Accumulated error in iterative use of FFT

Narimasa Sasa1a), Susumu Yamada1, Masahiko Machida1, and Toshiyuki Imamura2

1 CCSE, Japan Atomic Energy Agency
178-4-4 Wakashiba, Kashiwa, Chiba 277-0871, Japan

2 Advanced Institute for Computational Science, RIKEN
7-1-26 Minatojima-minami-machi, Chuo-ku, Kobe, Hyogo 650-0047, Japan

a) sasa.narimasa@jaea.go.jp

Received December 11, 2015; Revised April 25, 2016; Published July 1, 2016

Abstract: The roundoff error accumulation in the iterative use of the fast Fourier transform (FFT) is discussed. By using numerical simulations of partial differential equations, we numerically show that the roundoff error in the iterative use of the FFT tends to be accumulated. To avoid a lack of precision, we present numerical simulations in which a quadruple precision floating point number is used, which ensures sufficient precision against the accumulation of the roundoff errors in the FFT.

Key Words: FFT, roundoff error, PDE, symplectic integrator

1. Introduction
The fast Fourier transform (FFT) is one of the most important algorithms in computational science and engineering. It is applied as a key routine in many applications, including the direct numerical simulation of turbulent flows [1] and first principle band calculations [2]. In the numerical analysis of partial differential equations (PDEs), in particular, the FFT is used as a highly accurate method of evaluating derivatives in space, as compared with the difference or finite element methods. That is, when accurate calculations are required, as in the turbulence simulations, spectral methods with the FFT should be used.

The roundoff error in the FFT can be estimated by using theoretical error analysis. For example, for the 1-dimensional FFT implementation with \( L \) sites by the Cooley-Tukey algorithm, the roundoff error is found to grow as \( O(\sqrt{\log L}) \) on average [3] and as \( O(\log L) \) in the worst case [4].

On the other hand, in practical applications, it is difficult to estimate the total effect of the roundoff errors in the FFT. However, in a certain class of numerical scheme for PDEs, we find that an exact conservation law of the numerical scheme is substituted for a method of error estimation, because no quantity, except for the roundoff error, affects the exact conservative quantity of the numerical scheme.

In this study, we performed numerical time evolution of 1-D PDEs by symplectic integrators to monitor the exact conservative quantity and investigate whether the roundoff error affects it. Then, it was found that the roundoff error in the iterative use of the FFT tends to be accumulated as compared with the difference method. As mentioned above, since spectral methods with the FFT are used to perform accurate numerical calculations, the accumulation of the roundoff error may result in the...
reliability of the result being lost. As an attempt to recover the accuracy, we performed a numerical simulation using the FFT of quad precision.

In section 2, we give a simple numerical example, which is an iteration of a pair of the forward and inverse FFT. It is shown that the roundoff error is simply accumulated. The total error should be estimated as the worst case, i.e., as the product of the maximal roundoff error and the iteration number of the FFT. In section 3, we discuss the properties of the exact conservation laws in numerical time evolution by the symplectic integrator. It is known that the quadratic conservative quantities and the total momentum are conserved in numerical time evolutions if the original PDEs have such conserved quantities. In section 4, we describe a numerical simulation of the 1-D Schrödinger equation to show that the error accumulation occurs in a practical problem. We use the exact total probability density conservation law in the numerical time evolution by the symplectic integrator. By comparing the results yielded by the difference method and by using the FFT, we numerically confirm that error accumulation occurs in the FFT case. As another example, a numerical simulation for the nonlinear Schrödinger equation is considered in section 5. We use the momentum conservation law in the time evolution of PDEs by symplectic integrators. Furthermore, we describe numerical simulations using a quadruple precision floating point number to compare the results yielded by the double precision calculation in section 6. Section 7 is devoted to the concluding remarks.

2. Numerical test for iterative use of FFT

In this paper, the forward discrete Fourier transformation, $F$, is defined by

$$F[u_j] = \sqrt{\frac{1}{L}} \sum_{j=1}^{L} u_j e^{-ik_\ell(j-1)\Delta x} \quad (-\frac{L}{2} \leq \ell \leq \frac{L}{2} - 1),$$

and the inverse discrete Fourier transformation, $F^{-1}$, is defined by

$$F^{-1}[a_\ell] = \sqrt{\frac{1}{L}} \sum_{\ell=-L/2}^{L/2-1} a_\ell e^{ik_\ell(j-1)\Delta x} \quad (1 \leq j \leq L),$$

where $L \in \mathbb{N}$, even is a system size, $\Delta x$ is a spatial mesh size, and $k_\ell = 2\pi \ell / (L \Delta x)$.

