Numerical inverse transformation of double sided Laplace transform with parameter optimization[☆]

András Mészáros[a], Miklós Telek[a,b,*]

[a] *Department of Networked Systems and Services, Technical University of Budapest, Budapest, Hungary*
[b] *MTA-BME Information Systems Research Group, Budapest, Hungary*

Abstract

There are efficient numerical inverse Laplace transformation (NILT) procedures for double sided Laplace transforms (DSLT), which are based on some hyperparameters. If those hyperparameters are set properly, the computed inverse is accurate, while using an incorrect hyperparameter results in a completely incorrect inverse.

To the best of the authors knowledge, there is no NILT procedure for DSLT, which automatically sets the hyperparameters based on the DSLT and the point where the inverse needs to be computed. In this paper we propose a procedure which automatically optimizes the hyperparameters of the NILT method and this way it eliminates the unforeseen dependency of the computed inverse on the inherent hyperparameters.

Keywords: Double sided Laplace transform, numerical inverse Laplace transformation, Parameter optimization.

1. Introduction

There are many applications, where Laplace transform of functions over the whole real axis, referred to as double sided Laplace transform (DSLT), can be efficiently used, e.g., in financial mathematics [9], electronic circuits [8], etc.

Efficient numerical methods are available for the numerical inverse Laplace transformation (NILT) of DSLTs, e.g., in [9] and in [4], but these methods are based on inherent hyperparameters of the procedure. In Eq. 1 of [9], parameter A, while in Eq 3.1 of [4], parameters σ and C are the hyperparameters of the NILT procedure.

[☆]This work is partially supported by the Hungarian Scientific Research Fund OTKA K-138208 project.

[*]Corresponding author

Email addresses: meszarosa@hit.bme.hu (András Mészáros), telek@hit.bme.hu (Miklós Telek)

The density function of a simple normal distributed random variable, $h(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$, whose DSLT is $h^{\circledast}(s) = \int_{-\infty}^{\infty} h(t)e^{-st}dt = e^{sm+s^2\sigma/2}$, can showcase the strong dependence on the hyperparameters already. Let us consider the above density function with mean $m = -5$ and variance $\sigma = 0.1$. The order $N = 30$ approximation by Eq. 1 of [9],

$$h(T) \approx \tilde{h}_A(T) = \frac{e^{A/2}}{2|T|}\mathrm{Re}\left(h^{\circledast}\left(\frac{A}{2T}\right)\right) + \frac{e^{A/2}}{|T|}\sum_{j=1}^{N}(-1)^j \mathrm{Re}\left(h^{\circledast}\left(\frac{A-2j\pi\mathbb{i}}{2T}\right)\right),$$

with the hyperparameter choice of $A = 150$ gives $\tilde{h}_{A=150}(-7) = -4.08801 * 10^6$ and $\tilde{h}_{A=150}(-3) = 2.60028 * 10^{-9}$ and with $A = -300$ it gives $\tilde{h}_{A=-300}(-7) = 2.6003 * 10^{-9}$ and $\tilde{h}_{A=-300}(-3) = -4.08801 * 10^6$, while the exact density values are $h(-7) = h(-3) = 2.60028 * 10^{-9}$.

That is, low order approximations (e.g., $N = 30$) can be accurate, if the proper hyperparameter is applied, while the computation with improper parameter gives completely incorrect results with wrong sign and wrong order of magnitude. Additionally, as this example indicates, the optimal value of the parameter depends not only on the DSLT function to invert, but also on the point of interest. Eq 3.1 of [4] shows similar strong dependence on the hyperparameters σ and C.

In [9], the hyperparameter is set to be accurate in the considered application field ([9, page 385]: "We also set the value of the constant $A = 40$, which we have found in practice to provide the best results"). While in [4, page 773], the order of the approximation is increased to compensate the improper setting of the hyperparameters. This hyperparameter setting might be appropriate in some specific application fields, but as the introduced example indicates, it is inappropriate for our intended application in the numerical analysis of Markov modulated Brownian motion [3, 2].

To eliminate this hyperparameter dependency of NILT methods, we propose a procedure which optimizes the parameters of the NILT method automatically based on the DSLT function and the point where the inverse needs to be computed.

The scope of this paper is restricted to non-negative functions, which naturally applies to our intended application field, the PDF of random variables.

The rest of the paper is organized as follows. Section 2 introduces the basics of single and double sided Laplace transformation, while Section 3 introduces existing single sided NILT methods and their extensions to double sided NILT. Section 4 discusses the role of the scaling parameter in double sided NILT and presents an NILT algorithm whose numerical properties are investigated in Section 5. A refined approximation of the scaling parameter is introduced in Section 6 and its numerical properties are investigated in Section 7. Finally, the conclusion is provided in Section 8.

2. Laplace transformation

2.1. Single sided Laplace transformation

The single sided LT of the real function $h(t)$ is defined as

$$h^*(s) = \int_{t=0}^{\infty} e^{-st} h(t) \mathrm{d}t, \tag{1}$$

where $s \in \mathbb{C}$ and, consequently, $h^*(s) \in \mathbb{C}$. Most commonly, the term "Laplace transform" refers to single sided LT, and we adopt this terminology here and explicitly indicate when we refer to the double sided case.

2.2. Double sided Laplace transformation

The DSLT of the real function $h(t)$ is defined as

$$h^{\circledast}(s) = \int_{t=-\infty}^{\infty} e^{-st} h(t) \mathrm{d}t, \tag{2}$$

where $s \in \mathbb{C}$ and $h^{\circledast}(s) \in \mathbb{C}$. We use the superscript notation $^{\circledast}$ to distinguish DSLTs from single sided LTs, which are commonly annotated with the superscript *.

For later use we also introduce the Laplace transform for the negative real axis as

$$h^{\circ}(s) = h^{\circledast}(s) - h^*(s) = \int_{t=-\infty}^{0} e^{-st} h(t) \mathrm{d}t, \tag{3}$$

where $s \in \mathbb{C}$ and $h^{\circ}(s) \in \mathbb{C}$.

In this paper, we restrict our attention to the following set of functions

A1) $h(t)$ is not known, but it is known to be real, non-negative, it has no point mass at zero (that is, $\int_{0^-}^{0^+} h(t)dt = 0$), and it is non-vanishing on the negative and the positive half axes (that is, $\int_{-\infty}^{0} h(t)dt > 0$ and $\int_{0}^{\infty} h(t)dt > 0$).

A2) The integral in (2) is finite for any $s \in \mathbb{C}$.

Furthermore, we assume that $h^{\circledast}(s)$ is available (computable) for any $s \in \mathbb{C}$ and we intend to compute $h(t)$ in point $t = T$ based on $h^{\circledast}(s)$.

2.3. Properties of the set of considered functions

The $h(t)$ functions satisfying Assumptions A1) and A2) are such that

$$0 \leq \int_{t=0}^{\infty} e^{-st} h(t) \mathrm{d}t < \infty \text{ for all } s \in \mathbb{R}, \tag{4}$$

consequently, there exists $\hat{a} > 1$, $\hat{c} > 0$ such that

$$h(t) < \hat{c} e^{-t^{\hat{a}}} \text{ for } t > 0. \tag{5}$$

Similarly,

$$0 \leq \int_{t=-\infty}^{0} e^{-st} h(t) \mathrm{d}t < \infty, \text{ for all } s \in \mathbb{R}, \tag{6}$$

consequently, there exists $\check{a} > 1$, $\check{c} > 0$ such that

$$h(t) < \check{c} e^{t^{\check{a}}} \text{ for } t < 0. \tag{7}$$

That is, $h(t)$ decays faster than exponentially towards both, $-\infty$ and ∞. These properties are not applied directly in the sequel, but detail the consequences of Assumptions A1) and A2).

2.4. Properties of the transform integrals with Assumptions A1) and A2)

From its definition

$$\begin{aligned} \left|h^{\circledast}(s)\right| &= \left| \int_{t=-\infty}^{\infty} e^{-st} h(t) \mathrm{d}t \right| \\ &\leq \int_{t=-\infty}^{\infty} \left|e^{-st}\right| h(t) \mathrm{d}t = \int_{t=-\infty}^{\infty} e^{-\mathrm{Re}(s)t} h(t) \mathrm{d}t = h^{\circledast}(\mathrm{Re}(s)), \end{aligned} \tag{8}$$

where $\mathrm{Re}(s)$ denotes the real part of s. Similarly, $|h^*(s)| \leq h^*(\mathrm{Re}(s))$ and $|h^\circ(s)| \leq h^\circ(\mathrm{Re}(s))$.

