



Spheres Unions and Intersections and Some of their Applications in Molecular Modeling

Michel Petitjean

► To cite this version:

Michel Petitjean. Spheres Unions and Intersections and Some of their Applications in Molecular Modeling. Antonio Mucherino, Carlile Lavor, Leo Liberti, Nelson Maculan. Distance Geometry: Theory, Methods, and Applications., Springer New York, pp.61-83, 2013, 978-1-4614-5127-3. 10.1007/978-1-4614-5128-0_4 . hal-01955983

HAL Id: hal-01955983

<https://hal.science/hal-01955983>

Submitted on 14 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

Chapter 2

Spheres unions and intersections and some of their applications in molecular modeling

Michel Petitjean

Abstract The geometrical and computational aspects of spheres unions and intersections are described. A practical analytical calculation of their surfaces and volumes is given in the general case: any number of intersecting spheres of any radii. Applications to trilateration and van der Waals surfaces and volumes calculation are considered. The results are compared to those of other algorithms, such as Monte-Carlo methods, regular grid methods, or incomplete analytical algorithms. For molecular modeling, these latter algorithms are shown to give strongly overestimated values when the radii values are in the ranges recommended in the literature, while regular grid methods are shown to give a poor accuracy. Other concepts related to surfaces and volumes of unions of spheres are evoked, such as Connolly's surfaces, accessible surface areas, and solvent excluded volumes.

2.1 Introduction

We denote by E^d the d -dimensional Euclidean space. The relation between spheres intersections and distance geometry can be exemplified by the trilateration problem: given, in E^3 , three fixed points c_1, c_2, c_3 with known coordinates, locate an unknown point x from its respective distances $d(x, c_1), d(x, c_2), d(x, c_3)$ to these fixed points. This problem can be reformulated as a spheres intersection problem: given three spheres of respective centers c_1, c_2, c_3 and respective radii $R_1 = d(x, c_1), R_2 = d(x, c_2), R_3 = d(x, c_3)$, locate the points at the intersection of their boundaries. Such a reformulation allows us to realize immediately that, when the centers are not aligned and when the intersection of the three spheres is not void, there are in general two solution points which are mirror images through the plane containing c_1, c_2, c_3 . Then the experimentalist can decide which of these two solution points is relevant, e.g. via an appropriate determinant calculus.

Michel Petitjean
MTi, University Paris 7, INSERM UMR-S 973, e-mail: petitjean.chiral@gmail.com

Many applications occur in molecular modeling because atoms can be modeled as hard spheres and molecules can be modeled as unions of spheres. The values of the atomic radii depend on how they are defined and measured. Usually, they are needed to compute the van der Waals surface and the van der Waals volume of a molecule, i.e. the surface and the volume defined by the union of the atomic spheres of the molecule. Other molecular concepts which are based on atomic spheres include the Accessible Surface Areas, the Connolly surfaces, and the Solvent Excluded Volumes [4, 20, 24, 25, 36, 47, 54]. They all depend on an additional probe sphere assumed to modelize a solvent molecule. Discussions on the physical meaning of the molecular surfaces and volumes and on the adequate choice of radii values have been done in [5, 29, 30, 31], but this is not in the scope of this chapter. Practical radii values can be found in [5, 15, 23, 43, 46, 50, 55]. It happened also that whole molecules were implicitly or explicitly modeled by spheres. E.g., in the alpha-shape model of pockets and channels in proteins [13], the ligand is represented by a probe sphere. The alpha-shape model is strongly connected with Delaunay triangulations and Voronoi diagrams [11, 12]. More recently these channels were computed as union of spheres centered on the vertices of a grid [41, 42]. In any case, structural chemists know that the shapes of the molecules are generally far from being spherical: better models are minimal height and minimal radius enclosing cylinders [39], but spheres are much easier to handle so they are still much used for molecular modeling.

2.2 The analytical calculation

We consider n spheres in E^d of given centers and radii, and we look for the calculation of the surface and the volume of their union or of their intersection. Let V_i be the volume of the sphere i , $i \in \{1..n\}$, with fixed center c_i and fixed radius R_i . We denote by $V_{i_1 i_2}$ the volume of the intersection of the spheres i_1 and i_2 , $V_{i_1 i_2 i_3}$ the volume of the intersection of the spheres i_1 and i_2 and i_3 , etc. Similarly, S_i is the surface (i.e. its area) of the sphere i , $S_{i_1 i_2}$ is the one of the intersection of the spheres i_1 and i_2 , etc. V is the volume of the union of the n spheres and S is its surface. Although we will exhibit the full analytical calculation of V and S only for $d = 3$, it is enlighting to do some parts of this calculation in E^d . We will specify the dimension when needed.

2.2.1 Spheres and lens

We set $n = 1$: we consider one sphere of fixed radius R in E^d . We denote respectively by $V_d(r)$ and $S_d(r)$ the volume and the surface of this sphere as functions of the radius r . E.g., $V_1(R)$ is the length of a segment of half-length R , $S_2(R)$ is the perimeter of the circle of radius R and $V_2(R)$ is its surface, etc.

Theorem 2.1. *The volume and the surface of the sphere of radius R in E^d are respectively given in equations (2.1) and (2.2).*

$$V_d(R) = \frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2} + 1)} R^d, \quad d \geq 1 \quad (2.1)$$

$$S_d(R) = d \frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2} + 1)} R^{d-1}, \quad d \geq 2 \quad (2.2)$$

Proof. We know that $V_{-1}(R)$ and $S_{-2}(R)$ stand. We get $V_{-2}(R)$ by integration: $V_{-2}(R) = \int_0^R S_{-2}(r) dr$, and conversely $S_{-2}(R)$ is retrieved by derivating $V_{-2}(R)$. For similar reasons, it suffices to prove equation (2.1) to be true and equation (2.2) is proved to stand by derivation of $V_d(R)$. We proceed by recurrence and we calculate $V_d(R)$ by integration of $V_{-(d-1)}(h)$ with $h = \sqrt{R^2 - r^2}$, as indicated in Fig. 2.1: $V_d(R) = 2 \int_0^R V_{-(d-1)}((R^2 - r^2)^{\frac{1}{2}}) dr$.

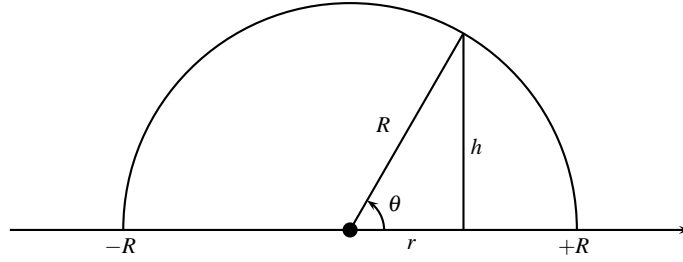


Fig. 2.1 Calculation of the sphere in E^d via summation of the volumes of spheres in E^{d-1}

Setting $r = R\sqrt{t}$, the integral is expressed with the β function:

$$V_d(R) = R^d \frac{\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d+1}{2})} \int_0^1 t^{-\frac{1}{2}} (1-t)^{\frac{d-1}{2}} dt = R^d \frac{\pi^{\frac{d-1}{2}}}{\Gamma(\frac{d+1}{2})} \beta\left(\frac{1}{2}, \frac{d+1}{2}\right).$$

Since $\beta\left(\frac{1}{2}, \frac{d+1}{2}\right) = \frac{\Gamma(\frac{1}{2})\Gamma(\frac{d+1}{2})}{\Gamma(\frac{d}{2} + 1)}$ and $\Gamma(\frac{1}{2}) = \pi^{\frac{1}{2}}$, we get the desired result.

Remark: expanding the expression of the Γ function shows that $V_d(R)$ and $S_d(R)$ are proportional to $\pi^{\frac{d}{2}}$ (when d is even) or to $\pi^{\frac{d-1}{2}}$ (when d is odd).

It is useful to calculate the volume $V_d(R, \theta)$ and the surface $S_d(R, \theta)$ of the spherical cap defined by the angle θ in Figure 2.1. They can be calculated by re-

currence as above via integration using respectively the expressions of $V_{\lfloor d-1 \rfloor}(R, \theta)$ and $S_{\lfloor d-1 \rfloor}(R, \theta)$, and with the help of incomplete β functions.

For clarity, we recall in equations (2.3) and (2.4) the results for $d = 3$, obtained, respectively, from summation of the elementary cylinders volumes $(\pi h^2)(dr)$ and of the elementary truncated cones surfaces $(2\pi h)(Rd\theta)$:

$$V_3(R, \theta) = \frac{\pi R^3}{3}(1 - \cos \theta)^2(2 + \cos \theta), \quad (2.3)$$

$$S_3(R, \theta) = 2\pi R^2(1 - \cos \theta). \quad (2.4)$$

Remark: in equations (2.3) and (2.4), θ takes values in $[0; \pi]$.