We numerically demonstrate that the roundoff error in the FFT tends to be accumulated in iterative calculations. To show this, a simple numerical test

![Accumulated roundoff errors](image)

Fig. 1. Accumulated roundoff errors $\Delta_m$ defined by Eq. (3). (a) shows the case of $N=10^2$ and (b) shows the case of $N=10^6$, with $L=256$. 

355
\[ \Delta_m = \max_{1 \leq j \leq L} [(FF^{-1})^N e(m)_j - e(m)_j], \quad (1 \leq m \leq L), \quad (3) \]

is performed by using an FFT routine, where \( N \) is an iteration number and \( e(m)_j \) denotes the \( j \)-th component of an \( L \)-dimensional unit vector \( e(m) \) in which only the \( m \)-th component has the value 1.0, namely, \( [e(m)]_{j=m}=1.0 \) and \( [e(m)]_{j\neq m}=0.0 \). \( \Delta_m \) indicates the accumulated roundoff error after the iterative calculation of the FFT. In an ideal calculation, \( \Delta_m \) should be zero.

Figure 1 shows the results of the numerical test of Eq. (3) in the case where \( L = 256 \) and (a) \( N = 10^2 \) or (b) \( N = 10^6 \). Our calculation was performed using the usual double precision floating point number. We used the FFT package, fft4g.f [5], in the implementation. Figure 1 shows that the roundoff error is accumulated in the iterative use of the FFT in Eq. (3). Furthermore, a pair of \( F \) and \( F^{-1} \) roughly has an error of the order of \( 10^{-16} \) at most per one iteration. Therefore, the results of the numerical simulation suggest that the roundoff error in the FFT is simply accumulated in the iterative calculation. Within our numerical experiments, similar results were obtained by varying \( L \) and \( N \) in Eq. (3), or by replacing the FFT routine.

3. Numerical scheme that preserves conserved quantities

To confirm that the roundoff error in the FFT is accumulated in real numerical simulations of PDEs, we use time integration algorithms that have a conservative property. Namely, we use an exact conservative quantity in a numerical time evolution of PDEs by geometric integrators [6]. Geometric integrators are known to possess the property of preserving the first integral or the geometrical symmetry of the original partial differential equations. For example, in the time evolution by the symplectic Gauss-Legendre Runge-Kutta method, quadratic conservative quantities of the original equations are preserved. In the following section, we describe the estimation of the accumulated error by using a quadratic conservative quantity.

As another example, we use the exact conservative property of the total momentum in the time evolution by the symplectic integrators for PDEs. One of the authors recently showed that time evolutions by symplectic integrators for PDEs have the exact conservative property of the total momentum in addition to the approximate conservative property of the energy [7]. Let us consider the

![Fig. 2. Time evolution of the increment of the total probability density, \( \Delta P \), defined by Eq. (12). (a) \( \Delta P \) is plotted in the time evolution of Eq. (9) without using the FFT. (b) \( \Delta P \) is plotted in the time evolution of Eq. (9) using the FFT. The parameters are chosen as and the initial condition is given by Eq. (13).](image-url)
Hamiltonian system,

\[
\frac{dq_j}{dt} = \frac{\partial H}{\partial p_j},
\]

(4)

\[
\frac{dp_j}{dt} = -\frac{\partial H}{\partial q_j},
\]

(5)

that is supposed to approximate the original spatially continuous PDEs. We apply symplectic integrators to Eqs. (4) and (5) to obtain the numerical time evolution. If the Hamiltonian \( H \) in Eqs. (4) and (5) is supposed to have the translational invariance under the permutation of the variables, \((q_j, p_j) \rightarrow (q_{j+1}, p_{j+1})\), the total momentum of the system

\[
I = \pi \frac{L}{2} - \frac{1}{2} \sum_{\ell=1}^{L/2-1} \left[ \ell(\alpha_{\ell}^2 + \beta_{\ell}^2) - \ell^2 (\alpha_{\ell}^2 + \beta_{\ell}^2) \right]
\]

(6)

is shown to be an exactly conserved quantity in the time evolution by the symplectic integrators, provided that the aliasing error does not occur. In Eq. (6), \((\alpha_{\ell}, \beta_{\ell})\) are defined by using Eq. (1),

\[
\alpha_{\ell} + i\beta_{\ell} = F[q_j + ip_j]
\]

(7)

By monitoring a time evolution of the total momentum Eq. (6) in the numerical simulation, we numerically estimate the accumulated error by the FFT in section 6.

