PROCEEDINGS A

royalsocietypublishing.org/journal/rspa

Research

Cite this article: Gnidovec A, Božič A, Jelerčič U, Čopar S. 2022 Measure of overlap between two arbitrary ellipses on a sphere. *Proc. R. Soc. A* **478**: 20210807. https://doi.org/10.1098/rspa.2021.0807

Received: 19 October 2021
Accepted: 30 March 2022

Subject Areas:
computational physics, applied mathematics

Keywords:
hard ellipse repulsion, collision detection, dense packing, curved substrate

Author for correspondence:
Andraž Gnidovec
e-mail: andraz.gnidovec@fmf.uni-lj.si

Measure of overlap between two arbitrary ellipses on a sphere

Andraž Gnidovec[1], Anže Božič[2], Urška Jelerčič[3] and Simon Čopar[1]

[1]Department of Physics, Faculty of Mathematics and Physics, University of Ljubljana, Ljubljana, Slovenia
[2]Department of Theoretical Physics, Jožef Stefan Institute, Ljubljana, Slovenia
[3]Department of Chemical Engineering, Ilse Kats Institute for Nanoscale Science and Technology, Ben Gurion University of the Negev, Beer-Sheva, Israel

AG, 0000-0001-5881-0960; AB, 0000-0001-6304-6637; SČ, 0000-0002-7566-0260

Various packing problems and simulations of hard and soft interacting particles, such as microscopic models of nematic liquid crystals, reduce to calculations of intersections and pair interactions between ellipsoids. When constrained to a spherical surface, curvature and compactness lead to non-trivial behaviour that finds uses in physics, computer science and geometry. A well-known idealized isotropic example is the Tammes problem of finding optimal non-intersecting packings of equal hard disks. The anisotropic case of elliptic particles remains, on the other hand, comparatively unexplored. We develop an algorithm to detect collisions between ellipses constrained to the two-dimensional surface of a sphere based on a solution of an eigenvalue problem. We investigate and discuss topologically distinct ways two ellipses may touch or intersect on a sphere, and define a contact function that can be used for construction of short- and long-range pair potentials.

1. Introduction

It comes as no surprise that packing of ellipses and ellipsoids is a very thoroughly researched topic that

THE ROYAL SOCIETY PUBLISHING

appears in many different fields of research, both in experimental realizations and in numerical models used to study them. Ellipsoids appear in Gay–Berne (anisotropic Lennard–Jones) models [1] of liquid crystals as a coarse-grained replacement for the full molecular structure [2–4], in colloidal dispersions with an anisotropic dispersed phase [5–7], and in granular and jammed matter [8–11], where random and optimal packings are of particular interest [12,13]. All these examples are, however, Euclidean—yet many experimental systems call for a confinement of particles to a curved surface, often that of a sphere. Recent examples include packings of rods [14] and ellipsoids [15], spherocylinder simulations of nematics [16] and proteins adsorbed on vesicles [17,18]. This calls for an adaptation of ellipse–ellipse intersection algorithms for use on a spherical surface. Such an algorithm would also allow answering the question of optimal packing: while the well-researched Tammes problem [19,20] considers optimal packings of circles on a sphere, a generalization from circles to ellipses of arbitrary aspect ratios can provide us with the packing fraction for hard ellipses, which so far remains an open question. Furthermore, an algorithm which can be applied to ellipses of different sizes and aspect ratios opens up the possibility to consider polydisperse systems.

The bread-and-butter of computing ellipse–ellipse interactions lies in detecting collisions and overlaps in simulations of hard particles [21], and, for long-range interactions, measuring the closest distance between them [22]. One of the widely used and cited algorithms developed by Perram *et al.* [23,24] has been used, optimized and adapted in numerous ways and for various applications—in two dimensions (for ellipses) [25,26], three dimensions (for ellipsoids) [27–29] and was even generalized to hyperellipsoids [30]. However, all these algorithms are limited to Euclidean space and cannot be applied to the spherical case without modification.

In this work, we present a new algorithm that tackles the previously unsolved question of computing the distance and detecting overlap of ellipses confined to the two-dimensional surface of a sphere. Spherical confinement poses interesting challenges to the algorithm. Stretching is not a linear operation on a sphere, and two ellipses can interact in topologically different ways—if they interact at all. These situations differ strongly from the Euclidean case. We explain the intricacies of the spherical ellipse–ellipse interaction with examples, discuss the performance of the numerical algorithm and conclude by showing a few packing solutions.

2. Numerical algorithm

(a) Problem formulation

First, we define what constitutes an ellipse on the surface of a sphere. We adopt the conventional definition of an ellipse as the set of points with a constant sum of distances to the foci. To generalize it to a sphere, we require a constant sum of geodesic distances $\gamma_1 + \gamma_2 = 2\eta < \pi$ to the foci $\boldsymbol{f}_{1,2}$: $\cos\gamma_{1,2} = \boldsymbol{r}\cdot\boldsymbol{f}_{1,2}$. Without loss of generality, we can set $\boldsymbol{f}_{1,2} = \{\pm\sin\psi, 0, \cos\psi\}$. To convert the trigonometric relations into an algebraic form, we work with the cosine of the focal property $\cos(\gamma_1+\gamma_2) = \cos 2\eta$, which introduces an extraneous solution—another ellipse at the opposite side of the sphere due to $\cos(2\pi - 2\eta) = \cos 2\eta$. With further manipulations, we obtain a quadratic form representing an elliptic cylinder,

$$x^2\frac{\sin^2\psi}{\sin^2\eta} + z^2\frac{\cos^2\psi}{\cos^2\eta} = 1, \tag{2.1}$$

which is oriented in the (x, z) plane. However, on the unit sphere, the set of solutions is invariant to addition of any multiple of $x^2 + y^2 + z^2 = 1$, which gives a whole family of quadratic forms that specify the same ellipse pair. The most natural representation among these is an elliptic cylinder in the (x, y) plane with zero z^2 term, which simply projects a planar ellipse onto the sphere along

royalsocietypublishing.org/journal/rspa Proc. R. Soc. A 478: 20210807

the axis through the centre of the spherical ellipse

$$x^2 \frac{1}{\sin^2 \eta} + y^2 \frac{\cos^2 \psi}{\sin(\eta + \psi)\sin(\eta - \psi)} = 1. \tag{2.2}$$

We will thus define a spherical ellipse as an intersection of the unit sphere and an elliptical cylinder represented as a degenerate positive semidefinite quadratic form $a(\boldsymbol{r})$

$$a(\boldsymbol{r}) = \boldsymbol{r}A\boldsymbol{r}; \quad A = T\mathrm{diag}(\square, \square, 0)T^{\mathrm{T}}, \tag{2.3}$$

where T is a rotation matrix that will not be explicitly needed. We assume from now on that A is a given quantity which can be computed from any representation of the ellipses, such as by rotating the focal representation (equation 2.2), or from centre vectors and major semiaxis directions. Looking for an intersection of two arbitrary spherical ellipses is therefore equivalent to looking for an intersection of two quadratic forms and the unit sphere

$$a(\boldsymbol{r}) = \boldsymbol{r}A\boldsymbol{r} = 1, \tag{2.4}$$

$$b(\boldsymbol{r}) = \boldsymbol{r}B\boldsymbol{r} = 1 \tag{2.5}$$

and

$$||\boldsymbol{r}|| = 1. \tag{2.6}$$

Quadratic forms are invariant to inversion and produce *a pair* of antipodal ellipses when intersected with the unit sphere. This poses an additional challenge for the collision detection algorithm, as we must specify which ellipse is the correct one and which collisions to ignore. The correct ellipses can be specified by vectors $\boldsymbol{r}_A$ and $\boldsymbol{r}_B$ corresponding to the centres of the ellipses—*signed* eigenvectors corresponding to the zero eigenvalue of the quadratic forms a and b. The dot product between the ellipse centre and any point on the ellipse is positive for the correct ellipse and negative for the antipode.

