Analytical formula for numerical evaluations of the Wigner rotation matrices at high spins

Naoki Tajima*

Department of Applied Physics, University of Fukui, Bunkyo 3-9-1, Fukui, 910-8507, Japan
(Received 2 November 2014; published 22 January 2015)

The Wigner d function, which is the essential part of an irreducible representation of SU(2) and SO(3) parameterized with Euler angles, has been know to suffer from a serious numerical errors at high spins, if it is calculated by means of the Wigner formula as a polynomial of cos and sin of half of the second Euler angle. This paper shows a way to avoid this problem by expressing the d functions as the Fourier series of the half angle. A precise numerical table of the coefficients of the series is provided as Supplemental Material.

I. INTRODUCTION

The matrix elements of the rotation operator between angular-momentum eigenstates are called the Wigner D function [1]. When the rotation is specified with the three Euler angles (ϕ,θ,ψ), and the eigenstates are labeled with the magnitude ($j = 0,\frac{1}{2},1,\frac{3}{2},\dots$) and the z component (m or $k = -j,-j+1,\dots,j$) of the angular momentum vector, the D function can be decomposed into three factors,

$$D_{mk}^j(\phi,\theta,\psi) = \langle jm|e^{-i\phi\hat{j}_z}e^{-i\theta\hat{j}_y}e^{-i\psi\hat{j}_z}|jk\rangle$$
$$= e^{-i(m\phi+k\psi)}d_{mk}^j(\theta), \qquad (1)$$

where

$$d_{mk}^j(\theta) = \langle jm|e^{-i\theta\hat{j}_y}|jk\rangle \qquad (2)$$

is the nontrivial part which needs some method to evaluate. It is called the Wigner (small) d function. In the standard phase convention for angular-momentum eigenstates, the matrix elements of $\hat{j}_y$ are purely imaginary and thus $d_{mk}^j(\theta)$ takes on real numbers.

One may be anxious about the fact that Bohr and Mottelson [2] define the rotation matrix D_{mk}^j as the complex conjugate of the right-hand side (r.h.s.) of Eq. (1). However, their definition for d_{mk}^j is identical with Wigner's and the definition of the d function is unique.

The Wigner D function is used in various fields of physics. In some applications, those for large values of j (say, >50) are necessary. An example in nuclear structure physics is the projection from spatially deformed solutions of modern realistic mean-field models to eigenstates of large angular momentum. For toy models, too, one occasionally needs d functions for very large j to confirm the validity of one's picture under extreme conditions. (e.g., a two-rotor model of Refs. [3] and [4]).

The explicit form of the d function is given by the Wigner formula [1], which can be written as

$$d_{mk}^j(\theta) = \sum_{n=n_{\min}}^{n_{\max}} (-1)^n W_n^{jmk}(\theta), \qquad (3)$$

where $n_{\min}$ and $n_{\max}$ are zero or positive integers,

$$n_{\min} = \max(0, k-m), \qquad (4)$$
$$n_{\max} = \min(j-m, j+k), \qquad (5)$$

and

$$W_n^{jmk}(\theta) = w_n^{jmk}\left(\cos\frac{\theta}{2}\right)^{2j+k-m-2n}\left(-\sin\frac{\theta}{2}\right)^{m-k+2n}, \qquad (6)$$

$$w_n^{jmk} = \frac{\sqrt{(j+m)!(j-m)!(j+k)!(j-k)!}}{(j-m-n)!(j+k-n)!(n+m-k)!n!}. \qquad (7)$$

However, this formula suffers from a serious loss of precision at high spins (i.e., for large j) except in the neighborhood of $\theta = 0,\pi$.

For example, assuming that j is a positive integer, $\theta = \frac{\pi}{2}$, and $m = k = 0$, one obtains

$$W_n^{j00}\left(\frac{\pi}{2}\right) = \frac{1}{2^j}\left[\frac{j!}{(j-n)!n!}\right]^2, \qquad (8)$$

which, if j is even, becomes maximum at $n = \frac{j}{2}$,

$$W_{j/2}^{j00}\left(\frac{\pi}{2}\right) = \frac{1}{2^j}\frac{j!}{[(j/2)!]^2} \approx \sqrt{\frac{2}{\pi j}}2^j. \qquad (9)$$

(The Stirling's formula is used in the last approximation.) The absolute value of the d function is not greater than one because it is a matrix element of a unitary operator between normalized states. Wigner's formula expresses the d function as a result of cancellation among terms of possibly huge size, $W_n^{jmk} \sim 2^j$. For $j \sim 54$, the precision of double-precision floating-point numbers (53-bit mantissa) is lost completely, and even quadruple precision float numbers (113-bit mantissa) is lost completely for $j \sim 114$.

