Technical Note—On Matrix Exponential Differentiation with Application to Weighted Sum Distributions

Milan Kumar Das, Henghsiu Tsai
Institute of Statistical Science, Academia Sinica, Taiwan (ROC) milan@stat.sinica.edu.tw, htsai@stat.sinica.edu.tw

Ioannis Kyriakou
Faculty of Actuarial Science & Insurance, Bayes Business School (formerly Cass), City, University of London, 106 Bunhill Row, London EC1Y 8TZ, UK ioannis.kyriakou@city.ac.uk

Gianluca Fusai
Dipartimento di Studi per l'Economia e l'Impresa, Università del Piemonte Orientale, Via Perrone 18, 28100 Novara, Italy & Faculty of Finance, Bayes Business School (formerly Cass), City, University of London, London EC1Y 8TZ, UK
gianluca.fusai@uniupo.it & gianluca.fusai.1@city.ac.uk

In this note, we revisit the innovative transform approach introduced by Cai et al. [Cai, N., Song, Y., Kou, S., 2015. A general framework for pricing Asian options under Markov processes] for accurately approximating the probability distribution of a weighted stochastic sum or time integral under general one-dimensional Markov processes. Since then, Song et al. [Song, Y., Cai, N., Kou, S., 2018. Computable error bounds of Laplace inversion for pricing Asian options] and Cui et al. [Cui, Z., Lee, C., Liu, Y., 2018. Single-transform formulas for pricing Asian options in a general approximation framework under Markov processes] have achieved an efficient reduction of the original double to a single transform approach. We move one step further by approaching the problem from a new angle and, by dealing with the main obstacle relating to the differentiation of the exponential of a matrix, we bypass the transform inversion. We highlight the benefit from the new result by means of some numerical examples.

Key words: Stochastic sum; probability distribution; matrix exponential and column vector differentiation; Pearson curve fit; pricing

1. Introduction

Continuous-time Markov chain (CTMC) approximations have gained popularity in the recent years in operations research, finance and medicine due to their ability to deliver efficient and accurate solutions to various problems. In finance, there has been a great research interest in applications to derivatives pricing, including, for example, Cai et al. (2015), Cui et al. (2018), Cui et al. (2017) and Kirkby et al. (2017). Most lately, Cui et al. (2021) proposed a novel Monte Carlo simulation method for stochastic differential equation systems based on CTMC with applications to stochastic local volatility models and queue processes. Pointing up the importance of regime-switching models in areas such as healthcare and financial engineering, Cai et al. (2020) also proposed an extended CTMC approximation to general regime-switching Markov models and presented relevant uses.

In this paper, we focus the spotlight on a key matrix function that appears in several CTMC applications, that is, a matrix exponential, which emerges, for example, in distributions of first passage times, running extrema and stochastic time integrals, in bond prices and generally option price formulations as well as their sensitivities (see Cai et al. 2020 and Ding et al. 2021). Here, we give prominence to a practically useful quantity that features in various applications, that of a stochastic time integral. Discrete or continuous additive functionals appear in numerous research problems in finance (net present value modelling, e.g., see Creemers 2018; average-based derivatives, e.g., see Fusai and Kyriakou 2016, Gambaro et al. 2020; stochastic volatility modelling, e.g., Cui et al. 2021, Kyriakou et al. 2021), insurance (Brignone et al. 2021), technology (see Nadarajah 2008), biomedical engineering (Baumann et al. 2019), and others. These problems become intricate in the lack of knowledge of the distribution of the sum.

Cai et al. (2015) pioneered a method for obtaining the unknown probability distribution of the stochastic sum (discrete and continuous) in general one-dimensional Markov models via an approximating CTMC based on the technique in Mijatović and Pistorius (2013). (This was extended later to general regime-switching Markov models in Cai et al. 2020.). As part of their application, they focused on the prices of Asian options which they recovered by numerical inversion of the double Laplace transform related to the constructed CTMC with respect to the strike price and the maturity (or number of monitoring dates in the case of discrete averaging). Based on the same principles, Cui et al. (2018) simplified to a single Laplace transform with respect to the strike, with consequent significant complexity and computational cost reductions. Song et al. (2018) derived computable bounds for the error in the Laplace transform inversion guaranteeing its accuracy. In doing this, they also obtained the closed-form single Laplace transforms which were derived independently by Cui et al. (2018).

In this paper, we revisit the underlying bedrock of the aforementioned contributions, that is, the Laplace transform of the random sum given by a column vector derived from a matrix exponential of the form $e^{A(\theta)}\mathbf{x}$, where $A(\theta)$ is an affine matrix-valued function in θ and $\mathbf{x}$ is a column vector. Our method requires access to the integer moments of the sum, hence requires differentiation of the exponential map; this is a notoriously difficult mathematical problem that has preoccupied many researchers. For example, in their early contribution, Tsai and Chan (2003) derived, under the assumption of a matrix with distinct eigenvalues, a closed-form solution for the first order parameter differentiation of the matrix exponential in terms of minors, polynomials, the exponential of the matrix and a matrix inversion. Although an algebraically manageable solution, it is undeniably complicated and particularly challenging, especially when considering adapting to matrices with repeated eigenvalues and computation of higher order derivatives. Separate contributions by Cai and Yang (2018, 2021) focused on techniques for the derivative of a column vector aiming to obtain several useful deterministic expressions related to the first passage time of reflected jump diffusion processes as elegant matrix functions. In this paper, we approach the problem differently showing that the *exact* derivatives of the matrix exponential satisfy a system of ordinary differential equations and derive closed-form solutions for exponential diagonalizable matrices. We also suggest, first, a possible extension beyond the diagonalizable case and, second, we propose an efficient technique for the direct differentiation of the column vector $e^{A(\theta)}\mathbf{x}$, reducing the computational cost of differentiating the full exponential matrix. Although we focus on the exponential of an affine matrix function, our method is easily adaptable to any general matrix via a straightforward modification of the parent recurrence relation (see later Proposition 3) of our approach. This way we are able to generalize to solving various problems that involve derivatives of a matrix exponential, such as the statistical inference of continuous-time auto-regressive moving average (CARMA) models (see Tsai and Chan 2003 and references therein), or the log-likelihood function maximization using a quasi-Newton method in a panel data analysis under the CTMC assumption (see Kalbfleisch and Lawless 1985).

In this paper, we concentrate on curve-fitting algorithms in moment-determinate problems based on moments that we derive using our technique; having at hand high order integer moments removes the major block to our application enabling us to obtain a *bona fide* moment-based distribution approximation based on a Pearson curve fit. Again, several possible applications may originate from this. For example, Cui et al. (2021) describe the CTMC construction for the diffusion limits of a M/M/s or a GI/M/s queue. This paves the way for obtaining the required higher order moments via our proposed technique for uses such as in the estimation and prediction of tail behaviour (e.g., see Choudhury and Lucantoni 1996, Abate et al. 1995). The problem of bounding tail probabilities under moment constraints is still of interest today and considered in several recent works, such as Chen et al. (2021) and Tian et al. (2017), following early contributions by Bertsimas and Popescu (2005), given the first three moments, and the fourth-moment approach of He et al. (2010). In addition, moment problems in finance are studied in Bertsimas and Sethuraman (2000) (see also Bertsimas and Popescu 2002 and Lo 1987), such as the formulation of optimal bounds on the price of an option given distributional moment information. Besides, being able to derive the moments of occupation times or Parisian stopping times can lead via curve-fitting to further uses in pricing step or Parisian options (see Yang et al. 2021, Zhang and Li 2021); also a distribution-fitting procedure based on moments of integral functionals of variance processes obtained from the Laplace transforms of the corresponding integrated CTMC processes (see Cui et al. 2021) can be used to facilitate the simulation of various volatility models via inverse transform sampling.

The remainder of this paper is organized as follows. In Section 2, we describe the model framework. Section 3 is devoted to our theoretical results. Section 4 presents our application, computational complexity analysis, error analysis and numerical examples for different models which illustrate the speed and accuracy of our approach. Section 5 concludes the paper. Several proofs and algorithms are deferred to the e-companion.

2. The Model

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space satisfying the usual conditions and supporting the process $S := \{S_t\}_{t\geq 0}$. Consider the sums $A_c(t) := \int_0^t S_w \, dw$ and $A_d(m) := \sum_{i=0}^m S_{t_i}$ where the latter is based on $m+1$ future recordings of S at the equidistant times $t_0 = 0, t_1 = \Delta, \ldots, t_m = m\Delta$.

As in Cai et al. (2015) and Cui et al. (2018), S is represented by a non-negative CTMC process with finite state space $\mathcal{X} := \{x_1, x_2, \ldots, x_p\}$ which is constructed via the technique in Mijatović and Pistorius (2013); we define as D a $p \times p$ diagonal matrix whose entries are the elements of $\mathcal{X}$. In addition, let $P(t)$ and G be, respectively, the $p \times p$ transition probability matrix and $p \times p$ transition rate matrix of S. Then, it is known from the aforementioned contributions that, for any complex number θ with positive real part,

$$\mathbb{E}_{\mathbb{P}}\left[e^{-\theta A_d(m)}\right] = E_d(m,\theta)\mathbf{1}, \text{ where } E_d(m,\theta) := \left(e^{-\theta D} P(\Delta)\right)^m e^{-\theta D}, \tag{1}$$

$$\mathbb{E}_{\mathbb{P}}\left[e^{-\theta A_c(t)}\right] = E_c(t,\theta)\mathbf{1}, \text{ where } E_c(t,\theta) := e^{(G-\theta D)t}, \tag{2}$$

and $\mathbf{1}$ is the $p \times 1$ column vector with all the entries equal to 1. The readers can refer to the original paper of Cai et al. (2015) for the details of the construction and evaluation of (1) and (2).

2.1. Pearson Curve Fit

Our ultimate goal in this paper is to construct the unknown probability distribution of A_d and A_c using an efficient distribution fit.

