t \otimes x_t\} \otimes H^{\mathrm{T}}\right]\right] \otimes H^{\mathrm{T}}\right) \\
& + \left\{\left(\Phi^{\theta}\right)^{\mathrm{T}} \otimes I_{\ell}\right\}\left\{M_{n^2,\ell}\left(\left[\mathbb{E}\left\{\left(x_t^{\theta}\right)^{\mathrm{T}} \otimes x_t\right\}\Phi^{\mathrm{T}}H^{\mathrm{T}}\right] \otimes H^{\mathrm{T}}\right)M_{m,m}\right\} \\
& + \left\{\left(\Phi^{\theta}\right)^{\mathrm{T}} \otimes I_{\ell}\right\}\left\{M_{n^2,\ell}\left(\left[\mathbb{E}\left\{\left(w_t^{\theta}\right)^{\mathrm{T}} \otimes x_t\right\}F^{\mathrm{T}}H^{\mathrm{T}}\right] \otimes H^{\mathrm{T}}\right)M_{m,m}\right\} \\
& + \left\{I_{\ell} \otimes \left(\Phi^{\theta}\right)^{\mathrm{T}}\right\}\left(\left[\mathbb{E}\left\{\left(x_t^{\theta}\right)^{\mathrm{T}} \otimes x_t\right\}\Phi^{\mathrm{T}}H^{\mathrm{T}}\right] \otimes H^{\mathrm{T}}\right) \\
& + \mathbb{E}\left\{x_t^{\theta} \otimes x_t^{\theta}\right\}^{\mathrm{T}}(H\Phi \otimes H\Phi)^{\mathrm{T}} + \mathbb{E}\left\{x_t^{\theta} \otimes w_t^{\theta}\right\}^{\mathrm{T}}(H\Phi \otimes HF)^{\mathrm{T}} \\
& + \left\{F^{\theta} \otimes F^{\theta}\right\}^{\mathrm{T}}\left\{\left(M_{mn,m}\left[\operatorname{vec} Q \otimes H^{\mathrm{T}}\right]\right) \otimes H^{\mathrm{T}}\right\} \\
& + \left\{I_{\ell} \otimes \left(\Phi^{\theta}\right)^{\mathrm{T}}\right\}\left(\left[\mathbb{E}\left\{\left(w_t^{\theta}\right)^{\mathrm{T}} \otimes x_t\right\}F^{\mathrm{T}}H^{\mathrm{T}}\right] \otimes H^{\mathrm{T}}\right) \\
& + \mathbb{E}\left[w_t^{\theta} \otimes x_t^{\theta}\right]^{\mathrm{T}}(HF \otimes H\Phi)^{\mathrm{T}} + \mathbb{E}\left[w_t^{\theta} \otimes w_t^{\theta}\right]^{\mathrm{T}}(HF \otimes HF)^{\mathrm{T}},
\end{aligned}
\qquad (57)
$$