A few remedies have been proposed [5,6] but the results are not completely free of the precision loss. In this paper, I investigate the details of this loss of significance and then present a perfect remedy to avoid such numerical difficulty.

II. THE FOURIER-SERIES EXPRESSION OF d FUNCTIONS

One can see easily that terms like $\cos^\lambda\frac{\theta}{2}\sin^\mu\frac{\theta}{2}$ appearing in the r.h.s. of Eq. (6), where λ and μ are zero or positive integers such that $\lambda + \mu \leqslant 2j$, can be expressed as linear

*tajima@quantum.apphy.u-fukui.ac.jp

	Even $m-k$	Odd $m-k$
Even $2j$	0	1
Odd $2j$	$\frac{1}{2}$	$\frac{1}{2}$

combinations of terms $\sin \frac{\kappa\theta}{2}$ (when μ is odd) and $\cos \frac{\kappa\theta}{2}$ (when μ is even) with integers κ in $0 \leqslant \kappa \leqslant \lambda + \mu$ and $\kappa \equiv \lambda + \mu$ (mod 2), by means of repeated applications of the elementary trigonometric identities called the product-to-sum identities or the prosthaphaeresis formulas.

Because the power of $\sin \frac{\theta}{2}$ is $\mu = m - k + 2n$ in Eq. (6), one can see that $\mu \equiv m - k$ (mod 2) and that the d function is an even (odd) function if $m - k$ is even (odd), which can be expanded only with cos (sin) function. This may also be deduced from Eq. (A3), one of the properties of the d function which I have enumerated in the Appendix.

From these considerations, one can conclude that the Fourier expansion of the d function has the following form:

$$d_{mk}^{j}(\theta) = \sum_{\nu} t_{\nu}^{jmk} f(\nu\theta), \tag{10}$$

where

$$f = \begin{Bmatrix} \cos \\ \sin \end{Bmatrix} \quad \text{for} \quad \begin{Bmatrix} \text{even} \\ \text{odd} \end{Bmatrix} m - k, \tag{11}$$

and the summation runs over

$$\nu = \nu_{\min}, \nu_{\min} + 1, \ldots, j, \tag{12}$$

with the values of $\nu_{\min}$ given in Table I.

For example, for $j = \frac{7}{2}$ and $m = -k = \frac{1}{2}$, the Wigner formula (3) gives an expression,

$$d_{\frac{1}{2},-\frac{1}{2}}^{\frac{7}{2}}(\theta) = \sin^7 \frac{\theta}{2} - 12 \cos^2 \frac{\theta}{2} \sin^5 \frac{\theta}{2}$$
$$+ 18 \cos^4 \frac{\theta}{2} \sin^3 \frac{\theta}{2} - 4 \cos^6 \frac{\theta}{2} \sin \frac{\theta}{2}, \tag{13}$$

which can be rewritten in the form (10) as

$$d_{\frac{1}{2},-\frac{1}{2}}^{\frac{7}{2}}(\theta) = -\frac{35 \sin \frac{7\theta}{2} - 5 \sin \frac{5\theta}{2} + 15 \sin \frac{3\theta}{2} - 9 \sin \frac{\theta}{2}}{64}. \tag{14}$$

By utilizing the orthogonality of $\cos \nu\theta$ and $\sin \nu\theta$ over $0 \leqslant \theta \leqslant 4\pi$ (considering that ν can take both integer and half-integer values), one can express the coefficients t_{ν}^{jmk} by an integral

$$t_{\nu}^{jmk} = \frac{1}{2\pi(1 + \delta_{\nu 0})} \int_0^{4\pi} d_{mk}^{j}(\theta) f(\nu\theta) d\theta, \tag{15}$$

where $\delta_{\nu 0} = 1$ for $\nu = 0$ and $\delta_{\nu 0} = 0$ for $\nu = \frac{1}{2}, 1, \frac{3}{2}, \ldots$.

