High order low variance matrix-exponential distributions*

Gábor Horváth
Dept. of Networked Systems and Services
Budapest University of Technology and Economics
Magyar Tudósok Körútja 2,
1117 Budapest, Hungary
ghorvath@hit.bme.hu

Illés Horváth
MTA-BME Information Systems Research Group
Magyar Tudósok Körútja 2,
1117 Budapest, Hungary
pollux@math.bme.hu

Miklós Telek
Dept. of Networked Systems and Services
Budapest University of Technology and Economics
Magyar Tudósok Körútja 2,
1117 Budapest, Hungary
telek@hit.bme.hu

ABSTRACT

This paper presents matrix-exponential (ME) distributions, whose squared coefficient of variation (SCV) is very low. Currently there is no symbolic construction available to obtain the most concentrated ME distributions, and the numerical optimization-based approaches to construct them have many pitfalls too. We present a numerical optimization-based procedure which avoids numerical issues.

Keywords: Non-negative matrix-exponential functions, Matrix-exponential distributions, Numerical optimization, Coefficient of variation

1. INTRODUCTION

Highly concentrated matrix exponential functions play an important role in many research fields, for example, they turned out to be essential for numerical inverse Laplace transform methods as well [6].

The least varying phase type (PH) distribution of order N is known to be the Erlang distribution [1] with SCV=$1/N$ (defined as $\frac{\mu_0\mu_2}{\mu_1^2}-1$, where $\mu_i, i=0,1,2$, are the moments of the distribution). The least varying ME distribution for order N much less known. It is known that for order 2 the class of ME distributions is identical to the class of PH distributions, and it is also known that there exists order 3 ME distribution with SCV=$0.200902 < 1/3$, but it is still only a conjecture that this is the least varying order 3 ME distribution. Concentrated ME distributions are provided in [2] up to order 17 and in [5] up to order 47. These preliminary results indicate that the minimal SCV of order N ME distributions tends to be less than $2/N^2$. In this work, we propose numerical procedures by which much higher order concentrated ME distributions can be computed and based on that we refine the dependence of the minimal SCV on the order.

*This work is supported by the OTKA K-123914 project

MAM10 2019, Hobart, Australia

2. CONCENTRATED ME DISTRIBUTIONS

DEFINITION 1. *Order N* ME functions *(referred to as ME(N)) are given by*

$$f(t) = \underline{\alpha} e^{\mathbf{A}t}(-\mathbf{A})\mathbb{1}, \tag{1}$$

where $\underline{\alpha}$ is a real row vector of size N, $\mathbf{A}$ is a real matrix of size $N \times N$ and $\mathbb{1}$ is the column vector of ones of size N, and $\underline{\alpha}$ is such that $\underline{\alpha}\mathbb{1} > 0$.

If $f(t) \geq 0, \forall t \geq 0$ and $\underline{\alpha}\mathbb{1} = 1$ then $f(t)$ is the probability density function of a ME distribution.

According to (1), vector $\underline{\alpha}$ and matrix $\mathbf{A}$ define the matrix exponential function. We refer to the pair $(\underline{\alpha}, \mathbf{A})$ as *matrix representation* in the sequel.

An ME distribution is said to be concentrated when its squared coefficient of variation

$$SCV(f(t)) = \frac{\mu_0\mu_2}{\mu_1^2} - 1, \tag{2}$$

is low. In (2), μ_i denotes the ith moment, defined by $\mu_i = \int_{t=0}^{\infty} t^i f(t)dt$ for $i = 0, 1, 2$. SCV is insensitive to multiplication and scaling, i.e. $SCV(f(t)) = SCV(c_1 f(c_2 t))$.

Although matrix-exponential functions have been used for many decades, there are still many questions open regarding their properties. Such an important question is how to decide efficiently if a matrix-exponential function is non-negative $\forall t > 0$. In general, $f(t) \geq 0, \forall t > 0$ does not necessarily hold for given $(\underline{\alpha}, \mathbf{A})$ parameters, unless it has been constructed to be always non-negative. In this paper, we are going to restrict our attention to such a special construction, the exponential-cosine square functions.