$$
\begin{aligned}
\mathbb{E}\left\{\left(\tilde{x}_{t+1}^{\theta}\right)^{\mathrm{T}} \otimes x_{t+1}\right\} = {} & \left\{\left(\overline{\Phi}_t^{\theta}\right)^{\mathrm{T}} \otimes \Phi\right\}\left(M_{n,n^2}\left[\mathbb{E}\left\{x_t \otimes \tilde{x}_t\right\} \otimes I_n\right]\right) \\
& + \left\{\left(\bar{F}_t^{\theta}\right)^{\mathrm{T}} \otimes F\right\}\left(M_{nm,m}[\operatorname{vec} Q \otimes I_n]\right) \\
& + \left[(I_\ell \otimes \Phi)\mathbb{E}\left\{\left(\tilde{x}_t^{\theta}\right)^{\mathrm{T}} \otimes x_t\right\}\right]\overline{\Phi}_t^{\mathrm{T}} \\
& + \left[(I_\ell \otimes \Phi)\mathbb{E}\left\{\left(w_t^{\theta}\right)^{\mathrm{T}} \otimes x_t\right\}\right]\bar{F}_t^{\mathrm{T}}, \qquad (58)
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left\{\left(x_{t+1}^{\theta}\right)^{\mathrm{T}} \otimes \widetilde{x}_{t+1}\right\} = {} & \left\{\left(\Phi^{\theta}\right)^{\mathrm{T}} \otimes \overline{\Phi}_t\right\}\left(M_{n^2,n}[\mathbb{E}\{\widetilde{x}_t \otimes x_t\} \otimes I_n]\right) \\
& + \left(I_\ell \otimes \overline{\Phi}_t\right)\left[\mathbb{E}\left\{\left(x_t^{\theta}\right)^{\mathrm{T}} \otimes \widetilde{x}_t\right\}\Phi^{\mathrm{T}}\right] \\
& + \left(I_\ell \otimes \overline{\Phi}_t\right)\left[\mathbb{E}\left\{\left(w_t^{\theta}\right)^{\mathrm{T}} \otimes \widetilde{x}_t\right\}F^{\mathrm{T}}\right] \\
& + \left\{\left(F^{\theta}\right)^{\mathrm{T}} \otimes \overline{F}_t\right\}\left(M_{mn,m}[\operatorname{vec} Q \otimes I_n]\right), \qquad (59)
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left\{\left(x_{t+1}^{\theta}\right)^{\mathrm{T}} \otimes x_{t+1}\right\} = {} & \left\{\left(\Phi^{\theta}\right)^{\mathrm{T}} \otimes \Phi\right\}\left(M_{n^2,n}[\mathbb{E}\{x_t \otimes x_t\} \otimes I_n]\right) \\
& + (I_\ell \otimes \Phi)\left[\mathbb{E}\left\{\left(x_t^{\theta}\right)^{\mathrm{T}} \otimes x_t\right\}\Phi^{\mathrm{T}}\right] \\
& + (I_\ell \otimes \Phi)\left[\mathbb{E}\left\{\left(w_t^{\theta}\right)^{\mathrm{T}} \otimes x_t\right\}F^{\mathrm{T}}\right] \\
& + \left\{\left(F^{\theta}\right)^{\mathrm{T}} \otimes F\right\}\left(M_{mn,m}[\operatorname{vec} Q \otimes I_n]\right), \qquad (60)
\end{aligned}
$$

$$
\begin{aligned}
-\mathbb{E}\left\{\left(w_{t+1}^{\theta}\right)^{\mathrm{T}} \otimes x_{t+1}\right\} = {} & \left\{\left(\Phi^{\theta}\right)^{\mathrm{T}} \otimes \Phi\right\}\left(M_{n^2,n}\left[\mathbb{E}\{x_t \otimes x_t\} \otimes H^{\mathrm{T}}\right]\right) \\
& + (I_\ell \otimes \Phi)\left[\mathbb{E}\left\{\left(x_t^{\theta}\right)^{\mathrm{T}} \otimes x_t\right\}\Phi^{\mathrm{T}}H^{\mathrm{T}}\right] \\
& + (I_\ell \otimes \Phi)\left[\mathbb{E}\left\{\left(w_t^{\theta}\right)^{\mathrm{T}} \otimes x_t\right\}F^{\mathrm{T}}H^{\mathrm{T}}\right] \\
& + \left\{\left(F^{\theta}\right)^{\mathrm{T}} \otimes F\right\}\left(M_{mn,m}\left[\operatorname{vec} Q \otimes H^{\mathrm{T}}\right]\right), \qquad (61)
\end{aligned}
$$

$$
\begin{aligned}
-\mathbb{E}\left\{\left(w_{t+1}^{\theta}\right)^{\mathrm{T}} \otimes \tilde{x}_{t+1}\right\} = {} & \left\{\left(\Phi^{\theta}\right)^{\mathrm{T}} \otimes \overline{\Phi}_t\right\}\left(M_{n^2,n}\left[\mathbb{E}\left\{\tilde{x}_t \otimes x_t\right\} \otimes H^{\mathrm{T}}\right]\right) \\
& + \left(I_\ell \otimes \overline{\Phi}_t\right)\left[\mathbb{E}\left\{\left(x_t^{\theta}\right)^{\mathrm{T}} \otimes \tilde{x}_t\right\}\Phi^{\mathrm{T}} H^{\mathrm{T}}\right] \\
& + \left(I_\ell \otimes \overline{\Phi}_t\right)\left[\mathbb{E}\left\{\left(w_t^{\theta}\right)^{\mathrm{T}} \otimes \tilde{x}_t\right\}F^{\mathrm{T}} H^{\mathrm{T}}\right] \\
& + \left\{\left(F^{\theta}\right)^{\mathrm{T}} \otimes \bar{F}_t\right\}\left(M_{mn,m}\left[\operatorname{vec} Q \otimes H^{\mathrm{T}}\right]\right). \qquad (62)
\end{aligned}
$$