By substituting the d function in Eq. (15) with Eqs. (3)–(7), I have derived a more useful expression for t_{ν}^{jmk} containing

only four elementary operations of arithmetic,

$$t_{\nu}^{jmk} = \frac{2(-1)^{m-k}}{1 + \delta_{\nu 0}} \sum_{n=n_{\min}}^{n_{\max}} (-1)^n w_n^{jmk} \sum_{r=0}^{[\nu - \frac{1}{2}p]} (-1)^r \binom{2\nu}{2r + p}$$
$$\times \frac{1}{2\pi} I_{2(j+\nu-n-r)-m+k-p, 2(n+r)+m-k+p}, \tag{16}$$

where $n_{\min}$ and $n_{\max}$ are those already defined by Eqs. (4) and (5), the square brackets are the floor function, i.e., $[l + x] = l$ for integer l and real x in $[0,1)$,

$$p \equiv |m - k| \pmod 2, \tag{17}$$

i.e.,

$$p = 0 \quad \text{for} \quad k = m, m \pm 2, m \pm 4, \ldots, \tag{18}$$
$$p = 1 \quad \text{for} \quad k = m \pm 1, m \pm 3, \ldots, \tag{19}$$

and

$$I_{\lambda\mu} = \int_0^{2\pi} \cos^\lambda x \sin^\mu x \, dx \tag{20}$$

with zero or positive integers for λ and μ. If both λ and μ are even,

$$I_{\lambda\mu} = \frac{2\pi(\lambda - 1)!!(\mu - 1)!!}{(\lambda + \mu)!!}, \tag{21}$$

while $I_{\lambda\mu} = 0$ otherwise.

Unlike the r.h.s. of Eq. (3), where the terms can have huge sizes and thus the numerical error is a serious problem, the r.h.s. of Eq. (10) is a summation of terms of order one or less and hence the problem is expected to disappear. This can be seen by calculating the integrals of the squares of the both sides of Eq. (10),

$$\int_0^{4\pi} d_{mk}^{j}(\theta)^2 d\theta = \sum_{\nu} \sum_{\mu} t_{\nu}^{jmk} t_{\mu}^{jmk} \int_0^{4\pi} f(\nu\theta) f(\mu\theta) d\theta$$
$$= \sum_{\nu} \frac{4\pi}{2 - \delta_{\nu 0}} (t_{\nu}^{jmk})^2. \tag{22}$$

Because the absolute values of d functions are $\leqslant 1$, the left-hand side is $\leqslant 4\pi$ and, consequently, it holds $|t_{\nu}^{jmk}| \leqslant 2 - \delta_{\nu 0}$.

A further study from the numerical point of view has indicated that the maximum (among all the possible combinations of $m, k,$ and ν) value of $|t_{\nu}^{jmk}|$ is 1 for $j \leqslant 1$, decreases as j increases from an integer to the next half integer, and does not change as j increases from a half integer to the next integer. For the interval $50 \leqslant j \leqslant 100$, the maximum value for integer j behaves as $\approx 1.13/\sqrt{j}$.

III. COMPUTATION OF THE NUMERICAL VALUES OF THE COEFFICIENTS

Unfortunately, Eq. (16) also suffers from a serious loss of significant digits in ordinary floating-point numerical calculations. Indeed, for even integer j, the term having the maximum magnitude among those in the r.h.s. of Eq. (16) occurs at $m = k = 0, n = r = \frac{1}{2}j, \nu = j$, with the maximum

value of

$$\frac{1}{\pi} w_{j/2}^{j00} \binom{2j}{j} I_{2j,2j} \approx \frac{2^{2(j+1)}}{(\pi j)^2}, \qquad (23)$$

which is roughly 2^j times as large as the value given by Eq. (9).

To avoid this problem, I first evaluated the r.h.s. of Eq. (16) rigorously as rational numbers or square root of rational numbers by means of a formula-manipulation software MAXIMA. Numerical values to be used in programs coded in Fortran, C, etc., can be calculated from such rigorous numbers to the full precision of the 64-bit floating-point number.

