A new numerically stable implementation of the T-matrix method for electromagnetic scattering by spheroidal particles

W.R.C. Somerville, B. Auguié, E.C. Le Ru *

The MacDiarmid Institute for Advanced Materials and Nanotechnology, School of Chemical and Physical Sciences, Victoria University of Wellington, PO Box 600, Wellington 6140, New Zealand

A R T I C L E   I N F O

A B S T R A C T

We propose, describe, and demonstrate a new numerically stable implementation of the extended boundary-condition method (EBCM) to compute the T-matrix for electromagnetic scattering by spheroidal particles. Our approach relies on the fact that for many of the EBCM integrals in the special case of spheroids, a leading part of the integrand integrates exactly to zero, which causes catastrophic loss of precision in numerical computations. This feature was in fact first pointed out by Waterman in the context of acoustic scattering and electromagnetic scattering by infinite cylinders. We have recently studied it in detail in the case of electromagnetic scattering by particles. Based on this study, the principle of our new implementation is therefore to compute all the integrands without the problematic part to avoid the primary cause of loss of precision. Particular attention is also given to choosing the algorithms that minimise loss of precision in every step of the method, without compromising on speed. We show that the resulting implementation can efficiently compute in double precision arithmetic the T-matrix and therefore optical properties of spheroidal particles to a high precision, often down to a remarkable accuracy (10^{-10} relative error), over a wide range of parameters that are typically considered problematic. We discuss examples such as high-aspect ratio metallic nanorods and large size parameter (≈ 35) dielectric particles, which had been previously modelled only using quadruple-precision arithmetic codes.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

The theoretical modelling of light scattering by particles has a long history, with the first formal analytical solutions provided by Lorenz and Mie for spherical scatterers [1]. Despite the success of the Lorenz–Mie theory in describing many aspects of light scattering in a wide variety of contexts, ranging from astrophysics to biology (for a comprehensive review, see the anniversary monograph of the Mie theory by Mischenko [1]), there are fundamental aspects of the interaction between an electromagnetic wave and a particle that can only be accounted for with a non-spherical particle geometry [2]. With arguably the simplest departure from a spherical shape, spheroidal particles have been thoroughly investigated with a variety of methods, starting from the simplest Rayleigh approximation based on an electrostatics solution [3], which has been refined to include higher order corrections [4] such as radiative corrections [5,6] and dynamic depolarisation [7,8] through to exact analytical [9–12] or numerical [13–15] solutions of the fully retarded Maxwell equations. Spheroids have thus provided the archetype problem of scattering by a non-spherical particle against which a new numerical method can be tested for accuracy and performance. Indeed, the existence of

* Corresponding author. Tel.: +64 44635233.
  E-mail addresses: walter.somerville@vuw.ac.nz (W.R.C. Somerville),
baptiste.auguie@vuw.ac.nz (B. Auguié), eric.leru@vuw.ac.nz (E.C. Le Ru).

rigorous, practically analytical solutions of Maxwell equations [11], supplemented by a wealth of reliable benchmark results using the T-matrix method [16], provide ample justification for the study of spheroidal scatterers, even as a first approximation of actual particle shapes.

It is in this context that we revisit here the problematic numerical calculation of T-matrix integrals for spheroids with Waterman's extended boundary-condition method (EBCM), widely used in theoretical studies for its similarities with the analytical Lorenz–Mie theory, as well as in applied problems thanks to the availability of efficient numerical implementations [17]. This work aims to improve the calculation of light scattering by spheroids in the EBCM, such that it may be considered a reliable benchmark over a large range of particle sizes, composition and aspect ratios, as with Lorenz–Mie theory for spheres. In its current formulation [18], the EBCM requires the calculation of matrix elements as integrals over the geometrical shape of the particle. A matrix inversion is then required to obtain the T-matrix. This scheme suffers from a number of numerical problems, especially for large and/or high aspect ratio particles. To date, the only way around these is to carry out computations with extended precision arithmetic [19,20], although these increase computational complexity and times. Revisiting the original EBCM formulation, Waterman was first to remark [21] that in the case of spheroids, the numerical evaluation of such integrals may suffer a dramatic loss of precision associated with exact cancellations of (large parts of) the integrand. A formal account of such cancellations, and their detrimental consequence on the numerical precision of the EBCM, was recently given in [20]. From this study, it was also revealed that, provided the matrix elements for spheroids could be accurately determined, the resulting matrices would be well-behaved, and in particular could be inverted without introducing additional numerical problems, even at large aspect ratios.

The detailed study of the cancellations highlighted that particular terms in the Laurent series expansion of the integrand are responsible for the problematic behaviour; in this work we seek to reformulate the integrals such that these problematic terms are removed. The integrals can then be computed to high accuracy using standard double precision arithmetic over a wide range of particle parameters.

2. T-matrix method

2.1. General principle

The T-matrix method [22] is a general approach to electromagnetic scattering aimed at finding the linear relationship between incident and scattered fields for scatterers of arbitrary shape, and for in principle any incident field. The most common approach is to expand the fields (respectively incident, scattered and internal) in a basis of vector spherical wavefunctions (VSWFs), as

$$\mathbf{E}_{\text{Inc}} = E_0 \sum_{\substack{n=1\ldots\infty \\ |m| \le n}} a_{nm}\mathbf{M}_{nm}^{(1)}(k_1\mathbf{r}) + b_{nm}\mathbf{N}_{nm}^{(1)}(k_1\mathbf{r}), \tag{1}$$

$$\mathbf{E}_{\text{Sca}} = E_0 \sum_{\substack{n=1\ldots\infty \\ |m| \le n}} p_{nm}\mathbf{M}_{nm}^{(3)}(k_1\mathbf{r}) + q_{nm}\mathbf{N}_{nm}^{(3)}(k_1\mathbf{r}), \tag{2}$$

$$\mathbf{E}_{\text{Int}} = E_0 \sum_{\substack{n=1\ldots\infty \\ |m| \le n}} c_{nm}\mathbf{M}_{nm}^{(1)}(k_2\mathbf{r}) + d_{nm}\mathbf{N}_{nm}^{(1)}(k_2\mathbf{r}), \tag{3}$$

where k_1 is the wavevector in the surrounding medium, k_2 is the wavevector in the scatterer, $\mathbf{M}^{(1)}$ and $\mathbf{N}^{(1)}$ are the magnetic and electric regular (at the origin) VSWFs, and $\mathbf{M}^{(3)}$ and $\mathbf{N}^{(3)}$ are the irregular magnetic and electric VSWFs that satisfy the radiation condition for outgoing spherical waves (see Ref. [16] for further details). The subscripts m and n index the projected and total angular momentum, respectively.

The matrix $\mathbf{T}$ expresses the linear relation between the coefficients $(\mathbf{p},\mathbf{q})$ and $(\mathbf{a},\mathbf{b})$, as

$$\begin{pmatrix} \mathbf{p} \\ \mathbf{q} \end{pmatrix} = \mathbf{T} \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix}, \tag{4}$$

where the expansion coefficients are grouped in column vectors, e.g. $\mathbf{a} = (a_{nm})$, and $\mathbf{T}$ in principle contains all the information about the scatterer (at a given wavelength). It allows for example for problems of analytical averaging over all orientations [18,23–27] or solving multiple scattering by ensemble of particles [26–28]. The most-used method to calculate the T-matrix is the extended boundary-condition method (EBCM) also called the null-field method, where the incident and internal fields, and the internal and scattered fields are related by some matrices $\mathbf{P}$, $\mathbf{Q}$, as

$$\begin{pmatrix} \mathbf{p} \\ \mathbf{q} \end{pmatrix} = -\mathbf{P} \begin{pmatrix} \mathbf{c} \\ \mathbf{d} \end{pmatrix}, \tag{5}$$

$$\begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} = \mathbf{Q} \begin{pmatrix} \mathbf{c} \\ \mathbf{d} \end{pmatrix}. \tag{6}$$

We then obtain $\mathbf{T}$ from

$$\mathbf{T} = -\mathbf{P}\mathbf{Q}^{-1}. \tag{7}$$

The matrices are typically written in block notation as

$$\mathbf{Q} = \begin{pmatrix} \mathbf{Q}^{11} & \mathbf{Q}^{12} \\ \mathbf{Q}^{21} & \mathbf{Q}^{22} \end{pmatrix}, \tag{8}$$

where each block is an infinite square matrix, which is in practice truncated to a $2N \times 2N$ matrix (N then represents the maximum multipole order n in the series expansions). The matrix elements in $\mathbf{P}$, $\mathbf{Q}$ are given by surface integrals over the particle boundary as derived for example in Ref. [16, Section 5.8].

2.2. Particles with symmetry of revolution

For particles with symmetry of revolution, the expansion coefficients with different m values are entirely decoupled, and one can then solve the problem for each value of m, where m can be viewed as a fixed parameter (which will therefore be implicit in most of our notations). Moreover, we have [16, Section 5.2.2]

$$T_{nk|-m}^{ij} = (-1)^{i+j}T_{nk|m}^{ij} \tag{9}$$

and therefore only $m \geq 0$ values need to be considered in the calculations of **T**. Note that we also have

$$T_{nk}^{12} = T_{nk}^{21} = 0 \quad \text{if } m = 0. \tag{10}$$

In addition, the surface integrals reduce to line integrals and we have recently proposed a number of further simplifications [29].

Following Ref. [29], we define

$$K_{nk}^1 = \int_0^\pi d\theta \sin\theta \pi_n d_k x_\theta \xi_n \psi_k', \tag{11}$$

$$K_{nk}^2 = \int_0^\pi d\theta \sin\theta \pi_n d_k x_\theta \xi_n' \psi_k, \tag{12}$$

$$L_{nk}^1 = \int_0^\pi d\theta \sin\theta \tau_n d_k x_\theta \xi_n \psi_k, \tag{13}$$

$$L_{nk}^2 = \int_0^\pi d\theta \sin\theta d_n \tau_k x_\theta \xi_n \psi_k, \tag{14}$$

where $x = k_1 r(\theta)$ is the size parameter, $r(\theta)$ describes the generatrix in polar coordinates, $x_\theta = dx/d\theta$, $d_n \equiv d_{0m}^n(\theta)$ is the Wigner d-function (proportional to the associated Legendre functions), $\tau_n \equiv dd_n/d\theta$, $\pi_n = md_n/\sin\theta$ (note that $\pi_n d_k = \pi_k d_n$), $\psi_k \equiv \psi_k(sx), \xi_n \equiv \xi_n(x)$, where ψ and ξ are the regular Riccati–Bessel function, and the Riccati–Hankel function of the first kind, respectively (see Ref. [16] for further details), and $s = k_2/k_1$ is the relative refractive index. The integrals K_{nk}^1 and K_{nk}^2 are used to calculate $\mathbf{Q}^{12}$ and $\mathbf{Q}^{21}$ as

$$Q_{nk}^{12} = A_n A_k \frac{s^2-1}{s} K_{nk}^1, \tag{15}$$

$$Q_{nk}^{21} = A_n A_k \frac{1-s^2}{s} K_{nk}^2, \tag{16}$$

where $A_n = \sqrt{(2n+1)/(2n(n+1))}$. Off-diagonal $(n \neq k)$ terms of $\mathbf{Q}^{11}$ will be calculated from

$$Q_{nk}^{11} = \frac{iA_n A_k (s^2-1)/s}{n(n+1)-k(k+1)} L_{nk}^5, \tag{17}$$

where

$$L_{nk}^5 = [n(n+1)L_{nk}^2 - k(k+1)L_{nk}^1]$$
$$= \int_0^\pi d\theta \sin\theta [n(n+1)d_n \tau_k - k(k+1)\tau_n d_k] x_\theta \xi_n \psi_k. \tag{18}$$

