Exact Gradient Vector of Loglikelihood for Linear Gaussian State Space Model

(A former title is "A recursion formula for calculating the exact gradient vector of the loglikelihood for linear Gaussian state space models")

Daisuke Nagakura
Faculty of Economics, Keio University,

Abstract

Linear Gaussian state space models have been widely used in a variety of fields. In estimation and testing of these state space models, the gradient vector of the log-likelihood plays important roles. One usually calculates an approximate value of the gradient vector by numerically differentiating the log-likelihood. In this paper, we propose a recursive formula for computing the exact value of the gradient vector for a general form of linear Gaussian state space models. We also explicitly consider how to handle the initial condition in the context of computing the exact gradient vector when state variables are stationary, which is the issue that has not been considered in the the existing literature. We give two examples to illustrate how to apply the proposed formula.

Key Words: State Space Model; Gradient Vector; Initial Condition; Commutation Matrix; Quasi Maximum Likelihood Estimation.

1 Introduction

Linear Gaussian state space models have been widely used in a variety of fields. Statistical analyses, such as estimation and testing, of these state space models often require the calculation of the gradient vector, or the vector of the first partial derivatives, of the log-likelihood for these models with respect to unknown parameters. One may calculate an approximate value of the gradient vector by numerically differentiating the log-likelihood. However, in some statistical analyses, it is more desirable to obtain the exact value of the gradient vector, or the exact gradient vector, of the log-likelihood for these state space models.

There are some papers, such as Harvey [2], Hooker [3], Koopman and Shephard [4], and Segal and Weinstein [8], that propose formulas for calculating the exact gradient vector. It seems that these formulas are not fully appreciated despite the importance of the gradient vector in statistical analyses. One possible reason behind this is that coding these formulas are somewhat cumbersome and some of them work only under restrictive situations, which limits the practical usefulness of these formulas. In this paper, we aim at providing a new formula, that is easy to code and works in a very general situation, for calculating (not an approximate but) the exact gradient vector of the log-likelihood for a general form of linear Gaussian state space models.

We derive a new formula similar to the formula in Harvey [2]. The formula in Harvey [2] is for calculating only one component of the gradient vector in a single pass. This implies that, to calculate all components of the gradient vector, the formula in Harvey [2] needs to be applied as many times as the number of different components of the gradient vector. The other existing formulas mentioned above also share this characteristic.[1] A major difference between the formulas in Harvey [2] and this paper is that the proposed formula computes all components of the gradient vector in a single pass, which simplifies the coding significantly, in particular, for some matrix computing languages. Furthermore, the formula itself is of some theoretical interest because it explicitly shows how all components of the gradient vector are calculated at specific values of the unknown parameters.

Although our emphasis is not a reduction of computational time achieved by the proposed formula, our simple computational experiment suggests the superiority of the proposed formula to the formula in Harvey [2] in terms of total computational time of calculating all components of the gradient vector. We also consider how to handle the initial condition in the context of computing the exact gradient vector when state variables are stationary, which is the issue that has not been addressed explicitly in the existing literature.

The rest of the paper proceeds as follows. In Section 2, we introduce a general form of linear Gaussian state space models. In Section 3, we present a recursive formula for calculating the exact gradient vector and consider the problem of the initial condition. In Section 4, we provide two examples for illustrating how the formula is applied. In Section 5, we conduct a simple computational experiment to assess the computational time of the proposed formula relative to the formula in Harvey [2]. The last section gives some concluding remarks. In Appendix A, we briefly review some matrix operators, their properties, and formulas of matrix calculus used in this paper. In Appendix B, we give mathematical proofs of Propositions.

2 State Space Model

Let $\mathbf{y}_t$ and $\boldsymbol{\alpha}_t$, for $t = 1, ..., n$, be $v \times 1$ observation and $m \times 1$ state vectors at time t, respectively. We consider the following linear Gaussian state space model.

$$\begin{array}{lll} \mathbf{y}_t = \mathbf{d}_t + \mathbf{Z}_t\boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t, & \boldsymbol{\varepsilon}_t \sim \text{i.i.d.}N(\mathbf{0}, \mathbf{S}_t), & \\ \boldsymbol{\alpha}_{t+1} = \mathbf{c}_{t+1} + \mathbf{T}_{t+1}\boldsymbol{\alpha}_t + \mathbf{R}_{t+1}\boldsymbol{\eta}_{t+1}, & \boldsymbol{\eta}_s \sim \text{i.i.d.}N(\mathbf{0}, \mathbf{Q}_s), & \boldsymbol{\alpha}_1 \sim N(\mathbf{a}_0, \mathbf{P}_0), \\ E(\boldsymbol{\varepsilon}_t\boldsymbol{\eta}_s') = \mathbf{0}_{v\times w}, \quad E(\boldsymbol{\varepsilon}_t\boldsymbol{\alpha}_1') = \mathbf{0}_{v\times m}, & E(\boldsymbol{\eta}_s\boldsymbol{\alpha}_1') = \mathbf{0}_{w\times m}, & \text{for } t = 1, ..., n, s = 2, ..., n+1, \end{array} \tag{1}$$

where $\mathbf{d}_t$, $\mathbf{Z}_t$, $\mathbf{S}_t$, $\mathbf{c}_s$, $\mathbf{T}_s$, $\mathbf{R}_s$ and $\mathbf{Q}_s$ (hereafter, we call these matrices *system matrices*) are $v \times 1$, $v \times m$, $v \times v$, $m \times 1$, $m \times m$, $m \times w$ and $w \times w$ matrices, respectively, $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\eta}_s$ are $v \times 1$ and $w \times 1$ vectors of unobserved errors, respectively, and $\mathbf{0}_{k\times j}$ denotes the $k \times j$ matrix whose elements are all zero. Additionally, $\mathbf{I}_k$, which appears later, denotes the $k \times k$ identity matrix. The first equation is called the *observation equation* and the second equation is called the *state equation*. We note that $\mathbf{d}_t$ and $\mathbf{c}_t$ can be exogenous or

[1] Moreover, one usually has to alter the code for calculating different components in the gradient vector, which makes the coding these formulas very cumbersome as the number of components increases.

predetermined inputs into the state and observation equations. For example, $\mathbf{d}_t$ may be set as $\mathbf{d}_t = \mathbf{D}\mathbf{X}_t$, where $\mathbf{D}$ is a constant coefficient matrix and $\mathbf{X}_t$ is a matrix of exogenous or predetermined variables.

We assume that each element of the system matrices is a function of an unknown $h \times 1$ parameter vector $\boldsymbol{\theta} = (\theta_1, ..., \theta_h)' \in \Theta \subseteq \mathbb{R}^h$, the functional form at time t is known before time t, and $\mathbf{S}_t$ and $\mathbf{Q}_s$ are positive semidefinite at any point of $\boldsymbol{\theta} \in \Theta$. We allow the functional forms at different time points to be different. Under the Gaussian assumption on $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\eta}_s$, the joint density of $\mathbf{Y}_n \equiv (\mathbf{y}_n', \cdots \mathbf{y}_1')'$ is also Gaussian, and one can calculate the exact log-likelihood applying the well known Kalman filter. A statistical analysis, such as estimation and testing, regarding the unknown parameter vector $\boldsymbol{\theta}$, often requires the calculation of the gradient vector of the log likelihood. One may obtain an approximate gradient vector by numerically differentiating the log-likelihood. However, depending on the situation, it is more desirable to use the exact gradient vector than approximate ones.

In view of this, Harvey [2], Hooker [3], Koopman and Shephard [4], and Segal and Weinstein [8] among others, propose formulas for calculating the exact gradient vector of the log-likelihood for linear Gaussian state space models. They assume different state space models, all of which are included as special cases of the general form given in (1). These formulas are for calculating only one component of the gradient vector, and hence one needs to repeatedly apply their formulas h times to obtain all components of the gradient vector.[2] In contrast, the formula we propose in the next section can calculate all components of the gradient vector in a single pass. In addition to this, unlike the formulas of Hooker [3], Koopman and Shephard [4], and Segal and Weinstein [8], the proposed formula does not have to assume either time-invariant system matrices or non-singularities of $\mathbf{S}_t$ and $\mathbf{R}_t\mathbf{Q}_t\mathbf{R}_t'$. Our simple computational experiment suggests that the proposed formula takes less total computational time than the formula in Harvey [2] for calculating all components of the gradient vector.

3 Exact Gradient Vector

In this section, we propose a recursive formula for calculating the exact gradient vector of the log-likelihood for the linear Gaussian state space model in (1). In the subsequent section, first, we consider the case in that system matrices can vary over time. In this case, the state vector $\boldsymbol{\alpha}_t$ need not be stationary and initial condition should be set by researchers accordingly. Next, in Section 3.2, we consider how to handle the initial conditions when system matrices are time-invariant and the state vector $\boldsymbol{\alpha}_t$ is stationary. In this case, it is natural to set the initial conditions as in (18), and we derive the derivative matrices for those initial conditions.

3.1 A recursive formula for the exact gradient vector

Suppose that $\mathbf{y}_t$ is generated by the system given in (1). Define $\mathbf{Y}_t \equiv (\mathbf{y}_t', ..., \mathbf{y}_1')'$. Let $p(\mathbf{y}_t|\mathbf{Y}_u)$ denote the conditional pdf of $\mathbf{y}_t$ conditional on $\mathbf{Y}_u$. Then, L_n, the log likelihood of $\mathbf{Y}_n$, is defined as $L_n \equiv \sum_{t=1}^n \ell_t$, where $\ell_t \equiv \log p(\mathbf{y}_t|\mathbf{Y}_{t-1})$ and $p(\mathbf{y}_1|\mathbf{Y}_0) \equiv p(\mathbf{y}_1)$. Under the assumption of Gaussian errors, the conditional distribution of $\mathbf{y}_t$ conditional on $\mathbf{Y}_{t-1}$ is also Gaussian. Let $\mathbf{a}_{t|u} \equiv E(\boldsymbol{\alpha}_t|\mathbf{Y}_u)$ and $\mathbf{P}_{t|u} \equiv \mathrm{var}(\boldsymbol{\alpha}_t|\mathbf{Y}_u)$. Then, ℓ_t is given by

$$\ell_t = -\frac{v}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{F}_t| - \frac{1}{2}\left(\mathbf{v}_t'\mathbf{F}_t^{-1}\mathbf{v}_t\right), \tag{2}$$