However, the computation time turned out to be excessively long for large values of j. To speed up the computation, I have changed the method to evaluate Eq. (16) not rigorously but in terms of high-precision floating-point numbers (of MAXIMA). This does not seem to be a major drawback because numerical values are sufficient for most of practical purposes.

I change the precision of floating-point numbers depending on j in such a way that the number of digits equals the common logarithm of the value of Eq. (23) divided by 10^{-18}. It increases with j, reaching 74 digits for $j = 100$. Precise 64-bit floating-point numbers can be obtained simply by truncating the high-precision results.

Empirically, the time necessary to compute all the coefficients t_ν^{jmk} for each j increases as j^4. The new method takes 43 h for $j = 100$ with a personal computer with a CPU Intel core-i7 3960X running at 3.3 GHz, using one physical core.

I use the obtained 64-bit floating-point number coefficients to evaluate the Fourier-series formula for d functions. I provide data files of the numerical values of the coefficients, together with a sample FORTRAN90 program to read the data and calculate the values of the d function, as Supplemental Material [7].

For each value of j, there are $(2j + 1)^2$ possible combinations of the values of m and k (because $-j \leqslant m \leqslant j, -j \leqslant k \leqslant j$). Only about a quarter of them are independent, however, because of the properties of the d function expressed by Eqs. (A4), (A6), and (A7) in the Appendix. Therefore, I consider only such combinations as $m \geqslant 0$ and $k \leqslant |m|$ in the following analysis of the numerical precision.

In other words, the coefficients t_ν^{jmk} have the following symmetry:

$$t_\nu^{jkm} = (-1)^{m-k} t_\nu^{jmk}, \qquad (24)$$

$$t_\nu^{j,-m,-k} = (-1)^{m-k} t_\nu^{jmk}, \qquad (25)$$

$$t_\nu^{j,-k,-m} = t_\nu^{jmk}. \qquad (26)$$

Hence, the numerical data for t_ν^{jmk} in the Supplemental Material are given only for $m \geqslant 0$ and $k \leqslant |m|$.

There are $\approx \frac{1}{2}(j_{max} + \frac{3}{2})^4$ coefficients for $j = 0, \frac{1}{2}, 1, \ldots, j_{max}$. The size of the memory to store them as 64-bit floating-point numbers amounts to 27 (404) MiB for $j_{max} = 50$ (100). My data are given as text files for the sake of compatibility, whose sizes are not very different from the above memory sizes after they are compressed (with the software GZIP).
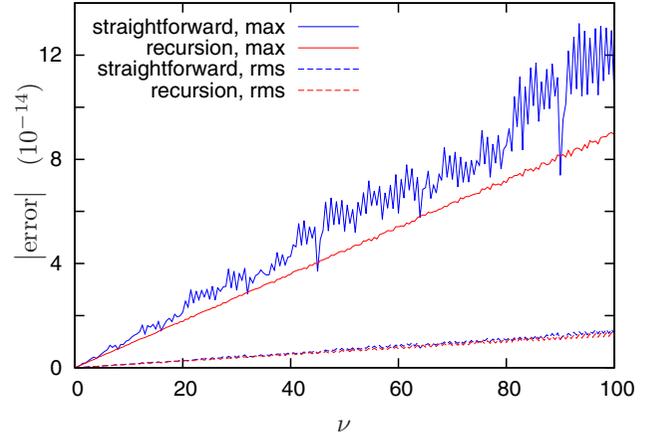
FIG. 1. (Color online) Maximum and root-mean-square errors in the values of $\cos \nu\theta$ and $\sin \nu\theta$, numerically calculated with 64-bit floating-point numbers, over $0 \leqslant \theta \leqslant \pi$. The values of θ are sampled with a uniform spacing of 10^{-5} degree. The abscissa is $\nu = 0, \frac{1}{2}, 1, \ldots, 100$. Recursion means the usage of Eq. (27) together with the initial values described in the following sentences.

The magnitude of the coefficient t_ν^{jmk} becomes smaller for larger values of j. However, even at $j = 100$, 80% (90%) of the coefficients are larger than 10^{-5} (10^{-10}).

Some of the coefficients vanish according to rules unmentioned so far. For example, $t_\nu^{jmk} = 0$ if j is an integer, $m = 0$ and/or $k = 0$, and $j - \nu \equiv 1$ (mod 2). One can prove this using Eq. (A8). I do not use these additional rules but simply give zero values in the data files.

