A NOTE ON PARAMETER DIFFERENTIATION OF MATRIX EXPONENTIALS, WITH APPLICATIONS TO CONTINUOUS-TIME MODELING

(SHORT RUNNING TITLE: PARAMETER DIFFERENTIATION OF MATRIX EXPONENTIALS)

HENGHSIU TSAI

Institute of Statistical Science, Academia Sinica, Taipei, Taiwan 115, R.O.C.

htsai@stat.sinica.edu.tw

K. S. CHAN

Department of Statistics & Actuarial Science, University of Iowa, Iowa City, IA 52242, U.S.A.

kchan@stat.uiowa.edu

December 27, 2001

Summary

We have derived a new analytic formula for evaluating the derivatives of a matrix exponential. In contrast to some existing methods, eigenvalues and eigenvectors do not appear explicitly in the formulae, although we show that a necessary and sufficient condition for the validity of the formulae is that the matrix has distinct eigenvalues. The new formula expresses the derivatives of a matrix exponential in terms of minors, polynomials, exponential of the matrix as well as matrix inver-

sion, and hence is algebraically more manageable. For sparse matrices, the formula can be further simplified. Two examples are discussed in some details. For the companion matrix of a continuous-time autoregressive moving average process, the derivatives of the exponential of the companion matrix can be computed recursively. We illustrate an use of these formulae in the conditional least square estimation of a CAR(p) model that leads to a numerically stable estimation procedure. The second example concerns the exponential of the tridiagonal transition intensity matrix of a finite-state-space continuous-time Markov chain whose instantaneous transitions must be between adjacent states.
Keywords: Cayley-Hamilton Theorem; CARMA models; companion matrix; finite-state-space continuous-time Markov processes; maximum likelihood estimation; minimal polynomial; tridiagonal matrix.

1 Introduction

Various methods of parameter differentiation of a matrix exponential have been studied in statistical mechanics and quantum theory (see, e.g., Wilcox, 1967), as well as in the mathematics, economics and statistics literature, see e.g., Jennrich and Bright (1976), Van Loan (1978), Kalbfleisch and Lawless (1985), Graham (1986), Horn and Johnson (1991), Chan and Munoz-Hernandez (1997), Chen and Zadrozny (2001). For continuous/discrete state space modelling (see, e.g., Jazwinski, 1970 and Singer, 1995), parameter differentiation of a matrix exponential is needed for computing the analytical score function. For continuous-time Markov modeling, efficient algorithm for the computation of the transition probability matrix and its derivatives with respect to the transition intensity parameters is needed for maximum likelihood estimation. For example, see Kalbfleisch and Lawless (1985) for an approach of analyzing a panel of categorical data by assuming that the data are obtained from sampling a latent continuous-time finite-state-space Markov process.

We propose in this note an alternative method for computing the derivatives of a matrix exponential. In contrast to some existing methods, eigenvalues and eigenvectors do not appear explicitly in the formulae, although we show that a necessary and sufficient condition for the validity of the formulae is that the matrix has distinct eigenvalues. The new formula expresses the derivatives of a matrix exponential in terms of minors, polynomials, exponential of the matrix as well as matrix inversion, and hence is algebraically more manageable. When the matrix has repeated eigenvalues, it seems hard to extend the results. See the end of section 3 for discussion. Fortunately, in most statistical applications that involve matrix exponentials, the distinct eigenvalue assumption often holds. For example, in continuous-time Markov chain modelling, for most models of interest, the transition intensity matrix has distinct eigenvalues for almost all parameter values (see, e.g., Kalbfleisch and Lawless, 1985).

This note is organized as follows. In § 2, we derive the new formula for computing the derivatives of a matrix exponential and a necessary and sufficient condition for the validity of the formula. For sparse matrices, the formula may be further simplified. Two interesting examples are the exponential of the companion matrix arising from a continuous-time autoregressive moving average process and that of the tridiagonal transition intensity matrix arising from a continuous-time Markov chain whose instantaneous transitions must be jumps between adjacent categories. The simplified formulae for these two examples are given in §3.

2 Main results

Let $A = [a_{ij}]$ be a $p \times p$ matrix whose elements are functions of $\vartheta = (\vartheta_1, ..., \vartheta_r)'$. By equation (2.1) of Wilcox (1967), we have that, for $i = 1, ..., r$,

$$\frac{\partial e^{tA}}{\partial \vartheta_i} = \int_0^t e^{(t-u)A} \left(\frac{\partial A}{\partial \vartheta_i} \right) e^{uA} du. \tag{1}$$

Alternatively, if we assume A has distinct eigenvalues $d_1, \cdots, d_p$ and X is the $p \times p$ matrix whose jth column is a right eigenvector corresponding to d_j, then $A = XDX^{-1}$, where $D = \mathrm{diag}(d_1, \cdots, d_p)$. Then $e^{tA} = X\ \mathrm{diag}(e^{d_1 t}, \cdots, e^{d_p t})X^{-1}$, and

$$\frac{\partial e^{tA}}{\partial \vartheta_u} = XV_uX^{-1}, \qquad u = 1, \cdots, r, \tag{2}$$

where V_u is a $p \times p$ matrix with (i, j) entry

$$g_{ij}^{(u)}(e^{d_i t} - e^{d_j t})/(d_i - d_j), \qquad i \neq j,$$

$$g_{ii}^{(u)} t e^{d_i t}, \qquad i = j,$$

and $g_{ij}^{(u)}$ is the (i, j) entry in $G^{(u)} = X^{-1}(\partial A/\partial\theta_u)X$. See Kalbfleisch and Lawless (1985) for the above formula and related discussions. See also Jennrich and Bright (1976) and Chan and Munos-Hernandez (1997). When A has repeated eigenvalues, an analogous decomposition of A to Jordan canonical form is possible (see chapter 4 of Cox and Miller, 1965). But as pointed out by Kalbfleisch and Lawless (1985), this is rarely necessary, since for most models of interest in continuous-time Markov modelling, A has distinct eigenvalues for almost all parameters.

One of the main results of this paper is to derive another closed form solution for $\partial e^{tA}/\partial\vartheta_i$. For $r = 1, ..., p$, define δ_r to be a $p \times 1$ vector with 1 in position r and 0 elsewhere. For $1 \le i, j \le p$, let $B_{ij} = \delta_i \delta_j'$, and define

$$\Sigma_{ij} = \int_0^t e^{(t-u)A} B_{ij} e^{uA} du. \tag{3}$$

Note that B_{ij} is a zero p by p matrix except for its (i, j)th element being unity, and, for $1 \le k \le r$,

$$\begin{aligned} \frac{\partial e^{tA}}{\partial\vartheta_k} &= \int_0^t e^{(t-u)A} \frac{\partial A}{\partial\vartheta_k} e^{uA} du \\ &= \int_0^t e^{(t-u)A} \left(\sum_{i=1}^p \sum_{j=1}^p \frac{\partial a_{ij}}{\partial\vartheta_k} B_{ij} \right) e^{uA} du \\ &= \sum_{i=1}^p \sum_{j=1}^p \frac{\partial a_{ij}}{\partial\vartheta_k} \int_0^t e^{(t-u)A} B_{ij} e^{uA} du \\ &= \sum_{i=1}^p \sum_{j=1}^p \frac{\partial a_{ij}}{\partial\vartheta_k} \Sigma_{ij}. \end{aligned} \tag{4}$$

A closed form solution for Σ_{ij} in terms of minors, polynomials, exponential of the matrix A as well as matrix inversion is given in Theorem 1.

Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be two $p \times p$ matrices. Define $[A, B] = AB - BA$ as the commutator of A and B, and let $|A|$ be the determinant of the matrix A. For vectors $\alpha = [\alpha_1, \cdots, \alpha_q]$ and $\beta = [\beta_1, \cdots, \beta_q]$, where $\alpha_j \in \{1, \cdots, p\}$ and $\beta_j \in \{1, \cdots, p\}$, for $j = 1, \cdots, q (\le p)$, we denote the (sub)matrix that lies in the rows of A indexed by α and the columns indexed by β as $A(\alpha, \beta)$. For example,

$$A([1, 3], [2, 1, 3]) = \begin{bmatrix} a_{12} & a_{11} & a_{13} \\ a_{32} & a_{31} & a_{33} \end{bmatrix}.$$

If $\beta = \alpha$, the submatrix $A(\alpha, \alpha)$ is called a principal submatrix of A and is abbreviated $A(\alpha)$, see, e.g., p. 17 of Horn and Johnson (1985). Let $R_0^p = 1$, and for $1 \leq k \leq p$, let

$$R_k^p = \sum_{1 \leq l_1 < \cdots < l_k \leq p} |A([l_1, \cdots, l_k])|. \tag{5}$$

Note that $q(\lambda) = \sum_{k=0}^{p} (-1)^{p-k} R_{p-k}^p \lambda^k$ is the characteristic polynomial of the matrix A, and $q(A) = 0$ by Cayley-Hamilton Theorem. Theorem 1 below essentially results from differentiating the preceding equality w.r.t. $a_{i,j}$. Let $q'(\lambda)$ be the (first) derivative of q w.r.t. λ. Then, $q'(A) = \sum_{k=0}^{p-1} (-1)^{p-k-1} (k+1) R_{p-k-1}^p A^k$ and it is independent of t. This fact may result in simpler inferential procedures as will be illustrated in an example below. The derivatives of the matrix exponential is trivial when $p = 1$. For $p \geq 2$, we have the following results:

THEOREM 1 *For $p \geq 2$ and assuming that $q'(A) = \sum_{k=0}^{p-1} (-1)^{p-k-1} (k+1) R_{p-k-1}^p A^k$ is invertible,*

$$\begin{aligned} \Sigma_{ij} \;=\; & \left\{ \sum_{k=0}^{p-1} (-1)^{p-k-1} (k+1) R_{p-k-1}^p A^k \right\}^{-1} \left[\left\{ \sum_{k=0}^{p-1} (-1)^{p-k+1} \left(\frac{\partial R_{p-k}^p}{\partial a_{ij}} \right) A^k \right\} t e^{tA} \right. \\ & \left. - \sum_{u=0}^{p-2} \sum_{k=u+2}^{p} (-1)^{p-k} (k-u-1) R_{p-k}^p A^{k-u-2} [B_{ij}, e^{tA}] A^u \right]. \end{aligned}$$

Theorem 2 gives an explicit representation of the partial derivatives of R_k^p's with respect to a_{ij}s, whereas Theorem 3 gives a necessary and sufficient condition for $q'(A)$ to be invertible.

THEOREM 2 *(a) For $1 \leq i \neq j \leq p$,*

$$\frac{\partial R_1^p}{\partial a_{ij}} = 0.$$

(b) For $1 \leq i \neq j \leq p$,

$$\frac{\partial R_2^p}{\partial a_{ij}} = -|A([j], [i])| = -a_{ji}.$$

(c) For $3 \leq k \leq p$, and $1 \leq i \neq j \leq p$,

$$\frac{\partial R_k^p}{\partial a_{ij}} = - \sum_{\substack{1 \leq l_1 < \cdots < l_{k-2} \leq p \\ i \notin \{l_1, \cdots, l_{k-2}\} \\ and\ j \notin \{l_1, \cdots, l_{k-2}\}}} |A([j, l_1, \cdots, l_{k-2}], [i, l_1, \cdots, l_{k-2}])|.$$

(d) For $1 \leq i \leq p$,

$$\frac{\partial R_1^p}{\partial a_{ii}} = 1.$$

(e) For $2 \leq k \leq p$, and $1 \leq i \leq p$,

$$\frac{\partial R_k^p}{\partial a_{ii}} = \sum_{\substack{1 \leq l_1 < \cdots < l_{k-1} \leq p \\ i \notin \{l_1, \cdots, l_{k-1}\}}} |A([l_1, \cdots, l_{k-1}])|.$$

THEOREM 3 *For $p \geq 2$, $q'(A) = \sum_{k=0}^{p-1} (-1)^{p-k-1}(k+1) R_{p-k-1}^p A^k$ is invertible if and only if the matrix A has p distinct eigenvalues.*

In the case that A has repeated eigenvalues, Theorem 3 implies that $q'(A)$ is singular so that Theorem 1 is inapplicable. Now, Theorem 1 may be generalized by considering the equation $m(A) = 0$ where $m(\lambda)$ is the minimal polynomial of A. Indeed, if A is diagonalizable, its minimal polynomial equals $m(\lambda) = \prod(\lambda - \lambda_j)$, where the product is over distinct eigenvalues, in which case, even though the eigenvalues are not distinct, they do not repeat in the minimal polynomial so that $m'(A)$ is invertible. This suggests that the preceding results may be extended to the more general case that A is diagonalizable, or equivalently, its minimal polynomial is of the form $m(\lambda) = \prod(\lambda - \lambda_j)$, where all λ_j's are distinct. However, the coefficients of the minimal polynomial may not admit a simple form. Moreover, Theorem 2 and related results seem not be easily generalizable in this more general situation.

3 Applications

3.1 Continuous-time autoregressive moving average processes

For continuous/discrete state space modelling (see, e.g., Jazwinski, 1970 and Singer, 1995), parameter differentiation of a matrix exponential is needed in computing the analytical score function; indeed, it is also required in other methods of estimation, e.g., least squares. The continuous/discrete state space model is defined by two equations:

$$
\begin{aligned}
dx_n(t) &= Ax_n(t)dt + Bz_n(t)dt + GdW_n(t), \qquad t \in [t_0, t_T] && (6)\\
y_n(t_i) &= H_{ni}x_n(t_i) + D_{ni}z_n(t_i) + \epsilon_{ni}, && (7)
\end{aligned}
$$

where $\epsilon_{ni} \sim N(0, R_{ni})$ is a discrete time white noise disturbance (measurement error), $W_n(t)$ is the standard Brownian motion, z_n is a covariate and the matrices H_{ni}, D_{ni} and R_{ni} are obtained from H, D, and R by dropping the respective rows (and columns) if the datum $y_n(t_i)$ contains missing values. State equation (6) is a linear stochastic differential equation in the sense of Itô (cf. Arnold, 1974). See Singer (1995) for further discussions.

A continuous-time autoregressive moving average (CARMA(p,q)) process is defined as the solution of a differential equation that can be cast in the state-space form (below, the Ys are the observations with Xs being the state vectors; see, e.g., Brockwell, 1993 and Brockwell and Stramer, 1995 for further discussions):

$$
\begin{aligned}
Y_{t_i} &= \beta^T X_{t_i}, \qquad i = 1, 2, \cdots, n,\\
dX_t &= (AX_t + \alpha_0 l)dt + \sigma l dW_t,
\end{aligned}
$$

where

$$
A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ \alpha_1 & \alpha_2 & \alpha_3 & \cdots & \alpha_p \end{bmatrix}, \quad X_t = \begin{bmatrix} X_t^{(0)} \\ X_t^{(1)} \\ \vdots \\ X_t^{(p-2)} \\ X_t^{(p-1)} \end{bmatrix}, \quad l = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}, \beta = \begin{bmatrix} 1 \\ \beta_1 \\ \vdots \\ \beta_{p-2} \\ \beta_{p-1} \end{bmatrix},
$$

and the superscript ^{T} denotes the transpose of a vector.