where $\mathbf{v}_t = \mathbf{y}_t - \mathbf{d}_t - \mathbf{Z}_t\mathbf{a}_{t|t-1}$ and $\mathbf{F}_t = \mathbf{Z}_t\mathbf{P}_{t|t-1}\mathbf{Z}_t' + \mathbf{S}_t$. One can calculate $\mathbf{a}_{t|t-1}$ and $\mathbf{P}_{t|t-1}$ by running the Kalman filter, which is given as the following set of recursions for $t = 1, ..., n$ (see, e.g., Harvey [2]):

$$\begin{aligned} &\mathbf{a}_{t|t} = \mathbf{a}_{t|t-1} + \mathbf{M}_t\mathbf{v}_t, && \mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{M}_t\mathbf{F}_t\mathbf{M}_t', && \mathbf{M}_t = \mathbf{P}_{t|t-1}\mathbf{Z}_t'\mathbf{F}_t^{-1}, \\ &\mathbf{a}_{t+1|t} = \mathbf{c}_{t+1} + \mathbf{T}_{t+1}\mathbf{a}_{t|t}, && \mathbf{P}_{t+1|t} = \mathbf{T}_{t+1}\mathbf{P}_{t|t}\mathbf{T}_{t+1}' + \mathbf{R}_{t+1}\mathbf{Q}_{t+1}\mathbf{R}_{t+1}', && \end{aligned} \tag{3}$$

with initial condition $\mathbf{a}_{1|0} = \mathbf{a}_0$ and $\mathbf{P}_{1|0} = \mathbf{P}_0$. The gradient vector of L_n is given as (see Appendix A on the notations, such as $\partial L_n/\partial\boldsymbol{\theta}$):

$$\frac{\partial L_n}{\partial \boldsymbol{\theta}} = \sum_{t=1}^{n}\frac{\partial \ell_t}{\partial \boldsymbol{\theta}}, \quad \frac{\partial \ell_t}{\partial \boldsymbol{\theta}} = \frac{\partial(\mathbf{v}_t')}{\partial \boldsymbol{\theta}}\frac{\partial \ell_t}{\partial \mathbf{v}_t} + \frac{\partial[\mathrm{vec}(\mathbf{F}_t)']}{\partial \boldsymbol{\theta}}\frac{\partial \ell_t}{\partial \mathrm{vec}(\mathbf{F}_t)}. \tag{4}$$

[2] The formulas of Koopman and Shephard [4] and Segal and Weinstein [8] may also be extended for calculating all elements of the gradient vector in a single pass. However, such an extension requires the derivation of explicit forms of the derivatives with respect to matrices and vectors, which are not straightforward.

From (2) and (j), (k) and (l) in (A.2) in Appendix A, it follows that

$$\frac{\partial \ell_t}{\partial \mathbf{v}_t} = -\mathbf{F}_t^{-1}\mathbf{v}_t \quad \text{and} \quad \frac{\partial \ell_t}{\partial \text{vec}(\mathbf{F}_t)} = \frac{1}{2}\text{vec}\left(\mathbf{F}_t^{-1}\mathbf{v}_t\mathbf{v}_t'\mathbf{F}_t^{-1} - \mathbf{F}_t^{-1}\right). \tag{5}$$

Thus, because we can obtain $\mathbf{F}_t$ and $\mathbf{v}_t$ by running the Kalman filter, we can calculate $\partial L_n/\partial\boldsymbol{\theta}$ from (4) and (5) if we know the values of $\partial(\mathbf{v}_t')/\partial\boldsymbol{\theta}$ and $\partial[\text{vec}(\mathbf{F}_t)']/\partial\boldsymbol{\theta}$. In Proposition 1 below, first, we propose a formula for calculating $\partial(\mathbf{v}_t')/\partial\boldsymbol{\theta}$ and $\partial[\text{vec}(\mathbf{F}_t)']/\partial\boldsymbol{\theta}$, $t = 1, ..., n$, recursively.

For a matrix $\mathbf{A}$, we denote $\partial[\text{vec}(\mathbf{A})']/\partial\boldsymbol{\theta}$ by $\mathbf{G}_\theta(\mathbf{A})$. For example, consider a 2×2 matrix $\mathbf{A}$,

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad \left(\text{and hence,} \quad \text{vec}(\mathbf{A})' = \begin{bmatrix} a_{11} & a_{21} & a_{12} & a_{22} \end{bmatrix}\right).$$

For the matrix $\mathbf{A}$, $\mathbf{G}_\theta(\mathbf{A})$ denotes the $h \times 4$ matrix of partial derivatives,

$$\mathbf{G}_\theta(\mathbf{A}) = \begin{bmatrix} \frac{\partial a_{11}}{\partial\theta_1} & \frac{\partial a_{21}}{\partial\theta_1} & \frac{\partial a_{12}}{\partial\theta_1} & \frac{\partial a_{22}}{\partial\theta_1} \\ \frac{\partial a_{11}}{\partial\theta_2} & \frac{\partial a_{21}}{\partial\theta_2} & \frac{\partial a_{12}}{\partial\theta_2} & \frac{\partial a_{22}}{\partial\theta_2} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial a_{11}}{\partial\theta_h} & \frac{\partial a_{21}}{\partial\theta_h} & \frac{\partial a_{12}}{\partial\theta_h} & \frac{\partial a_{22}}{\partial\theta_h} \end{bmatrix}.$$

With this notation, $\partial L_t/\partial\boldsymbol{\theta}$, $\partial\ell_t/\partial\boldsymbol{\theta}$, $\partial(\mathbf{v}_t')/\partial\boldsymbol{\theta}$, and $\partial[\text{vec}(\mathbf{F}_t)']/\partial\boldsymbol{\theta}$ are denoted by $\mathbf{G}_\theta(L_n)$, $\mathbf{G}_\theta(\ell_t)$, $\mathbf{G}_\theta(\mathbf{v}_t)$, and $\mathbf{G}_\theta(\mathbf{F}_t)$, respectively. Throughout the paper, we adapt the rule that matrix multiplication always precede the Kronecker product, and we omit parentheses, namely, expressions such as $(\mathbf{AB}) \otimes (\mathbf{CD})$ are simply written like $\mathbf{AB} \otimes \mathbf{CD}$.

Proposition 1 *Suppose that all elements of* $\mathbf{G}_\theta(\mathbf{d}_t)$, $\mathbf{G}_\theta(\mathbf{Z}_t)$, $\mathbf{G}_\theta(\mathbf{S}_t)$, $\mathbf{G}_\theta(\mathbf{c}_s)$, $\mathbf{G}_\theta(\mathbf{T}_s)$, $\mathbf{G}_\theta(\mathbf{R}_s)$, $\mathbf{G}_\theta(\mathbf{Q}_s)$, $\mathbf{G}_\theta(\mathbf{a}_0)(= \mathbf{G}_\theta(\mathbf{a}_{1|0}))$, *and* $\mathbf{G}_\theta(\mathbf{P}_0)(= \mathbf{G}_\theta(\mathbf{P}_{1|0}))$ *for* $t = 1, ..., n$ *and* $s = 2, ..., n+1$, *exist and finite at a point of* $\boldsymbol{\theta}$. *Then, for* $t = 1, ..., n$, $\mathbf{G}_\theta(\mathbf{v}_t)$ *and* $\mathbf{G}_\theta(\mathbf{F}_t)$ *evaluated at this point can be calculated recursively by*

$$\mathbf{G}_\theta(\mathbf{v}_t) = -\mathbf{G}_\theta(\mathbf{d}_t) - \mathbf{G}_\theta(\mathbf{Z}_t)\left(\mathbf{a}_{t|t-1} \otimes \mathbf{I}_v\right) - \mathbf{G}_\theta(\mathbf{a}_{t|t-1})\mathbf{Z}_t', \tag{6}$$

$$\mathbf{G}_\theta(\mathbf{F}_t) = \mathbf{G}_\theta(\mathbf{Z}_t)(\mathbf{P}_{t|t-1}\mathbf{Z}_t' \otimes \mathbf{I}_v)\mathbf{N}_v + \mathbf{G}_\theta(\mathbf{P}_{t|t-1})\left(\mathbf{Z}_t' \otimes \mathbf{Z}_t'\right) + \mathbf{G}_\theta(\mathbf{S}_t), \tag{7}$$

$$\begin{aligned} \mathbf{G}_\theta(\mathbf{a}_{t|t}) = {} & \mathbf{G}_\theta(\mathbf{a}_{t|t-1}) + \mathbf{G}_\theta(\mathbf{P}_{t|t-1})(\mathbf{Z}_t'\mathbf{w}_t \otimes \mathbf{I}_m) + \mathbf{G}_\theta(\mathbf{Z}_t)(\mathbf{P}_{t|t-1} \otimes \mathbf{w}_t) \\ & -\mathbf{G}_\theta(\mathbf{F}_t)(\mathbf{w}_t \otimes \mathbf{M}_t') + \mathbf{G}_\theta(\mathbf{v}_t)\mathbf{M}_t', \end{aligned} \tag{8}$$

$$\mathbf{G}_\theta(\mathbf{P}_{t|t}) = \mathbf{G}_\theta(\mathbf{P}_{t|t-1})[\mathbf{I}_{m^2} - (\mathbf{Z}_t'\mathbf{M}_t' \otimes \mathbf{I}_m)\mathbf{N}_m] - \mathbf{G}_\theta(\mathbf{Z}_t)(\mathbf{P}_{t|t-1} \otimes \mathbf{M}_t')\mathbf{N}_m + \mathbf{G}_\theta(\mathbf{F}_t)(\mathbf{M}_t' \otimes \mathbf{M}_t'), \tag{9}$$

$$\mathbf{G}_\theta(\mathbf{a}_{t+1|t}) = \mathbf{G}_\theta(\mathbf{c}_{t+1}) + \mathbf{G}_\theta(\mathbf{T}_{t+1})(\mathbf{a}_{t|t} \otimes \mathbf{I}_m) + \mathbf{G}_\theta(\mathbf{a}_{t|t})\mathbf{T}_{t+1}', \tag{10}$$

$$\begin{aligned} \mathbf{G}_\theta(\mathbf{P}_{t+1|t}) = {} & \mathbf{G}_\theta(\mathbf{P}_{t|t})(\mathbf{T}_{t+1}' \otimes \mathbf{T}_{t+1}') + \mathbf{G}_\theta(\mathbf{Q}_{t+1})(\mathbf{R}_{t+1}' \otimes \mathbf{R}_{t+1}') \\ & +[\mathbf{G}_\theta(\mathbf{T}_{t+1})(\mathbf{P}_{t|t}\mathbf{T}_{t+1}' \otimes \mathbf{I}_m) + \mathbf{G}_\theta(\mathbf{R}_{t+1})(\mathbf{Q}_{t+1}\mathbf{R}_{t+1}' \otimes \mathbf{I}_m)]\mathbf{N}_m, \end{aligned} \tag{11}$$

where $\mathbf{w}_t = \mathbf{F}_t^{-1}\mathbf{v}_t$, $\mathbf{N}_m = \mathbf{I}_{m^2} + \mathbf{K}_m$, *and* $\mathbf{K}_{mk}$ *is a unique* $mk \times mk$ *matrix that satisfies* $\mathbf{K}_{mk}\text{vec}(\mathbf{A}) = \text{vec}(\mathbf{A}')$ *for a* $m \times k$ *matrix* $\mathbf{A}$. *If* $m = k$, *we write* $\mathbf{K}_{mm}$ *as* $\mathbf{K}_m$.