For the other cases, we will use alternative expressions (to those of [29]) that are better suited to our problem here. Explicitly, for the off-diagonal terms of Q_{nk}^{22}, we use

$$Q_{nk}^{22} = \frac{iA_n A_k (s^2-1)/s}{n(n+1)-k(k+1)} L_{nk}^6, \tag{19}$$

where

$$L_{nk}^6 = [n(n+1)L_{nk}^8 - k(k+1)L_{nk}^7] \tag{20}$$

with

$$L_{nk}^7 = \int_0^\pi d\theta \sin\theta \tau_n d_k x_\theta \left(\xi_n' \psi_k' + n(n+1)\frac{\xi_n \psi_k}{sx^2}\right), \tag{21}$$

$$L_{nk}^8 = \int_0^\pi d\theta \sin\theta d_n \tau_k x_\theta \left(\xi_n' \psi_k' + k(k+1)\frac{\xi_n \psi_k}{sx^2}\right). \tag{22}$$

For the diagonal terms, we use the original expressions given in [16, Section 5.8.3] to obtain

$$Q_{nn}^{11} = -\frac{i}{s}(A_n)^2 \tilde{L}_n^1, \tag{23}$$

$$Q_{nn}^{22} = -\frac{i}{s}(A_n)^2 [\tilde{L}_n^2 + (s^2-1)n(n+1)\tilde{L}_n^3], \tag{24}$$

with

$$\tilde{L}_n^1 = \int_0^\pi d\theta \sin\theta (\pi_n \pi_n + \tau_n \tau_n)(\xi_n' \psi_n - s\xi_n \psi_n'), \tag{25}$$

$$\tilde{L}_n^2 = \int_0^\pi d\theta \sin\theta (\pi_n \pi_n + \tau_n \tau_n)(s\xi_n' \psi_n - \xi_n \psi_n'), \tag{26}$$

$$\tilde{L}_n^3 = \int_0^\pi d\theta \sin\theta \tau_n d_n x_\theta \frac{\xi_n \psi_n}{sx^2}. \tag{27}$$

To calculate **P** we use the same integrals but replace the Riccati–Hankel functions $\xi_n(x)$ with the regular Riccati-Bessel functions $\psi_n(x)$. For the purpose of this work, it will also be beneficial to calculate **Q** from:

$$\mathbf{Q} = \mathbf{P} + i\mathbf{U}, \tag{28}$$

where, since $\xi_n(x) = \psi_n(x) + i\chi_n(x)$, **U** is calculated as **Q** by replacing $\xi_n(x)$ with the irregular Riccati-Bessel functions $\chi_n(x)$. In Ref. [20], we have shown that there is no loss of precision associated with computing **P**. By using $\mathbf{Q} = \mathbf{P} + i\mathbf{U}$, we therefore isolate the problems in the calculation of **U** only. We note in passing that we recently suggested [6] that **U** itself may be more physically meaningful than **Q** in some instances.

Finally, in this paper, we will only consider spheroidal particles for which

$$x(\theta) = \frac{k_1 ac}{\sqrt{a^2 \cos^2\theta + c^2 \sin^2\theta}}, \tag{29}$$

with semi-axes a and c (axis of symmetry).

2.3. Particles with mirror symmetry

For spheroids, the reflection symmetry with respect to the equatorial plane results in a number of further simplifications (see Ref. [16, Section 5.2.2]). Half of the matrix entries are zero because of the symmetry in changing $\theta \to \pi - \theta$ and the other integrals are simply twice the integrals evaluated over the half-range 0 to $\pi/2$. Explicitly, we have

$$P_{nk}^{11} = P_{nk}^{22} = Q_{nk}^{11} = Q_{nk}^{22} = 0 \quad \text{if } n+k \text{ odd},$$
$$P_{nk}^{12} = P_{nk}^{21} = Q_{nk}^{12} = Q_{nk}^{21} = 0 \quad \text{if } n+k \text{ even}. \tag{30}$$

We then have a complete decoupling between the even magnetic and odd electric multipoles on the one hand and the odd magnetic and even electric multipoles on the other hand (i.e. the matrices are block-diagonal upon rearranging the multipoles in these two groups). We can therefore rewrite Eq. (6) as two independent sets of equations, one for each. For illustration, we do it explicitly in the following for even magnetic and odd electric multipoles. The corresponding vectors of

coefficients are denoted as

$$\mathbf{a}_e = \begin{pmatrix} a_2 \\ a_4 \\ \vdots \end{pmatrix}, \quad \mathbf{b}_o = \begin{pmatrix} b_1 \\ b_3 \\ \vdots \end{pmatrix}, \tag{31}$$

and similarly for $\mathbf{c}$, $\mathbf{d}$, $\mathbf{p}$, $\mathbf{q}$.

We then have

$$\begin{pmatrix} \mathbf{p}_e \\ \mathbf{q}_o \end{pmatrix} = -\mathbf{P}^{eo} \begin{pmatrix} \mathbf{c}_e \\ \mathbf{d}_o \end{pmatrix}, \tag{32}$$

$$\begin{pmatrix} \mathbf{a}_e \\ \mathbf{b}_o \end{pmatrix} = \mathbf{Q}^{eo} \begin{pmatrix} \mathbf{c}_e \\ \mathbf{d}_o \end{pmatrix}, \tag{33}$$

where $\mathbf{Q}^{eo}$ (similarly $\mathbf{P}^{eo}$) is defined as

$$\mathbf{Q}^{eo} = \begin{pmatrix} \mathbf{Q}^{11}_{ee} & \mathbf{Q}^{12}_{eo} \\ \mathbf{Q}^{21}_{oe} & \mathbf{Q}^{22}_{oo} \end{pmatrix}, \tag{34}$$

where $\mathbf{Q}^{12}_{eo}$ denotes the sub-matrix of $\mathbf{Q}^{12}$ with even row indices and odd column indices, and similarly for the others. For the T-matrix, we therefore have

$$\begin{pmatrix} \mathbf{p}_e \\ \mathbf{q}_o \end{pmatrix} = \mathbf{T}^{eo} \begin{pmatrix} \mathbf{a}_e \\ \mathbf{b}_o \end{pmatrix}, \tag{35}$$

where

$$\mathbf{T}^{eo} = -\mathbf{P}^{eo}(\mathbf{Q}^{eo})^{-1}. \tag{36}$$

A similar expression is obtained for the second group of multipoles to define $\mathbf{T}^{oe}$. The problem of finding the $2N \times 2N$ T-matrix up to multipole order N therefore reduces to finding the two decoupled T-matrices $\mathbf{T}^{eo}$ and $\mathbf{T}^{oe}$, each of size $N \times N$. This approach is particularly useful to reduce computation times for the inversion step.

3. A problem with spheroids

3.1. Loss of precision in numerical integration

In the EBCM we have a method that is particularly appealing in the case of axisymmetric particles. However, in some situations such as metallic particles with a high aspect ratio, numerical problems limit the accuracy of the method. Often it is assumed that this is due to the matrix inversion step (Eq. (7)) [19,30,31], yet it was shown recently that in the case of spheroids, the problem lies primarily in the calculation of the integrals prior to the inversion [20]. Notably, leading terms in the Laurent series[1] expansion of the integrands can dominate by many orders of magnitude, yet integrate identically to zero. When numerically evaluating these integrals, the limited precision available means that these terms dominate the integral through a lack of proper cancellation, even though they should make no contribution.

There are two main ways to formulate a solution to the cancellation problem. The conceptually simpler one is to implement the calculations using code with arbitrary

---

[1] We refer to these expansions as Laurent series, since they are power series with negative powers, but note that the expansions only have a finite number of negative terms.

precision (AP), so the precision of the results can be maintained. The drawbacks of this approach are that arbitrary precision calculations are much slower than double precision (DP) calculations, and they often require rewriting the computer code used to carry out the calculations. They are useful though in being able to calculate the correct results, for checking the results of other code. The other option to remove the cancellations is to use different integrands which do not have cancellations while integrating, but give the same integral. In this case, where the cancellations are due to terms in the series expansion that integrate identically to zero, this amounts to rewriting the integrands so that the problematic terms are removed. This is the approach that was suggested by Waterman [21] in the case of elliptic cylinders, and it is also the approach taken in this work.

3.2. Example of integral cancellations

The problematic term in the integrals is the product of the radial functions, that is the Riccati–Bessel functions ξ_n, ψ_k, and their derivatives ξ'_n, ψ'_k. We here provide a rough outline of the approach we used for the product $\xi_n \psi_k$ (as in the integral L^1_{nk}, L^2_{nk}, or L^5_{nk} needed for Q^{11}_{nk}), and the other products behave similarly. As mentioned earlier, we start by noting that $\xi_n = \psi_n + i\chi_n$, and that $\mathbf{P}$ has no cancellations, so the product $\psi_n \psi_k$ has no cancellations. Hence we focus on the product $\chi_n \psi_k$ in the calculation of $\mathbf{U}$, which is the part that exhibits cancellations.

In this section we provide an outline of the method, with further details provided in the Appendix. We use a method based on the Laurent-series expansions of the functions to calculate their product. The series expansions are fully described in Appendix A, and are

$$\chi_n(x) = \frac{1}{x^n} \sum_{i=0}^{\infty} \tilde{a}_i x^{2i}, \tag{37}$$

$$\psi_k(sx) = x^{k+1} \sum_{i=0}^{\infty} \tilde{b}_i x^{2i} s^{2i+k+1}, \tag{38}$$

for some coefficients $\tilde{a}_i$, $\tilde{b}_i$ whose value is not important in this section. The product of these can be written as

$$\chi_n(x)\psi_k(sx) = \sum_{q=0}^{\infty} \tilde{c}_q x^{2q+k-n+1}, \tag{39}$$

where

$$\tilde{c}_q = \sum_{i=0}^{q} \tilde{a}_i \tilde{b}_{q-i} s^{2q-2i+k+1}. \tag{40}$$

If we focus on a specific example, such as $n = 5$, $k = 1$, then the integral L^1_{51} takes the form

$$L^1_{51} = \int_0^\pi d\theta \sin\theta \tau_5 d_1 x_\theta \left(\frac{\tilde{c}_0}{x^3} + \frac{\tilde{c}_1}{x} + \tilde{c}_2 x + \tilde{c}_3 x^3 + \cdots \right). \tag{41}$$

For particles with a size parameter less than 1, the first few terms strongly dominate the integrand. However, as demonstrated numerically and analytically in Ref. [20], the x^{-3} term actually exactly integrates to zero in the special case of spheroids.

As a result, the integral is often several orders of magnitude smaller than the integrand, which causes serious loss of precision issues when evaluating it numerically. As $(n-k)$ increases, the problem becomes particularly acute and even affects larger values of x, since an increasing number of dominant terms in the series integrate exactly to zero [20].

4. A new implementation

4.1. Series-based implementation

The solution to this problem is *a priori* simple. Rather than computing the integrand, one should compute only the part of it that does not integrate to zero, i.e. exclude the leading terms that have been shown to integrate exactly to zero. After integration, this should give the same value, but not suffer from loss of precision issues.