2.2.2 Lens and radical hyperplanes

We set $n = 2$. The intersection of two spheres in E^d of respective centers c_1 and c_2 and radius R_1 and R_2 is either empty, or reduces to one point, or is a lens, or is the smallest sphere in the case it is included in the largest sphere. The case of interest is the one of one lens.

The lens exists when:

$$|R_1 - R_2| \leq \|c_2 - c_1\| \leq R_1 + R_2. \quad (2.5)$$

This lens is bounded by two spherical caps separated by a $(d - 1)$ -hyperplane orthogonal to the direction $c_2 - c_1$ and intersecting the axis $c_2 - c_1$ at the point t_{12} . This $(d - 1)$ -hyperplane is called a *radical hyperplane*, or simply a *radical plane* when $d = 3$.

Theorem 2.2. *The location of the intersection point t_{12} is given in equations (2.6)–(2.8):*

$$t_{12} = \left(\frac{c_1 + c_2}{2} \right) + \frac{R_1^2 - R_2^2}{\|c_2 - c_1\|^2} \left(\frac{c_2 - c_1}{2} \right) \quad (2.6)$$

$$\|t_{12} - c_1\| = \frac{1}{2} \left(\|c_2 - c_1\| + \frac{R_1^2 - R_2^2}{\|c_2 - c_1\|} \right) \quad (2.7)$$

$$\|t_{12} - c_2\| = \frac{1}{2} \left(\|c_2 - c_1\| + \frac{R_2^2 - R_1^2}{\|c_2 - c_1\|} \right) \quad (2.8)$$

Proof. We denote with quotes the transposed vectors, e.g. t'_{12} is the transposed of t_{12} , and $t'_{12}t_{12} = \|t_{12}\|^2$. The intersection of the lens with its radical hyperplane defines a $(d - 1)$ -sphere (i.e. a disk when $d = 3$), of radius L_{12} to be calculated further. We define y_{12} as being any point on the boundary of this $(d - 1)$ -sphere in the radical hyperplane. Considering the right triangles c_1, t_{12}, y_{12} and c_2, t_{12}, y_{12} , we have $L_{12}^2 = R_1^2 - (t_{12} - c_1)'(t_{12} - c_1) = R_2^2 - (t_{12} - c_2)'(t_{12} - c_2)$. We express t_{12} with its

barycentric coordinates relative to c_1 and c_2 : $t_{12} = \alpha_1 c_1 + \alpha_2 c_2$, $\alpha_1 + \alpha_2 = 1$. Solving for the unknown quantity α_1 and after elimination of the term $t'_{12}t_{12}$, we get $\alpha_1 = \frac{1}{2} + \frac{R_2^2 - R_1^2}{2\|c_2 - c_1\|^2}$ and then $\alpha_2 = \frac{1}{2} + \frac{R_1^2 - R_2^2}{2\|c_2 - c_1\|^2}$.

Moreover, θ_1 and θ_2 being the angles respectively associated to each spherical cap (see Figure 2.1), the barycentric coefficients of t_{12} are the respective cosine of these angles.

$$\alpha_1 = \cos \theta_1 = \frac{1}{2} + \frac{R_2^2 - R_1^2}{2\|c_2 - c_1\|^2} \quad (2.9)$$

$$\alpha_2 = \cos \theta_2 = \frac{1}{2} + \frac{R_1^2 - R_2^2}{2\|c_2 - c_1\|^2} \quad (2.10)$$

Then the radius L_{12} of the $(d-1)$ -sphere bounding the lens is:

$$4L_{12}^2 = 2(R_1^2 + R_2^2) - \left(\frac{R_1^2 - R_2^2}{\|c_2 - c_1\|} \right)^2 - \|c_2 - c_1\|^2 \quad (2.11)$$

The surface of the triangle defined by c_1, c_2, y_{12} is $L_{12}\|c_2 - c_1\|/2$. Then we express this surface from R_1, R_2 , and $\|c_2 - c_1\|$ with the Heron formula [53]: the expression of L_{12} above comes after expansion and squaring.

There is a major difference about the validity of equations (2.6)–(2.8) and (2.11): the latter is valid for non concentric spheres if and only if the inequalities (2.5) stand, although the former stand if and only if $c_1 \neq c_2$. Thus, the radical plane exists for any radius values, even null ones, discarding whether or not the intersection of the two spheres is empty and discarding if one sphere is included in the other one. In equations (2.9)–(2.10), α_1 and α_2 are defined for any pair of non concentric spheres, although the existence of θ_1 and θ_2 need that inequalities (2.5) stand. When $R_1 = R_2$, the radical hyperplane is just the $(d-1)$ -hyperplane mediator of the segment $c_2 - c_1$.

Then, the volume V and the surface S of the union of the two spheres are respectively expressed from the volume and the surface of their intersection:

$$V = V_1 + V_2 - V_{12}, \quad (2.12)$$

$$S = S_1 + S_2 - S_{12}. \quad (2.13)$$

V_{12} and S_{12} can be respectively calculated from the volume and the surface of the two spherical caps bounding the lens. Knowing t_{12} from equation (2.6), the calculation is done as indicated at the end of Section 2.2.1 with the angles θ_1 and θ_2 respectively associated to each spherical cap (see Figure 2.1). These angles are known from equations (2.7) and (2.8): $|\cos \theta_1| = \|t_{12} - c_1\|/R_1$ and $|\cos \theta_2| = \|t_{12} - c_2\|/R_2$. In the case $d = 3$, we simply use equations (2.3) and (2.4). The values $\cos \theta_1$ and $\cos \theta_2$ are taken from equations (2.9) and (2.10). Remark: they can be negative.

2.2.3 More about radical hyperplanes

We consider n spheres in E^d , not two of them being concentric, so that we have defined $n(n-1)/2$ radical hyperplanes.

The intersection of the radical hyperplanes orthogonal to $c_1 - c_2$ and $c_2 - c_3$ is a $(d-2)$ -flat, for which any point z satisfies to the two orthogonality conditions $(z - t_{12})'(c_1 - c_2) = 0$ and $(z - t_{23})'(c_2 - c_3) = 0$, where t_{12} and t_{23} are known from equation (2.6). By expanding these two equations and by adding them, and by expressing t_{13} from equation (2.6), we get $(z - t_{13})'(c_1 - c_3) = 0$. That proves Lemma 2.1.

Lemma 2.1. *Assuming that c_1, c_2, c_3 are not aligned, the common intersection of the three radical hyperplanes defined by c_1, c_2, c_3 exists and is a unique $(d-2)$ -flat.*

When the three radii are equal, we retrieve for $d = 2$ that the three perpendicular bisectors of the sides of a triangle intersect at a common single point.

More generally, we look for the existence of the points z at the intersection of the $n(n-1)/2$ radical planes defined by $c_i - c_j$, $1 \leq i < j \leq n$. Applying repeatedly Lemma 2.1 to all triplets of centers, it appears that the set of the points z is the intersection of the $n-1$ radical planes defined by $c_1 - c_i$, $i = 2, \dots, n$. If existing, this intersection is a unique $(d+1-n)$ -flat. This flat is orthogonal to the $(n-1)$ -flat containing c_1, c_2, \dots, c_n , thus it can be located from the projection z_0 of any of its points z on the $(n-1)$ -flat containing c_1, c_2, \dots, c_n . Let γ_j , $j = 1, \dots, n$, be the barycentric coordinates of z_0 related to c_1, \dots, c_n . We get the following linear system of order n :

$$\begin{pmatrix} 1 & \dots & 1 \\ c'_1(c_1 - c_2) & \dots & c'_n(c_1 - c_2) \\ \vdots & \vdots & \vdots \\ c'_1(c_1 - c_n) & \dots & c'_n(c_1 - c_n) \end{pmatrix} \cdot \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 \\ c'_1 c_1 - c'_2 c_2 + R_1^2 - R_2^2 \\ \vdots \\ c'_1 c_1 - c'_n c_n + R_1^2 - R_n^2 \end{pmatrix}. \quad (2.14)$$

The system (2.14) receives a unique solution when the determinant Δ of the matrix above is not zero. The value of Δ is obtained by subtracting columns 2, ..., n from column 1 and by developing from the first column:

$$\Delta = (-1)^n \cdot \det \begin{pmatrix} (c_2 - c_1)'(c_2 - c_1) & \dots & (c_n - c_1)'(c_2 - c_1) \\ \vdots & \vdots & \vdots \\ (c_2 - c_1)'(c_n - c_1) & \dots & (c_n - c_1)'(c_n - c_1) \end{pmatrix}.$$

Denoting by $\Delta_{1,2,\dots,n}$ the determinant of the simplex $c_2 - c_1, \dots, c_n - c_1$, we get:

$$\Delta = (-1)^n \Delta_{1,2,\dots,n}^2. \quad (2.15)$$

It means that z_0 and the intersection we are looking for exists if and only if the simplex c_1, \dots, c_n is not degenerated in the subspace of dimension $n-1$ defined by

c_1, \dots, c_n (in which case there are neither two concentric spheres nor three aligned centers). The conclusion is still valid if we consider the $n - 1$ radical planes defined by $c_k - c_i, i = 1, \dots, n, i \neq k, 2 \leq k \leq n$, and we get Theorem 2.3.