The matrix $\mathbf{K}_{mk}$ is called a *commutation matrix.* Magnus and Neudecker [6] show that

$$\mathbf{K}_{mk} = \sum\nolimits_{i=1}^{m}\sum\nolimits_{j=1}^{k}(\mathbf{E}_{i,j} \otimes \mathbf{E}_{i,j}'),$$

which can be used for computing $\mathbf{K}_{mk}$, where $\mathbf{E}_{i,j} = e_i^{(m)}e_j^{(k)\prime}$ and $e_i^{(m)}$ is the $m \times 1$ vector whose i-th element is 1 and all other elements are 0. See Magnus and Neudecker [6] and Magnus and Neudecker [7] for properties of the commutation matrix.

The formula in the above Proposition is useful if one wants to obtain $\mathbf{G}_\theta(\mathbf{a}_{t|t})$ and $\mathbf{G}_\theta(\mathbf{P}_{t|t})$ too. However, if one's goal is only to calculate the gradient vector $\mathbf{G}_\theta(L_n) = \sum_{t=1}^{n}\mathbf{G}_\theta(\ell_t)$, it is more convenient to combine the results in (4) $\sim$ (11) as in the following Proposition

Proposition 2 *Suppose that all elements of* $\mathbf{G}_\theta(\mathbf{d}_t)$, $\mathbf{G}_\theta(\mathbf{Z}_t)$, $\mathbf{G}_\theta(\mathbf{S}_t)$, $\mathbf{G}_\theta(\mathbf{c}_s)$, $\mathbf{G}_\theta(\mathbf{T}_s)$, $\mathbf{G}_\theta(\mathbf{R}_s)$, $\mathbf{G}_\theta(\mathbf{Q}_s)$, $\mathbf{G}_\theta(\mathbf{a}_0)(=\mathbf{G}_\theta(\mathbf{a}_{1|0}))$, *and* $\mathbf{G}_\theta(\mathbf{P}_0)(=\mathbf{G}_\theta(\mathbf{P}_{1|0}))$ *for* $t=1,...,n$ *and* $s=2,...,n+1$, *exist and finite at a point of* $\boldsymbol{\theta}$. *Then, for* $t=1,....,n$, $\mathbf{G}_\theta(\ell_t)$ *evaluated at this point can be recursively calculated by*

$$\begin{aligned}\mathbf{G}_\theta(\ell_t) = {} & \mathbf{G}_\theta(\mathbf{a}_{t|t-1})\mathbf{Z}_t'\mathbf{w}_t + \frac{1}{2}\mathbf{G}_\theta(\mathbf{P}_{t|t-1})\mathrm{vec}(\mathbf{Z}_t'\mathbf{w}_t\mathbf{w}_t'\mathbf{Z}_t - \mathbf{Z}_t'\mathbf{F}_t^{-1}\mathbf{Z}_t) + \mathbf{G}_\theta(\mathbf{d}_t)\mathbf{w}_t \\ & + \mathbf{G}_\theta(\mathbf{Z}_t)\mathrm{vec}(\mathbf{w}_t\mathbf{a}_{t|t-1}' + \mathbf{w}_t\mathbf{v}_t'\mathbf{M}_t' - \mathbf{M}_t') + \frac{1}{2}\mathbf{G}_\theta(\mathbf{S}_t)\mathrm{vec}(\mathbf{w}_t\mathbf{w}_t' - \mathbf{F}_t^{-1}),\end{aligned} \tag{12}$$

$$\begin{aligned}\mathbf{G}_\theta(\mathbf{a}_{t+1|t}) = {} & \mathbf{G}_\theta(\mathbf{a}_{t|t-1})\mathbf{L}_t' + \mathbf{G}_\theta(\mathbf{P}_{t|t-1})(\mathbf{Z}_t'\mathbf{w}_t \otimes \mathbf{L}_t') + \mathbf{G}_\theta(\mathbf{c}_{t+1}) - \mathbf{G}_\theta(\mathbf{d}_t)\mathbf{J}_t' \\ & + \mathbf{G}_\theta(\mathbf{Z}_t)[\mathbf{P}_{t|t-1}\mathbf{L}_t' \otimes \mathbf{w}_t - (\mathbf{a}_{t|t-1} + \mathbf{M}_t\mathbf{v}_t) \otimes \mathbf{J}_t'] - \mathbf{G}_\theta(\mathbf{S}_t)(\mathbf{w}_t \otimes \mathbf{J}_t') \\ & + \mathbf{G}_\theta(\mathbf{T}_{t+1})[(\mathbf{a}_{t|t-1} + \mathbf{M}_t\mathbf{v}_t) \otimes \mathbf{I}_m],\end{aligned} \tag{13}$$

and

$$\begin{aligned}\mathbf{G}_\theta(\mathbf{P}_{t+1|t}) = {} & \mathbf{G}_\theta(\mathbf{P}_{t|t-1})(\mathbf{L}_t' \otimes \mathbf{L}_t') + \mathbf{G}_\theta(\mathbf{S}_t)(\mathbf{J}_t' \otimes \mathbf{J}_t') + \mathbf{G}_\theta(\mathbf{Q}_{t+1})(\mathbf{R}_{t+1}' \otimes \mathbf{R}_{t+1}') \\ & + [\mathbf{G}_\theta(\mathbf{T}_{t+1})(\mathbf{P}_{t|t-1}\mathbf{L}_t' \otimes \mathbf{I}_m) - \mathbf{G}_\theta(\mathbf{Z}_t)(\mathbf{P}_{t|t-1}\mathbf{L}_t' \otimes \mathbf{J}_t') + \mathbf{G}_\theta(\mathbf{R}_{t+1})(\mathbf{Q}_{t+1}\mathbf{R}_{t+1}' \otimes \mathbf{I}_m)]\mathbf{N}_m,\end{aligned} \tag{14}$$

where $\mathbf{L}_t = \mathbf{T}_{t+1} - \mathbf{J}_t\mathbf{Z}_t$, $\mathbf{J}_t = \mathbf{T}_{t+1}\mathbf{M}_t$, *and* $\mathbf{w}_t$ *and* $\mathbf{N}_m$ *are the same as in* Proposition 1.

See Appendix B for the proof of Proposition 2.

Note that the formula in (3) can be concisely written as recursions for calculating only $\mathbf{a}_{t+1|t}$ and $\mathbf{P}_{t+1|t}$:

$$\mathbf{a}_{t+1|t} = \mathbf{c}_{t+1} + \mathbf{T}_{t+1}\mathbf{a}_{t|t-1} + \mathbf{J}_t\mathbf{v}_t, \quad \mathbf{P}_{t+1|t} = \mathbf{T}_{t+1}\mathbf{P}_{t|t-1}\mathbf{L}_t' + \mathbf{R}_{t+1}\mathbf{Q}_{t+1}\mathbf{R}_{t+1}'.$$

This may be a more convenient expression in some cases.

Remark 1. Unlike the formula in Harvey [2] (and papers mentioned in Section 1), which can calculate only one component of the gradient vector in a single pass, the above formula can obtain all components of the gradient vector in a single pass. A simple computational experiment in Section 5 suggests that the total computational time of the proposed formula for computing all components of the gradient vector is less than that of the formula in Harvey [2] .

Remark 2. $\mathbf{G}_\theta(\ell_t)$ is called the gradient vector of t-th contribution to the log likelihood. This quantity is often used to construct an asymptotic covariance matrix estimator in the context of estimation and testing. The proposed formula can calculate $\mathbf{G}_\theta(\ell_t)$ for all t in a single pass.

Remark 3. In some cases, it is more convenient to calculate $\mathbf{G}_\theta(\mathbf{Z}_t')$, $\mathbf{G}_\theta(\mathbf{T}_s')$, and $\mathbf{G}_\theta(\mathbf{R}_s')$ rather than $\mathbf{G}_\theta(\mathbf{Z}_t)$, $\mathbf{G}_\theta(\mathbf{T}_s)$, and $\mathbf{G}_\theta(\mathbf{R}_s)$ (see the first example in Section 4). In the above formula, one can replace $\mathbf{G}_\theta(\mathbf{Z}_t)$, $\mathbf{G}_\theta(\mathbf{T}_s)$, and $\mathbf{G}_\theta(\mathbf{R}_s)$ with $\mathbf{G}_\theta(\mathbf{Z}_t')\mathbf{K}_{vm}$, $\mathbf{G}_\theta(\mathbf{T}_s')\mathbf{K}_m$, and $\mathbf{G}_\theta(\mathbf{R}_s')\mathbf{K}_{mw}$, respectively. See Lemma 1(c) in Appendix B.