For $s \in \mathbb{R}$, $h^*(s)$ is monotone increasing such that

$$\lim_{s \to -\infty} h^*(s) = \infty \text{ and } \lim_{s \to \infty} h^*(s) = 0 \tag{9}$$

and $h^\circ(s) = h^{\circledast}(s) - h^*(s)$ is monotone decreasing such that

$$\lim_{s \to -\infty} h^\circ(s) = 0 \text{ and } \lim_{s \to \infty} h^\circ(s) = \infty, \tag{10}$$

and from $h^\circ(s) + h^*(s) = h^{\circledast}(s)$ it follows that

$$\lim_{s \to -\infty} h^{\circledast}(s) = \infty \text{ and } \lim_{s \to \infty} h^{\circledast}(s) = \infty. \tag{11}$$

We summarize these relations in the following theorem.

Theorem 1. *For any $\epsilon > 0$ there exists a $\sigma \in \mathbb{R}$ such that*

$$\left|h^{\circledast}(s) - h^*(s)\right| \leq \epsilon \text{ when } \mathrm{Re}(s) \leq \sigma. \tag{12}$$

Proof. The theorem follows from $|h^{\circledast}(s) - h^*(s)| = |h^\circ(s)| \leq h^\circ(\mathrm{Re}(s))$ and $\lim_{s \to -\infty} h^\circ(s) = 0$ for $s \in \mathbb{R}$. □

The primary importance of Theorem 1 is that single sided NILT methods can be used to approximate $h(t)$ at positive T points, when the applied NILT method evaluates the known transform function $h^{\circledast}(s)$ only in those points where the known $h^{\circledast}(s)$ closely approximates the unknown $h^*(s)$.

It seems to be a restriction, that T needs to be positive in order to apply single sided NILT methods, but the following relation relaxes this limitation.

2.5. Horizontal sliding of double sided Laplace transform

Let $\hat{h}(t,\Delta) = h(t-\Delta)$. Then the effect of horizontally sliding $h(t)$ with Δ can be expressed in DSLT domain as

$$\begin{aligned}\hat{h}^{\circledast}(s,\Delta) &= \int_{t=-\infty}^{\infty} e^{-st}\hat{h}(t,\Delta)\mathrm{d}t = \int_{t=-\infty}^{\infty} e^{-st}h(t-\Delta)\mathrm{d}t \\ &= e^{-s\Delta}\int_{t=-\infty}^{\infty} e^{-st}h(t)\mathrm{d}t = e^{-s\Delta}h^{\circledast}(s).\end{aligned} \tag{13}$$

As a consequence, $h(T)$ can also be approximated based on $h(T) = \hat{h}(T+\Delta,\Delta)$. That is, instead of the NILT of $h^{\circledast}(s)$ at point T, we can evaluate the NILT of $\hat{h}^{\circledast}(s,\Delta)$ at point $T+\Delta$.

3. Numerical inverse Laplace transformation

3.1. Single sided numerical inverse Laplace transformation

Let $NILT(h^*(s),T)$ denote the single sided numerical inverse Laplace transform of $h^*(s)$ at point T. In this paper, we assume that the applied NILT method is an Abate-Whitt framework (AWF) method of order N with *nodes* $\beta_k\,(0 \le k \le N-1)$ and *weights* $\eta_k\,(0 \le k \le N-1)$ [1] and, consequently, the NILT procedure that approximates $h(t)$ in point T based on $h^*(s)$ is

$$h(T) \approx NILT(h^*(s),T,N) = \sum_{k=0}^{N-1}\frac{\eta_k}{T}h^*\left(\frac{\beta_k}{T}\right), \tag{14}$$

where T has to be non-negative and $\mathrm{Re}(\beta_k/T)$ should be in the ROC of $h^*(s)$. In the AWF methods, the β_k and η_k $(0 \le k \le N-1)$ coefficients depend on the order (N), and the applied NILT method.

An *integral interpretation* of the AWF methods can be obtained from (14) by substituting (1) [6]:

$$\begin{aligned}NILT(h^*(s),T,N) &= \sum_{k=0}^{N-1}\frac{\eta_k}{T}h^*\left(\frac{\beta_k}{T}\right) = \sum_{k=0}^{N-1}\frac{\eta_k}{T}\int_0^{\infty} h(t)\cdot e^{-\beta_k t/T}\mathrm{d}t \\ &= \int_0^{\infty} h(t)\cdot\frac{1}{T}w_N(t/T)\mathrm{d}t = \int_0^{\infty} h(tT)\cdot w_N(t)\mathrm{d}t,\end{aligned} \tag{15}$$

where

$$w_N(t) = \sum_{k=0}^{N-1}\eta_k e^{-\beta_k t}. \tag{16}$$

That is, the result of an AWF NILT procedure, computed according to (14), is equivalent to the integral in (15), where $w_N(t)$ is an appropriately selected *weight function*, which depends on the applied AWF method.

Among the AWF methods we distinguish sign preserving and non-sign preserving methods [6]. Sign preserving methods are such that the sign of the NILT is non-negative (non-positive) if the function to approximate is non-negative (non-positive) for $t > 0$. If the weight function of an AWF method is non-negative then the associated NILT method is sign preserving [6], as it is readable also from (15). Throughout this paper, we apply the CME method, whose weight function is non-negative, because our proposed method makes use of its sign preserving property.

The weight function approximates the unit impulse function. If $w_N(t)$ was the unit impulse function at one, then the integral in (15) would result in a perfect Laplace inversion. The weight functions of the sign preserving AWF methods are such that $\text{argmax}_{t\in\mathbb{R}^+} w_N(t) \approx 1$ and $w_N(t)$ is negligibly small when t is far from 1. Furthermore, the main peak of $w_N(t)$ at $t = 1$ is enclosed by a zero from either side (i.e., where $w_N(t) = 0$), from which the error of the Laplace inversion can be interpreted based on

$$\begin{aligned} NILT(h^*(s), T, N) &= \int_0^\infty h(tT) \cdot w_N(t)\mathrm{d}t \qquad (17)\\ &= \underbrace{\int_0^{z_I} h(tT)w_N(t)\mathrm{d}t}_{\varepsilon_{\text{left}}} + \underbrace{\int_{z_I}^{z_{I+1}} h(tT)w_N(t)\mathrm{d}t}_{h_{\text{main}}} + \underbrace{\int_{z_{I+1}}^{\infty} h(tT)w_N(t)\mathrm{d}t}_{\varepsilon_{\text{right}}}, \end{aligned}$$

where z_I and z_{I+1} denote the borders (e.g., the closest zeros before and after $t = 1$) of the main impulse of $w_N(t)$ at $t = 1$. In this interpretation, the main term, $h_{\text{main}}(\theta)$, approximates $h(T)$, and the left error term, $\varepsilon_{\text{left}}(\theta)$, and the right error term, $\varepsilon_{\text{right}}(\theta)$ causes the error of the approximation.

Any AWF method can be extended with a so called *shifting* parameter, θ, which results in the NILT procedure [7]

$$\begin{aligned} NILT_\theta(h^*(s), T, N, \theta) &= \sum_{k=0}^{N-1} \frac{e^\theta \eta_k}{T} h^*\left(\frac{\beta_k + \theta}{T}\right) \qquad (18)\\ &= \underbrace{\int_0^{z_I} h(tT)w_{N,\theta}(t)\mathrm{d}t}_{\varepsilon_{\text{left}}(\theta)} + \underbrace{\int_{z_I}^{z_{I+1}} h(tT)w_{N,\theta}(t)\mathrm{d}t}_{h_{\text{main}}(\theta)} + \underbrace{\int_{z_{I+1}}^{\infty} h(tT)w_{N,\theta}(t)\mathrm{d}t}_{\varepsilon_{\text{right}}(\theta)}, \end{aligned}$$

where $\text{Re}((\beta_k + \theta)/T)$ should be in the ROC.

The effect of the shifting parameter on the weight function provides an intuitive explanation on the behavior of the associated NILT method

$$\begin{aligned} w_{N,\theta}(t) &= \sum_{k=0}^{N-1} \eta_k(\theta) e^{-\beta_k(\theta)t} = \sum_{k=0}^{N-1} (e^\theta \eta_k) e^{-(\beta_k+\theta)t}\\ &= e^{-\theta(t-1)} \sum_{k=0}^{N-1} \eta_k e^{-\beta_k t} = e^{-\theta(t-1)} w_N(t). \qquad (19) \end{aligned}$$

Obviously, for $\theta = 0$, $w_{N,\theta}(t) = w_N(t)$ and we obtain the original AWF method with coefficients η_k, β_k. When $\theta \neq 0$, we still have $w_{N,\theta}(1) = w_N(1)$. If $\theta > 0$, then $w_{N,\theta}(t)$ is suppressed for $t > 1$ and amplified for $t < 1$, compared to $w_N(t)$. That is, $\varepsilon_{\text{left}}(\theta) > \varepsilon_{\text{left}}$, $h_{\text{main}}(\theta) \approx h_{\text{main}}$, and $\varepsilon_{\text{right}}(\theta) < \varepsilon_{\text{right}}$. If $\theta < 0$, the inequalities are reversed.

