Optimized numerical inverse Laplace transformation*

Illés Horváth[1], András Mészáros[2], Miklós Telek[1,2]
[1] MTA-BME Information Systems Research Group, Budapest, Hungary
[2] Department of Networked Systems and Services,
Budapest University of Technology and Economics, Budapest, Hungary
email: horvath.illes.antal@gmail.com, {meszarosa,telek}@hit.bme.hu

ABSTRACT

Among the numerical inverse Laplace transformation (NILT) methods, those that belong to the Abate–Whitt framework (AWF) are considered to be the most efficient ones currently. It is a characteristic feature of the AWF NILT procedures that they are independent of the transform function and the time point of interest.

In this work we propose an NILT procedure that goes beyond this limitation and optimize the accuracy of the NILT utilizing also the transform function and the time point of interest.

Keywords: numerical inverse Laplace transformation, shifting, Abate–Whitt framework, Euler method, CME method.

1. INTRODUCTION

Due to the widespread use of Laplace transforms in various scientific fields, a large number of numerical inverse Laplace transformation (NILT) methods have been developed. Recent surveys are available, e.g., in [3, 2].

Among these methods, the most efficient and widely applied ones belong to a subset which is referred to as Abate–Whitt framework (AWF) [1]. For a given order N, each method in the AWF uses a predefined set of η_k, β_k (potentially complex) coefficients independent of the transform function to invert ($h^*(s)$) and the time point of interest (T). Based on these parameters the AWF NILT procedure is

$$h(T) \approx h_N(T) = \sum_{k=0}^{N-1} \frac{\eta_k}{T} \ h^*\left(\frac{\beta_k}{T}\right), \tag{1}$$

where $h^*(s) = \int_{t=0}^{\infty} e^{-st} h(t) \mathrm{d}t$ is the Laplace transform function and $h_N(T)$ is the order N approximate of its inverse transform ($h(t)$) at point T. Different β_k and η_k parameters define different NILT methods of the AWF. The most efficient ones are the Euler method [1] and the CME method [2].

In this paper, we propose a generalization of the AWF such that the NILT method is optimized also for the given transform function to invert ($h^*(s)$) and for the time point of interest (T). This proposed approach is composed by the following elements:

- *a parametric set of AWF methods*
  When the η_k, β_k coefficients depend on a parameter θ, the $\eta_k(\theta)$, $\beta_k(\theta)$ coefficients define an NILT method of the AWF with NILT procedure
  $$h_N(T,\theta) = \sum_{k=0}^{N-1} \frac{\eta_k(\theta)}{T} \ h^*\left(\frac{\beta_k(\theta)}{T}\right), \tag{2}$$
  which is a function of parameter θ.
- *an error indicator* ($Err(h_N(T,\theta))$)
  A parameter computed by a numerical procedure that indicates the error of the approximation $h(T) \approx h_N(T,\theta)$ for a given $h^*(s)$ and T.
- *an optimization method*
  A method to find the optimal value of the parameter
  $$\hat{\theta} = \arg\min_{\theta} Err(h_N(T,\theta)).$$

This rather straight forward enhancement of the AWF methods was not considered up to now, because of the lack of reasonably accurate error measures associated with the available AWF methods.

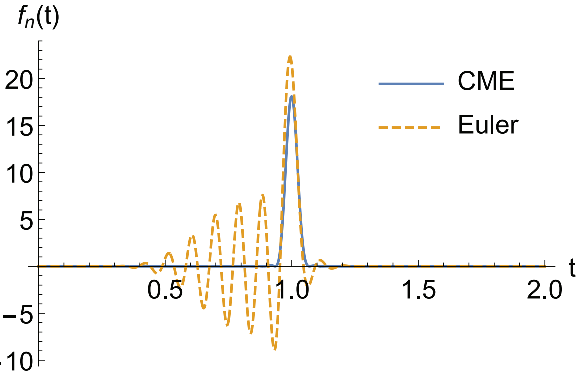
The recently introduced CME method, which is sign preserving [2], offers new opportunities for measuring the error of NILT for non-negative $h(t)$ functions. This assumption holds in many practical applications, e.g., when $h(t)$ represents an intrinsically nonnegative physical quantity like a probability or the level of fluid in a container. The framework can also be extended to lower bounded $h(t)$ functions with known lower bound $m = \min_{\tau\geq 0} h(\tau)$, since in this case $h(t)+m$ with Laplace transform $h^*(s)+m/s$ is a nonnegative function.