Theorem 2.3. *Let n spheres in $E^d, 2 \leq n \leq d + 1$, their n centers c_1, \dots, c_n being the vertices of a non degenerated simplex in the $(n - 1)$ -flat defined by these n centers. The intersection of the resulting $n(n - 1)/2$ radical planes is a $(d + 1 - n)$ -flat orthogonal to the $(n - 1)$ -flat containing the n centers and its orthogonal projection z_0 on this $(n - 1)$ -flat satisfies to equations (2.14) and (2.15).*

The result given above was reported in [37]. When $n = d + 1$, the $d(d + 1)/2$ radical planes all intersect at a common single point z_0 . When all radii are equal, we retrieve for $d = 3$ that the six planes mediators of the edges of a tetrahedron intersect at a common single point.

The case of $n > d + 1$ spheres has interesting connections with Voronoi diagrams [10, 11], but Voronoi diagrams are out of scope of this chapter, because they do not lead to practical surfaces and volumes computations.

2.2.4 Intersections of order 3

We set $n = 3$. When two spheres are tangent, either one is included in the other one or their intersection reduces to one point. For surfaces and volumes calculations, this latter case can be viewed as if the intersection is empty. Thus, in both cases we can neglect the existence of tangent spheres. The enumeration of the possible topological configurations is done in the plane containing the three centers: we have to enumerate the configurations encountered from the intersections of the three great circles of the spheres.

When the intersection of two spheres is a lens, their great circles intersect at two contact points. Otherwise there is no contact point. For three spheres, there are either 0, or 2, or 4, or 6 contact points (we neglect the cases of contact points with multiplicities greater than 1). We enumerated in Figures 2.2–2.5 the 14 possible configurations for 3 spheres.

Not all configurations are relevant in chemistry and for trilateration (see Section 2.1), but all must be considered by the programmer willing to build a software computing V and S in the general case. Exception made for the configuration in Figure 2.5d, V_{123} and S_{123} are trivial to calculate, and so are V and S from equations (2.16) and (2.17), obtained by iterating equations (2.12) and (2.13):

$$V = V_1 + V_2 + V_3 - V_{12} - V_{13} - V_{23} + V_{123}, \quad (2.16)$$

$$S = S_1 + S_2 + S_3 - S_{12} - S_{13} - S_{23} + S_{123}. \quad (2.17)$$

Now we set $d = 3$ and we consider the case of Figure 2.5d. The convex domain at the intersection of the three spheres is symmetric through the plane containing c_1, c_2, c_3 , and the three arcs on the boundary of this domain intersect at two points z_+

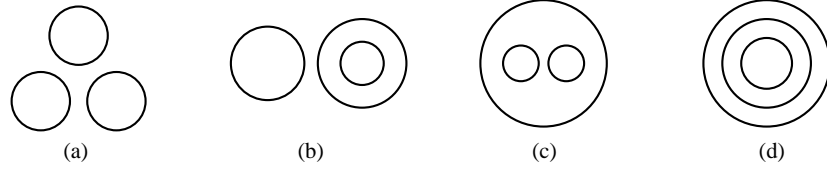


Fig. 2.2 The four configurations in the case there is no contact point

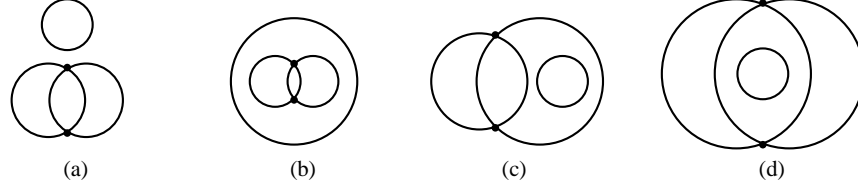


Fig. 2.3 The four configurations in the case there are 2 contact points

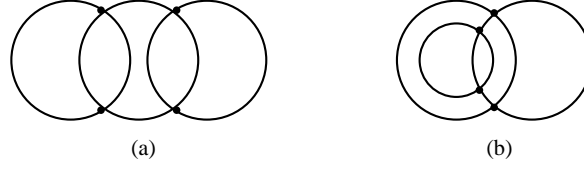


Fig. 2.4 The two configurations in the case there are 4 contact points

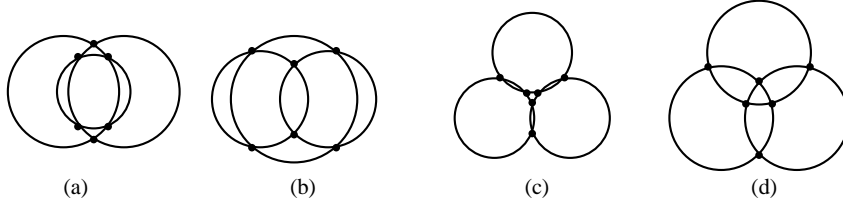


Fig. 2.5 The four configurations in the case there are 6 contact points

and z_- which both lie at the intersection of the three radical planes. The barycentric coordinates γ_i , $i \in \{1, 2, 3\}$, of $z_0 = (z_+ + z_-)/2$ are calculated in equation (2.18) from equations (2.14) and (2.15), where T_{123} is the surface of the triangle c_1, c_2, c_3 and i, j, k are circular permutations of 1, 2, 3. We get also the length of the segment $z^+ - z^-$ in equation (2.19), and then we deduce Theorem 2.4:

$$\begin{aligned}
16T_{123}^2 \gamma_i = & -2R_i^2 \|c_j - c_k\|^2 \\
& + R_j^2 (\|c_j - c_k\|^2 + \|c_k - c_i\|^2 - \|c_i - c_j\|^2) \\
& + R_k^2 (\|c_j - c_k\|^2 - \|c_k - c_i\|^2 + \|c_i - c_j\|^2) \\
& + \|c_j - c_k\|^2 (-\|c_j - c_k\|^2 + \|c_k - c_i\|^2 + \|c_i - c_j\|^2), \quad (2.18)
\end{aligned}$$

$$\begin{aligned}
\left\| \frac{z_+ - z_-}{2} \right\|^2 = & \frac{R_1^2 + R_2^2 + R_3^2}{3} - \left\| z_0 - \frac{c_1 + c_2 + c_3}{3} \right\|^2 \\
& - \frac{\|c_2 - c_1\|^2 + \|c_3 - c_2\|^2 + \|c_1 - c_3\|^2}{9}. \quad (2.19)
\end{aligned}$$

Theorem 2.4. *The convex domain defined by the configuration of Figure 2.5d exists if and only if the quantity at the right member of equation (2.19) is positive.*

In the case of Figure 2.5d, the point $z_0 = (z_+ + z_-)/2$ is interior to the intersection of the three spheres because this intersection is convex. Then, z_0, z_+, z_- and the three contact points on the boundary of the sphere intersection in the plane c_1, c_2, c_3 define a partition of the intersection into six parts, each one being the intersection between one sphere and one trihedron. Each of the six trihedra originates in z_0 and is bounded by two radical planes and by the plane c_1, c_2, c_3 . We denote by $x_{ij}^{(k)}$ each of the contact points at the intersection of the great circles of spheres i and j located in the interior of the sphere k , where i, j, k are circular permutations of 1, 2, 3. Thus the trihedra are defined by the half lines sets:

$$\begin{aligned}
& \{z_0 \rightarrow x_{31}^{(2)}, z_0 \rightarrow x_{12}^{(3)}, z_0 \rightarrow z_+\} \\
& \{z_0 \rightarrow x_{12}^{(3)}, z_0 \rightarrow x_{23}^{(1)}, z_0 \rightarrow z_+\} \\
& \{z_0 \rightarrow x_{23}^{(1)}, z_0 \rightarrow x_{31}^{(2)}, z_0 \rightarrow z_+\},
\end{aligned}$$

and their mirror symmetric images through the plane c_1, c_2, c_3 :

$$\begin{aligned}
& \{z_0 \rightarrow x_{31}^{(2)}, z_0 \rightarrow x_{12}^{(3)}, z_0 \rightarrow z_-\} \\
& \{z_0 \rightarrow x_{12}^{(3)}, z_0 \rightarrow x_{23}^{(1)}, z_0 \rightarrow z_-\} \\
& \{z_0 \rightarrow x_{23}^{(1)}, z_0 \rightarrow x_{31}^{(2)}, z_0 \rightarrow z_-\}.
\end{aligned}$$

The six above trihedra intersect respectively with the sphere 1, 2, 3, 1, 2, 3.