More formally, we define the operator $\mathcal{P}^+$ such that it returns the part of a series expansion with non-negative powers of x. That is, for example:

$$\mathcal{P}^+\left(\frac{1}{x^2}+\frac{1}{x}+1+x+x^2+\cdots\right)$$
$$=1+x+x^2+\cdots. \tag{42}$$

Similarly $\mathcal{P}^-$ returns terms with negative powers of x, and $(\mathcal{P}^+ + \mathcal{P}^-)f = f$. For convenience, we also denote

$$f^+ \equiv \mathcal{P}^+ f, \tag{43}$$

$$f^- \equiv \mathcal{P}^- f. \tag{44}$$

Initially, we shall focus on the product $\chi_n\psi_k$, which appears for the calculation of Q_{nk}^{11} in integral L_{nk}^5 given in Eq. (18); the other cases follow naturally and we will come back to them later. In Ref. [20] we demonstrated that the part of the Laurent series of this product that contributes to the integral is that with powers of x greater or equal to -1. We therefore define

$$F_{nk}(s,x) = x\chi_n(x)\psi_k(sx) \tag{45}$$

and wish to calculate $F_{nk}^+(s,x) = \mathcal{P}^+(F_{nk})$, so that F^+/x will be used in the cancellation-free integrand. Because of the reflection symmetry for spheroids, we only need to compute the L_{nk}^j integrals, and therefore F_{nk}^+, for $(n+k)$ even. Moreover, the dominant term in the series expansion of F_{nk} is of order x^{k+2-n}; for $k+2-n \geq 0$, we then simply have $F_{nk}^+ = F_{nk}$, which can computed directly by calculating the Riccati–Bessel functions. We therefore focus on cases where $n \geq k+4$ (the bottom-left part of the matrix). It is worth noting that the calculation of F^+ should be carried out without ever using any part of F^-, else subtracting those terms (which dominate in magnitude) will reduce the precision of our result. In other words, calculating $F^+ = F - F^-$ would result in severe loss of precision as in most cases of interest $F \approx F^-$ and $F^+ \ll F^-, F$. The main difficulty is therefore to devise a method of computing F^+ directly.

The most direct approach, which is the one used by Waterman in Ref. [21], is to write explicitly the series expansion of the product and keep only the part $(^+)$ devoid of cancellations, i.e. from Eq. (39) we deduce

$$F_{nk}^+ = (x\chi_n(x)\psi_k(sx))^+ = \sum_{q=q_{\min}}^{\infty} \tilde{c}_q x^{2q+k-n+2}, \tag{46}$$

where $q_{\min} = (n-k)/2 - 1$ (note that $q_{\min} \geq 1$ in cases of interest where cancellations are present, i.e. when $n \geq k+4$). Full expressions for this series are given and discussed in Appendix B. The above series expansion for F_{nk}^+ can be computed numerically by truncating it when additional terms are smaller than the required accuracy (machine epsilon for double precision in our implementation). To check the numerical stability of this approach, we have compared in Fig. 1 the double precision results for $F_{nk}^+(s,x)$ with a similar computation in arbitrary precision arithmetic, where the exact accuracy to double precision (around 16 digits) can be ensured. From Fig. 1, it is clear that the computations suffer from loss of precision for larger x (typically $x \gtrsim 10$), which is not unexpected for a Taylor expansion. However, even for small x where such series expansions should converge rapidly, there is also a clear numerical problem when $|s|$ approaches 1. In fact, if high precision is required, this problem even affects values of $s \approx 1.5$, which is clearly undesirable. Further investigations revealed that this problem is related to additional analytic cancellations occurring in the computation of the coefficients $\tilde{c}_q$ when $s=1$, namely $\tilde{c}_q = 0$ for $q \leq n-k-2$ if $s=1$ (see Appendix C for further discussions of these).

In order to circumvent this problem, we developed a method of computing this series, which isolates the problems associated with $s=1$ and avoids the associated numerical loss of precision. The method is rather technical and it is described in full in Appendix B. As shown in Fig. 2, the problem around $|s| \approx 1$ can be fully overcome. With a relative accuracy as high as 10^{-13}, this approach then appears to be valid for all s, for small x and up to $x \approx 10$. We have thus obtained a method of computing F_{nk}^+ to high accuracy. However, two issues remain to be addressed. First, for larger size parameter x, the method still suffers from severe loss of precision in part of the matrix (small n). Second, the computation of the series for $\mathbf{F}^+$ can be relatively time-consuming and will scale quadratically as N^2, where N is the maximum multipole order for the T-matrix computation. In the following section we present a complementary method of calculation for the matrix elements based on recursion, and show that these two methods can work together to alleviate these limitations.

### 4.2. Stable recursion relation between matrix elements

In order to reduce the computation demands to a linear scaling with N, we have also developed additional schemes based on recursion relations on F_{nk}^+. As we shall see, this approach will also circumvent the problem of loss of precision for large x/small n.

One possible approach is to use recursion relations of the Bessel functions, such as

$$\chi_{n-1}(x)+\chi_{n+1}(x) = \frac{2n+1}{x}\chi_n(x), \tag{47}$$
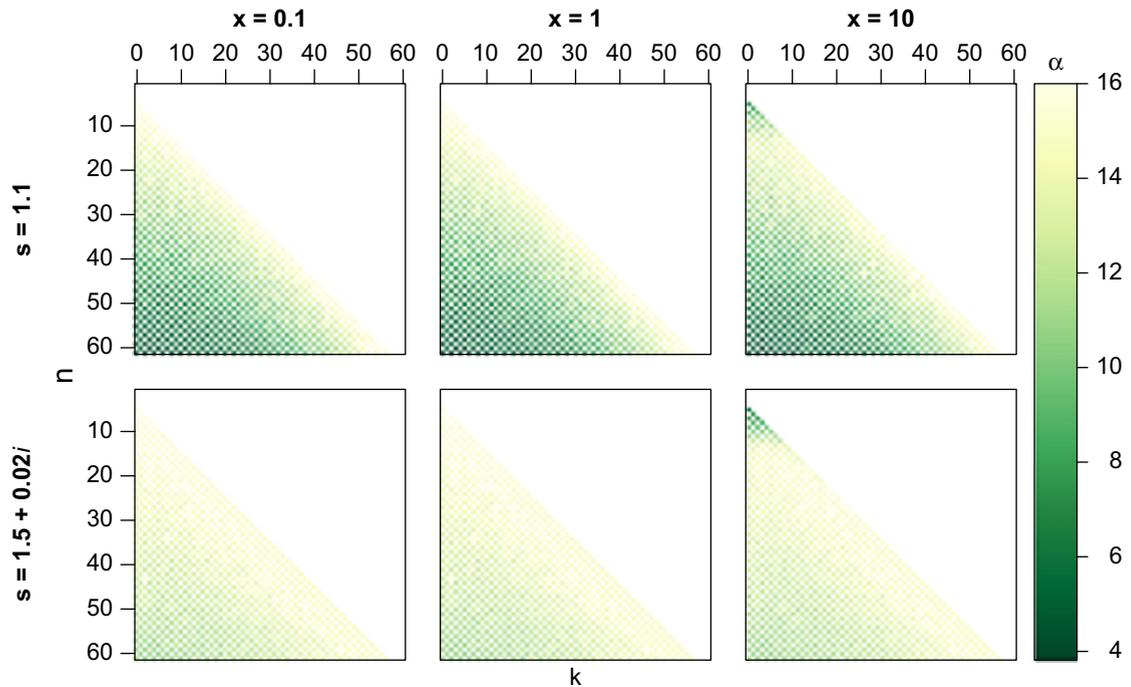
Fig. 1. Relative error in computing $F_{nk}^+(s,x)$ by a straight calculation of the series expansion (Eq. (46)) in double precision (DP), compared to the arbitrary precision results (AP). The relative error is defined as $\alpha = -\log_{10}|(DP-AP)/AP|$, which represents the number of digits of agreement with respect to the arbitrary precision result. Only relevant cases are shown: $n+k$ even and $n \geq k+4$, up to relatively large n (max of $N+1 = 61$). Three values of x (size parameter) are tested here $x=0.1$ (left), $x=1$ (middle), and $x=10$ (right); along with two values of the relative refractive index $s = 1.5 + 0.02i$ (bottom) and $s = 1.1$ (top). The latter highlights the problem of loss of precision for low k, high n, which can be traced back to specific cancellations when $s=1$.
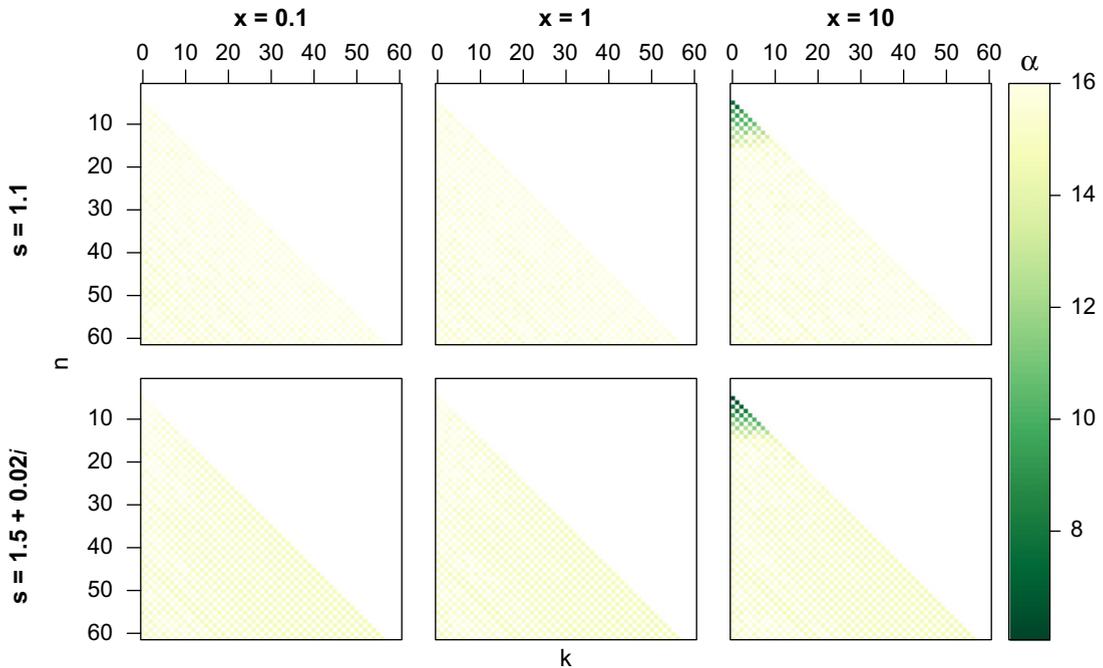
Fig. 2. Same as Fig. 1, but with the optimised implementation described in Appendix B designed to analytically compute the problematic terms associated with $s=1$. Extremely high accuracy (13 digits) is now achieved across the whole matrix, except for large x (typically ≥ 10), where the low n part of the matrix still suffers from serious loss of precision. These can be circumvented by using the recursion-based scheme presented in Section 4.2.

which implies

$$F_{n+1,k} = \frac{2n+1}{x} F_{n,k} - F_{n-1,k}, \qquad (48)$$

from which we deduce by considering Taylor expansions that

$$F_{n+1,k}^{+}(x) = \frac{2n+1}{x}(F_{n,k}^{+}(x) - F_{n,k}^{+}(0)) - F_{n-1,k}^{+}(x). \qquad (49)$$

However, for small x this approach fails, as computing the difference $F_{n,k}^{+}(x) - F_{n,k}^{+}(0)$ leads to a loss of precision, which accumulates during the recursion. To avoid this problem, it is necessary to obtain a recursion relation on F_{nk} that does not involve any multiplication or division by x. This is achieved by combining Eq. (47) with the equivalent one for ψ, namely

$$\psi_{k+1}(sx) + \psi_{k-1}(sx) = \frac{2k+1}{sx} \psi_k(sx), \qquad (50)$$

which leads to

$$F_{n+1,k}^{+} + F_{n-1,k}^{+} = s\frac{2n+1}{2k+1}(F_{n,k+1}^{+} + F_{n,k-1}^{+}). \qquad (51)$$

Note that the above recursion is also valid for F_{nk} itself.

