



Full paper/Mémoire

Analytical algorithms for ligand cone angles calculations. Application to triphenylphosphine palladium complexes

*Algorithmes analytiques pour le calcul des angles coniques de ligands.
Application à des complexes palladium triphénylphosphine*

Michel Petitjean

MTi, UMR-S 973, INSERM, Université Paris-Diderot (Paris-7), France

ARTICLE INFO

Article history:

Received 17 February 2015

Accepted after revision 14 April 2015

Available online 20 May 2015

Keywords:

Tolman cone angle

Ligand cone angle

Minimal enclosing cone

Best fitting d -dimensional cone

Least squares

Conicity index

Mots clés :

Angle conique de Tolman

Angle conique de ligand

Cône minimal englobant

Meilleur cône moyen d -dimensionnel

Moindre carrés

Indice de conicité

ABSTRACT

We defined the smallest enclosing cone angle as the Tolman cone angle for null atomic spheres radii. Then we provide a simple analytical algorithm to compute the smallest enclosing cone at fixed apex, which works in the case of unsymmetrical ligand. We applied it to compute ligand cones for a family of triphenylphosphine palladium complexes, and we showed that both the angle of the cone and its resulting solid angle strongly correlate with the Tolman cone angle, thus suggesting that there is no more need for atomic radii. We also defined the best cone of fixed apex fitting a population of unit vectors. We proposed a simple analytical algorithm to compute it, which is proved to work in any d -dimensional Euclidean space. We defined the conicity index κ to evaluate quantitatively the pertinence of the best fitting cone. We used this best fit cone to define a mean ligand cone, and thus a mean cone angle and a mean cone axis. We applied it to our family of triphenylphosphine palladium complexes and we observed that the axis of the individual cones deviated from the mean cone axis by at most 13.2° . The observed conicity index was small ($\kappa = 0.0177$), indicating a very good fit for the whole family of complexes.

© 2015 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

R É S U M É

Nous définissons l'angle du plus petit cône englobant comme étant l'angle conique de Tolman à rayons atomiques nuls. Puis, nous fournissons un algorithme analytique simple de calcul du plus petit cône englobant à apex fixé, qui fonctionne dans le cas des ligands non symétriques. Nous l'appliquons aux cônes de ligands pour une famille de complexes palladium triphénylphosphine et nous montrons qu'à la fois l'angle du cône et l'angle solide qui en résulte sont fortement corrélés avec l'angle conique de Tolman, suggérant ainsi qu'il n'y a plus besoin des rayons atomiques. Nous définissons aussi le meilleur cône moyen d'apex fixé pour une population de vecteurs unitaires. Nous proposons un algorithme analytique simple pour le calculer, que nous prouvons être valide dans tout espace euclidien d -dimensionnel. Nous définissons l'indice de conicité κ pour évaluer quantitativement la pertinence du meilleur cône. Nous utilisons ce meilleur cône pour définir un cône moyen de ligand, et donc un angle moyen de cône et un axe moyen de cône. Nous l'appliquons à notre famille de complexes palladium triphénylphosphine et nous

E-mail addresses: michel.petitjean@univ-paris-diderot.fr, petitjean.chiral@gmail.com.

<http://dx.doi.org/10.1016/j.crci.2015.04.004>

1631-0748/© 2015 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

observons que les axes individuels des cônes dévient de l'axe moyen de cône d'au plus 13,2°. L'indice de conicité observé est faible ($\kappa = 0,0177$), indiquant un très bon ajustement à l'ensemble de la famille de complexes.

© 2015 Académie des sciences. Publié par Elsevier Masson SAS. Tous droits réservés.