Unlike in the Euclidean case, scaling the semiaxes of the quadratic form has an important effect on the topology of its intersection with the unit sphere. When the semiaxes are small compared with the radius of the sphere, the ellipses are similar to Euclidean ellipses. If the semiaxes are scaled to be comparable with the sphere radius, the apexes become sharper and converge to a 'lemon wedge' shape in the limit where the large semiaxis of the quadratic form matches the sphere radius. In this configuration, the antipodes touch at two 'poles', forming two intersecting great circles. Beyond this size, the intersection with the sphere splits again into a new pair of ellipses, but now their centres are directed along the shorter of the quadratic form semiaxes. At this crossover, the former antipodal pair recombines, and no longer corresponds to elliptical particles centred at $\boldsymbol{r}_{A,B}$. These cases with *inverted* ellipses will play a role in our theoretical analysis, but have no physical significance.

The goal of our algorithm is to detect when two ellipses are tangent or overlapping by defining a contact function and to obtain the contact point $\boldsymbol{v}$. If forces at the contact point are required, the direction of the force should be along the normal to the ellipse, which is given by the gradient of the quadratic form (magnitudes can be normalized—here we halve the expression to simplify notation)

$$\boldsymbol{n} = \tfrac{1}{2}\nabla_{\perp} \boldsymbol{r}A\boldsymbol{r}\big|_{\boldsymbol{r}=\boldsymbol{v}} = A\boldsymbol{r} - \boldsymbol{r}(\boldsymbol{r}A\boldsymbol{r})\big|_{\boldsymbol{r}=\boldsymbol{v}} = (A - I)\boldsymbol{v}. \tag{2.7}$$

From the force and the intersection point, we can also compute torques acting on the ellipse, which is useful for molecular dynamics simulations.

(b) Solving for ellipse contacts

Following the same steps as Perram and Wertheim [11,23,24], we define a linear interpolation of the quadratic forms,

$$q(\boldsymbol{r}, t) = \boldsymbol{r}Q(t)\boldsymbol{r}, \quad Q(t) = A(1 - t) + Bt, \tag{2.8}$$

with the parameter $t \in [0, 1]$, so that $q(\boldsymbol{r}, t) \geq 0$ on the whole sphere. Solution for contacts of spherical ellipses, i.e. level sets $a = 1$ and $b = 1$, is based on finding the minimal values of this

interpolation at each t. Constraining the solutions to the unit sphere, the problem can be restated in terms of finding the stationary points of the Lagrangian function

$$\mathcal{L} = q(\boldsymbol{r}, t) - \lambda(\boldsymbol{r} \cdot \boldsymbol{r} - 1). \tag{2.9}$$

Equation $\nabla\mathcal{L} = 0$ reduces to solving the eigenvalue problem $Q(t)\boldsymbol{r} = \lambda\boldsymbol{r}$, with solutions $\{\lambda_i(t)\}$ and eigenvectors $\{\boldsymbol{r}_i(t)\}$ that satisfy $||\boldsymbol{r}_i|| = 1$ (for $i = 1, 2, 3$). Let $\boldsymbol{r}_1$ be the eigenvector corresponding to the smallest eigenvalue λ_1. Plugging $\boldsymbol{r}_1$ back into expression (2.8), we get the minimum value of $q(\boldsymbol{r}, t)$ on the sphere,

$$q_{\min}(t) = q(\boldsymbol{r}_1(t), t) = \lambda_1(t). \tag{2.10}$$

Consider that the value of the quadratic form $q(t)$ is always greater than 1 in the part of the sphere that is outside both ellipses, as it is an interpolation of two values greater than 1. As t is varied from 0 to 1, $\boldsymbol{r}_1(t)$ will trace a continuous path on the sphere from $\boldsymbol{r}_1(0) = \boldsymbol{r}_A$ to $\boldsymbol{r}_1(1) = \boldsymbol{r}_B$ where $q_{\min}(0) = q_{\min}(1) = 0$. If the ellipses do not overlap, this means that $\boldsymbol{r}_1(t)$ will have to cross the region outside both ellipses for some t, where consequently $q_{\min}(t) > 1$. On the other hand, for overlapping ellipses, $q(\boldsymbol{r}, t)$ will always be smaller than 1 in the intersection region, meaning that $q_{\min}(t) < 1$ for all t. It follows from equation (2.10) that ellipses $a = 1$ and $b = 1$ intersect on the unit sphere if and only if the smallest eigenvalue of $Q(t)$ never exceeds 1 on the interval $t \in [0, 1]$. We denote this extremum Λ_1 and the corresponding eigenvector $\boldsymbol{v}_1$,

$$\Lambda_1 = \max \lambda_1(t) \quad \text{and} \quad Q\boldsymbol{v}_1 = \Lambda_1\boldsymbol{v}_1. \tag{2.11}$$

Positive definiteness ensures there are always three non-negative real eigenvalues, corresponding to the *casus irreducibilis* of the cubic equation, which is solvable in closed form through trigonometry. To find the maximum Λ_1, any one-dimensional maximization algorithm can be used, such as the golden section search. We can rely on this function being anticonvex with a single maximum, which ensures reliable and fast convergence.

To gain additional understanding of the relation between eigenvalue extrema and ellipse contacts, one can consider that the geometric representation of the unconstrained three-dimensional level set $q(t) = 1$ is a generic ellipsoid (or possibly a degenerate elliptical cylinder when one eigenvalue of $Q(t)$ is zero). The eigenvalues of $Q(t)$ correspond to inverse squares of its semiaxes. This ellipsoid is thus completely contained within the unit sphere if *all* its eigenvalues are greater than 1 and intersects the unit sphere if this is not the case. For non-overlapping ellipses, where $\Lambda_1 > 1$, the level set $q = 1$, constrained to the surface of the sphere, will therefore be empty for some t. This supports the fact that the space of allowed level set locations on the sphere is discontinuous if the ellipses do not overlap and the level set therefore cannot evolve continuously from $a = q(0) = 1$ to $b = q(1) = 1$ as t is increased from 0 to 1. Conversely, if the ellipses intersect, the level set on the sphere is always non-empty as $\Lambda_1 < 1$, with the intersection points of $a = 1$ and $b = 1$ a part of the level set for each t. Examples of disjoint, touching and overlapping ellipses, and the level sets of $q(t)$, are shown in figure 1*a–c*.