The last recurrences needed are

$$\mathbb{E}\left\{\tilde{x}_{t+1} \otimes \tilde{x}_{t+1}\right\} = \left\{\overline{\Phi}_t \otimes \overline{\Phi}_t\right\}\mathbb{E}\left\{\tilde{x}_t \otimes \tilde{x}_t\right\} + \left\{\bar{F}_t \otimes \bar{F}_t\right\}\operatorname{vec} Q, \qquad (63)$$

$$\mathbb{E}\left\{\tilde{x}_{t+1} \otimes x_{t+1}\right\} = \left\{\overline{\Phi}_t \otimes \Phi\right\}\mathbb{E}\left\{\tilde{x}_t \otimes x_t\right\} + \left\{\bar{F}_t \otimes F\right\}\operatorname{vec} Q, \qquad (64)$$

$$\mathbb{E}\{x_{t+1} \otimes x_{t+1}\} = \{\Phi \otimes \Phi\}\mathbb{E}\{x_t \otimes x_t\} + \{F \otimes F\}\operatorname{vec} Q. \qquad (65)$$

Note that (63) is a variant of the Riccati equation (33) which also appears in the Kalman filter. Its form is however simple because of the structure of $\overline{\Phi}_t$ and $\overline{F}_t$. Also, the following matrix is used: $\mathbb{E}\{w_t \otimes w_t\} = \operatorname{vec} Q$. The recurrences in this section require initial values which are presented in Section 5.

5. Initialization and computation of the likelihood

To start the computation of the relations (5–11), we need initial values. We have adapted the initialization given by [20] to our slightly different notation (his K_t is equivalent to our $K_t B_t$)

$$
\begin{aligned}
\widehat{x}_{1|0} &= 0, \\
B_1 &= H P_{1|0} H^{\mathrm{T}} + Q, \\
Y_1 &= \Phi P_{1|0} H^{\mathrm{T}} + FQ, \\
K_1 &= Y_1 B_1^{-1}, \\
X_1 &= -B_1^{-1}.
\end{aligned}
$$

$$(P_{1|0} H^{\mathrm{T}})^{(i)} = \sum_{j=i}^{p} \alpha_j \Gamma(j-i+1) - \sum_{i=1}^{h} \sum_{j=i}^{s} \beta_j \delta^{\mathrm{T}}(j-i+1), \qquad (66)$$

where

$$\delta(j) = -\beta_j Q + \sum_{i=1}^{p} \alpha_i \delta(j-i), \quad j = 1, \ldots, s \ (\delta(0) = Q),$$
$$\Gamma(j) = \sum_{i=1}^{p} \Gamma(j-i)\alpha_i^{\mathrm{T}} - \sum_{i=j}^{s} \delta(i-j)\beta_i^{\mathrm{T}}, \quad j = 0, \ldots, h.$$

Remark 1. $\Gamma(j)$ and $\partial \text{ vec } \Gamma(j)/\partial\theta^{\mathrm{T}}$ represent respectively, the covariance

$$\mathbb{E}\left\{z_t z_{t-j}^{\mathrm{T}}\right\}$$

and

$$\partial \text{ vec } \mathbb{E}\left\{z_t z_{t-j}^{\mathrm{T}}\right\}/\partial\theta^{\mathrm{T}}.$$

We also need the derivatives of the initial values of (66)

$$\frac{\partial \text{ vec}B_1}{\partial\theta^{\mathrm{T}}} = (I_m \otimes H)\frac{\partial \text{ vec}P_{1|0}H^{\mathrm{T}}}{\partial\theta^{\mathrm{T}}},$$

$$\frac{\partial \text{ vec}Y_1}{\partial\theta^{\mathrm{T}}} = (I_m \otimes \Phi)\frac{\partial \text{ vec}P_{1|0}H^{\mathrm{T}}}{\partial\theta^{\mathrm{T}}} + \left[\left(P_{1|0}H^{\mathrm{T}}\right)^{\mathrm{T}} \otimes I_n\right]\frac{\partial \text{ vec}\Phi}{\partial\theta^{\mathrm{T}}} + (Q \otimes I_n)\frac{\partial \text{ vec}F}{\partial\theta^{\mathrm{T}}},$$