1. Introduction

Ligand cone angles were introduced by Tolman to measure the size of phosphine derivatives and other phosphorus ligands [1]. This size is the solid angle defined by the smallest angle cone having its apex lying at 2.28 Å from the phosphorus atom and circumscribing the ligand atoms, usually modelled by spheres. The solid angle, expressed in steradians, is $\theta = 2\pi(1 - \cos\alpha)$, where α is the angle between the generatrix of the cone and its axis (see general definitions in Section 3.1). In the case of symmetric ligands PR_3 , the Tolman cone angle is easy to compute because its axis is in the direction of the mean of the three vectors defined by the P–R bonds. Then, given the radius of the spherical ligands, α is retrieved by elementary geometry calculus. This approach has to be refined for unsymmetrical ligands such as $MPR_1R_2R_3$ and more generally for $MXR_1R_2R_3$, where $M=H$ or is a metal atom, and $X=P, N, CH$, or is any atom having a tetrahedral hybridisation such as sp^3 or sd^3 . It was pointed out that even symmetrically substituted bulky phosphines may offer unsymmetrical conformations [2]. The difficulty of the calculation arises (i) when the spheres radii are unequal, and (ii) when the X– R_i bonds are not symmetrically arranged around the M–X axis. Tolman approximated θ as $(\hat{\theta}_1 + \hat{\theta}_2 + \hat{\theta}_3)/3$, where $\hat{\theta}_i$ is the acute angle between the directions of the X– R_i bond and of the M–X bond [3,4]. This method was criticized because the obtained values may not reflect the properties of the ligand, particularly when the substituent groups differ greatly [5]. It seems that few geometric tools are available to measure steric effects in organometallic chemistry, and that could explain why ligand cone angles were much used in this field [6–9]. Recently, Bilibrey et al. [10–12] proposed an analytic solution to the ligand cone angle calculation.

It was also proposed to measure the steric size of ligands and substituents by the solid angle generated by the union of the atomic spheres [13], rather than the one generated by their enclosing cone. This approach gives rise to an analytical calculation of the solid angle, provided that the intersections of more than two spheres could be neglected [13]. There is a non-linear relationship between the cone angle and the solid angle, which was measured quantitatively [14]. To evaluate the importance of sphere overlaps, an exact analytical calculation of sphere intersections was done with the ASV freeware [15] using the atomic radii recommended by Gavezzotti [16], and showed that intersections between 6 or 7 atoms are commonly observed in organic molecules [15,18]. These atomic radii are sometimes slightly larger than those given by Bondi [17], but it is recalled that an increase in the sphere radii does not guarantee an increase of the van der Waals surfaces. Running ASV on a database of 70 diverse ligands

showed that neglecting the intersections of more than two atoms induced a mean error on van der Waals surface calculations of 249%, and that neglecting the intersections of more than three atoms led to a mean error of 87%, and neglecting the ones of more than four atoms led to a mean error to 16%, the maximal observed error in this case being 37% [18]. Despite that van der Waals surfaces are not used in ligand solid angles calculations, these numbers show the importance of atomic spheres overlaps.

An improved ligand solid angle algorithm was proposed, which takes into account spheres intersections of orders 3 and 4 [19], but it needs a complex numerical integration. Recently, Bilibrey et al. [12,20] proposed an analytic solution to the solid angle calculation, based on the decomposition of the solid angle contributions between those due to spherical polygon parts and those due to the resulting truncated spherical sectors parts. This algorithm, implemented by the authors in their Mathematica FindSolidAngle package, is effective for the simple geometrical arrangements expected to be encountered in chemistry. However, it is not specified how it works in general. E.g., the detection of potential multiple connected components got by projection at the surface of the unit sphere is not evoked, the solid angle subtended by an internal spherical polygon may be not void and may even be not unique, etc. The detection and the management of such situations let the algorithm rather difficult to implement.

The impact of conformational variations was evoked early [21], leading to use weighted average cone angles [22], while it was considered that this problem was overcome by the use of the solid angle methodology [23]. At the same time, Müller and Mingos noticed also that the Tolman cone angle definition does not take into account the variations due to conformational changes [24,25], and they used the atomic centers of the ligand atoms rather than their van der Waals spheres. Then they applied their algorithm to perform statistics on thousands of phosphine structures found in the Cambridge Crystallographic Data Base [26], and observed a variation in cone angles for specific ligands, which is much larger than had previously been suspected.