If the applied NILT method is sign preserving ($w_N(t) > 0$) and $h(t)$ is non-negative then the optimal shifting parameter, which minimizes the error of the approximation, can be obtained according to [7] as

$$NILT_{\text{opt}}(h^*(s), T, N) = \min_{\theta} NILT_{\theta}(h^*(s), T, N, \theta). \tag{20}$$

As it can bee seen from (19), $\varepsilon_{\text{left}}(\theta)$ ($\varepsilon_{\text{right}}(\theta)$) increases (decreases) exponentially fast with θ and the minimum in (20) is attained at around the θ value for which $\varepsilon_{\text{left}}(\theta) = \varepsilon_{\text{right}}(\theta)$. When $w_N(t)$ and $h(t)$ are non-negative, $NILT_{\theta}(h^*(s), T, N, \theta)$ is proved to be a convex function [7, Theorem 2], whose minimum can be found efficiently using simple optimization methods (e.g., the golden section search).

3.2. Double sided numerical inverse Laplace transformation

The integral interpretation in (15) allows to interpret the AWF-like method with nodes β_k and weights η_k for NILT of a DSLT function, which we denote as NILT2,

$$\begin{aligned} NILT2(h^{\circledast}(s), T, N) &= \sum_{k=0}^{N-1} \frac{\eta_k}{T} h^{\circledast}\left(\frac{\beta_k}{T}\right) = \int_{-\infty}^{\infty} h(tT) \cdot w_N(t) \mathrm{d}t \qquad (21) \\ &= \underbrace{\int_{-\infty}^{z_I} h(tT) w_N(t) \mathrm{d}t}_{\varepsilon_{\text{left}}} + \underbrace{\int_{z_I}^{z_{I+1}} h(tT) w_N(t) \mathrm{d}t}_{h_{\text{main}}} + \underbrace{\int_{z_{I+1}}^{\infty} h(tT) w_N(t) \mathrm{d}t}_{\varepsilon_{\text{right}}}, \end{aligned}$$

where $w_N(t)$ is defined in (16) and z_I and z_{I+1} are the same as in (17). That is, the obtained NILT2 method is the same as the related NILT method, except that the integral in the left error term starts from $-\infty$. If $w_N(t)$ was negligible also for $t \in \{-\infty, 0\}$, then the accuracy of the approximation would be similar to the regular LT case. Unfortunately, it is hard to find such a weight function in the form of (16), but the concept of shifting can be applied in a similar manner as for the single sided NILT case:

$$\begin{aligned} NILT2_{\theta}(h^{\circledast}(s), T, N, \theta) &= \sum_{k=0}^{N-1} \frac{e^{\theta}\eta_k}{T} h^{\circledast}\left(\frac{\beta_k + \theta}{T}\right) \qquad (22) \\ &= \underbrace{\int_{-\infty}^{z_I} h(tT) w_{N,\theta}(t) \mathrm{d}t}_{\varepsilon_{\text{left}}(\theta)} + \underbrace{\int_{z_I}^{z_{I+1}} h(tT) w_{N,\theta}(t) \mathrm{d}t}_{h_{\text{main}}(\theta)} + \underbrace{\int_{z_{I+1}}^{\infty} h(tT) w_{N,\theta}(t) \mathrm{d}t}_{\varepsilon_{\text{right}}(\theta)}. \end{aligned}$$

The related result of the single sided case, [7, Theorem 2], easily extends to the double sided case, which we present here for completeness.

Theorem 2. *If $h(t)$ and $w_N(t)$ are non-negative, then $NILT2_\theta(h^{\circledast}(s), T, N, \theta)$ is a convex function of θ.*

Proof. Using $w_{N,\theta}(t) = e^{-\theta(t-1)} w_N(t)$, the second derivative of (22) is

$$\frac{d^2}{d\theta^2} NILT2_\theta(h^{\circledast}(s), T, N, \theta) = \int_{-\infty}^{\infty} h(tT) \cdot (1-t)^2 e^{\theta(1-t)} w_N(t) \mathrm{d}t \geq 0.$$

□

From the parametric set of $NILT2_\theta$ methods, we use the one that is minimal at the given T value:

$$NILT2_{\text{opt}}(h^{\circledast}(s), T, N) = \min_\theta NILT2_\theta(h^{\circledast}(s), T, N, \theta). \tag{23}$$

Similarly to the single sided case [7], for the convex optimization in (23), we use golden section search and the θ value at the optimum is such that $\varepsilon_{\text{left}}(\theta) \approx \varepsilon_{\text{right}}(\theta)$, but in this case the integral of $\varepsilon_{\text{left}}(\theta)$ starts from $-\infty$.

The computational complexity of this $NILT2_{\text{opt}}$ procedure is characterized by the number of evaluations of $h^{\circledast}(s)$. In (22), for a given θ value, $h^{\circledast}(s)$ is evaluated N times, where N is the order of the NILT approximation, and the optimization in (23) requires approximately 20 evaluations for different θ values. (The precise number depends on the accuracy goal, but it was less than 20 in all examples presented in this paper.) That is, $NILT2_{\text{opt}}$ requires approximately $20N$ evaluations of $h^{\circledast}(s)$.

4. The scaling parameter

In case of DSLT with ROC over the whole complex plane, (13) offers an additional free parameter to optimize the NILT procedure, since

$$\begin{aligned} h(T) = \hat{h}(T+\Delta, \Delta) &\approx NILT2_{\text{opt}}(\hat{h}^{\circledast}(s, \Delta), T+\Delta, N) \\ &= NILT2_{\text{opt}}(e^{-s\Delta} h^{\circledast}(s), T+\Delta, N), \end{aligned} \tag{24}$$

where Δ is a free parameter to optimize the approximation.

For a better intuitive understanding of the role of the Δ parameter, we utilize

the integral interpretation again.

$$
\begin{aligned}
&NILT2_{\text{opt}}(e^{-s\Delta}h^{\circledast}(s), T+\Delta, N)\\
&\quad= \min_{\theta} \sum_{k=0}^{N-1} \frac{e^{\theta}\eta_k}{T+\Delta} e^{-\frac{\beta_k+\theta}{T+\Delta}\Delta} h^{\circledast}\left(\frac{\beta_k+\theta}{T+\Delta}\right)\\
&\quad= \min_{\theta} \sum_{k=0}^{N-1} \frac{e^{\theta}\eta_k}{T+\Delta} e^{-\frac{\beta_k+\theta}{T+\Delta}\Delta} \int_{-\infty}^{\infty} h(t)\cdot e^{-\frac{(\beta_k+\theta)t}{T+\Delta}}\,\mathrm{d}t\\
&\quad= \min_{\theta} \sum_{k=0}^{N-1} \frac{e^{\theta}\eta_k}{T+\Delta} \int_{-\infty}^{\infty} h(t)\cdot e^{-\frac{(\beta_k+\theta)(t+\Delta)}{T+\Delta}}\,\mathrm{d}t\\
&\quad= \min_{\theta} \sum_{k=0}^{N-1} \frac{e^{\theta}\eta_k}{T+\Delta} \int_{-\infty}^{\infty} h(t-\Delta)\cdot e^{-\frac{(\beta_k+\theta)t}{T+\Delta}}\,\mathrm{d}t\\
&\quad= \min_{\theta} \int_{-\infty}^{\infty} h(t-\Delta)\cdot \frac{1}{T+\Delta} w_{N,\theta}\left(\frac{t}{T+\Delta}\right)\mathrm{d}t\\
&\quad= \min_{\theta} \int_{-\infty}^{\infty} h(t(T+\Delta)-\Delta)\cdot w_{N,\theta}\left(t\right)\mathrm{d}t\ . \qquad (25)
\end{aligned}
$$

That is, the result of the obtained $NILT2_{\text{opt}}$ method is the integral of the appropriately shifted weight function $w_{N,\theta}(t)$ multiplied by a scaled version of the unknown $h(t)$ function. This is why we refer to Δ as the scaling parameter.

We note again, if $w_{N,\theta}(t)$ was the unit impulse function in 1, then the $NILT2_{\text{opt}}$ method would be exact independent of Δ.