Note that the companion matrix A is a function of the parameters $\alpha_1, \cdots, \alpha_p$, and due to the simplicity of the matrix, parameter differentiation of the corresponding matrix exponential can be easily computed by the recursive procedure:

$$
\frac{\partial e^{tA}}{\partial \alpha_i} = \left(\frac{\partial e^{tA}}{\partial \alpha_{i-1}}\right) A, \qquad 2 \le i \le p,
$$

see Theorem 4 (c) and the appendix for a proof. The partial derivative of e^{tA} with respect to α_1 is given by parts (a) and (b) of the following theorem.

THEOREM 4 *(a) For* $p=1$, $\partial e^{tA}/\partial\alpha_1 = te^{tA}$.
(b) For $p \geq 2$,

$$\frac{\partial e^{tA}}{\partial \alpha_1} = K_{p,0}^{-1}\left\{te^{tA} - \sum_{r=1}^{p-1} K_{p,r}[B_{pr}, e^{tA}]\right\}, \tag{8}$$

where

$$K_{p,r} = \begin{cases} (p-r)A^{p-r-1} - \sum_{k=r+2}^{p}(k-r-1)\alpha_k A^{k-r-2}, & 0 \leq r \leq p-2, \\ I, & r = p-1, \end{cases}$$

where $[B_{pr}, e^{tA}] = B_{pr}e^{tA} - e^{tA}B_{pr}$ *is the commutator of* B_{pr} *and* e^{tA}.
(c) For $2 \leq i \leq p$,

$$\frac{\partial e^{tA}}{\partial \alpha_i} = \left(\frac{\partial e^{tA}}{\partial \alpha_{i-1}}\right) A. \tag{9}$$

For clarification, the expressions of the matrix $\partial e^{tA}/\partial\alpha_1$, for $p = 1, ..., 4$, are illustrated as follows:

$$\begin{aligned}
p=1 \quad &: \quad \frac{\partial e^{tA}}{\partial \alpha_1} = te^{tA}, \\
p=2 \quad &: \quad \frac{\partial e^{tA}}{\partial \alpha_1} = (2A - \alpha_2 I)^{-1}(te^{tA} - [B_{21}, e^{tA}]), \\
p=3 \quad &: \quad \frac{\partial e^{tA}}{\partial \alpha_1} = (3A^2 - 2\alpha_3 A - \alpha_2 I)^{-1}\{te^{tA} - (2A - \alpha_3 I)[B_{31}, e^{tA}] - [B_{32}, e^{tA}]\} \\
p=4 \quad &: \quad \frac{\partial e^{tA}}{\partial \alpha_1} = (4A^3 - 3\alpha_4 A^2 - 2\alpha_3 A - \alpha_2 I)^{-1}\{te^{tA} \\
&\qquad -(3A^2 - 2\alpha_4 A - \alpha_3 I)[B_{41}, e^{tA}] - (2A - \alpha_4 I)[B_{42}, e^{tA}] - [B_{43}, e^{tA}]\}.
\end{aligned}$$

We now present an example illustrating the use of the new formulae. Suppose that we observed the states X_t from a CAR(p) model over (possibly) unequally spaced epoches, say, t_is, and we desire to compute the conditional least squares estimators of the parameters. First note that, the sum of squared predictive residuals equals

$$g(\alpha_0, \cdots, \alpha_p) = \sum_{i=1}^{N}\{x_{t_i} - \mu - e^{\Delta_i A}(x_{t_{i-1}} - \mu)\}^T\{x_{t_i} - \mu - e^{\Delta_i A}(x_{t_{i-1}} - \mu)\},$$

where $\Delta_i = t_i - t_{i-1}$ and $\mu = (-\alpha_0/\alpha_1, 0, \cdots, 0)^T$. For simplicity asssume that $\alpha_0 = 0$ so that $\mu = 0$. Therefore, for $1 \leq j \leq p$,

$$\frac{\partial g(\alpha_0, \cdots, \alpha_p)}{\partial \alpha_j}$$

$$
\begin{aligned}
&= -2\sum_{i=1}^{N}\left(\frac{\partial e^{\Delta_i A}}{\partial \alpha_j}x_{t_{i-1}}\right)^T (x_{t_i} - e^{\Delta_i A}x_{t_{i-1}}) \\
&= -2\sum_{i=1}^{N} x_{t_{i-1}}^T (A^T)^{j-1}\left(\Delta_i e^{\Delta_i A} - \sum_{r=1}^{p-1} K_{p,r}[B_{pr}, e^{\Delta_i A}]\right)^T (K_{p,0}^{-1})^T (x_{t_i} - e^{\Delta_i A}x_{t_{i-1}}) \\
&= -2\sum_{i=1}^{N} \operatorname{tr}\left\{(x_{t_i} - e^{\Delta_i A}x_{t_{i-1}})x_{t_{i-1}}^T (A^T)^{j-1}\left(\Delta_i e^{\Delta_i A} - \sum_{r=1}^{p-1} K_{p,r}[B_{pr}, e^{\Delta_i A}]\right)^T (K_{p,0}^{-1})^T\right\} \\
&= -2\left[\sum_{i=1}^{N} \operatorname{vec}\left\{\left(\Delta_i e^{\Delta_i A} - \sum_{r=1}^{p-1} K_{p,r}[B_{pr}, e^{\Delta_i A}]\right) A^{j-1}x_{t_{i-1}}(x_{t_i} - e^{\Delta_i A}x_{t_{i-1}})^T\right\}\right]^T \operatorname{vec}[(K_{p,0}^{-1})^T].
\end{aligned}
$$

By replacing $K_{p,0}^{-1}$ by the adjoint of $K_{p,0}$ in the preceding expression, high numerical accuracy can be attained even when some of the eigenvalues are nearly identical.

3.2 Tridiagnal intensity matrix in continuous-time Markov processes

Kalbfleisch and Lawless (1985) proposed methods for the analysis of panel data under a continuous-time Markov model with a finite state space. Let Q be a $p \times p$ transition intensity matrix that is constant over an interval of length t, and e^{tQ} is the corresponding transition probability matrix. For some applications, the matrix Q is a sparse matrix in the sense that only a few elements of Q are non-zero. See Kalbfleisch and Lawless (1985) for examples. Chan and Munos-Hernandez (1997) adopted the continuous-time Markov processes to model longitudinal data consisting of transitional frequencies classified according to an ordered categorical response variable. The ordering of the categories implies that the continuous-time Markov chain can only jump between adjacent categories over an infinitesimal period, resulting in a tridiagonal transition intensity matrix. For the tridiagonal transition intensity matrix, the coefficients R_k^p's and their partial derivatives with respect to $q_{i,j}$'s, which have closed form solutions given in Theorem 2, can be further simplified as in Theorem 5 below. From now on, assume $p \geq 2$, and write the tridiagonal transition intensity matrix as

$$Q_p = \begin{bmatrix} -q_1 & q_1 & 0 & \cdots & 0 & 0 & 0 \\ q_2 & -q_2 - q_3 & q_3 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & q_{2p-4} & -q_{2p-4} - q_{2p-3} & q_{2p-3} \\ 0 & 0 & 0 & \cdots & 0 & q_{2p-2} & -q_{2p-2} \end{bmatrix}.$$

For $1 \leq i \leq p$, let

$$Q_p^i = \begin{bmatrix} -q_1^i & q_1^i & 0 & \cdots & 0 & 0 & 0 \\ q_2^i & -q_2^i - q_3^i & q_3^i & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & q_{2p-4}^i & -q_{2p-4}^i - q_{2p-3}^i & q_{2p-3}^i \\ 0 & 0 & 0 & \cdots & 0 & q_{2p-2}^i & -q_{2p-2}^i \end{bmatrix},$$

where, for $1 \leq k \leq 2p-2$,

$$q_k^i = \begin{cases} q_k, & \text{if } i \notin \{2i-2, 2i-1\}, \\ 0, & \text{if } k \in \{2i-2, 2i-1\}. \end{cases}$$