The evaluations of $\cos \nu\theta$ and $\sin \nu\theta$ should be calculated by means of a recursion relation,

$$\begin{pmatrix} \cos(\nu+1)\theta \\ \sin(\nu+1)\theta \end{pmatrix} = \begin{pmatrix} \cos\theta, & -\sin\theta \\ \sin\theta, & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\nu\theta \\ \sin\nu\theta \end{pmatrix}, \qquad (27)$$

which is nothing but the trigonometric (angle) addition theorem. For integer j, the initial values are $\cos 0 = 1$ and $\sin 0 = 0$. For half integer j, one has to calculate, first, the initial values $\cos \frac{\theta}{2}$ and $\sin \frac{\theta}{2}$ and, second, $\cos \theta$ and $\sin \theta$ using identities $\cos \theta = \cos^2 \frac{\theta}{2} - \sin^2 \frac{\theta}{2}$ and $\sin \theta = 2 \sin \frac{\theta}{2} \cos \frac{\theta}{2}$.

Straightforward evaluation of $\cos \nu\theta$ and $\sin \nu\theta$ (i.e., passing the value of $\nu\theta$ to the internal functions cos and sin) requires about j calls to the functions and is computationally very inefficient. Moreover, as shown in Fig. 1, such straightforward evaluation causes slightly larger numerical errors probably due to the loss of significant digits in reducing the value of $\nu\theta$ to an interval such as $[-\frac{\pi}{4}, \frac{\pi}{4}]$, especially when the value of $|\nu\theta|$ is large. The reason why the recursion formula (27) does not suffer from large errors even after hundred steps may be attributed to the fact that the magnitudes of cos and sin functions are always not greater than one.

IV. PRECISION OF THE NUMERICAL VALUES OF THE d FUNCTION

In this section, I compare the errors of the values of $d_{mk}^j(\theta)$ calculated with 64-bit floating-point numbers according to the Wigner formula (3) and the Fourier series expression (10). The errors have been calculated as the differences from the
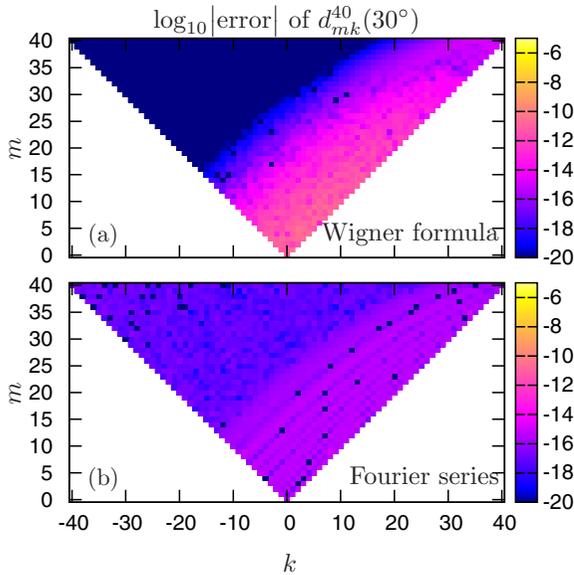
FIG. 2. (Color online) The common logarithm of the error in the 64-bit floating-point numerical value of the d function $d_{mk}^{j}(\theta)$ as a function of m and k for $j = 40$ and $\theta = 30°$. The value of the d function is calculated by means of the Wigner formula in part (a) and the Fourier series expression in part (b). Only a quarter of the possible combinations of m and k are plotted because the error is symmetric in the lines $m = \pm k$. Errors smaller than 10^{-20} are painted with the same color as that for the error of 10^{-20}.

exact values calculated by applying the formula-manipulation software MAXIMA to the Wigner formula.