As a consequence of Assumption A2, $h(t)$ quickly decays towards both ∞ and $-\infty$ (c.f. (5) and (7)). The scaling parameter allows to adjust the width of the interval where $h(t(T+\Delta)-\Delta)$ has significant contribution to the integral in (25). When $T+\Delta$ is large this interval is "narrow", and vice versa.

4.1. Numerical demonstration

The following general parameters are applied in the numerical examples of this paper. If it is not indicated differently, the numerical examples are computed with order $N = 30$ NILT approximation using the CME method [6] and the local approximations of Algorithm 1 and 2 are computed with $\delta = 10^{-6}$.

We demonstrate the effect of the scaling parameter through a simple example in Table 1. We consider the PDF of the normal distribution with mean $\mu = 3$ and variance $\sigma^2 = 1$. That is, $h(t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ is the Gaussian density function (GDF) and $h^{\circledast}(s) = \int_{-\infty}^{\infty} e^{-st}h(t)dt = e^{-s\mu+s^2\sigma^2/2}$, and we approximate $h(T)$ for $T = 1, 2, 3, 4, 5$ based on $h^{\circledast}(s)$ using the $NILT2_{\text{opt}}(e^{-s\Delta}h^{\circledast}(s), T+\Delta, N)$ method with order $N = 30$ and various $T+\Delta$ parameters. The table indicates that the result is a monotone decreasing function of the scaling parameter for each evaluated T value and the error of the approximation could be significant, especially, when the scaling parameter is too small. We highlight that the most accurate approximation is obtained at $T+\Delta = 4$ in each evaluated case.

T	*exact*	$NILT2_{\mathrm{opt}}(e^{-s\Delta}h^{\circledast}(s), T+\Delta, 30)$ with $T+\Delta=$				
		1/4	1	**4**	16	64
1	0.053991	0.813426	0.116573	**0.0542341**	0.051022	0.0315184
2	0.241971	2.058	0.52161	**0.243032**	0.228665	0.141256
3	0.398942	3.88404	0.84537	**0.400682**	0.377011	0.232891
4	0.241971	2.25507	0.50848	**0.243043**	0.228665	0.141256
5	0.053991	0.616544	0.114859	**0.0542263**	0.051022	0.0315186

Table 1: The effect of the shifting parameter on the $NILT2_{\mathrm{opt}}$ method applied for the GDF with mean 3 and variance 1

We explain the monotone decreasing behaviour in Table 1 using Figure 1, which shows the weight function $w_{N,\theta}(t)$, the scaled object function $h(t(T+\Delta)-\Delta)$, and the product of the two functions $h(t(T+\Delta)-\Delta)\cdot w_{N,\theta}(t)$. These determine the NILT2 approximation according to (25). Figure 1 uses a linear x and a logarithmic y axis. A GDF on such a plot appears as a parabola opening to the bottom. Its axis of symmetry is at the mean (μ), and its width (latus rectum) is determined by the variance (σ^2). In the plots, the weight function, is the order 30 CME weight function with appropriate node shifting (c.f. (23)).

At $T+\Delta=1/4$, the $h(t(T+\Delta)-\Delta)$ function is scaled such that it is rather flat in the observed interval and, consequently, the $h(t(T+\Delta)-\Delta)\cdot w_{N,\theta}(t)$ product has many comparable peaks which add up in the integral in (25). Assuming that $h_{\mathrm{main}}(\theta)$ should approximate $h(T)$ in (22), the large peaks on the left and the right of $t=1$ result in significant left and right errors ($\varepsilon_{\mathrm{left}}(\theta)$ and $\varepsilon_{\mathrm{right}}(\theta)$) and the overall integral ($\varepsilon_{\mathrm{left}}(\theta)+h_{\mathrm{main}}(\theta)+\varepsilon_{\mathrm{right}}(\theta)$) becomes larger than $h(T)$.

On the contrary, when $T+\Delta=64$, the scaled $h(t(T+\Delta)-\Delta)$ function becomes so narrow that, while the left and right errors are negligible, $h_{\mathrm{main}}(\theta)$ also becomes significantly smaller than $h(T)$. This is because $h(t(T+\Delta)-\Delta)$ is narrower than (z_I, z_{I+1}), where the main peak of $w_{N,\theta}(t)$ is located.

Between these two extreme cases, at $T+\Delta=4$, the scaling is such that $h(t(T+\Delta)-\Delta)$ is negligibly small at the peaks of $w_{N,\theta}(t)$ left and right of $t=1$, but still wide enough to have significant values in the whole (z_I, z_{I+1}) interval. That is, $h(z_I(T+\Delta)-\Delta)\approx h(T)$ and $h(z_{I+1}(T+\Delta)-\Delta)\approx h(T)$ and consequently, $h_{\mathrm{main}}(\theta)\approx h(T)$.

4.2. The proposed scaling parameter

Since our primary application for the proposed $NILT2_{\mathrm{opt}}$ method is the analysis of the Markov modulated Brownian motion, which is strongly related with normal distribution, we propose a heuristic setting of the scaling parameter, such that the proposed scaling parameter results in close approximation for GDFs and we investigate its applicability in various different cases in Section 5.

If $h(t)$ is the GDF then $\log(h(t(T+\Delta)-\Delta))$ is a parabola, whose width increases with the variance of the normal distribution and decreases with the scaling parameter $T+\Delta$. As Figure 1 demonstrates, there is an optimal width of $h(t(T+\Delta)-\Delta)$, which result in the most accurate $NILT2_{\mathrm{opt}}$.

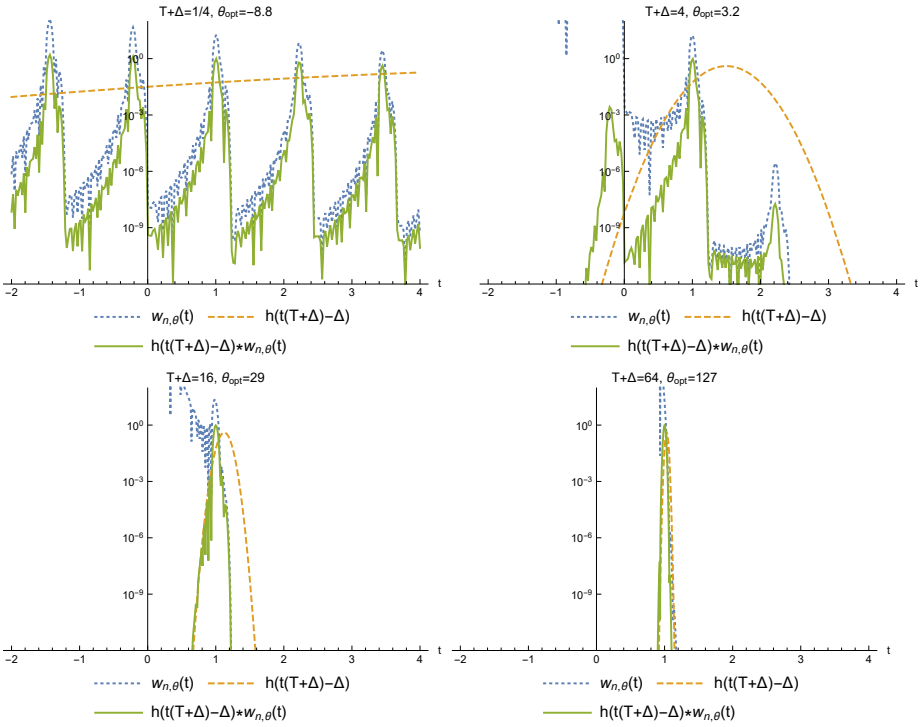

Figure 1: Effect of the scaling parameter on the integral in (25) at $T = 1$

According to our numerical experiments, e.g., the ones in Table 1, the ratio of the normalized variance of $h(t)$ and the scaling parameter has its optimum at 1/4. That is,

$$\Delta = 4\sqrt{\frac{m_2 m_0 - m_1^2}{m_0^2}} - T, \tag{26}$$

where $m_i = \int_{-\infty}^{\infty} t^i h(t)dt$ for $i = 0, 1, 2$. Since $h(t)$ is not known, for $i = 0, 1, 2$, we compute m_i based on $h^{\circledast}(s)$ as

$$m_i = \int_{-\infty}^{\infty} t^i h(t)dt = (-1)^i \left.\frac{d^i}{ds^i} h^{\circledast}(s)\right|_{s=0}. \tag{27}$$

For $i = 0$, we have $m_0 = h^{\circledast}(0)$. For $i = 1, 2$, we approximate the derivatives using the order 2 finite difference coefficients as

$$m_1 = \frac{h^{\circledast}(-\delta) - h^{\circledast}(\delta)}{2\delta}, \quad m_2 = \frac{h^{\circledast}(-\delta) - 2h^{\circledast}(0) + h^{\circledast}(\delta)}{\delta^2}.$$

Algorithm 1 presents the steps of the resulted NILT method. The computational complexity of the algorithm is characterized by approximately $20N + 3$ evaluations of $h^{\circledast}(s)$: 3 evaluations in line 1 and approximately $20N$ evaluations in line 4.