For the least varying ME(N) distributions only conjectures are available for $N \geq 3$ [2]. According to the current conjecture for odd N, the most concentrated ME(N) distribution belongs to a special subset of ME(N) given by the definition below.

DEFINITION 2. *The set of* exponential cosine-square *functions of order n has the form*

$$f^+(t) = e^{-t} \prod_{i=1}^{n} \cos^2\left(\frac{\omega t - \phi_i}{2}\right). \tag{3}$$

An exponential cosine-square function is defined by $n+1$ parameters: ω and ϕ_i for $i = 1, \ldots, n$. An exponential cosine-square function is a matrix exponential function. Although

the representation in (3), which we refer to as the *cosine-square representation*, is not a matrix representation, [5, Appendix A] presents the associated matrix representation of size $N = 2n+1$. Consequently, the set of exponential cosine-square functions of order n is a special subset of ME(N) (where $N = 2n+1$) which, by construction, is non-negative. The SCV of an exponential cosine-square function is a complicated function of the parameters, whose minimum does not exhibit a closed analytic form. That is why we have resorted to the following numerical problem. For a given odd order $N = 2n+1$, we are looking for efficient numerical methods for finding the ω and ϕ_i ($i = 1, \ldots, n$) parameters which result in a low SCV. For efficient numerical minimization of the SCV for $N > 47$ (i.e., $n > 23$) we need

i) an accurate computation of the SCV based on the parameters with low computational cost and

ii) an efficient optimization procedure with low computational cost.

In this paper we present a method that addresses i) in Section 3, and one that addresses ii) in Section 4.

3. EFFICIENT COMPUTATION OF THE SQUARED COEFFICIENT OF VARIATION

To evaluate the objective function of the optimization, namely the SCV, we need efficient methods to compute μ_0, μ_1 and μ_2. Deriving the μ_i parameters based on (3) is difficult (for large N). Hence we propose to compute them based on a different representation.

3.1 The hyper-trigonometric representation

The following theorem defines the *hyper-trigonometric form* of the exponential cosine-square functions and provides a recursive procedure to obtain its parameters from $\omega, \phi_i, i = 1, \ldots, n$.

THEOREM 1. *An order $N = 2n+1$ exponential cosine-square function can be transformed to a hyper-trigonometric representation of form*

$$f^+(t) = c^{(n)} \cdot e^{-t} + e^{-t} \sum_{k=1}^{n} a_k^{(n)} \cos(k\omega t) + e^{-t} \sum_{k=1}^{n} b_k^{(n)} \sin(k\omega t), \tag{4}$$

where $c^{(n)} = \frac{1}{2} a_0^{(n)}$ and the coefficients $a_k^{(n)}$, $b_k^{(n)}$ are calculated recursively:

- *for $n = 1$:*

$$a_0^{(1)} = 1, \quad b_0^{(1)} = 0, \quad a_1^{(1)} = \frac{1}{2}\cos\phi_1, \quad b_1^{(1)} = \frac{1}{2}\sin\phi_1, \tag{5}$$

- *for $k > n, n \geq 1$:*

$$a_k^{(n)} = b_k^{(n)} = 0,$$

- *for $k = 0, n \geq 1$:*

$$a_0^{(n)} = \frac{1}{2} a_0^{(n-1)} + \frac{1}{2} a_1^{(n-1)} \cos\phi_n + \frac{1}{2} b_1^{(n-1)} \sin\phi_n, \tag{6}$$