$$\frac{\partial \text{ vec}K_1}{\partial\theta^{\mathrm{T}}} = \left(B_1^{-1} \otimes I_n\right)\frac{\partial \text{ vec}Y_1}{\partial\theta^{\mathrm{T}}} - (I_m \otimes Y_1)\left(B_1^{-1} \otimes B_1^{-1}\right)\frac{\partial \text{ vec}B_1}{\partial\theta^{\mathrm{T}}},$$

$$\frac{\partial \text{ vec}X_1}{\partial\theta^{\mathrm{T}}} = \left(B_1^{-1} \otimes B_1^{-1}\right)\frac{\partial \text{ vec}B_1}{\partial\theta^{\mathrm{T}}},$$

$$\begin{aligned}\frac{\partial \text{ vec}(P_{1|0}H^{\mathrm{T}})^{(i)}}{\partial\theta^{\mathrm{T}}} = &\sum_{j=i}^{p}\left\{\left(I_m \otimes \alpha_j\right)\frac{\partial \text{ vec}\Gamma(j-i+1)}{\partial\theta^{\mathrm{T}}} + \left(\Gamma^{\mathrm{T}}(j-i+1) \otimes I_m\right)\frac{\partial \text{ vec}\alpha_j}{\partial\theta^{\mathrm{T}}}\right\} \\ &- \sum_{i=1}^{h}\sum_{j=i}^{s}\left\{(\delta(j-i+1) \otimes I_m)\frac{\partial \text{ vec}\beta_j}{\partial\theta^{\mathrm{T}}} + \left(I_m \otimes \beta_j\right)\frac{\partial \text{ vec}\delta^{\mathrm{T}}(j-i+1)}{\partial\theta^{\mathrm{T}}}\right\},\end{aligned}$$

where

$$\frac{\partial\ \mathrm{vec}\delta(j)}{\partial\theta^{\mathrm{T}}} = -(Q\otimes I_m)\frac{\partial\ \mathrm{vec}\beta_j}{\partial\theta^{\mathrm{T}}} + \sum_{i=1}^{p}\left\{\left(\delta^{\mathrm{T}}(j-i)\otimes I_m\right)\frac{\partial\ \mathrm{vec}\alpha_j}{\partial\theta^{\mathrm{T}}} + \left(I_m\otimes\alpha_j\right)\frac{\partial\ \mathrm{vec}\delta(j-i)}{\partial\theta^{\mathrm{T}}}\right\},$$

$$j = 1,\ldots,s$$

$$\frac{\partial\ \mathrm{vec}\Gamma(j)}{\partial\theta^{\mathrm{T}}} = \sum_{i=1}^{p}\left\{(\alpha_i\otimes I_m)\frac{\partial\ \mathrm{vec}\,\Gamma(j-i)}{\partial\theta^{\mathrm{T}}} + (I_m\otimes\Gamma(j-i))\frac{\partial\ \mathrm{vec}\alpha_i^{\mathrm{T}}}{\partial\theta^{\mathrm{T}}}\right\} - \sum_{i=j}^{s}\left\{(\beta_i\otimes I_m)\frac{\partial\ \mathrm{vec}\,\delta(i-j)}{\partial\theta^{\mathrm{T}}} + (I_m\otimes\delta(i-j))\frac{\partial\ \mathrm{vec}\beta_i^{\mathrm{T}}}{\partial\theta^{\mathrm{T}}}\right\},$$

$j = 1,\ldots,p$ and recursively for $j = p+1,\ldots,h$.

Remark 2. We have to adapt some expressions because of vectorization. For example, for $\partial\ \mathrm{vec}\,P_{1|0}H^{\mathrm{T}}/\partial\theta^{\mathrm{T}}$, we know that $P_{1|0}H^{\mathrm{T}}$ is a block matrix with the following structure

$$\underset{(mh\times m)}{P_{1|0}H^{\mathrm{T}}} = \begin{bmatrix} \underset{(m\times m)}{(P_{1|0}H^{\mathrm{T}})^{(1)}} \\ \cdots \\ \underset{(m\times m)}{(P_{1|0}H^{\mathrm{T}})^{(h)}} \end{bmatrix},$$