Remark 4. Suppose that the unknown parameter vector $\boldsymbol{\theta}$ is divided so that $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$, and one wants to obtain only the gradient vector with respect to $\boldsymbol{\theta}_1$. In this case, one just has to apply the formula only for $\mathbf{G}_{\theta_1}(\cdot)$, where $\mathbf{G}_{\theta_1}(\mathbf{B}) = \partial[\mathrm{vec}(\mathbf{B})']/\partial\boldsymbol{\theta}_1$ for a matrix $\mathbf{B}$. This often simplifies the formula. For example, if only $\mathbf{T}_s$ depends on $\boldsymbol{\theta}_1$, i.e., $\mathbf{G}_{\theta_1}(\mathbf{X}_s) = \mathbf{0}$ for $\mathbf{X}_s \neq \mathbf{T}_s$, then, the formula in Proposition 2 is simplified to

$$\mathbf{G}_{\theta_1}(\ell_t) = \mathbf{G}_{\theta_1}(\mathbf{a}_{t|t-1})\mathbf{Z}_t'\mathbf{w}_t + \frac{1}{2}\mathbf{G}_{\theta_1}(\mathbf{P}_{t|t-1})\mathrm{vec}(\mathbf{Z}_t'\mathbf{w}_t\mathbf{w}_t'\mathbf{Z}_t - \mathbf{Z}_t'\mathbf{F}_t^{-1}\mathbf{Z}_t), \tag{15}$$

$$\mathbf{G}_{\theta_1}(\mathbf{a}_{t+1|t}) = \mathbf{G}_{\theta_1}(\mathbf{a}_{t|t-1})\mathbf{L}_t' + \mathbf{G}_{\theta_1}(\mathbf{P}_{t|t-1})(\mathbf{Z}_t'\mathbf{w}_t \otimes \mathbf{L}_t') + \mathbf{G}_{\theta_1}(\mathbf{T}_{t+1})[(\mathbf{a}_{t|t-1} + \mathbf{M}_t\mathbf{v}_t) \otimes \mathbf{I}_m], \tag{16}$$

and

$$\mathbf{G}_{\theta_1}(\mathbf{P}_{t+1|t}) = \mathbf{G}_{\theta_1}(\mathbf{P}_{t|t-1})(\mathbf{L}_t' \otimes \mathbf{L}_t') + \mathbf{G}_{\theta_1}(\mathbf{T}_{t+1})[(\mathbf{P}_{t|t-1}\mathbf{L}_t') \otimes \mathbf{I}_m]\mathbf{N}_m. \tag{17}$$

3.2 Time-invariant system matrices and stationary state vector

In this subsection, we explicitly consider how to handle the initial condition when the system matrices are all time-invariant and the state vector $\boldsymbol{\alpha}_t$ is stationary, in the context of computing the exact gradient vector. This issue does not seem to be fully addressed in the previous literature. Harvey [2], Hooker [3], and Koopman and Shephard [4] explicitly (or implicitly) assume that the initial condition $\mathbf{a}_0$ and $\mathbf{P}_0$ do not depend on the parameter vector $\boldsymbol{\theta}$. This assumption is; however, restrictive in many practical applications.

Consider the case that $\mathbf{d}_t$, $\mathbf{Z}_t$, $\mathbf{S}_t$, $\mathbf{c}_s$, $\mathbf{T}_s$, $\mathbf{R}_s$ and $\mathbf{Q}_s$ are all time-invariant (e.g., $\mathbf{c}_s = \mathbf{c}$ for all). Then, if the eigenvalues of $\mathbf{T}_t = \mathbf{T}$ are all less than one in absolute value, the state vector $\boldsymbol{\alpha}_t$ is (asymptotically) covariance stationary. In this case, $\mathbf{a}_0$ and $\mathbf{P}_0$ are often set to the stationary mean vector and covariance matrix of $\boldsymbol{\alpha}_t$, namely,

$$\mathbf{a}_0 = (\mathbf{I}_m - \mathbf{T})^{-1}\mathbf{c} \quad \text{and} \quad \mathrm{vec}(\mathbf{P}_0) = (\mathbf{I}_{m^2} - \mathbf{T}\otimes\mathbf{T})^{-1}\mathrm{vec}(\mathbf{RQR}'). \tag{18}$$

Under this setting, changes in the value of $\boldsymbol{\theta}$ affect the value of the gradient vector via the initial conditions because $\mathbf{a}_0$ and $\mathbf{P}_0$ are functions of $\boldsymbol{\theta}$. Hence, one has to take its effects into account in calculating the gradient vector.

In what follows, we focus on the case where the initial condition is set as in (18). The following proposition gives $\mathbf{G}_\theta(\mathbf{a}_0)$ and $\mathbf{G}_\theta(\mathbf{P}_0)$ when $\mathbf{a}_0$ and $\mathbf{P}_0$ are set as in (18).

Proposition 3 ***(Derivative Matrices of Stationary Initial Conditions in (18))***
Suppose that the conditions stated in Proposition 1 are satisfied, and that $\mathbf{a}_0$ *and* $\mathbf{P}_0$ *are set as in (18). Then*

$$\mathbf{G}_\theta(\mathbf{a}_0) = [\mathbf{G}_\theta(\mathbf{c}) + \mathbf{G}_\theta(\mathbf{T})(\mathbf{a}_0\otimes\mathbf{I}_m)](\mathbf{I}_m - \mathbf{T}')^{-1}, \tag{19}$$

and

$$\mathbf{G}_\theta(\mathbf{P}_0) = \{[\mathbf{G}_\theta(\mathbf{T})(\mathbf{P}_0\mathbf{T}'\otimes\mathbf{I}_m) + \mathbf{G}_\theta(\mathbf{R})(\mathbf{QR}'\otimes\mathbf{I}_m)]\mathbf{N}_m + \mathbf{G}_\theta(\mathbf{Q})(\mathbf{R}'\otimes\mathbf{R}')\}(\mathbf{I}_{m^2} - \mathbf{T}'\otimes\mathbf{T}')^{-1}. \tag{20}$$

Note that the formulas given in Propositions 1 and 2 need not assume time-invariant system matrices, while the formula in Proposition 3 is valid only when the system matrices are time-invariant.

4 Examples

4.1 ARMA(p, q) model

Here, we consider a stationary ARMA(p, q) model as a first example. One can write a stationary ARMA(p, q) model as

$$y_t - \mu = \phi_1(y_{t-1}-\mu) + \cdots + \phi_m(y_{t-m}-\mu) + \varepsilon_t + \beta_1\varepsilon_{t-1} + \cdots + \beta_{m-1}\varepsilon_{t-m+1}, \quad \varepsilon_t \sim \text{i.i.d.}N(0,\sigma^2),$$

where $m \equiv \max\{p, q+1\}$, $\phi_j = 0$ for $j > p$ and $\beta_j = 0$ for $j > q$. Let $\boldsymbol{\phi}_k = [\phi_1 \cdots \phi_k]'$, and $\boldsymbol{\beta}_k = [\beta_1 \cdots \beta_k]'$. The above ARMA($p$, q) model can be cast into the following state space form (see Hamilton [5, p.375]):

$$y_t = \mathbf{d} + \mathbf{Z}\boldsymbol{\alpha}_t, \quad \boldsymbol{\alpha}_{t+1} = \mathbf{T}\boldsymbol{\alpha}_t + \mathbf{R}\boldsymbol{\eta}_{t+1}, \quad \boldsymbol{\eta}_t \sim i.i.d.N(\mathbf{0},\, \mathbf{Q}), \quad \boldsymbol{\alpha}_1 \sim N(\mathbf{a}_0,\, \mathbf{P}_0), \quad t = 1, ..., n, \tag{21}$$

where $\mathbf{d} = \mu$, $\boldsymbol{\eta}_t = \varepsilon_t$, $\mathbf{Q} = \sigma_\varepsilon^2$,

$$\underset{(1\times m)}{\mathbf{Z}} = \left[\begin{array}{cc} 1 & \boldsymbol{\beta}_{m-1}' \end{array}\right], \; \underset{(m\times m)}{\mathbf{T}} = \left[\begin{array}{cc} \multicolumn{2}{c}{\boldsymbol{\phi}_m'} \\ \mathbf{I}_{m-1} & \mathbf{0}_{(m-1)\times 1} \end{array}\right], \quad \text{and} \quad \underset{(m\times 1)}{\mathbf{R}} = \left[\begin{array}{c} 1 \\ \mathbf{0}_{(m-1)\times 1} \end{array}\right].$$

The notation $(a \times b)$ implies that the numbers of rows and columns of the matrix are a and b, respectively. See de Jong and Penzer [1] on other state space forms for ARMA(p, q) models. This state space form is an example of time-invariant system matrices and stationary state variable. Hence, one often sets the initial condition, $\mathbf{a}_0$ and $\mathbf{P}_0$, as in (18). Collect unknown parameters into a vector $\boldsymbol{\theta}$ so that $\boldsymbol{\theta} = [\boldsymbol{\phi}_p', \boldsymbol{\beta}_q', \mu, \sigma^2]'$. Then, for example, $\mathbf{G}_\theta(\mathbf{T})$ is given as

$$\underset{((p+q+2)\times m^2)}{\mathbf{G}_\theta(\mathbf{T})} = \begin{bmatrix} \mathbf{G}_\phi(\mathbf{T}) \\ \mathbf{G}_\beta(\mathbf{T}) \\ \mathbf{G}_\mu(\mathbf{T}) \\ \mathbf{G}_{\sigma^2}(\mathbf{T}) \end{bmatrix} \quad \text{where} \quad \mathbf{G}_\phi(\mathbf{T}) = \frac{\partial[\text{vec}(\mathbf{T})']}{\partial \boldsymbol{\phi}_p},$$

and $\mathbf{G}_\beta(\mathbf{T})$, $\mathbf{G}_\mu(\mathbf{T})$, and $\mathbf{G}_{\sigma^2}(\mathbf{T})$ are defined similarly. For this state space form, however, it is more convenient to work with $\mathbf{G}_\theta(\mathbf{T}')$ rather than $\mathbf{G}_\theta(\mathbf{T})$ because the former can be explicitly expressed in a much simpler matrix form than the latter (see Remark 3), namely,

$$\mathbf{G}_\theta(\mathbf{T}') = \begin{bmatrix} \mathbf{I}_p & \mathbf{0}_{p\times(m^2-p)} \\ \multicolumn{2}{c}{\mathbf{0}_{(q+2)\times m^2}} \end{bmatrix}.$$

Similarly, other derivative matrices are given as

$$\mathbf{G}_\theta(\mathbf{d}) = \begin{bmatrix} \mathbf{0}_{(p+q)\times 1} \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{G}_\theta(\mathbf{Q}) = \begin{bmatrix} \mathbf{0}_{(p+q+1)\times 1} \\ 1 \end{bmatrix}, \quad \mathbf{G}_\theta(\mathbf{R}) = \mathbf{0}_{(p+q+2)\times m},$$

and

$$\mathbf{G}_\theta(\mathbf{Z}) = \begin{bmatrix} \multicolumn{3}{c}{\mathbf{0}_{p\times m}} \\ \mathbf{0}_{q\times 1} & \mathbf{I}_q & \mathbf{0}_{q\times(m-q-1)} \\ \multicolumn{3}{c}{\mathbf{0}_{2\times m}} \end{bmatrix} \text{ when } m > q+1, \quad \text{and} \quad \mathbf{G}_\theta(\mathbf{Z}) = \begin{bmatrix} \multicolumn{2}{c}{\mathbf{0}_{p\times m}} \\ \mathbf{0}_{q\times 1} & \mathbf{I}_q \\ \multicolumn{2}{c}{\mathbf{0}_{2\times m}} \end{bmatrix} \text{ when } m = q+1.$$

Given these values, one can apply the formula given in Propositions 1 to 3 for calculating the exact gradient vector of the log likelihood of the ARMA (p, q) model.