The value of Λ_1 has a clear geometric meaning: if we observe the intersection with a sphere $r^2 = 1/\Lambda_1$ instead of the unit sphere, the ellipses $a = 1$ and $b = 1$ touch at a single point of tangency, given by the appropriately scaled eigenvector $\boldsymbol{v}_1$. Scaling the system back to the unit sphere by a factor of $\sqrt{\Lambda_1}$, we see that Λ_1 is the factor by which the orthogonal projected area of both ellipses must be grown to make them tangent (scaling the semiaxes by $\sqrt{\Lambda_1}$). Values of $\Lambda_1 > 1$ signify non-overlapping ellipses that become tangent when grown, and values of $\Lambda_1 < 1$ overlapping ellipses that become tangent when shrunk. This property makes Λ_1 an appropriate choice for a contact function, with the same meaning it has in the Euclidean case (see the work of Perram and Wertheim [11,23,24]). However, without additional tests, the value of Λ_1 does not distinguish between the two antipodal ellipses represented by the same quadratic form and thus signals an overlap even when the ellipses in question are on the opposite sides of the sphere. For a usable algorithm, collisions with the antipodes of the ellipses represented by the quadratic forms must be ignored. This is handled in the following section.

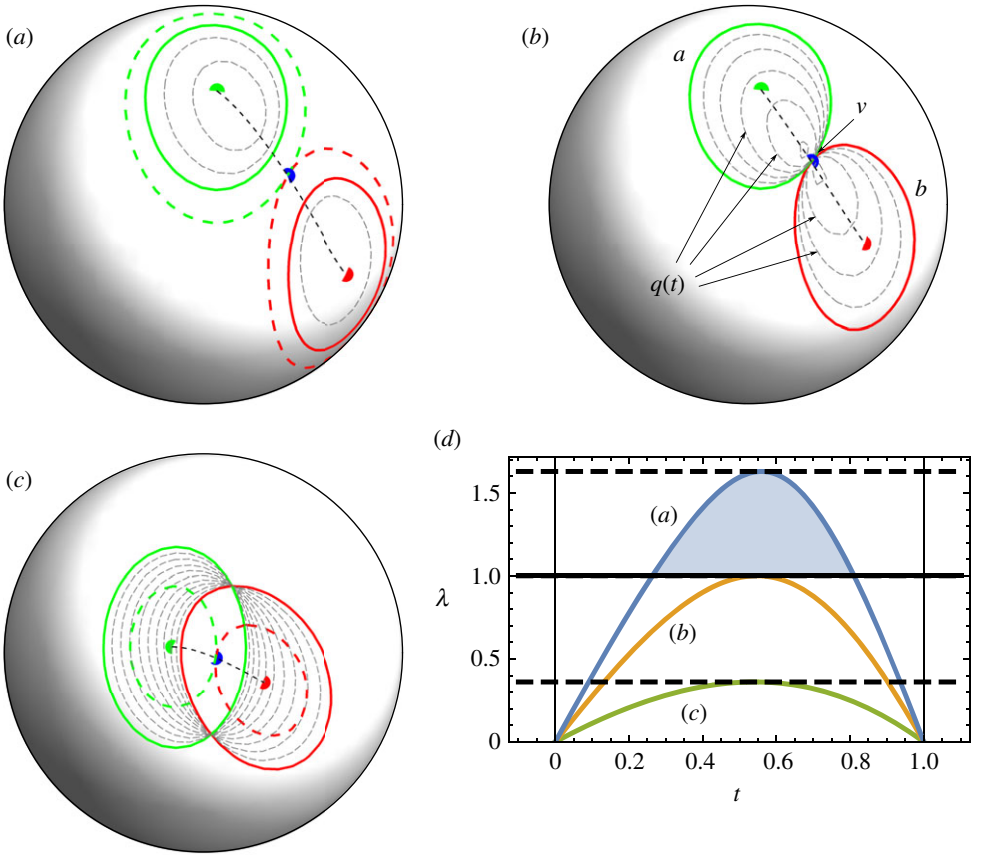

Figure 1. Cases of (*a*) disjoint, (*b*) touching and (*c*) overlapping ellipses on a sphere, with ellipses stretched to achieve tangency shown by dashed coloured lines. Level sets of $q(t) = 1$ (equation (2.8)) at different t are shown by dashed grey lines. Black dashed lines represent the lines described by the eigenvector corresponding to the smallest eigenvalue of $Q(t)$ when t runs from 0 to 1, and $\boldsymbol{v}$ marks the intersection point found when the eigenvalue is maximized. (*d*) Smallest eigenvalue with respect to t for the three cases in (*a–c*), showing that the smallest eigenvalue exceeds 1 when the ellipses are disjoint. (Online version in colour.)

(c) Solution branches and secondary contacts

Points of tangency of ellipses on the sphere can be defined in terms of the full intersection set of two elliptical cylinders in three dimensions,

$$\mathcal{S} = \{\boldsymbol{r}, a(\boldsymbol{r}) = b(\boldsymbol{r}) = 1\}. \tag{2.12}$$

The ellipses, obtained as intersections of a and b with the sphere of radius r, are intersecting at points on $\mathcal{S}$ at radius r and are tangent in *critical points* on $\mathcal{S}$ with locally extremal distance r from the origin. The maximized smallest eigenvalue Λ_1, which we derived in the previous section, simply corresponds to the critical point of $\mathcal{S}$ farthest from the origin; but this is just one of the critical points.

Degenerate cases aside, the set $\mathcal{S}$ consists of an antipodal pair of two disjoined loops. Each loop can have at most four critical points—two with locally maximal and two with locally minimal distance to the origin, corresponding to four values $r^{-2} = \{\Lambda_1, \Lambda_2, \Lambda_3, \Lambda_4\}$ (figure 2*a*). Depending on the relative orientation and size of the ellipses, there may be only two critical points, $r^{-2} = \{\Lambda_1, \Lambda_4\}$ (figure 2*b*). At the transition between these two regimes, the critical points Λ_2 and Λ_3 merge into an inflection point before disappearing. In other borderline cases with zero measure,

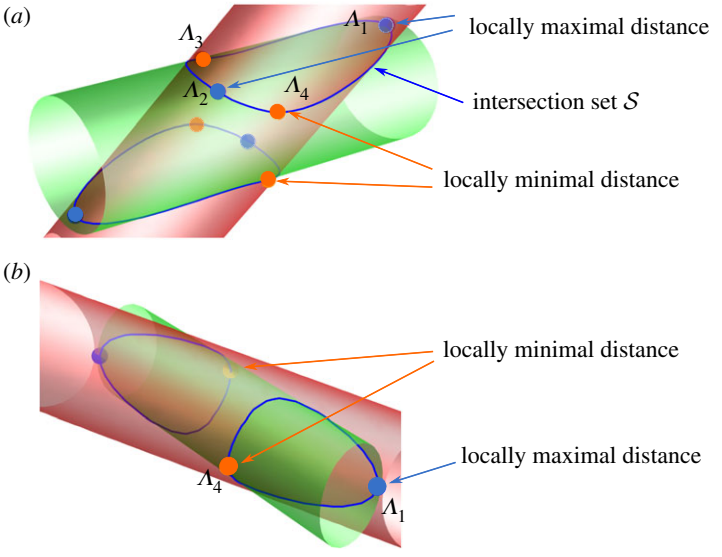

Figure 2. (*a*) A generic intersection set $\mathcal{S}$ of two obliquely intersecting elliptical cylinders. The intersection consists of two antipodal loops, with two points of maximal distance and two points of minimal distance from the origin. These represent the four tangency cases; only Λ_1 and Λ_2 are relevant for our analysis. (*b*) In exceptional cases, the two loops might be joined in a four-way junction. (Online version in colour.)

the intersection set may be a 'basket' with two fourfold junctions, or may have whole arcs at constant distance from the origin. These can all be understood as limiting cases with degenerate maxima and minima.