For $1 \leq i \leq p$, let $R_{0,i}^p = 1$. For $p \geq 2$ and $1 \leq i, k \leq p$, define

$$R_{k,i}^p = \sum_{i_1=1}^{2p-2k} \sum_{i_2=i_1+2}^{2p-2k+2} \cdots \sum_{i_k=i_{k-1}+2}^{2p-2} (-1)^k q_{i_1}^i \cdots q_{i_k}^i.$$

Also, for $1 \leq i \leq p$, let

$$\tilde{Q}_p^i = \begin{bmatrix} -\tilde{q}_1^i & \tilde{q}_1^i & 0 & \cdots & 0 & 0 & 0 \\ \tilde{q}_2^i & -\tilde{q}_2^i - \tilde{q}_3^i & \tilde{q}_3^i & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \tilde{q}_{2p-4}^i & -\tilde{q}_{2p-4}^i - \tilde{q}_{2p-3}^i & \tilde{q}_{2p-3}^i \\ 0 & 0 & 0 & \cdots & 0 & \tilde{q}_{2p-2}^i & -\tilde{q}_{2p-2}^i \end{bmatrix},$$

where, for $1 \leq k \leq 2p-2$,

$$\tilde{q}_k^i = \begin{cases} q_k, & \text{if } k \notin \{2i-2, 2i-1, 2i, 2i+1\}, \\ 0, & \text{if } k \in \{2i-2, 2i-1, 2i, 2i+1\}. \end{cases}$$

For $1 \leq i \leq p$, let $\tilde{R}_{0,i}^p = 1$. For $p \geq 2$ and $1 \leq i, k \leq p$, define

$$\tilde{R}_{k,i}^p = \sum_{i_1=1}^{2p-2k} \sum_{i_2=i_1+2}^{2p-2k+2} \cdots \sum_{i_k=i_{k-1}+2}^{2p-2} (-1)^k \tilde{q}_{i_1}^i \cdots \tilde{q}_{i_k}^i.$$

By equation (4) and the tridiagonality of the matrix, we only need to compute $\Sigma_{i,j}$ for $|i-j| \leq 1$, and so only $\partial R_k^p / \partial q_{i,j}$ for $|i-j| \leq 1$ are needed. Theorem 5 gives a closed form of the R_k^p's and the required derivatives.

THEOREM 5 *(a)* $R_p^p = 0$*, for* $p \geq 2$*.*

(b) For $p \geq 2$ *and* $1 \leq k \leq p-1$*,*

$$R_k^p = \sum_{i_1=1}^{2p-2k} \sum_{i_2=i_1+2}^{2p-2k+2} \cdots \sum_{i_k=i_{k-1}+2}^{2p-2} (-1)^k q_{i_1} \cdots q_{i_k}.$$

(c) For $p \geq 2$ *and* $1 \leq i \leq p$*,*

$$\frac{\partial R_1^p}{\partial q_{i,i}} = R_{0,i}^p = 1.$$

(d) For $p \geq 2$*,* $1 \leq i \leq p$ *and* $2 \leq k \leq p$*,*

$$\frac{\partial R_k^p}{\partial q_{i,i}} = R_{k-1,i}^p = \sum_{i_1=1}^{2p-2k+2} \sum_{i_2=i_1+2}^{2p-2k+4} \cdots \sum_{i_k=i_{k-1}+2}^{2p-2} (-1)^{k-1} q_{i_1}^i \cdots q_{i_{k-1}}^i.$$

(e) For $p \geq 2$*,*

$$\frac{\partial R_1^p}{\partial q_{i,i+1}} = \frac{\partial R_1^p}{\partial q_{i+1,i}} = 0.$$

(f) For $p \geq 2$*,*

$$\frac{\partial R_2^p}{\partial q_{i,i+1}} = -q_{i+1,i} = -q_{2i},$$

and

$$\frac{\partial R_2^p}{\partial q_{i+1,i}} = -q_{i,i+1} = -q_{2i-1}.$$

(g) For $p \geq 2$*,* $1 \leq i \leq p-1$ *and* $3 \leq k \leq p$*,*

$$\frac{\partial R_k^p}{\partial q_{i,i+1}} = -q_{2i} \tilde{R}_{k-2,i}^p = -q_{2i} \sum_{i_1=1}^{2p-2k+4} \sum_{i_2=i_1+2}^{2p-2k+6} \cdots \sum_{i_k=i_{k-1}+2}^{2p-2} (-1)^{k-2} \tilde{q}_{i_1}^i \cdots \tilde{q}_{i_{k-2}}^i,$$

and

$$\frac{\partial R_k^p}{\partial q_{i+1,i}} = -q_{2i-1} \tilde{R}_{k-2,i}^p = -q_{2i-1} \sum_{i_1=1}^{2p-2k+4} \sum_{i_2=i_1+2}^{2p-2k+6} \cdots \sum_{i_k=i_{k-1}+2}^{2p-2} (-1)^{k-2} \tilde{q}_{i_1}^i \cdots \tilde{q}_{i_{k-2}}^i.$$

Example: For clarification, we write out the parameter differentiation of the matrix exponential for an example from Chan and Munoz-Hernandez (1997). Let

$$Q = \begin{bmatrix} -q_1 & q_1 & 0 \\ q_2 & -q_2 - q_3 & q_3 \\ 0 & q_4 & -q_4 \end{bmatrix} = \begin{bmatrix} -e^{\theta_5 t+\theta_1} & e^{\theta_5 t+\theta_1} & 0 \\ e^{\theta_6 t+\theta_3} & -e^{\theta_6 t+\theta_3} - e^{\theta_5 t+\theta_2} & e^{\theta_5 t+\theta_2} \\ 0 & e^{\theta_6 t+\theta_4} & -e^{\theta_6 t+\theta_4} \end{bmatrix}.$$

Then

$$\begin{aligned}
\Sigma_{11} &= W^{-1}\Big[\{Q^2 + (q_2+q_3+q_4)Q + q_2 q_4 I\}te^{tQ} \\
&\quad -\{2Q + (q_1+q_2+q_3+q_4)I\}[B_{11}, e^{tQ}] - [B_{11}, e^{tQ}]Q\Big], \\
\Sigma_{22} &= W^{-1}\Big[\{Q^2 + (q_1+q_4)Q + q_1 q_4 I\}te^{tQ} \\
&\quad -\{2Q + (q_1+q_2+q_3+q_4)I\}[B_{22}, e^{tQ}] - [B_{22}, e^{tQ}]Q\Big], \\
\Sigma_{33} &= W^{-1}\Big[\{Q^2 + (q_1+q_2+q_3)Q + q_1 q_3 I\}te^{tQ} \\
&\quad -\{2Q + (q_1+q_2+q_3+q_4)[B_{33}, e^{tQ}] - [B_{33}, e^{tQ}]Q\Big], \\
\Sigma_{12} &= W^{-1}\Big[(q_2 Q + q_2 q_4 I)te^{tQ} \\
&\quad -\{2Q + (q_1+q_2+q_3+q_4)[B_{12}, e^{tQ}] - [B_{12}, e^{tQ}]Q\Big], \\
\Sigma_{21} &= W^{-1}\Big[(q_1 Q + q_1 q_4 I)te^{tQ} \\
&\quad -\{2Q + (q_1+q_2+q_3+q_4)I\}[B_{21}, e^{tQ}] - [B_{21}, e^{tQ}]Q\Big], \\
\Sigma_{23} &= W^{-1}\Big[(q_4 Q + q_1 q_4 I)te^{tQ} \\
&\quad -\{2Q + (q_1+q_2+q_3+q_4)I\}[B_{23}, e^{tQ}] - [B_{23}, e^{tQ}]Q\Big], \\
\Sigma_{32} &= W^{-1}\Big[(q_3 Q + q_1 q_3 I)te^{tQ} \\
&\quad -\{2Q + (q_1+q_2+q_3+q_4)I\}[B_{32}, e^{tQ}] - [B_{32}, e^{tQ}]Q\Big],
\end{aligned}$$