4.2 Stochastic Coefficient Model

As a next example, we consider a model known as a stochastic coefficient (STC) model. We restrict our attention to the case where the STC vector, $\boldsymbol{\alpha}_t$, follows an vector autoregressive (VAR) process of order 1, or VAR(1). Specifically, we consider the following STC model:

$$\mathbf{y}_t = \mathbf{Z}_t \boldsymbol{\alpha}_t, \quad \boldsymbol{\alpha}_{t+1} = \mathbf{c} + \mathbf{T}\boldsymbol{\alpha}_t + \boldsymbol{\eta}_{t+1}, \quad \boldsymbol{\eta}_t \sim i.i.d.N(\mathbf{0}, \mathbf{Q}), \quad \boldsymbol{\alpha}_1 \sim N(\mathbf{a}_0, \mathbf{P}_0), \quad t = 1, ..., n, \tag{22}$$

where $\mathbf{y}_t$, $\mathbf{Z}_t$, $\boldsymbol{\alpha}_t$, and $\boldsymbol{\eta}_t$ are $v \times 1$ dependent vector, $v \times m$ explanatory matrices, $m \times 1$ STC vector, and $m \times 1$ vector of unobserved errors, respectively, $\mathbf{c}$, $\mathbf{T}$, and $\mathbf{Q}$ are $m \times 1$, $m \times m$, $m \times m$ constant matrices (or vectors), respectively. We assume that the elements of $\mathbf{c}$, $\mathbf{T}$, and $\mathbf{Q}$, are functions of an unknown $h \times 1$ parameter vector $\boldsymbol{\theta} = (\theta_1, ..., \theta_h)'$.

The model specified in (22) is enough general to include the usual form of STC model in that i-th element of $\mathbf{y}_t$, $y_{i,t}$, is given as

$$y_{i,t} = \mathbf{x}_{i,t}' \boldsymbol{\beta}_{i,t} + \varepsilon_{i,t}, \quad \text{for} \quad i = 1, ..., v,$$

where $\mathbf{x}_{i,t}$ is a $k_i \times 1$ vector of non-constant explanatory variables, $\boldsymbol{\beta}_{i,t}$ is a $k_i \times 1$ STC vector of $\mathbf{x}_{i,t}$, and $\varepsilon_{i,t}$ is a disturbance term. Here, as the notation implies, k_i, the number of the explanatory variables for $y_{i,t}$, depends on i, and the STC vector of $\mathbf{x}_{i,t}$ can be different across i. This usual form of STC model can be written in the form in (22) by setting

$$\mathbf{Z}_t = \begin{bmatrix} \mathbf{x}_{1,t}' & \mathbf{0}_{1\times k_2} & \cdots & \mathbf{0}_{1\times k_v} & 1 & 0 & \cdots & 0 \\ \mathbf{0}_{1\times k_1} & \mathbf{x}_{2,t}' & \cdots & \mathbf{0}_{1\times k_v} & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{1\times k_1} & \mathbf{0}_{1\times k_2} & \cdots & \mathbf{x}_{v,t}' & 0 & 0 & \cdots & 1 \end{bmatrix} \quad \boldsymbol{\alpha}_t = \begin{bmatrix} \boldsymbol{\beta}_{1,t} \\ \boldsymbol{\beta}_{2,t} \\ \vdots \\ \boldsymbol{\beta}_{v,t} \\ \boldsymbol{\varepsilon}_t \end{bmatrix},$$

where $\boldsymbol{\varepsilon}_t = [\varepsilon_{1,t}, \varepsilon_{2,t}, ..., \varepsilon_{v,t}]'$ is a $v \times 1$ vector of disturbance terms. Here, $m = k_1 + \cdots + k_v + v$.

The unknown parameter vector is

$$\boldsymbol{\theta} = [\mathbf{c}', \mathrm{vec}(\mathbf{T})', \mathrm{vech}(\mathbf{Q})']' = [c_1, ..., c_m, t_{11}, t_{21}, ..., t_{mm}, q_{11}, q_{21}, ..., q_{m1}, q_{22}, ..., q_{mm}]'. \tag{23}$$

The lengths of the vectors, $\mathbf{c}$, $\mathrm{vec}(\mathbf{T})$, and $\mathrm{vech}(\mathbf{Q})$ are m, m^2, and $m(m+1)/2$, respectively. $\mathbf{G}_\theta(\mathbf{c})$, $\mathbf{G}_\theta(\mathbf{T})$, and $\mathbf{G}_\theta(\mathbf{Q})$ are given as

$$\mathbf{G}_\theta(\mathbf{c}) = \begin{bmatrix} \mathbf{I}_m \\ \mathbf{0}_{\frac{m(3m+1)}{2}\times m} \end{bmatrix}, \quad \mathbf{G}_\theta(\mathbf{T}) = \begin{bmatrix} \mathbf{0}_{m\times m^2} \\ \mathbf{I}_{m^2} \\ \mathbf{0}_{\frac{m(m+1)}{2}\times m^2} \end{bmatrix}, \quad \text{and} \quad \mathbf{G}_\theta(\mathbf{Q}) = \begin{bmatrix} \mathbf{0}_{m(m+1)\times m^2} \\ \mathbf{D}'_m, \end{bmatrix},$$

where $\mathbf{D}_m$ is a $m^2 \times \frac{1}{2}m(m+1)$ matrix called the *duplication matrix*, which has the property that $\mathbf{D}_m\mathrm{vech}(\mathbf{A}) = \mathrm{vec}(\mathbf{A})$ for a symmetric $m \times m$ matrix $\mathbf{A}$. Note that $\mathbf{G}_\theta(\mathbf{Q})$ is given as above because

$$\mathbf{G}_\theta(\mathbf{Q}) = \begin{bmatrix} \mathbf{0}_{(m+m^2)\times m^2} \\ \frac{\partial[\mathrm{vec}(\mathbf{Q})']}{\partial \mathrm{vech}(\mathbf{Q})} \end{bmatrix} \quad \text{and} \quad \frac{\partial[\mathrm{vec}(\mathbf{Q})']}{\partial \mathrm{vech}(\mathbf{Q})} = \frac{\partial\{[\mathbf{D}_m \mathrm{vech}(\mathbf{Q})]'\}}{\partial \mathrm{vech}(\mathbf{Q})} = \frac{\partial[\mathrm{vech}(\mathbf{Q})']}{\partial \mathrm{vech}(\mathbf{Q})}\mathbf{D}'_m = \mathbf{D}'_m.$$

See Magnus and Neudecker [7] for more details on $\mathbf{D}_m$.

5 Comparison of Computational Time

In this section, we conduct a simple computational experiment, in that we generate samples from a simple state space model, and compare the computational time by the formulas in Harvey [2] and Proposition 2, for calculating all components of the gradient vector of the Gaussian log-likelihood.

Specifically, we generate samples from the following time-invariant state space model.

$$\begin{aligned} &\mathbf{y}_t = \mathbf{Z}\boldsymbol{\alpha}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\alpha}_{t+1} = \mathbf{T}\boldsymbol{\alpha}_t + \boldsymbol{\eta}_{t+1}, \quad \boldsymbol{\alpha}_1 \sim N(\mathbf{a}_0,\ \mathbf{P}_0), \\ &\boldsymbol{\varepsilon}_t, \sim \text{i.i.d.}N(\mathbf{0}, \mathbf{S}), \quad \boldsymbol{\eta}_t \sim \text{i.i.d.}N(\mathbf{0}, \mathbf{Q}), \quad t = 1, ..., n, \end{aligned} \tag{24}$$

where $\mathbf{Z} = [z_{i,j}]$, $\mathbf{T} = [t_{i,j}]$, $\mathbf{S} = [s_{i,j}]$, and $\mathbf{Q} = [q_{i,j}]$, $i, j = 1, ..., m$. We set the number of samples as $n = 100$. We calculate the gradient vector with respect to the $(3m^2 + m) \times 1$ unknown parameter vector $\boldsymbol{\theta} = [\mathrm{vec}(\mathbf{Z})', \mathrm{vec}(\mathbf{T})', \mathrm{vech}(\mathbf{S})', \mathrm{vech}(\mathbf{Q})']'$. We fix the initial conditions at $\mathbf{a}_0 = \mathbf{0}_{m\times 1}$ and $\mathbf{P}_0 = \mathbf{I}_m$. We consider the following three cases: (a) $m = 1$, (b) $m = 2$, and (c) $m = 3$. The number of unknown parameters are 4, 14, and 30, for Cases (a), (b), and (c), respectively. We set the true values of parameters as $\mathbf{Z} = \mathbf{I}_m$, $\mathbf{T} = 0.8 \times \mathbf{I}_m$ and $\mathbf{S} = \mathbf{Q} = \mathbf{I}_m$. We calculate the gradient vector at this point.

We compare the formulas in Harvey [2] and Propositions 2. We conduct the following experiment for each m:

(S1) generate n samples from the state space model in (24).

(S2) record the computational time for each formula to calculate the gradient vector.

(S3) repeat (S1) and (S2) 100 times and take the average of the computational time for each formula.

We code these formulas by MATLAB programing language version 7.1. Table 1 reports the ratios of the average computational time of the formula in Harvey [2] to that of the formula in Proposition 2.

We observe that the computational time of the proposed formula is less than the formula in Harvey [2] and the ratio increases as the number of parameter increases, which suggests that the proposed formula even has computational advantages over the formula in Harvey [2].[3]

[3]We recognize that our computational experiment is not a formal comparison of the computational time of these formulas because the computational time depends on how these formulas are programed. On the other hand, although it is difficult to formally compare the costs of codings of these formulas, it is usually much easier to code the proposed formula than the formula in Harvey [2].

Table 1: Comparison of computational time

m	1(4)	2(14)	3(30)
ratio	1.83	2.55	5.17

Note: the number of parameters for each m is inside the parenthesis.

6 Concluding Remarks

In this paper, extending the formula in Harvey [2], we proposed a recursive formula for calculating the exact gradient vector of the log-likelihood for a general form of linear Gaussian state space models. Two of the advantages of the proposed formula compared with some existing formulas are that it is easy to code for some matrix computing languages and that its computational time is faster than the formula in Harvey [2]. Additionally, we explicitly addressed the issue on the initial condition in the case of stationary state variable.

One can combine the proposed formula with the usual numerical differentiation techniques to calculate the Hessian matrix of the log-likelihood more accurately, which is a key ingredient of the robust covariance matrix estimation. It may be the case for the test statistics involving the calculation of the gradient vector such as the score test that using not an approximate but the exact gradient vector reduces the size distortions and enhances the power.

Appendix

A Some Matrix Operators and Formulas of Matrix Calculus

In this Appendix, we briefly review some matrix operators, their properties, and formulas in matrix calculus used in the text. See Magnus and Neudecker [7] for more details.