Algorithm 1 Global NILT2 method

Input: $h^{\circledast}(s), T, N, \delta$.

1: Numerical approximation of the derivatives in (27):
$f_i \leftarrow h^{\circledast}(i\delta)$, for $i = \{-1, 0, 1\}$.
2: Global moments approximation:
$m_0 \leftarrow f_0, m_1 \leftarrow \frac{f_{-1} - f_1}{2\delta}, m_2 \leftarrow \frac{f_1 - 2f_0 + f_{-1}}{\delta^2}$.
3: $\Delta \leftarrow 4\sqrt{\frac{m_2 m_0 - m_1^2}{m_0^2}} - T$ according to (26).
4: **return** $NILT2_{\text{opt}}(e^{-s\Delta} h^{\circledast}(s), T + \Delta, N)$.

5. Numerical examples

In this section we investigate the behaviour of the prosed NILT2 method for different families of functions with analytic DSLT on the whole complex plane.

5.1. PDF of normally distributed random variables

The $T + \Delta = 4$ column of Table 1 already presents an example on the accuracy of the proposed NILT method (because Algorithm 1 sets $\Delta = 4 - T$ for that example), when applied for that GDF. Further examples are depicted in Figure 2, where the relative error, computed as $r_{rel} = \frac{|h_{NILT2}(T) - h(T)|}{h(T)}$, remains below 0.005 in all cases. Based on these experiments, and further unreported ones, we conclude that the NILT2 procedure with the proposed setting of the scaling parameter can be used to obtain a fairly accurate approximation of GDF.

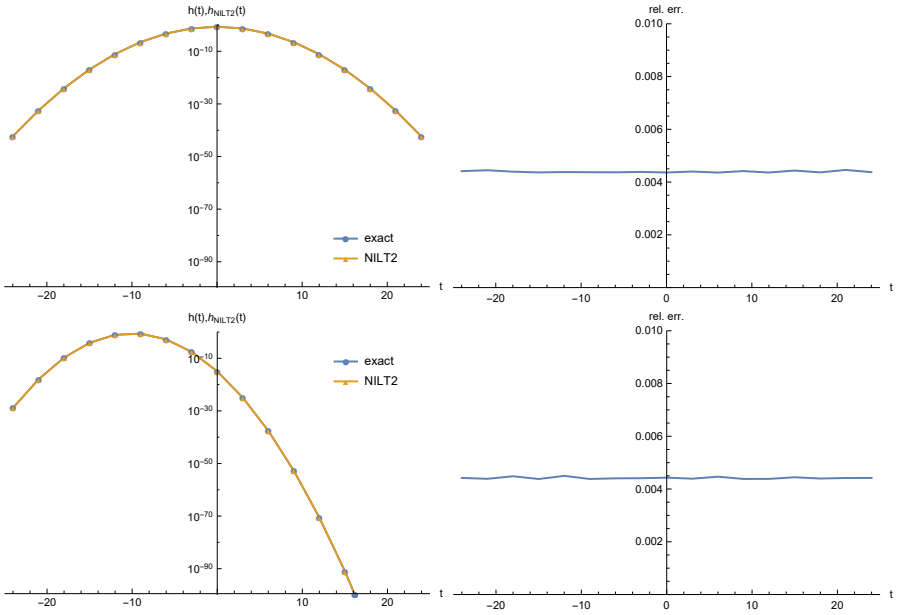

Figure 2: NILT2 of the density of normal distribution with different parameters (upper line: $\mu = 0,\ \sigma = 3$; lower line: $\mu = -10,\ \sigma = 1.5$)

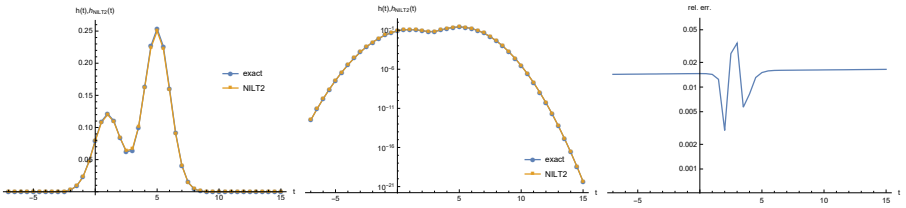

Figure 3: NILT2 approximation of mixture of normal distributions with $I = 2$, $m_1 = 1$, $\sigma_1^2 = 1.2$, $m_2 = 5$, $\sigma_2^2 = 1.1$, $p_1 = 1/3$, $p_2 = 2/3$

5.2. PDF of mixtures of normal distributed random variables

The set of functions representing the mixtures of normal distributed random variables can be efficiently used for testing our procedure, because the PDF and its DSLT are easily obtained from the linear property of LT. If $h_i(t) = \frac{1}{\sigma_i\sqrt{2\pi}} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}}$ are GDFs for $i = \{1, 2, \ldots, I\}$ and $h(t) = \sum_{i=1}^{I} p_i h_i(t)$, then $h_i^{\circledast}(s) = e^{-s\mu_i + s^2\sigma_i^2/2}$ for $i = \{1, 2, \ldots, I\}$ and $h^{\circledast}(s) = \sum_{i=1}^{I} p_i h_i^{\circledast}(s)$.

For the case when $I = 2$, $m_1 = 1$, $\sigma_1^2 = 1.2$, $m_2 = 5$, $\sigma_2^2 = 1.1$, $p_1 = 1/3$, $p_2 = 2/3$, Figure 3 plots the exact and the NILT2 values with linear and logarithmic y axis. The NILT2 approximation of this example with Algorithm 1 provides fairly accurate results, with less than 0.05 relative error for all T.

Modifying only a single parameter of the mixture and setting $m_2 = 15$ (instead of $m_2 = 5$), we obtain a function whose NILT2 approximation with

significantly lower accuracy according to Figure 4. Our numerical investigations suggest that there are two different kinds of approximation errors in Figure 4, which have two different reasons.

The first kind of error is the ≈ 0.1 relative error in case of order 30 NILT2 (e.g., $h_{NILT2}(15) = 0.222352$ instead of $h(15) = 0.253584$), which is dominant everywhere, except for the $(4, 13)$ interval. The reason of this error is the improper setting of the scaling parameter. Setting the scaling parameter according to (26) gives an accurate NILT2 for functions which are similar to the GDF. The mixture in Figure 3 is rather similar to the GDF, thus the approximation of the scaling parameters and consequently the NILT2 are fairly accurate. In contrast, the mixture in Figure 4 significantly differs from the GDF (it is much "wider" than a GDF), and (26) does not provide the appropriate scaling parameter, which results in the ≈ 0.1 relative error. To overcome this error, in Section 6, we propose a procedure for a refined approximation of the scaling parameter.

The second kind of error, which is far more significant, appears in the $(4, 13)$ interval. For $T \in (4, 13)$, the function is such that it has significant peaks both right and left to T. That is, in (25), $h(t(T + \Delta) - \Delta)$ takes much larger values than $h(T)$ both for $t < 1$ and $t > 1$. To suppress those large values for $t < 1$, θ has to be negative, but to suppress the large values for $t > 1$, θ has to be positive. As a result, no θ setting can suppress both large peaks for $t < 1$ and $t > 1$, therefore the integral in (25) is dominated by these large peaks. Figure 5 plots the elements of the integral interpretation in (25) for $T = 8$ with order 30 and $T + \Delta = 1, 12, 64$. In all the 3 cases, the optimal θ value results in similar $\varepsilon_{\text{left}}(\theta)$ and $\varepsilon_{\text{right}}(\theta)$ errors. While $h_{\text{main}}(\theta)$ properly approximates $h(T)$ in the first two cases ($T + \Delta = 1, 12$), the computed NILT2 values are much higher than $h(T)$ due to the significant $\varepsilon_{\text{left}}(\theta)$ and $\varepsilon_{\text{right}}(\theta)$ errors. In the third case, when $T + \Delta = 64$, the scaled function, $h(t(T + \Delta) - \Delta)$, is so narrow that $\varepsilon_{\text{left}}(\theta)$ and $\varepsilon_{\text{right}}(\theta)$ are negligible, but unfortunately, $h_{\text{main}}(\theta)$ overestimates $h(T)$ because the peaks of $h(t(T + \Delta) - \Delta)$, which are much larger than $h(T)$, dominate the integral in the (z_I, z_{I+1}) interval.