The Pearson system is a family of solutions $g(z)$ to the differential equation

$$\frac{1}{g(z)} \frac{dg(z)}{dz} = -\frac{\beta_0 + z}{\beta_1 + \beta_2 z + \beta_3 z^2}$$

whereby well-defined density functions can be derived with general form

$$g(x) = \mathcal{C}\left(\beta_1 + \beta_2 x + \beta_3 x^2\right)^{-\frac{1}{2\beta_3}} \exp\left\{ \frac{(\beta_2 - 2\beta_0\beta_3) \arctan\left(\frac{\beta_2 + 2\beta_3 x}{\sqrt{4\beta_1\beta_3 - \beta_2^2}}\right)}{\beta_3\sqrt{4\beta_1\beta_3 - \beta_2^2}} \right\},$$

where $\mathcal{C}$ is the normalizing constant and $\{\beta_0, \beta_1, \beta_2, \beta_3\}$ the parameters that control the shape of the distribution. These are estimated in the distribution fitting using the first four finite integer moments $\mu_n := \mathbb{E}_{\mathbb{P}}[A_\cdot^n(\cdot)]$, $n = 1, 2, 3, 4$, and are given by

$$\beta_0 = \beta_2 := \frac{\sqrt{\alpha\gamma}\,(\varepsilon+3)}{10\varepsilon - 12\gamma - 18}, \quad \beta_1 := \frac{(4\varepsilon - 3\gamma)\,\alpha}{10\varepsilon - 12\gamma - 18}, \quad \beta_3 := \frac{2\varepsilon - 3\gamma - 6}{10\varepsilon - 12\gamma - 18},$$

where

$$\alpha := \mu_2 - \mu_1^2, \quad \gamma := \frac{\left(\mu_3 - 3\mu_1\mu_2 + 2\mu_1^3\right)^2}{\left(\mu_2 - \mu_1^2\right)^3}, \quad \varepsilon := \frac{\mu_4 - 4\mu_1\mu_3 + 6\mu_1^2\mu_2 - 3\mu_1^4}{\left(\mu_2 - \mu_1^2\right)^2}$$

are, respectively, the variance, squared skewness and kurtosis of the Pearson random variable. The required moments follow from the next proposition.

PROPOSITION 1. *The n-th raw moment of the random variable $A_d(m)$ and $A_c(t)$ given $S_0 = x_i \in \mathcal{X}$ follows, respectively, from $(-1)^n \frac{d^n}{d\theta^n} \left\{ (\mathbf{e}_i^p)^* E_d(m,\theta)\mathbf{1} \right\} \big|_{\theta=0}$ and $(-1)^n \frac{d^n}{d\theta^n} \left\{ (\mathbf{e}_i^p)^* E_c(m,\theta)\mathbf{1} \right\} \big|_{\theta=0}$, where $E_d(m,\theta)$ and $E_c(t,\theta)$ are as in (1)–(2), $\mathbf{e}_i^p$ is, for any positive integer p, the $p \times 1$ column vector with the i-th entry only non-zero and equal to 1, and * denotes the transpose operation.*

Proof. It is obvious from (1) that $\mathbb{E}_{\mathbb{P}} \left[e^{-\theta A_d(m)} \middle| S_0 = x_i \right] = (\mathbf{e}_i^p)^* E_d(m,\theta)\mathbf{1}$. Thus, the n-th raw moment is given by $(-1)^n \frac{d^n}{d\theta^n} \left\{ (\mathbf{e}_i^p)^* E_d(m,\theta)\mathbf{1} \right\} \big|_{\theta=0}$. The same argument holds for $A_c(t)$. $\square$

Given knowledge of the first four moments, we can construct a density function g that is consistent with these and can be used to evaluate quantities of interest, as shown, for example, later in Section 4. Our preference towards the system of Pearson curves is driven by the simplicity, fast family selection and parameter estimation, ability to adapt to varying levels of skewness and kurtosis, and accuracy based on a first four-moment fit. Its accuracy has been verified in researches such as Solomon and Stephens (1978) and, more recently, Kyriakou et al. (2021). We have considered and applied alternatives to the Pearson system, such as a Gram–Charlier or a Cornish–Fisher series expansion or Johnson systems, but have excluded them because of encountered cases of non-convergence with increasing number of moments, or a non-guaranteed well-defined density, or because they have just been slower. Pearson is highly performant as we demonstrate in our numerical application in Section 4.

The following sections are devoted to the derivation of closed-form expressions for the derivatives of $E_d(m,\theta)$, $E_c(t,\theta)$ and $E_c(t,\theta)\mathbf{1}$.

3. Derivatives of Matrix Exponential

Throughout the paper, $M_p(\mathbb{C})$ will denote the space of all $p \times p$ matrices over the complex field $\mathbb{C}$.

3.1. The Discrete Case

Let $E_d(m,\theta)$ be given by (1) and $E_d^{(n)}(m,\theta)$ the n-th derivative of $E_d(m,\theta)$ with respect to θ for any non-negative integer n; in addition, $E_d^{(0)}(m,\theta) \equiv E_d(m,\theta)$. We start by establishing a key differential-difference relation in the following proposition.

PROPOSITION 2. *For any two non-negative integers m, n, we have that*

$$E_d^{(n)}(m+1,\theta) = \sum_{k=0}^{n} \binom{n}{k} (-1)^{n-k} D^{n-k} e^{-\theta D} P(\Delta) E_d^{(k)}(m,\theta),$$

where $E_d^{(n)}(0,\theta) = (-1)^n D^n e^{-\theta D}$ *and* $\binom{n}{k}$ *denotes the binomial coefficient.*

Proof. See e-companion section EC.1. $\square$

3.2. The Continuous Case

In what follows, we are interested in studying the computation of the higher order derivatives of the exponential map $E_c(t,\theta) = e^{t(G-\theta D)}$ with respect to θ, where G and D are elements on $M_p(\mathbb{C})$, and $t \geq 0$. In the next proposition, we establish an important recurrence relation between the $(n+1)$-th and n-th order derivatives of $E_c(t,\theta)$. We use $E_c^{(0)}(t,\theta)$ to mean $E_c(t,\theta)$.

PROPOSITION 3. *For any non-negative integer n and any real $t \geq 0$,*

$$E_c^{(n+1)}(t,\theta) = -(n+1)\int_0^t E_c(t-u,\theta) D E_c^{(n)}(u,\theta)\, du.$$

Proof. See e-companion section EC.1. $\square$

The closed-form expressions of the derivatives of the exponential maps are presented in the next two sections. First, we study the case of the set of diagonalizable matrices; we then extend to the more general class of matrices. Finally, we provide the exact expression for the derivative of a column vector derived from a matrix exponential map.

3.2.1. Diagonalizable Matrices. We present first a closed-form formula for calculating a higher order derivative of the exponential of a diagonalizable matrix with distinct eigenvalues. Let $\mathcal{D}$ be the set of all diagonalizable $p \times p$ matrices with distinct eigenvalues.

Assuming $d_1, \ldots, d_p$ are distinct complex numbers, we define a $p \times p$ matrix Γ with the (i,j)-th element given by

$$\Gamma_{ij}(t,\theta) := \mathcal{L}_{d_i} e^{t d_j}, \text{ for any } i,j \in \{1,\ldots,p\}, \tag{3}$$

where, for $i = 1,\ldots,p$, the linear operator $\{\mathcal{L}_{d_i}\}$ is defined as

$$\mathcal{L}_{d_i} f(t) = \int_0^t e^{(t-u)d_i} f(u)\, du,$$

for all integrable functions f on $[0,\infty)$. Before proceeding further, we establish some auxiliary results related to important properties of the operator $\mathcal{L}_{d_i}$, when acting on some integrable functions, that are required for the upcoming results of this paper. With slight extension of our notation, we use, for any integer $n \geq 0$, $\Gamma(t,\theta;n)$ to denote a multidimensional matrix of dimension $\underbrace{p \times \cdots \times p}_{n+2}$ and, for $i_1,\ldots,i_n,j,k \in \{1,\ldots,p\}$, the $(i_1,i_2,\ldots,i_n,k,j)$-th element satisfies the recurrence relation

$$\Gamma_{kj}^{i_1,i_2,\ldots i_n}(t,\theta;n) := \mathcal{L}_{d_{i_1}} \Gamma_{kj}^{i_2,\ldots i_n}(t,\theta;n-1), \quad n \geq 1, \tag{4}$$

where $\Gamma(t,\theta;0) \equiv \Gamma(t,\theta)$. The exact expression for $\Gamma(t,\theta;n)$ is given in the following proposition.

PROPOSITION 4. *Let $\{d_1,\ldots,d_p\}$ be the set of distinct complex numbers. For any integer $n\geq 0$, let $\Gamma(t,\theta;n)$ be as in (4). Then,*

1. *for $i,j=1,2,\ldots,p$, we have that*

$$\Gamma_{ij}(t,\theta;0)=\begin{cases}\frac{e^{td_j}-e^{td_i}}{d_j-d_i}, & \text{if } i\neq j\\ te^{td_i}, & \text{if } i=j\end{cases} \tag{5}$$

2. *for $i_1,\ldots,i_n,j,k\in\{1,\ldots,p\}$ and $n>1$, we have that*

$$\Gamma_{kj}^{i_1\ldots i_n}(t,\theta;n)=\begin{cases}\frac{\Gamma_{i_nj}^{i_1\ldots i_{n-1}}(t,\theta;n-1)-\Gamma_{i_nk}^{i_1\ldots i_{n-1}}(t,\theta;n-1)}{d_j-d_k}, & \text{if } \ k\neq j\\ \frac{\Gamma_{jj}^{i_1\ldots i_{n-1}}(t,\theta;n-1)-\Gamma_{i_nj}^{i_1\ldots i_{n-1}}(t,\theta;n-1)}{d_j-d_{i_n}}, & \text{if } \ \ i_n\neq k=j\\ \vdots & \\ \frac{\Gamma_{jj}^{j\ldots j}(t,\theta;n-1)-\Gamma_{i_1j}^{i_1\ldots j}(t,\theta;n-1)}{d_j-d_{i_1}}, & \text{if } \ i_1\neq i_2=\cdots=i_n=k=j\\ \frac{t}{n+1}\Gamma_{jj}^{j\ldots j}(t,\theta;n-1), & \text{if } \ i_1=i_2=\cdots=i_n=k=j\end{cases}.$$

Proof. See e-companion section EC.1. □

The closed-form expression for the derivatives of $E_c(t,\theta)\in\mathcal{D}$ is presented in the following theorem. It is worth noting that our result for the first order derivative is equivalent to that of Kalbfleisch and Lawless (1985). For convenience, we present the algorithm which computes the exact formulae for the first four derivatives in the e-companion section EC.2.