This slight change in the ligand cone angle calculation, that we retain here (see Fig. 1) offers two other major advantages: (i) the cone angle can be generalized to complex polyatomic ligands R_i via the calculation of the fixed apex minimal cone enclosing any desired number of atoms, and (ii) this calculation can be done analytically, as shown in Section 3 of the present paper. We emphasize that this generalization allows us to model molecular shapes and structural fragments with cones although it is usual to work with spherical models. Despite that is easy to compute spherical shapes, the spherical model was shown

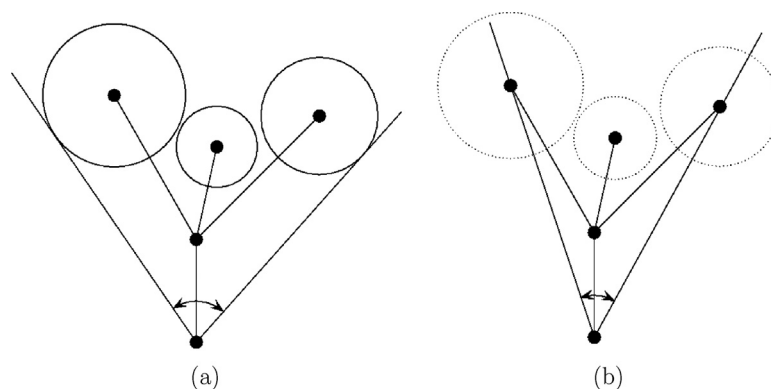


Fig. 1. (a) Ligand cone, defined by Tolman [1]. (b) Smallest enclosing cone, defined here.

to be unrealistic and a cylindrical model was preferred for drug design applications [27]. Fortunately, minimal height enclosing cylinders and minimal radius enclosing cylinders are computable analytically [28]. However, it seems that apart cylinders and cones, it is hard to find the use of non-spherical molecular shapes in the literature: it may be due to the lack of simple analytical calculation algorithms.

There are several ways to take into account the conformational changes of the ligands. We propose the following one. For each conformer, we know from the present analytical minimal enclosing cone algorithm which ligand atom centers are on the surface of the cone (see Section 3.5). We mark the atomic centers of these ligand atoms. These marked ligand atoms can differ from one conformer to another conformer, even for simple ligands such as Me or Et. Then, assuming a common origin in M, we are left with the problem of finding the best cone of fixed apex fitting all the marked atomic centers. We give in Section 3.3 an analytical solution to this problem, formulated as a least squares one. Then we define in Section 3.4 the conicity index κ , which takes values in the interval $[0,1]$, the value $\kappa=0$ meaning that all marked atomic centers lies on the surface of the cone, and the value $\kappa=1$ being reached in the worst cases, characterized in Section 3.4. It is emphasized that, compared to our best fitting cone algorithm, computing some mean cone angle such as the arithmetic mean of individual cone angles, has drawbacks: such mean cone angle does not produce a mean axis, and computing some mean axis would not be coherent with the arithmetic mean of the cone angles. Furthermore, such a method would not permit to define a conicity index, although this latter provides quantitative information about the impact of conformational changes on steric effects. The axis of the best fitting cone is of interest because it gives rise to a second quantitative parameter: the acute planar angle between the axis of the smallest enclosing cone and the axis of the best fitting cone. This parameter indicates how the ligand size of $\text{MXR}_1\text{R}_2\text{R}_3$ deviates from the mean ligand size of the family. At the opposite of the well-known RMS (Root Mean Square) deviation, it does not need the knowledge of a mean conformer.

A minor problem is to suppress the impact on a best fit cone calculation of free rotations around the M–X axis before aligning the conformers in a common Cartesian