It follows that the calculation of V_{123} and S_{123} can be done through the calculation of the portion of the area of a sphere defined by its intersection with a trihedron having its origin in the interior of the sphere. This latter calculation will be shown to be required in the case of four intersecting spheres and is done analytically in Section 2.2.6 using the Gauss-Bonnet theorem.

Several analytical treatments of up to three intersecting spheres appeared in the literature. Some do not involve the Gauss-Bonnet theorem [17, 18, 49], but most involve it [2, 7, 8, 24, 36, 47]. These latter are mainly based on [7] for the surfaces and on [8] for the volumes. As shown in Section 2.2.7, neglecting the intersections

of order 4 and higher is numerically not acceptable for van der Waals surfaces and volumes calculation.

2.2.5 Intersections of order 4

We set $n = 4$ and we iterate again equations (2.12) and (2.13):

$$\begin{aligned} V = & V_1 + V_2 + V_3 + V_4 - V_{12} - V_{13} - V_{14} - V_{23} - V_{24} - V_{34} \\ & + V_{123} + V_{124} + V_{134} + V_{234} - V_{1234}, \end{aligned} \quad (2.20)$$

$$\begin{aligned} S = & S_1 + S_2 + S_3 + S_4 - S_{12} - S_{13} - S_{14} - S_{23} - S_{24} - S_{34} \\ & + S_{123} + S_{124} + S_{134} + S_{234} - S_{1234}. \end{aligned} \quad (2.21)$$

We set $d = 3$. For clarity, we do not enumerate all topological configurations generated by four intersecting spheres. If any of the four triplets of spheres is not in the configuration of Figure 2.5d, we know how to calculate V_{1234} and S_{1234} .

Assuming that the four sphere triplets are in the configuration of Figure 2.5d, we can classify the topological configurations via locating the four pairs of two contact points $z_+^{(ijk)}$ and $z_-^{(ijk)}$ at the intersections of the boundaries of the three spheres i, j, k , $1 \leq i < j < k \leq 4$. Each of the two contact points defined by a triplet of spheres can be inside or outside the fourth sphere (we still neglect tangencies and multiplicities). We consider the possible locations of the six remaining contact points. When all of them are inside a sphere, one sphere is included in the union of the three other ones. When all the six remaining contact points are outside a sphere, either the intersection of three spheres is included in the remaining one, or the 4-order intersection is empty. This case of empty intersection was called an Empty Simplicial Topology (EST) in [37]. If it happens for some triplet i, j, k that $z_+^{(ijk)}$ is inside the fourth sphere and $z_-^{(ijk)}$ is outside, or conversely, then that happens for all the other triplets: there are three contact points inside a sphere and three outside (this is the general case of intersection of four spheres). The two remaining cases correspond in fact to the same configuration: the intersection of two spheres is included in the union of the two other ones.

To summarize, when the four sphere triplets are in the configuration of Figure 2.5d, the possible configurations for a non-empty 4-order intersection are:

1. One sphere is included in the union of the three other ones.
2. The intersection of two spheres is included in the union of the two other ones.
3. The intersection of three spheres is included in the remaining one.
4. None of the above ones: general case.

Calculating the 4-order intersection reduces to calculate the 3-order intersections in the cases of configurations 1,2,3. The total number of pairs of interior contact

points for these latter configurations are respectively 3,2,1, and the total number of pairs of exterior contact points are respectively 1,2,3. Only the EST configuration has 4 exterior contact point pairs.

In the general case, the domain of the 4-order intersection is bounded by four spherical triangles separated by six arcs of circle and intersecting at four contact points: $z^{(123)}$, $z^{(124)}$, $z^{(134)}$, $z^{(234)}$. This convex domain is topologically similar to a tetrahedron. From Theorem 2.3, there is a unique point z_0 at the intersection of the 6 radical planes. Its barycentric coordinates are computed from equations (2.14) and (2.15). The point z_0 is interior to the convex domain of the 4-order intersection. This latter is partitioned into four parts, each one being the intersection of one sphere and one trihedron originating in z_0 .

These trihedra, which intersect respectively the spheres 1, 2, 3, 4, are:

$$\begin{aligned} &\{z_0 \rightarrow x^{(123)}, z_0 \rightarrow x^{(124)}, z_0 \rightarrow z^{(134)}\}, \\ &\{z_0 \rightarrow x^{(123)}, z_0 \rightarrow x^{(124)}, z_0 \rightarrow z^{(234)}\}, \\ &\{z_0 \rightarrow x^{(123)}, z_0 \rightarrow x^{(134)}, z_0 \rightarrow z^{(234)}\}, \\ &\{z_0 \rightarrow x^{(124)}, z_0 \rightarrow x^{(134)}, z_0 \rightarrow z^{(234)}\}. \end{aligned}$$

It follows that the calculation of V_{1234} and S_{1234} can be done through the calculation of the portion of the area of a sphere defined by its intersection with a trihedron having its origin in the interior of the sphere. This latter calculation is done analytically in Section 2.2.6 using the Gauss-Bonnet theorem. An analytical treatment of the four spheres intersection without invoking the Gauss-Bonnet theorem was reported [27].

2.2.6 Intersection of a sphere with a dihedron or with trihedron

We consider a sphere of center c and radius R intersecting a salient trihedron of origin z inside the sphere. The trihedron intersects the boundary of the sphere at x_1, x_2, x_3 , these three points being ordered in the direct sense when referred to z as origin. We look for the calculation of the volume V_{st} of the intersection and for the surface S_{st} of the spherical triangle on the boundary of the sphere. This spherical triangle is bounded by the oriented arcs of circle (x_1, x_2) , (x_2, x_3) , (x_3, x_1) . These arcs have respective radii h_{12} , h_{23} , h_{31} and respective angles b_{12} , b_{23} , b_{31} , each of these angles being in $[0; \pi]$. The trihedron intersects each of the planes respectively tangent to the sphere at x_1, x_2, x_3 , thus defining the respective angles a_1, a_2, a_3 , each of these angles being in $[0; \pi]$. We define c_{12}, c_{23}, c_{31} as the projections of c on the planes containing the respective triplets (x_1, z, x_2) , (x_2, z, x_3) , (x_3, z, x_1) . We also define the angles associated to the three lens: $\theta_{ij} = \widehat{x_i, c, c_{ij}} = \widehat{x_j, c, c_{ij}}$, $(i, j) = (1, 2)$ or $(i, j) = (2, 3)$ or $(i, j) = (3, 1)$, such that $\cos \theta_{ij}$ is positive when the vector $c_{ij} - c$ of origin c has the same sense than the direct normal to the plane of the oriented arc (x_i, x_j) , and $\cos \theta_{ij}$ is negative when $c_{ij} - c$ has a sense opposite to this normal. We have also $\sin \theta_{ij} = h_{ij}/R$. The sphere has a constant Gaussian curvature and each arc of circle has a constant geodesic curvature. Thus, the Gauss-Bonnet theorem

[26, 52] leads to a simple expression of S_{st} in equation (2.22), in which the first term in parenthesis is the spherical excess:

$$\frac{S_{st}}{R^2} = (a_1 + a_2 + a_3 - \pi) - (b_{12} \cos \theta_{12} + b_{23} \cos \theta_{23} + b_{31} \cos \theta_{31}). \quad (2.22)$$

V_{zx} being the signed volume of the oriented tetrahedron $(x_1 - z, x_2 - z, x_3 - z)$ and V_{cx} being the signed volume of the oriented tetrahedron $(x_1 - c, x_2 - c, x_3 - c)$, the volume V_{st} is given below [37]:

$$V_{st} = \frac{R^3}{6} [\cos \theta_{12} \sin^2 \theta_{12} (\sin b_{12} - b_{12}) + \cos \theta_{23} \sin^2 \theta_{23} (\sin b_{23} - b_{23}) + \cos \theta_{31} \sin^2 \theta_{31} (\sin b_{31} - b_{31})] + V_{zx} - V_{cx} + \frac{S_{st}R}{3}. \quad (2.23)$$

When z lies at c , we retrieve the expression given by Girard's theorem [48]: $\frac{S_{st}}{R^2}$ is the spherical excess and $V_{st} = \frac{S_{st}R}{3}$.

