Journal of Quantitative Spectroscopy & Radiative Transfer 160 (2015) 29–35

Contents lists available at ScienceDirect

Journal of Quantitative Spectroscopy & Radiative Transfer

journal homepage: www.elsevier.com/locate/jqsrt

Accurate and convergent *T*-matrix calculations of light scattering by spheroids

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

ARTICLE INFO

Article history:
Received 20 January 2015
Received in revised form
13 March 2015
Accepted 17 March 2015
Available online 25 March 2015

Keywords:
Scattering
Mie theory
T-matrix
Electromagnetic optics
Numerical approximation and analysis

ABSTRACT

The convergence behavior of the *T*-matrix method as calculated by the extended boundary condition method (EBCM) is studied, in the case of light scattering by spheroidal particles. By making use of a new formulation of the EBCM integrals specifically designed to avoid numerical cancellations, we are able to obtain accurate matrices up to high multipole order, and study the effect of changing this order on both the individual matrix elements and derived physical observables. Convergence of near- and far-field scattering properties with a relative error of 10^{-15} is demonstrated over a large parameter space in terms of size, aspect ratio, and particle refractive index. This study demonstrates the capability of the *T*-matrix/EBCM method for fast, efficient, and numerically stable electromagnetic calculations on spheroidal particles with an accuracy comparable to Mie theory.

1. Introduction

The *T*-matrix method, as originally formulated by Waterman [1], also known as the extended boundary condition method (EBCM) or null-field method, is considered to be one of the most efficient semi-analytical approaches to model electromagnetic scattering by particles [2]. It is particularly suited to the calculation of orientation-averaged properties, which can be extremely time-consuming with fully numerical methods. The EBCM has been applied across many fields to calculate the optical properties of particles, e.g. in atmospheric measurements [3], astronomical studies [4], nano-science and particularly plasmonics [5–8], with a recent emphasis on near-field calculations [9–11]. With a theoretical footing closely related to Mie theory, but generalized to nonspherical particles, the *T*-matrix method lends itself naturally to analytical studies and improvements, e.g. with the introduction of discrete symmetries using group theory [12], or the enforcement of energy conservation and the study of radiative corrections [13]. As shown recently, the convergence properties of Mie theory are relatively simple and highly accurate results (e.g. 10^{-15} relative error in double precision) can be straightforwardly obtained over a large parameter range (of size and material) [14]. In contrast to Mie theory however, the EBCM suffers from a number of numerical instabilities, typically attributed to inversion of ill-conditioned matrices, which severely limits its range of applicability [15,16]. These problems are usually evidenced as a failure to obtain convergent results for physical observables, such as far-field cross-sections, with respect to the number of higher-order multipoles included in the simulation. This undesirable behavior – lack of convergence entails poor estimates of the accuracy of the calculations, and offers no possibility of reaching an arbitrary level of precision – is most easily noticeable for particles much larger than the wavelength, with a large

* Corresponding author.
E-mail addresses: wrcsomerville@gmail.com (W.R.C. Somerville), baptiste.auguie@gmail.com (B. Auguié), eric.leru@vuw.ac.nz (E.C. Le Ru).

http://dx.doi.org/10.1016/j.jqsrt.2015.03.020
0022-4073/

refractive index, or with highly anisotropic shapes (high aspect ratio). These numerical problems have so far precluded any detailed study of the convergence properties of the EBCM or *T*-matrix method itself, as they often arise before full convergence can be established. Some of these effects can be partly mitigated by increasing the floating point precision to quadruple [15] or arbitrary [17] precision, by using general symmetry relations for particles with point group symmetries [18], or by using the null-field method with multiple discrete sources [19,20]. However, these approaches significantly increase computational times and complexity, arguably negating the main advantages of the EBCM over fully numerical techniques. Alternatively, we have recently identified the cause of numerical cancellations in the special case of spheroidal particles [17] and developed new methods to overcome them [21]. This new approach opens up the possibility to study the convergence properties of the EBCM for spheroids with no interference from numerical issues, as demonstrated in this paper. The convergence of *T*-matrix elements, as well as both far-field and near-field properties is studied as a function of the maximum multipole order. We show that the convergence properties are more complicated than those of Mie theory and for example depend not only on size, but also on refractive index. Perhaps unexpectedly, convergence does not depend significantly on aspect ratio, however. We demonstrate the potential of the method for ultra-high precision calculations (10^{-15} relative error) over a large parameter space, including challenging cases such as highly elongated spheroids with aspect ratio 100. It is hoped that the tools and results presented in this work will demystify the convergence/numerical problems of the *T*-matrix/EBCM method and encourage its widespread application for the calculations of the optical properties of spheroids across diverse fields.