so if the vectorization is used

$$\underset{(m^2h\times 1)}{\mathrm{vec}\,P_{1|0}H^{\mathrm{T}}} = \underset{(m^2h\times m^2h)}{M_{h,m}^{b}} \begin{bmatrix} \underset{(m^2\times 1)}{\mathrm{vec}\,(P_{1|0}H^{\mathrm{T}})^{(1)}} \\ \cdots \\ \underset{(m^2\times 1)}{\mathrm{vec}\,(P_{1|0}H^{\mathrm{T}})^{(h)}} \end{bmatrix},$$

where $M_{h,m}^{b}$ is a block-permutation matrix, where the (i,j)th block is given by

$$\left[M_{h,m}^{b}\right]^{ij} = \begin{cases} I_m & \text{for } i = 1,h \text{ and } j = 1+(i-1)m \bmod[hm-1], \\ 0_m & \text{otherwise.} \end{cases} \tag{67}$$

We can now write

$$\frac{\partial\ \mathrm{vec} P_{1|0} H^{\mathrm{T}}}{\partial \theta^{\mathrm{T}}} = M^{b}_{h,m} \begin{bmatrix} \dfrac{\partial\ \mathrm{vec}(P_{1|0} H^{\mathrm{T}})^{(1)}}{\partial \theta^{\mathrm{T}}} \\ \cdots \\ \dfrac{\partial\ \mathrm{vec}(P_{1|0} H^{\mathrm{T}})^{(h)}}{\partial \theta^{\mathrm{T}}} \end{bmatrix}.$$

This matrix $M^{b}_{h,m}$ is used also in order to compute other elements for example $\left(\partial\ \mathrm{vec} F/\partial\theta^{\mathrm{T}}\right), \left(\partial\ \mathrm{vec} \Phi/\partial\theta^{\mathrm{T}}\right)$.

Remark 3. For computing $\left(\partial\ \mathrm{vec} Y_t^{\mathrm{T}}/\partial\theta^{\mathrm{T}}\right)$, where Y_t^{T}, is a block matrix with the following structure

$$\underset{(h\times hm)}{Y_t^{\mathrm{T}}} = \left[\underset{(m\times m)}{(Y_t)^{(1)T}} \cdots \underset{(m\times m)}{(Y_t)^{(h)T}} \right],$$

we introduce an element-permutation matrix called $M^{\mathrm{e}}_{h,m}$, such that

$$\underset{(m^2h\times 1)}{\mathrm{vec}\ Y_t^{\mathrm{T}}} = \underset{(m^2h\times m^2h)}{M^{\mathrm{e}}_{h,m}} \underset{(m^2h\times 1)}{\mathrm{vec}\ Y_t},$$

where the element $\left(M^{\mathrm{e}}_{h,m}\right)_{ij}$ follows

$$\left(M^{\mathrm{e}}_{h,m}\right)_{ij} = \begin{cases} 1 & \text{for } i = 1, m^2h \text{ and } j = 1 + (i-1)m \ \mathrm{mod}[m^2h - 1] \\ 0 & \text{otherwise} \end{cases}. \tag{68}$$

We can also write

$$\frac{\partial\ \mathrm{vec} Y_t^{\mathrm{T}}}{\partial \theta^{\mathrm{T}}} = M^{\mathrm{e}}_{h,m} \frac{\partial\ \mathrm{vec} Y_t}{\partial \theta^{\mathrm{T}}}.$$

Remark 4. We have also to compute the initial values of the recurrence expressions (51–65)

$$\mathbb{E}\left\{\tilde{z}_1^{\theta} \otimes \tilde{z}_1^{\theta}\right\}^{\mathrm{T}}, \mathbb{E}\left\{\left(\tilde{x}_1^{\theta}\right)^{\mathrm{T}} \otimes \tilde{x}_1\right\}, \mathbb{E}\left\{\tilde{x}_1^{\theta} \otimes \tilde{x}_1^{\theta}\right\}^{\mathrm{T}}, \mathbb{E}\left\{w_1^{\theta} \otimes \tilde{x}_1^{\theta}\right\}^{\mathrm{T}},$$

$$\mathbb{E}\{\tilde{x}_1 \otimes x_1\}, \mathbb{E}\left\{\left(\tilde{x}_1^{\theta}\right)^{\mathrm{T}} \otimes x_1\right\}, \mathbb{E}\left\{\left(x_1^{\theta}\right)^{\mathrm{T}} \otimes \tilde{x}_1\right\}, \mathbb{E}\left\{x_1^{\theta} \otimes \tilde{x}_1^{\theta}\right\}^{\mathrm{T}},$$