In order to implement this recursion, we still need initial conditions, which require evaluation of some of the F_{nk}^{+} using the series-based method discussed earlier. For example, if we solve for $F_{n+1,k}^{+}$, the matrix is filled from top (small n,k) to bottom (large n) and only requires to compute as initial conditions the first column ($k=1$) and the sub-diagonal ($n=k+2$, which is trivial since $F_{nk}^{+} = F_{nk}$). However, this approach, arguably the most natural, results in a significant accumulation of loss of precision during the recursion as shown in Fig. 3(d). There are at least three alternative approaches to fill the matrix recursively, and their respective accuracies are compared in Fig. 3. From this, we have therefore chosen to solve for $F_{n,k-1}^{+}$ (Fig. 3(c)), which appears to be numerically stable and requires the fewest initial entries (the last row, $n=N+1$, and the sub-diagonal). Moreover, it provides a method of calculating the entire F^{+} matrix for large x, where the series-based calculation failed for small n. One simply needs to choose N large enough so that the last row $n=N+1$ can be computed accurately via series expansions. The recursion scheme is then used to fill the rest of the matrix with no loss of precision. This is further illustrated in Fig. 4 in the case where $x=35$. With this method, F_{nk}^{+} can therefore be computed rapidly, in double precision arithmetic, with almost no loss of precision for all the n,k values where cancellations occur.

4.3. Other products and integrals

The other blocks of the matrix **Q** (or **U**) can be computed in a similar way. For L_{nk}^{5}, we wrote the integral with the cancellation-free integrand for **U** as

$$L_{nk}^{5} = \int_0^{\pi} d\theta \sin\theta [n(n+1)d_n\tau_k - k(k+1)\tau_n d_k] \frac{x_\theta}{x} F_{nk}^{+}(s,x). \qquad (52)$$

Following the findings of Ref. [20], we may rewrite the other integrals that we need with cancellation-free
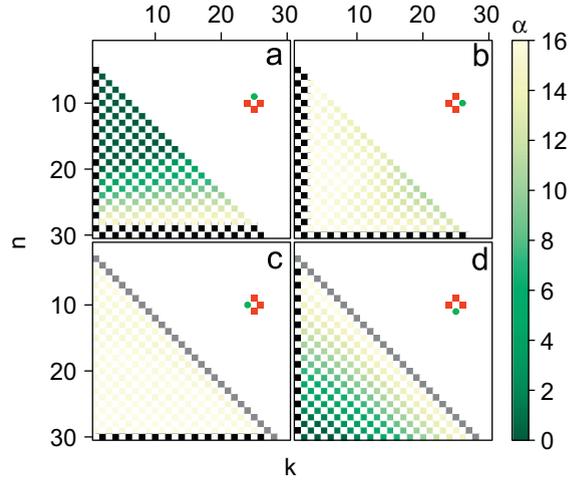
Fig. 3. A comparison of the accumulated error (represented as α, the number of digits of agreement as with Fig. 1) from different recursion directions, compared to values calculated by the series in AP, for the elements of F^{+} for $x=0.1$, $s=1.5+0.02i$. The different directions all have different initial values that must be populated using the series (Section 4.1), which are marked in black. Entries which do not exhibit cancellations (or are not calculated as $n+k$ is odd) are shown in white, and those which do not have cancellations, but which are needed for the recursion, are in grey. For each direction, there is a schematic, showing how to calculate an entry (green circles) from three known values (red squares) using the recursion relation of Eq. (51). These are the relations used to populate the matrix. For the (a) direction, we need the last two rows, and the first column. For the (b) direction we require the first two columns, and the last row. The (c) direction needs the last row and the subdiagonal, while the (d) direction needs the first column and the subdiagonal. Values of α less than zero (for which the order of magnitude is incorrect) are shown as zero. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

integrands for **U** as follows:

$$K_{nk}^{1} = \int_0^{\pi} d\theta \sin\theta \pi_n d_k \frac{x_\theta}{x}[x\chi_n\psi_k']^{+}, \qquad (53)$$

$$K_{nk}^{2} = \int_0^{\pi} d\theta \sin\theta \pi_n d_k \frac{x_\theta}{x}[x\chi_n'\psi_k]^{+}, \qquad (54)$$

$$L_{nk}^{7} = \int_0^{\pi} d\theta \sin\theta \tau_n d_k \frac{x_\theta}{x}\left[x\left(\chi_n'\psi_k' + n(n+1)\frac{\chi_n\psi_k}{sx^2}\right)\right]^{+}, \qquad (55)$$
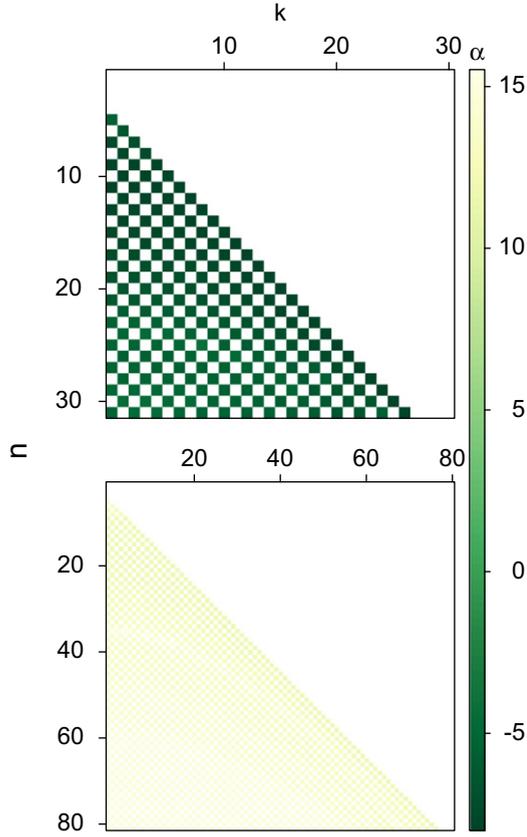
$$L_{nk}^{8} = \int_0^{\pi} d\theta \sin\theta d_n\tau_k \frac{x_\theta}{x}\left[x\left(\chi_n'\psi_k' + k(k+1)\frac{\chi_n\psi_k}{sx^2}\right)\right]^{+}. \qquad (56)$$

To obtain these, all the cancellation-free integrands can be deduced from $F_{nk}^{+}(x) = [x\chi_n(x)\psi_k(sx)]^{+}$ computed for $n+k$ even as explained before.

For K_{nk}^{1} and K_{nk}^{2}, which are non-zero for $n+k$ odd, we may use the relations for Riccati–Bessel functions:

$$(2k+1)\psi_k'(sx) = (k+1)\psi_{k-1}(sx) - k\psi_{k+1}(sx), \qquad (57)$$

$$(2n+1)\chi_n'(x) = (n+1)\chi_{n-1}(x) - n\chi_{n+1}(x). \qquad (58)$$

Fig. 4. Error in F_{nk}^+ for $s=1.5+0.02i$ and a relatively large $x=35$, computed using series expansions for the last row ($n=N+1$) and the recursion scheme (c) of Fig. 3. We here compare two cases: $N=30$ (top) and $N=80$ (bottom). Because of the loss of precision in the series expansion at low N for large x, one must ensure that N is large enough. In this case, $N=80$ results in an accuracy of better than 11.5 digits over the entire matrix, while for $N=30$ the result is never correct to within 4 orders of magnitude.

from which we deduce

$$[x\chi_n\psi'_k]^+ = \frac{(k+1)F_{n,k-1}^+ - kF_{n,k+1}^+}{2k+1}, \tag{59}$$

$$[x\chi'_n\psi_k]^+ = \frac{(n+1)F_{n-1,k}^+ - nF_{n+1,k}^+}{2n+1}. \tag{60}$$

These expressions only need to be used when cancellations are present, i.e. for $n \geq k+3$ [20].

For, L_{nk}^7 and L_{nk}^8, which are non-zero for $(n+k)$ even, we in addition use Eq. (47) for $\chi_n(x)$ and Eq. (50) to deduce

$$\left[x\left(\chi'_n\psi'_k + k(k+1)\frac{\chi_n\psi_k}{sx^2}\right)\right]^+$$
$$= \frac{1}{(2n+1)(2k+1)}$$
$$\times \{(n+k+1)[(k+1)F_{n-1,k-1}^+ + kF_{n+1,k+1}^+]$$
$$+ (k-n)[(k+1)F_{n+1,k-1}^+ + kF_{n-1,k+1}^+]\}, \tag{61}$$

$$\left[x\left(\chi'_n\psi'_k + n(n+1)\frac{\chi_n\psi_k}{sx^2}\right)\right]^+$$

$$= \frac{1}{(2n+1)(2k+1)}$$
$$\times \{(n+k+1)[(n+1)F_{n-1,k-1}^+ + nF_{n+1,k+1}^+]$$
$$+ (n-k)[(n+1)F_{n-1,k+1}^+ + nF_{n+1,k-1}^+]\}. \tag{62}$$

Again, these expressions only need to be used when cancellations are present, i.e. for $n \geq k+2$ [20] (note that this range is larger than for L^5).

All the cancellation-free integrands can therefore be computed from the knowledge of $F_{nk}^+(x)$ for $0 \leq n$, $k \leq N+1$, the computation of which was described in Section 4.2.

When implementing this approach, we noted for small size parameters a minor loss of precision (compared to those normally seen in the lower part of **Q**) on the diagonal of $\mathbf{P}^{11}$, i.e. in integral $\tilde{L}_n^1$ (Eq. (25)). This can be traced to the cancellation of the leading order term ($\propto x^{2n+1}$) in the radial part of the integrand: $\psi'_n(x)\psi_n(sx) - s\psi_n(x)\psi'_n(sx)$ (note that it cancels out when computing the integrand itself, not upon integration as before). To avoid this issue, we simply rewrite this integrand in a form where the leading order term does not appear. For this, we use the relations:

$$\psi'_n(x) = -\psi_{n+1}(x) + \frac{n+1}{x}\psi_n(x), \tag{63}$$

$$\psi'_n(sx) = -\psi_{n+1}(sx) + \frac{n+1}{sx}\psi_n(sx), \tag{64}$$

to arrive at

$$\psi'_n(x)\psi_n(sx) - s\psi_n(x)\psi'_n(sx)$$
$$= s\psi_n(s)\psi_{n+1}(sx) - \psi_{n+1}(x)\psi_n(sx). \tag{65}$$

Note that both quantities on the right-hand-side are now of order x^{2n+3}, and this solves the problem.