$$b_0^{(n)} = 0, \tag{7}$$

- *for $1 \leq k \leq n, n \geq 2$*

$$a_k^{(n)} = \frac{1}{2} a_k^{(n-1)} + \frac{1}{2} \frac{a_{k-1}^{(n-1)} + a_{k+1}^{(n-1)}}{2} \cos\phi_n + \frac{1}{2} \frac{b_{k+1}^{(n-1)} - b_{k-1}^{(n-1)}}{2} \sin\phi_n, \tag{8}$$

$$b_k^{(n)} = \frac{1}{2} b_k^{(n-1)} + \frac{1}{2} \frac{b_{k-1}^{(n-1)} + b_{k+1}^{(n-1)}}{2} \cos\phi_n + \frac{1}{2} \frac{a_{k-1}^{(n-1)} - a_{k+1}^{(n-1)}}{2} \sin\phi_n. \tag{9}$$

The hyper-trigonometric representation makes it possible to express the Laplace transform (LT) and the moments μ_i in a simple and compact way.

COROLLARY 1. *The LT and the $\mu_i, i = 0, 1, 2$ moments of the exponential cosine-square function are given by*

$$f^*(s) = \frac{c^{(n)}}{1+s} + \sum_{k=1}^{n} \frac{a_k^{(n)}(1+s) + b_k^{(n)} k\omega}{(1+s)^2 + (k\omega)^2}, \tag{10}$$

and

$$\mu_0 = c^{(n)} + \sum_{k=1}^{n} \frac{a_k^{(n)} + b_k^{(n)} k\omega}{1 + (k\omega)^2}, \tag{11}$$

$$\mu_1 = c^{(n)} + \sum_{k=1}^{n} \frac{a_k^{(n)} + 2b_k^{(n)} k\omega - a_k^{(n)}(k\omega)^2}{(1+(k\omega)^2)^2},$$

$$\mu_2 = 2c^{(n)} + \sum_{k=1}^{n} \frac{2a_k^{(n)} + 6b_k^{(n)} k\omega - 6a_k^{(n)}(k\omega)^2 - 2b_k^{(n)}(k\omega)^3}{(1+(k\omega)^2)^3}.$$

3.2 Numerical computation of the moments

Theorem 1 together with Corollary 1 provides a very efficient explicit method to compute the SCV based on the parameters $\omega, \phi_i, i = 1, \ldots, n$.

There is one numerical issue that has to be taken care of when applying this numerical procedure with floating point arithmetic for large values of n. To evaluate the SCV, coefficients $a_k^{(n)}, b_k^{(n)}, c^{(n)}$ need to be obtained from the ω and $\phi_i, i = 1, \ldots, n$ parameters. The recursion defined in Theorem 1 involves multiplications between bounded numbers (sine and cosine always fall into $[-1, +1]$), which is beneficial from the numerical stability point of view, but subtractions are unfortunately also present, leading to loss of precision. To overcome this loss of precision, we introduced increased precision floating point arithmetic both in our Mathematica and C++ implementations[1]. Mathematica can quantify the precision loss, enabling us to investigate this issue experimentally. According to Figure 1, the number of accurate decimal digits lost when evaluating the SCV from the ω, ϕ_i parameters (computed by the `Precision` function of Mathematica), denoted by L_n, is nearly linear and can be approximated by