THEOREM 1. *Let $\{d_1,\ldots,d_p\}$ be the set of eigenvalues of a matrix $(G-\theta D)\in\mathcal{D}$ for $\theta\in\mathbb{C}$. Furthermore, assume that $Q(\theta)$ and $M(\theta)$ are such that $G-\theta D=Q(\theta)M(\theta)Q^{-1}(\theta)$, where $M(\theta)$ is a diagonal matrix with $M_{ii}(\theta)=d_i$ for any $i\in\{1,\ldots,p\}$. Let $\Gamma(t,\theta;\cdot)$ be as in Proposition 4. Then, for $n\geq 1$ and any $i,j\in\{1,\ldots,p\}$, the n-th order derivative can be expressed in the form*

$$E_c^{(n)}(t,\theta)=(-1)^n n!\,Q(\theta)\tilde{\Gamma}^{(n)}(t,\theta)Q^{-1}(\theta),$$

where

$$\tilde{\Gamma}_{ij}^{(n)}(t,\theta):=\begin{cases}\sum\limits_{i_1,\ldots,i_{n-1}}\Gamma_{i_{n-1}j}^{ii_1\ldots i_{n-2}}(t,\theta;n-1)L_{ii_1}(\theta)\cdots L_{i_{n-1}j}(\theta), \ \text{for } n\geq 2\\ \Gamma_{ij}(t,\theta;0)L_{ij}(\theta), \ \text{for } n=1\end{cases}$$

and $L(\theta)=[L_{ij}(\theta)]_{p\times p}:=Q^{-1}(\theta)DQ(\theta)$.

Proof. We prove the result by induction. To compute the first order derivative of $E_c(t,\theta)$, we apply Proposition 3:

$$E_c^{(1)}(t,\theta)=-\int_0^t e^{(t-u)(G-\theta D)}De^{u(G-\theta D)}\,du.$$

Using the decomposition $G-\theta D=Q(\theta)M(\theta)Q^{-1}(\theta)$, the above equation can be rewritten as

$$\begin{aligned} E_c^{(1)}(t,\theta) &= -\int_0^t Q(\theta)e^{(t-u)M(\theta)}Q^{-1}(\theta)DQ(\theta)e^{uM(\theta)}\,du\,Q^{-1}(\theta) \\ &= -Q(\theta)\int_0^t e^{(t-u)M(\theta)}L(\theta)e^{uM(\theta)}\,du\,Q^{-1}(\theta). \end{aligned}$$

It is obvious that, for any $i,j=1,2,\dots,p$, the (i,j)-th element of $\int_0^t e^{(t-u)M(\theta)}L(\theta)e^{uM(\theta)}\,du$ is given by $\int_0^t e^{(t-u)M_{ii}(\theta)}L_{ij}(\theta)e^{uM_{jj}(\theta)}\,du$, i.e., $\mathcal{L}_{d_i}\left(L_{ij}(\theta)e^{td_j}\right)$. Again, since $L_{ij}(\theta)$ is constant in t, $\mathcal{L}_{d_i}\left(L_{ij}(\theta)e^{td_j}\right)=L_{ij}(\theta)\mathcal{L}_{d_i}e^{td_j}=L_{ij}(\theta)\Gamma_{ij}(t,\theta;0)=\tilde{\Gamma}_{ij}^{(1)}(t,\theta)$. Thus, we prove our claim for $n=1$.

Now suppose that for $n=l$ the statement is true, i.e.,

$$E_c^{(l)}(t,\theta)=(-1)^l l!\,Q(\theta)\tilde{\Gamma}^{(l)}(t,\theta)Q^{-1}(\theta)$$

with

$$\tilde{\Gamma}_{ij}^{(l)}(t,\theta)=\sum_{i_1,\dots,i_{l-1}}\Gamma_{i_{l-1}j}^{ii_1\dots i_{l-2}}(t,\theta;l-1)L_{ii_1}\cdots L_{i_{l-1}j}.$$

Again from Proposition 3, we get that

$$\begin{aligned} E_c^{(l+1)}(t,\theta) &= -(l+1)\,Q(\theta)\int_0^t e^{(t-u)M(\theta)}Q^{-1}(\theta)D\left((-1)^l l!Q(\theta)\tilde{\Gamma}^{(l)}(u,\theta)Q^{-1}(\theta)\right)\,du \\ &= (-1)^{l+1}(l+1)!\,Q(\theta)\int_0^t e^{(t-u)M(\theta)}L(\theta)\tilde{\Gamma}^{(l)}(u,\theta)\,du\,Q^{-1}(\theta). \end{aligned}$$

Similarly to the case $n=1$, the (i,j)-th element of the matrix $\int_0^t e^{(t-u)M(\theta)}L(\theta)\tilde{\Gamma}^{(l)}(u,\theta)\,du$ can be obtained from the expression $\mathcal{L}_{d_i}[L(\theta)\tilde{\Gamma}^{(l)}(t,\theta)]_{ij}$; using Proposition 4, we get that

$$\begin{aligned} \mathcal{L}_{d_i}\left(\sum_k L_{ik}(\theta)\tilde{\Gamma}_{kj}^{(l)}(t,\theta)\right) &= \mathcal{L}_{d_i}\left(\sum_k L_{ik}(\theta)\sum_{i_1,\dots,i_{l-1}}\Gamma_{i_{l-1}j}^{ki_1\cdots i_{l-2}}(t,\theta;l-1)L_{ki_1}(\theta)\cdots L_{i_{l-1}j}(\theta)\right) \\ &= \sum_{k,i_1,\dots,i_{l-1}}\Gamma_{i_{l-1}j}^{iki_1\cdots i_{l-2}}(t,\theta;l)L_{ik}(\theta)L_{ki_1}(\theta)\cdots L_{i_{l-1}j}(\theta)=\tilde{\Gamma}_{ij}^{(l+1)}(t,\theta), \end{aligned}$$

hence the result is proved. □

The closed-form derivatives of exponential diagonalizable matrices with repeated eigenvalues follow from a straightforward modification of Proposition 4 and re-establishment of Theorem 1 accordingly.

3.2.2. Beyond Diagonalizable Matrices. We present next an approach to finding the closed-form representation for the non-diagonalizable class of matrices; a further rigorous analysis is currently in progress.

It is well-known that, for any matrix $A \in M_p(\mathbb{C})$ of the form $G - \theta D$, we can find a diagonalizable matrix $A_\varepsilon \in \mathcal{D}$ such that $\|A - A_\varepsilon\|_F < \varepsilon$, for any $\varepsilon > 0$, where $\|\cdot\|_F$ denotes the matrix Frobenius-norm or simply F-norm. (For more details, see Horn and Johnson 2012, Theorem 2.4.7.1.) Therefore, we conjecture that

$$\lim_{\varepsilon\to 0} \frac{d^n}{d\theta^n} e^{tA_\varepsilon(\theta)} = \frac{d^n}{d\theta^n} e^{tA(\theta)} \quad \text{in } F\text{-norm},$$

where, as $\varepsilon \to 0$, $A_\varepsilon(\theta) \to A(\theta)$ in F-norm. This idea is illustrated in the following simple example by computing the first order derivative.

Example 1. Consider $G = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and $D = I_2$. Let $A(\theta) = G - \theta D$. Then, $e^{A(\theta)} = \begin{pmatrix} e^{-\theta} & e^{-\theta} \\ 0 & e^{-\theta} \end{pmatrix}$. Therefore, the first order derivative of $e^{A(\theta)}$ with respect to θ is $\begin{pmatrix} -e^{-\theta} & -e^{-\theta} \\ 0 & -e^{-\theta} \end{pmatrix}$. As $A(\theta)$ is not diagonalizable, we consider the perturbed matrix $A_\varepsilon(\theta) = \begin{pmatrix} -\theta + \varepsilon & 1 \\ 0 & -\theta \end{pmatrix}$. Hence, $A_\varepsilon(\theta)$ is diagonalizable with

$$A_\varepsilon(\theta) = Q(\theta) M(\theta) Q^{-1}(\theta),$$

where

$$M(\theta) = \begin{pmatrix} -\theta & 0 \\ 0 & -\theta + \varepsilon \end{pmatrix}, \quad Q(\theta) = \begin{pmatrix} -\frac{1}{\varepsilon} & 1 \\ 1 & 0 \end{pmatrix}, \quad Q^{-1}(\theta) = \begin{pmatrix} 0 & 1 \\ 1 & \frac{1}{\varepsilon} \end{pmatrix}.$$

Now, we apply Theorem 1 to find the first order derivative of $e^{A_\varepsilon(\theta)}$. Also, from the first part of Proposition 4, the matrix Γ is given by $\begin{pmatrix} e^{-\theta} & \frac{e^{-\theta+\varepsilon} - e^{-\theta}}{\varepsilon} \\ \frac{e^{-\theta+\varepsilon} - e^{-\theta}}{\varepsilon} & e^{-\theta+\varepsilon} \end{pmatrix}$ and $L = Q^{-1}(\theta) D Q(\theta) = I_2$. Therefore, the first order derivative of $e^{A_\varepsilon(\theta)}$ is given by

$$-Q[\Gamma_{ij} L_{ij}]_{2\times 2} Q^{-1} = \begin{pmatrix} -e^{-\theta+\varepsilon} & -\frac{e^{-\theta+\varepsilon} - e^{-\theta}}{\varepsilon} \\ 0 & -e^{-\theta} \end{pmatrix},$$

and, as $\varepsilon \to 0$, we get the result.

In many applications it is required to compute the derivatives of a column vector of the form $e^{t(G-\theta D)}\mathbf{x}$, where $(G - \theta D) \in M_p(\mathbb{C})$ and $\mathbf{x}$ is a column vector that does not depend on θ, rather than the derivatives of the full exponential matrix. In the next section, we establish the combined representation of the first n derivatives of this column vector.

3.2.3. Derivatives of a Column Vector of the Form $E_c(t,\theta)\mathbf{x}$. We start by introducing some notation.

DEFINITION 1. 1. Let B_q be a $q\times q$ matrix whose (i,j)-th element is given by $[B_q]_{ij} := i\delta_j^{i+1}$, where $\delta_j^i = \begin{cases} 1 & \text{if} \quad i=j \\ 0 & \text{otherwise} \end{cases}$ for $i,j=1,\dots,q$.

2. $C_{p,q}^{\mathbf{x}}$ denotes a $p\times q$ matrix whose only first column is non-zero and is given by $\mathbf{x}$.