$$\mathbb{E}\left\{w_1^{\theta} \otimes w_1^{\theta}\right\}^{\mathrm{T}}, \mathbb{E}\{x_1 \otimes x_1\}, \mathbb{E}\left\{\left(x_1^{\theta}\right)^{\mathrm{T}} \otimes x_1\right\}, \mathbb{E}\left\{\left(w_1^{\theta}\right)^{\mathrm{T}} \otimes x_1\right\},$$

$$\mathbb{E}\left\{x_1^{\theta} \otimes x_1^{\theta}\right\}^{\mathrm{T}}, \mathbb{E}\left\{x_1^{\theta} \otimes w_1^{\theta}\right\}^{\mathrm{T}}, \mathbb{E}\left\{\left(w_1^{\theta}\right)^{\mathrm{T}} \otimes \tilde{x}_1\right\}, \mathbb{E}\{\tilde{x}_1 \otimes \tilde{x}_1\}.$$

The initial values of the Chandrasekhar relations are

$$\widehat{x}_{1|0} = 0, \quad x_0 = 0.$$

By using the Eqs. (9) and (11), we have

$$\widetilde{x}_1 = x_1 = Fw_0,$$

$$\frac{\partial \widetilde{x}_1}{\partial \theta^{\mathrm{T}}} = \frac{\partial x_1}{\partial \theta^{\mathrm{T}}} = \left(w_0^{\mathrm{T}} \otimes I_n\right) \frac{\partial \ \mathrm{vec}\, F}{\partial \theta^{\mathrm{T}}},$$

$$\frac{\partial w_1}{\partial \theta^{\mathrm{T}}} = -H\left\{ \left(w_0^{\mathrm{T}} \otimes I_n\right) \frac{\partial \ \mathrm{vec}\, F}{\partial \theta^{\mathrm{T}}} \right\}.$$

We know that $\mathbb{E}\{w_1 \otimes w_1\} = \mathrm{vec}\, Q$. We now give all the initial values we need for the expressions (51–65).

$$\mathbb{E}[x_1 \otimes x_1] = \mathbb{E}\{\widetilde{x}_1 \otimes x_1\} = \mathbb{E}\{\widetilde{x}_1 \otimes \widetilde{x}_1\} = (F \otimes F)\mathrm{vec}\, Q,$$

$$\begin{aligned}\mathbb{E}\left\{ \left(\widetilde{x}_1^{\theta}\right)^{\mathrm{T}} \otimes \widetilde{x}_1 \right\} &= \mathbb{E}\left\{ \left(\widetilde{x}_1^{\theta}\right)^{\mathrm{T}} \otimes x_1 \right\} = \mathbb{E}\left\{ \left(x_1^{\theta}\right)^{\mathrm{T}} \otimes \widetilde{x}_1 \right\} = \mathbb{E}\left\{ \left(x_1^{\theta}\right)^{\mathrm{T}} \otimes x_1 \right\} \\ &= \left\{ \left(\mathrm{vec}\, F^{\theta}\right)^{\mathrm{T}} \otimes F \right\} (M_{mn,m}[\mathrm{vec}\, Q \otimes I_n]),\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left\{ \left(w_1^{\theta}\right)^{\mathrm{T}} \otimes x_1 \right\} &= \mathbb{E}\left\{ \left(w_1^{\theta}\right)^{\mathrm{T}} \otimes \widetilde{x}_1 \right\} \\ &= -\left[\left\{ \left(\mathrm{vec}\, F^{\theta}\right)^{\mathrm{T}} \otimes F \right\} (M_{mn,m}[\mathrm{vec}\, Q \otimes I_n]) \right] H^{\mathrm{T}},\end{aligned}$$

$$\begin{aligned}\mathbb{E}\left\{ \widetilde{x}_1^{\theta} \otimes \widetilde{x}_1^{\theta} \right\}^{\mathrm{T}} &= \mathbb{E}\left\{ x_1^{\theta} \otimes \widetilde{x}_1^{\theta} \right\}^{\mathrm{T}} = \mathbb{E}\left\{ x_1^{\theta} \otimes x_1^{\theta} \right\}^{\mathrm{T}} \\ &= \left\{ \mathrm{vec}\, F^{\theta} \otimes \mathrm{vec}\, F^{\theta} \right\}^{\mathrm{T}} \{ (M_{mn,m}[\mathrm{vec}\, Q \otimes I_n]) \otimes I_n \},\end{aligned}$$