2. INVERSE LAPLACE TRANSFORMATION WITH THE AWF

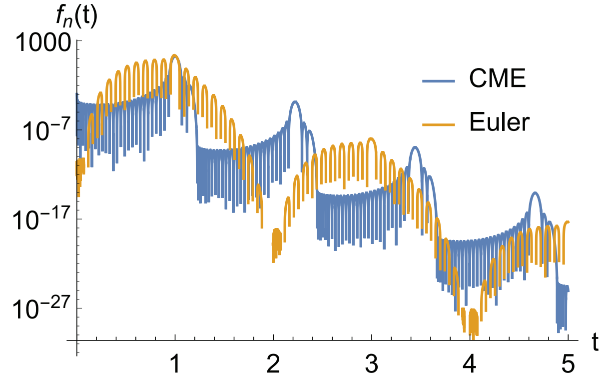
We build on the following *integral interpretation* of the AWF methods which is obtained from (1) by substituting definition of the Laplace transform:

$$\begin{aligned} h_N(T) &= \sum_{k=0}^{N-1} \frac{\eta_k}{T} h^*\left(\frac{\beta_k}{T}\right) = \sum_{k=0}^{N-1} \frac{\eta_k}{T} \int_0^\infty h(t) \cdot e^{-\beta_k t/T} \mathrm{d}t \\ &= \int_0^\infty h(t) \cdot \frac{1}{T} f_N(t/T) \mathrm{d}t = \int_0^\infty h(tT) \cdot f_N(t) \mathrm{d}t, \end{aligned} \tag{3}$$

*This work is partially supported by the OTKA K-138208.



a) linear y-axis



b) logarithmic y-axis

Figure 1: The $f_N(t)$ for the Euler and the CME methods for order 30 with linear and logarithmic y-axis. The negative parts of the Euler $f_N(t)$ are not visible with logarithmic scaling.

where

$$f_N(t) = \sum_{k=0}^{N-1} \eta_k e^{-\beta_k t}. \tag{4}$$

That is, the result of an AWF NILT procedure according to (1), is equivalent to the integral in (3), where $f_N(t)$ is a so called *weight function*, which approximates the unit impulse. If $f_N(t)$ was the unit impulse function at one (also referred to as Dirac function), then the integral in (3) would result in a perfect Laplace inversion. The weight functions of the Euler and the CME methods are depicted in Figure 1.

3. A PARAMETRIC SET OF ABATE–WHITT FRAMEWORK METHODS

Let η_k, β_k be the set of coefficients associated with an AWF method. Starting from this set of coefficients, we define

$$\eta_k(\theta) = e^{\theta}\eta_k, \quad \beta_k(\theta) = \beta_k + \theta. \tag{5}$$

as a function of parameter θ, which we refer to as the shifting parameter.

To gain an intuitive understanding on the effect of θ we write the associated weight function as

$$f_{N,\theta}(t) = \sum_{k=0}^{N-1} \eta_k(\theta) e^{-\beta_k(\theta)t} = \sum_{k=0}^{N-1} (e^{\theta}\eta_k) e^{-(\beta_k+\theta)t}$$
$$= e^{-\theta(t-1)} \sum_{k=0}^{N-1} \eta_k e^{-\beta_k t} = e^{-\theta(t-1)} f_N(t). \tag{6}$$

Obviously, for $\theta = 0$, we obtain the original AWF method with coefficients η_k, β_k. If $\theta > 0$, then $f_{N,\theta}(t)$ is suppressed for $t > 1$ and amplified for $t < 1$, compared to $f_N(t)$. If $\theta < 0$, these relations are reversed.

4. ERROR INDICATOR

We look for information about the accuracy of an NILT method defined by the $\eta_k(\theta), \beta_k(\theta)$ parameters. I.e., the error of the approximation $h(T) \approx h_N(T,\theta)$, where $h_N(T,\theta)$ is computed according to (2) and $h(T)$ is not known.