4. Numerical simulation I : 1-D Schrödinger equation

In this section, we present a numerical simulation to show that the roundoff error in the FFT accumulates in the time evolution of PDEs. Let us consider the time evolution of the 1-D Schrödinger equation with a potential

\[
i\frac{\partial \Psi}{\partial t} = -\frac{\partial^2 \Psi}{\partial x^2} + a_1 [1 - \cos(2\pi a_2 x/L \Delta x)] \Psi,
\]

(8)

Fig. 3. Time evolution of the increment of the total momentum, \(\Delta I = I(t) - I(0)\), defined by Eq. (6). (a) \(\Delta I\) is plotted in the time evolution of Eq. (19) without using the FFT. (b) \(\Delta I\) is plotted in the time evolution of Eq. (19) using the FFT. The parameters are chosen as \(L=1024\), \(\Delta x=0.125\), \(\Delta t=0.005\) and the initial condition is given by a 2-soliton solution multiplied by 1.2.
where \( a_1(=0.25) \) and \( a_2(=4.0) \) are real constants. Equation (8) is spatially approximated by the difference method

\[
\frac{i\partial \Psi_j}{\partial t} = -\frac{\Psi_{j+1} - 2\Psi_j + \Psi_{j-1}}{\Delta x^2} + a_1[1 - \cos(2\pi a_2 j/L)]\Psi_j, \quad (1 \leq j \leq L),
\]

(9)

where \( \Psi_j \) is the spatially discretized variable

\[
\Psi_j(t) = \Psi(j\Delta x, t), \quad (1 \leq j \leq L),
\]

(10)

with the cyclic boundary condition

\[
\Psi_{\pm i}(t) = \Psi_{L \pm i}(t) \quad (i = 0, 1, 2, \cdots).
\]

(11)

It is found that the total probability density

\[
P = \sum_j |\Psi_j|^2
\]

(12)

is an exactly conservative quantity in the time evolution of Eq. (9) by the symplectic Gauss-Legendre Runge-Kutta method, as mentioned in the previous section. We use the lowest order one, the midpoint rule, as the time evolution scheme of Eq. (9).

Figure 2 (a) shows the time evolution of the total probability density Eq. (12) from an initial condition

\[
\Psi_j(0) = a_3\exp[i a_4 (j - 1)\Delta x]/\cosh[a_5(j - 1)\Delta x],
\]

(13)

with the parameters \( a_3=1.0, a_4=0.5, a_5=1.0, L=1024, \Delta x=0.125, \) and \( \Delta t=0.005. \) In the numerical simulation, our calculation was performed using the usual double precision floating point number. In Fig. 2 (a), it can be seen that the probability density Eq. (12) is sufficiently conserved and the roundoff error does not accumulate within the double precision calculation.

To compare the result of the difference method, we consider the time evolution by using FFT. We split Eq. (9) into two parts:

\[
\begin{align*}
\frac{i\partial \Psi_j}{\partial t} & = -\frac{\Psi_{j+1} - 2\Psi_j + \Psi_{j-1}}{\Delta x^2}, \\
\frac{i\partial \Psi_j}{\partial t} & = a_1[1 - \cos(2\pi a_2 j/L)]\Psi_j.
\end{align*}
\]

(14)

(15)

In Eq. (14), the difference term is estimated by the FFT. \( S_1(\Delta t) \) denotes the temporal progress operator by Eq. (14),

\[
S_1(\Delta t)\Psi_j(0) = F D_2 F^{-1}[\Psi_j(0)]
\]

(16)

where \( D_2 \) is a diagonal matrix in which the \((\ell, \ell)\)-component is equal to \( \exp[-i\Delta t(2\sin(k\Delta x/2)/\Delta x)^2] \). \( S_2(\Delta t) \) denotes the temporal progress operator by Eq. (15). Then, the second-order explicit symplectic integrator is represented by

\[
\Psi_j(\Delta t) = S_2(\Delta t/2)S_1(\Delta t)S_2(\Delta t/2)\Psi_j(0).
\]