Consider $E_c(t,\theta)=e^{t(G-\theta D)}$ for $(G-\theta D)\in M_p(\mathbb{C})$ and $t\geq 0$. For any non-negative integer n, we define the $p\times(n+1)$ matrix

$$\mathcal{E}_c^{(0,\dots,n)}(t,\theta) := \left[E_c(t,\theta)\mathbf{x}, \frac{d}{d\theta}\left\{E_c(m,\theta)\mathbf{x}\right\}, \dots, \frac{d^n}{d\theta^n}\left\{E_c(m,\theta)\mathbf{x}\right\}\right], \tag{6}$$

i.e., the j-th column of the matrix $\mathcal{E}_c^{(0,\dots,n)}(t,\theta)$ is $\frac{d^j}{d\theta^j}\left\{E_c(m,\theta)\mathbf{x}\right\}$ for any $j=1,\dots,n+1$. We show in the following lemma that $\mathcal{E}_c^{(0,\dots,n)}(t,\theta)$ satisfies a matrix differential equation. The proof is sketched in the e-companion.

LEMMA 1. *For any non-negative integer n, let $\mathcal{E}_c^{(0,\dots,n)}(t,\theta)$ be as in (6). Then, we have that*

$$\begin{aligned} &\frac{d}{dt}\mathcal{E}_c^{(0,\dots,n)}(t,\theta) = (G-\theta D)\mathcal{E}_c^{(0,\dots,n)}(t,\theta) - D\mathcal{E}_c^{(0,\dots,n)}(t,\theta)B_{n+1} \\ &\textit{with } \mathcal{E}_c^{(0,\dots,n)}(0,\theta) = C_{p,(n+1)}^{\mathbf{x}}, \end{aligned} \tag{7}$$

where B_{n+1} and $C_{p,(n+1)}^{\mathbf{x}}$ follow Definition 1 with dimensions $(n+1)\times(n+1)$ and $p\times(n+1)$, respectively.

Proof. See e-companion section EC.1. □

Before solving the matrix ordinary differential equation (7), we recall some definitions and properties of the Kronecker product and the vectorization operation. (For more details, refer to Magnus and Neudecker 2019, Chapter 2.) Let X and Y be $m\times n$ and $p\times q$ matrices, respectively. Then, the Kronecker product of X and Y is given by the $mp\times nq$ matrix $X\otimes Y := (x_{ij}Y)_{\substack{i=1,\dots,m \\ j=1,\dots,n}}$; the vectorization of X is the $mn\times 1$ column vector $\text{vec}(X)=(x_1^*, x_2^*, \dots, x_n^*)^*$, where $x_1, x_2, \dots, x_n$ are the column vectors of X. Finally, for an additional matrix Z, we have that

$$\text{vec}(XYZ) = (Z^*\otimes X)\text{vec}(Y). \tag{8}$$

THEOREM 2. *The solution (in terms of vectorization) to the matrix differential equation* (7) *is given by*

$$vec(\mathcal{E}_c^{(0,\dots,n)}(t,\theta)) = \exp\left[t\left(I_{n+1}\otimes(G-\theta D) - (B_{n+1}^*\otimes D)\right)\right]\cdot\left(\mathbf{e}_1^{n+1}\otimes\mathbf{x}\right), \tag{9}$$

where B_{n+1} is given in the first part of Definition 1 and $\mathbf{e}_1^{n+1}$ in Proposition 1.

Proof. From property (8), the matrix differential equation (7) can be rewritten as

$$\frac{d}{dt}\text{vec}(\mathcal{E}_c^{(0,\dots,n)}(t,\theta)) = \left(I_{n+1}\otimes(G-\theta D) - B_{n+1}^*\otimes D\right)\text{vec}(\mathcal{E}_c^{(0,\dots,n)}(t,\theta)), \tag{10}$$

with the initial condition $\text{vec}(\mathcal{E}_c^{(0,\dots,n)}(0,\theta)) = \mathbf{e}_1^{n+1}\otimes\mathbf{x}$. The proof is completed by solving (10). □

4. Application

As explained in Section 2.1, we are interested in building the density function of the random sum based on its moments, which we compute using the theory developed in the previous sections and which in its implementation is much simpler than it may look. We can evaluate any moments, however we focus on the first four as required for the Pearson curve fit.

4.1. The Discrete Sum

We define the $p \times 5$ matrix $\mathcal{E}_d^{(0,\ldots,4)}(m,\theta) = \left[E_d^{(0)}(m,\theta)\mathbf{1},\ldots,E_d^{(4)}(m,\theta)\mathbf{1}\right]$. Then, from Proposition 2, we get that

$$\mathcal{E}_d^{(0,\ldots,4)}(m+1,\theta) = e^{-\theta D}\sum_{k=0}^{4}\frac{(-1)^k}{k!}D^k P(\Delta)\mathcal{E}_d^{(0,\ldots,4)}(m,\theta)B_5^k, \tag{11}$$

with $\mathcal{E}_d^{(0,\ldots,4)}(0,\theta) = e^{-\theta D}\sum_{k=0}^{4}\frac{(-1)^k}{k!}D^k C_{p,5}^{\mathbf{1}} B_5^k$ for B_5 and $C_{p,5}^{\mathbf{1}}$ as in Definition 1.

4.2. The Time Integral

If $(G-\theta D) \in \mathcal{D}$, we can derive the required moments from Theorem 1. If, instead, $(G-\theta D) \notin \mathcal{D}$, we can use the limit of small perturbation as explained in Section 3.2.2. Nevertheless, alternatively to both, the first four derivatives of the vector $E_c(t,\theta)\mathbf{1}$ with respect to θ, for *any* matrix $G-\theta D$, can be obtained from Theorem 2.

4.3. Computational Complexity

This section presents the time complexity of evaluating up to the fourth order derivative of $E_d(m,\theta)$, $E_d(m,\theta)\mathbf{1}$ (discrete case) and $E_c(t,\theta)$, $E_c(t,\theta)\mathbf{1}$ (continuous case) at $\theta = 0$, of interest to us, using our method. We provide the algorithms for computing $E_c(t,\theta)$ and $E_c(t,\theta)\mathbf{1}$ in the e-companion section EC.2, whereas we discuss a time-efficient computational approach to $E_d(m,\theta)\mathbf{1}$ in this section.

The time complexity of the derivatives in the discrete cases is described in items a)–b), whereas the details of the continuous cases are given in items c)–d):

a) The differentiation of $E_d(m,\theta)$ requires evaluation of the formula presented in Proposition 2 involving matrix multiplications of the order $p \times p$, whose worst-case complexity is well known to be $O(p^3)$.

b) In computing the combined derivatives of $E_d(m,\theta)\mathbf{1}$ using formula (11), we have considered both the suggestion of backward multiplication in (Cui et al. 2018, Remark 2) and forward multiplication. First, note that, for any $k = 0,\ldots,4$, $e^{-\theta D}D^k$ is diagonal and the operation $L := e^{-\theta D}D^k P(\Delta)$ costs $O(kp)$. Also, the multiplication $M := \mathcal{E}_d^{(0,\ldots,4)}(m,\theta)B_5^k$ costs $O(5^2kp)$ as B_5 is a 5×5 square matrix. Again, as M is a $p \times 5$ matrix, the multiplication LM costs $O(5p^2)$. Therefore, the total cost of evaluating $\sum_{k=0}^{4}((-1)^k/k!)e^{-\theta D}D^k P(\Delta)\mathcal{E}_d^{(0,\ldots,4)}(m,\theta)B_5^k$ is $O(p^2)$.

c) The most expensive part of the computation of the derivatives of $E_c(t,\theta)$ up to the fourth order using Theorem 1 is the construction of a $p \times p \times p \times p \times p$ matrix, that is, $\Gamma(t,\theta;3)$ of Proposition 4, whose worst-case complexity is $O(p^5)$ (see Algorithm EC.2.1 in the e-companion section EC.2).

d) The combined derivative of the column vector of the form $E_c(t,\theta)\mathbf{1}$ requires evaluation of the exponential of a $5p \times 5p$ matrix (see expression 9), which we implement using the `@expm` function in Matlab. We have from Moler and Van Loan (2003) that the worst-case complexity of computing a matrix exponential is $O(p^3)$, which holds also in our case of Theorem 2 (see Algorithm EC.2.2 in the e-companion section EC.2). For a faster multiplication of a matrix exponential and a column vector, it is possible to use the `@expokit` function (see Sidje 1998 for more details).

In summary, the direct evaluation of the derivatives of $E_d(m,\theta)\mathbf{1}$ reduces the computational cost by $O(p)$ compared to the derivatives of $E_d(m,\theta)$. Similarly, $O(p^2)$ cost reduction results from computing the derivatives of $E_c(t,\theta)\mathbf{1}$ instead of $E_c(t,\theta)$. Matlab codes linked to these computations are made available from `https://github.com/milan30/DME`.

4.4. Illustrative Examples

We consider the expected values

$$C_c(t,k;S_0) := \mathbb{E}_{\mathbb{P}}\left[\left(\int_0^t S_w\,dw - k\right)^+ \middle| S_0\right] \text{ and } C_d(m,k;S_0) := \mathbb{E}_{\mathbb{P}}\left[\left(\sum_{i=0}^m S_{t_i} - k\right)^+ \middle| S_0\right], \qquad (12)$$

where $y^+ := \max(y,0)$, k is a non-negative constant and S represents some asset price process such that $\mathbb{E}_{\mathbb{P}}(S_t|S_0) = S_0 e^{(r-\lambda)t}$ with r denoting the risk-free interest rate and λ the dividend rate. For strike price K and maturity time T, the quantities $(e^{-rT}/T)C_c(T,TK;S_0)$ and $(e^{-rT}/(m+1))C_d(m,(m+1)K;S_0)$ correspond to the prices at time 0 of Asian call options with, respectively, continuous monitoring of the underlying asset price S; discrete monitoring at the equidistant times $t_0 = 0, t_1, \ldots, t_m = m\Delta = T$.