Remark: rather than looking for the validity of equation (2.23) from integrating S_{st} as stated in [37], we give a hint of proof as follows. We define the point x_0 at the intersection of the half line of origin c passing through z with the boundary of the sphere, and we add the three signed contributions to V_{st} due to the trihedra $\{z \rightarrow x_1, z \rightarrow x_2, z \rightarrow x_0\}$ and $\{z \rightarrow x_2, z \rightarrow x_3, z \rightarrow x_0\}$ and $\{z \rightarrow x_3, z \rightarrow x_1, z \rightarrow x_0\}$, a contribution being negative when the bounding arcs are not oriented in the conventional sense. For each trihedra, say $\{z \rightarrow x_1, z \rightarrow x_2, z \rightarrow x_0\}$, the volume of the intersection with the sphere is $V_{sc012} - V_{cz12} - V_{c12}$, in which V_{sc012} is the volume of the domain bounded by $c - x_0$, $c - x_1$, $c - x_2$ and by the spherical triangle (x_0, x_1, x_2) intercepted by the trihedron $\{z \rightarrow x_1, z \rightarrow x_2, z \rightarrow x_0\}$ and of area computable by equation (2.22). Despite that this domain is not bounded by a trihedron, V_{sc012} is computable by trivial integration in respect to the radius of the sphere. V_{cz12} is the volume of the tetrahedron (c, z, x_1, x_2) . V_{c12} is the volume of the portion of the cone of origin c with height $R \cos \theta_{12}$ and of basis the circular segment of area $(b_{12} - \sin b_{12})(R \sin \theta_{12})^2/2$, i.e. $V_{c12} = \cos \theta_{12} \sin^2 \theta_{12} (b_{12} - \sin b_{12})R^3/6$. The rest of the calculation is easy.

The case of the six trihedra listed at the end of Section 2.2.4 is of interest. It is such that one of the arcs, say, (x_2, x_3) , lies on a great circle of the sphere: $z, x_2, x_3, c, c_{12}, c_{31}$, are coplanar and c_{23} coincides with c , then $a_2 = \pi/2$, $a_3 = \pi/2$, and $\theta_{23} = \pi/2$.

Grouping each trihedra with its symmetric image through the plane of the great circle let us to the expressions of the volume V_{sd} of the intersection of a sphere and a dihedron and for the surface S_{sd} of the associated portion of spherical area. Keeping the notations used in equations (2.22) and (2.23):

$$\frac{S_{sd}}{R^2} = 2a_1 - 2(b_{12} \cos \theta_{12} + b_{31} \cos \theta_{31}), \quad (2.24)$$

$$V_{sd} = \frac{R^3}{3} [\cos \theta_{12} \sin^2 \theta_{12} (\sin b_{12} - b_{12}) + \cos \theta_{31} \sin^2 \theta_{31} (\sin b_{31} - b_{31})] + 2(V_{zx} - V_{cx}) + \frac{S_{sd}R}{3}. \quad (2.25)$$

Care: in equations (2.24) and (2.25), b_{12} and b_{31} are the angles of the half arcs (not the full arcs) bounding the intersection of the sphere with the dihedron. It is also recalled that the arcs must be correctly oriented and that $\cos \theta_{12}$ and $\cos \theta_{31}$ can be negative.

When z is outside the sphere, the intersection with a dihedron reduces to simpler cases such as a sphere minus two lens, and the intersection with a trihedron can be reduced to intersections with appropriate dihedra or trihedra for which equations (2.22)–(2.25) can be used. More generally we can extend volumes and surfaces computations to the intersections of a sphere with a tetrahedron, a convex polyhedron, and with any union of convex polyhedra. These computations are based on equations (2.22)–(2.25) but are not needed in the framework of this chapter.

2.2.7 Intersections of order 5 and higher

By iterating equations (2.12) and (2.13) until order n , the volume and the surface of the union of the n spheres can be expressed from the inclusion-exclusion principle as follows:

$$V = \sum_{1 \leq i \leq n} V_i - \sum_{1 \leq i_1 < i_2 \leq n} V_{i_1 i_2} + \sum_{1 \leq i_1 < i_2 < i_3 \leq n} V_{i_1 i_2 i_3} + \cdots + (-1)^{n-1} V_{i_1 i_2 \dots i_n}, \quad (2.26)$$

$$S = \sum_{1 \leq i \leq n} S_i - \sum_{1 \leq i_1 < i_2 \leq n} S_{i_1 i_2} + \sum_{1 \leq i_1 < i_2 < i_3 \leq n} S_{i_1 i_2 i_3} + \cdots + (-1)^{n-1} S_{i_1 i_2 \dots i_n}. \quad (2.27)$$

Setting $d = 3$ and $n \geq 5$, the existence of intersections of order n is deduced from Helly's theorem [14]. It means that the intersection of 5 spheres is not empty if and only if each of the 5 subsets of 4 spheres gives rise to a non-empty intersection. If $n \geq 6$, the intersection of 6 spheres is not empty if and only if each of the 6 subsets of 5 spheres gives rise to a non-empty intersection. Etc.

For practical computations, we need Theorem 2.5:

Theorem 2.5. *When $n \geq 5$ spheres in E^3 have a common non-empty intersection, there are m (with $1 \leq m \leq 3$) of these spheres such that their union J contains the intersection I of the $n - m$ remaining spheres.*

Proof. Assume initially that $n = 5$ spheres have a common non-empty intersection. From Helly’s theorem, each of the 5 subsets of 4 spheres gives rise to a non-empty intersection. If one of these subsets of 4 spheres is such that the union of $m = 1, 2$, or 3 spheres contains the intersection of the $4 - m$ remaining spheres, the union of these m spheres contains the intersection of the $5 - m$ remaining spheres and the theorem stands. As stated in Section 2.2.5, a set of 4 spheres giving rise to a non-empty intersection which is not in one of these latter configurations is in the general configuration.

Now we need to prove the theorem when the 5 subsets of 4 spheres are in the case of a general configuration. For this latter, the four contact points pairs between three spheres are such that one of these contact points lies inside the fourth sphere and the other contact point lies outside this fourth sphere. Thus, enumerating the spatial arrangements of the five spheres leads to only two possible configurations:

1. One sphere contains the intersection of the four other ones, and simultaneously the union of these four spheres contains the first one. The common 5-order intersection is reduced to a 4-order intersection of the four spheres in the general configuration, which is topologically related to a tetrahedron.
2. The union of two spheres contain the intersection of the three other ones, and simultaneously the union of these three spheres contains the intersection of the two first ones. The common 5-order intersection is a spherical polyhedron topologically related to a triangular prism, i.e. bounded by two triangles and three tetragons, nine arcs and six vertices.

Obviously the theorem stands for both configurations and thus it stands always for $n = 5$. Then it stands for $n > 5$ because the intersection of the $n - m$ spheres is itself included in the intersection of the $5 - m$ spheres.

Theorem 2.5 is known as the *three spheres theorem* [37]. A more general expression of this theorem is given in the appendix.

Then, applying the inclusion-exclusion principle to both members of the equality $I \cup J = J$ provides a relation between the n -order intersection and the $(n - 1)$ -order intersections. Starting from 4-order intersections surfaces and volumes, we can compute 5-order and then higher order intersections surfaces and volumes.

Some authors attempted to fully describe spheres intersections [9, 16, 20, 37] or to produce softwares dealing with more than $n = 3$ intersecting spheres [21, 40]. Since many authors did an analytical treatment with the assumption that no 4-order intersections exist (see Section 2.2.4), we analyzed a set of 70 molecules offering a wide structural diversity [28] with the ASV freeware [37, 40]. The number of atoms ranged from 16 to 186 (median: 45). The atomic radii (in Å) were taken from [15]: H=1.17, C=1.75, N=1.55, O=1.40, F=1.30, I=2.10, and from [23]: P=1.75, S=2.55.

The results for volumes are in Table 2.1 and those for surfaces are in Table 2.2.

All molecules gave rise to 5-order intersections, 69 to 6-order intersections, 13 to 7-order intersections and 2 to 8-order intersections. Owing to the results in Tables 2.1 and 2.2, truncating the calculation after the 3-order intersections is unacceptable, both for volumes and surfaces computations: the values are strongly overestimated. Correcting the overestimated volumes via the use of smaller radii is

Table 2.1 Distribution of the volume and of the errors on the volume.

	Mean	Std. dev.	Min	Max	Median
Volume: exact value V in \AA^3	336.181	165.917	115.518	1162.450	386.702
Error $\hat{V}^{(3)} - V$, from eq. (2.16)	35.022	15.329	6.389	88.783	33.581
Error $\hat{V}^{(4)} - V$, from eq. (2.20)	-6.173	2.745	-12.768	-0.498	-5.180
$100 \cdot (\hat{V}^{(3)} - V)/V$	10.655	2.973	3.154	19.107	8.889
$100 \cdot (\hat{V}^{(4)} - V)/V$	-1.987	0.838	-5.283	-0.161	-1.427

Table 2.2 Distribution of the surface and of the errors on the surface.