4.4. Summary of the procedure

For clarity, we here summarise the entire procedure for calculating the matrices **P** and **Q** with minimal loss of precision, up to multipole order N. The integrals are computed as in most T-matrix implementations using a Gaussian quadrature with T points denoted θ_i and associated weights w_i ($i=1..T$). Thanks to the reflection symmetry, half of the entries are zero, and for the others the integration range can be reduced to $0 \leq \theta \leq \pi/2$ and we choose T even, such that

$$P_{nk}^{ij} = 2\sum_{i=1}^{T/2} f(\theta_i)w_i, \tag{66}$$

where the integrand for P_{nk}^{11} for example is

$$f(\theta_i) = C_{nk}x_\theta(\theta_i)A_{nk}(\theta_i)[\psi_n(x_i)\psi_k(sx_i)]. \tag{67}$$

Note that the $\sin\theta$ factor is contained in the weights, that C_{nk} is a constant and that $A_{nk}(\theta_i)$ represents the angular function dependence, which in the case of P_{nk}^{11} is

$$A_{nk}(\theta_i) = n(n+1)\tau_k(\theta_i)d_n(\theta_i) - k(k+1)\tau_n(\theta_i)d_k(\theta_i). \tag{68}$$

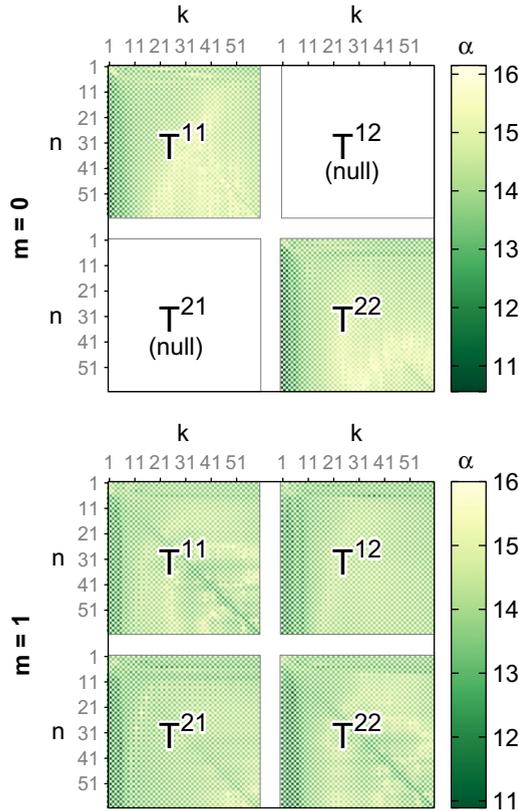For **P**, there is no loss of precision, and the entire matrix can be computed with these standard methods.

Q_{nk}^{ij} is then obtained from $Q_{nk}^{ij} = P_{nk}^{ij} + iU_{nk}^{ij}$ where for U_{nk}^{ij}, we now use the cancellation-free integrands described in
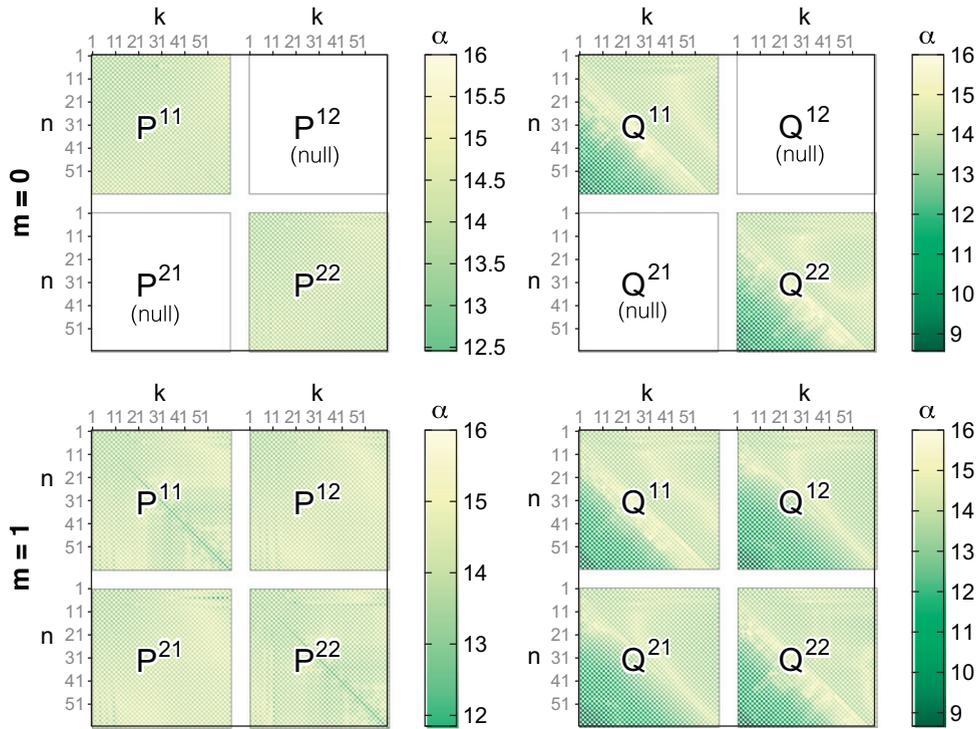
Fig. 5. The relative error (characterised by α) for **P** (left) and **Q** (right) when computed in double precision with our proposed method, compared to arbitrary precision, for $m=0$ (top) and $m=1$ (bottom). This is for a prolate spheroid of aspect ratio $h=4$, with a large size parameter $x_{\max}=10$, and $s=1.5+0.02i$.
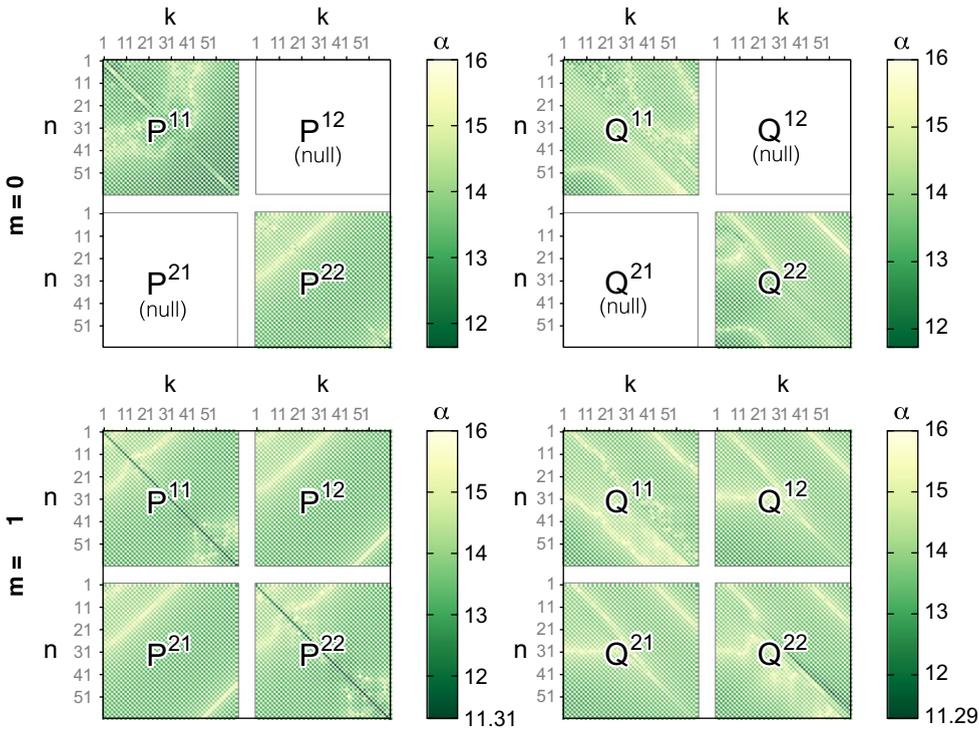
Fig. 6. The relative error (characterised by α) for **P** (left) and **Q** (right) when computed in double precision with our proposed method, compared to arbitrary precision, for $m=0$ (top) and $m=1$ (bottom). This is for a silver nanorod in a solvent of refractive index $n=1.33$ at its dipolar localised plasmon resonance ($\lambda=2560$ nm), modelled as a prolate spheroid with $a=10$ nm, $c=200$ nm (i.e. aspect ratio $h=20$ and size parameter $x_{\max}=0.65$), and $s=1.02+13.5i$.

Fig. 7. The error in T computed fully in double precision compared to arbitrary precision, for $m=0$ (top) and $m=1$ (bottom). This is for a prolate spheroid of aspect ratio $h=4$, with a large size parameter $x_{max}=10$, and $s=1.5+0.02i$.
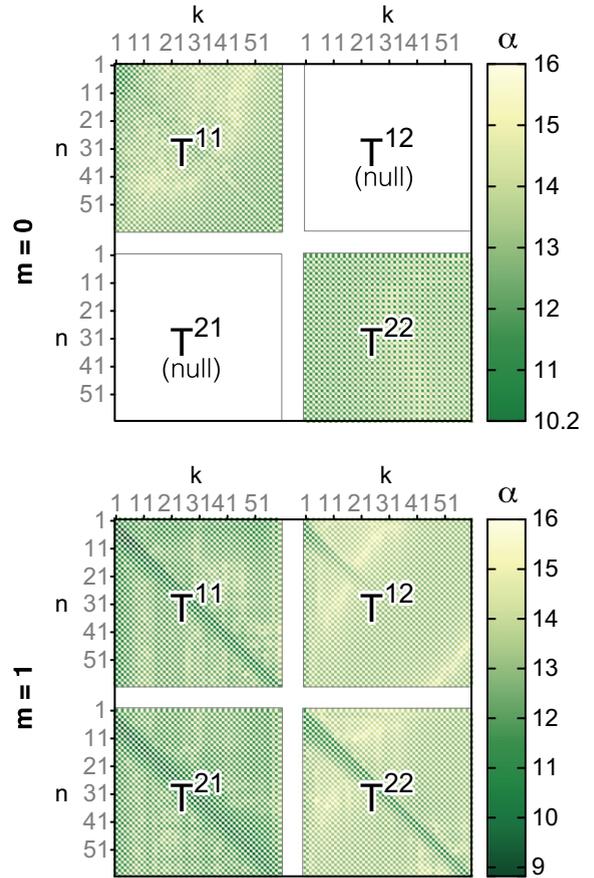

Fig. 8. The error in T computed fully in double precision compared to arbitrary precision, for $m=0$ (top) and $m=1$ (bottom). This is for a silver nanorod in a solvent of refractive index $n=1.33$ at its dipolar localised plasmon resonance ($\lambda=2560$ nm), modelled as a prolate spheroid with $a=10$ nm, $c=200$ nm (i.e. aspect ratio $h=20$ and size parameter $x_{max}=0.65$), and $s=1.02+13.5i$.

Section 4.3, for example for U_{nk}^{11}:

$$f(\theta_i) = C_{nk}x_\theta(\theta_i)A_{nk}(\theta_i)F_{nk}^+(x_i)/x_i. \qquad (69)$$

In Eq. (69), we have replaced $\chi_n(x_i)\psi_k(sx_i)$ with the cancellation-free integrand $F_{nk}^+(x_i)/x_i$; this will not affect the integrals as the part that has been removed integrates identically to zero as shown in Ref. [20].

For each x_i, we calculate the matrix $F_{nk}^+(x_i)$ as follows (note that $n+k$ is even, and we will compute it for $0 \le n$, $k \le N+1$, which include extra terms that are needed for the other integrals):

- For $n \le k+2$ (upper-right of the matrix), we simply calculate $F_{nk}(x_i)=x_i\chi_n(x_i)\psi_k(sx_i)$ and use $F_{nk}^+(x_i)=F_{nk}(x_i)$.
- For the last row, i.e. $n=N+1$ and $0 \le k \le N-4$, we then calculate F_{nk}^+ using the series expansion as described in Appendix B. Note that for large size parameter, N must be large enough for the series-based calculation to be accurate. The accuracy of this step is easy to verify during the calculation of the series, by comparing the order of magnitude of the terms in the summation to the final result. A warning can be issued if a loss of precision occurred, and larger N values should be used.