The maximal critical points $\Lambda_{1,2}$ correspond to the tangency with appearance of two new intersections when ellipses are stretched past the tangency condition. The minimal critical points $\Lambda_{3,4}$ correspond to the disappearance of intersections when stretching ellipses past the tangency condition. Only the maxima—the critical points $\Lambda_{1,2}$—are relevant for detecting ellipse contacts. The remaining two critical points $\Lambda_{3,4}$ involve inverted ellipses, as they describe points on $\mathcal{S}$ with locally minimal distance to the origin and are thus closer than at least one of the quadratic form semiaxes.

The antipodal doubling of ellipses means that the tangency at Λ_1 may correspond to the contact with the antipode of the second ellipse, so it might not be the one we are looking for. If there are only two critical points, there is no other possible contact. If there are four critical points, growing the ellipses further makes them touch again at the next locally maximal critical point (Λ_2). This contact might be between the correct pair of ellipses, or it could be between the same pair of ellipses as the Λ_1 critical point, in which case it is not a candidate for a true contact either.

As already discussed, the maximum of the lowest eigenvalue, Λ_1, solves for the first contact. The rest of the contacts can also be tied to extrema of the eigenvalues of $Q(t)$ over t. The values Λ_2 and Λ_3 correspond to the minimum and the maximum of the middle eigenvalue, and Λ_4 to the minimum of the largest eigenvalue (figure 3). Unlike the lowest eigenvalue of $Q(t)$, which is guaranteed to have a local maximum between $t = 0$ and $t = 1$, the remaining eigenvalues can have extrema outside the interval $t \in (0, 1)$, or none at all. In these cases, constrained minimization returns one of the edge points of the interval.

If there are only two critical points on each loop of the intersection manifold $\mathcal{S}$, then the middle eigenvalue has no local extrema, neither inside the interval $(0, 1)$ nor anywhere else on the real line, and $\Lambda_{2,3}$ are undefined. If there are four critical points, the local extrema may lie outside the interval $t \in (0, 1)$. This corresponds to a second contact between the same pair of ellipses as Λ_1,

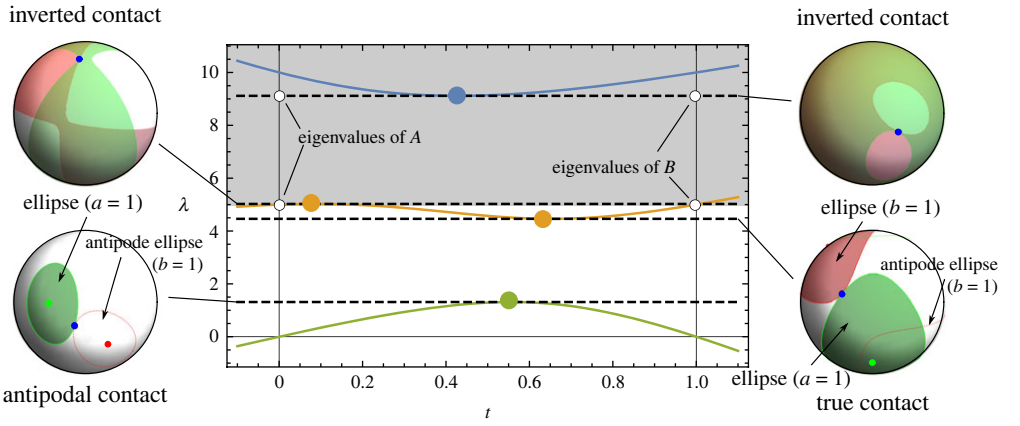

Figure 3. Eigenvalue spectrum of a generic case, with all four extrema Λ_i occurring inside the interval $t \in (0, 1)$. If the first contact is between the antipodes (lower left inset), the true contact and thus the correct value of the contact function is found by the minimum of the second eigenvalue. Observe that the green ellipse ($a = 1$) still intersects the red ($b = 1$) antipode ellipse, which we are ignoring. If the first contact is between the correct ellipses, then the lowest eigenvalue is the correct solution—we need information about the correct antipode to test for that. The upper two extrema correspond to inverted contacts (the shaded area lies above the lowest non-zero eigenvalue of A and B). (Online version in colour.)

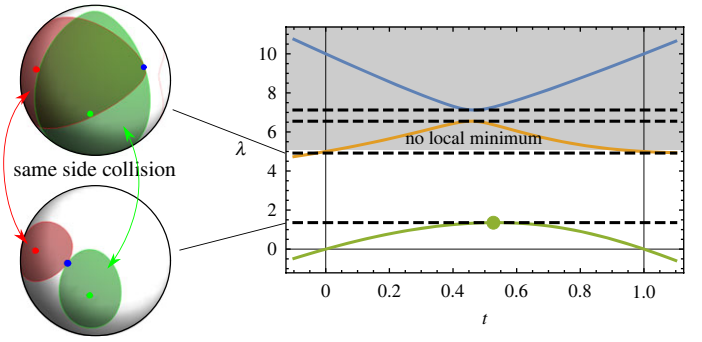

Figure 4. Eigenvalue spectrum of a case where the first two contacts are between the same ellipses, signifying that either the first contact is correct or neither of them is (as in the latter case, both contacts are between antipodes). Such situations are characterized by the middle eigenvalue not having a local minimum in the interval $0 < t < 1$. Top two contacts are not depicted. The shaded area corresponds to inverted ellipses. (Online version in colour.)

meaning that either both critical points signify contact between the true ellipses or both signify contact with the antipode, in which case there is no contact (figure 4). This is convenient, as simply checking for the existence of a minimum of the middle eigenvalue inside the interval $t \in (0, 1)$ includes all cases in which the critical point Λ_2 can constitute a real contact. Finally, if the resulting Λ_1 or Λ_2 exceed any of the eigenvalues of A or B (which coincide with the non-zero eigenvalues of $Q(t = 0)$ and $Q(t = 1)$), it signifies a contact where at least one ellipse is inverted. We can test this by finding the minimum non-zero eigenvalue Ω of A and B, corresponding to the largest semiaxis of the largest ellipse. Critical points that exceed this value, $\Lambda_{1,2} > \Omega$, do not correspond to valid contacts, nor can their values be unambiguously used as an analitical continuation of the contact function, because mixing of antipodes into the inverted ellipse makes the choice between the branches impossible.

