Accuracy improvement of the FFT-based numerical inversion of Laplace transforms

A. Yonemoto, T. Hisakado and K. Okumura

Abstract: The authors present an improvement on the FFT-based numerical inversion of Laplace transforms. Since inversions obtained by the FFT-based method contain large errors, especially for the latter half of the computed region, only the former half is acceptable. The truncation error, which is the greater part of the errors, is analysed and an acceleration method is proposed that uses a property of the complex frequency variable $s$. This acts as the differential operator in the time domain. The errors are significantly reduced by the proposed method, and thus the inversions become acceptable for almost the entire region.

1 Introduction

Various methods for the numerical inversion of Laplace transforms have been proposed [1, 2] using Legendre functions [3], Laguerre functions [4], Fourier series [5] and others [6-9]. Among them, only with the FFT-based method is the corresponding numerical Laplace transform observed. A pair of the numerical transform enables an efficient method to analyse nonlinear circuits by dealing with linear elements in the frequency domain in the knowledge that the pair has small valid regions [10, 11].

The FFT-based method has originally been introduced by Dubner and Abate [12] although their method uses only the Fourier cosine series. Afterwards, several authors have devised the so-called FFT-based method by using Fourier sine series as well [5, 13-16]. As some of the authors pointed out, the accuracy of the FFT-based method is better than the Fourier cosine series method. Indeed, the valid region of the obtained inversion is [0, $T/2$] for the Fourier cosine series method, and [0, $T$] for the FFT-based method, for which the period is $2T$. However, the FFT-based method still involves fairly large errors, and consequently, we must discard almost the latter half of the calculated result. In this paper, we analyse the errors in more detail, and propose a simple yet effective method to extend the valid range to [0, $2T$].

We first analyse the errors involved in the FFT-based method, especially for the truncation error which causes the large errors. Based on the error analysis, we propose an acceleration method [17] which effectively reduces the errors using a property of the complex frequency variable $s$ that acts as the differential operator in the time domain. We also analyse the errors of our proposed method particularly for $F(s) = 1$. Numerical examples show that the truncation error is remarkably reduced, and the obtained inversions are valid for almost entire region [0, $2T$].

2 FFT-based numerical inversion of Laplace transforms

Let $f(t)$ be a real function of $t$, with $f(t) = 0$ for $t < 0$, the Laplace transform of $f(t)$ is defined by

$$F(s) = \mathcal{L}[f(t)] = \int_{0}^{\infty} f(t)e^{-st}dt$$

(1)

where $s$ is the complex frequency variable. The inverse Laplace transform of the complex function $F(s)$ is defined by the Bromwich integral as

$$f(t) = \mathcal{L}^{-1}[F] = \frac{1}{2\pi j} \int_{a-j\infty}^{a+j\infty} F(s)e^{st}ds$$

(2)

where $a$ is an arbitrary real number which is greater than the real part of any singularity of $F(s)$, and $j$ denotes the imaginary unit $\sqrt{-1}$.

Durbin [5] and some other authors [13-16] have proposed the following formula for numerical inversion of a Laplace transform $F(s)$: for $0 \leq t < 2T$,

$$f_N(t) = \frac{e^{st}}{2T} \sum_{k=-N}^{N-1} \hat{F}(a + ik/T) e^{ikT}$$

(3)

$$= \frac{e^{st}}{T} \left\{ \sum_{k=0}^{N-1} \hat{F}(a + ik/T) e^{ikT} \right\} - \hat{F}(a)$$

(4)

Equation (3) is obtained by applying the trapezoidal quadrature to the Bromwich integral (2) and by truncating the resultant infinite sum. The formula (4) is the so-called FFT-based numerical inversion of $F(s)$, in which the summation can be computed by the FFT algorithm for equally spaced points

$$t = nh, \quad n = 0, 1, \cdots, N - 1, \quad h = \frac{2T}{N}$$

(5)

Figures 1-4, respectively, show numerical inversions of the Laplace transform of the step, the cosine, the sine function, and $F(s) = 1/\sqrt{s}$ computed by (4). The errors for the first two inversions become quite large especially for $t > T$. Hence we should discard the latter half of the result, $f_N(t)$ for $t > T$, in general cases if we cannot predict the size of the errors. In addition, there are some types of functions which...
Sections, we analyse the errors and propose another error reduction method.