2. *T*-matrix method

A detailed description of the *T*-matrix/EBCM method can be found in e.g. [2] and we here only recall the definitions most relevant to this work. As for Mie theory, the field solutions are expressed as infinite series of the vector spherical wavefunctions (VSWFs), which represent magnetic ($\mathbf{M}$) and electric ($\mathbf{N}$) multipole fields. For example, for the scattered field, we have

$$\mathbf{E}_{\mathrm{sca}}(\mathbf{r}) = E_0 \sum_{n,m} p_{nm}\mathbf{M}_{nm}(k\mathbf{r}) + q_{nm}\mathbf{N}_{nm}(k\mathbf{r}), \tag{1}$$

where $n = 1\ldots\infty$ is the multipolar order, $|m| \le n$ the projected angular momentum number, and k the wavevector in the medium. The incident and internal fields are also expressed as similar expansions in terms of regular VSWFs (Rg$\mathbf{M}$ and Rg$\mathbf{N}$) with coefficients (a_{nm}, b_{nm}) and (c_{nm}, d_{nm}) respectively.

The EBCM expresses the linear relationship between these coefficients with two infinite matrices $\mathbf{P}$ and $\mathbf{Q}$,

$$\begin{pmatrix}\mathbf{p}\\ \mathbf{q}\end{pmatrix} = -\mathbf{P}\begin{pmatrix}\mathbf{c}\\ \mathbf{d}\end{pmatrix} \quad \text{and} \quad \begin{pmatrix}\mathbf{a}\\ \mathbf{b}\end{pmatrix} = \mathbf{Q}\begin{pmatrix}\mathbf{c}\\ \mathbf{d}\end{pmatrix}, \tag{2}$$

which are then used to form the *T*-matrix (*R*-matrix) linking the coefficients of the incident and scattered (internal) fields:

$$\begin{pmatrix}\mathbf{p}\\ \mathbf{q}\end{pmatrix} = \mathbf{T}\begin{pmatrix}\mathbf{a}\\ \mathbf{b}\end{pmatrix} \quad \text{and} \quad \begin{pmatrix}\mathbf{c}\\ \mathbf{d}\end{pmatrix} = \mathbf{R}\begin{pmatrix}\mathbf{a}\\ \mathbf{b}\end{pmatrix}, \quad \text{with} \begin{cases}\mathbf{T} = -\mathbf{P}\mathbf{Q}^{-1}\\ \mathbf{R} = \mathbf{Q}^{-1}.\end{cases} \tag{3}$$

Note that for convenience, the matrices are often described in 2×2 block form, separating the magnetic, electric, and coupled magneto-electric terms as follows:

$$\mathbf{T} = \begin{pmatrix}\mathbf{T}^{11} & \mathbf{T}^{12}\\ \mathbf{T}^{21} & \mathbf{T}^{22}\end{pmatrix}. \tag{4}$$

From these matrices, all physical properties can be expressed as series involving either the expansion coefficients (like the scattered field in Eq. (1)) or the *T*-matrix elements for orientation-averaged properties (e.g. scattering cross-section). In practice those series must be truncated at a maximum multipole order N and we will here focus on the convergence of the series as N is increased (as for Mie theory [14,22]). Note that in this work we use all values of $|m| \le n$ at each multipole order. This study will be limited to prolate spheroids and the relevant parameters of the problem will be the relative refractive index $s = n_2/n_1$, the size parameter $x_{\max} = kr_{\max}$, and the aspect ratio $h = r_{\max}/r_{\min}$ where $r_{\max}$ ($r_{\min}$) is the long (short) semi-axis, $k = 2\pi n_1/\lambda$ the wavevector in the incident medium, and n_1 (n_2) is the refractive index in the medium (particle).

Within the EBCM, the matrix elements of $\mathbf{P}$ and $\mathbf{Q}$ are calculated as surface integrals on the particle surface [2,23]. In the case of axisymmetric particles, a number of simplifications arise as a result of symmetries, most notably there is entire decoupling between subspaces corresponding to different m, i.e. all matrices are block-diagonal and the integrals are reduced to simple one-dimensional integrals, which are computed using standard Gaussian quadrature [2,23]. The emphasis here is not on these quadratures, and the number of quadrature points is always chosen large enough to ensure that the results do not depend on it within double precision accuracy. As recently shown [17], the computation of a subset of those integrals is the primary source of numerical instability for the special case of spheroids, as severe cancellations occur, resulting in errors by potentially many orders of magnitude. We will therefore use the algorithms developed in [21] and which have been shown to overcome those problems. Thanks to this new approach, we can therefore assume that $\mathbf{P}$ and $\mathbf{Q}$ are accurate to a high precision up to large multipole order. The focus here is on whether this high precision is retained in the computation of the *T*-matrix and the physical properties.

3. *T*-matrix convergence

In contrast to Mie theory where only the series need truncating (at $n=N$), we here have one additional source of potential problems associated with the inversion of

infinite matrices and their necessary truncation *before* inversion. In practice, once **P** and **Q** are calculated to a high precision up to multipole order N_Q, the T- and R-matrices are derived from Eq. (3) also up to multipole order N_Q using the inversion procedure described in [21]. From this step, there are two separate factors that may result in incorrect entries. Firstly, because these matrices are in theory infinite, there is no guarantee that the T-matrix is correct up to order N_Q. Secondly, any matrix inversion, especially for large matrices, can result in loss of precision because of ill-conditioning (whereby minute changes in the value of the Q-matrix may result in large errors in its computed inverse).