(d) Contact function

To define a well-behaved contact function F to use as a test for ellipse–ellipse intersections, the correct eigenvalue must be selected. This is done with the help of the eigenvectors, which correspond to critical points. We denote with $\pm\boldsymbol{v}_1$ and $\pm\boldsymbol{v}_2$ the eigenvectors corresponding to Λ_1 and Λ_2, respectively, and $\boldsymbol{r}_A$ and $\boldsymbol{r}_B$ are the true centres of the ellipses on the unit sphere, their signs picking the correct ellipse of the antipodal pair. If the true ellipse collides with the antipode of the second one, the projections of the intersection vector onto the vectors of ellipse centres are of opposite signs, and vice versa. If there is no contact, or the contact is with an inverted ellipse, assigning the value $F = \Omega$ makes the function continuous under variations of the relative position of the ellipses. The full algorithm for computing the contact function is described in algorithm 1.

Algorithm 1. Ellipse-ellipse contact function.

Result: Contact function F
$\Omega \leftarrow \min \text{eigenvalue}_{2,3}(Q(0), Q(1))$;
$t_1 \leftarrow \text{argmax}_{t\in[0,1]} \text{eigenvalue}_1(Q(t))$;
$\Lambda_1 \leftarrow Q(t_1)$;
$\boldsymbol{v}_1 \leftarrow \text{eigenvector}(Q(t_1), \Lambda_1)$;
if $(\mathbf{r}_A \cdot \mathbf{v}_1)(\mathbf{r}_B \cdot \mathbf{v}_1) > 0$ **then**
  if $\Lambda_1 < \Omega$ **then**
    return Λ_1;
  else
    return Ω;
  end
else
  $t_2 \leftarrow \text{argmin}_{t\in[0,1]} \text{eigenvalue}_2(Q(t))$;
  $\Lambda_2 \leftarrow Q(t_2)$;
  if $0 < t_2 < 1$ **and** $\Lambda_2 < \Omega$ **then**
    return Λ_2;
  else
    return Ω;
  end
end

The contact function F, which is according to the above criterion equal to Λ_1, Λ_2, or Ω, can be used either to directly detect when ellipses overlap ($F < 1$) or to construct a pair potential. Instead of a hard core repulsion, a soft repulsion potential for overlapping cases $F < 1$ can be defined based on the value of F, such as $-\ln F$, F^{-1}, $F^{-1} + F - 2$ or $1 - F$, the last being a soft potential of finite strength at complete overlap. On the other hand, long-range values of $F > 1$ could act as a distance metric, e.g. in a Lennard–Jones-like potential, as they do in Euclidean space [22]. Setting the function to Ω in cases for which the ellipses cannot intersect no matter the stretch factor, ensures a constant potential and zero force on the particles for that entire region, and makes the function well-behaved for use in methods that require a potential (e.g. Monte Carlo methods). Even though there is no correspondence between such an artificially fashioned potential and any physical phenomena we know of, such an academic exercise could provide a reasonable approximation to medium-range behaviour that could match empirical observations in certain physical systems.

3. Examples

(a) Intersection of unequal circles

The simplest example that can be used for interpretation of the contact function F is a pair of unequal circles. Define the following pair of quadratic forms:

$$\left.\begin{aligned} a &= \alpha(x^2 + y^2) \\ \text{and} \qquad b &= \beta(x^2\cos^2\theta + z^2\sin^2\theta - 2xz\cos\theta\sin\theta + y^2), \end{aligned}\right\} \tag{3.1}$$

with $\alpha > \beta > 1$ and θ the angular separation of the circle centres. In this case, the extremal eigenvalues (without applying the restriction to $0 < t < 1$) have a relatively simple closed form, and the contact function can be expressed as

$$F(\theta) = \frac{\alpha\beta\sin^2\theta}{\alpha + \beta + 2\sqrt{\alpha\beta}\cos\theta}. \tag{3.2}$$

The function's behaviour with respect to θ is depicted in figure 5 for a few combinations of circle sizes α and β. At $\theta > \pi/2$, we have $F(\theta) = \Lambda_2$, corresponding to the second contact, as the first contact is with the antipode. We observe that the crossover between the branches is continuously differentiable. However, with the exception of equal circles, we see that the function reaches a maximum at $\cos^2\theta = \beta/\alpha$ and then goes back to zero at $\theta = \pi$. This part of the contact function corresponds to the second collision also being with the antipode. The collision is internal (non-facing normals), and the interpolation parameter at minimal middle eigenvalue is $t > 1$. In our algorithm, we assign these collisions $F = \Omega$.

(b) Computational cost of the algorithm

The algorithm itself is fast, as the eigenvalue calculation can be expressed in a closed form, although it uses trigonometric functions that are slower than simple multiplications. One-dimensional minimization and maximization routines are available in any number of numerical libraries. We implemented two such routines, the golden section search (GSS) and the Brent method (GSS with quadratic interpolation), and compared both the numbers of eigenvalue evaluations N_{eval} to achieve the desired accuracy (tolerance of 10^{-7} in t) as well as calculation times. In figure 6, we show the results for a pair of ellipses on a unit sphere with major and minor semiaxes $\xi_1 = 0.5$ and $\xi_2 = 0.15$, respectively (aspect ratio $\varepsilon = \xi_1/\xi_2 \approx 3.33$). At a given angular separation θ, the contact function and the computational cost required to determine it with the Brent method depend on orientations of both ellipses as shown for $\theta = \pi/3$ in figure 6*a* and 6*b* for the number of first and second eigenvalue evaluations, respectively. The number of evaluations for Λ_1 mostly lies between 10 and 20, with the exceptions of diagonals with fewer evaluations and two loops with $N_{\text{eval}} \sim 30$ that correspond to the cases near the Λ_1 and Λ_2 crossover. For relatively high aspect ratios ε, as is the case in the demonstrated example, the second derivative close to the crossover becomes large, which is unfavourable for the Brent minimization. Inside these loops, the second eigenvalue becomes relevant for the contact function, as indicated in figure 6*b* (Λ_2 is only evaluated in regions where the Λ_1 eigenvector test fails, see algorithm 1). Additionally, closer to the diagonals, the local minimum of $\lambda_2(t)$ inside the interval $[0, 1]$ disappears and the algorithm returns Ω. Note again that despite the algorithm branch changes, the contact function is continuous in the whole configuration space.