3 Error analysis of the FFT-based method

The FFT-based numerical inversion of Laplace transforms involves two kinds of errors excluding the computational round off errors. They are the discretisation error \( e_1(t) \) and the truncation error \( e_2(t) \) introduced in the derivation of (3) as

\[
f_s(t) = f(t) + e_1(t) + e_2(t)
\]

The discretisation error \( e_1(t) \) originates from the application of the trapezoidal quadrature to the Bromwich integral, namely, the discretisation of \( F(s) \) in the \( s \) domain. The discretisation causes the periodisation in the time domain, and hence \( e_1(t) \) is given by [5]

\[
e_1(t) = \sum_{k=1}^N f(t + 2kT)e^{-2\pi ik}
\]

Suppose \( f(t) \) is bounded by a constant \( C \) as \( |f(t)| \leq C \), then it can be easily shown that \( e_1(t) \) is also bounded as

\[
|e_1(t)| \leq \sum_{k=1}^N Ce^{-2\pi ik} = \frac{C}{e^{2\pi i} - 1}
\]

Since the value of \( aT \) is usually taken to be 3-5, this error bound is small enough. For example, with \( aT = 3.5 \), we have

\[
|e_1(t)| \leq \frac{C}{e^{2\pi i} - 1} = 0.000912C
\]

Therefore, this error does not contribute to the large errors in the preceding examples for which \( C = 1 \).

The truncation error \( e_2(t) \) stems from the truncation of the infinite sum resulting from the trapezoidal quadrature, and is given by

\[
e_2(t) = \frac{e^{at}}{2T} \sum_{|k| \geq N} F(a + \frac{ik}{T})e^{ikT}
\]

To evaluate \( e_2(t) \), let us define a function \( E_2(t) \) as the absolute summation of \( e_2(t) \) by

\[
E_2(t) = \frac{e^{at}}{2T} \sum_{|k| \geq N} |F(a + \frac{ik}{T})|
\]
Suppose $F(s) = 1/s^m$ (m is a positive real number), $E_2(t)$ is then given by

$$E_2(t) = \sum_{k=-\infty}^{\infty} F(a + jkT) \left( e^{j\pi/k^m} - 1 \right)$$

and the bounds for $E_2(t)$ are given by, since $aT \neq 0$,

$$|e_2(t)| \leq E_2(t) < \sum_{k=-\infty}^{\infty} \frac{1}{k^m} \left( e^{j\pi/k^m} - 1 \right)$$

If $T < \pi(N-1)$, this upper bound for $E_2(t)$ becomes smaller for the larger value of $m$. In other words, since the high-frequency components are small for $1/s^m$ if the value of $m$ is large, the truncation error becomes small as well. The same argument would be concluded if $F(s) = \mathcal{L}(a^s), s \to \infty$. 

4 Acceleration of convergence

4.1 Acceleration method

The above error analysis implies that $R(s) = 1/s^i$ (i is a positive integer) can be inverted by the FFT-based method with better accuracy than $R(s)$, since the factor $1/s^i$ suppresses the high-frequency components of $F(s)$ and thereby the truncation error becomes small. For example, let us consider the Laplace transform of the sine function. It is the Laplace transform of the cosine function multiplied by $as$, and accordingly, the error for the sine function is expected to be smaller than that for the cosine function. This expectation is confirmed by Figs. 2 and 3.

Here, we recall that the complex frequency variable $s$ corresponds to the differential operator in the time domain, namely

$$f(t) = \mathcal{L}^{-1}[F(s)] = \mathcal{L}^{-1}\left[ s^i \cdot \frac{F(s)}{s^i} \right]$$

We propose the use of this equality to compute the inversion of $F(s)$ by following steps:

Step 1. Compute $G(s) = F(s)e^{j\pi/k^m}$.

Step 2. Invert $G(s)$ by the FFT-based method and have $g_N(hn)$ for $n = 0, \ldots, N-1$. The function $g_N(t)$ is defined as (3) by

$$g_N(t) = \sum_{k=0}^{N-1} a_k e^{j\pi/k^m}$$

Step 3. Compute numerical $i$th differential of $g_N(hn)$ by

$$\mathcal{F}^{-1}[g_N(hn)] = \frac{1}{n^i} \mathcal{F}^{-1}[g_N(h(n + i/2 - f))]$$

for odd $i$

$$\mathcal{F}^{-1}[g_N(hn)] = \frac{1}{n^i} \mathcal{F}^{-1}[g_N(h(n + i/2 - f))]$$

for even $i$

Figure 5 illustrates maximum absolute errors, $\max |\hat{f}_N(t) - f(t)|$, for the step function $F(s) = 1/s$ with $i = 0, 1, \ldots, 7$. Note that it reverts to the conventional FFT-based method when $i = 0$. The errors are decreasing along with $i$ up to $i = 4$ for the moderate number of points $N < 1024$, and up to $i = 3$ for $N > 1024$. For $i$ more than 4 or 3, the errors become large and are mostly the round-off errors originating from the numerical $i$th differentiation. Thus we can confirm the effectiveness of the proposed method for the step function with $i < 4$.
4.2 Error analysis of the proposed method