We first consider issues arising from the truncation of the matrix. The truncation of **P** and **Q** at some order N_Q may affect the matrix elements of **T** (and **R**) at some lower order, perhaps even at all orders. For the EBCM method to be valid and convergent, the matrix elements must scale in such a way that the higher order elements do not affect the lowest order elements of the inverted matrix (within some precision). One can then in principle choose N_Q large enough to ensure that the inverted matrix is correct (and unaffected by N_Q) up to at least some order $N < N_Q$. To quantify this, we will study the relative convergence of a given matrix element or physical property A, by computing the relative error ϵ of A (as a function of N or N_Q) with respect to the converged value A_∞ (obtained for large N or N_Q), namely:

$$\epsilon_A(N_Q) = \left|\frac{A(N_Q) - A_\infty}{A_\infty}\right|. \tag{5}$$

To illustrate this, we study in Fig. 1 the relative convergence as a function of N_Q of the first row and column of $\mathbf{T}^{22,m=0}$, which relates incident and scattered electric multipoles (**b** and **q**). For example, $T_{11}^{22,m=0}$ relates q_{10} to b_{10} and physically corresponds to the electric dipolar polarizability along the symmetry axis. It is clear that to obtain an accurate value for this low order element (shown in bold in Fig. 1(a)), larger **P** and **Q** matrices must be calculated, in stark contrast with Mie theory where this term is always correct from $N=1$ (for spherical shapes **P** and **Q** become diagonal, hence Eq. (3) does not mix multipolar orders). Denoting by Δ the number of extra multipoles required for **P** and **Q** to ensure that $T_{11}^{22,m=0}$ has fully converged, we see that Δ is of the order of 32 in the example of Fig. 1(a) ($x_{\max}=10$, $h=10$, $s=1.5+0.02i$). The same applies to $R_{11}^{22,m=0}$ (bold line in Fig. 1(c)), with the same number of extra multipoles required.

In practice, we may need a T- or R-matrix that is accurate for all matrix elements T_{nk}^m, R_{nk}^m up to some multipole order N. From the results of Fig. 1, and similar studies for other parts of the matrices (i.e. other blocks, other m, and other indices n,k), the following conclusions can be drawn:

- For matrix elements with $k \geq n$ (i.e. upper triangular parts of the matrices, see Fig. 1(a) and (c)), the relative convergence is very similar to that of $T_{11}^{22,m=0}$. Although $N_Q = k$ is the minimum required to calculate T_{nk}, a larger value of $N_Q = k+\Delta$ is needed to obtain full convergence, where Δ is slightly smaller but of the same order as for $T_{11}^{22,m=0}$ (it goes down from $\Delta=32$ for $k=1$ to $\Delta=24$ for larger k). Convergence is obtained with a high accuracy close to the best obtainable in double precision.
- For matrix elements with $n \geq k$ (i.e. lower triangular parts of the matrices, see Fig. 1(b) and (d)), the best relative convergence is obtained for $N_Q \approx \max(n,\Delta)$. This means that for $N \geq \Delta$, no extra multipoles (i.e. $N_Q = N$) are necessary to obtain the best possible accuracy. This limiting error, corresponding to the plateauing of the relative error in Fig. 1(b) and (d)), is higher than that obtained for the elements in the upper part of the matrix, of the order of 10^{-12} for T_{nk} and further increasing with n for R_{nk}.

This plateau in convergence can be attributed to loss of precision in the inversion step, i.e. even if N_Q is large enough, the linear system to obtain **T** is slightly ill-conditioned, resulting in a loss of precision of 3–4 orders of magnitude in this case. At large N_Q, the inclusion of extra matrix elements in **Q** should not influence the lower-order elements, but this "noise" of the order of the machine epsilon ($\sim 10^{-16}$) is magnified by the ill-conditioning, resulting in a plateauing of the relative error at 10^{-12}. This figure should be viewed as the accuracy of the method (and is parameter dependent). Rather than relying on the imperfection of double precision floating point arithmetic and to make sure that this problem is always detected, one can instead deliberately add noise to the matrix elements of **Q**. This is the procedure we adopted in the rest of our convergence studies with a random noise with a maximum relative magnitude of 10^{-15} added to every element of **Q**. This small noise is magnified by any ill-conditioning during inversion and allows us to obtain the true converged precision from the plateau region at large N_Q.

In this context, we also note that inversion in quadruple [15] or arbitrary precision [17] arithmetic or pre-conditioning of the linear system would be needed to further improve the T-matrix precision. It is worth noting here that the condition number, which is often used to assess how well a matrix may be inverted, does not provide the full picture when it comes to inversion. For example in the case of Fig. 1, the condition number of $\mathbf{Q}^{22,m=0}$ is terrible at $\sim 10^{54}$, yet only 3 digits precision is lost since the worst precision for the $\mathbf{T}^{22,m=0}$ matrix elements is 10^{-12} (when N_Q is large enough). In fact, for the linear system in Eq. (3) for spheroidal particles, the inversion algorithm plays a critical role in minimizing ill-conditioning effects. We here use the `mldivide` operator of MATLAB on the transposed matrices as recommended in [21]. Using `mrdivide` results in extreme errors (by many orders of magnitude). The choice of the right inversion algorithm is here equivalent to pre-conditioning the linear system.