For $h(t)$ functions which are even more different from the GDF and have sharper multi-modal alternation than the one in Figure 4, these two types of errors can get even more significant.

A potential way to deal with such errors, or at least reduce their effect, is to increase the order of the applied NILT2 procedure, which improves the approximation of the unit impulse function by the weight function [6]. Figure 4 also demonstrates the gain of using order 60 NILT2 approximation instead of order 30.

5.3. Functions with non quadratic decays

In the previous examples, we investigated NILT2 of functions whose asymptotic decay is the same as the one of GDF. Here we investigate the accuracy of the NILT2 procedure for a function with different decay rate. For $h(t) = e^{-|x|^3}$,

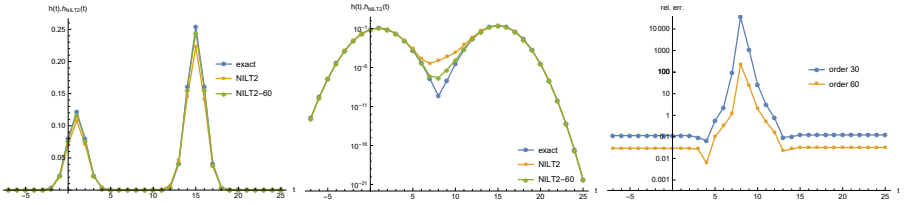

Figure 4: Order 30 and 60 NILT2 approximation of mixture of normal distributions with $I = 2$, $m_1 = 1$, $\sigma_1^2 = 1.2$, $m_2 = 15$, $\sigma_2^2 = 1.1$, $p_1 = 1/3$, $p_2 = 2/3$ with linear and logarithmic y-axis, and their relative errors

Figure 6 depicts the NILT2 approximation of $h(t)$[1] and its relative error. The approximation is surprisingly accurate at $T = 0$ and it gets less accurate than in case of GDFs (relative error < 0.005) for $|T| > 1$. In the previous examples with quadratic decay toward $t \to -\infty$ and $t \to \infty$, the relative error got constant towards the limits. In this example, where the decay is faster than quadratic, the relative error increases towards both limits.

We also tried to check the behaviour at a slower decay (for $h(t) = e^{-|x|^{3/2}}$), but unfortunately, we were unable to compute $h^{\circledast}(s)$ with any available computational platform in this case, which made the application of NILT2 impossible.

5.4. PDF of the Markov modulated Brownian motion

The Markov modulated Brownian motion $\{\mathcal{Z}(t) = \{\mathcal{J}(t), \mathcal{X}(t)\}, t > 0\}$ is a compound stochastic process consisting of a background continuous time Markov chain $\{\mathcal{J}(t), t > 0\}$ and a Brownian motion $\{\mathcal{X}(t), t > 0\}$ modulated by this Markov chain [3]. When the Markov chain stays in state i for the $(t, t+\Delta)$, interval $\mathcal{X}(t)$ increases with a normal distributed amount with mean $r_i\Delta$ and variance $\sigma_i^2\Delta$, that is, when $\mathcal{J}(\tau) = i, \forall\tau \in (t, t+\Delta)$

$$\frac{d}{dx}Pr(\mathcal{X}(t+\Delta) - \mathcal{X}(t) < x) = \mathcal{N}(r_i\Delta, \sigma_i^2\Delta, x), \tag{28}$$

where $\mathcal{N}(\mu, \sigma^2, x) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ is the GDF. We denote the generator matrix of the Markov chain by $\mathbf{Q}$, and the diagonal matrix of the means and the variances by $\mathbf{R} = \mathrm{diag}(r_i)$ and $\mathbf{\Sigma} = \mathrm{diag}(\sigma_i^2)$.

Assuming $\mathcal{X}(0) = 0$, the state dependent density of $\mathcal{X}(t)$ is

$$[\mathbf{N}(t, x)]_{i,j} = \frac{\partial}{\partial x}Pr(\mathcal{X}(t) < x, \mathcal{J}(t) = j | \mathcal{J}(0) = i) \tag{29}$$

and its DSLT is [5]

$$\mathbf{N}^*(t, v) = \int_{x=-\infty}^{\infty} e^{-xv}\mathbf{N}(t, x)dx = e^{(\mathbf{Q} - v\mathbf{R} - v^2\mathbf{\Sigma}/2)t}. \tag{30}$$

[1] where $h^{\circledast}(s)$ is efficiently computed in Mathematica based on (2).

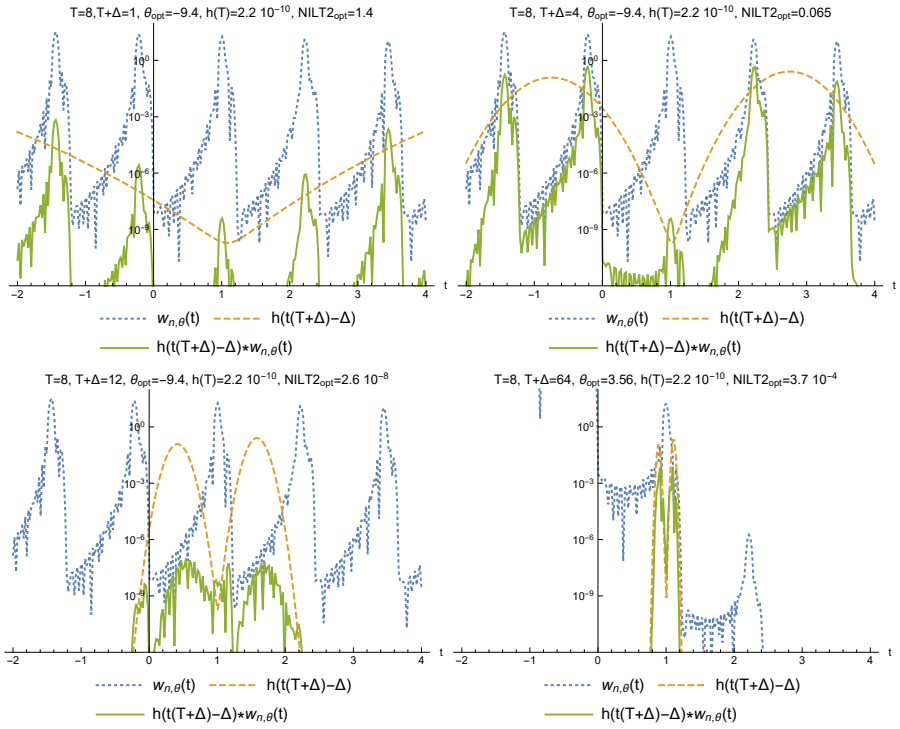

Figure 5: The integral interpretation of the order 30 NILT2 approximation for the mixture of normal distributions with $I = 2$, $m_1 = 1$, $\sigma_1^2 = 1.2$, $m_2 = 15$, $\sigma_2^2 = 1.1$, $p_1 = 1/3$, $p_2 = 2/3$

The authors are not aware of any efficient numerical method to calculate $\mathbf{N}(t, x)$ for general $\mathbf{Q}$, $\mathbf{R}$ and $\mathbf{\Sigma}$ matrices, except the NILT2 transformation of $\mathbf{N}^*(t, v)$ with respect to the transform variable v.