We analyse the errors of the proposed method particularly for $F(s) = 1/s$. In this case, we have

$$y_n(t) = g(t) + e_1(t) + e_2(t) \ldots (24)$$

$$g(t) = e^{-t} \int_0^t e^{\alpha N} \frac{1}{N} e^{-\alpha h N} \frac{1}{h} \ldots (25)$$

$$e_1(t) = \sum_{k=1}^{\infty} \frac{(t + 2kt)^i}{i!} e^{-2\alpha R} \ldots (26)$$

$$e_2(t) = e^{\omega_T} \sum_{|k| \leq N} \frac{1}{2T} \sum_{|j| \leq N} e^{i kn/T} \ldots (27)$$

Since $g(t)$ and $e_1(t)$ are monotonic smooth functions between the sampling points $t = nh$, the $i$th difference of them can be approximated by the $i$th differential as

$$\int_0^t \frac{1}{h} d\frac{d^i g(hn)}{dh^i} = g(hn) + O(h^i) \ldots (28)$$

$$\int_0^t \frac{1}{h} d\frac{d^i e_1(hn)}{dh^i} = g(hn) + O(h^i) \ldots (29)$$

Thus, together with (19) and (24), we have

$$\int_0^t f_n(hn) = g(hn) + \int_0^t \frac{1}{h} d\frac{d^i e_1(hn)}{dh^i} + O(h^i) \ldots (30)$$

On the other hand, $e_2(t)$ is an oscillatory function between the sampling points as shown in Fig. 6 because it is composed of the truncated high-frequency components, and hence the $i$th difference, $d^i e_2(t)$, cannot be approximated by the $i$th differential. Instead, we evaluate $d^i e_2(t)$ by (20) and (21). Since $d^i e_2(hn)$ is given by less than $2^i$ times additions/subtractions of $e_2(hn)$ in the relations, we have

$$\left| \frac{1}{h^i} \sum_1^N e_2(hn) \right| \leq 2^i \left| \frac{1}{h^i} \sum_1^N e_2(hn) \right| \ldots (31)$$

From (27) and (16) with $m = i + 1$,

$$\frac{1}{h^i} \sum_1^N e_2(hn) = \frac{e^{\omega_T}}{h^i} \int_0^T \frac{1}{(N-1)^i} \ldots (32)$$

$$\left| \frac{1}{h^i} \sum_1^N e_2(hn) \right| = \frac{e^{\omega_T}}{h^i} \int_0^T \frac{1}{(N-1)^i} \ldots (33)$$

In Fig. 5, the dashed line gives this bound, $e^{\omega_T}(i+1)$. Although the bound is an overestimation of the truncation error, the declining of the error along with $i$ is well captured.

5 Numerical examples

5.1 Comparison with the FFT-based method

Figures 7-10 illustrate the numerical inversions of the step, the cosine, the sine function, and $F(s) = 1/s$ by the proposed method.

Fig. 7 Proposed inversion of the step function $F(s) = 1/s$, $(N = 256, T = 1, a = 3.5)$

Fig. 8 Proposed inversion of the cosine function $F(s) = \cos(x^2 + \omega^2)$, $(N = 256, T = 1, a = 3.5)$

Fig. 9 Proposed inversion of the sine function $F(s) = \sin(x^2 + \omega^2)$, $(N = 256, T = 1, a = 3.5)$

The proposed method with $N = 256$, $i = 4$, 1, 0. The truncation errors are significantly reduced so that the line for $i = 4$ and the line for the analytical method are overlapped in the Figures. 

Looking closely at Fig. 7, the maximum error is 0.00149 at $n = N - 1$. If we set the relative error tolerance to be 0.5%, the inversion by the proposed method is valid for the entire region $0 < t < 2$, while that by the FFT-based method is valid only for $0.09 < t < 1.09$, namely, the former half of the region. Furthermore, as shown in Fig. 10, the proposed method enables the inversion of functions which cannot be inverted by the FFT-based method, although first and last several points contain errors.