Finally, we note that in the case of the lower part of **R**, the effect of ill-conditioning is more extreme, and the loss of precision during inversion increases with increasing n, with a loss of up to 13 orders of magnitude for $n=49$ in the example shown in Fig. 1(d). We will come back to this later since **R** is only used as an intermediate to compute

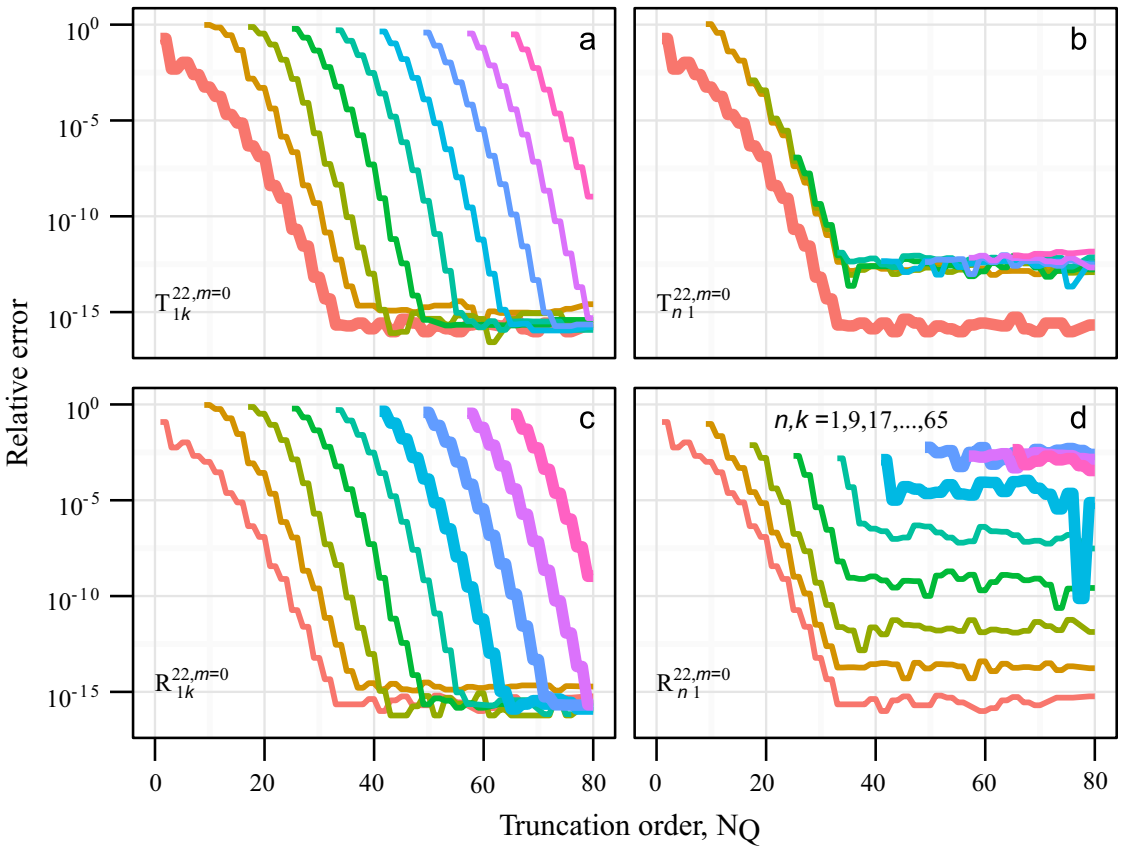

Fig. 1. Convergence of relative error in the matrix elements $T_{nk}^{22,m=0}$ in (a) the first row ($n=1$) and (b) first column ($k=1$), and the same for $R_{nk}^{22,m=0}$ (c, d). n and k are increased in steps of 8, from 1 to 65. The geometry is that of a large dielectric ($s=1.5+0.02i$) prolate spheroid with size parameter $x_{\max}=10$ and aspect ratio $h=10$, and all integrals are calculated with 500 quadrature points. The relative error for each matrix element is plotted against the size N_Q of the P- and Q-matrices and computed against converged results obtained using $N_Q=110$ (much more than needed in this case).

near-fields and focus on the more commonly studied T-matrix for the moment.

The conclusions presented so far for a specific scatterer were further confirmed by considering a wider range of parameters size, aspect ratio, and refractive index. We in particular studied how Δ and the converged precision ϵ vary with the properties of the scatterer. This can be assessed to a good approximation by considering the convergence of a single matrix element, for example $T_{11}^{22,m=0}$ or $T_{11}^{22,m=1}$. Examples are presented in Fig. 2 and Table 1. From those (and more extensive studies not shown here), we conclude that the size $x_{\max}$ and magnitude of the refractive index $|s|$, and to a lesser extent the aspect ratio h, all influence the value of Δ, which increases as the value of those parameters increases. Their respective effect does not however appear to be independent of each other. The converged precision remains very good over a wide parameter range and then quickly deteriorates when reaching certain limits, which can be viewed as the current range of validity of the algorithms presented in [21]. This limit typically corresponds to regimes where N_Q and Δ become too large for the EBCM to remain competitive anyway. For low-index dielectric particles, this occurs around $x_{\max}=30, h=10$ and $x_{\max}=20, h=100$. For higher refractive index materials like the metallic particles, this occurs earlier around $x_{\max}=20, h=2$ and $x_{\max}=10, h=100$. It should be noted that thanks to the improvements of [21], these limits are much larger than what is typically considered computable with the T-matrix EBCM method [15].