$$\mathbb{E}\left\{ x_1^{\theta} \otimes w_1^{\theta} \right\}^{\mathrm{T}} = -\left\{ \mathrm{vec}\, F^{\theta} \otimes \mathrm{vec}\, F^{\theta} \right\}^{\mathrm{T}} \{ (M_{mn,m}[\mathrm{vec}\, Q \otimes I_n]) \otimes I_n \} (I_n \otimes H)^{\mathrm{T}},$$

$$\mathbb{E}\left\{ w_1^{\theta} \otimes \widetilde{x}_1^{\theta} \right\}^{\mathrm{T}} = -\left\{ \mathrm{vec}\, F^{\theta} \otimes \mathrm{vec}\, F^{\theta} \right\}^{\mathrm{T}} \{ (M_{mn,m}[\mathrm{vec}\, Q \otimes I_n]) \otimes I_n \} (H \otimes I_n)^{\mathrm{T}},$$

$$\mathbb{E}\left\{ w_1^{\theta} \otimes w_1^{\theta} \right\}^{\mathrm{T}} = \left\{ \mathrm{vec}\, F^{\theta} \otimes \mathrm{vec}\, F^{\theta} \right\}^{\mathrm{T}} \{ (M_{mn,m}[\mathrm{vec}\, Q \otimes I_n]) \otimes I_n \} (H \otimes H)^{\mathrm{T}}.$$

Finally, also using these initial values, the logarithm of the likelihood is computed by

$$\frac{\partial l(\theta)}{\partial \theta^{\mathrm{T}}} = -\frac{\partial \log L(\theta)}{\partial \theta^{\mathrm{T}}} = \sum_{t=1}^{N} \left\{ \frac{1}{2} \frac{\partial \log |B_t|}{\partial \theta^{\mathrm{T}}} + \frac{1}{2} \frac{\partial \left(\tilde{z}_t^{\mathrm{T}} B_t^{-1} \tilde{z}_t \right)}{\partial \theta^{\mathrm{T}}} \right\}$$

$$= \sum_{t=1}^{N} \frac{1}{2} \left(\mathrm{vec}\, B_t^{-T} \right)^{\mathrm{T}} \frac{\partial \mathrm{vec}\, B_t}{\partial \theta^{\mathrm{T}}} + \frac{1}{2} \left\{ \left[B_t^{-1} \tilde{z}_t \right]^{\mathrm{T}} \frac{\partial \tilde{z}_t}{\partial \theta^{\mathrm{T}}} \right.$$

$$\left. \times \tilde{z}_t^{\mathrm{T}} \left[B_t^{-1} \frac{\partial \tilde{z}_t}{\partial \theta^{\mathrm{T}}} - (\tilde{z}_t^{\mathrm{T}} \otimes I_m)(B_t^{-1} \otimes B_t^{-1}) \frac{\partial B_t}{\partial \theta^{\mathrm{T}}} \right] \right\},$$

where we use

$$\frac{\partial \tilde{z}_t}{\partial \theta^{\mathrm{T}}} = -\frac{\partial \widehat{z}_{t|t-1}}{\partial \theta^{\mathrm{T}}} = -H \frac{\partial \widehat{x}_{t|t-1}}{\partial \theta^{\mathrm{T}}}.$$

6. Conclusion

With respect to the asymptotic information matrix, severe algebraic problems are raised by the efficient computation, using closed form recurrence relations, of the exact information matrix of a Gaussian linear model in state space form. Contrary to the literature, we have chosen to evaluate the information matrix as a whole, not element-wise. Computational efficiency is obtained by the use of the Chandrasekhar equations instead of the Kalman filter recurrences, provided that sparsity of Φ and H is taken into account. The model is general enough so that there is no need to specify the parameters. We have however treated the VARMA model as a special case, where the parameters are the elements of the coefficients of the autoregressive and moving average polynomials.

This has been made possible by applying appropriate matrix differential rules, combined with Kronecker products and vectorization of matrices. It can be seen as an illustration of the assertions put forth by [23] and [17] about matrix differentiation concepts, for the former, and about the relation between the algebra of Kronecker products and the abstract algebra of tensor analysis, for the latter. It is also a successful application of the [10] approach. The equations obtained cannot easily be simplified except that some common factors can still be found but removing them would produce less appealing equations.