where

$$\begin{aligned}
W &= 3R_0^3 Q^2 - 2R_1^3 Q + R_2^3 I \\
&= 3Q^2 + 2(q_1+q_2+q_3+q_4)Q + (q_1 q_3 + q_1 q_4 + q_2 q_4)I,
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial e^{tQ}}{\partial \theta_1} &= -q_1(\Sigma_{11} - \Sigma_{12}), \\
\frac{\partial e^{tQ}}{\partial \theta_2} &= -q_3(\Sigma_{22} - \Sigma_{23}),
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial e^{tQ}}{\partial \theta_3} &= q_2(\Sigma_{21} - \Sigma_{22}), \\
\frac{\partial e^{tQ}}{\partial \theta_4} &= q_4(\Sigma_{32} - \Sigma_{33}), \\
\frac{\partial e^{tQ}}{\partial \theta_5} &= -tq_1\Sigma_{11} + tq_1\Sigma_{12} - tq_3\Sigma_{22} + tq_3\Sigma_{23} \\
&= t\left(\frac{\partial e^{tQ}}{\partial \theta_1} + \frac{\partial e^{tQ}}{\partial \theta_2}\right), \\
\frac{\partial e^{tQ}}{\partial \theta_6} &= tq_2\Sigma_{21} - tq_2\Sigma_{22} + tq_4\Sigma_{32} - tq_4\Sigma_{33} \\
&= t\left(\frac{\partial e^{tQ}}{\partial \theta_3} + \frac{\partial e^{tQ}}{\partial \theta_4}\right).
\end{aligned}
$$

Acknowledgement

The research of Tsai was partially supported by the National Science Council (NSC 90-2118-M-001-039), R.O.C.

APPENDIX

Proof of Theorem 1: By Cayley-Hamilton Theorem (see, e.g., page 86 of Horn and Johnson, 1985),

$$\sum_{k=0}^{p}(-1)^{p-k}R_{p-k}^{p}A^{k} = 0. \tag{10}$$

Taking partial derivative with respect to a_{ij} on both sides, we have

$$\sum_{k=0}^{p}(-1)^{p-k}\left(\frac{\partial R_{p-k}^{p}}{\partial a_{ij}}\right)A^{k} + \sum_{k=1}^{p}(-1)^{p-k}R_{p-k}^{p}\sum_{r=0}^{k-1}A^{k-r-1}\left(\frac{\partial A}{\partial a_{ij}}\right)A^{r} = 0, \tag{11}$$

which implies

$$\sum_{k=1}^{p}(-1)^{p-k}R_{p-k}^{p}\sum_{r=0}^{k-1}A^{k-r-1}B_{ij}A^{r} = \sum_{k=0}^{p-1}(-1)^{p-k+1}\left(\frac{\partial R_{p-k}^{p}}{\partial a_{ij}}\right)A^{k}. \tag{12}$$

Pre-multiplying $e^{(t-u)A}$ and post-multiplying e^{uA} to both sides of equation (12) and then integrating u from $u = 0$ to $u = t$ to get

$$
\begin{aligned}
&\int_0^t e^{(t-u)A}\left\{\sum_{k=1}^{p}(-1)^{p-k}R_{p-k}^{p}\sum_{r=0}^{k-1}A^{k-r-1}B_{ij}A^{r}\right\}e^{uA}du \\
= &\int_0^t e^{(t-u)A}\left\{\sum_{k=0}^{p-1}(-1)^{p-k+1}\left(\frac{\partial R_{p-k}^{p}}{\partial a_{ij}}\right)A^{k}\right\}e^{uA}du.
\end{aligned} \tag{13}
$$

Using the fact that (see, e.g., page 95 of Hale, 1969),

$$Ae^{tB} = e^{tB}A \text{ if and only if } AB = BA,$$

we have

$$\sum_{k=1}^{p}(-1)^{p-k}R^p_{p-k}\sum_{r=0}^{k-1}A^{k-r-1}\Sigma_{ij}A^r = \left\{\sum_{k=0}^{p-1}(-1)^{p-k+1}\left(\frac{\partial R^p_{p-k}}{\partial a_{ij}}\right)A^k\right\}te^{tA}. \tag{14}$$

Applying integration by parts to equation (3) to get

$$\Sigma_{ij}A = A\Sigma_{ij} + [B_{ij}, e^{tA}],$$

and then post-multiplying A to both sides to get

$$\begin{aligned}\Sigma_{ij}A^2 &= A\Sigma_{ij}A + [B_{ij}, e^{tA}]A \\ &= A(A\Sigma_{ij} + [B_{ij}, e^{tA}]) + [B_{ij}, e^{tA}]A \\ &= A^2\Sigma_{ij} + A[B_{ij}, e^{tA}] + [B_{ij}, e^{tA}]A.\end{aligned}$$

Applying the same technique recursively to get

$$\Sigma_{ij}A^r = A^r\Sigma_{ij} + \sum_{u=0}^{r-1}A^{r-u-1}[B_{ij}, e^{tA}]A^u, \text{ where } r \geq 1. \tag{15}$$

It follows from equation (15) that the left hand side of (14) can be rewritten as

$$\begin{aligned}& (-1)^{p-1}R^p_{p-1}\Sigma_{ij} \\ & + \sum_{k=2}^{p}(-1)^{p-k}R^p_{p-k}\left\{A^{k-1}\Sigma_{ij} + \sum_{r=1}^{k-1}A^{k-r-1}\left(A^r\Sigma_{ij} + \sum_{u=0}^{r-1}A^{r-u-1}[B_{ij}, e^{tA}]A^u\right)\right\} \\ = \;& (-1)^{p-1}R^p_{p-1}\Sigma_{ij} + \sum_{k=2}^{p}(-1)^{p-k}R^p_{p-k}A^{k-1}\Sigma_{ij} \\ & + \sum_{k=2}^{p}(-1)^{p-k}R^p_{p-k}\sum_{r=1}^{k-1}A^{k-1}\Sigma_{ij} + \sum_{k=2}^{p}(-1)^{p-k}R^p_{p-k}\sum_{r=1}^{k-1}\sum_{u=0}^{r-1}A^{k-u-2}[B_{ij}, e^{tA}]A^u \\ = \;& \left\{\sum_{k=1}^{p}(-1)^{p-k}kR^p_{p-k}A^{k-1}\right\}\Sigma_{ij} + \sum_{k=2}^{p}(-1)^{p-k}R^p_{p-k}\sum_{u=0}^{k-2}\sum_{r=u+1}^{k-1}A^{k-u-2}[B_{ij}, e^{tA}]A^u \\ = \;& \left\{\sum_{r=0}^{p-1}(-1)^{p-r-1}(r+1)R^p_{p-r-1}A^r\right\}\Sigma_{ij} \\ & + \sum_{k=2}^{p}(-1)^{p-k}R^p_{p-k}\sum_{u=0}^{k-2}(k-u-1)A^{k-u-2}[B_{ij}, e^{tA}]A^u \\ = \;& \left\{\sum_{r=0}^{p-1}(-1)^{p-r-1}(r+1)R^p_{p-r-1}A^r\right\}\Sigma_{ij} \\ & + \sum_{u=0}^{p-2}\sum_{k=u+2}^{p}(-1)^{p-k}(k-u-1)R^p_{p-k}A^{k-u-2}[B_{ij}, e^{tA}]A^u.\end{aligned}$$

This proves the result.

Proof of Theorem 2:

The proofs for (a), (b) and (d) are trivial.