4. Symmetry and physical properties

In order to further support these conclusions, we will now extend these results to more conventional convergence tests based on the convergence of symmetry properties of the T-matrix [24–27] or of the computed physical properties [2]. We start with the symmetry properties of the calculated T-matrix that arise from optical reciprocity [2,28]. In the case of spheroids, these are ($\mathbf{M}^{\mathbf{T}}$ denotes matrix transpose of $\mathbf{M}$)

$$\begin{cases} \left(\mathbf{T}^{11}\right)^{\mathbf{T}} = \mathbf{T}^{11} \\ \left(\mathbf{T}^{22}\right)^{\mathbf{T}} = \mathbf{T}^{22} \end{cases} \quad \text{and} \quad \begin{cases} \left(\mathbf{T}^{12}\right)^{\mathbf{T}} = -\mathbf{T}^{21} \\ \left(\mathbf{T}^{21}\right)^{\mathbf{T}} = -\mathbf{T}^{12} \end{cases} \tag{6}$$

The relative errors in the symmetry for each matrix element of $\mathbf{T}^{22,m=1}$ are shown in Fig. 3(a) for the same particle as in Fig. 1. This plot confirms the previous conclusions: firstly, although the T-matrix is calculated

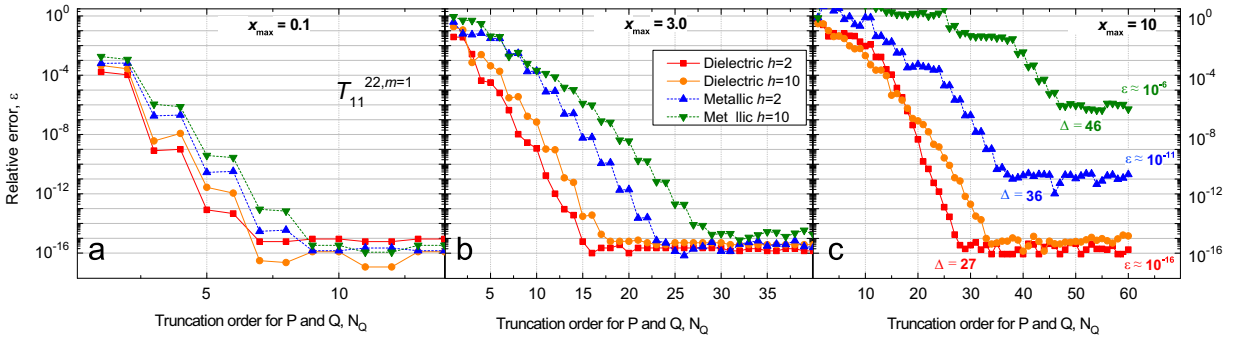

Fig. 2. Convergence of one of the lowest order element of the T-matrix, $T_{11}^{22,m=1}$, as a function of the size of the matrices P and Q. Four types of prolate spheroids are considered, either mildly absorbing dielectric ($s=1.5+0.02i$) or metallic ($s=\sqrt{-4+0.5i}$), both for small ($h=2$) and large ($h=10$) aspect ratios. For each, three sizes are shown from left to right: (a) $x_{max}=0.1$, (b) $x_{max}=3.0$, and (c) $x_{max}=10$. An increasingly larger N_Q is needed to obtain full convergence as either x_{max}, h, or $|s|$ increases. The relative precision is computed with respect to the value obtained for a large N_Q with a small random noise added to **P** and **Q** (a maximum of 10^{-15} relative amplitude is added to each matrix element).

Table 1
Summary of parameters governing the convergence of the EBCM method for spheroids of various sizes (x_{max} from 0.1 to 30), aspect ratios (h from 1.01 to 100), and refractive indices (s). These are extracted automatically from plots of the convergence of $T_{11}^{22,m=1}$, examples of which are shown in Fig. 2. They are here given as Δ (α), where $\Delta+1$ is the multipole order for which the converged plateau has been reached and α represents the accuracy of the converged value in number of digits, i.e. $\alpha=-\log_{10}\epsilon$ with ϵ being the limiting value of the relative error. Note that Δ is here likely to be slightly overestimated because it is extracted automatically from the plots. Entries marked – are beyond the capabilities of our algorithm, as they require large N_Q for which some entries are beyond double precision floating points (i.e. smaller than 10^{-300} or larger than 10^{300}). It is notable that the dependence of Δ on h is rather mild.