Let $\mathbf{A}$ and $\mathbf{B}$ be $m \times k$ and $p \times q$ matrices, respectively. Let $\mathbf{C}$ and $\mathbf{D}$ be matrices whose numbers of rows are q and k, respectively. Let a_{ij} denote (i,j) element of $\mathbf{A}$. The commutation matrix $\mathbf{K}_{mk}$ is a unique $mk \times mk$ matrix that satisfies $\mathbf{K}_{mk}\text{vec}(\mathbf{A}) = \text{vec}(\mathbf{A}')$ (see Magnus and Neudecker [6]). When $k = m$, we denote $\mathbf{K}_{mk}$ by $\mathbf{K}_m$. The Kronecker product, denoted by $\otimes$, the vec operator, and the commutation matrix have the following properties:

$$\begin{array}{lll} \text{(a) } (\mathbf{A}\otimes\mathbf{B})' = \mathbf{A}'\otimes\mathbf{B}', & \text{(b) } \text{vec}(\mathbf{ABC}) = (\mathbf{C}'\otimes\mathbf{A})\text{vec}(\mathbf{B}) \text{ if } k=p, & \\ \text{(c) } \mathbf{K}'_{mk} = \mathbf{K}^{-1}_{mk} = \mathbf{K}_{km}, & \text{(d) } \mathbf{K}_{pm}(\mathbf{A}\otimes\mathbf{B}) = (\mathbf{B}\otimes\mathbf{A})\mathbf{K}_{qk}, & \text{(e) } \mathbf{K}_m\mathbf{N}_m = \mathbf{N}_m = \mathbf{N}_m\mathbf{K}_m, \\ \text{(f) } \mathbf{K}_{m1} = \mathbf{K}_{1m} = \mathbf{I}_m, & \text{(g) } (\mathbf{A}\otimes\mathbf{B})(\mathbf{D}\otimes\mathbf{C}) = \mathbf{AD}\otimes\mathbf{BC}, & \end{array} \tag{A.1}$$

where $\mathbf{N}_m = \mathbf{I}_{m^2} + \mathbf{K}_m$.

Let f be a function of elements of $\mathbf{A}$. Let $\mathbf{f} = [f_1, ..., f_q]'$ be a $q \times 1$ vector whose elements are functions of elements of a vector $\mathbf{c} = [c_1, ..., c_p]'$. Let $\mathbf{F} = [f_{ij}]$ be a $p \times q$ matrix whose elements are functions of a variable d. The notations $\partial f/\partial\mathbf{A}$, $\partial\mathbf{f}'/\partial\mathbf{c}$, and $\partial\mathbf{F}/\partial d$ mean

$$\underset{(m\times k)}{\frac{\partial f}{\partial \mathbf{A}}} = \begin{bmatrix} \frac{\partial f}{\partial a_{11}} & \cdots & \frac{\partial f}{\partial a_{1k}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial a_{m1}} & \cdots & \frac{\partial f}{\partial a_{mk}} \end{bmatrix}, \quad \underset{(p\times q)}{\frac{\partial \mathbf{f}'}{\partial \mathbf{c}}} = \begin{bmatrix} \frac{\partial f_1}{\partial c_1} & \cdots & \frac{\partial f_q}{\partial c_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial c_p} & \cdots & \frac{\partial f_q}{\partial c_p} \end{bmatrix}, \quad \text{and} \quad \underset{(p\times q)}{\frac{\partial \mathbf{F}}{\partial d}} = \begin{bmatrix} \frac{\partial f_{11}}{\partial d} & \cdots & \frac{\partial f_{1q}}{\partial d} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_{p1}}{\partial d} & \cdots & \frac{\partial f_{pq}}{\partial d} \end{bmatrix}.$$

It immediately follows that $\text{vec}(\partial f/\partial\mathbf{A}) = \partial f/\partial\text{vec}(\mathbf{A})$ and $\text{vec}(\partial\mathbf{F}/\partial d) = \partial\text{vec}(\mathbf{F})/\partial d$.

Let $\mathbf{S}$ be a $p \times p$ square matrix. Suppose that $\mathbf{AB}$ is well defined. Suppose further that the elements of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{S}$ are functions of a variable d and that these functions are differentiable at a point of d. Then, at this point, the followings hold:

$$\begin{array}{llll} \text{(h)} & \frac{\partial(\mathbf{AB})}{\partial d} = \frac{\partial\mathbf{A}}{\partial d}\mathbf{B} + \mathbf{A}\frac{\partial\mathbf{B}}{\partial d}, & \text{(i)} \quad \frac{\partial\mathbf{S}^{-1}}{\partial d} = -\mathbf{S}^{-1}\frac{\partial\mathbf{S}}{\partial d}\mathbf{S}^{-1}, & \text{(j)} \quad \frac{\partial\log\|\mathbf{S}\|}{\partial\mathbf{S}} = (\mathbf{S}')^{-1}, \\ \text{(k)} & \frac{\partial(\mathbf{c}'\mathbf{Sc})}{\partial\mathbf{c}} = (\mathbf{S}+\mathbf{S}')\mathbf{c}, & \text{(l)} \quad \frac{\partial(\mathbf{c}'\mathbf{S}^{-1}\mathbf{c})}{\partial\mathbf{S}} = -(\mathbf{S}')^{-1}\mathbf{cc}'(\mathbf{S}')^{-1}, & \end{array} \tag{A.2}$$

where $\|\mathbf{S}\|$ denote the absolute value of the determinant of $\mathbf{S}$. For the results of (i), (j), and (l), we assumed that $\mathbf{S}$ is non-singular at that point of d.

B Proofs

In the proof of Propositions, we use the following Lemma, which shows some properties of $\mathbf{G}_\theta(\cdot)$.

Lemma 1 *Let $\mathbf{A}$ and $\mathbf{C}$ be $m \times k$ and $k \times k$ matrices, respectively. Let $\mathbf{A}_i$ $(i = 1, ..., n)$ be matrices of the same size, where $n > 1$ is a positive integer. Let $\mathbf{B}_i$ $(i = 1, ..., n)$ be $p_i \times q_i$ matrices such that $\prod_{i=1}^n \mathbf{B}_i = \mathbf{B}_1\mathbf{B}_2\cdots\mathbf{B}_n$ is well defined (which implies $q_{i-1} = p_i$ for $i = 2, ..., n$). Suppose that each element of these matrices is a function of $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^h$, and whose first order partial derivatives with respect to each element of $\boldsymbol{\theta}$ exist and finite at a point of $\boldsymbol{\theta}$. Then, at this point,*

$$(a) \quad \mathbf{G}_\theta\left(\sum_{i=1}^n \mathbf{A}_i\right) = \sum_{i=1}^n \mathbf{G}_\theta(\mathbf{A}_i),$$

$$(b) \quad \mathbf{G}_\theta\left(\prod_{i=1}^n \mathbf{B}_i\right) = \sum_{i=1}^n \mathbf{G}_\theta(\mathbf{B}_i)\left[\left(\prod_{j=i+1}^{n+1}\mathbf{B}_j\right)\otimes\left(\prod_{k=n-i}^{n-1}\mathbf{B}'_{n-1-k}\right)\right], \ \textit{where } \mathbf{B}_0 = \mathbf{I}_{p_1} \textit{ and } \mathbf{B}_{n+1} = \mathbf{I}_{q_n},$$

$$(c) \quad \mathbf{G}_\theta(\mathbf{A}') = \mathbf{G}_\theta(\mathbf{A})\mathbf{K}_{km}.$$

If $\mathbf{C}$ *is non-singular at this point, then*

(*d*) $\quad \mathbf{G}_\theta(\mathbf{C}^{-1}) = -\mathbf{G}_\theta(\mathbf{C})(\mathbf{C}^{-1} \otimes \mathbf{C}'^{-1})$.

If $\mathbf{C}$ *is symmetric at this point, then,*

(*e*) $\quad \mathbf{G}_\theta(\mathbf{ACA}') = \mathbf{G}_\theta(\mathbf{A})(\mathbf{CA}' \otimes \mathbf{I}_m)\mathbf{N}_m + \mathbf{G}_\theta(\mathbf{C})(\mathbf{A}' \otimes \mathbf{A}')$,

(*f*) $\quad \mathbf{G}_\theta(\mathbf{C})(\mathbf{A}' \otimes \mathbf{A}')\mathbf{N}_m = 2\mathbf{G}_\theta(\mathbf{C})(\mathbf{A}' \otimes \mathbf{A}')$,

(*g*) $\quad \mathbf{G}_\theta(\mathbf{C})(\mathbf{I}_{k^2} - \mathbf{K}_k) = \mathbf{0}_{h\times k^2}$.

Proof of Lemma 1.
(a) It follows immediately from the properties of vec operator.
(b) Let θ_j be j-th element of $\boldsymbol{\theta}$. First, consider the case of $n = 2$. From (h) in (A.2), we have

$$\frac{\partial(\mathbf{B}_1\mathbf{B}_2)}{\partial\theta_j} = \frac{\partial\mathbf{B}_1}{\partial\theta_j}\mathbf{B}_2 + \mathbf{B}_1\frac{\partial\mathbf{B}_2}{\partial\theta_j}.$$

From (b) in (A.1), we have

$$\frac{\partial[\text{vec}(\mathbf{B}_1\mathbf{B}_2)']}{\partial\theta_j} = \frac{\partial[\text{vec}(\mathbf{B}_1)']}{\partial\theta_j}(\mathbf{B}_2 \otimes \mathbf{I}_{p_1}) + \frac{\partial[\text{vec}(\mathbf{B}_2)']}{\partial\theta_j}(\mathbf{I}_{q_2} \otimes \mathbf{B}_1'),$$

which implies

$$\mathbf{G}_\theta(\mathbf{B}_1\mathbf{B}_2) = \mathbf{G}_\theta(\mathbf{B}_1)(\mathbf{B}_2 \otimes \mathbf{I}_{p_1}) + \mathbf{G}_\theta(\mathbf{B}_2)(\mathbf{I}_{q_2} \otimes \mathbf{B}_1'). \tag{B.1}$$