4.4.1. Error Analysis. Before moving to the computation of (12), we brief on the error associated with this. The expected values in (12) are given, by virtue of straightforward equivalence result, by

$$\varrho S_0 - k + \int_0^{\mu_1+k} (\mu_1 + k - z) f(z) dz = \varrho S_0 - k + \int_0^{\mu_1+k} F(z) dz, \qquad (13)$$

where $\mu_1 = \mathbb{E}_{\mathbb{P}}\left[\int_0^t S_w dw\right]$ in the continuous case or $\mathbb{E}_{\mathbb{P}}\left[\sum_{i=0}^m S_{t_i}\right]$ in the discrete case, F and f denote, respectively, the true distribution and density functions, $\varrho \equiv (e^{(r-\lambda)t} - 1)/(r-\lambda)$ in the continuous case and $\varrho \equiv (1 - e^{(r-\lambda)(m+1)\Delta})/(1 - e^{(r-\lambda)\Delta})$ in the discrete case. The first type of error is incurred by approximating the true distribution by the moment-based Pearson curve fit with associated distribution and density functions G and g, respectively. Based on a result due to (Akhiezer 1965, Corollary 2.5.4), which is revisited by (Lindsay and Basak 2000, Theorems 1, 2), we get that

$$\left|\int_0^{\mu_1+k}[F(z)-G(z)]\,dz\right|\le\int_0^{\mu_1+k}|F(z)-G(z)|\,dz\le\int_0^{\mu_1+k}\varpi_{\tilde n}(z)\,dz=:\hat e_1(k;\mu_1,\dots,\mu_{2\tilde n}),$$

where $\varpi_{\tilde n}(z):=\left\{P'_{\tilde n}(z)W_{\tilde n}^{-1}P_{\tilde n}(z)\right\}^{-1}$, $P_{\tilde n}(z):=(1,z,z^2,\dots,z^{\tilde n})'$ and $W_{\tilde n}:=\|\mu_{i+j}\|_{i,j=0}^{\tilde n}$ is a Hankel symmetric matrix defined by the first $2\tilde n$ moments. $\varpi_{\tilde n}(z)$ goes to 0 at the rate $z^{-2\tilde n}$ as $z\to\infty$ giving relatively sharp tail information and guaranteeing accuracy. For example, for $\tilde n=1$,

$$\hat e_1(k;\mu_1,\mu_2)=\sqrt{\mu_2}\arctan\frac{\mu_1+k}{\sqrt{\mu_2}},$$

but $\hat e_1$ can be computed easily and accurately numerically for any $\tilde n$. An improvement to this bound is due to Khamis (1954) who introduces a constant non-negative multiplier that is smaller than the unity and is given by $1+\min\{\text{l.u.b}_{c\le z\le d}(-f(z)/g(z)),\,\text{l.u.b}_{c\le z\le d}(-g(z)/f(z))\}$ if it exists. For more on the proximity of distributions with shared moments as well as the estimation of their closeness in different moment-based metrics, interested readers may refer to Kyriakou et al. (2021).

The second source of error is due to the numerical evaluation of the integral in (13) based on the approximating distribution function G. For this, we use adaptive quadrature (e.g., see Shampine et al. 1997, Chapter 5; Shampine 2008) where the interval $[0,\mu_1+k]$ is partitioned into subintervals $0=:\eta_1<\vartheta_1=\eta_2<\vartheta_2=\eta_3<\dots<\vartheta_N:=\mu_1+k$ on which the *basic* quadrature rule R is sufficiently accurate. This yields

$$\int_0^{\mu_1+k}G(z)dz=\sum_{j=1}^N\int_{\eta_j}^{\vartheta_j}G(z)dz=\sum_{j=1}^N R_j+\hat e_2.$$

The error $\hat e_2$ is estimated by comparing it to a more accurate result. To this end, an additional rule $\bar R$ is introduced which is believed to be more accurate. In particular, R is given by the three-point Gaussian quadrature formula and $\bar R$ by the seven-point Kronrod formula, for which we have

$$\hat e_2:=\sum_{j=1}^N\left[\bar R_j-R_j+7.14\times10^{-17}(\vartheta_j-\eta_j)^{13}G^{(12)}(\xi_j)\right]$$

for some $\xi_j\in(\eta_j,\vartheta_j)$, where $G^{(n)}$ denotes the n-th derivative of G and $\bar R_j-R_j$ is an error estimate of R_j. If the current approximation is not sufficiently accurate, subintervals, ranked by largest error, are further refined for improvement until the intended error tolerance is satisfied.

Finally, approximating the targeted Markov model with a CTMC induces also some error. However, this is common to both Cai et al. (2015) and Cui et al. (2018) and decreases with increasing number of states for the approximating CTMC; readers may refer to (Cai et al. 2015, Section 6.2) for illustrations of this.

4.4.2. Numerical Results. In Tables 1–2, we present a battery of numerical results corresponding to Asian option prices for varying strikes, monitoring frequencies and underlying model assumptions, including the Cox–Ingersoll–Ross (CIR) and constant elasticity of variance (CEV) diffusions, the double-exponential jump diffusion (DEJD), the Merton jump diffusion (MJD), the Carr–Geman–Madan–Yor (CGMY) and the variance gamma (VG) model. We also consider various parameter values and benchmarks as in Cai et al. (2015) and Cui et al. (2018), more details of which can be found in the tables. For a fair comparison, we use the same equipment as Cui et al. (2018) for the execution of the numerical experiments, that is, Matlab on a machine with an Intel Core 2 i7 CPU @ 2GHz and 8GB of RAM.

Comparisons with the benchmarks and the earlier methods of Cai et al. (2015) and Cui et al. (2018) divulge the superiority of our moment-based approximation. In almost all cases under consideration, we achieve smaller (absolute) error than both the other techniques, or on a few occasions nearby errors, implying generally larger total error involved in the Laplace transform inversion than the error from our moment-based approximation. This is also confirmed by inspection of the Q-Q plots in Figure 1 comparing, for different kinds of driving dynamics, the simulated true distribution of the arithmetic average and the corresponding Pearson distribution approximation. It is obvious that the points in the Q-Q plots follow very closely the line $y = x$ with minor departures in the tails (subject to simulation error). Also, two-sample Kolmogorov–Smirnov tests lead to acceptance of the null hypothesis that the two samples come from the same distribution with p-values above 10%. In addition, bypassing the transform inversion using our method depletes the CPU time as shown in Figure 2 on log-scale. Reducing double to single transform inversion reduces the time it takes to compute this. Our approach results in further reduction at an increasing rate with the averaging frequency that goes above a factor of 4. This translates to CPU time of one to two hundredths of a second for $m = 250$ and $m = \infty$, while warranting almost higher precision which suffices for practical purposes. This is particularly welcome news for highly intensive problems involving integrated stochastic processes (case $m = \infty$) as highlighted in our concluding remarks.

5. Conclusions

In this article, we have derived closed-form expressions for the derivatives of matrix exponentials for diagonalizable matrices. We have also discussed extensions to non-diagonalizable matrices and derivatives of vectors.

In our application, we have focused on the typical example of Asian option evaluation. Nevertheless, our method offers an attractive speed–accuracy package that is transferable to other applications of interest in financial engineering and beyond, as discussed in the introduction. We quickly recall a typical case that arises in simulation problems, that of the simulation of integrated

K	Benchmark	Moment-based approximation	Cai et al.	Cui et al.	K	Benchmark	Moment-based approximation	Cai et al.	Cui et al.
		CIR					Lognormal (CEV $\beta=0$)		
0.9	0.21575	-9.2E-05	-2.3E-04	1.7E-04	2	0.05599	-1.8E-04		
0.95	0.18958	-5.7E-05	-2.1E-04	1.8E-04	2	0.21839	-9.7E-05		
1	0.16580	-2.6E-05	-1.8E-04	2.0E-04	2	0.19317	4.9E-04		
1.05	0.14433	4.3E-06	-1.5E-04	2.4E-04	2	0.24642	4.3E-04		
1.1	0.12506	4.0E-05	-1.0E-04	2.8E-04	2	0.30622	3.2E-04		
		CEV ($\beta=0.25$)					DEJD ($\sigma=0.1$)		
80	21.59408	1.3E-02	1.7E-02	1.7E-02	90	13.55964	-8.1E-04	4.5E-03	4.3E-03
90	13.15109	-2.1E-06	8.2E-03	8.1E-03	95	9.41962	-5.5E-03	9.7E-03	5.1E-03
100	6.83859	-1.1E-03	-7.3E-03	-7.1E-03	100	5.91537	7.8E-04	-1.7E-03	2.4E-03
110	3.07333	6.1E-04	-1.2E-02	-1.2E-02	105	3.35071	2.9E-03	-2.4E-03	7.2E-04
120	1.23175	-3.9E-04	-4.1E-03	-4.1E-03	110	1.74896	1.6E-04	5.3E-03	4.7E-04
		CEV ($\beta=-0.25$)					DEJD ($\sigma=0.3$)		
80	21.66618	1.1E-02	1.5E-02	1.5E-02	90	15.33688	-2.5E-03	-1.4E-03	-1.1E-03
90	13.26741	-1.3E-03	4.1E-03	4.0E-03	95	12.10723	-6.0E-03	-3.1E-03	-2.8E-03
100	6.85150	-7.5E-04	-1.2E-02	-1.2E-02	100	9.35336	-8.3E-03	-4.5E-03	-4.2E-03
110	2.93166	1.3E-03	-1.1E-02	-1.1E-02	105	7.08059	-8.9E-03	-5.4E-03	-5.1E-03
120	1.04453	2.6E-03	-2.4E-04	-3.3E-04	110	5.26109	-7.8E-03	-5.5E-03	-5.2E-03
		CEV ($\beta=-0.5$)					DEJD ($\sigma=0.5$)		
80	21.71118	9.4E-03	1.3E-02	1.3E-02	90	18.46259	-8.5E-03	-9.7E-03	-1.1E-03
90	13.32850	4.9E-03	2.0E-03	1.9E-03	95	15.75006	-5.0E-03	-1.1E-02	-4.3E-03
100	6.85984	-4.2E-03	-1.6E-02	-1.6E-02	100	13.36027	-2.6E-03	-1.3E-02	-6.4E-03
110	2.86666	-1.3E-03	-1.4E-02	-1.4E-02	105	11.27716	-1.4E-03	-1.4E-02	-7.7E-03
120	0.95995	2.8E-03	8.9E-04	7.5E-04	110	9.47826	-1.3E-03	-1.4E-02	-8.2E-03
		MJD					CGMY		
90	12.74587	7.6E-04	1.2E-03	1.1E-03	90	12.74788	-1.0E-04	-9.9E-04	-8.0E-05
100	5.05974	3.1E-04	-2.3E-03	1.2E-03	100	5.08865	-7.5E-03	-8.5E-03	-7.3E-03
110	1.08413	1.4E-03	8.2E-03	3.0E-03	110	1.05810	-4.8E-03	2.2E-03	-5.9E-04
		VG					VG		
100	5.59320	3.2E-03	-6.3E-03		100	6.46870	1.2E-03	-1.4E-03	
100	5.85850	3.1E-03	-5.2E-03		100	6.22890	5.4E-04	6.0E-04	