To see the applicability of the proposed method, we compare it with the conventional method on various types of functions used in [1]. The accuracy of numerical inversions is measured in their root-mean-square deviation

$$L_N = \left[ \frac{1}{N-1} \sum_{n=1}^{N-1} (f_n(h_n) - f(h_n))^2 \right]^{1/2}$$

where $n = 0$ is excluded because it is a singular point for some of the functions. Table I presents the results where $L_N$ and $\hat{L}_N$ are represented in their mantissa and exponent of 10. In the Table, the positive number $m$ shows the order of $F_j(s)$:

$$F_j(s) = O \left( \frac{1}{s^m} \right) \quad s \to \infty, \quad j = 1, \ldots, 16$$

The correspondence between $m$ and $L_N, \hat{L}_N$ can be observed in the Table. That is, the error $L_N$ is small for the large value of $m$, and the error is remarkably reduced by the proposed method especially for worse cases, $m \leq 1$. It must be noted that the errors are not improved significantly for $F_0(s)$ and $F_2(s)$, the functions which have discontinuities. This is because their discontinuities are smoothened by the fourth difference used in the proposed method, and cause the errors to be seemingly large at the discontinuities.

The integers $K_1$, $K_2$ in the last two columns are the minimum multipliers for the window length $N$ [5] to achieve comparable small errors to the proposed method. Defining the deviation of the conventional inversion for enlarged window length $K \cdot N$ by

$$L_N^K = \left[ \frac{1}{N-1} \sum_{n=1}^{N-1} (f_k(h_n) - f(h_n))^2 \right]^{1/2}$$

Table 1: Functions used in comparisons and the results ($N = 256$, $T = 15$, $\alpha T = 3.5$, $i = 4$)

<table>
<thead>
<tr>
<th>$f(n)$</th>
<th>$g(n)$</th>
<th>$m$</th>
<th>$L_N$</th>
<th>$\hat{L}_N$</th>
<th>$K_1$</th>
<th>$K_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_1(s) = \frac{1}{\sqrt{s}}$</td>
<td>$f(t) = \Delta(t)$</td>
<td>1</td>
<td>4.03(+0)</td>
<td>6.17(-4)</td>
<td>—</td>
<td>608</td>
</tr>
<tr>
<td>$F_2(s) = \frac{1}{\sqrt{\pi}}$</td>
<td>$g(t) = \sin(\alpha t)$</td>
<td>1/2</td>
<td>1.90(+1)</td>
<td>1.51(-2)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$F_3(s) = \frac{1}{\sqrt{\pi}}$</td>
<td>$f(t) = e^{\alpha t}$</td>
<td>1</td>
<td>4.02(+0)</td>
<td>1.91(-4)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$F_4(s) = \frac{1}{\sqrt{t}}$</td>
<td>$f(t) = e^{-\alpha t} \sin t$</td>
<td>2</td>
<td>1.95(-2)</td>
<td>6.03(-4)</td>
<td>—</td>
<td>2</td>
</tr>
<tr>
<td>$F_5(s) = \frac{1}{\sqrt{t}}$</td>
<td>$f(t) = 1$</td>
<td>1</td>
<td>4.03(+0)</td>
<td>9.25(-4)</td>
<td>—</td>
<td>425</td>
</tr>
<tr>
<td>$F_6(s) = \frac{1}{\sqrt{t}}$</td>
<td>$f(t) = t$</td>
<td>2</td>
<td>4.99(-2)</td>
<td>4.18(-2)</td>
<td>—</td>
<td>1</td>
</tr>
<tr>
<td>$F_7(s) = \frac{1}{\sqrt{t}^2}$</td>
<td>$f(t) = t^{-\alpha}$</td>
<td>2</td>
<td>2.14(-2)</td>
<td>1.03(-3)</td>
<td>—</td>
<td>2</td>
</tr>
<tr>
<td>$F_8(s) = \frac{1}{\sqrt{t}^2}$</td>
<td>$f(t) = \sin t$</td>
<td>2</td>
<td>1.91(-2)</td>
<td>2.99(-3)</td>
<td>—</td>
<td>1</td>
</tr>
<tr>
<td>$F_9(s) = \frac{1}{\sqrt{t}}$</td>
<td>$g(t) = \sin t$</td>
<td>1/2</td>
<td>1.94(+1)</td>
<td>1.30(-2)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$F_{10}(s) = \frac{1}{\sqrt{t}}$</td>
<td>$f(t) = e^{(t-5)}$</td>
<td>1</td>
<td>8.16(-2)</td>
<td>1.98(-2)</td>
<td>—</td>
<td>1</td>
</tr>
<tr>
<td>$F_{11}(s) = \frac{(t)}{\sqrt{t}^2}$</td>
<td>$f(t) = -t + \ln(t)$</td>
<td>—</td>
<td>1.53(+1)</td>
<td>1.44(-2)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$F_{12}(s) = \frac{1}{\sqrt{t}^2}$</td>
<td>$f(t) = \text{square wave}$</td>
<td>1</td>
<td>5.31(+0)</td>
<td>1.24(-1)</td>
<td>—</td>
<td>5</td>
</tr>
<tr>
<td>$F_{13}(s) = \frac{1}{\sqrt{t}^2}$</td>
<td>$f(t) = \cos t$</td>
<td>2</td>
<td>3.24(-2)</td>
<td>4.10(-2)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$F_{14}(s) = \sqrt{s^2 + 1/2} - s^2 + 1/4$</td>
<td>$f(t) = e^{\alpha t - \alpha n}/\sqrt{N}$</td>
<td>1/2</td>
<td>2.42(+0)</td>
<td>1.67(-3)</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$F_{15}(s) = e^{\alpha t} + \frac{1}{2}$</td>
<td>$f(t) = e^{\alpha t - \alpha n}/\sqrt{N}$</td>
<td>—</td>
<td>3.47(-6)</td>
<td>3.53(-5)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$F_{16}(s) = \arctan t$</td>
<td>$f(t) = e^{\alpha t}$</td>
<td>1</td>
<td>4.03(+1)</td>
<td>3.34(-4)</td>
<td>—</td>
<td>1061</td>
</tr>
</tbody>
</table>