(17)

Figure 2 (b) shows the time evolution of the total probability density Eq. (12) by Eq. (17) from the initial condition Eq. (13). The parameters are the same as those in Fig. 2 (a) and the FFT package; FFTW [8] is used in the implementation of \( S_1(\Delta t) \). In the time evolution by Eq. (17), the probability density Eq. (12) is known to be exactly conserved. However, the increment of the total probability density, \( \Delta P[=P(t) - P(0)] \), is found to increase. Hence, we conclude that the roundoff error in the FFT is accumulated in the numerical simulation.

5. Numerical simulation II: nonlinear Schrödinger equation

As a second example, we consider the numerical time evolutions of the 1-D nonlinear Schrödinger equation
which implementation of the FFT [10] is performed by using Bailey’s double-double algorithm [11].

To compare the results in the double precision case. In the quadruple precision calculation, the

where \( \Psi_j(=\Psi(j\Delta x, t)) \) is the spatially discretized dependent variable.

To confirm that the roundoff error in the FFT is accumulated, we use the exact conservative property of the total momentum in the time evolution by symplectic integrators. To give a symplectic integrator without the FFT, we split Eq. (19) into

\[
\frac{i}{\partial t} \frac{\partial \Psi_j}{\partial t} = -\frac{\Psi_{j+1} - 2\Psi_j + \Psi_{j-1}}{\Delta x^2} - 2|\Psi_j|^2 \Psi_j \quad (1 \leq j \leq L),
\]

where \( \Psi_j=\Psi_j+i\eta_j \). We define that \( S_{11}(\Delta t) \), \( S_{12}(\Delta t) \) and \( S_{13}(\Delta t) \) denote the temporal progress operator by Eq. (20), Eq. (21), and Eq. (22), respectively. Then, the second-order explicit symplectic integrator is given by

\[
\Psi_j(\Delta t) = S_{13}(\Delta t/2)S_{12}(\Delta t/2)S_{11}(\Delta t)S_{12}(\Delta t/2)S_{13}(\Delta t/2)\Psi_j(0).
\]

Figure 3 (a) shows the time evolution of the total momentum Eq. (6) by Eq. (23) with the parameters, \( L=1024, \Delta x=0.125, \) and \( \Delta t=0.005 \). The initial condition is chosen as a 2-soliton solution multiplied by 1.2 with parameters \( P_1 = 0.6 - 0.25i, \) \( P_2 = 0.3 + 0.2i, \) \( \alpha = \gamma = 0, \) \( \beta = 1, \) and \( \delta = 2 \), defined in [9]. In Fig. 3 (a), the time evolution of the total momentum (6) is found to be sufficiently conserved, and the roundoff error is not accumulated in the double precision calculation.

On the other hand, the symplectic integrator that uses the FFT is obtained by splitting Eq. (19) into

\[
\frac{i}{\partial t} \frac{\partial \Psi_j}{\partial t} = -\frac{\Psi_{j+1} - 2\Psi_j + \Psi_{j-1}}{\Delta x^2},
\]

where \( \Psi_j=\Psi_j+i\eta_j \). We define that \( S_{21}(\Delta t) \) and \( S_{22}(\Delta t) \) are defined as the temporal progress operator by Eq. (24) and Eq. (25), respectively. In Eq. (24), the difference term is estimated by the FFT. The second-order explicit symplectic integrator is given by

\[
\Psi_j(\Delta t) = S_{22}(\Delta t/2)S_{21}(\Delta t)S_{22}(\Delta t/2)\Psi_j(0).
\]

Figure 3 (b) shows the time evolution of the total momentum Eq. (6) by Eq. (26). The deviation from the exact conservation is caused by the roundoff error in the FFT, since other errors do not affect the conservation of Eq. (6).

6. Numerical simulation by quadruple precision fast fourier transform

In this section, we describe numerical simulations using a quadruple precision floating point number to compare the results in the double precision case. In the quadruple precision calculation, the implementation of the FFT [10] is performed by using Bailey’s double-double algorithm [11].