By repeatedly applying (B.1) and (g) in (A.1), we have

$$\begin{aligned}
&\mathbf{G}_\theta(\mathbf{B}_1\mathbf{B}_2\cdots\mathbf{B}_n)\\
&= \mathbf{G}_\theta(\mathbf{B}_1)(\mathbf{B}_2\cdots\mathbf{B}_n \otimes \mathbf{I}_{p_1}) + \mathbf{G}_\theta(\mathbf{B}_2\cdots\mathbf{B}_n)(\mathbf{I}_{q_n} \otimes \mathbf{B}_1')\\
&= \mathbf{G}_\theta(\mathbf{B}_1)(\mathbf{B}_2\cdots\mathbf{B}_n \otimes \mathbf{I}_{p_1}) + [\mathbf{G}_\theta(\mathbf{B}_2)(\mathbf{B}_3\cdots\mathbf{B}_n \otimes \mathbf{I}_{p_2}) + \mathbf{G}_\theta(\mathbf{B}_3\cdots\mathbf{B}_n)(\mathbf{I}_{q_n} \otimes \mathbf{B}_2')](\mathbf{I}_{q_n} \otimes \mathbf{B}_1')\\
&= \mathbf{G}_\theta(\mathbf{B}_1)(\mathbf{B}_2\cdots\mathbf{B}_n \otimes \mathbf{I}_{p_1}) + \mathbf{G}_\theta(\mathbf{B}_2)(\mathbf{B}_3\cdots\mathbf{B}_n \otimes \mathbf{B}_1') + \mathbf{G}_\theta(\mathbf{B}_3\cdots\mathbf{B}_n)(\mathbf{I}_{q_n} \otimes \mathbf{B}_2'\mathbf{B}_1')\\
&\quad\vdots\\
&= \sum_{i=1}^{n}\mathbf{G}_\theta(\mathbf{B}_i)\left[\left(\prod_{j=i+1}^{n+1}\mathbf{B}_j\right) \otimes \left(\prod_{k=n-i}^{n-1}\mathbf{B}_{n-1-k}'\right)\right],
\end{aligned}$$

which is the desired result.
(c) From the properties of $\mathbf{K}_{mk}$, we have

$$\frac{\partial\text{vec}(\mathbf{A}')}{\partial\theta_j} = \frac{\partial[\mathbf{K}_{mk}\text{vec}(\mathbf{A})]}{\partial\theta_j} = \mathbf{K}_{mk}\frac{\partial\text{vec}(\mathbf{A})}{\partial\theta_j}. \tag{B.2}$$

Then, the result follows from (c) in (A.1) with the same argument used (B.1).
(d) If follows from (a), (b) in (A.1), (i) in (A.2), and the same argument in (B.1).
(e) From Lemma 1(b), Lemma 1(c), and (d) in (A.1), we have

$$\begin{aligned}
\mathbf{G}_\theta(\mathbf{ACA}') &= \mathbf{G}_\theta(\mathbf{A})(\mathbf{CA}' \otimes \mathbf{I}_m) + \mathbf{G}_\theta(\mathbf{C})(\mathbf{A}' \otimes \mathbf{A}') + \mathbf{G}_\theta(\mathbf{A}')(\mathbf{I}_m \otimes \mathbf{CA}'),\\
&= \mathbf{G}_\theta(\mathbf{A})(\mathbf{CA}' \otimes \mathbf{I}_m) + \mathbf{G}_\theta(\mathbf{C})(\mathbf{A}' \otimes \mathbf{A}') + \mathbf{G}_\theta(\mathbf{A})\mathbf{K}_{km}(\mathbf{I}_m \otimes \mathbf{CA}')\\
&= \mathbf{G}_\theta(\mathbf{A})(\mathbf{CA}' \otimes \mathbf{I}_m) + \mathbf{G}_\theta(\mathbf{C})(\mathbf{A}' \otimes \mathbf{A}') + \mathbf{G}_\theta(\mathbf{A})(\mathbf{CA}' \otimes \mathbf{I}_m)\mathbf{K}_m\\
&= \mathbf{G}_\theta(\mathbf{A})(\mathbf{CA}' \otimes \mathbf{I}_m)\mathbf{N}_m + \mathbf{G}_\theta(\mathbf{C})(\mathbf{A}' \otimes \mathbf{A}'),
\end{aligned}$$

which completes the proof of Lemma 1(e).
(f) From (d) in (A.1) and Lemma 1(c), we have

$$\begin{aligned}
\mathbf{G}_\theta(\mathbf{C})(\mathbf{A}' \otimes \mathbf{A}')\mathbf{N}_m &= \mathbf{G}_\theta(\mathbf{C})(\mathbf{A}' \otimes \mathbf{A}')(\mathbf{I}_{m^2} + \mathbf{K}_m)\\
&= \mathbf{G}_\theta(\mathbf{C})(\mathbf{A}' \otimes \mathbf{A}') + \mathbf{G}_\theta(\mathbf{C})(\mathbf{A}' \otimes \mathbf{A}')\mathbf{K}_m\\
&= \mathbf{G}_\theta(\mathbf{C})(\mathbf{A}' \otimes \mathbf{A}') + \mathbf{G}_\theta(\mathbf{C})\mathbf{K}_k(\mathbf{A}' \otimes \mathbf{A}')\\
&= \mathbf{G}_\theta(\mathbf{C})(\mathbf{A}' \otimes \mathbf{A}') + \mathbf{G}_\theta(\mathbf{C})(\mathbf{A}' \otimes \mathbf{A}')\\
&= 2\mathbf{G}_\theta(\mathbf{C})(\mathbf{A}' \otimes \mathbf{A}'),
\end{aligned}$$

which completes the proof of Lemma 1(f).
(g) From Lemma 1(c), we have

$$\begin{aligned}\mathbf{G}_\theta(\mathbf{C})(\mathbf{I}_{k^2}-\mathbf{K}_k) &= \mathbf{G}_\theta(\mathbf{C})-\mathbf{G}_\theta(\mathbf{C})\mathbf{K}_k \\ &= \mathbf{G}_\theta(\mathbf{C})-\mathbf{G}_\theta(\mathbf{C}) \\ &= \mathbf{0}_{h\times k^2},\end{aligned}$$

which completes the proof of Lemma 1(g). □

Proof of Proposition 1

We show only the derivation of (9) in detail because the derivations of other equations are entirely analogous. First, from Lemma 1(b) with $n=3$, we have

$$\mathbf{G}_\theta(\mathbf{B}_1\mathbf{B}_2\mathbf{B}_3)=\mathbf{G}_\theta(\mathbf{B}_1)(\mathbf{B}_2\mathbf{B}_3\otimes\mathbf{I}_{p_1})+\mathbf{G}_\theta(\mathbf{B}_2)(\mathbf{B}_3\otimes\mathbf{B}_1')+\mathbf{G}_\theta(\mathbf{B}_3)(\mathbf{I}_{q_3}\otimes\mathbf{B}_2'\mathbf{B}_1'). \tag{B.3}$$

Applying Lemma 1(a) and Lemma 1(e) to $\mathbf{P}_{t|t}$ in (3), we have

$$\mathbf{G}_\theta(\mathbf{P}_{t|t})=\mathbf{G}_\theta(\mathbf{P}_{t|t-1})-\mathbf{G}_\theta(\mathbf{M}_t)(\mathbf{Z}_t\mathbf{P}_{t|t-1}\otimes\mathbf{I}_m)\mathbf{N}_m-\mathbf{G}_\theta(\mathbf{F}_t)(\mathbf{M}_t'\otimes\mathbf{M}_t'). \tag{B.4}$$

From (B.3), Lemma 1(c), Lemma 1(d), and (d) and (g) in (A.1), we have

$$\begin{aligned}\mathbf{G}_\theta(\mathbf{M}_t) &= \mathbf{G}_\theta(\mathbf{P}_{t|t-1}\mathbf{Z}_t'\mathbf{F}_t^{-1}) \\ &= \mathbf{G}_\theta(\mathbf{P}_{t|t-1})(\mathbf{Z}_t'\mathbf{F}_t^{-1}\otimes\mathbf{I}_m)+\mathbf{G}_\theta(\mathbf{Z}_t')(\mathbf{F}_t^{-1}\otimes\mathbf{P}_{t|t-1})+\mathbf{G}_\theta(\mathbf{F}_t^{-1})(\mathbf{I}_v\otimes\mathbf{Z}_t\mathbf{P}_{t|t-1}) \\ &= \mathbf{G}_\theta(\mathbf{P}_{t|t-1})(\mathbf{Z}_t'\mathbf{F}_t^{-1}\otimes\mathbf{I}_m)+\mathbf{G}_\theta(\mathbf{Z}_t)\mathbf{K}_{mv}(\mathbf{F}_t^{-1}\otimes\mathbf{P}_{t|t-1})-\mathbf{G}_\theta(\mathbf{F}_t)(\mathbf{F}_t^{-1}\otimes\mathbf{F}_t^{-1})(\mathbf{I}_v\otimes\mathbf{Z}_t\mathbf{P}_{t|t-1}) \\ &= \mathbf{G}_\theta(\mathbf{P}_{t|t-1})(\mathbf{Z}_t'\mathbf{F}_t^{-1}\otimes\mathbf{I}_m)+\mathbf{G}_\theta(\mathbf{Z}_t)(\mathbf{P}_{t|t-1}\otimes\mathbf{F}_t^{-1})\mathbf{K}_{mv}-\mathbf{G}_\theta(\mathbf{F}_t)(\mathbf{F}_t^{-1}\otimes\mathbf{M}_t').\end{aligned} \tag{B.5}$$

From (B.5), (d), (e), and (g) in (A.1), and Lemma 1(f), we have

$$\begin{aligned}&\mathbf{G}_\theta(\mathbf{M}_t)(\mathbf{Z}_t\mathbf{P}_{t|t-1}\otimes\mathbf{I}_m)\mathbf{N}_m \\ &= \mathbf{G}_\theta(\mathbf{P}_{t|t-1})(\mathbf{Z}_t'\mathbf{M}_t'\otimes\mathbf{I}_m)\mathbf{N}_m+\mathbf{G}_\theta(\mathbf{Z}_t)(\mathbf{P}_{t|t-1}\otimes\mathbf{M}_t')\mathbf{N}_m-\mathbf{G}_\theta(\mathbf{F}_t)(\mathbf{M}_t'\otimes\mathbf{M}_t')\mathbf{N}_m \\ &= \mathbf{G}_\theta(\mathbf{P}_{t|t-1})(\mathbf{Z}_t'\mathbf{M}_t'\otimes\mathbf{I}_m)\mathbf{N}_m+\mathbf{G}_\theta(\mathbf{Z}_t)(\mathbf{P}_{t|t-1}\otimes\mathbf{M}_t')\mathbf{N}_m-2\mathbf{G}_\theta(\mathbf{F}_t)(\mathbf{M}_t'\otimes\mathbf{M}_t').\end{aligned} \tag{B.6}$$