	Mean	Std. dev.	Min	Max	Median
Surface: exact value S in \AA^2	388.162	185.526	147.386	1324.635	461.750
Error $\hat{S}^{(3)} - S$, from eq. (2.17)	337.120	170.906	71.409	869.994	362.754
Error $\hat{S}^{(4)} - S$, from eq. (2.21)	-59.374	27.288	-130.446	-6.716	-53.672
$100 \cdot (\hat{S}^{(3)} - S)/S$	86.511	25.691	35.609	139.625	79.243
$100 \cdot (\hat{S}^{(4)} - S)/S$	-16.176	7.064	-37.001	-3.407	-12.130

of course inappropriate, and lowering the radii even does not guarantee a decrease of the surfaces values. Truncating the calculation after the 4-order intersections give underestimated values which are still not acceptable for surfaces, although it gives a rough approximation of the volumes which could be tolerated in some contexts.

Another set was analyzed [37], containing 63 molecules: even though the hydrogens were discarded, similar conclusions were derived. It is amazing to see the success encountered during the last three decades by the analytical algorithms neglecting the existence of intersections between more than 3 spheres (see references cited at the end of Section 2.2.4). Unless working with sufficiently small atomic radii, the resulting software tools should return strongly erroneous results, including those which compute derivatives [24, 36, 47]

2.3 Numerical methods

The full analytical calculation of spheres unions volumes and surfaces can be used to compute numerical approximations of derivatives, which in turn are useful in some optimization problems [38]. However, for many QSAR (Quantitative Structure Activity Relationships) applications involving the molecular surface or volume as an input variable for regression, an approximate value based on a numerical calculus may be acceptable. Several numerical methods were developed [3, 4, 25, 32, 34, 35, 54]. The brute approach is based on a regular grid, i.e. a mesh of N nodes defining

cubes to be counted in order to estimate the volume of the surface. Despite that more or less sophisticated possible variants of this approach are possible, these are $O(N^3)$ processing time algorithms. Thus, Monte-Carlo methods should be preferred (see further).

2.4 Monte-Carlo methods

These methods are highly attractive due to their great simplicity. In general, they are used in awkward situations where analytical methods are unavailable and where numerical methods are inefficient, e.g. due to a multiple integral over a domain with an untractable boundary calculation.

As shown below, Monte-Carlo methods are $O(N^2)$ processing time algorithms, although regular grid methods are generally $O(N^d)$ ones. Thus, for $d > 2$, Monte-Carlo methods should be preferred. Furthermore, they are fully adequate for QSAR applications and easy to programme.

2.4.1 Monte-Carlo calculations of volumes

The Monte-Carlo measure of the volume V of a finite domain \mathcal{D} of E^d is performed as follows:

1. Enclose the domain in a d -dimensional parallelepipedic window of volume W .
2. Build a function returning the status of one point: inside \mathcal{D} or not.
3. Generate a sample of N independent random points in E^d , each one following the uniform law in the window.
4. Count the number $N_{\mathcal{D}}$ of these random points which fell inside \mathcal{D} .
5. The Monte-Carlo estimate of V is $\hat{V} = W \cdot N_{\mathcal{D}}/N$.

We consider the N random variables $X_i = \mathbf{1}_{\{U_i \in \mathcal{D}\}}$, $i = 1, \dots, N$, where U_i is the i^{th} random vector following the uniform law in the window and $\mathbf{1}_{\{U_i \in \mathcal{D}\}}$ is the indicator function of the event $U_i \in \mathcal{D}$. In other words, $X_i = 1$ when U_i takes a value in \mathcal{D} and $X_i = 0$ elsewhere. Assuming that U_1, \dots, U_N are independent, then X_1, \dots, X_N are independent and identically distributed (i.i.d.). The probability for U_i to take a value in \mathcal{D} is $p = V/W$. Setting $q = 1 - p$, $\text{Prob}(X_i = 1) = p$ and $\text{Prob}(X_i = 0) = q$, i.e. each of the X_i follow the Bernouilli law of parameter p . The sum of these N i.i.d. random variables X_i follow a binomial law $B(N, p)$, of expectation Np and variance

Npq . Their mean is the random variable $\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$, with expectation $p = \frac{V}{W}$ and

variance $\frac{pq}{N} = \frac{1}{N} \left(\frac{V}{W} \right) \left(\frac{W-V}{W} \right)$. It follows that the observed mean $\frac{\hat{V}}{W} = \frac{N_{\mathcal{D}}}{N}$

is a consistent and unbiased estimator of \bar{X} . Thus, \hat{V} is a consistent and unbiased estimator of V , of expectation V and variance $V(W - V)/N$. Furthermore, we know

Table 2.3 Monte-Carlo measures of the van der Waals volume of the cyclosporine. Analytical value: $V = 1162.450152$.

N	$N_{\mathcal{D}}$	W	\hat{V}	$\hat{\sigma}_V$	$100 \cdot \hat{\sigma}_V / V$
100	23	5230.333	1202.977	220.109	18.297
10000	2220	5230.333	1161.134	21.737	1.872
1000000	222642	5230.333	1164.492	2.176	0.187

from de Moivre-Laplace theorem [19, 45] that the law of $\frac{N\bar{X} - Np}{\sqrt{Npq}}$ converges to the normal law $\mathcal{N}(0,1)$ of expectation 0 and variance 1.

We notice that the best possible window is the one minimizing $V(W - V)$, so that it should be the smallest possible one containing \mathcal{D} . The theory still works with a non parallelepipedic window, but this latter needs an increase of the computational cost to generate the U_i following the uniform law in the window.

The precision is evaluated through a confidence interval. Several binomial proportion confidence intervals were proposed [1, 6, 44]. Among them, we selected the so-called Wald interval, which is symmetric and based under the normal approximation: $[\hat{p} \pm \xi_{(1-\alpha/2)} \sqrt{\frac{\hat{p}(1-\hat{p})}{N}}]$, where $\hat{p} = \hat{V}/W$ and $\xi_{(1-\alpha/2)}$ is the $(1 - \alpha/2)$ percentile of the normal law $\mathcal{N}(0,1)$ corresponding to the error α (e.g., for a 95% confidence level, $\alpha = 0.05$, $1 - \alpha/2 = 0.975$ and $\xi_{(1-\alpha/2)} \approx 1.96$). Thus, having estimated the center of the confidence interval for \hat{V} , its length is estimated from the observed standard deviation $\hat{\sigma}_V = W \hat{\sigma}_p$, with $\hat{\sigma}_p = \sqrt{\frac{\hat{p}(1-\hat{p})}{N}}$. The Wald confidence interval is appropriate as long as neither Np nor Nq is too small, which is the case in our context.

Examples of Monte-Carlo computations of a molecular volume are given in Table 2.3. The molecule is the cyclosporine. It is the largest of the dataset used at the end of Section 2.2.7 and it contains 186 atomic spheres.

Clearly, multiplying by 100 the number of observations led to an increase of the precision by a factor 10. This is in agreement with the proportionality of $\hat{\sigma}_V$ to $1/\sqrt{N}$.

2.4.2 Monte-Carlo calculations of surfaces

The Monte-Carlo measure of the surface S of an union of n spheres in E^d is similar to the one for volumes, except that we consider the ratio of the surface of the union of the spheres to the union of their surfaces, i.e. the sum $T = S_1 + \dots + S_n$ of the individual surfaces of the n spheres plays the role of W for the volumes, and the domain \mathcal{D} is now defined by the surface of the union of the n spheres.

The Monte-Carlo measure of S is performed as follows:

1. Compute the total surface T .
2. Build a function returning the status of one point: inside a sphere or not.
3. Generate a sample of N independent random points, each one following the uniform law over the union of the n surfaces. It is done as follows:
 - Select one of the n spheres such that each sphere i has a probability S_i/T to be selected, i.e. get a random number v following the uniform law over $[0; T]$ and retain the sphere index i as the smallest one such that $v < S_1 + \dots + S_i$.
 - Generate a random point of E^d following the uniform distribution on the surface of the sphere i . It will fall in \mathcal{D} if it is not interior to any of the $n - 1$ other spheres.
4. Count the number $N_{\mathcal{D}}$ of these random points which fell in \mathcal{D} .
5. The Monte-Carlo estimate of S is $\hat{S} = T \cdot N_{\mathcal{D}}/N$.

The analysis of the algorithm is identical to the one for volumes, with $\hat{\sigma}_S = T \hat{\sigma}_p$, $\hat{\sigma}_p = \sqrt{\frac{\hat{p}(1-\hat{p})}{N}}$ and $\hat{p} = \hat{S}/T$.