Table 1 Pricing continuous Asian options in the CIR, CEV, lognormal, DEJD, MJD, CGMY and VG models via the CTMC approximation based on finite state space with number of states $p=50$ (as in Cai et al. 2015 and Cui et al. 2018) using our moment-based method and the methods from the aforementioned papers. K denotes the strike price. "Error" columns report the differences of the indicated method with respect to the benchmark. CIR parameters: (Cai et al. 2015, Table 3); benchmark: Fusai et al. (2008). CEV parameters: (Cai et al. 2015, Table 4); benchmark: Cai et al. (2014). Lognormal parameters: (Fusai and Kyriakou 2016, Table 10); benchmark: Cai and Kou (2012). DEJD parameters: (Cai et al. 2015, Table 5); benchmark: Cai and Kou (2012). MJD parameters: (Cai et al. 2015, Table 6); benchmark: Monte Carlo (MC) price estimates. VG parameters: (Cai et al. 2015, Tables 7–8); benchmark: MC price estimates. CGMY parameters: (Cai et al. 2015, Table 9); benchmark: MC price estimates. All MC estimates are based on 10^6 simulation trials and 10^4 time steps and are from Cai et al. (2015) with standard errors reported there.

functionals of stochastic volatility (see Cui et al. 2021 for generalized SABR and stochastic local volatility models), which can be slow especially when generating entire asset price trajectories and where the bias from their approximation, that is hard to quantify in practice, would be desired to be safely assumed negligible. We are currently exploring further speed-up and precision enhance-

K	Benchmark	Moment-based approximation	Cai et al.	Cui et al.	K	Benchmark	Moment-based approximation	Cai et al.	Cui et al.
		CIR ($m=12$)					CIR ($m=25$)		
0.9	0.21279	-9.0E-05	-2.2E-04	2.1E-04	0.9	0.21428	-8.9E-05	-2.2E-04	2.1E-04
0.95	0.18659	-5.4E-05	-2.1E-04	1.5E-04	0.95	0.18810	-6.1E-05	-2.1E-04	1.3E-04
1	0.16282	-1.5E-05	-1.8E-04	1.5E-04	1	0.16432	-2.3E-05	-1.8E-04	1.3E-04
1.05	0.14140	2.4E-05	-1.4E-04	1.8E-04	1.05	0.14287	1.4E-05	-1.4E-04	1.6E-04
1.1	0.12223	5.5E-05	-1.0E-04	2.2E-04	1.1	0.12365	4.6E-05	-1.0E-04	2.0E-04
		CIR ($m=50$)					CIR ($m=100$)		
0.9	0.21501	-9.7E-05	-9.5E-04	2.0E-04	0.9	0.21538	-9.8E-05	-2.3E-04	2.0E-04
0.95	0.18883	-6.1E-05	-2.1E-04	1.3E-04	0.95	0.18920	-5.7E-05	-2.1E-04	1.3E-04
1	0.16505	-2.7E-05	-1.8E-04	1.2E-04	1	0.16542	-2.5E-05	-1.8E-04	1.2E-04
1.05	0.14359	6.5E-06	-1.5E-04	1.5E-04	1.05	0.14395	1.2E-05	-1.4E-04	1.5E-04
1.1	0.12434	4.5E-05	-1.0E-04	1.9E-04	1.1	0.12470	3.9E-05	-1.0E-04	1.9E-04
		CIR ($m=250$)					CEV ($\beta=0.25, m=250$)		
0.9	0.21560	-9.3E-05	-2.3E-04	2.1E-04	80	21.60167	4.7E-03	8.1E-03	8.1E-03
0.95	0.18943	-6.0E-05	-2.1E-04	1.3E-04	90	13.15550	-8.6E-04	-2.0E-05	1.0E-05
1	0.16565	-2.8E-05	-1.8E-04	1.3E-04	100	6.84034	-4.4E-03	-1.4E-02	-1.4E-02
1.05	0.14418	4.6E-06	-1.5E-04	1.4E-04	110	3.07180	-2.3E-03	-1.5E-02	-1.5E-02
1.1	0.12492	3.5E-05	-1.1E-04	1.8E-04	120	1.22841	2.2E-03	-3.4E-03	-3.4E-03
		CEV ($\beta=-0.25, m=250$)					CEV ($\beta=-0.5, m=250$)		
80	21.67122	5.0E-03	8.6E-03	8.6E-03	80	21.71428	4.9E-03	8.1E-03	8.1E-03
90	13.26903	1.3E-03	-1.4E-03	-1.4E-03	90	13.32877	6.1E-04	-2.0E-03	-2.0E-03
100	6.84853	-3.0E-03	-1.4E-02	-1.4E-02	100	6.85365	-3.2E-03	-1.5E-02	-1.5E-02
110	2.92962	-1.2E-03	-1.4E-02	-1.4E-02	110	2.86119	-4.6E-04	-1.3E-02	-1.3E-02
120	1.04072	3.7E-03	8.0E-04	8.2E-04	120	0.95542	4.7E-03	2.6E-03	2.6E-03
		VG					VG		
12	5.51930	-1.9E-03	-1.3E-02		50	6.45410	-9.7E-04	-1.1E-02	
12	5.77730	2.4E-03	-8.1E-03		50	6.21460	-3.0E-03	-1.3E-02	
12	6.38730	2.2E-03	-7.1E-03		250	5.59750	-5.2E-03	-1.6E-02	
12	6.15910	-8.4E-03	-1.9E-02		250	5.86060	-3.2E-03	-1.4E-02	
50	5.57400	-5.2E-03	-1.6E-02		250	6.47520	-4.7E-03	-1.4E-02	
50	5.84550	-3.2E-03	-1.4E-02		250	6.22460	3.6E-03	-6.9E-03	

Table 2 Pricing discrete Asian options in the CIR, CEV and VG models via the CTMC approximation based on finite state space with number of states $p=50$ (as in Cai et al. 2015 and Cui et al. 2018) using our moment-based method and the methods from the aforementioned papers. K denotes the strike price and m the number of monitoring dates. "Error" columns report the differences of the indicated method with respect to the benchmark. CIR parameters: (Cai et al. 2015, Table 3); benchmark: Fusai et al. (2008). CEV parameters: (Cai et al. 2015, Table 4); benchmark: Cai et al. (2014). VG parameters: (Cai et al. 2015, Tables 7–8); benchmark: MC price estimates. All MC estimates are based on 10^6 simulation trials and are from Cai et al. (2015) with standard errors reported there.

ments when computing the derivatives of the exponential of matrices of large dimension, which are particularly relevant in such problems.

Acknowledgments

We thank the Area co-Editor, Steven Kou, for his feedback and for giving us the opportunity to respond to our reviewers' comments and revise our paper accordingly. We are grateful to an anonymous Associate Editor and two anonymous reviewers for their fruitful comments. The first author's research is supported by the Ministry of Science and Technology (MOST 108-2811-M-001-628, MOST 108-2118-M-001-003-MY2) and the Academia Sinica of the Republic of China. The second author's research is supported by the Academia

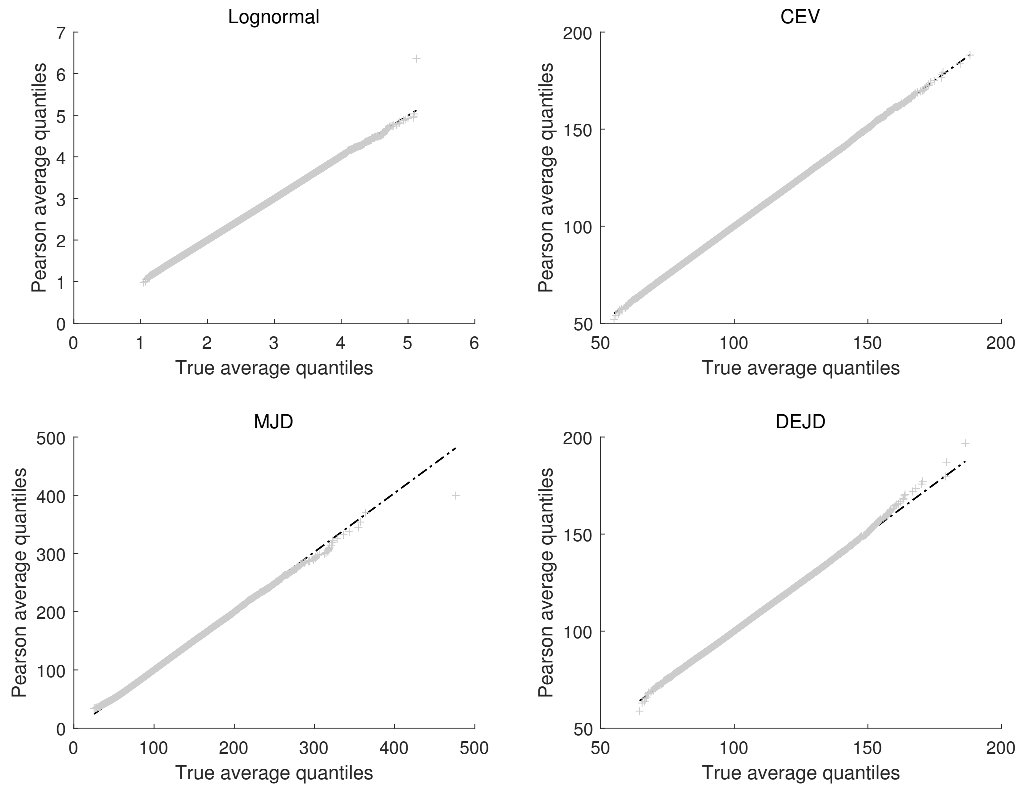

Figure 1 Plot of the quantiles of the sample of averages drawn based on the true underlying asset price process (lognormal, CEV, exponential MJD, exponential DEJD) versus the quantiles of the sample of averages drawn from the fitted Pearson distribution.

Sinica and the Ministry of Science and Technology (MOST 108-2118-M-001-003-MY2). The usual disclaimer applies.

References

Abate J, Choudhury GL, Lucantoni DM, Whitt W (1995) Asymptotic analysis of tail probabilities based on the computation of moments. *The Annals of Applied Probability* 5(4):983–1007.

Akhiezer NI (1965) *The Classical Moment Problem and Some Related Questions in Analysis*. University Mathematical Monographs (Oliver & Boyd).

Baumann B, Merkle CW, Leitgeb RA, Augustin M, Wartak A, Pircher M, Hitzenberger CK (2019) Signal averaging improves signal-to-noise in OCT images: But which approach works best, and when? *Biomedical Optics Express* 10(11):5755–5775.

Bertsimas D, Popescu I (2002) On the relation between option and stock prices: A convex optimization approach. *Operations Research* 50(2):358–374.

Bertsimas D, Popescu I (2005) Optimal inequalities in probability theory: A convex optimization approach. *SIAM Journal on Optimization* 15(3):780–804.