h	Dielectric, $s=1.5+0.02i$					Metal, $s=\sqrt{-4+0.5i}$				
	$x_{max}=0.1$	1	3	10	30	0.1	1	3	10	20
1.01	5 (15)	7 (16)	8 (16)	10 (16)	14 (16)	7 (16)	8 (16)	9 (16)	10 (16)	13(17)
1.2	7 (15)	10 (16)	13 (16)	22 (16)	53 (15)	9 (16)	13 (16)	17 (16)	25 (16)	26 (15)
2	8 (15)	13 (16)	18 (17)	29 (15)	82 (12)	10 (16)	16 (15)	26 (15)	39 (11)	–
10	8 (17)	15 (16)	21 (17)	36 (15)	–	10 (17)	20 (15)	30 (15)	48 (6)	–
30	9 (17)	15 (16)	21 (17)	35 (15)	–	11 (16)	20 (15)	29 (14)	50 (7)	–
100	9 (16)	15 (16)	22 (17)	36 (15)	–	10 (16)	20 (15)	30 (15)	52 (8)	–

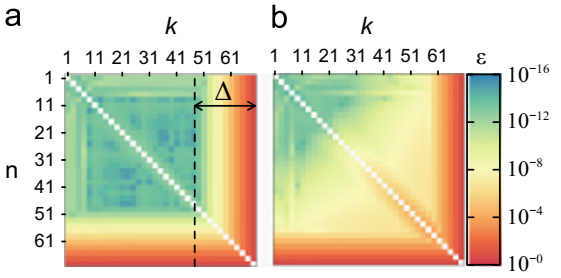

Fig. 3. Relative error in the symmetry of $\mathbf{T}^{22,m=1}$ for the same prolate spheroid as considered in Fig. 1: $x_{max}=10$, $h=10$, and $s=1.5+0.02i$. This relative error is computed as $\epsilon=2\left|T_{nk}^{22}-T_{kn}^{22}\right|/\left|T_{nk}^{22}+T_{kn}^{22}\right|$. **T** is computed with $N_Q=70$ multipoles, but the symmetry relations are only satisfied to high precision up to $N=47$. 500 quadrature points are used in (a), but only 150 in (b) to demonstrate the effect of insufficient quadrature precision. Similar plots are obtained for the other three blocks of the matrix.

with $N_Q=70$, it only satisfies the optical reciprocity symmetry relations up to about $N\approx 47$; and secondly the method developed in [21] appears to be extremely accurate, ensuring symmetry of all elements of the T-matrix (up to N) with a relative accuracy of 10^{-12} or better. We also use this example to caution about the effect of an insufficient quadrature precision in the calculations of the integrals in the P- and Q-matrices. This is exemplified in Fig. 3(b), where a reduction in the number of quadrature points clearly affects the precision of the results. It is therefore paramount to ensure that the number of quadrature points is large enough, for example by checking the matrix symmetry or that the results are independent of the precise number chosen.

It is satisfying to know that one can ensure that every single element of the T-matrix is correct to a high accuracy and this could indeed be important in specific fundamental studies. In practice, however, one rarely requires accurate convergence of *all* the elements of the T-matrix, but only of the derived physical properties, such as extinction cross-section or near fields. Some matrix elements in **T** or **R** may not be correct for larger n, but their contribution to the actual properties of interest may in fact be negligible. It is therefore equally important to study the convergence of those properties and we will here focus on a representative selection of far- and near-field properties, namely extinction and scattering cross-sections (orientation-averaged or not), surface-averaged electric field intensity, $\langle M_{Loc}\rangle=\langle|E|^2/|E_0|^2\rangle$, and electric field intensity on the surface at the tip (A) of the prolate spheroid, $M_{Loc}(A)$. To avoid any issues with the Rayleigh hypothesis [28], the surface fields are calculated from the internal