$$L_n \approx 1.487 + 0.647n. \tag{12}$$

[1]In C++ we used to `mpfr` library for multi-precision computations

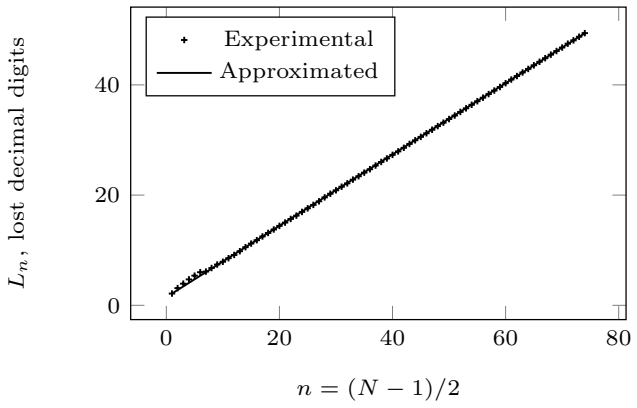

Figure 1: The precision loss while computing the SCV

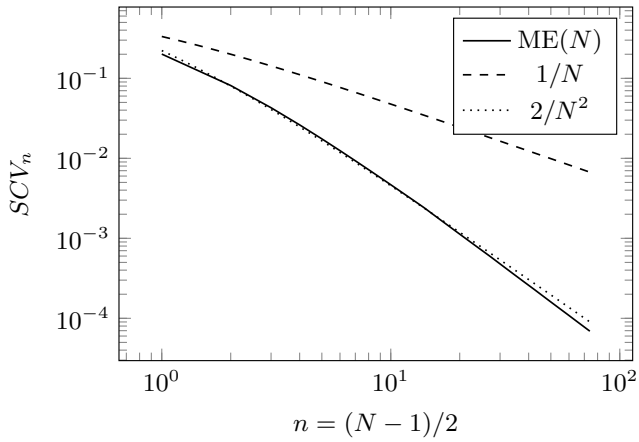

Figure 2: The minimal SCV of the exponential cosine-square functions as the function of n in log-log scale

In the forthcoming numerical experiments we have set the floating point precision to $L_n + 16$ decimal digits to obtain an accuracy of results up to 16 decimal digits, and this precision setting eliminated all numerical issues.

It is important to note that the high precision is needed only to calculate the $a_k^{(n)}, b_k^{(n)}, c^{(n)}$ coefficients and the SCV itself. Representing parameters ω, ϕ_i themselves does not need extra precision, and the resulting exponential cosine-square function $f(t)$ can be evaluated with machine precision as well (in the range of our interest, $n \leq 1000$).

A basic pseudo-code of the computation of the SCV with the indications where high precision is needed is provided by Algorithm 1.

Algorithm 1 Pseudo-code for the computation of the SCV

procedure COMPUTESCV($n, \omega, \phi_i, i = 1, \ldots, n$)
  Compute the required precision, L_n, from (12)
  Convert $\omega, \phi_i, i = 1, \ldots, n$ to $L_n + 16$ digits precision
  Calculate $a_k^{(n)}, b_k^{(n)}, c^{(n)}, k = 1, \ldots, n$, recursively by Theorem 1 (**high precision**)
  Calculate moments μ_0, μ_1, μ_2 according to (11) (**high precision**)
  Calculate $SCV = \frac{\mu_0 \mu_2}{\mu_1^2} - 1$ (**high precision**)
  Convert SCV to machine precision
  return SCV
end procedure

4. MINIMIZING THE SQUARED COEFFICIENT OF VARIATION

Given the size of the representation $N = 2n + 1$, the $f^+(t)$ function providing the minimal SCV is obtained by minimizing (2) subject to ω and $\phi_i, i = 1, \ldots, n$. The form of the SCV does not allow a symbolic solution, and its numerical optimization is challenging too. The surface to optimize has many local optima, hence simple gradient descent procedures failed to find the global optimum and are sensitive to the initial guess.

4.1 Optimizing the parameters

In the numerical optimization of the parameters, we had success with evolutionary optimization methods, in particular with *evolution strategies*. The results introduced in [5] were obtained by one of the simplest evolution strategies, the Rechenberg method [7]. In [5], it was the high computational demand of the numerical integration needed to obtain the SCV and its reduced accuracy that prevented the optimization for $N > 47$ ($n > 23$).

However, computing the SCV based on the hyper-trigonometric representation using the results of Section 3.1 allows us to evaluate the moments orders of magnitudes faster and more accurately, enabling the optimization for higher n values. With the Rechenberg method ([7], also referred to as (1+1)-ES in the literature) it is possible to obtain low SCV values relatively quickly for orders as high as $n = 125$, but these values are suboptimal in the majority of cases.