Figures 2, 3, and 4 show the errors for $\theta = 30°, 60°,$ and $90°$, respectively. The errors are expressed as a function of (m,k) while j is fixed at 40. I have found that similar plots for $\theta > 90°$ look almost indistinguishable from those for $180° - \theta$ except that the sign of k is reversed, as could be foreseen from Eq. (A9).
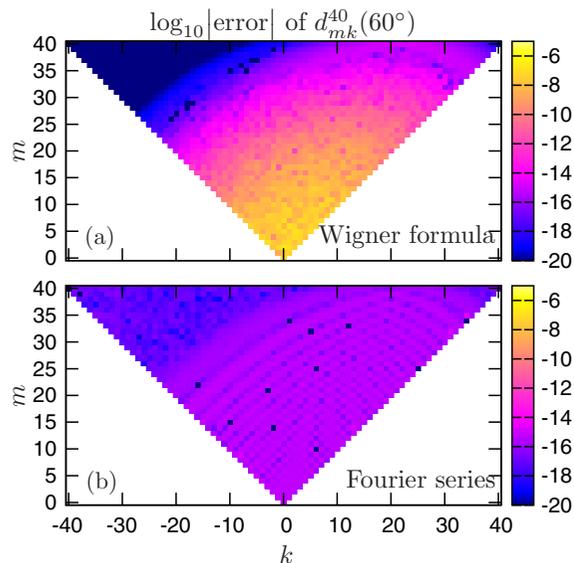
FIG. 3. (Color online) The same as in Fig. 2 but for $\theta = 60°$.
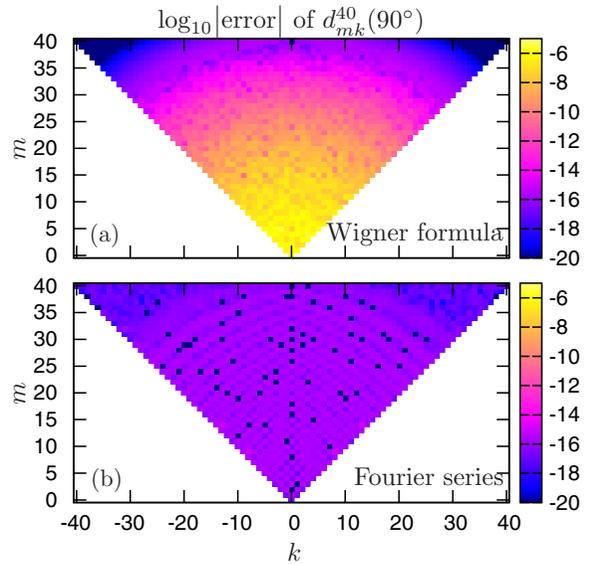
FIG. 4. (Color online) The same as in Fig. 2 but for $\theta = 90°$.

One can see that the Wigner formula results in very large errors, up to $\sim 10^{-5}$ for $\theta \sim 90°$ and $m \sim k \sim 0$, while the Fourier series expression gives precisely 15 digits irrespective of the values of $\theta, j, m,$ and k. I confidently recommend the Fourier series expression over the Winger formula already at $j \sim 40$.

For a special purpose, however, the Wigner formula still has an advantage. For regions of the arguments $\theta \sim 0°$ (180°) and $|m + k| \sim 0$ (2j), the Wigner formula has smaller errors than 10^{-15}, i.e., than the level of the almost constant error of the Fourier series expression. In such regions of the arguments, the magnitude of $d_{mk}^{j}(\theta)$ is very small and a small number of terms dominate in the summation of the Wigner formula, while many terms of order 1 cancel among themselves to give the small value in the Fourier series expression. Therefore, if
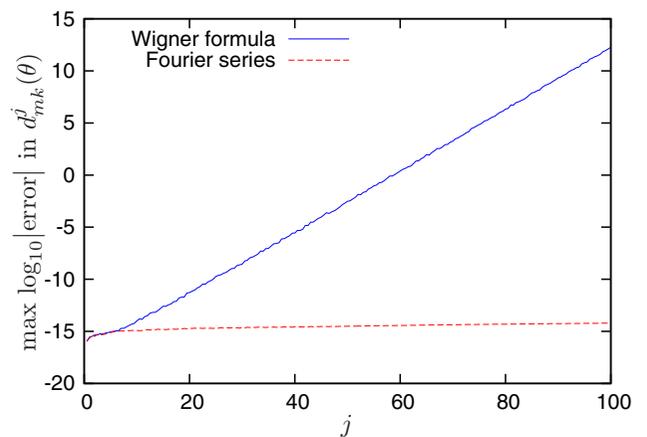
FIG. 5. (Color online) The common logarithm of the maximum error in the numerical value of the d function $d_{mk}^{j}(\theta)$ vs j calculated with the Wigner formula (blue [gray] solid line) and the Fourier series expression (red [gray] dashed line). The maximum is taken over the values of $m, k,$ and θ for each j.

one needs such high precision at those regions, it would be a good idea to develop a program which switches between the two formulas depending on the values of θ, j, m, and k.