There are several methods to generate a random point following the uniform law on the boundary of a sphere. For clarity, we assume that the center of the sphere lies at the origin. The rejection method is the simplest one: generate a point following the uniform law in the smallest cube containing the sphere, and accept the point if it falls inside the sphere; if it falls outside, generate an other one, until it falls inside the sphere. Once done, normalize the corresponding vector to set its length equal to the radius of the sphere. The rejection method is rather inefficient for high d values because the ratio of the volume of the sphere to its smallest enclosing cube tends to zero when d increases to infinity (see Section 2.2.1). Nevertheless, it can be retained for $d = 3$. There are other methods, such as normalizing a vector following the isotropic multinormal law, i.e. such that its d components follow d uncorrelated normal laws of null expectation and of identical standard deviations. Observations from the normal law can themselves be generated from various methods, such as Box-Muller or Marsaglia [22].

Examples of Monte-Carlo computations of a molecular surface are given in Table 2.4. The molecule is the cyclosporine. It is the largest of the dataset used at the end of Section 2.2.7 and it contains 186 atomic spheres.

Again, multiplying by 100 the number of observations led to an increase of the precision by a factor 10, which is in agreement with the proportionality of $\hat{\sigma}_S$ to $1/\sqrt{N}$.

Table 2.4 Monte-Carlo measures of the van der Waals surface area of the cyclosporine. Analytical value: $S = 1324.635145$.

N	$N_{\mathcal{S}}$	T	\hat{S}	$\hat{\sigma}_S$	$100 \cdot \hat{\sigma}_S / S$
100	20	4751.111	950.222	190.044	20.000
10000	2715	4751.111	1289.927	21.130	1.638
1000000	277944	4751.111	1320.543	2.128	0.161

2.5 Discussion and conclusion

Selecting the analytical calculation vs. the Monte-Carlo calculation of the volume or surface of a union of spheres depends on two criteria:

1. A high precision is required, e.g. for computing derivatives via finite differences.
2. A small computing time is required, due to the need of numerous repeated calls.

It is clear from equations (2.26)–(2.27) that, when all spheres intersect, the computing time of the analytical calculation grows exponentially with the number of spheres. Practically, we measured computing times with the linux 64 bits Intel version of the freeware ASV [40], which can perform both analytical and Monte-carlo calculations. For the cyclosporine data mentioned in Tables 2.3 and 2.4 and containing 186 atoms, the analytical calculation took 0.23s, although the Monte-Carlo calculation took 36s. However, for the 5188 "ATOM" set of the human prostate antigen (PDB code 2ZCH; contains hydrogens) the analytical calculation took 1053s, although the Monte-Carlo calculation took only 131s. Thus, for small molecules or small sets of spheres, the analytical calculation is recommended. For large sets of spheres with a huge of intersections, the Monte-Carlo calculation is useful as long as a moderate accuracy suffices. Since it is the case for most molecular modeling applications, the f77 sources of the Monte-Carlo calculations in E^d , plus a f77 implementation of a pseudo-random generator [22] of period 2^{57} based on the congruential method used in [33], are provided with ASV.

Geometry programs are often subject to potential numerical instabilities. It is the case of the calculation of the intersection of the radical planes via Theorem 2.3, which assumes that the centers of the spheres are the vertices of a non degenerate simplex (see Section 2.2.3). Alas, chemistry data offer a huge of triplets of aligned atoms and of quadruplets of coplanar atoms. E.g. benzene derivatives offer 12 coplanar atoms and three sets of 4 aligned atoms due to the benzene ring, which is by far the ring the most frequently encountered by chemists [51]. In fact, any sp^2 carbon (e.g. a carbon connected with one double bond and two single bonds) gives rise to four coplanar atoms, a very common situation. Fortunately, in most cases calculating the intersection of the radical planes is not required. Furthermore, due to the presence of a limited number of significant digits in molecular files, most cases of alignment or coplanarity are avoided. In any case, to prevent instabilities, it is

possible to let the user perturbate the atomic coordinates with a given magnitude [40].

Appendix: segments and disks

We give here a stronger version of Theorem 2.5.

Theorem 2.6. *When $n \geq d + 2$ spheres in E^d , $d \leq 3$, have a common non-empty intersection, there are m (with $1 \leq m \leq \lfloor \frac{d+1}{2} \rfloor + 1$) of these spheres such that their union contains the intersection of the $n - m$ remaining spheres and simultaneously the union of these $n - m$ spheres contains the intersection of the first m ones.*

Proof. We assume initially that $n = d + 2$ spheres have a common non-empty intersection.

We set $d = 1$. A sphere in E^1 is a segment. The $n = 3$ segments have a common 3-order intersection. The theorem stands if a segment is included in an other one. If not, it is easy to check that there is only one possible configuration: one segment contains the intersection of the two other ones, and simultaneously the union of these two segments contains the first one.

We set $d = 2$. A sphere in E^2 is a disk. The $n = 4$ disks have a common 4-order intersection. The possible configurations for three intersecting disks are enumerated in Figures 2.2 and 2.5 (Section 2.2.4). If any of the four triplets of disks is not in the general case of intersection of Figure 2.5d, the theorem stands. We assume that the four triplets of disks are in this general case (Figure 2.5d). We consider the four bounding circles. There are two contact points at the intersection of each of the six pairs of circles. Enumerating the arrangements of these four circles can be done with the help of their contact points and with the two 2D-lens at the respective intersections of the disks 1,2 and 3,4. It leads to only two possible configurations:

1. One disk contains the intersection of the three other ones, and simultaneously the union of these three disks contains the first one. The common 4-order intersection is a curvilinear triangle (Figure 2.6a).
2. The union of two disks contains the intersection of the two other ones, and simultaneously the union of the two latter ones contains the intersection of the two former ones. The common 4-order intersection is a curvilinear tetragon (Figure 2.6b).

The theorem stands in both cases and thus it stands always for $n = 4$. For $d = 3$, the theorem was proved in Section 2.2.7.

For $n = d + 2$ and $d \leq 3$ the theorem stands and we found the $1 \leq m \leq \lfloor \frac{d+1}{2} \rfloor + 1$ required spheres. For $n > d + 2$ and $d \leq 3$, we consider $n - d - 2$ additional spheres. The theorem still stands because (a) the union of the m spheres contains the intersection of the $d + 2 - m$ spheres which in turn contains the intersection of the $(n - d - 2) + (d + 2 - m) = n - m$ spheres, and (b) the union of these $n - m$ spheres contains the union of the $d + 2 - m$ spheres which in turn contains the intersection of the m ones.

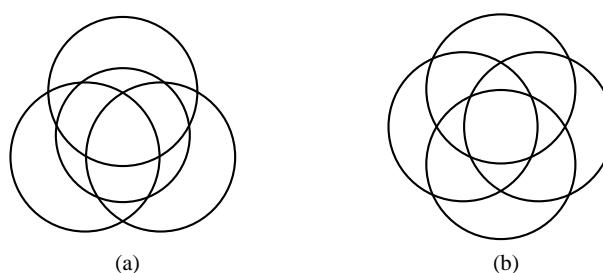


Fig. 2.6 The two configurations in the case of 4 intersecting disks

It is conjectured that Theorem 2.6 stands for $d > 3$. Although useful in the case $d = 3$ for spheres unions surfaces and volumes computations, this theorem can also be used in the case $d = 2$ for computing surfaces and exposed arcs lengths of disks unions.

References

1. A. Agresti, A.B. Coull, *Approximate is better than “exact” for interval estimation of binomial proportions*, The American Statistician **52**(2), 119–126, 1998.
2. P. Alard, S.J. Wodak, *Detection of cavities in a set of interpenetrating spheres*, Journal of Computational Chemistry **12**(8), 918–922, 1991.
3. A.A. Blyznyuk, J.E. Gready, *Numerical calculation of molecular surface area. I. Assessment of errors*, Journal of Computational Chemistry **17**(8), 962–969, 1996.
4. R.S. Bohacek, C. McMartin, *Definition and display of steric, hydrophobic, and hydrogen-bonding properties of ligand binding sites in proteins using Lee and Richards accessible surface: Validation of a high-resolution graphical tool for drug design*, Journal of Medicinal Chemistry **35**(10), 1671–1684, 1992.
5. A. Bondi, *Van der Waals volumes and radii*, Journal of Physical Chemistry **68**(3), 441–451, 1964.
6. L.D. Brown, T.T. Cai, A. DasGupta, *Interval estimation for a binomial proportion*, Statistical Science **16**(2), 101–133, 2001.
7. M.L. Connolly, *Analytical molecular surface calculation*, Journal of Applied Crystallography **16**(5), 548–558, 1983.
8. M.L. Connolly, *Computation of molecular volume*, Journal of the American Chemical Society **107**(5), 1118–1124, 1985.
9. L.R. Dodd, D.N. Theodorou, *Analytical treatment of the volume and surface area of molecules formed by an arbitrary collection of unequal spheres intersected by planes*, Molecular Physics **72**(6), 1313–1345, 1991.
10. H. Edelsbrunner, *Algorithms in Combinatorial Geometry*, EATCS Monographs on Theoretical Computer Science, volume 10, Springer-Verlag, Berlin, 1987.
11. H. Edelsbrunner, *The union of balls and its dual shape*, Discrete & Computational Geometry **13**(1), 415–440, 1995.
12. H. Edelsbrunner, M. Facello, J. Liang, *On the definition and the construction of pockets in macromolecules*, Discrete Applied Mathematics **88**(1–3), 83–102, 1998.