In the columns $L_N$ and $\hat{L}_N$, the representation $x \pm y$ means the value $x \times 10^\pm y$. In the columns $K_1$ and $K_2$, a dash means that the value is more than 100,000.
we defined \( K_1, K_2 \) as the minimum integers to satisfy
\[
L_N^{(K_1)} < L_N \quad (38)
\]
\[
L_N^{(K_2)} < 10L_N \quad (39)
\]
In the two columns, a dash shows that the condition was not satisfied for \( K \) less than 100000. These results show that the proposed technique to accelerate the convergence is effective and computation time required to attain the same accuracy would also be much reduced.

5.2 Application to an electric circuit

In this Section, we apply the proposed method to transient analysis of an electric circuit including a transmission line as shown in Fig. 11. The lossless single-phase transmission line is connected to the step voltage source with the resistance \( R_0 \). The Laplace transform of the right-hand end voltage is given by
\[
V(s) = \frac{Z_0}{Z_0 \cosh(q(s)\ell) + R_0 \sinh(q(s)\ell)} E(s) \quad (40)
\]
where \( Z_0 \) denotes the characteristic impedance of the line, \( q(s) = s/c \) is the propagation constant (\( c \) is the light speed), \( \ell \) is the length of the line and \( E(s) \) is the Laplace transform of the step source.

![Electric circuit with a lossless transmission line](image)

Figure 12 shows inversions of \( V(s) \) obtained by the conventional FFT-based method and the proposed method. The parameters are \( Z_0 = 75\Omega, \ Z_0 = 5\Omega, \ \ell = 30m \) and thereby the impedances at the left-hand end of the line are mismatched in order to exaggerate the reflections of traveling waves. The proposed method gives the more accurate result for the latter half. Moreover, it avoids the Gibbs phenomena which are involved by the FFT-based method at the discontinuous points, due to the integration (i.e. the division by a power of \( s \)) of the discontinuous waveform.

6 Conclusion

We analysed the truncation error, which is the most part of the errors involved in the FFT-based numerical inversion of Laplace transforms, and clarified that the bound for the error depends on asymptotic behaviors of the Laplace transforms. Then, we proposed a simple improvement method in which the 4th power of \( 1/s \) suppresses the high-frequency components of \( H(s) \) to reduce the truncation error and the 4th differentiation cancels the 4th power of \( 1/s \) in the time domain. In addition, we derived the upper bound for the improved truncation error for \( |H(s)| = 1/s \) and showed the proposed method is effective. Numerical examples showed that the inversions obtained by the proposed method are acceptable for almost entire region \([0, 27]\).

7 Acknowledgment

The authors wish to thank Dr. S. Ichikawa for collaborating in this work.

8 References