4.1 Properties of the weight functions

Let $z_1, z_2, \ldots$ denote the zeros of $f_{N,\theta}(t)$ for $t > 0$ in increasing order. We set $z_0 = 0$ (regardless of whether $f_{N,\theta}(0) = 0$ or not). According to (6), the z_i parameters do not depend on θ. The index of the largest zero less than one is denoted by I, that is, $1 \in [z_I, z_{I+1}]$. We decompose $h_N(T)$ from (3) as

$$h_N(T,\theta) = \int_0^\infty h(tT) \cdot f_{N,\theta}(t)\mathrm{d}t = \underbrace{\int_0^{z_I} h(tT) f_{N,\theta}(t)\mathrm{d}t}_{\varepsilon_{\text{left}}(\theta)}$$
$$+ \underbrace{\int_{z_I}^{z_{I+1}} h(tT) f_{N,\theta}(t)\mathrm{d}t}_{h_{\text{main}}(\theta)} + \underbrace{\int_{z_{I+1}}^{\infty} h(tT) f_{N,\theta}(t)\mathrm{d}t}_{\varepsilon_{\text{right}}(\theta)}, \tag{7}$$

and refer to these terms as the main term, $h_{\text{main}}(\theta)$, the left error term, $\varepsilon_{\text{left}}(\theta)$, and the right error term, $\varepsilon_{\text{right}}(\theta)$. This naming convention comes from the fact that, if $f_{N,\theta}(t)$ was the unit impulse function at one, then we would have $\varepsilon_{\text{left}}(\theta) = \varepsilon_{\text{right}}(\theta) = 0$ and $h_{\text{main}}(\theta) = h_N(T,\theta) = h(T)$.

We can decompose the associated weight functions similarly

$$\int_0^\infty f_{N,\theta}(t)\mathrm{d}t = \tag{8}$$
$$\underbrace{\int_0^{z_I} f_{N,\theta}(t)\mathrm{d}t}_{f_{\text{left}}} + \underbrace{\int_{z_I}^{z_{I+1}} f_{N,\theta}(t)\mathrm{d}t}_{f_{\text{main}}} + \underbrace{\int_{z_{I+1}}^{\infty} f_{N,\theta}(t)\mathrm{d}t}_{f_{\text{right}}} \quad (= 1).$$

For the Euler weight function $f_{\text{main}} >> 1$ and $f_{\text{left}} + f_{\text{right}} << 0$, where the $<<$ relation indicates "significant" differences. In contrast, the CME weight function is nonnegative, consequently, f_{main}, f_{left} and f_{right} are all nonnegative, furthermore $f_{\text{main}} \approx 1$, therefore $1 - f_{\text{main}} < 0.01$, as it is demonstrated by Table 1. The integrals f_{main}, f_{left}, and f_{right} hardly change with the order, while the (z_I, z_{I+1}) interval, where the main peak of the weight function is located, decreases significantly with increasing order.

According to the assumption that $h(t)$ is nonnegative, we can interpret the $h_{\text{main}}(\theta)$, $\varepsilon_{\text{left}}(\theta)$, $\varepsilon_{\text{right}}(\theta)$ terms depending on the sign of the weight function.

- If $f_{N,\theta}(t)$ is nonnegative (like for the CME method and its parametric variants), the terms $h_{\text{main}}(\theta)$, $\varepsilon_{\text{left}}(\theta)$,

Euler					
n	z_I	z_{I+1}	f_{left}	f_{main}	f_{right}
30	0.9534	1.0465	-0.1492	1.1967	-0.0475
60	0.9772	1.0227	-0.1528	1.2012	-0.0483
CME					
n	z_I	z_{I+1}	f_{left}	f_{main}	f_{right}
30	0.9344	1.0698	0.0028	0.9950	0.0021
60	0.9689	1.0322	0.0026	0.9949	0.0023

Table 1: Properties of the weight function for the Euler and the CME method

and $\varepsilon_{\text{right}}(\theta)$ are all nonnegative. In this case $h_{\text{main}}(\theta)$ approximates $h(T)$, and $\varepsilon_{\text{left}}(\theta)$ and $\varepsilon_{\text{right}}(\theta)$, represents the error of the approximation.

- If $f_{N,\theta}(t)$ has alternating sign (like in the case of the Euler method and its parametric variants), such clear interpretation of the $h_{\text{main}}(\theta)$, $\varepsilon_{\text{left}}(\theta)$, $\varepsilon_{\text{right}}(\theta)$ terms is not available. In this case $h_{\text{main}}(\theta) >> h(T)$, $\varepsilon_{\text{left}}(\theta) << 0$ and $\varepsilon_{\text{right}}(\theta) << 0$ for "smooth" functions.