13. H. Edelsbrunner, E.P. Mücke, *Three-dimensional alpha shapes*, ACM Transactions on Graphics **13**(1), 43–72, 1994.
14. H.G. Eggleston, *Convexity*, Cambridge Tracts in Mathematics and Mathematical Physics **47**, chapter 2. F. Smithies, J.A. Todd, (Eds.), Cambridge University Press, Cambridge, 1966.
15. A. Gavezzotti, *The calculation of molecular volumes and the use of volume analysis in the investigation of structured media and of solid-state organic reactivity*, Journal of the American Chemical Society **105**(16), 5220–5225, 1983.
16. K.D. Gibson, H.A. Scheraga, *Exact calculation of the volume and surface area of fused hard-spheres molecules with unequal atomic radii*, Molecular Physics **62**(5), 1247–1265, 1987.
17. K.D. Gibson, H.A. Scheraga, *Volume of the intersection of three spheres of unequal size. A simplified analytical formula*, Journal of Physical Chemistry **91**(15), 4121–4122, 1987. Additions and corrections: **91**(24), 6326.
18. K.D. Gibson, H.A. Scheraga, *Surface area of the intersection of three spheres with unequal radii. A simplified analytical formula*, Molecular Physics **64**(4), 641–644, 1988.
19. Girardin, V., Limnios, L.: *Probabilités en vue des applications: cours et exercices corrigés*, Volume 1, chap 4.3. Vuibert, Paris (2008)
20. V. Gogonea, E. Osawa, *An improved algorithm for the analytical computation of solvent-excluded volume. The treatment of singularities in solvent-accessible surface area and volume functions*, Journal of Computational Chemistry **16**(7), 817–842, 1995.
21. G. Guerrero-Ruiz, A. Ocadiz-Ramirez, R. Garduño-Juárez, *ESFERA: a program for exact calculation of the volume and surface area of fused hard-sphere molecules with unequal atomic radii*, Journal of Computational Chemistry **15**(4), 351–352, 1991.
22. J.M. Hammersley, D.C. Handscomb, *Monte Carlo Methods*, In: “Monographs on Statistics and Applied Probability”, Chapman & Hall, London, 1983.
23. A.J. Hopfinger, *Conformational properties of macromolecules*, Molecular Biology Series **17**, chap. 2, sect. II, Academic Press, New York, 1973.
24. C.E. Kundrot, J.W. Ponder, F.M. Richards, *Algorithms for calculating excluded volume and its derivatives as a function of molecular conformation and their use in energy minimization*, Journal of Computational Chemistry **12**(3), 402–409, 1991.
25. B. Lee, F.M. Richards, *The interpretation of protein structures: Estimation of static accessibility*, Journal of Molecular Biology **55**(3), 379–400, 1971.
26. J. Lelong-Ferrand, *Geometrie Differentielle*, Chaps. 10.5–10.6. Masson, Paris, 1983.
27. R. Lustig, *Geometry of four hard Fused spheres in an arbitrary spatial configuration*, Journal of Computational Chemistry **59**(2), 195–207, 1986.
28. J.E. Meslamani, F. André, M. Petitjean, *Assessing the geometric diversity of cytochrome P450 ligand conformers by hierarchical clustering with a stop criterion*, Journal of Chemical Information and Modeling **49**(2), 330–337, 2009.
29. A.Y. Meyer, *The size of molecules*, Chemical Society Reviews **15**(4), 449–474, 1986.
30. A.Y. Meyer, *Molecular mechanics and molecular shape. V. On the computation of the bare surface area of molecules*, Journal of Computational Chemistry **9**(1), 18–24, 1988.
31. A.Y. Meyer, *More on the size of molecules*, Journal of Structural Chemistry **1**(2–3), 265–279, 1990.
32. J.J. Müller, *Calculation of scattering curves for macromolecules in solution and comparison with results of methods using effective atomic scattering factors*, Journal of Applied Crystallography **16**(1), 74–82, 1983.
33. NAG Fortran Library Routine Document
<http://www.nag.co.uk/numeric/fl/manual/xhtml/G05/g05caf.xml>.
34. R. Pavanì, G. Ranghino, *A method to compute the volume of a molecule*, Journal of Computational Chemistry **6**(3), 133–135, 1982.
35. M.Y. Pavlov, B.A. Fedorov, *Improved technique for calculating X-ray scattering intensity of biopolymers in solution: Evaluation of the form, volume, and surface of a particle*, Biopolymers **22**(6), 1507–1522, 1983.

36. G. Perrot, B. Cheng, K.D. Gibson, J. Vila, K.A. Palmer, A. Nayeem, B. Maigret, H.A. Scheraga, *MSEED: a program for the rapid analytical determination of accessible surface areas and their derivatives*, Journal of Computational Chemistry **13**(1), 1–11, 1992.
37. M. Petitjean, *On the analytical calculation of van der Waals surfaces and volumes: some numerical aspects*, Journal of Computational Chemistry **15**(5), 507–523, 1994.
38. M. Petitjean, *Geometric molecular similarity from volume-based distance minimization: application to saxitoxin and tetrodotoxin*, Journal of Computational Chemistry **16**(1), 80–90, 1995.
39. M. Petitjean, *About the algebraic solutions of smallest enclosing cylinders problems*, arXiv:1008.5259, 2010.
40. M. Petitjean, *ASV freeware*,
<http://petitjeanmichel.free.fr/itoweb.petitjean.freeware.html>.
41. M. Petřek, P. Košinová, J. Koča, M. Otyepka, *MOLE: a Voronoi diagram-based explorer of molecular channels, pores, and tunnels*, Structure **15**(11), 1357–1363, 2007.
42. M. Petřek, M. Otyepka, P. Banáš, P. Košinová, J. Koča, D. Damborský, *CAVER: a new tool to explore routes from protein clefts, pockets and cavities*, BMC Bioinformatics **7**(316), 1–9, 2006.
43. G.N. Ramachandran, V. Sasisekharan, *Conformations of polypeptides and proteins*. In: “Advances in Protein Chemistry”, vol. 23, pp. 283–437 (appendix by the Editors p. 438), Academic Press, New-York, 1968.
44. J. Reiczigel, *Confidence intervals for the binomial parameter: some new considerations*, Statistics in Medicine **22**(4), 611–621, 2003.
45. A. Renyi, *Probability theory*, North Holland Publishing Company, Amsterdam, 1970.
46. F.M. Richards, *Calculation of molecular volumes and areas for structures of known geometry*. In: “Diffraction Methods for Biological Macromolecules, part B”, H.W. Wyckoff, C.H.W. Hirs, S.N. Timasheff (Eds.), volume 115 of *Methods in Enzymology*, 440–464, Academic Press, London, 1985.
47. T.J. Richmond, *Solvent accessible surface area and excluded volume in proteins. Analytical equations for overlapping spheres and implications for the hydrophobic effect*, Journal of Molecular Biology **178**(1), 63–89, 1984.
48. B.A. Rosenfeld, *A history of non-euclidean geometry. Evolution of the concept of a geometric space*, Springer-Verlag, New-York, 1988.
49. J.S. Rowlinson, *The triplet distribution function in a fluid of hard spheres*, Journal of Molecular Physics **6**(5), 517–524, 1963.
50. R.A. Scott, H.A. Scheraga, *Conformational analysis of macromolecules. III. Helical structures of polyglycine and poly-L-alanine*, Journal of Chemical Physics **45**(6), 2091–2101, 1966.
51. R.E. Stobaugh, *Chemical abstracts service chemical registry system. 11. Substance-related statistics: Update and additions*, Journal of Chemical Information and Computer Sciences **28**(4), 180–187, 1988.
52. J.J. Stoker, *Differential geometry*, Wiley, New-York, 1969.
53. R. Vakili, *A mathematical mosaic. Patterns and problem solving*. Chapter: “Geometry Revisited”, p. 160. Brendan Kelly Publishing Inc. Burlington, Ontario, 1996.
54. S.J. Wodak, J. Janin, *Analytical approximation to the accessible surface area of proteins*, Proceedings of the National Academy of Sciences of the USA **77**(4), 1736–1740, 1980.
55. Y.V. Zefirov, P.M. Zorkii, *Van der Waals radii and their application in chemistry*, Russian Chemical Review **58**(5), 421–440, 1989. Uspekhi Khimii **58**(5), 713–746, 1989 (in Russian).