In Fig. 5, I compare the maximum errors of $d_{mk}^j(\theta)$ as functions of j. The maximum is taken over the values of θ from $0°$ to $180°$ with an increment of $5°$ and all possible combinations of m and k. For both formulas, the maximum error increases as exponential functions of j. The Wigner formula increases its error, however, far faster than the Fourier series expression. For the interval $20 \leqslant j \leqslant 100$, the error of the Wigner formula can be approximated by $\log_{10} |\text{error}| \approx 0.294 j - 17.2$ while that of the Fourier series expression can be approximated by $\log_{10} |\text{error}| \approx 0.006 j - 14.8$. In other words, by increasing j by one, the error of the Wigner formula is doubled, while that of the Fourier series expression increases only by 1.4%.

V. SUMMARY

I have shown that the Wigner formula for the d function results in intolerable large numerical errors for large values of the angular momentum quantum number j. On the other hand, the Fourier series expression for the d function is shown to be free of such errors, providing precision of $\sim 10^{-14}$ even at $j = 100$. An analytic expression for the coefficients of the Fourier series is given. Their numerical values, which are precise as far as 64-bit floating-point numbers can express, are provided as electric files in Supplemental Material. Sample programs in FORTRAN90 to use the data files are also provided.

APPENDIX: PROPERTIES OF THE d FUNCTION UTILIZED IN THIS PAPER

I enumerate the symmetries of the d function to be referred to in this paper.

First, the unitarity of rotations (i.e., the Hermite conjugate operator is the inverse operator) means

$$d_{mk}^j(\theta) = d_{km}^j(-\theta). \tag{A1}$$

Second, the composition of two rotations can be rewritten as multiplication of matrices to represent them,

$$d_{mk}^j(\theta_1 + \theta_2) = \sum_{\nu=-j}^{j} d_{m\nu}^j(\theta_1) d_{\nu k}^j(\theta_2). \tag{A2}$$

I need three more relations, whose easiest derivation may be to use the Wigner formula (3) as in Ref. [1],

$$d_{mk}^j(\theta) = (-1)^{m-k} d_{mk}^j(-\theta), \tag{A3}$$

$$d_{mk}^j(\theta) = d_{-k,-m}^j(\theta), \tag{A4}$$

$$d_{mk}^j(\pi) = (-1)^{j+m} \delta_{m,-k}. \tag{A5}$$

From Eqs. (A1) and (A3), one can prove

$$d_{mk}^j(\theta) = (-1)^{m-k} d_{km}^j(\theta), \tag{A6}$$

from Eqs. (A4) and (A6),

$$d_{mk}^j(\theta) = (-1)^{m-k} d_{-m,-k}^j(\theta), \tag{A7}$$

and from Eqs. (A2), (A5), and (A6),

$$d_{mk}^j(\pi + \theta) = (-1)^{j+m} d_{-m,k}^j(\theta), \tag{A8}$$

$$d_{mk}^j(\pi - \theta) = (-1)^{j+m} d_{m,-k}^j(\theta). \tag{A9}$$

[1] M. Rose, *Elementary Theory of Angular Momentum* (John Wiley and Sons, New York, 1957).

[2] A. Bohr and B. Mottelson, *Nuclear Structure* (Benjamin, New York, 1969), Vol. 1.

[3] N. Tajima, J. Phys.: Conf. Ser. **445**, 012014 (2013).

[4] N. Tajima and T. Otsuka, Phys. Rev. C **84**, 064316 (2011).

[5] C. Choi, J. Ivanic, M. Gordon, and K. Ruedenberg, J. Chem. Phys. **111**, 8825 (1999).

[6] H. Dachsel, J. Chem. Phys. **124**, 144115 (2006).

[7] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevC.91.014320 for the numerical values of the coefficients of the Fourier series of the d functions and sample programs to use the data.