The algorithm in the VARMA case programmed in the MATLAB environment is available from the authors. It can be translated in any other matrix environment where the Kronecker product is defined.

Acknowledgements

The authors wish to thank Heinz Neudecker, two anonymous referees and the editor Simo Puntanen for helpful and valuable comments and suggestions.

References

[1] B.D.O. Anderson, J.B. Moore, Optimal Filtering, Prentice-Hall, Englewood Cliffs, New Jersey, 1979.
[2] B. Friedlander, On the computation of the Cramér–Rao bound for ARMA parameter estimation, IEEE Trans. Acoust. Speech, Signal Processing 32 (1984) 721–727.
[3] E.J. Hannan, M. Deistler, The Statistical Theory of Linear Systems, Wiley, New York, 1988.
[4] A.C. Harvey, Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge University Press, Cambridge, 1989.
[5] R.E. Kalman, A new approach to linear filtering and predicting problems, Transactions of the ASME, J. Basic Engineering 82D (1960) 35–45.
[6] A. Klein, G. Mélard, The information matrix of multiple input single output time series models, J. Comput. Appl. Math. 51 (1994) 349–356.
[7] A. Klein, G. Mélard, T. Zahaf, Computation of the exact information matrix of Gaussian dynamic regression time series models, Ann. Statist. 26 (1998) 1636–1650.
[8] A. Klein, P. Spreij, On Fisher's information matrix of an ARMAX process and Sylvester's resultant matrices, Linear Algebra Appl. 237/238 (1996) 579–590.
[9] P. Lancaster, M. Tismenetsky, The Theory of Matrices with Applications, 2nd ed., Academic Press, Orlando, 1985.
[10] J.R. Magnus, H. Neudecker, Matrix Differential Calculus with Applications in Statistics and Econometrics, Wiley, New York, 1988.
[11] G. Mélard, A. Klein, On a fast algorithm for the exact information matrix of a Gaussian ARMA time series, IEEE Trans. Signal Processing 42 (1994) 2201–2203.
[12] S. Mittnik, P.A. Zadrozny, Asymptotic distributions of impulse responses, step responses, and variance decompositions of estimated linear dynamic model, Econometrica 61 (1993) 857–870.
[13] M. Morf, G.S. Sidhu, T. Kailath, Some new algorithms for recursive estimation on constant, linear, discrete-time systems, IEEE Trans. Automat. Contr. 19 (1974) 315–323.
[14] H. Neudecker, Some theorems on matrix differentiation with special reference to Kronecker matrix products, J. Amer. Statist. Assoc. 64 (1969) 953–963.
[15] H. Neudecker, T. Wansbeek, Some results on commutation matrices, with statistical applications, Canad. J. Statist. 11 (1983) 221–231.
[16] H.J. Newton, The information matrices of the parameters of multiple mixed time series, J. Multivariate Anal. 8 (1978) 317–323.
[17] D.S.G. Pollock, Tensor products and matrix differential calculus, Linear Algebra Appl. 67 (1985) 169–193.
[18] B. Porat, B. Friedlander, Computation of the exact information matrix of Gaussian time series with stationary random components, IEEE Trans. Acoust., Speech, Signal Processing 14 (1986) 118–130.
[19] F.C. Schweppe, Evaluation of likelihood functions for Gaussian signals, IEEE Trans. Inf. Theor. 11 (1965) 61–70.
[20] B.L. Shea, The exact likelihood of a vector autoregressive moving average model, J. Royal Statist. Soc. Ser. C 38 (1989) 161–184.
[21] J. Terceiro Lomba, Estimation of Dynamic Econometric Models with Errors in Variables, Springer, Berlin, 1990.

[22] P. Whittle, Estimation and information in time series, Ark. Math. Fys. Astr. 2 (1953) 423–434.
[23] C.S. Wong, On the use of differentials in statistics, Linear Algebra Appl. 70 (1985) 285–299.
[24] P.A. Zadrozny, Analytical derivatives for estimation of linear dynamic models, Comput. Math. Appl. 18 (1989) 539–553.
[25] P.A. Zadrozny, Errata to Analytical derivatives for estimation of linear dynamic models, Comput. Math. Appl. 24 (1992) 289–290.
[26] P.A. Zadrozny, S. Mittnik, Kalman filtering methods for computing information matrices for time-invariant periodic and generally time-varying VARMA models and samples, Comput. Math. Appl. 28 (1994) 107–119.