Substituting (B.6) into (B.4), we obtain (9). □

Proof of Proposition 2

From (5), (6), (7), (b) in (A.1), and noting that $(\mathbf{N}_v/2)\mathrm{vec}(\mathbf{S})=\mathrm{vec}(\mathbf{S})$ for a v by v symmetric matrix $\mathbf{S}$, we have

$$\frac{\partial(\mathbf{v}_t')}{\partial\boldsymbol{\theta}}\frac{\partial\ell_t}{\partial\mathbf{v}_t}=-\mathbf{G}_\theta(\mathbf{v}_t)\mathbf{w}_t=\mathbf{G}_\theta(\mathbf{d}_t)\mathbf{w}_t+\mathbf{G}_\theta(\mathbf{Z}_t)\mathrm{vec}(\mathbf{w}_t\mathbf{a}_{t|t-1}')+\mathbf{G}_\theta(\mathbf{a}_{t|t-1})\mathbf{Z}_t'\mathbf{w}_t, \tag{B.7}$$

and

$$\begin{aligned}\frac{\partial[\mathrm{vec}(\mathbf{F}_t)']}{\partial\boldsymbol{\theta}}\frac{\partial\ell_t}{\partial\mathrm{vec}(\mathbf{F}_t)} &= \mathbf{G}_\theta(\mathbf{F}_t)\left[\frac{1}{2}\mathrm{vec}(\mathbf{w}_t\mathbf{w}_t'-\mathbf{F}_t^{-1})\right] \\ &= \mathbf{G}_\theta(\mathbf{Z}_t)\mathrm{vec}(\mathbf{w}_t\mathbf{v}_t'\mathbf{M}_t'-\mathbf{M}_t')+\tfrac{1}{2}\mathbf{G}_\theta(\mathbf{S}_t)\mathrm{vec}(\mathbf{w}_t\mathbf{w}_t'-\mathbf{F}_t^{-1}) \\ &\quad+\tfrac{1}{2}\mathbf{G}_\theta(\mathbf{P}_{t|t-1})\mathrm{vec}(\mathbf{Z}_t'\mathbf{w}_t\mathbf{w}_t'\mathbf{Z}_t-\mathbf{Z}_t'\mathbf{F}_t^{-1}\mathbf{Z}_t).\end{aligned} \tag{B.8}$$

From (4), (B.7), and (B.8), we have (12). Substituting (6) and (7) into the right hand sides in (8) and (9) and noting that $\mathbf{N}_v(\mathbf{M}_t'\otimes\mathbf{M}_t')=(\mathbf{M}_t'\otimes\mathbf{M}_t')\mathbf{N}_m$, $\mathbf{B}=1\otimes\mathbf{B}=\mathbf{B}\otimes 1$, $\mathbf{B}\otimes\mathbf{C}\pm\mathbf{B}\otimes\mathbf{D}=\mathbf{B}\otimes(\mathbf{C}\pm\mathbf{D})$ and $\mathbf{C}\otimes\mathbf{B}\pm\mathbf{D}\otimes\mathbf{B}=(\mathbf{C}\pm\mathbf{D})\otimes\mathbf{B}$ for matrices $\mathbf{C}$ and $\mathbf{D}$ of the same size, we have

$$\begin{aligned}\mathbf{G}_\theta(\mathbf{a}_{t|t}) &= \mathbf{G}_\theta(\mathbf{a}_{t|t-1})(\mathbf{I}_m-\mathbf{Z}_t'\mathbf{M}_t')+\mathbf{G}_\theta(\mathbf{P}_{t|t-1})[\mathbf{Z}_t'\mathbf{w}_t\otimes(\mathbf{I}_m-\mathbf{Z}_t'\mathbf{M}_t')]-\mathbf{G}_\theta(\mathbf{d}_t)\mathbf{M}_t' \\ &\quad-\mathbf{G}_\theta(\mathbf{S}_t)(\mathbf{w}_t\otimes\mathbf{M}_t')+\mathbf{G}_\theta(\mathbf{Z}_t)[\mathbf{P}_{t|t-1}(\mathbf{I}_m-\mathbf{Z}_t'\mathbf{M}_t')\otimes\mathbf{w}_t-(\mathbf{M}_t\mathbf{v}_t+\mathbf{a}_{t|t-1})\otimes\mathbf{M}_t'],\end{aligned} \tag{B.9}$$

and

$$\begin{aligned}\mathbf{G}_\theta(\mathbf{P}_{t|t}) &= \mathbf{G}_\theta(\mathbf{P}_{t|t-1})[\mathbf{I}_{m^2}-(\mathbf{Z}_t'\mathbf{M}_t'\otimes\mathbf{I}_m)\mathbf{N}_m+(\mathbf{Z}_t'\mathbf{M}_t'\otimes\mathbf{Z}_t'\mathbf{M}_t')] \\ &\quad+\mathbf{G}_\theta(\mathbf{S}_t)(\mathbf{M}_t'\otimes\mathbf{M}_t')-\mathbf{G}_\theta(\mathbf{Z}_t)[\mathbf{P}_{t|t-1}(\mathbf{I}_m-\mathbf{Z}_t'\mathbf{M}_t')\otimes\mathbf{M}_t']\mathbf{N}_m.\end{aligned} \tag{B.10}$$

Substituting (B.9) and (B.10) into (10) and (11), noting $\mathbf{P}_{t|t}\mathbf{T}'_{t+1} = \mathbf{P}_{t|t-1}\mathbf{L}'_t$, and applying Lemma 1(g), we have (13) and

$$\begin{aligned}
&\mathbf{G}_\theta(\mathbf{P}_{t+1|t}) \\
&= \mathbf{G}_\theta(\mathbf{P}_{t|t-1})[(\mathbf{T}'_{t+1} \otimes \mathbf{T}'_{t+1}) - (\mathbf{Z}'_t\mathbf{J}'_t \otimes \mathbf{T}'_{t+1})\mathbf{N}_m + (\mathbf{Z}'_t\mathbf{J}'_t \otimes \mathbf{Z}'_t\mathbf{J}'_t)] \\
&\quad + \mathbf{G}_\theta(\mathbf{S}_t)(\mathbf{J}'_t \otimes \mathbf{J}'_t) - \mathbf{G}_\theta(\mathbf{Z}_t)(\mathbf{P}_{t|t-1}\mathbf{L}_t \otimes \mathbf{J}'_t)\mathbf{N}_m + \mathbf{G}_\theta(\mathbf{Q}_{t+1})(\mathbf{R}'_{t+1} \otimes \mathbf{R}'_{t+1}) \\
&\quad + [\mathbf{G}_\theta(\mathbf{T}_{t+1})(\mathbf{P}_{t|t}\mathbf{T}'_{t+1} \otimes \mathbf{I}_m) + \mathbf{G}_\theta(\mathbf{R}_{t+1})(\mathbf{Q}_{t+1}\mathbf{R}'_{t+1} \otimes \mathbf{I}_m)]\mathbf{N}_m \\
&= \mathbf{G}_\theta(\mathbf{P}_{t|t-1})[(\mathbf{L}'_t \otimes \mathbf{L}'_t) + (\mathbf{I}_{m^2} - \mathbf{K}_m)(\mathbf{T}'_{t+1} \otimes \mathbf{Z}'_t\mathbf{J}'_t)] + \mathbf{G}_\theta(\mathbf{S}_t)(\mathbf{J}'_t \otimes \mathbf{J}'_t) + \mathbf{G}_\theta(\mathbf{Q}_{t+1})(\mathbf{R}'_{t+1} \otimes \mathbf{R}'_{t+1}) \\
&\quad + [\mathbf{G}_\theta(\mathbf{T}_{t+1})(\mathbf{P}_{t|t-1}\mathbf{L}'_t \otimes \mathbf{I}_m) + \mathbf{G}_\theta(\mathbf{R}_{t+1})(\mathbf{Q}_{t+1}\mathbf{R}'_{t+1} \otimes \mathbf{I}_m) - \mathbf{G}_\theta(\mathbf{Z}_t)(\mathbf{P}_{t|t-1}\mathbf{L}'_t \otimes \mathbf{J}'_t)]\mathbf{N}_m \\
&= \mathbf{G}_\theta(\mathbf{P}_{t|t-1})(\mathbf{L}'_t \otimes \mathbf{L}'_t) + \mathbf{G}_\theta(\mathbf{S}_t)(\mathbf{J}'_t \otimes \mathbf{J}'_t) + \mathbf{G}_\theta(\mathbf{Q}_{t+1})(\mathbf{R}'_{t+1} \otimes \mathbf{R}'_{t+1}) \\
&\quad + [\mathbf{G}_\theta(\mathbf{T}_{t+1})(\mathbf{P}_{t|t-1}\mathbf{L}'_t \otimes \mathbf{I}_m) - \mathbf{G}_\theta(\mathbf{Z}_t)(\mathbf{P}_{t|t-1}\mathbf{L}'_t \otimes \mathbf{J}'_t) + \mathbf{G}_\theta(\mathbf{R}_{t+1})(\mathbf{Q}_{t+1}\mathbf{R}'_{t+1} \otimes \mathbf{I}_m)]\mathbf{N}_m,
\end{aligned} \tag{B.11}$$

where $\mathbf{L}_t = \mathbf{T}_{t+1} - \mathbf{J}_t\mathbf{Z}_t$, and $\mathbf{J}_t = \mathbf{T}_{t+1}\mathbf{M}_t$, which completes the proof of Proposition 2. □

Proof of Proposition 3

The result can be obtained by an argument entirely analogous to the ones in the proof of Propositions 1 and 2. Therefore the proof is omitted. □

References

[1] P. de Jong, J. Penzer, The ARMA model in state space form. Statist. Probab. Lett. 70(2004) 119-125.

[2] A.C. Harvey, Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge, Academic Press, 1989.

[3] M.A. Hooker, Analytic first and second derivatives for the recursive prediction error algorithm's log likelihood function, IEEE Trans. Autom. Control. 39(3) (1994) 662-664.

[4] S.J. Koopman, N. Shephard, Exact score for time series models in state space form. Biometrika 79(1992) 823-826.

[5] D.J. Hamilton, Time Series Analysis. Princeton University Press, New Jersey, 1994.

[6] J.R. Magnus, H. Neudecker, The commutation matrix: some properties and applications. Ann. Statist. 7(1979) 381-394.

[7] J.R. Magnus, H. Neudecker, Matrix Differential Calculus with Applications in Statistics and Econometrics, revised edition. John Wiley & Suns: New York, 1999.

[8] M. Segal, E. Weinstein, A new method for evaluating the log-likelihood gradient, the Hessian, and the Fisher information matrix for linear dynamic systems. IEEE Trans. Theory 35(1989) 682-687.