- We then use the recursion relation given in Eq. (51) (solving for $F_{n,k-1}^+$) and illustrated in Fig. 3(c) to populate the rest of the matrix.
- The $F_{nk}^+(x_i)$ matrices can then be used to compute the U_{nk}^{11} for each desired value of the projected angular momentum number m (typically $0 \le m \le N$) using the Gaussian quadrature described above and in Eq. (66).
- The $F_{nk}^+(x_i)$ matrices are also used to compute the cancellation-free integrands for the other blocks of the matrix **U** as described in the previous section.

The entire Q-matrix can therefore be computed using this method to a high accuracy using only double precision arithmetic. The P-matrix can also be computed to a high precision using standard methods, as it does not suffer from loss of precision in the integrals.

As an example, the accuracy of the P- and Q-matrices for $m=0$ and $m=1$ computed with this method, as compared to the arbitrary precision result, is shown in Fig. 5 for a prolate spheroid with an aspect ratio of 4, maximum size parameter $x_{max}=10$, and $s=1.5+0.02i$, up to $N=60$. It is clear that this procedure results in

extremely accurate integrals even with double precision arithmetic. It is very efficient at avoiding the troublesome cancellations discussed earlier. Relative accuracies of 10^{-12} or better are achieved up to $n = 25$ and these remain better than 10^{-8} for the largest n of the order of 60. A similar plot is shown in Fig. 6 for a high aspect ratio metallic nanoparticle (i.e. a nanorod).

4.5. Matrix inversion

A number of approaches have been proposed [30–32] to solve the linear system to obtain the T-matrix (Eq. (7)), and their respective numerical stability may strongly depend on the structure of the Q-matrix, and therefore

on particle shape and refractive index. As pointed out in Ref. [20], the matrix inversion step required to obtain the T-matrix can be carried out with minimal loss of precision in the special case of spheroids, provided that the Q-matrix does not suffer from loss of precision during integral evaluation and that the appropriate inversion method is used. We found that an efficient and numerically stable method of carrying out this inversion is to obtain $\mathbf{T}^{eo}$ and $\mathbf{T}^{oe}$ separately using block matrix inversions [31,33,34] where matrix inversions are carried out through a LU factorisation obtained from Gaussian elimination with partial pivoting. This corresponds to the
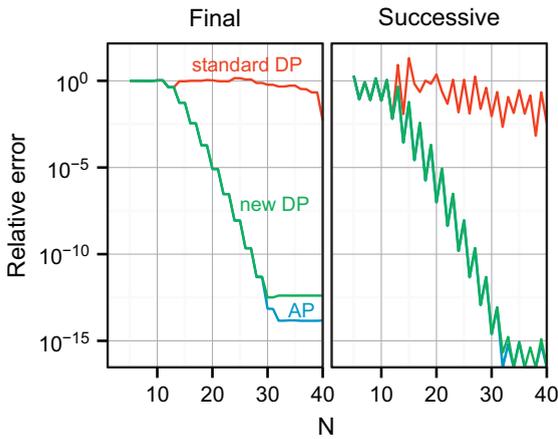
Fig. 9. A demonstration of the convergence behaviour for the problematic case of high aspect ratio metallic nanoparticles (as in Fig. 8). The extinction coefficient (Q_{Ext}, which is the extinction cross-section, normalised by the geometric cross-section πac) at resonance is calculated (only the dominant $m = 0$ case is included) using a standard DP code, our new DP code, and AP codes (with inversion in double precision). The left panel shows the relative error compared to an AP value including inversion in AP (which is assumed to be the correct result), as the multipole order N is increased from 4 to 40 in steps of 1. The right panel shows the relative successive error (from N to $N+1$), which corresponds to a typical convergence test in T-matrix computations. The values of Q_{Ext} for $N = 41$ (to full double precision) are 14.0408863566895 (standard DP), 44.7992522913149 (new DP) and 44.7992522913331 (AP, including inversion).
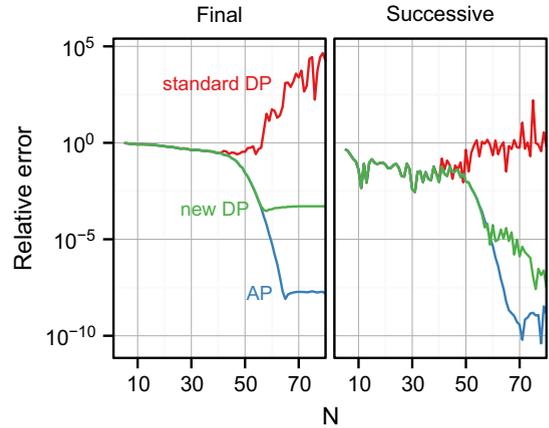
Fig. 10. A demonstration of the convergence behaviour for the problematic case of large size parameter particle. This is for $x_{max} = 35$, $h = 4$, $s = 1.5 + 0.02i$, and focuses on Q_1 (the orientation averaged extinction coefficient, which is the extinction cross-section normalised by the average projected surface area, when considering $m = 0$ only). A similar plot was obtained in Ref. [26] using quadruple-precision codes. The left panel shows the relative error compared to an AP value with inversion in AP (which is assumed to be the correct result), as the multipole order N is increased from 4 to 80 in steps of 1. The right panel shows the relative successive error (from N to $N+1$), which corresponds to a typical convergence test in T-matrix computations. In all cases with the new code, $F_{nk}^{+}(x)$ first were obtained up to $N+1 = 81$, then truncated to the N shown in the figure for integration, inversion and calculation of Q_1. As shown in Fig. 4, for large x the calculation of the Bessel functions requires going to high N, hence starting from $N = 80$ in this case. The values of Q_1 for $N = 81$ are 278729.622910794 (standard DP), 5.64964087428721 (new DP) and 5.64970167842576 (AP, including inversion).

Table 1
Values of orientation-averaged scattering and extinction coefficients, albedo ($\omega = Q_{Sca}/Q_{Ext}$), and value of N used, for (A) silver spheroid of aspect ratio $h = 20$ as in Fig. 9, and (B) for the $x = 35$ spheroid as in Fig. 10, considering all $|m| \leq N$. The value of the dielectric function of the spheroid (A) is $\epsilon_{in} = -318 + 48.5i$ (i.e. silver at 2560 nm). The spheroid (A) is embedded in a solvent with refractive index $n = 1.33$. The apparent low value of N to reach convergence for the AP case is due to the fact the AP inversion was calculated once on a large ((A) – $N = 41$, (B) – $N = 81$) matrix, after which $\mathbf{T}$ was truncated.

Model		Q_{sca}	Q_{ext}	ω	N
A	Standard DP	–	–	–	Not converging
	New DP	2.91644692911172	18.6590264037057	0.156302202805849	31
	AP	2.91644692910476	18.6590264036833	0.156302202805662	7
B	Standard DP	–	–	–	Not converging
	New DP	1.68191804051559	2.34118758579468	0.718403792468721	81
	AP	1.68179883363743	2.3411277033233	0.718371249381958	51

Table 2
Values of scattering and extinction coefficients, albedo ($\omega = Q_{Sca}/Q_{Ext}$), and value of N when reaching convergence or pseudo-convergence (lowest relative difference in coefficients for successive values of N), for the Model 1 of [35]. This is a prolate spheroid with $s = 1.55 + 0.01i$, $x_{max} = 10.079368$, $\lambda = 632.8$ nm and $h = 4$. All the new numbers provided are given to double precision, to allow direct comparison.

Source	Q_{sca}	Q_{ext}	ω	N
Ref. [35]	3.212906	3.367213	0.954174	
Standard DP	3.21290557751034	3.36721182146481	0.954173882685129	21
New DP	3.21290554203156	3.36721292620922	0.954173559094947	31
AP	3.21290554203154	3.36721292620919	0.954173559094946	22

`mrdivide` function in Matlab,[2] also denoted /. Using this notation, **T** is computed from the following steps (similar to those in Ref. [31]):

$$\mathbf{F}_1 = \mathbf{I}/\mathbf{Q}^{11},$$

$$\mathbf{G}_1 = \mathbf{P}^{11}\mathbf{F}_1, \quad \mathbf{G}_3 = \mathbf{P}^{21}\mathbf{F}_1, \quad \mathbf{G}_5 = \mathbf{Q}^{21}\mathbf{F}_1,$$

$$\mathbf{F}_2 = \mathbf{I}/[\mathbf{Q}^{22} - \mathbf{G}_5\mathbf{Q}^{12}],$$

$$\mathbf{G}_2 = \mathbf{P}^{22}\mathbf{F}_2, \quad \mathbf{G}_4 = \mathbf{P}^{12}\mathbf{F}_2, \quad \mathbf{G}_6 = \mathbf{Q}^{12}\mathbf{F}_2,$$

$$\mathbf{T}^{12} = \mathbf{G}_1\mathbf{G}_6 - \mathbf{G}_4,$$

$$\mathbf{T}^{22} = \mathbf{G}_3\mathbf{G}_6 - \mathbf{G}_2,$$

$$\mathbf{T}^{11} = -\mathbf{G}_1 - \mathbf{T}^{12}\mathbf{G}_5,$$

$$\mathbf{T}^{21} = -\mathbf{G}_3 - \mathbf{T}^{22}\mathbf{G}_5. \tag{70}$$

For further efficiency, these steps are performed separately for $\mathbf{T}^{eo}$ and $\mathbf{T}^{oe}$.

It appears that this algorithm does not introduce major loss of precision in all the cases we have investigated even if the matrices to be inverted are badly conditioned (condition numbers of the order of 10^{90} in the example of a high aspect ratio silver spheroid studied here).

This is illustrated in Figs. 7 and 8 where the *T*-matrix computed in double precision using the method presented here and the inversion of Eq. (70) is compared to arbitrary precision result. The *T*-matrix can be obtained with a remarkable relative accuracy of at least 10^{-9} up to a large N in both cases. We note that if direct inversion is used instead of block inversion, the relative error in the obtained *T*-matrix for $m = 1$ can be as bad as 30 orders of magnitude, demonstrating the severe ill-conditioning of the linear system and the efficiency of the proposed inversion method in circumventing the numerical problems.

5. Discussion and conclusion

In conclusion, our proposed new implementation is capable of calculating in double precision reliable scattering information in challenging cases where traditional double precision code fails. One of these difficult cases is that of high aspect ratio metallic nanoparticles in the visible/infrared regime. In such cases, the small size parameter means that the integral cancellations are particularly spectacular even at relatively small n, as shown in Ref. [20]. Traditional codes then typically fail

---

[2] Note that we have noticed that in some implementations of Matlab, `mrdivide(A,B)` seems to result in unwanted loss of precision. When this is the case, it can be replaced by the equivalent expression `transpose(mldivide(transpose(B),transpose(A)))`, which for some reason does not suffer from the same problem.

for N as small as $N \approx 10$. This may be sufficient for low-precision prediction of far-field properties. But if higher precision is required, or if near-field information is desired, as often the case for plasmonics applications, then larger N are necessary. As shown in Fig. 8, our new code can handle such cases up to very large N. This is further illustrated in Fig. 9 where we compare the performance of our new code with earlier codes in terms of the convergence (as a function of N) of the computed extinction coefficient at resonance for a silver nanorod. The standard code starts failing around $N \approx 10-12$ and never approaches the correct result. In contrast, the new code starts converging around $N = 12$ and reaches a remarkable maximum relative accuracy of 10^{-12} for $N = 30$. Converged values of the scattering and extinction coefficients for this nanorod are given in Table 1.