Since we are interested in the accuracy of the NILT2 approximation of $\mathbf{N}(t, x)$, we consider a special case for which $\mathbf{N}(t, x)$ can be also computed directly. When $\mathbf{Q} = \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix}$, $\mathbf{R} = \begin{bmatrix} 1 & \\ & 15 \end{bmatrix}$, $\mathbf{\Sigma} = \begin{bmatrix} 2 & \\ & 1.1 \end{bmatrix}$, and the modulating Markov chain starts from state 1 ($\mathcal{J}(0) = 1$), the $\mathcal{J}(t)$ process performs a single state transition at time τ, which is exponentially distributed with parameter $q_{12} = 1$. Conditioning on the time of this state transition, the PDF of $\mathcal{X}(t)$

$$\frac{\partial}{\partial x} Pr(\mathcal{X}(t) < x | \mathcal{J}(0) = 1) = [\mathbf{N}(t, x)]_{1,1} + [\mathbf{N}(t, x)]_{1,2} \tag{31}$$

can be computed as follows. If $\tau > t$, then $\mathcal{X}(t)$ is normally distributed with mean $r_1 t$ and variance $\sigma_1^2 t$. If $\tau < t$, then $\mathcal{X}(t)$ is the sum of two normally distributed random variable with mean $r_1 \tau$ and variance $\sigma_1^2 \tau$ and with mean

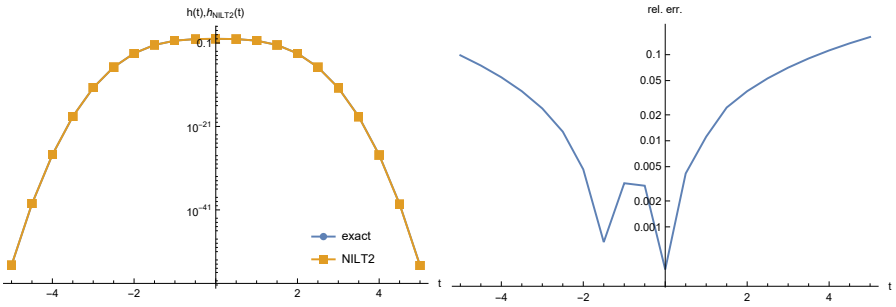

Figure 6: NILT2 approximation of $f(t) = e^{-|x|^3}$ and its relative error

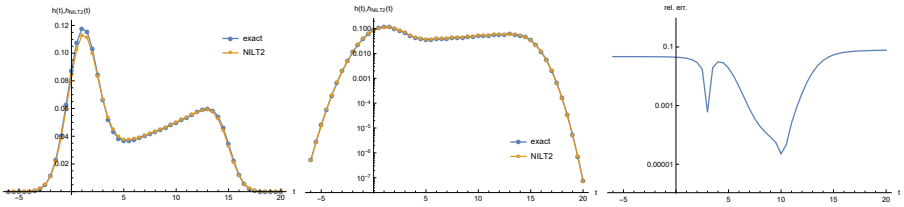

Figure 7: Distribution of the Markov modulated Brownian motion with $\mathbf{Q}$, $\mathbf{R}$ and $\mathbf{\Sigma}$ matrices at $t = 1$ and its order 30 NILT2 approximation

$r_2(t-\tau)$ and variance $\sigma_2^2(t-\tau)$. That is,

$$\frac{\partial}{\partial x} Pr(\mathcal{X}(t) < x | \mathcal{J}(0) = 1, \tau > t) = \frac{1}{\sqrt{2\pi\sigma_1^2 t}} e^{-\frac{(x-r_1 t)^2}{2\sigma_1^2 t}} \tag{32}$$

$$\begin{aligned} &\frac{\partial}{\partial x} Pr(\mathcal{X}(t) < x | \mathcal{J}(0) = 1, \tau < t) \\ &= \int_{y=-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_1^2 \tau}} e^{-\frac{(y-r_1\tau)^2}{2\sigma_1^2\tau}} \frac{1}{\sqrt{2\pi\sigma_2^2(t-\tau)}} e^{-\frac{(x-y-r_2(t-\tau))^2}{2\sigma_2^2(t-\tau)}} dy \end{aligned} \tag{33}$$

and

$$\begin{aligned} \frac{\partial}{\partial x} Pr(\mathcal{X}(t) < x | \mathcal{J}(0) = 1) = &\int_{\tau=0}^{t} \frac{\partial}{\partial x} Pr(\mathcal{X}(t) < x | \mathcal{J}(0) = 1, \tau < t) q_{12} e^{q_{11}\tau} d\tau \\ &+ \int_{\tau=t}^{\infty} \frac{\partial}{\partial x} Pr(\mathcal{X}(t) < x | \mathcal{J}(0) = 1, \tau > t) q_{12} e^{q_{11}\tau} d\tau \end{aligned}$$

Figure 7 plots the analytical and NILT2 approximation of the distribution of $\mathcal{X}(1)$ for the Markov modulated Brownian motion with $\mathbf{Q}$, $\mathbf{R}$ and $\mathbf{\Sigma}$ matrices starting from $\mathcal{J}(0) = 1, \mathcal{X}(0) = 0$. We assume that the ≈ 0.05 relative error (e.g., at $x = 1$ exact$= 0.117741$ and NILT2$= 0.112869$) comes from the improper scaling parameter computed according to (26), which is due to the fact that the scaled $h(t(T+\Delta)-\Delta)$ function is much wider than the GDF.

6. Refined approximation of the scaling parameter

Based on our experiments, the relative error of the NILT2 approximation of GDFs is less than 0.005, as demonstrated in Figure 2. If we compare the two examples with Gaussian mixtures in Section 5.2, we have a reasonably good approximation when $m_2 = 5$, i.e., when the two peaks of the PDF are relatively close to each other (with relative error ≈ 0.015 at the limits, c.f. Figure 4) and a much worse one when $m_2 = 15$ and the peaks are farther (relative error ≈ 0.1 at the limits). In these examples, the shape of the curves left of the left side peak and right of the right side peak are similar to the GDF, but their $\sqrt{\frac{m_2 m_0 - m_1^2}{m_0^2}}$ parameters differ a lot, consequently, according to (26), different scaling parameters are used for their NILT2 approximation.

To address this issue, we propose a refined method, which approximates the scaling parameter based on the local behaviour of the function around the point of interest. Algorithm 2 presents the pseudo code of the refined procedure.

Algorithm 2 Refined NILT2 method

Input: $h^{\circledast}(s), T, N, \delta$.

1: Compute m_0, m_1, m_2 as in Algorithm 1.

2: Set Δ according to (26):
$\Delta \leftarrow 4\sqrt{\frac{m_2 m_0 - m_1^2}{m_0^2}} - T$.

3: Approximate $h(T-\delta)$, $h(T)$ and $h(T+\delta)$ with this initial Δ:
$\tilde{h}(T + i\delta) \leftarrow NILT2_{\text{opt}}(e^{-s\Delta} h^{\circledast}(s), T + \Delta + i\delta, N), \quad i = \{-1, 0, 1\}$.

4: Approximate the shape of $h(t)$ around T:
$a \leftarrow \frac{\log(\tilde{h}(T-\delta)) - 2\log(\tilde{h}(T)) + \log(\tilde{h}(T+\delta))}{2\delta^2}$.

5: Refine Δ based on the local shape:
$$\Delta \leftarrow \begin{cases} 4\sqrt{\frac{-1}{2a}} - T, & \text{if } a < -\frac{m_0^2}{2(m_2 m_0 - m_1^2)}, \\ 4\sqrt{\frac{m_2 m_0 - m_1^2}{m_0^2}} - T, & \text{if } a > -\frac{m_0^2}{2(m_2 m_0 - m_1^2)}. \end{cases}$$

6: **return** $NILT2_{\text{opt}}(e^{-s\Delta} h^{\circledast}(s), T + \Delta, N)$.

The refined method is based on the following considerations. When $f(t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2}}$ is the Gaussian(μ, σ^2) PDF with mean μ and variance σ^2, then $m_0 = 1, m_1 = \mu, m_2 = \mu^2 + \sigma^2$ (where $m_i = \int_{-\infty}^{\infty} t^i f(t) dt$) and $\sqrt{\frac{m_2 m_0 - m_1^2}{m_0^2}} = \sigma$, furthermore, $\log(f(t)) = \frac{-1}{2\sigma^2} t^2 + \frac{\mu}{\sigma^2} t - \frac{\mu^2}{2\sigma^2} - \log(\sigma) - \frac{1}{2}\log(2\pi)$.

That is, $\log(f(t))$ is a parabola, and the coefficient of its second order term is $a = \frac{-1}{2\sigma^2}$. According to (26), the proposed shifting parameter for $f(t)$ is $\Delta = 4\sqrt{\frac{m_2 m_0 - m_1^2}{m_0^2}} - T = 4\sigma - T$. Consequently, the coefficient of the second order term of $\log(f(t))$ can also be used to compute the optimal shifting as $\Delta_{opt} = 4\sigma - T = 4\sqrt{\frac{-1}{2a}} - T$.