(c) First note that, for $3 \leq k \leq p$,

$$
\begin{aligned}
R_k^p &= \sum_{1\leq l_1<\cdots<l_k\leq p} |A([l_1,\cdots,l_k])| \\
&= \sum_{\substack{1\leq l_1<\cdots<l_k\leq p \\ i,j\in\{l_1,\cdots,l_k\}}} |A([l_1,\cdots,l_k])| + \sum_{\substack{1\leq l_1<\cdots<l_k\leq p \\ i\notin\{l_1,\cdots,l_k\} \\ or\ j\notin\{l_1,\cdots,l_k\}}} |A([l_1,\cdots,l_k])| \\
&= -\sum_{\substack{1\leq l_1<\cdots<l_{k-2}\leq p \\ i\notin\{l_1,\cdots,l_{k-2}\} \\ and\ j\notin\{l_1,\cdots,l_{k-2}\}}} |A([i,j,l_1,\cdots,l_{k-2}],[j,i,l_1,\cdots,l_{k-2}])| \\
&\quad + \sum_{\substack{1\leq l_1<\cdots<l_k\leq p \\ i\notin\{l_1,\cdots,l_k\} \\ or\ j\notin\{l_1,\cdots,l_k\}}} |A([l_1,\cdots,l_k])|,
\end{aligned}
$$

which implies that, for $1 \leq i \neq j \leq p$,

$$
\frac{\partial R_k^p}{\partial a_{ij}} = -\sum_{\substack{1\leq l_1<\cdots<l_{k-2}\leq p \\ i\notin\{l_1,\cdots,l_{k-2}\} \\ and\ j\notin\{l_1,\cdots,l_{k-2}\}}} |A([j,l_1,\cdots,l_{k-2}],[i,l_1,\cdots,l_{k-2}])|.
$$

(e) For $2 \leq k \leq p$,

$$
\begin{aligned}
R_k^p &= \sum_{1\leq l_1<\cdots<l_k\leq p} |A([l_1,\cdots,l_k])| \\
&= \sum_{\substack{1\leq l_1<\cdots<l_k\leq p \\ i\in\{l_1,\cdots,l_k\}}} |A([l_1,\cdots,l_k])| + \sum_{\substack{1\leq l_1<\cdots<l_k\leq p \\ i\notin\{l_1,\cdots,l_k\}}} |A([l_1,\cdots,l_k])|,
\end{aligned}
$$

which implies that, for $1 \leq i \leq p$,

$$
\frac{\partial R_k^p}{\partial a_{ii}} = \sum_{\substack{1\leq l_1<\cdots<l_{k-1}\leq p \\ i\notin\{l_1,\cdots,l_{k-1}\}}} |A([l_1,\cdots,l_{k-1}])|
$$

Proof of Theorem 3:

The charactreistic polynomial can be written as

$$
q(\lambda) = |\lambda I - A| = \prod_{i=1}^{p}(\lambda - \lambda_i),
$$

where the $\lambda_i's$ are the eigenvalues of A. Now, the derivative of the characteristic polynomial equals $q'(\lambda) = \sum_{i=1}^{p}\prod_{j\neq i}(\lambda-\lambda_j)$. Hence, $q'(A) = \sum_{i=1}^{p}\prod_{j\neq i}(A-\lambda_j I)$. If v_k is an eigenvector of A corresponding to the eigenvalue λ_k, then

$$
q'(A)v_k = \sum_{i=1}^{p}\prod_{j\neq i}(\lambda_k - \lambda_j)v_k.
$$

In other words, the eigenvalues of $q'(A)$ are $\sum_{i=1}^{p}\prod_{j\neq i}(\lambda_k-\lambda_j)=\prod_{j\neq k}(\lambda_k-\lambda_j), k=1,2,..,p$, which are non-zero if and only if all the eigenvalues of A are distinct.

Proof of Theorem 4:

(a) The proof is trivial.

(b) The proof for $p=2$ is trivial. For $p\geq 3$, first note that,

$$R^p_{p-k-1} = \begin{cases} (-1)^{p-k}\alpha_{k+2}, & \text{for } -1\leq k\leq p-2, \\ 1, & \text{for } k=p-1. \end{cases} \tag{16}$$

$$\begin{aligned} \frac{\partial R^p_{p-k}}{\partial a_{p1}} &= \begin{cases} \frac{\partial}{\partial \alpha_1}\{(-1)^{p-k+1}\alpha_{k+1}\}, & \text{for } 0\leq k\leq p-1, \\ 0, & \text{for } k=p, \end{cases} \\ &= \begin{cases} (-1)^{p+1}, & \text{for } k=0, \\ 0, & \text{for } 1\leq k\leq p, \end{cases} \end{aligned} \tag{17}$$

and

$$[B_{p1},e^{tA}]A^u = [B_{p1}A^u,e^{tA}] = [B_{p(u+1)},e^{tA}], \text{ for } 0\leq u\leq p-1. \tag{18}$$

Thus,

$$\begin{aligned} \frac{\partial e^{tA}}{\partial \alpha_1} &= \int_0^t e^{(t-u)A}\left(\frac{\partial A}{\partial \alpha_1}\right)e^{uA}du \qquad \text{(by equation (1))} \\ &= \int_0^t e^{(t-u)A}B_{p1}e^{uA}du \\ &= \Sigma_{p1} \qquad \text{(by equation (3))} \\ &= \left\{pA^{p-1}+\sum_{k=0}^{p-2}(-1)^{p-k-1}(k+1)(-1)^{p-k}\alpha_{k+2}A^k\right\}^{-1}\Big[(-1)^{p+1}(-1)^{p+1}te^{tA} \\ &\quad -\sum_{u=0}^{p-3}\sum_{k=u+2}^{p-1}(-1)^{p-k}(-1)^{p-k+1}\alpha_{k+1}(k-u-1)A^{k-u-2}[B_{p1},e^{tA}]A^u \\ &\quad -\sum_{u=0}^{p-2}(-1)^{p-p}(p-u-1)A^{p-u-2}[B_{p1},e^{tA}]A^u\Big] \\ &\qquad \text{(by Theorem 1, equations (16) and (17))} \\ &= \left\{pA^{p-1}-\sum_{k=0}^{p-2}(k+1)\alpha_{k+2}A^k\right\}^{-1} \\ &\quad \Big\{te^{tA}+\sum_{u=0}^{p-3}\sum_{k=u+2}^{p-1}(k-u-1)\alpha_{k+1}A^{k-u-2}[B_{p(u+1)},e^{tA}] \\ &\quad -\sum_{u=0}^{p-2}(p-u-1)A^{p-u-2}[B_{p(u+1)},e^{tA}]\Big\} \qquad \text{(by equation (18))} \end{aligned}$$

$$
\begin{aligned}
&= K_{p,0}^{-1}\left\{te^{tA}+\sum_{r=1}^{p-2}\sum_{k=r+1}^{p-1}(k-r)\alpha_{k+1}A^{k-r-1}[B_{pr},e^{tA}]-\sum_{r=1}^{p-1}(p-r)A^{p-r-1}[B_{pr},e^{tA}]\right\} \\
&\qquad\qquad\qquad(\text{ by letting } r=u+1) \\
&= K_{p,0}^{-1}\left\{te^{tA}+\sum_{r=1}^{p-2}\sum_{v=r+2}^{p}(v-r-1)\alpha_{v}A^{v-r-2}[B_{pr},e^{tA}]-\sum_{r=1}^{p-1}(p-r)A^{p-r-1}[B_{pr},e^{tA}]\right\} \\
&\qquad\qquad(\text{ by letting } v=k+1) \\
&= K_{p,0}^{-1}\Bigg[te^{tA}-\sum_{r=1}^{p-2}\left\{(p-r)A^{p-r-1}-\sum_{k=r+2}^{p}(k-r-1)\alpha_k A^{k-r-2}\right\}[B_{pr},e^{tA}] \\
&\quad -[B_{p(p-1)},e^{tA}]\Bigg] \\
&= K_{p,0}^{-1}\left\{te^{tA}-\sum_{r=1}^{p-1}K_{p,r}[B_{pr},e^{tA}]\right\}.
\end{aligned}
$$

This proves the result.