coordinate system. When R_1 , R_2 and R_3 are different, a 3D rotation performed to optimally superpose each conformer on a common reference conformer solves the problem. It is proposed to set the pivot at M and to restrict this optimal rotation to X and to the respective three atoms of R_1 , R_2 and R_3 that are bonded to X, rather to involve more atoms when the R_i are polyatomic. The reason is that extending the optimal superposition to more atoms may give poor alignments in the neighborhood of X in the case of bulky ligands, while for usual applications of cone angles the neighborhood of X is assumed to be more important than the rest of the ligands. Furthermore, the restriction to X and to its neighbours permits potential extensions to superpositions of different molecules $\text{MXR}_1\text{R}_2\text{R}_3$ rather than to different conformers of a common molecule $\text{MXR}_1\text{R}_2\text{R}_3$, thus generalizing the definition of the best fitting cone. After translating the M atom of each conformer at the origin, each desired optimal 3D rotation can be found by minimizing the RMS deviation by the least squares method implemented in the ARMS freeware, which is based on quaternions (see appendix in [29], or appendix A.5 in [30] for more general results about optimal rotations). When two or three ligands are identical, there are respectively two or six pairwise correspondences between the ligands atoms bonded to X. In this situation, the one with the smallest minimized RMS is retained.

2. Results and discussion

We exemplify our minimal enclosing cone algorithm using a family of palladium triphenylphosphines complexes (Table 1). The resulting cone is equivalent to the Tolman cone for null atomic radii. The angle values we got are in the range $57.7\text{--}64.6^\circ$, and should be compared with the half of the Tolman cone angle values, which ranged in the interval $150.3\text{--}173.6^\circ$. This difference of a factor 2 is due to our mathematical definition of the cone angle, which stands in E^d (see Section 3.1).

The observed correlation coefficient between the ligand cone angle in [10] and our minimum enclosing cone angle α is $r_\alpha = 0.9800$, and with our solid angle θ is $r_\theta = 0.9796$, while α and θ are highly correlated (0.99995). Since the ligand cone angles encountered in the literature are almost all times used for empirical correlations with physical data, it is simpler to calculate α rather than the usual ligand cone

Table 1

Minimal enclosing cone angles for a family of palladium triphenylphosphines complexes. α : angle between the axis and the generatrix. θ : solid angle of the cone.

	Palladium complex (data from ref. [10])	$\cos \alpha$	α (degrees)	θ (steradians)	Cone angle from ref. [10]
1	Pd(PPh ₃)	0.441571	63.796	3.509	170.0
2	Pd(PPh ₃) ₂ (SN ₂ C ₃) ₂ Cl	0.532900	57.798	2.935	150.3
3	Pd(PPh ₃) ₂ (SN ₂ C ₃) ₂ Cl	0.521229	58.585	3.008	155.4
4	Pd(PPh ₃)(P ₂ OC ₁₄ H ₉)Cl	0.534833	57.667	2.923	151.4
5	Pd(PPh ₃)(SN ₄ O ₂ C ₈ H ₇)Cl	0.514252	59.053	3.052	155.2
6	Pd(PPh ₃)(SN ₃ C ₉ H ₁₀)Cl	0.501834	59.879	3.130	156.9
7	Pd(PPh ₃)(SH)	0.483577	61.081	3.245	160.8
8	Pd(PPh ₃) ₂ (S ₃ NO ₂ C ₇ H ₅)	0.489504	60.692	3.208	160.9
9	Pd(PPh ₃) ₂ (S ₃ NO ₂ C ₇ H ₅)	0.434539	64.244	3.553	173.6
10	Pd(PPh ₃)(SNC ₅ H ₄)	0.474321	61.685	3.303	163.1
11	Pd(PPh ₃)(NFC ₁₅ H ₁₅)Cl	0.485598	60.948	3.232	165.9
12	Pd(PPh ₃)(SN ₃ C ₁₀ H ₁₁)	0.458999	62.677	3.399	167.4
13	Pd(PPh ₃)(S ₂ N ₃ C ₈ H ₉)	0.458199	62.729	3.404	167.7
14	Pd(PPh ₃)(SN ₄ C ₈ H ₁₀)	0.451533	63.158	3.446	170.6
15	Pd(PPh ₃)(NFC ₁₁ H ₁₅)Cl	0.429580	64.559	3.584	172.2

angles because there is no need for atomic radii. The correlation between α and θ is not surprising because θ is a function of α . Then, the high value of the correlation coefficient indicates that the relationship can be estimated as linear for the considered ligands.