If $\log(h(t)) \approx g(t) = at^2 + bt + c$, in $t \in (T - \delta, T + \delta)$, i.e., $g(t)$ is a parabola which closely approximates $\log(h(t))$ around the point of interest, then

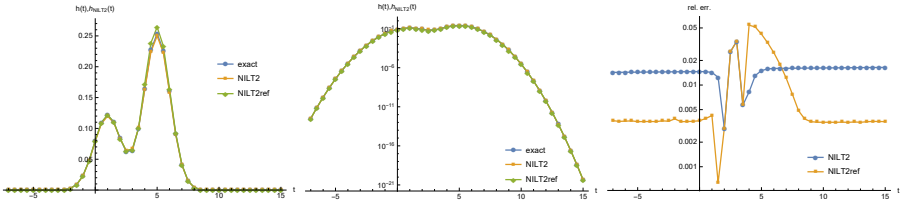

Figure 8: Refined NILT2 approximation of mixture of normal distributions with $I = 2$, $m_1 = 1$, $\sigma_1^2 = 1.2$, $m_2 = 5$, $\sigma_2^2 = 1.1$, $p_1 = 1/3$, $p_2 = 2/3$

$\frac{d^2}{dt^2}\log(h(t))|_{t=T} \approx \frac{d^2}{dt^2}g(t)|_{t=T} = 2a$ and the second derivative of $g(t)$ at T can be approximated as $2a \approx \frac{g(T-\delta)-2g(T)+g(T+\delta)}{\delta^2}$. That is, we approximate $\log(h(t))$ with a parabola in the $(T-\delta, T+\delta)$ interval, approximate the coefficient of its second order term using order 2 finite difference coefficients, and compute Δ_{opt} based on this approximated coefficient a. Unfortunately, when the parabola approximating $\log(h(t))$ opens upward, a is positive and $4\sqrt{\frac{-1}{2a}}$ is complex. That is, we need to restrict the application of the local shape based computation of Δ_{opt} for some ranges of parameter a. Our numerical investigations indicate that the approximation of shape parameter a is numerically sensitive when $a > -\frac{m_0^2}{2(m_2 m_0 - m_1^2)}$. (e.g., in between the peaks of the examples in Section 5.2). As a consequence, we apply the local shape based setting of Δ_{opt} only, when the approximation of a is reasonably accurate, i.e., when $a < -\frac{m_0^2}{2(m_2 m_0 - m_1^2)}$. That is,

$$\Delta_{opt} = \begin{cases} 4\sqrt{\frac{-1}{2a}} - T, & \text{if } a < -\frac{m_0^2}{2(m_2 m_0 - m_1^2)}, \\ 4\sqrt{\frac{m_2 m_0 - m_1^2}{m_0^2}} - T, & \text{if } a > -\frac{m_0^2}{2(m_2 m_0 - m_1^2)}. \end{cases} \tag{34}$$

The computational complexity of Algorithm 2 is characterized by approximately $4 \cdot 20N + 3$ evaluations of $h^{\circledast}(s)$; 3 evaluations happen in line 1, $\approx 3 \cdot 20N$ in line 3 and $\approx 20N$ in line 6. That is, the computational time of Algorithm 2 is approximately 4 times longer than the one of Algorithm 1.

7. Numerical experiments with the refined method

Figure 8, 9, 10, and 11 plot also the refined NILT2 approximation for the examples depicted in Figure 3, 4, 6, and 7.

The relative error curves suggest that the refined method gives better approximates when the logarithm of the function is concave and, according to (34), it resorts to the same approximation as the global NILT2 method when the logarithm of the function is convex. At the transition between the concave and convex intervals the refined method is subject to unpredictable numerical errors.

Based on these observations, for an unknown function, a human assisted NILT approach could be to plot the logarithm of the NILT2 approximation and

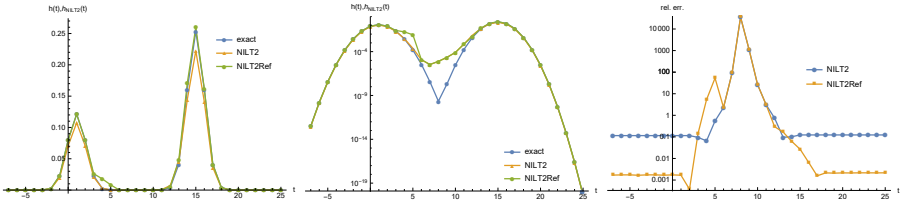

Figure 9: Refined NILT2 approximation of mixture of normal distributions with $I = 2$, $m_1 = 1$, $\sigma_1^2 = 1.2$, $m_2 = 15$, $\sigma_2^2 = 1.1$, $p_1 = 1/3$, $p_2 = 2/3$

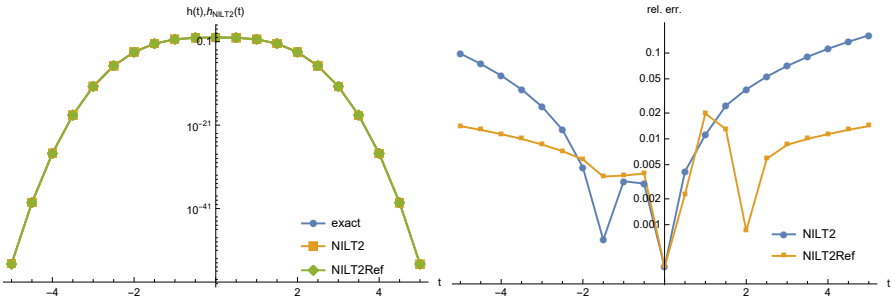

Figure 10: Refined NILT2 approximation $f(t) = e^{-|x|^3}$ and its relative error

for the steadily decaying convex parts refine the approximation using the refined method.

8. Conclusion

The paper proposes a NILT method which optimizes the parameters, referred to as the shifting and the scaling parameters, based on the Laplace domain function and the point where the inverse is computed. The shifting parameter is set based on numerical optimization of a convex function, while the scaling parameter is set heuristically. One of the heuristic setting is based on the global normalized variance of the object function, which is accurate for functions similar to GDFs. The other heuristic setting, is based on the local approximation

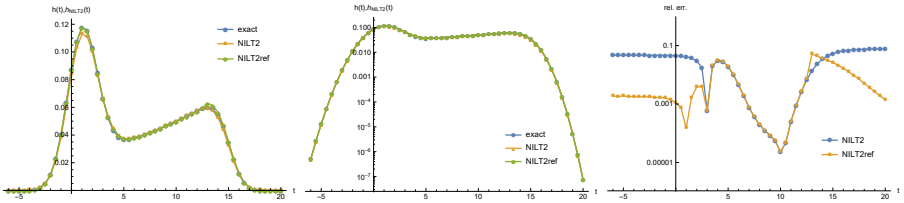

Figure 11: Distribution of the Markov modulated Brownian motion with $\mathbf{Q}$, $\mathbf{R}$ and $\mathbf{\Sigma}$ at $t = 1$ and its refined NILT2 approximation

of the object function. The second approach is more accurate when the object function is concave and its shape differs from GDFs.

The Mathematica implementation of the proposed method is available at `https://github.com/ghorvath78/iltcme`.

9. Declarations

9.1. Ethical Approval

Not Applicable.

9.2. Availability of supporting data

The manuscript has no associated data.

9.3. Competing interests

The authors have no competing interest as defined by Springer, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

9.4. Funding

This work is partially supported by the Hungarian Scientific Research Fund OTKA K-138208 project.

9.5. Authors' contributions

A. Meszaros and M. Telek wrote the main manuscript text, participated in the implementation of the proposed procedures and prepared the figures. Both authors reviewed the manuscript.

9.6. Acknowledgments

Not Applicable.

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

[2] Almousa, S. A.-D., Horváth, G., and Telek, M. (2022). Parameter estimation for Markov modulated fluid arrival processes. *Performance Evaluation*, 157-158:102316.

[3] Asmussen, S. (1995). Stationary distributions for fluid flow models with or without brownian noise. *Communications in Statistics. Stochastic Models*, 11(1):21–49.

[4] Cai, N., Kou, S. G., and Liu, Z. (2014). A two-sided laplace inversion algorithm with computable error bounds and its applications in financial engineering. *Advances in Applied Probability*, 46(3):766âĂŞ789.

[5] Horváth, G., Rácz, S., and Telek, M. (2004). Analysis of second-order Markov reward models. In *The International Conference on Dependable Systems and Networks, DSN/PDS 2004*, pages 845–854, Florence, Italy. IEEE CS Press.

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

[7] Horváth, I., Mészáros, A., and Telek, M. (2022). Numerical inverse Laplace transformation beyond the Abate-Whitt framework. *Journal of Computational and Applied Mathematics*, page 114651.

[8] Matsuzuka, I., Nagasawa, K., and Kitahama, A. (1999). A proposal for two-sided laplace transforms and its application to electronic circuits. *Applied Mathematics and Computation*, 100(1):1–11.

[9] Petrella, G. (2004). An extension of the euler laplace transform inversion algorithm with applications in option pricing. *Operations Research Letters*, 32(4):380–389.