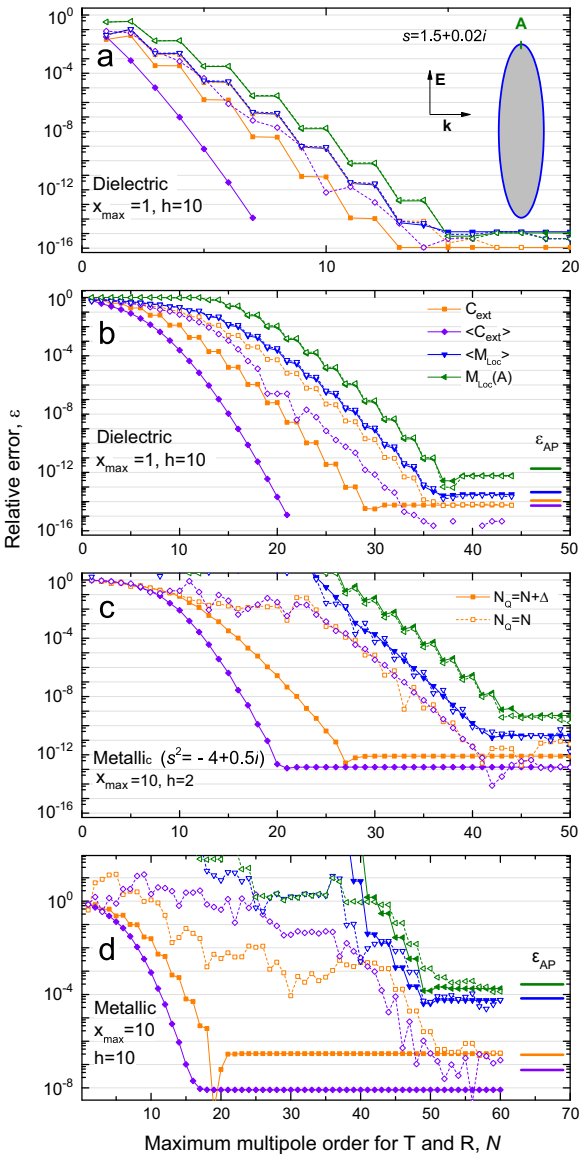

Fig. 4. Convergence of predicted physical properties as a function of maximum multipole order N (equivalent to truncation order for **T** and **R**) for four types of prolate spheroids. Far-field cross-sections are calculated for an incident field along x and polarized along the symmetry axis z. Note that only the extinction cross-section C_{ext} and its orientation-averaged analog $\langle C_{\text{ext}}\rangle$ are shown; absorption and scattering give similar results. We also study the surface-averaged field intensity $\langle M_{\text{Loc}}\rangle$ and the surface-field intensity at the tip of the spheroid $M_{\text{Loc}}(A)$. All these physical properties are calculated in two ways: in the first, **T** and **R** are calculated from **P** and **Q** with $N_Q = N+\Delta$, and are then truncated down to $N < N_Q$. In this case, the matrices are fully converged as shown earlier in this work. In the second approach, **T** and **R** are directly calculated with $N_Q = N$ (i.e. no truncation after inversion). The relative error is obtained by comparing with the values obtained for the largest N using $N_Q = N+\Delta$. In (b) and (d), we also show as horizontal bold lines the error ϵ_{AP} between this reference and the correct result computed in arbitrary precision.

field expansion and the boundary conditions are then used to deduce the outside surface fields [8].

Such a study is summarized in Fig. 4 for a number of representative cases. The convergence is again studied by considering the relative error compared to the converged result (i.e. obtained for large N with $N_Q = N+\Delta$) with the introduction of random noise in **P** and **Q** to assess the stability of the method. As before, convergence to a high precision is obtained in the majority of cases, with some loss of precision in the most challenging cases at the boundary of the applicability of the method (e.g. $x_{\max} = 10$, $h = 10$, $s = \sqrt{-4+0.5i}$). We also compare in Fig. 4 the exact approach derived here (using $N_Q = N+\Delta$) with a more direct approach where $N = N_Q$. It is interesting to see that, despite the fact that the higher order elements of the T- and R-matrices are incorrect when $N = N_Q$, the correct convergence is nevertheless obtained as N increases, indicating that the incorrect elements do not affect the results. This suggests that checking convergence with $N_Q = N$ would be more efficient in practice, as convergence would be obtained with a smaller value of N_Q, thus saving computing resources. $N_Q = N+\Delta$ should only be necessary in fundamental studies where the validity of the entire T-matrix is required.

Finally, one should also consider the possibility that the entire method converges to an incorrect solution, because of systematic numerical errors. To exclude this, the converged values were compared to those obtained with arbitrary precision arithmetic, where numerical problems can be detected and avoided by increasing the floating-point precision as needed [17]. As shown in Fig. 4, the double-precision results do converge to the correct solution, with exactly the precision predicted by introducing a small noise in **P** and **Q**.

A number of additional observations can be made from the results of Fig. 4. Once the numerical difficulties and the problems of matrix truncation are overcome, we have access to the intrinsic convergence properties of the various series as a function of multipole order. As is the case for Mie theory [14], more multipoles are needed as the particle size is increased, as expected from physical arguments. Also, some properties require more multipoles, for example the near fields when compared to far-field properties. Interestingly, the orientation-averaged cross-sections, which have no equivalent in Mie theory, converge faster than their non-averaged analogs. A major difference with Mie theory is that the number of multipoles is here strongly dependent on the refractive index: more are required for larger $|s|$. This could make the applicability of the method to large refractive index spheroidal particles more difficult.

We also note that the high accuracy of these calculations enables the detailed study of the validity of the Rayleigh hypothesis [28] in the vicinity of the particle. We have deliberately avoided this discussion here as it is outside the scope of this paper and would merit a dedicated study.

5. Discussion and conclusion

The results in this paper would naturally extend to the case of oblate spheroids, but not necessarily to other shapes. As shown in [17], the dominant matrix elements of **Q** show a scaling behavior that is very specific to spheroids (and this is at the heart of the numerical

problems of conventional EBCM implementations for spheroids). The convergence as a function of truncation N_Q could therefore be expected to be very different for other shapes. This would require a dedicated study although no numerical-problem-free implementations of the EBCM, similar to the one used here for spheroids, exist in the general case.

Our results demonstrate that in the special case of spheroids, the *T*-matrix/EBCM method exhibits convergence to a very high accuracy, even for relatively large and elongated particles (i.e. $x_{\max} = 10, h = 10 \ldots 100$). Contrary to what has been reported so far [15,17], this can moreover be achieved using standard double-precision arithmetic, providing an adequate implementation is used to avoid numerical errors in the integral computations and matrix inversion. Moreover, we have shown that the precision of those results can be assessed without needing to use other methods to compare against, by studying the convergence of the relative error or the symmetry of the *T*-matrix.