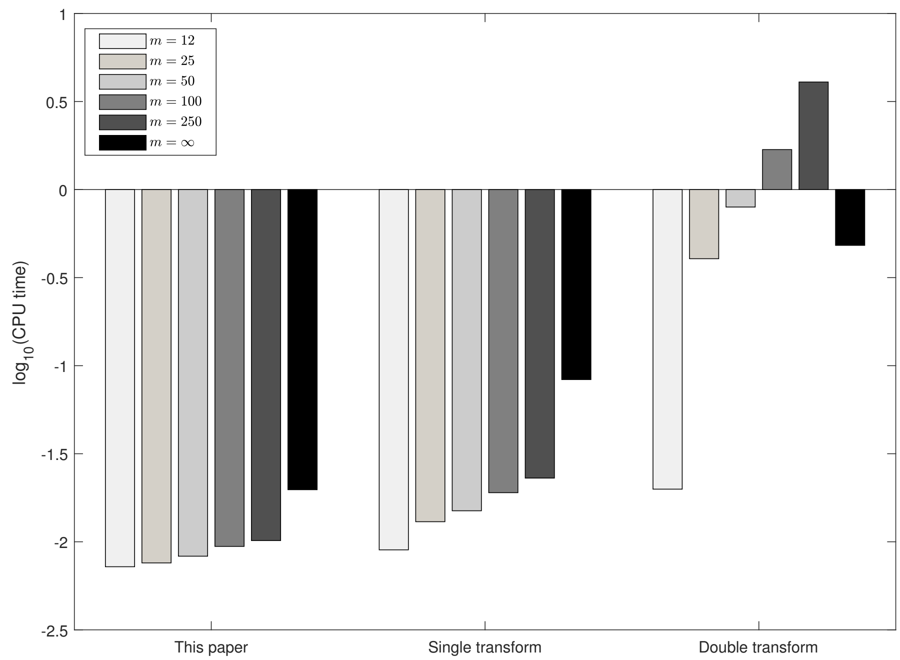

Figure 2 Comparing CPU times of the different methods. The bars are based on average times across models for given monitoring frequencies.

Bertsimas D, Sethuraman J (2000) Moment problems and semidefinite optimization. Wolkowicz H, Saigal R, Vandenberghe L, eds., *Handbook of Semidefinite Programming*, volume 27 of *International Series in Operations Research & Management Science*, 469–509 (Boston, MA: Springer).

Brignone R, Kyriakou I, Fusai G (2021) Moment-matching approximations for stochastic sums in non-Gaussian Ornstein–Uhlenbeck models. *Insurance: Mathematics and Economics* 96:232–247.

Cai N, Kou S (2012) Pricing Asian options under a hyper-exponential jump diffusion model. *Operations Research* 60(1):64–77.

Cai N, Kou S, Song Y (2020) A unified framework for regime-switching models, working paper.

Cai N, Li C, Shi C (2014) Closed-form expansions of discretely monitored Asian options in diffusion models. *Mathematics of Operations Research* 39(3):789–822.

Cai N, Song Y, Kou S (2015) A general framework for pricing Asian options under Markov processes. *Operations Research* 63(3):540–554.

Cai N, Yang X (2018) International reserve management: A drift-switching reflected jump-diffusion model. *Mathematical Finance* 28(1):409–446.

Cai N, Yang X (2021) A computational approach to first passage problems of reflected hyperexponential jump diffusion processes. *INFORMS Journal on Computing* 33(1):216–229.

Chen X, He S, Jiang B, Ryan CT, Zhang T (2021) The discrete moment problem with nonconvex shape constraints. *Operations Research* 69(1):279–296.

Choudhury GL, Lucantoni DM (1996) Numerical computation of the moments of a probability distribution from its transform. *Operations Research* 44(2):368–381.

Creemers S (2018) Moments and distribution of the net present value of a serial project. *European Journal of Operational Research* 267(3):835–848.

Cui Z, Kirkby JL, Nguyen D (2017) A general framework for discretely sampled realized variance derivatives in stochastic volatility models with jumps. *European Journal of Operational Research* 262(1):381–400.

Cui Z, Kirkby JL, Nguyen D (2021) Efficient simulation of generalized SABR and stochastic local volatility models based on Markov chain approximations. *European Journal of Operational Research* 290(3):1046–1062.

Cui Z, Lee C, Liu Y (2018) Single-transform formulas for pricing Asian options in a general approximation framework under Markov processes. *European Journal of Operational Research* 266(3):1134–1139.

Ding K, Cui Z, Wang Y (2021) A Markov chain approximation scheme for option pricing under skew diffusions. *Quantitative Finance* 21(3):461–480.

Fusai G, Kyriakou I (2016) General optimized lower and upper bounds for discrete and continuous arithmetic Asian options. *Mathematics of Operations Research* 41(2):531–559.

Fusai G, Marena M, Roncoroni A (2008) Analytical pricing of discretely monitored Asian-style options: Theory and application to commodity markets. *Journal of Banking & Finance* 32(10):2033–2045.

Gambaro AM, Kyriakou I, Fusai G (2020) General lattice methods for arithmetic Asian options. *European Journal of Operational Research* 282(3):1185–1199.

He S, Zhang J, Zhang S (2010) Bounding probability of small deviation: A fourth moment approach. *Mathematics of Operations Research* 35(1):208–232.

Horn RA, Johnson CR (2012) *Matrix Analysis* (New York: Cambridge University Press), ISBN 9781139788885.

Kalbfleisch JD, Lawless JF (1985) The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association* 80(392):863–871.

Khamis SH (1954) On the reduced moment problem. *The Annals of Mathematical Statistics* 25(1):113–122.

Kirkby JL, Nguyen D, Cui Z (2017) A unified approach to Bermudan and barrier options under stochastic volatility models with jumps. *Journal of Economic Dynamics and Control* 80:75–100.

Kyriakou I, Brignone R, Fusai G (2021) Unified moment-based modelling of integrated stochastic processes, working paper.

Lindsay BG, Basak P (2000) Moments determine the tail of a distribution (but not much else). *The American Statistician* 54(4):248–251.

Lo AW (1987) Semi-parametric upper bounds for option prices and expected payoffs. *Journal of Financial Economics* 19(2):373–387.

Magnus JR, Neudecker H (2019) *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley Series in Probability and Statistics (Chichester: John Wiley & Sons, Ltd.), 3 edition.

Mijatović A, Pistorius M (2013) Continuously monitored barrier options under Markov processes. *Mathematical Finance* 23(1):1–38.

Moler C, Van Loan C (2003) Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Review* 45(1):3–49.

Nadarajah S (2008) A review of results on sums of random variables. *Acta Applicandae Mathematicae* 103(2):131–140.

Shampine LF (2008) Vectorized adaptive quadrature in MATLAB. *Journal of Computational and Applied Mathematics* 211(2):131–140.

Shampine LF, Allen, Jr RC, Pruess S (1997) *Fundamentals of Numerical Computing* (New York: John Wiley & Sons, Inc.).

Sidje RB (1998) Expokit: A software package for computing matrix exponentials. *ACM Transactions on Mathematical Software (TOMS)* 24(1):130–156.

Solomon H, Stephens MA (1978) Approximations to density functions using Pearson curves. *Journal of the American Statistical Association* 73:153–160.

Song Y, Cai N, Kou S (2018) Computable error bounds of Laplace inversion for pricing Asian options. *INFORMS Journal on Computing* 30(4):625–786.

Tian R, Cox SH, Zuluaga LF (2017) Moment problem and its applications to risk assessment. *North American Actuarial Journal* 21(2):242–266.

Tsai H, Chan KS (2003) A note on parameter differentiation of matrix exponentials, with applications to continuous-time modelling. *Bernoulli* 9(5):895–919.

Yang W, Ma J, Cui Z (2021) Analysis of Markov chain approximation for Asian options and occupation-time derivatives: Greeks and convergence rates. *Mathematical Methods of Operations Research* 93(2):359–412.

Zhang G, Li L (2021) A general approach for Parisian stopping times under Markov processes, working paper.

Proofs and Algorithms

EC.1. Proofs of Auxiliary Results

Proof of Proposition 2. Note that for any complex number θ, $E_d(0,\theta) = e^{-\theta D}$. Thus, we prove our claim $E_d^{(n)}(0,\theta) = (-1)^n D^n e^{-\theta D}$ for any non-negative integer n. It is easy to see that, for $m = 0, 1, 2, \ldots$,

$$E_d(m+1,\theta) = e^{-\theta D} P(\Delta) E_d(m,\theta); \tag{EC.1}$$

this proves the result for $n = 0$. Then, using the general Leibniz rule on the recurrence relation (EC.1), we get our result. $\square$

Proof of Proposition 3. We have that $E_c(t,\theta)$ satisfies the ordinary differential equation

$$\frac{\partial}{\partial t} E_c(t,\theta) = (G - \theta D) E_c(t,\theta). \tag{EC.2}$$

We also see that $E_c(t,\theta)$ is infinitely differentiable with respect to θ. Therefore, differentiating (EC.2) with respect to θ yields

$$\frac{\partial^2}{\partial t \partial \theta} E_c(t,\theta) = -D E_c(t,\theta) + (G - \theta D) \frac{\partial}{\partial \theta} E_c(t,\theta), \tag{EC.3}$$

where we have used the fact that $\frac{\partial}{\partial \theta}(G - \theta D) = -D$. Note that equation (EC.3) can be rewritten as

$$\frac{\partial}{\partial t} E_c^{(1)}(t,\theta) = -D E_c(t,\theta) + (G - \theta D) E_c^{(1)}(t,\theta).$$

Therefore, for $n = 0, 1, 2, \ldots$, it follows trivially that the $(n+1)$-th order derivative satisfies the recurrence equation

$$\frac{\partial}{\partial t} E_c^{(n+1)}(t,\theta) = -(n+1) D E_c^{(n)}(t,\theta) + (G - \theta D) E_c^{(n+1)}(t,\theta). \tag{EC.4}$$

It is also easy to see that, for any $\theta \in \mathbb{C}$ and non-negative integer n,

$$E_c(0,\theta) = I_p, \quad E_c^{(n+1)}(0,\theta) = \mathbf{0}_{p,p}, \tag{EC.5}$$

where $\mathbf{0}_{p,p}$ denotes the $p \times p$ zero matrix and I_p the $p \times p$ identity matrix in $M_p(\mathbb{C})$. Using the initial condition (EC.5), it follows that the solution to (EC.4) has integral representation

$$E_c^{(n+1)}(t,\theta) = -(n+1) \int_0^t E_c(t-u,\theta) D E_c^{(n)}(u,\theta)\, du,$$

which completes the proof. $\square$

Proof of Proposition 4. 1. This holds trivially by definition.