Published values of ligand cone angles are close to 180° (see Table 1) and can be greater than 180° for some nickel or platinum complexes [10], while 2α is around 120°. This difference is due to the exclusion of the ligand atomic spheres in the calculation of α . On the other hand, modeling the molecular shape by a cone should either take in account all atomic spheres, including the one of the metal (estimated to 1.63 Å for Pd [17]), or ignore all atomic spheres. Locating the apex of the cone enclosing all atomic spheres (including the metal one) would lead to a complex algorithm, and worse, would need the knowledge of adequate atomic radii. Ignoring only the atomic sphere of the metal gives large angle cones, close to a half plane, which are not realistic from the molecular shape point of view. Ignoring all atomic spheres leads to a very simple analytical algorithm (Section 3.5), it does not require atomic radii, the angle values are still pertinent for establishing empirical calculations, and the conical molecular shape is physically more realistic than a half plane. E.g., for the tetrakis Pd(PPh₃)₄ a solid angle close to 0.5 half space per Pd(PPh₃) part is more realistic than a solid angle around one half space because for this latter the sum of the four contributions of the Pd(PPh₃) parts is around two full spaces, thus indicating excessive cone intersections.

In order to define a mean cone for the 15 complexes of Table 1 and to evaluate quantitatively the dispersion around this mean cone, we operated as follows. The mean cone has sense only in a common Cartesian coordinate system: we selected Pd(PPh₃) (see Table 1) as the reference complex to perform a 3D superposition of each of the 14 other complexes onto this reference one. We set the

palladium atom as the common origin for the 15 complexes and we computed the 14 optimal rotations as indicated at the end of Section 1. There was an additional difficulty due to the differences in the atom numbering of the complexes. Thus, to retrieve each of the 14 pairwise correspondences for the phosphorus and its neighboring carbons, we used the CSR freeware, based on an automatic 3D motif recognition [31].

For each of the 15 complexes, we got 3 contact points on the surface of their individual smallest enclosing cone, all with apex at the origin. These 45 points are in a common Cartesian coordinate system and thus we computed the best cone fitting these 45 points (see Section 3.3). This best fit cone has an angle $\bar{\alpha} = 60.616^\circ$ while the smallest cone enclosing the 45 points has an angle of 68.090°. The resulting 15 angles between the axis \bar{u} of the best fit cone and the 15 individual cone axes are given in Table 2. This angle, denoted by γ , indicates how the ligand size deviates from the mean one of the family. At the opposite of the smallest enclosing cone angle, it is poorly correlated with the Tolman cone angle (correlation coefficient: 0.511).

We measured the global dispersion of the directions of the 45 points around the surface of the best fit cone with the conicity index κ (see Section 3.4), which takes values in 0;1. We found $\kappa = 0.0177$. A null value would have meant that all 45 points are on the surface on the cone, although only 2 or 3 are expected to be found on the surface, in general (see Section 3.5). It indicates that the observed differences of conformations of the phenyl groups in the input structural files have little effect on the cone calculation. The largest angle was the one of Pd(PPh₃)(S₂N₃C₈H₉).

It is emphasized that the knowledge of the mean cone leads us to define not only a mean angle, but also a mean axis and deviations from this mean axis: that was not possible with usual ligand cone angle approaches.

Table 2

Angles $\gamma = (u, \bar{u})$ in degrees, between the axis \bar{u} of the best fit cone and the cone axis u of each of the 15 complexes, numbered as in Table 1.

Complex	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
γ	4.4	1.9	6.7	2.2	8.8	6.6	7.2	8.3	10.3	4.6	6.4	8.4	13.2	12.3	5.0

The minimal angle enclosing cone algorithm and the best fitting d -dimensional cone algorithm were implemented in the freeware CONE. Sources are written in portable f77. Documentation and binaries for Mac OS 10 and 64 bits Intel linux platforms are available free of charge on a software repository located at <http://petitjeanmichel.free.fr/itoweb.petitjean.freeware.html>.

Running CONE on Windows platforms can be done through the installation of a linux emulator such as Cygwin (free). When needed, convex hull calculations (see Section 3.5) can be done with the freeware RAD1. This latter can be found on the same software repository than CONE together with the freewares ARMS and ASV mentioned in Section 1, and with the CSR freeware mentioned in Section 2.