With more advanced evolution strategies the optimal SCV can be approached better. Our implementation supports the covariance matrix adoption evolution strategy (CMA-ES [3]), and one of its variants, the BIPOP-CMA-ES with restarts [4]. Starting from a random initial guess, we got very low SCV values much quicker with the CMA-ES than with the (1+1)-ES with similar suboptimal minimum values (cf. Fig. 4). The limit of applicability of CMA-ES is about $n = 180$. The best solution (lowest SCV for the given order), however, was always provided by the BIPOP-CMA-ES method, although it is by far the slowest among the three methods we studied. In fact, we believe that BIPOP-CMA-ES returned the global optimum for $n = 1, \ldots, 74$, and we investigate the properties of those solutions in the next sections. For $n > 74$, we can still compute low SCV functions with the BIPOP-CMA-ES method, but its computation time gets to be prohibitive, and we are less confident about the global minimality of the results.

For our particular problem, the running time, T, and the quality of the minimum, Q (how low the SCV is), obtained by the different optimization methods can be summarized as follows

$$T_{\text{CMA-ES}} < T_{(1+1)\text{-ES}} << T_{\text{BIPOP-CMA-ES}},$$
$$Q_{\text{CMA-ES}} \sim Q_{(1+1)\text{-ES}} < Q_{\text{BIPOP-CMA-ES}}.$$

4.2 Properties of the minimal SCV solutions

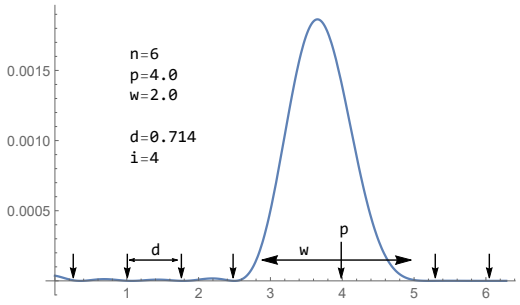

Figure 3: The spike and the zeros of $f^+(t)$

The minimal SCV values obtained by the BIPOP-CMA-ES optimization, which we consider as being optimal for $n = 1, \ldots, 74$, are depicted in Figure 2. Apart from the minimal SCV values of the exponential cosine-square functions, Figure 2 also plots $1/N$ and $2/N^2$, for comparison. The $SCV = 1/N$ is known to be the minimal SCV value of phase-type (PH) distributions of order N [1], which form a subset in the set of ME distributions by assuming positive off-diagonal elements for $\mathbf{A}$ and nonnegative elements for $\underline{\alpha}$. The $2/N^2$ curve is reported to be the approximate decay rate in [5], up to $n = 23$ ($N = 47$).

Figure 2 indicates that the SCV decreases much faster than $1/N$ and a bit faster than $2/N^2$. Indeed, $2/N^2$ is a good approximation up to $n = 23$, but the decay seems to decrease below $2/N^2$ for $n > 23$.

5. HEURISTIC OPTIMIZATION WITH 3 PARAMETERS

According to the previously discussed approach the number of parameters to optimize increases with n. This drawback limits the applicability of the general optimization procedures to about $n \leq 74$ in case of BIPOP-CMA-ES and about $n \leq 180$ in case of the basic CMA-ES. By these n values the optimization procedure takes several days to terminate on our average PC clocked at 3.4 GHz.

While the $f^+(t)$ function obtained this way for $n = 180$ have an extremely low ($\approx 10^{-5}$) SCV already, some applications might benefit from ME distributions with even lower SCV. To overcome this limitation we developed a suboptimal heuristic procedure, that aims to obtain low SCV for a given large order n.