Another type of challenging calculation is that of large size parameter dielectric particles, which normally require quadruple-precision codes. In this case, while the effect of cancellations on a given element of **Q** is not as extreme as for small particles, the number of multipole orders required to calculate the scattering properties of a large particle is larger. As shown in Fig. 4, using our implementation the accuracy of the calculated terms actually *increases* as more terms are calculated, for large x. We can compute the optical properties of a spheroid with $x_{max} = 35$, and see that our new code produces results correct to 10^{-4} (compared to arbitrary precision results). We also provide converged values of the scattering and extinction coefficients for this particle in Table 1.

Finally we have also compared the scattering parameters calculated with our new code to results obtained previously in benchmark calculations [35]. The emphasis for benchmarking is not on challenging cases as we have considered so far, but on obtaining reliable and high-precision results for standard cases. As highlighted in Ref. [35], this can be difficult and time-consuming. In the special case of spheroids, our new implementation provides an easy and fast way of calculating such benchmark results. To illustrate this, we also provide results in Table 2 from standard DP code, and AP code (where the calculation of **T**, including inversion, is performed in AP). It is clear that our new code produces results in agreement with the previous results, as do the other two codes. We note that the new DP code agrees to much higher precision with the AP results than the standard DP code. Because the new DP code is numerically stable, it is also easier to assess the convergence compared to standard DP codes, which may initially converge at intermediate N, but eventually diverge at sufficiently high N.

Hence, this method is well-suited to calculations of the optical properties of spheroids in cases ranging from plasmonic nanoparticles (including high aspect-ratio ones) up to large dielectric particles, using code that operates entirely in double precision. This opens up the possibility of routine calculations of the optical properties of spheroidal particles over a large parameter space with a high accuracy.

Appendix A. Series expansions for Riccati–Bessel functions

The radial functions used in the integrals are the Riccati–Bessel functions. These are

$$\psi_n(x) = xj_n(x), \quad \chi_n(x) = xy_n(x), \tag{A.1}$$

$$\xi_n(x) = xh_n^{(1)}(x) = \psi_n(x) + i\chi_n(x), \tag{A.2}$$

where $j_n(x)$ is the regular spherical Bessel function, $y_n(x)$ is the irregular spherical Bessel function, and $h_n^{(1)}(x)$ is the spherical Hankel function of the first kind. The series expansions for these are [36]

$$\chi_n(x) = \frac{-1}{x^n} \sum_{i=0}^{\infty} \frac{(-\frac{1}{2})^i}{i!} a_{i,n} x^{2i}, \tag{A.3}$$

$$\psi_k(sx) = (sx)^{k+1} \sum_{i=0}^{\infty} \frac{(-\frac{1}{2})^i}{i!} b_{i,k} x^{2i} s^{2i}, \tag{A.4}$$

where

$$
a_{i,n} = (2n-1)!! \prod_{\substack{j=1 \\ \text{odd}}}^{2i-1} \frac{1}{j-2n}
$$
$$
= \begin{cases} (-1)^i (2n-2i-1)!! & \text{for } i \le n, \\ \dfrac{(-1)^n}{(2i-1-2n)!!} & \text{for } i > n, \end{cases} \tag{A.5}
$$

$$
b_{i,k} = \frac{1}{(2k+1)!!} \prod_{\substack{j=1 \\ \text{odd}}}^{2i-1} \frac{1}{2k+j+2}
$$
$$
= \frac{1}{(2k+2i+1)!!}, \tag{A.6}
$$

and $\ell!! = 1 \times 3 \times 5 \times \cdots \times \ell$ is the double factorial function.

Appendix B. Calculating $F_{nk}^+(x)$ using series expansions

We explain here in detail how $F_{nk}^+ = (x\chi_n(x)\psi_k(sx))^+$ can be computed in practice using a series expansion where there are cancellations, i.e. for $n+k$ even and $n \ge k+4$. In our proposed implementation, we will typically compute these series for a given fixed n (corresponding to the last row), for all k such as $n+k$ even and $n \ge k+4$, and for a number of x (for the integration).

The proposed solution is therefore optimised with this in mind.

From the series expansion of the radial functions given in Appendix A, we deduce that

$$F_{nk}^+ = -s^{k+1} \sum_{q=q_{\min}}^{\infty} \gamma_{qnk} \frac{(-1)^q}{2^q q!} x^{2q+k-n+2}, \tag{B.1}$$

where

$$q_{\min} = \frac{n-k}{2} - 1 \tag{B.2}$$

and

$$\gamma_{qnk} = \sum_{i=0}^{q} c_{iqnk}, \tag{B.3}$$

where

$$c_{iqnk} = \binom{q}{i} a_{i,n} b_{q-i,k} s^{2(q-i)}, \tag{B.4}$$

$a_{i,n}$ and $b_{j,k}$ have been defined earlier in Appendix A.

As shown in Fig. 1, some loss of precision occurs when s is close to 1 when using Eq. (B.3). These cancellations can be traced back to the fact that when $i \le q \le n-k-1$, $a_{i,n} b_{q-i,k}$ can be written as $(-1)^i$ times a polynomial of variable i and of degree $n-k-q-1$.

In fact for a general polynomial P of degree less than q, there is a general identity:

$$\sum_{i=0}^{q} (-1)^i \binom{q}{i} P(i) = 0. \tag{B.5}$$

From this we deduce that when $s=1$

$$\gamma_{qnk}(s=1) = 0, \quad \text{if } \frac{n-k}{2} \le q \le n-k-1. \tag{B.6}$$

To avoid these severe cancellations, which in fact remain to some degree when s is not exactly 1, we have developed an alternative method to compute γ_{qnk} when $q \le n-k-1$.

B.1. Method for $q_{\min} \le q \le n-k-1$

To isolate the singular behaviour at $s=1$, we will replace $s^{2(q-i)}$ in Eq. (B.4) by

$$s^{2(q-i)} = (s^2-1+1)^{q-i} = \sum_{j=0}^{q-i} \binom{q-i}{j} (s^2-1)^j \tag{B.7}$$

meaning that

$$\gamma_{qnk} = \sum_{i=0}^{q} \binom{q}{i} a_{i,n} b_{q-i,k} \sum_{j=0}^{q-i} \binom{q-i}{j} (s^2-1)^j.$$

This double-sum may be re-arranged using $\sum_{i=0}^{q} \sum_{j=0}^{q-i} = \sum_{j=0}^{q} \sum_{i=0}^{q-j}$. Substituting the expressions for $a_{i,n}$ (noting that $i \le n$ here) and $b_{q-i,k}$, and combining the binomial coefficients, we arrive at

$$\gamma_{qnk} = \sum_{j=0}^{q} \binom{q}{j} (s^2-1)^j w_{q-j}, \tag{B.8}$$

where

$$w_r = \sum_{i=0}^{r}(-1)^i \binom{r}{i} \frac{(2n-2i-1)!!}{(2k+2q-2i+1)!!}. \tag{B.9}$$

As noted earlier $(2n-2i-1)!!/(2k+2q-2i+1)!!$ is polynomial in i of degree $n-k-q-1$, and therefore $w_r=0$ if $n-k-q-1 < r$, such that the sum only needs to be carried out from $j = \max(0, 2(q-q_{\min})-1)$. To calculate w_r in other cases, we introduce (the n-dependence is implicit)

$$u_{rb} = 2^b \left(\frac{d}{dX}\right)^b \left[X^{n-1/2}\left(1-\frac{1}{X}\right)^r\right]\Bigg|_{X=1}$$

$$= 2^b \left(\frac{d}{dX}\right)^b \left[\sum_{i=0}^{r}(-1)^i\binom{r}{i}X^{n-1/2-i}\right]\Bigg|_{X=1}. \tag{B.10}$$

By carrying successive differentiations, one can notice that for $0 \le b \le n-r-1$:

$$u_{rb} = \sum_{i=0}^{r}(-1)^i\binom{r}{i}\frac{(2n-2i-1)!!}{(2n-2i-2b-1)!!} \tag{B.11}$$

from which we deduce that

$$w_r = u_{r,n-k-q-1} \tag{B.12}$$

and

$$\gamma_{qnk} = \sum_{j=\max(0,2(q-q_{\min})-1)}^{q} \beta_{jq}u_{q-j,n-k-q-1}, \tag{B.13}$$

where

$$\beta_{jq} = \binom{q}{j}(s^2-1)^j. \tag{B.14}$$

In practice, we use this expression for $(n-k)/2-1 \le q \le n-k-1$ and therefore need to calculate u_{rb} for $0 \le b \le (n-k)/2$ and $0 \le r \le b$; and β_{jq} for $0 \le l \le q$.

The latter can easily be computed for a given q through the recursion

$$\beta_{0,q} = 1,$$

$$\frac{\beta_{j,q}}{\beta_{j-1,q}} = \frac{q-j+1}{j}(s^2-1). \tag{B.15}$$

Note that some β_{jq} are reused in the computations at different k, and it is not necessary to compute them multiple times (they do not depend on k).

Using the definition of u_{rb}, one can find the following relations:

$$u_{0,0} = 1, \quad u_{r,0} = 0 \text{ for } r > 0$$

$$u_{r,b+1} = (n-\tfrac{1}{2}-2r)u_{rb}-(n-\tfrac{1}{2}-r)u_{r+1,b}+ru_{r-1,b}, \tag{B.16}$$

from which u_{rb} can be computed by recursion on b. Note that we can easily deduce from this recursion that $u_{rb} = 0$ if $r > b$, as obtained before by different means. Since u_{rb} does not depend on k, for a given n, it can be calculated once for all $0 \le b \le n/2$ and $r \le b$, and then used as necessary for all k.

In summary, all the γ_{qnk} for all $0 \le k \le n-4$ and $(n-k)/2-1 \le q \le n-k-1$ can be computed efficiently and without loss of precision using Eqs. (B.13), (B.15), and (B.16).

B.2. Method for $q \ge n-k$

In the case where $q \ge n-k$, there does not seem to be any serious numerical problems when computing γ_{qnk} using Eq. (B.3). One should nevertheless make sure that the numerical implementation is efficient in time (i.e. minimising operations) and in precision (i.e. for example avoiding to compute factorials of large numbers, which can be problematic in double precision). We briefly present here our chosen implementation. The goal here is to compute γ_{qnk} from Eq. (B.3) for a fixed n, for $0 \le k \le n-4$, and for $q \ge q_{\text{int}} = n-k$.

We therefore need to calculate c_{iqnk} for $0 \le i \le q$. For this, we use (indices n, k are implicit here)

$$\frac{c_{i,q}}{c_{i+1,q}} = \frac{s^2(i+1)(2i+1-2n)}{(q-i)(2k+2q-2i+1)}. \tag{B.17}$$

Assuming we have the initial values $c_{q,q,n,k}$, we can then calculate all c_{iqnk} for $0 \le i \le q$ by downward recursion on i.