2. We only prove for $n=1$ as the rest of the result can be shown based on similar arguments. Let $i,j,k\in\{1,\dots,p\}$. In view of (5), we consider the following cases. First, let $j\neq k$. Then, from the first part of the proposition,

$$\Gamma^i_{kj}(t,\theta;1)=\mathcal{L}_{d_i}\Gamma_{kj}(t,\theta)=\mathcal{L}_{d_i}\left(\frac{e^{td_j}-e^{td_k}}{d_j-d_k}\right).$$

By linearity of $\mathcal{L}_{d_i}$ and from (3), we get that

$$\Gamma^i_{kj}(t,\theta;1)=\frac{\mathcal{L}_{d_i}e^{d_jt}-\mathcal{L}_{d_i}e^{d_kt}}{d_j-d_k}=\frac{\Gamma_{ij}(t,\theta)-\Gamma_{ik}(t,\theta)}{d_j-d_k}.$$

Second, consider the case $i\neq j=k$. Integration by parts yields

$$\Gamma^i_{jj}(t,\theta;1)=\mathcal{L}_{d_i}\Gamma_{jj}(t,\theta)=\int_0^t e^{(t-u)d_i}ue^{ud_j}du=\frac{te^{d_jt}}{d_j-d_i}-\frac{1}{d_j-d_i}\mathcal{L}_{d_i}e^{d_jt}=\frac{\Gamma_{jj}(t,\theta)}{d_j-d_i}-\frac{\Gamma_{ij}(t,\theta)}{d_j-d_i}.$$

Finally, for $i=j=k$, we have that

$$\Gamma^i_{ii}(t,\theta;1)=\mathcal{L}_{d_i}\Gamma_{ii}(t,\theta)=\int_0^t e^{(t-u)d_i}ue^{ud_i}du=\frac{t^2}{2}e^{td_i}.$$

□

Proof of Lemma 1. Upon post-multiplication of $\mathbf{x}$ to the differential equation (EC.2) and the equations (EC.4)–(EC.5), we deduce that

$$\begin{aligned}&\frac{\partial}{\partial t}\{E_c(t,\theta)\mathbf{x}\}=(G-\theta D)\{E_c(t,\theta)\mathbf{x}\},\\&\frac{\partial}{\partial t}\left(\frac{d^{n+1}}{d\theta^{n+1}}\{E_c(t,\theta)\mathbf{x}\}\right)=-(n+1)D\frac{d^n}{d\theta^n}\{E_c(t,\theta)\mathbf{x}\}+(G-\theta D)\frac{d^{n+1}}{d\theta^{n+1}}\{E_c(t,\theta)\mathbf{x}\},\end{aligned}\tag{EC.6}$$

with the condition

$$E_c(0,\theta)\mathbf{x}=\mathbf{x},\quad E_c^{(n+1)}(0,\theta)\mathbf{x}=\mathbf{0}_{p,1}.\tag{EC.7}$$

Rearrangement of (EC.6)–(EC.7) in the matrix form (6) gives the result. □

EC.2. Algorithms

EC.2.1. Algorithm: Computation of derivatives of matrix exponential with respect to parameter (Theorem 1)

Input: Matrices G,D and complex number θ with $G-\theta D$ containing distinct eigenvalues

Output: First four derivatives $E_c^{(1)},E_c^{(2)},E_c^{(3)},E_c^{(4)}$ of $e^{G-\theta D}$ with respect to θ

1. Diagonalize $G-\theta D$: $G-\theta D=Q(\theta)M(\theta)Q^{-1}(\theta)$
2. Let $\{d_1,\dots,d_p\}$ be the eigenvalues of $G-\theta D$
3. $L\leftarrow Q^{-1}DQ$
4. **Initialize** $\Gamma_0\leftarrow p\times p$ zero matrix, and $\tilde{\Gamma}^{(1)}\leftarrow p\times p$ zero matrix

for $i, j = 1, 2, \ldots, p$

  if $i \neq j$, $\Gamma_0(i,j) \leftarrow \frac{e^{td_j} - e^{td_i}}{d_j - d_i}$

  else $\Gamma_0(i,j) \leftarrow te^{td_i}$

for $i, j = 1, 2, \ldots, p$

  $\tilde{\Gamma}_0(i,j) \leftarrow \Gamma_0(i,j) L(i,j)$

$E_c^{(1)} \leftarrow -Q\tilde{\Gamma}^{(1)}Q^{-1}$ // first order derivative

Initialize $\Gamma_1 \leftarrow p \times p \times p$ zero matrix, and $\tilde{\Gamma}^{(2)} \leftarrow p \times p$ zero matrix

for $i, j, k = 1, 2, \ldots, p$

  if $k \neq j$, $\Gamma_1(i,j,k) \leftarrow \frac{\Gamma_0(i,j) - \Gamma_0(i,k)}{d_j - d_k}$

  else if $i \neq k = j$, $\Gamma_1(i,k,j) \leftarrow \frac{\Gamma_0(j,j) - \Gamma_0(i,j)}{d_j - d_i}$

  else $\Gamma_1(i,k,j) \leftarrow \frac{t}{2}\Gamma_0(j,j)$

for $i, j = 1, 2, \ldots, p$

  $\tilde{\Gamma}^{(2)}(i,j) \leftarrow \sum_k \Gamma_1(i,j,k) L(i,k) L(k,j)$

$E_c^{(2)} \leftarrow 2\,Q\tilde{\Gamma}^{(2)}Q^{-1}$ // second order derivative

Initialize $\Gamma_2 \leftarrow p \times p \times p \times p$ zero matrix, and $\tilde{\Gamma}^{(3)} \leftarrow p \times p$ zero matrix

for $i, j, k, l = 1, 2, \ldots, p$

  if $k \neq j$, $\Gamma_2(i,l,k,j) \leftarrow \frac{\Gamma_1(i,l,j) - \Gamma_1(i,l,k)}{d_j - d_k}$

  else if $l \neq k = j$, $\Gamma_2(i,l,k,j) \leftarrow \frac{\Gamma_1(i,j,j) - \Gamma_1(i,l,j)}{d_j - d_l}$

  else if $i \neq l = k = j$, $\Gamma_2(i,l,k,j) \leftarrow \frac{\Gamma_1(j,j,j) - \Gamma_1(i,i,j)}{d_j - d_i}$

  else $\Gamma_2(i,l,k,j) \leftarrow \frac{t}{3}\Gamma_1(j,j,j)$

for $i, j = 1, 2, \ldots, p$

  $\tilde{\Gamma}^{(3)}(i,j) \leftarrow \sum_{l,k} \Gamma_2(i,l,k,j) L(i,l) L(l,k) L(k,j)$

$E_c^{(3)} \leftarrow -6\,Q\tilde{\Gamma}^{(3)}Q^{-1}$ // third order derivative

Initialize $\Gamma_3 \leftarrow p \times p \times p \times p \times p$ zero matrix, and $\tilde{\Gamma}^{(4)} \leftarrow p \times p$ zero matrix

for $i, j, k, l, m = 1, 2, \ldots, p$

  if $k \neq j$, $\Gamma_3(i,l,m,k,j) \leftarrow \frac{\Gamma_2(i,l,m,j) - \Gamma_2(i,l,m,k)}{d_j - d_k}$

  else if $m \neq k = j$, $\Gamma_3(i,l,m,k,j) \leftarrow \frac{\Gamma_2(i,l,j,j) - \Gamma_2(i,l,m,j)}{d_j - d_m}$

  else if $l \neq m = k = j$, $\Gamma_3(i,l,m,k,j) \leftarrow \frac{\Gamma_2(i,l,j,j) - \Gamma_2(i,l,l,j)}{d_j - d_l}$

  else if $i \neq l = m = k = j$, $\Gamma_3(i,l,m,k,j) \leftarrow \frac{\Gamma_2(i,j,j,j) - \Gamma_2(i,j,i,j)}{d_j - d_i}$

  else $\Gamma_3(i,l,m,k,j) \leftarrow \frac{t}{4}\Gamma_2(j,j,j,j)$

for $i, j = 1, 2, \ldots, p$

  $\tilde{\Gamma}^{(4)}(i,j) \leftarrow \sum_{l,k,m} \Gamma_3(i,l,m,k,j) L(i,l) L(l,m) L(m,k) L(k,j)$

$E_c^{(4)} \leftarrow 24\,Q\tilde{\Gamma}^{(4)}Q^{-1}$ // fourth order derivative

EC.2.2. Algorithm: Computation of derivatives of matrix exponential times a vector with respect to parameter (Theorem 2)

Input: Matrices G, D and complex number θ

Output: First four derivatives $\mathcal{E}_c^{(1)}, \mathcal{E}_c^{(2)}, \mathcal{E}_c^{(3)}, \mathcal{E}_c^{(4)}$ of $e^{t(G-\theta D)}\mathbf{1}$ with respect to θ

1. **Initialize** $B_5 \leftarrow 5\times 5$ zero matrix, $\mathbf{e}_1 \leftarrow (1,0,0,0,0)^*$, $I_5 \leftarrow 5\times 5$ identity matrix, $\mathbf{1} \leftarrow$ column vector of ones of size p
2. for $i, j = 1, 2, \ldots, 5$
3. if $j = i+1$, $B_5(i,j) \leftarrow i$
4. $\Upsilon \leftarrow I_5 \otimes (G - \theta D)$ // $\otimes$ denotes Kronecker product
5. $\Theta \leftarrow B_5^* \otimes D$
6. $\zeta \leftarrow \mathbf{e}_1 \otimes \mathbf{1}$
7. $\mathcal{E}_c^{(0,\ldots,5)} \leftarrow e^{\Upsilon - \Theta}\zeta$
8. $\mathcal{E}_c^{(1)} \leftarrow \mathcal{E}_c^{(0,\ldots,5)}(p+1,\ldots,2p)$, $\mathcal{E}_c^{(2)} \leftarrow \mathcal{E}_c^{(0,\ldots,5)}(2p+1,\ldots,3p)$, $\mathcal{E}_c^{(3)} \leftarrow \mathcal{E}_c^{(0,\ldots,5)}(3p+1,\ldots,4p)$, $\mathcal{E}_c^{(4)} \leftarrow \mathcal{E}_c^{(0,\ldots,5)}(4p+1,\ldots,5p)$