(c) For $2 \le i \le p$,

$$
\begin{aligned}
\frac{\partial e^{tA}}{\partial \alpha_i} &= \int_0^t e^{(t-u)A}\delta_p\delta_i' e^{uA}du \\
&= \int_0^t e^{(t-u)A}\delta_p\delta_{i-1}' Ae^{uA}du \qquad (\text{ because } \delta_{i-1}'A=\delta_i') \\
&= \left(\frac{\partial e^{tA}}{\partial \alpha_{i-1}}\right)A \qquad (\text{ because } Ae^{uA}=e^{uA}A).
\end{aligned}
$$

Proof of Theorem 5:

The proofs for (c), (e) and (f) are trivial.

(a) & (b) Let $q_0 = q_{2p-1} = 0$, then the transition intensity matrix can be rewritten as

$$
Q_p=[q_{i,j}]=\begin{bmatrix}
-q_0-q_1 & q_1 & 0 & \cdots & 0 & 0 & 0 \\
q_2 & -q_2-q_3 & q_3 & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & q_{2p-4} & -q_{2p-4}-q_{2p-3} & q_{2p-3} \\
0 & 0 & 0 & \cdots & 0 & q_{2p-2} & -q_{2p-2}-q_{2p-1}
\end{bmatrix},
$$

and

$$
q_{i,j}=\begin{cases}
-q_{2i-2}-q_{2i-1}, & \text{if } 1\le i=j\le p, \\
q_{2i-1}, & \text{if } 1\le j=i+1\le p, \\
q_{2i-2}, & \text{if } 1\le j=i-1\le p-1, \\
0, & \text{otherwise.}
\end{cases}
$$

We will prove later that, for $p \geq 2$, and $1 \leq k \leq p$,

$$R_k^p = \sum_{i_1=0}^{2p-2k+1} \sum_{i_2=i_1+2}^{2p-2k+3} \cdots \sum_{i_k=i_{k-1}+2}^{2p-1} (-1)^k q_{i_1} \cdots q_{i_k}. \tag{19}$$

But because $q_0 = q_{2p-1} = 0$, we have, by equation (19),

$$R_p^p = \sum_{i_1=0}^{1} \sum_{i_2=i_1+2}^{3} \cdots \sum_{i_p=i_{p-1}+2}^{2p-1} (-1)^p q_{i_1} \cdots q_{i_p} = 0,$$

and, for $1 \leq k \leq p-1$,

$$R_k^p = \sum_{i_1=1}^{2p-2k} \sum_{i_2=i_1+2}^{2p-2k+2} \cdots \sum_{i_k=i_{k-1}+2}^{2p-2} (-1)^k q_{i_1} \cdots q_{i_k}.$$

This proves parts (a) and (b) of the theorem.

For equation (19), we will prove it by mathematical induction. First, it is easily seen that equation (19) holds for $p = 2$ and $1 \leq k \leq 2$. Now, suppose equation (19) holds for R_k^r, where $2 \leq r \leq p-1$ and $1 \leq k \leq r$, we want to show that, for $2 \leq k \leq p$, R_k^p is given by the right hand side of equation (19) (the proof for $k = 1$ is trivial).

Note that, for the tridiagonal matrix Q_p, we have

$$|\lambda I - Q_p| = (\lambda - q_{p,p})|\lambda I - Q_{p-1}| - q_{p,p-1}q_{p-1,p}|\lambda I - Q_{p-2}|,$$

which implies that

$$\begin{aligned} \sum_{k=0}^{p}(-1)^{p-k}R_{p-k}^p\lambda^k &= (\lambda + q_{2p-2} + q_{2p-1})\sum_{k=0}^{p-1}(-1)^{p-k-1}R_{p-k-1}^{p-1}\lambda^k \\ &\quad -q_{2p-2}q_{2p-3}\sum_{k=0}^{p-2}(-1)^{p-k-2}R_{p-k-2}^{p-2}\lambda^k. \end{aligned}$$

Comparing the coefficients of λ^k on both sides, we have, for $k = 1, ..., p-2$,

$$\begin{aligned} (-1)^{p-k}R_{p-k}^p &= (-1)^{p-k}R_{p-k}^{p-1} + (-1)^{p-k-1}(q_{2p-2} + q_{2p-1})R_{p-k-1}^{p-1} \\ &\quad -(-1)^{p-k-2}q_{2p-2}q_{2p-3}R_{p-k-2}^{p-2}. \end{aligned}$$

Equivalently, we have, for $k = 2, ..., p-1$,

$$\begin{aligned} (-1)^k R_k^p &= (-1)^k R_k^{p-1} + (-1)^{k-1}(q_{2p-2} + q_{2p-1})R_{k-1}^{p-1} \\ &\quad -(-1)^{k-2}q_{2p-2}q_{2p-3}R_{k-2}^{p-2}. \end{aligned} \tag{20}$$

But, for $k = 3, \cdots, p-1$,

$$
\begin{aligned}
R_k^{p-1} &= \sum_{i_1=0}^{2p-2k-1} \sum_{i_2=i_1+2}^{2p-2k+1} \cdots \sum_{i_k=i_{k-1}+2}^{2p-3} (-1)^k q_{i_1} \cdots q_{i_k}, \\
R_{k-1}^{p-1} &= \sum_{i_1=0}^{2p-2k+1} \sum_{i_2=i_1+2}^{2p-2k+3} \cdots \sum_{i_{k-1}=i_{k-2}+2}^{2p-3} (-1)^{k-1} q_{i_1} \cdots q_{i_{k-1}}, \\
R_{k-2}^{p-2} &= \sum_{i_1=0}^{2p-2k+1} \sum_{i_2=i_1+2}^{2p-2k+3} \cdots \sum_{i_{k-2}=i_{k-3}+2}^{2p-5} (-1)^{k-2} q_{i_1} \cdots q_{i_{k-2}},
\end{aligned}
$$

and so, for $k = 3, \cdots, p-1$, equation (20) becomes

$$
\begin{aligned}
(-1)^k R_k^p &= \sum_{i_1=0}^{2p-2k-1} \sum_{i_2=i_1+2}^{2p-2k+1} \cdots \sum_{i_{k-1}=i_{k-2}+2}^{2p-5} \sum_{i_k=i_{k-1}+2}^{2p-3} q_{i_1} \cdots q_{i_k} \\
&\quad + \sum_{i_1=0}^{2p-2k} \sum_{i_2=i_1+2}^{2p-2k+2} \cdots \sum_{i_{k-2}=i_{k-3}+2}^{2p-6} \sum_{i_{k-1}=i_{k-2}+2}^{2p-4} \sum_{i_k=2p-2}^{2p-1} q_{i_1} \cdots q_{i_k} \\
&\quad + \sum_{i_1=0}^{2p-2k+1} \sum_{i_2=i_1+2}^{2p-2k+3} \cdots \sum_{i_{k-2}=i_{k-3}+2}^{2p-5} \sum_{i_{k-1}=2p-3} \sum_{i_k=2p-2} q_{i_1} \cdots q_{i_k} \\
&\quad + \sum_{i_1=0}^{2p-2k+1} \sum_{i_2=i_1+2}^{2p-2k+3} \cdots \sum_{i_{k-2}=i_{k-3}+2}^{2p-5} \sum_{i_{k-1}=2p-3} \sum_{i_k=2p-1} q_{i_1} \cdots q_{i_k} \\
&\quad - \sum_{i_1=0}^{2p-2k+1} \sum_{i_2=i_1+2}^{2p-2k+3} \cdots \sum_{i_{k-2}=i_{k-3}+2}^{2p-5} \sum_{i_{k-1}=2p-3} \sum_{i_k=2p-2} q_{i_1} \cdots q_{i_k} \\
&= \sum_{i_1=0}^{2p-2k-1} \sum_{i_2=i_1+2}^{2p-2k+1} \cdots \sum_{i_{k-1}=i_{k-2}+2}^{2p-5} \sum_{i_k=i_{k-1}+2}^{2p-3} q_{i_1} \cdots q_{i_k} \\
&\quad + \sum_{i_1=0}^{2p-2k} \sum_{i_2=i_1+2}^{2p-2k+2} \cdots \sum_{i_{k-2}=i_{k-3}+2}^{2p-6} \sum_{i_{k-1}=i_{k-2}+2}^{2p-4} \sum_{i_k=2p-2}^{2p-1} q_{i_1} \cdots q_{i_k} \\
&\quad + \sum_{i_1=0}^{2p-2k+1} \sum_{i_2=i_1+2}^{2p-2k+3} \cdots \sum_{i_{k-2}=i_{k-3}+2}^{2p-5} \sum_{i_{k-1}=2p-3} \sum_{i_k=2p-1} q_{i_1} \cdots q_{i_k} \\
&= \sum_{i_1=0}^{2p-2k+1} \sum_{i_2=i_1+2}^{2p-2k+3} \cdots \sum_{i_k=i_{k-1}+2}^{2p-1} q_{i_1} \cdots q_{i_k}.
\end{aligned}
$$