Our heuristic procedure has to optimize only three parameters, independent of the order n. The procedure is based on the assumption that the location of the spike in the $(0, 2\pi)$ cycle of the cosine-squared function plays the most important role in the SCV, and the exact values of the ϕ_k parameters are less important, the only important feature is that the cosine-squared terms characterized by the ϕ_k parameters should suppress $f^+(t)$ uniformly in the $(0, 2\pi)$ cycle – apart from the spike (cf. Figure 3).

Based on this assumption we set the ϕ_k parameters of the cosine-squared terms equidistantly. This way the position of the spike (p) and its width (w) inside the $(0, 2\pi)$ interval completely define the ϕ_k values for a given order n.

The distance of the ϕ_k parameters (d) and the number of ϕ_k parameters before the spike (i) can be computed from p and w by

$$d = \frac{2\pi - w}{n}, \quad i = \left\lfloor \frac{p - w/2}{d} + \frac{1}{2} \right\rfloor, \tag{13}$$

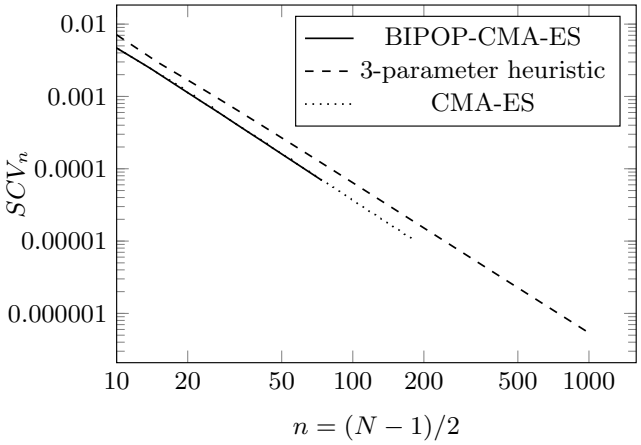

Figure 4: The minimal and the heuristic SCV as a function of order n in log-log scale

and for $k = 1, \ldots, n$ the ϕ_k parameters are

$$\phi_k + \pi = \begin{cases} (k - 1/2)d & \text{if } k \leq i, \\ (k - 1/2)d + w & \text{if } k > i. \end{cases} \tag{14}$$

Figure 4 depicts the SCV obtained by the heuristic procedure for large n values, compared with the outputs of the highly accurate BIPOP-CMA-ES and the faster CMA-ES optimization procedures. Figure 4 suggests that the heuristic optimizations remains very close to the minimum also for larger n values and the SCV obtained by the heuristic optimization maintains its polynomial decay between $n^{-2.1}$ and $n^{-2.2}$.

6. REFERENCES

[1] D. Aldous and L. Shepp. The least variable phase type distribution is Erlang. *Stochastic Models*, 3:467–473, 1987.

[2] T. Éltető, S. Rácz, and M. Telek. Minimal coefficient of variation of matrix exponential distributions. In *2nd Madrid Conference on Queueing Theory, Madrid, Spain (July 2006)*, 2006.

[3] N. Hansen. The cma evolution strategy: a comparing review. In *Towards a New Evolutionary Computation*, pages 75–102. Springer, 2006.

[4] N. Hansen. Benchmarking a bi-population cma-es on the bbob-2009 function testbed. In *Proceedings of the 11th Annual Conference Companion on Genetic and Evolutionary Computation Conference: Late Breaking Papers*, pages 2389–2396. ACM, 2009.

[5] I. Horváth, O. Sáfár, M. Telek, and B. Zámbó. Concentrated matrix exponential distributions. In *European Workshop on Performance Engineering*, pages 18–31. Springer, 2016.

[6] I. Horváth, Z. Talyigás, and M. Telek. An optimal inverse Laplace transform method without positive and negative overshoot – an integral based interpretation. *Electronic Notes in Theoretical Computer Science*, 337:87 – 104, 2018.

[7] I. Rechenberg. Evolutionsstrategien. In *Simulationsmethoden in der Medizin und Biologie*, pages 83–114. Springer, 1978.