We evaluate the necessary computational cost to determine the contact function both for the GSS and Brent methods. The results with respect to the angular separation θ are shown in figure 6*c*, where the value at each distance represents the average number of eigenvalue evaluations over the whole orientational domain (10^6 points, figure 6*a*,*b*). The number of Λ_1 evaluations with the GSS method remains (almost) constant for all distances, as a fixed number of interval divisions is necessary to achieve the desired precision. This number is also markedly

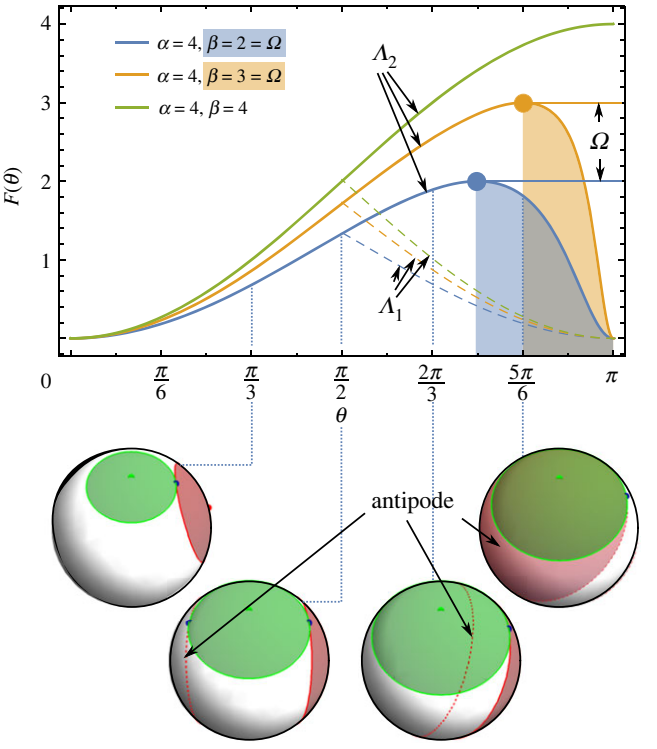

Figure 5. Contact function for circles of different relative radii separated by angle θ. Transition to the antipodal contact at $\theta > \pi/2$ is continuous; the insets show that the secondary contact is the correct one, while the contact with the antipode (dashed circle outline without infill) is ignored. If the circles are of the same size, the contact function is monotonously increasing, but if they are of different sizes, the decreasing part (shaded below the curve) corresponds to the case when the ellipses cannot be made to touch by stretching, and the corresponding eigenvalue detects the second contact between the 'wrong' pair of ellipses. In this region, the value of F is set to Ω (horizontal line), which corresponds to the inverse square radius of the larger circle. Parameters α and β correspond to inverse square radii (see equation (3.1)). (Online version in colour.)

higher compared with the Brent method, which shows that quadratic interpolation is highly effective for this problem (this could be expected from eigenvalue curves in figures 3 and 4). Note that the number of Λ_1 evaluations is symmetric around $\theta = \pi/2$, as elliptical cylinder configurations are invariant to coordinate transformation $\theta \to \pi - \theta$ and only the antipode interpretations for the correct/wrong ellipse are exchanged. The number of Λ_2 evaluations does not show this symmetry. At small angular separations, the first eigenvalue will always be the correct one and only for higher θ does the second eigenvalue evaluation become necessary in parts of the orientational space (figure 6*b*). These regions become larger as θ is increased (at some point, they consume the whole orientational domain), which in turn increases the average Λ_2 evaluation numbers.

In some situations, e.g. for simulations of hard particles, the calculation of the exact contact function is not needed. The optimization algorithm can be terminated immediately after a value of $\Lambda_1 > 1$ is encountered, as that means no overlap. The average number of Λ_1 evaluations with this early termination (ET) condition is shown in figure 6*c* with dashed lines and leads to a sharp decrease of the necessary calculations in a large part of the plot. As shown in (*a*) for $\theta = \pi/3$, more than one Λ_1 evaluation is necessary only inside the white contour which grows/shrinks for smaller/larger distances. Additionally, the grey region in the plot highlights the distances where the overlap appears only for certain ellipse orientations ($2\arcsin\xi_2 \leq \theta \leq 2\arcsin\xi_1$)—on the left side of this region, ellipses overlap for all orientations and on the right, overlap is not possible as they are too distant and the eigenvalue calculation can be skipped entirely.