To obtain $c_{q,q,n,k}$, we use an upward recursion on q:

$$\frac{c_{q+1,q+1}}{c_{q,q}} = \frac{1}{2q+1-2n}. \tag{B.18}$$

For this we only need the initial value of $c_{q,q}$ for $q = q_{\text{int}} = n-k$ (which is even), which is simply

$$c_{n-k,n-k,n,k} = a_{n-k,n}b_{0,k} = \frac{1}{2k+1}. \tag{B.19}$$

B.3. Computation of the series

Using the results of the previous sections, we can compute γ_{qnk} for all cases of interest. To calculate the sum in Eq. (B.1), we write it as

$$F_{nk}^{+}(x) = -s^{k+1}\sum_{q=q_{\min}}^{\infty}\gamma_{qnk}\alpha_{qnk}(x), \tag{B.20}$$

where

$$\alpha_{qnk}(x) = \frac{(-1)^q}{2^q q!}x^{2q+k-n+2}. \tag{B.21}$$

For efficiency, $\alpha_{qnk}(x)$ is computed as

$$\frac{\alpha_{q,n,k}(x)}{\alpha_{q-1,n,k}} = \frac{-x^2}{2q} \tag{B.22}$$

and the initial conditions $\overline{\alpha}_{nk} = \alpha_{q_{\min},n,k}(x)$ are obtained by recursion on k:

$$\overline{\alpha}_{n,k=n-2}(x) = 1, \tag{B.23}$$

$$\frac{\overline{\alpha}_{n,k}(x)}{\overline{\alpha}_{n,k+2}(x)} = \frac{-1}{n-k-2}. \tag{B.24}$$

In practice, the series are computed for a given n, by looping on k, starting at $k=n-4$ and decrementing by steps of 2. For each k, the partial series are initialised at $S_k(x) = 0$ for each x. We then loop on q, starting at $q_{\min}$ and compute γ_{qnk} using first the method of Appendix B.1 and then for $q \ge q_{\text{int}}$, that of Appendix B.2. At each step, we add $\alpha_{qnk}(x)\gamma_{qnk}$ to the partial series. If $S_k(x)$ does not change (within double precision) upon addition of three consecutive terms, the series is assumed to have

converged, and we simply deduce $F_{nk}^+(x) = -s^{k+1}S_k(x)$. Note that even if the series has converged before q reaches q_{int}, we will still restart the loop at q_{int} (with the second method) until convergence is ensured there. This is because for s close to 1, the added terms for $q_{min} < q < q_{int}$ can be very small, therefore mimicking convergence, even when the terms with $q \geq q_{int}$ do contribute to the final result (see also Appendix C for further details).

Appendix C. Additional comments on F_{nk}^+ when $s=1$

This final appendix is not necessary for our proposed implementation but may be useful in understanding the analytic cancellations in the series for $s=1$. It is with this in mind that it is included here.

We here consider further the special case of $s=1$, for which

$$F_{nk}^+(s=1, x) = \mathcal{P}^+(x\chi_n(x)\psi_k(x)). \tag{C.1}$$

We assume $n+k$ even and $n \geq k+4$, as these are the only cases where we need it. All functions depend on x and for conciseness this variable is implicit in the following.

We first define

$$\Delta_{nk} = \psi_n\chi_k - \chi_n\psi_k. \tag{C.2}$$

We will show that the series expansion of Δ_{nk} only contains negative odd powers of x, such that $(x\Delta_{nk})^+$ is only one term, for which we will find an analytic expression. In fact, Waterman wrote a similar series expansion in Ref. [21] (Eq. (28a)). We will here focus on the dominant term of the series to obtain a simpler expression.

Let us first note that

$$\Delta_{nn} = 0,$$
$$\Delta_{n+1,n} = \psi_{n+1}\chi_n - \chi_{n+1}\psi_n = 1. \tag{C.3}$$

The latter is Eq. (10.1.31) in Ref. [36] and a consequence of the Wronskian relation. These lead us to define

$$v_{pn} = x\Delta_{p+n,n} \tag{C.4}$$

and look for a recursion relation on p for v_{pn} at a fixed n. The initial conditions are simply

$$v_{0,n} = x\Delta_{nn} = 0,$$
$$v_{1,n} = x\Delta_{n+1,n} = x. \tag{C.5}$$

Moreover we have, using relations on Bessel functions,

$$\psi_{n+1}\chi_k + \psi_{n-1}\chi_k = \frac{2n+1}{x}\psi_n\chi_k, \tag{C.6}$$

$$\chi_{n+1}\psi_k + \chi_{n-1}\psi_k = \frac{2n+1}{x}\chi_n\psi_k \tag{C.7}$$

from which we deduce

$$\Delta_{n+1,k} + \Delta_{n-1,k} = \frac{2n+1}{x}\Delta_{nk} \tag{C.8}$$

or in terms of v_{pn}:

$$v_{p+1,n} = \frac{2n+2p+1}{x}v_{p,n} - v_{p-1,n}. \tag{C.9}$$

It is relatively straightforward to show by recursion using this relation and the initial conditions that

$$v_{2q-1,n}^+ = (-1)^{q+1}x,$$
$$v_{2q,n}^+ = (-1)^{q+1}q(2n+2q+1). \tag{C.10}$$

For $n+k$ even, we may therefore write

$$(x\Delta_{nk})^+ = v_{n-k,k}^+ = \frac{(-1)^{(n-k-1)/2}}{x}\frac{n-k}{2}(n+k+1). \tag{C.11}$$

In addition, for $n \geq k+4$, we have $(x\chi_k\psi_n)^+ = x\chi_k\psi_n$ (i.e. there is no negative powers of x in the series). We therefore deduce

$$F_{nk}^+(s=1, x) = x\chi_k(x)\psi_n(x) + (-1)^{(n-k)/2}(n+k+1)\frac{n-k}{2}. \tag{C.12}$$

Since the series expansion for $x\chi_k(x)\psi_n(x)$ starts with terms of order x^{n+1-k}, there is a potential large gap in the series expansion of $F_{nk}^+(s=1, x)$ between the zeroth order and the next non-zero order x^{n+1-k}. This corresponds to all the γ_{qnk} that are zero for $s=1$ and subject to loss of precision for s close to 1. The method presented in Appendix B.1 was developed specifically to circumvent these problems.

References

[1] Mishchenko MI, Travis LD. Gustav Mie and the evolving discipline of electromagnetic scattering by particles. Bull Amer Meteor Soc 2008;89:1853–61.
[2] Mishchenko MI, Hovenier JW, Travis LD, editors. Light scattering by nonspherical particles: theory, measurements, and applications. San Diego: Academic Press; 2000.
[3] Bohren CF, Huffman DR. Absorption and scattering of light by small particles. New York: John Wiley & Sons, Inc.; 1983.
[4] Stevenson A. Electromagnetic scattering by an ellipsoid in the third approximation. J Appl Phys 1953;24:1143–51.
[5] Wokaun A, Gordon JP, Liao PF. Radiation damping in surface-enhanced Raman scattering. Phys Rev Lett 1982;48:957–60.
[6] Le Ru EC, Somerville WRC, Auguié B. Radiative correction in approximate treatments of electromagnetic scattering by point and body scatterers. Phys Rev A 2013;87:012504.
[7] Meier M, Wokaun A. Enhanced fields on large metal particles: dynamic depolarization. Opt Lett 1983;8:581–3.
[8] Moroz A. Depolarization field of spheroidal particles. J Opt Soc Am B 2009;26:517–27.
[9] Asano S, Yamamoto G. Light-scattering by a spheroidal particle. Appl Opt 1975;14:29–49.
[10] Asano S. Light scattering properties of spheroidal particles. Appl Opt 1979;18:712–23.
[11] Farafonov V, Il'in V, Vinokurov A. Near-and far-field light scattering by nonspherical particles: applicability of methods that involve a spherical basis. Opt Spectrosc 2010;109:432–43.
[12] Farafonov V, Vinokurov A, Il'in V. Comparison of the light scattering methods using the spherical basis. Opt Spectrosc 2007;102:927–38.
[13] Barber P, Chang R, Massoudi H. Surface-enhanced electric intensities on large silver spheroids. Phys Rev Lett 1983;50:997–1000.
[14] Barber PW, Chang RK, Massoudi H. Electrodynamic calculations of the surface-enhanced electric intensities on large Ag spheroids. Phys Rev B 1983;27:7251–61.
[15] Boyack R, Le Ru EC. Investigation of particle shape and size effects in SERS using T-matrix calculations. Phys Chem Chem Phys 2009;11:7398–405.
[16] Mishchenko MI, Travis LD, Lacis AA. Scattering, absorption and emission of light by small particles. 3rd electronic release ed.Cambridge: Cambridge University Press; 2002.
[17] Mishchenko M, Travis L, Mackowski D. T-matrix computations of light scattering by nonspherical particles: a review. J Quant Spectrosc Radiat Transfer 1996;55:535–75.

[18] Mishchenko MI, Travis LD. Capabilities and limitations of a current FORTRAN implementation of the T-matrix method for randomly oriented, rotationally symmetric scatterers. J Quant Spectrosc Radiat Transfer 1998;60:309–24.
[19] Mishchenko MI, Travis LD. T-matrix computations of light scattering by large spheroidal particles. Opt Commun 1994;109: 16–21.
[20] Somerville WRC, Auguié B, Le Ru EC. Severe loss of precision in calculations of T-matrix integrals. J Quant Spectrosc Radiat Transfer 2012;113:524–35.
[21] Waterman PC. The T-matrix revisited. J Opt Soc Am A 2007;24: 2257–67.
[22] Waterman PC. Matrix formulation of electromagnetic scattering. Proc IEEE 1965;53:805–12.
[23] Mishchenko MI. Light scattering by randomly oriented axially symmetric particles. J Opt Soc Am A 1991;8:871–82.
[24] Khlebtsov N. Orientational averaging of light-scattering observables in the J-matrix approach. Appl Opt 1992;31:5359–65.
[25] Quirantes A. A T-matrix method and computer code for randomly oriented, axially symmetric coated scatterers. J Quant Spectrosc Radiat Transfer 2005;92:373–81.
[26] Mishchenko M. Light scattering by size-shape distributions of randomly oriented axially symmetric particles of a size comparable to a wavelength. Appl Opt 1993;32:4652–66.

[27] Mackowski DW, Mishchenko MI. Calculation of the T matrix and the scattering matrix for ensembles of spheres. J Opt Soc Am A 1996;13:2266–78.
[28] Peterson B, Ström S. T-matrix for electromagnetic scattering from an arbitrary number of scatterers and representations of E(3). Phys Rev D 1973;8:3661.
[29] Somerville WRC, Auguié B, Le Ru EC. Simplified expressions of the T-matrix integrals for electromagnetic scattering. Opt Lett 2011;36: 3482–4.
[30] Moroz A. Improvement of Mishchenko's T-matrix code for absorbing particles. Appl Opt 2005;44:3604–9.
[31] Petrov D, Shkuratov Y, Videen G. Optimized matrix inversion technique for the T-matrix method. Opt Lett 2007;32:1168–70.
[32] Wielaard D, Mishchenko M, Macke A, Carlson B. Improved T-matrix computations for large, nonabsorbing and weakly absorbing nonspherical particles and comparison with geometrical-optics approximation. Appl Opt 1997;36:4305–13.
[33] Strassen V. Gaussian elimination is not optimal. Numer Math 1969;13:354.
[34] Li L. Comment on "optimized matrix inversion technique for the T-matrix method". Opt Lett 2008;33:1366.
[35] Kuik F, DeHaan J, Hovenier J. Benchmark results for single scattering by spheroids. J Quant Spectrosc Radiat Transfer 1992;47:477–89.
[36] Abramowitz M, Stegun IA, editors. Handbook of mathematical functions. New York: Dover; 1972.