The proofs for $k = 2$ and p are similar to that of $3 \le k \le p-1$. This proves equation (19).

(d) First note that, for $1 \le l_1 < \cdots < l_{k-1} \le p$ and $i \notin \{l_1, \cdots, l_{k-1}\}$, $|Q_p([l_1, \cdots, l_{k-1}])| = |Q_p^i([l_1, \cdots, l_{k-1}])|$. For $1 \le l_1 < \cdots < l_{k-1} \le p$ and $i \in \{l_1, \cdots, l_{k-1}\}$, $|Q_p^i([l_1, \cdots, l_{k-1}])| = 0$, beause the matrix $Q_p^i([l_1, \cdots, l_{k-1}])$ contains a zero row vector. Thus, by Theorem 2 (e), we have that, for $2 \le k \le p$ and

$1 \le i \le p$,

$$\begin{aligned}
\frac{\partial R_k^p}{\partial q_{i,i}} &= \sum_{\substack{1\le l_1<\cdots<l_{k-1}\le p\\ i\notin\{l_1,\cdots,l_{k-1}\}}} |Q_p([l_1,\cdots,l_{k-1}])| \\
&= \sum_{\substack{1\le l_1<\cdots<l_{k-1}\le p\\ i\notin\{l_1,\cdots,l_{k-1}\}}} |Q_p^i([l_1,\cdots,l_{k-1}])| + \sum_{\substack{1\le l_1<\cdots<l_{k-1}\le p\\ i\in\{l_1,\cdots,l_{k-1}\}}} |Q_p^i([l_1,\cdots,l_{k-1}])| \\
&= \sum_{1\le l_1<\cdots<l_{k-1}\le p} |Q_p^i([l_1,\cdots,l_{k-1}])| \\
&= R_{k-1,i}^p,
\end{aligned}$$

where the last equality follows from equation (5), Theorem 5 (b) and the definition of $R_{k,i}^p$.

(g) By Theorem 2 (c), we have that, for $3 \le k \le p$ and $1 \le i \le p-1$,

$$\begin{aligned}
\frac{\partial R_k^p}{\partial q_{i,i+1}} &= -\sum_{\substack{1\le l_1<\cdots<l_{k-2}\le p\\ i\notin\{l_1,\cdots,l_{k-2}\}\\ and\ i+1\notin\{l_1,\cdots,l_{k-2}\}}} |Q_p([i+1,l_1,\cdots,l_{k-2}],[i,l_1,\cdots,l_{k-2}])| \\
&= -q_{i+1,i} \sum_{\substack{1\le l_1<\cdots<l_{k-2}\le p\\ i\notin\{l_1,\cdots,l_{k-2}\}\\ and\ i+1\notin\{l_1,\cdots,l_{k-2}\}}} |Q_p([l_1,\cdots,l_{k-2}],[l_1,\cdots,l_{k-2}])| \\
&= -q_{i+1,i} \sum_{\substack{1\le l_1<\cdots<l_{k-2}\le p\\ i\notin\{l_1,\cdots,l_{k-2}\}\\ and\ i+1\notin\{l_1,\cdots,l_{k-2}\}}} |\tilde{Q}_p^i([l_1,\cdots,l_{k-2}],[l_1,\cdots,l_{k-2}])| \\
&\quad -q_{i+1,i} \sum_{\substack{1\le l_1<\cdots<l_{k-2}\le p\\ i\in\{l_1,\cdots,l_{k-2}\}\\ or\ i+1\in\{l_1,\cdots,l_{k-2}\}}} |\tilde{Q}_p^i([l_1,\cdots,l_{k-2}],[l_1,\cdots,l_{k-2}])| \\
&= -q_{i+1,i} \sum_{1\le l_1<\cdots<l_{k-2}\le p} |\tilde{Q}_p^i([l_1,\cdots,l_{k-2}],[l_1,\cdots,l_{k-2}])| \\
&= -q_{2i}\tilde{R}_{k-2,i}^p.
\end{aligned}$$

The proof for $\partial R_k^p/\partial q_{i+1,i}$ is similar.

References

[1] Arnold, L. (1974) *Stochastic Differential Equations*. New York: Wiley.

[2] Brockwell, P. J. (1993) threshold ARMA processes in continuous time, 170-90 in *Dimension Estimation and Models*, ed. by H. Tong, World Scientific Publishing, River Edge, NJ.

[3] Brockwell, P. J. and Stramer, O. (1995) On the approximation of continuous time threshold ARMA processes. *Annals of the Institute of Statistical Mathematics* **47**, no. 1, 1–20.

[4] Chan, K.S. and Munoz-Hernandez, B. (1997) A generalized linear model for repeated ordered categorical response data. Technical Report 260, Department of Statistics & Actuarial Science, The University of Iowa.

[5] Chen, B. and Zadrozny, P. A. (2001) Analytic derivatives of the matrix exponential for estimation of linear continuous-time models. *Journal of Economic Dynamics and Control* **25**, 1867-79.

[6] Cox, D. R., and Miller, H. D. (1965) *The Theory of Stochastic Processes.* Londo: Methuen.

[7] Graham, A. (1986) *Kronecker Products and Matrix Calculus with Applications.* Prentice Hall.

[8] Hale, J. (1969) *Ordinary Differential Equations.* New York, Wiley-Interscience.

[9] Horn, R. A. & Johnson, C. R. (1985) *Matrix Analysis.* Cambridge: Cambridge University Press.

[10] Horn, R. A. & Johnson, C. R. (1991) *Topics in Matrix Analysis.* Cambridge: Cambridge University Press.

[11] Jazwinski, A. H. (1970) *Stochastic Processes and Filtering Theory.* New York: Academic Press.

[12] Jennrich, R. I. & Bright, P. B. (1976) Fitting systems of linear differential equations using computer generated exact derivatives. *Technometrics* **18**, 385-92.

[13] Kalbfleisch, J. D. & Lawless, J. F. (1985) The analysis of panel data under a Markov assumption. *JASA* **80**, 863-71.

[14] Singer, H. (1995) Analytical Score Function for Irregularly Sampled Continuous Time Stochastic Processes with Control Variables and Missing Values. *Econometric Theory* **11**, 721-735.

[15] Van Loan, C.F. (1978) Computing integrals involving the matrix exponential. *IEEE Transactions on Automatic Control* **23**, 395-404.

[16] Wilcox, R. M. (1967) Exponential Operators and Parameter Differentiation in Quantum Physics. *Journal of Mathematical Physics* **8**, no. 4, 962-82.