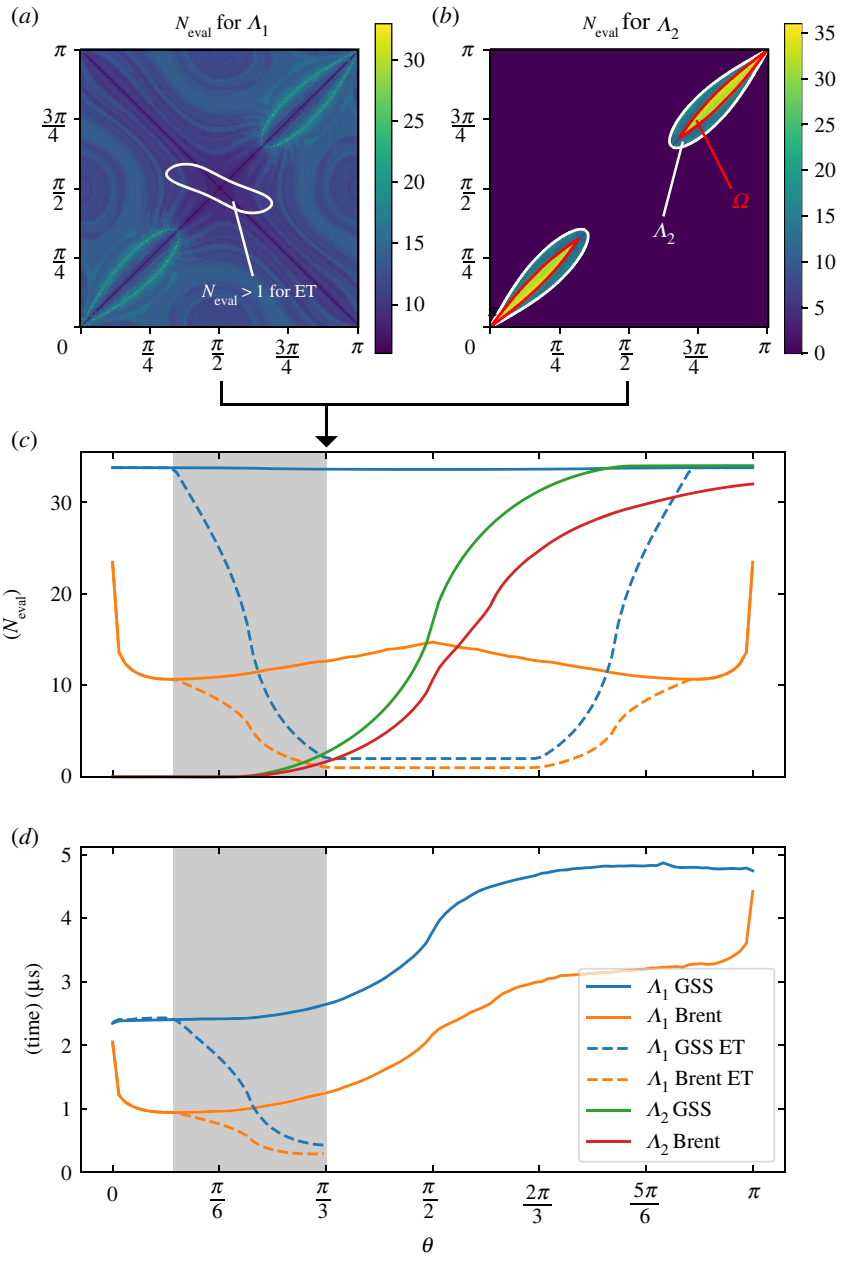

Figure 6. Number of eigenvalue evaluations and contact function calculation times for a pair of ellipses with $\xi_1 = 0.5$ and $\xi_2 = 0.15$. Number of evaluations needed to determine (*a*) Λ_1 and (*b*) Λ_2 for the Brent method at angular separation $\theta = \pi/3$ in the whole orientational domain. For Λ_2, $N_{\text{eval}} = 0$ in a large part of the domain where Λ_1 is the correct eigenvalue for determining the contact function (outside of the white contours in (*b*)). Around this eigenvalue crossover, the number of Λ_1 evaluations is increased. The white contour line in (*a*) surrounds the region where $N_{\text{eval}} > 1$ even with the ET enabled. The red contours in (*b*) show the border where the contact function transitions to the constant value of $F = \Omega$. (*c*) Number of contact evaluations for GSS and Brent line minimizators. The increase in Λ_2 evaluations is a consequence of growing regions where Λ_1 is not the correct eigenvalue. ET (dashed lines) significantly decreases the number of Λ_1 evaluations. The grey area corresponds to distances where the overlap of ellipses depends on their orientations; on the left side of this region, overlap is guaranteed, while there can be no overlap on the right side of the region. (*d*) Comparison of contact function calculation times for GSS and Brent methods. ET results are relevant only for $\theta < 2 \arcsin \xi_1$, as they can only be used to determine overlap/no overlap. (Online version in colour.)

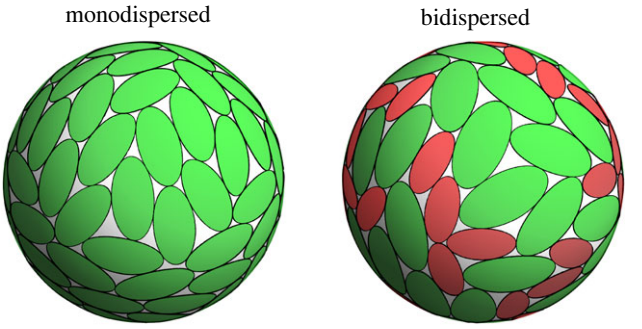

Figure 7. Examples of monodispersed and bidispersed dense packings of $N = 100$ spherical ellipses with aspect ratio $\varepsilon = 2$. In the bidispersed case, half of all ellipses are smaller by a factor of 1.4. (Online version in colour.)

Finally, figure 6*d* shows the average calculation time to evaluate the contact function. The results are on the order of µs and closely follow the combined number of Λ_1 and Λ_2 evaluations from (*c*), with the increase in calculation times corresponding to additional evaluations needed to determine the second eigenvalue at larger distances. If ET is enabled, the efficiency of the calculation is significantly improved.

(c) Dense packings of spherical ellipses

To demonstrate the use case of the proposed algorithm in multiparticle simulations, we calculated dense packings of $N = 100$ spherical ellipses with $\varepsilon = 2$ for both monodispersed and bidispersed systems (figure 7). We employed an energy minimization-based approach similar to the scheme used by Mailman *et al*. [31] where the system is randomly initialized at a packing fraction far from the jamming point, with subsequent iterative increases of particle sizes and relaxations to remove all overlaps. As angular separation θ between the centres of neighbouring (touching) ellipses is smaller than $\pi/2$ for our system parameters (N and ε), it is sufficient to calculate only the minimal first eigenvalue to determine the contacts—possible cases with antipodal contacts can be excluded based on ellipse separation alone.

4. Discussion

Depending on the requirements, the algorithm can be optimized further. For example, with SIMD instructions, evaluations at multiple $t \in (0, 1)$ could be performed with minimal overhead, allowing for faster determination of the correct eigenvalue branch and narrower initial bracket for the optimization algorithm. For the purposes of collision-driven molecular dynamics, the expensive $\mathcal{O}(n^2)$ complexity of evaluating pair interactions for a large number of particles can be alleviated by keeping track of nearest neighbours (e.g. by adapting pre-existing methods that make use of the contact functions [32]). Tracking and changing particle positions and orientations while keeping their shapes constant requires keeping track of the rotation matrices in a numerically stable form, which can be done either by tracking the ellipse centre vectors and the vector of its principal component (e.g. through Euler angles) or by using unit quaternions.

Our algorithm is largely based on the algorithm of [23] but has some important differences due to the differences between spherical and Euclidean geometry. On the one hand, spherical geometry of the problem makes it simpler, because in the Euclidean space, translations and rotations have to be considered separately, while on the sphere, the only parameter for the position and orientation of the ellipse is a single rotation matrix. Similarities can be partially restored by handling the Euclidean case in homogeneous (projective) coordinates, but then the confinement surface is a plane, not a sphere, resulting in a different algorithm. Due to this

difference, our algorithm requires solving an eigenvalue problem and not a linear system of equations. In general, the eigenvalue problem is numerically more expensive, but for 3×3 matrices, a closed-form solution is available.

From the aspect of finding the correct solutions, the spherical version of the algorithm is more involved, as the configuration space of possible intersections is topologically non-trivial and splits into different parts based on the behaviour of eigenvalue bands with respect to the parameter t. The antipodal doubling means we need additional information to treat different branches of the solution differently. However, as shown in our work, this can be done with a few trivial tests, with the only caveat that the long-range contact function is spliced and undefined (clipped to Ω) in parts of the configuration space.

5. Conclusion

The simplicity and speed of the presented algorithm makes it a viable workhorse for future simulations on a sphere, be it interactions of hard particles or general long-range interactions where distances are needed, although the concept of the contact function as a distance metric must be considered with care. Collision detection and generalized distance can be used for Monte Carlo simulations, while molecular dynamics can make use of the intersection vector and the normal vector to the surface as well. Elongation of particles is known to affect optimal packing fraction of random packings in Euclidean space [9,33], and with the presented algorithm, related questions can be answered for packings on a sphere.

Simulations can also be augmented with other potentials that do not use the contact function—for example, multipolar interactions, which may account for elliptical magnetic particles or electrostatically charged macromolecules. The algorithm is viable for particles of different aspect ratios and sizes, so it can be used for simulations of polydisperse particle systems. Another important use case is in representation of arbitrarily shaped objects as isosurfaces of Gaussian sums (called blobs or metaballs in three-dimensional graphics). A product of Gaussians, resulting directly in addition of quadratic forms when constraned to a sphere, also resembles posterior Bayesian update when handling probability models for directional or geographical data, which may be relevant in data processing and machine learning.

Finally, more fundamental questions can also be tackled. Recall that both the Tammes problem and its long-range potential cousin, the Thomson problem, have been well studied not only by physicists but also from the perspective of fundamental and applied mathematics and computer science. Generalization to an anisotropic case is a richer example, which without doubt hides many undiscovered facts about spherical packings.

Data accessibility. Code is available at https://git0.fmf.uni-lj.si/gnidovec/SphericalEllipseOverlap.

Authors' contributions. A.G.: conceptualization, formal analysis, methodology, software, visualization, writing—original draft, writing—review and editing. A.B.: funding acquisition, supervision, validation, writing—review and editing. U.J.: conceptualization, validation, writing—review and editing. S.Č.: conceptualization, funding acquisition, methodology, project administration, supervision, visualization, writing—original draft, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interests.

Funding. We acknowledge support by Slovenian Research Agency (ARRS) under contracts no. P1-0099 and no. J1-9149. The work is associated with the COST action no. CA17139.