We calculate the numerical time evolution of Eq. (19) by the symplectic integrator Eq. (26) in which \( S_{21}(\Delta t) \) is given by the same form as Eq. (16), replacing \( 2\sin(k_t\Delta x/2)/\Delta x \to k_t \). The initial condition is given by Eq. (13) and the parameters are chosen as \( L=1024, \Delta x=0.125, \) and \( \Delta t=0.005 \). Figure 4 (a) shows the time evolution of the total momentum Eq. (6) in the quadruple precision case,
Fig. 4. Time evolution of the increment of the total momentum, \( \Delta I = I(t) - I(0) \), defined by Eq. (6). (a) \( \Delta I \) is plotted in the time evolution of Eq. (19) with using the FFT in the quadruple precision calculation. (b) \( \Delta I \) is plotted in the time evolution of Eq. (19) with using the FFT in the double precision calculation. The parameters are chosen as \( L = 1024 \), \( \Delta x = 0.125 \), and \( \Delta t = 0.005 \) and the initial condition is given by Eq. (13).

In the simulation, the last values of \( \Delta I(t) = I(t) - I(0) \) [\( t = 1000 \)] are \( 1.8 \times 10^{-26} \) in Fig. 4 (a) and \( 1.2 \times 10^{-11} \) in Fig. 4 (b). Then, by a rough estimation, the roundoff errors per one iteration are \( 9.3 \times 10^{-32} \) in Fig. 4 (a) and \( 6.1 \times 10^{-17} \) in Fig. 4 (b). Since these values have the same order of the roundoff error by the use of the FFT at one time, the numerical experiments suggest that the roundoff error in the FFT is simply accumulated. Therefore, for a long time numerical simulation, we should use the quadruple precision floating point number if we need accuracy of at least the order of double precision.

7. Conclusions
In this paper, by using the numerical time evolutions of PDEs by the symplectic integrators, we showed that the roundoff error in the FFT is simply accumulated in iterative calculations. Although the roundoff error in the FFT per one use is as small as the machine epsilon, the roundoff error is accumulated in iterative calculations. Moreover, it becomes not negligible eventually. It was also demonstrated that a quadruple precision calculation is needed in order to avoid a lack of accuracy.

Our investigation relied mainly on the numerical simulations performed in this study. We will perform a theoretical analysis of the subject as a future study. The investigation of the effect of the accumulated error on a real application, such as the direct numerical simulation of turbulent flows, is also a future problem.

References
[1] Y. Kaneda, T. Ishihara, M. Yokokawa, K. Itakura, and A. Uno, “Energy dissipation rate and energy spectrum in high resolution direct numerical simulations of turbulence in a periodic box,” Physics of Fluids, vol. 15, no. 2, pp. L21–L24, February 2003.
[2] M.L. Cohen and V. Heine, “The fitting of pseudopotentials to experimental data and their subsequent application,” Solid State Physics, vol. 24, eds. H. Ehrenreich, F. Seitz, and D. Turnbull, Academic Press, New York, 1970.
[3] J.C. Schatzman, “Accuracy of the discrete Fourier transform and the fast Fourier transform,” *SIAM J. Sci. Comput.*, vol. 17, no. 5, pp. 1150–1166, September 1996.

[4] W.M. Gentleman and G. Sande, “Fast Fourier transforms—for fun and profit,” *Proc. 1966 Fall Joint Computer Conference AFIPS 29*, pp. 563–578, 1966.

[5] Ooura’s General Purpose FFT Package, http://www.kurims.kyoto-u.ac.jp/~ooura/fft.html

[6] E. Hairer, Ch. Lubich, and G. Wanner, *Geometric Numerical Integration*, Springer Verlag, Berlin, 2002.

[7] N. Sasa, “Momentum conservation law in symplectic integrators for partial differential equations,” *J. Phys. Soc. Jpn.*, vol. 83, no. 12, pp. 123003–1–123003–4, November 2014.

[8] FFTW, http://www.fftw.org

[9] R. Hirota, “Exact envelope-soliton solutions of a nonlinear wave equation,” *J. Math. Phys.*, vol. 14, no. 7, pp. 805–809, July 1973.

[10] To appear in http://ccse.jaea.go.jp/ja/download/software.html

[11] Bailey's High-Precision Software Package, http://crd-legacy.lbl.gov/~dhbailey/mpdist/