4.2 Measuring the error by the computed NILT value

When both $h(t)$ and $f_{N,\theta}(t)$ are known to be nonnegative, and consequently $\varepsilon_{\text{left}}(\theta)$, $\varepsilon_{\text{right}}(\theta)$, and $h_{\text{main}}(\theta)$ are known to be nonnegative, we can approximate the error of the NILT in a computationally efficient way.

For the parametric Euler and CME methods, the main peak of $f_{N,\theta}(t)$ and consequently $h_{\text{main}}(\theta) \approx \tilde{h}_{\text{main}}$ in (7) is fairly independent of θ. That is

$$\min_{\theta} h_N(T,\theta) = \min_{\theta}(\varepsilon_{\text{right}}(\theta) + h_{\text{main}}(\theta) + \varepsilon_{\text{right}}(\theta))$$
$$\approx \tilde{h}_{\text{main}} + \min_{\theta}(\varepsilon_{\text{right}}(\theta) + \varepsilon_{\text{right}}(\theta)),$$

thus minimizing $h_N(T,\theta)$ according to θ minimizes the error of the NILT as well. Consequently, with nonnegative weight function, the NILT value $h_{N,\theta}(T)$ itself can be used to indicate the error with different θ parameters.

5. OPTIMIZATION METHOD

The optimization problem defined in the previous section can be solved with various optimization approaches. To pick a computationally efficient one, we utilize the property, that increasing θ exponentially suppresses $f_{N,\theta}(t)$ for $t > 1$ and consequently, it exponentially reduces $\varepsilon_{\text{right}}(\theta)$ and at the same time it amplifies $f_{N,\theta}(t)$ for $t < 1$ and exponentially increases $\varepsilon_{\text{left}}(\theta)$. That is, to optimize the shifting parameter of the CME based NILT, we have a convex optimization problem to solve for which various ternary search methods can be applied. We apply the Golden-section search method in our implementation.

6. NUMERICAL ANALYSIS OF THE OPTIMIZED CME METHOD

Applying the proposed optimized NILT method starting from the CME method, referred to as CME-S method, we experience significant accuracy gain for computing small values, like the tail of a distribution. For $h(t) = \exp(-t^2)$ and $h^*(s) = \frac{1}{2}e^{(s/2)^2}\sqrt{\pi}\,\text{Erfc}(s/2)$ where $\text{Erfc}(z) = \frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}dt$

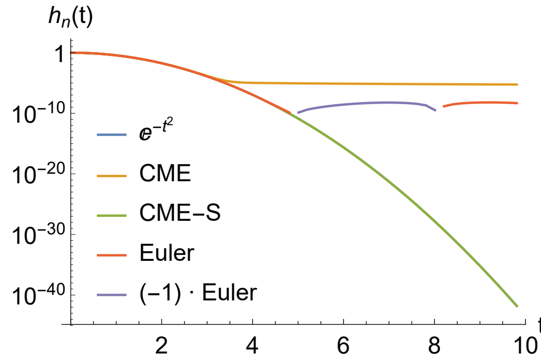


Figure 2: Order 30 NILT approximation of $h(t) = \exp(-t^2)$. "(-1) · Euler" indicates the range where the result of the Euler method is negative.

is the error function Figure 2 plots the computed NILT values in logarithmic scaling and some numerical values are provided in Table 2.

order	precise	CME	Euler	CME-S
$h(t) = \exp(-t^2), T = 5$				
30	$1.389E-11$	$8.739E-6$	$-1.221E-10$	$1.372E-11$
60	$1.389E-11$	$1.356E-6$	$1.389E-11$	$1.385E-11$
$h(t) = \exp(-t^2), T = 10$				
30	$3.720E-44$	$5.515E-6$	$3.889E-9$	$3.557E-44$
60	$3.720E-44$	$8.911E-7$	$3.205E-17$	$3.681E-44$

Table 2: Numerical NILT results for $h(t) = \exp(-t^2)$

7. REFERENCES

[1] J. Abate and W. Whitt. A Unified Framework for Numerically Inverting Laplace Transforms. *INFORMS Journal on Computing*, 18(4):408–421, Fall 2006.

[2] I. Horváth, G. Horváth, S. A.-D. Almousa, and M. Telek. Numerical inverse Laplace transformation using concentrated matrix exponential distributions. *Performance Evaluation*, 137:102067, 2020.

[3] Q. Wang and H. Zhan. On different numerical inverse Laplace methods for solute transport problems. *Advances in Water Resources*, 75:80 – 92, 2015.