References

1. Gay JG, Berne BJ. 1981 Modification of the overlap potential to mimic a linear site-site potential. *J. Chem. Phys.* **74**, 3316–3319. (doi:10.1063/1.441483)
2. Luckhurst G, Simmonds P. 1993 Computer simulation studies of anisotropic systems. *Mol. Phys.* **80**, 233–252. (doi:10.1080/00268979300102241)

3. Zannoni C. 2001 Molecular design and computer simulations of novel mesophases. *J. Mater. Chem.* **11**, 2637–2646. (doi:10.1039/b103923g)
4. Allen MP. 2019 Molecular simulation of liquid crystals. *Mol. Phys.* **117**, 2391–2417. (doi:10.1080/00268976.2019.1612957)
5. van Dillen T, van Blaaderen A, Polman A. 2004 Shaping colloidal assemblies. *Mater. Today* **7**, 40–46. (doi:10.1016/S1369-7021(04)00345-1)
6. Roller J, Geiger JD, Voggenreiter M, Meijer JM, Zumbusch A. 2020 Formation of nematic order in 3D systems of hard colloidal ellipsoids. *Soft Matter* **16**, 1021–1028. (doi:10.1039/C9SM01926J)
7. Roller J, Laganapan A, Meijer JM, Fuchs M, Zumbusch A. 2021 Observation of liquid glass in suspensions of ellipsoidal colloids. *Proc. Natl Acad. Sci. USA* **118**, e2018072118. (doi:10.1073/pnas.2018072118)
8. Donev A. 2006 Jammed packings of hard particles. PhD thesis, Princeton, NJ: Princeton University.
9. Donev A. 2004 Improving the density of jammed disordered packings using ellipsoids. *Science* **303**, 990–993. (doi:10.1126/science.1093010)
10. Man W, Donev A, Stillinger FH, Sullivan MT, Russel WB, Heeger D, Inati S, Torquato S, Chaikin PM. 2005 Experiments on random packings of ellipsoids. *Phys. Rev. Lett.* **94**, 141. (doi:10.1103/PhysRevLett.94.198001)
11. Donev A, Connelly R, Stillinger FH, Torquato S. 2007 Underconstrained jammed packings of nonspherical hard particles: ellipses and ellipsoids. *Phys. Rev. E* **75**, 6026. (doi:10.1103/PhysRevE.75.051304)
12. Chaikin PM, Donev A, Man W, Stillinger FH, Torquato S. 2006 Some observations on the random packing of hard ellipsoids. *Ind. Eng. Chem. Res.* **45**, 6960–6965. (doi:10.1021/ie060032g)
13. Jin W, Jiao Y, Liu L, Yuan Y, Li S. 2017 Dense crystalline packings of ellipsoids. *Phys. Rev. E* **95**, 033003. (doi:10.1103/PhysRevE.95.033003)
14. Smallenburg F, Löwen H. 2016 Close packing of rods on spherical surfaces. *J. Chem. Phys.* **144**, 164903. (doi:10.1063/1.4947256)
15. Xie Z, Atherton TJ. 2021 Elongation and percolation of defect motifs in anisotropic packing problems. *Soft Matter* **17**, 4426–4433. (doi:10.1039/D0SM02174A)
16. Bates MA. 2008 Nematic ordering and defects on the surface of a sphere: a Monte Carlo simulation study. *J. Chem. Phys.* **128**, 104707. (doi:10.1063/1.2890724)
17. Frost A, De Camilli P, Unger VM. 2007 F-BAR proteins join the BAR family fold. *Structure* **15**, 751–753. (doi:10.1016/j.str.2007.06.006)
18. Frost A, Perera R, Roux A, Spasov K, Destaing O, Egelman EH, De Camilli P, Unger VM. 2008 Structural basis of membrane invagination by F-BAR domains. *Cell* **132**, 807–817. (doi:10.1016/j.cell.2007.12.041)
19. Clare BW, Kepert DL. 1986 The closest packing of equal circles on a sphere. *Proc. R. Soc. Lond. A* **405**, 329–344. (doi:10.1098/rspa.1986.0056)
20. Saff EB, Kuilaars ABJ. 1997 Distributing many points on a sphere. *Math. Intell.* **19**, 5–11. (doi:10.1007/BF03024331)
21. Michele CD. 2010 Simulating hard rigid bodies. *J. Comput. Phys.* **229**, 3276–3294. (doi:10.1016/j.jcp.2010.01.002)
22. Everaers R, Ejtehadi MR. 2003 Interaction potentials for soft and hard ellipsoids. *Phys. Rev. E* **67**, 3316. (doi:10.1103/PhysRevE.67.041710)
23. Perram JW, Wertheim M. 1985 Statistical mechanics of hard ellipsoids. I. Overlap algorithm and the contact function. *J. Comput. Phys.* **58**, 409–416. (doi:10.1016/0021-9991(85)90171-8)
24. Perram JW, Rasmussen J, Præstgaard E, Lebowitz JL. 1996 Ellipsoid contact potential: theory and relation to overlap potentials. *Phys. Rev. E* **54**, 6565–6572. (doi:10.1103/PhysRevE.54.6565)
25. Paramonov L, Yaliraki SN. 2005 The directional contact distance of two ellipsoids: coarse-grained potentials for anisotropic interactions. *J. Chem. Phys.* **123**, 194111. (doi:10.1063/1.2102897)
26. Zheng X, Palffy-Muhoray P. 2007 Distance of closest approach of two arbitrary hard ellipses in two dimensions. *Phys. Rev. E* **75**, 061709. (doi:10.1103/PhysRevE.75.061709)
27. Zheng X, Iglesias W, Palffy-Muhoray P. 2009 Distance of closest approach of two arbitrary hard ellipsoids. *Phys. Rev. E* **79**, 057702. (doi:10.1103/PhysRevE.79.057702)
28. Guevara-Rodríguez FJ, Odriozola G. 2011 Hard ellipsoids: analytically approaching the exact overlap distance. *J. Chem. Phys.* **135**, 084508. (doi:10.1063/1.3626805)

29. Choi MG. 2020 Computing the closest approach distance of two ellipsoids. *Symmetry* **12**, 1302. (doi:10.3390/sym12081302)
30. Gilitschenski I, Hanebeck UD. 2014 A direct method for checking overlap of two hyperellipsoids. In *2014 Sensor data fusion: trends, solutions, applications (SDF)*, pp. 1–6. IEEE. (doi:10.1109/SDF.2014.6954724)
31. Mailman M, Schreck CF, O'Hern CS, Chakraborty B. 2009 Jamming in systems composed of frictionless ellipse-shaped particles. *Phys. Rev. Lett.* **102**, 255501. (doi:10.1103/PhysRevLett.102.255501)
32. Donev A, Torquato S, Stillinger FH. 2005 Neighbor list collision-driven molecular dynamics simulation for nonspherical hard particles. I. Algorithmic details. *J. Comput. Phys.* **202**, 737–764. (doi:10.1016/j.jcp.2004.08.014)
33. Delaney G, Weaire D, Hutzler S, Murphy S. 2005 Random packing of elliptical disks. *Phil. Mag. Lett.* **85**, 89–96. (doi:10.1080/09500830500080763)