We believe that this study brings the prospect of routine, fast, accurate, and numerically stable electromagnetic calculations on spheroidal particles and will bring the EBCM/*T*-matrix implementation of [21] alongside Mie theory in the electromagnetic modelling toolbox of scientists of diverse fields.

Acknowledgments

We acknowledge the support of the Royal Society of New Zealand (RSNZ) through a Marsden Grant (VUW1107) and Rutherford Discovery Fellowship (VUW1002).

References

[1] Waterman PC. Matrix formulation of electromagnetic scattering. Proc. IEEE 1965;53:805–12.
[2] Mishchenko MI, Travis LD, Lacis AA. Scattering, absorption and emission of light by small particles. 3rd ed.. Cambridge: Cambridge University Press; 2002.
[3] Nousiainen T, Vermeulen K. Comparison of measured single-scattering matrix of feldspar particles with *T*-matrix simulations using spheroids. J Quant Spectrosc Radiat Transf 2003;79:1031–42.
[4] Draine BT, Fraisse AA. Polarized far-infrared and submillimeter emission from interstellar dust. Astrophys J 2009;696(1):1.
[5] Barber P, Chang R, Massoudi H. Surface-enhanced electric intensities on large silver spheroids. Phys Rev Lett 1983;50(13):997–1000.
[6] Khlebtsov BN, Khlebtsov NG. Multipole plasmons in metal nanorods: scaling properties and dependence on particle size, shape, orientation, and dielectric environment. J Phys Chem C 2007;111(31):11516–27.
[7] Qiu L, Larson TA, Smith D, Vitkin E, Modell MD, Korgel BA, et al. Observation of plasmon line broadening in single gold nanorods. Appl Phys Lett 2008;93(15):153106.
[8] 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.
[9] Doicu A, Wriedt T. Near-field computation using the null-field method. J Quant Spectrosc Radiat Transf 2010;111(3):466–73.
[10] Forestiere C, Iadarola G, Dal Negro L, Miano G. Near-field calculation based on the *T*-matrix method with discrete sources. J Quant Spectrosc Radiat Transf 2011;112(14):2384–94.
[11] Nilsson A, Alsholm P, Karlsson A, Andersson-Engels S. T-matrix computations of light scattering by red blood cells. Appl Opt 1998;37(13):2735–48.
[12] Kahnert M. T-matrix computations for particles with high-order finite symmetries. J Quant Spectrosc Radiat Transf 2013;123(0):79–91.
[13] 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(1):012504.
[14] Allardice JR, Le Ru EC. Convergence of Mie theory series: criteria for far-field and near-field properties. Appl Opt. 2014;53:7224–9.
[15] Mishchenko MI, Travis LD. T-matrix computations of light scattering by large spheroidal particles. Opt Commun 1994;109(1–2):16–21.
[16] Moroz A. Improvement of Mishchenko's *T*-matrix code for absorbing particles. Appl Opt 2005;44(17):3604–9.
[17] Somerville WRC, Auguié B, Le Ru EC. Severe loss of precision in calculations of *T*-matrix integrals. J Quant Spectrosc Radiat Transf 2012;113(7):524–35.
[18] Volkov SN, Samokhvalov IV, Kim D. Assessing and improving the accuracy of *T*-matrix calculation of homogeneous particles with point-group symmetries. J Quant Spectrosc Radiat Transf 2013;123:169–75.
[19] Doicu A, Wriedt T, Eremin YA. Light scattering by systems of particles.Berlin Heidelberg: Springer-Verlag; 2006.
[20] Hellmers J, Schmidt V, Wriedt T. Improving the numerical stability of *T*-matrix light scattering calculations for extreme particle shapes using the nullfield method with discrete sources. J Quant Spectrosc Radiat Transf 2011;112:1679–86.
[21] Somerville WRC, Auguié B, Le Ru EC. A new numerically stable implementation of the *T*-matrix method for electromagnetic scattering by spheroidal particles. J Quant Spectrosc Radiat Transf 2013;123:153–68.
[22] Wiscombe WJ. Improved Mie scattering algorithms. Appl Opt 1980;19:1505–9.
[23] Somerville WRC, Auguié B, Le Ru EC. Simplified expressions of the *T*-matrix integrals for electromagnetic scattering. Opt Lett 2011;36(17):3482–4.
[24] Doicu A, Wriedt T. Calculation of the *T* matrix in the null-field method with discrete sources. J Opt Soc Am A 1999;16(10):2539–44.
[25] Farafonov VG, Il'in VB. On checking the calculations of optical properties of non-spherical particles. Meas Sci Technol 2002;13(3):331–5.
[26] Rother T, Wauer J. Case study about the accuracy behavior of three different *T*-matrix methods. Appl Opt 2010;49(30):5746–56.
[27] Schmidt K, Yurkin MA, Kahnert M. A case study on the reciprocity in light scattering computations. Opt Express 2012;20(21):23253–74.
[28] Rother T, Kahnert M. Electromagnetic wave scattering on nonspherical particles. Berlin/Heidelberg: Springer; 2009.