3. Appendix: analytical results and algorithms

3.1. Definitions and notations

Definition 1. In the the Euclidean space E^d , a cone of apex x_0 is a ruled surface generated by the set of all lines intersecting x_0 and having a constant angle α with a given axis containing x_0 . Each of these lines is called a generatrix.

This axis is defined by a unit vector u , and we set conventionnally $c = \cos \alpha$ as being a non-negative value. The case $c=0$ corresponds to the plane orthogonal to u and containing x_0 . The case $c=1$ corresponds to a degenerated cone reduced to its axis. Generalizations to non-constant angles (non-circular cones) are not considered here. Thus, a cone in E^d is defined by $2d$ free parameters: x_0 , c , and u . Its equation is the set of points x so that $u'(x - x_0) = c\|x - x_0\|$, where the quote indicates a transposition operation and where the norm of $(x - x_0)$ is $\|x - x_0\| = \sqrt{(x - x_0)'(x - x_0)}$.

Remark 1: Given the axis defined by u , the word *cone* applies in some contexts to the points x satisfying to the additional constraint $\sin \alpha \geq 0$, which in fact lets to consider only a half of the cone. In E^3 , this latter encompasses a solid angle equal to $2\pi(1 - \cos\alpha)$. Unless otherwise stated, we retain the definition corresponding to a full cone.

Remark 2: In some contexts the cone is defined as the convex set such that $0 \leq u'(x - x_0) \leq \|x - x_0\|\cos \alpha$, still with $\cos \alpha \geq 0$, in which case the half cone in the sense of Definition 1 is the boundary of this latter convex set. There are other variants. For convenience, we retain Definition 1.

We consider $n+1$ distinct given points x_i , $i = 0, 1, \dots, n$, in E^d . We define the unit vectors $v_i = (x_i - x_0)/\|x_i - x_0\|$, $i = 1, \dots, n$, and the matrix W with n lines and d columns, each line i containing v_i' , $i = 1, \dots, n$.

We use the following notations. I is the identity matrix of rank n . $\mathbf{1}$ is the vector having n components, all equal to 1. $A = I - \mathbf{1}\mathbf{1}'/n$ is the centering operator (thus $A = A' = A^2$). $T = W'AW$ is the inertia matrix associated with W , i.e. T is n times the covariance matrix of the v_i , $i = 1, \dots, n$. $P = I - uu'$ is the projection matrix on the $(d - 1)$ -dimensional subspace generated by the vectors orthogonal to u .

3.2. Calculation of a circumscribed cone

Definition 2. A cone circumscribed to k points is a cone such that the k points lie on its surface.

We assume that the apex x_0 is fixed and we would like to find the cone circumscribed to the points x_1, \dots, x_n . The resulting system has n equations $u'v_i = c$, $i = 1, \dots, n$ or, in matricial form, $Wu = c\mathbf{1}$. It has d unknowns and thus it is underdetermined when $n < d$ and it is overdetermined when $n > d$. Thus we consider the case $n=d$ and we assume that the square matrix W is invertible, i.e. no $(d - 1)$ -dimensional plane contains the n points.

Theorem 1. The cone of apex x_0 circumscribed to $n=d$ points in E^d has its axis in the direction of the unit vector $u = W^{-1}\mathbf{1}/\sqrt{\mathbf{1}'(WW')^{-1}\mathbf{1}}$ and the cosine of its angle is $\cos \alpha = 1/\sqrt{\mathbf{1}'(WW')^{-1}\mathbf{1}}$.

Proof. W is invertible, thus $u = cW^{-1}\mathbf{1}$. Because u is a unit vector and c was conventionnally set non-negative, we get $c = \mathbf{1}'/\sqrt{\mathbf{1}'(WW')^{-1}\mathbf{1}}$ and $u = W^{-1}\mathbf{1}/\sqrt{\mathbf{1}'(WW')^{-1}\mathbf{1}}$.

Remark 3: It is checked that we have indeed the solution value $c^2 \leq 1$ as follows. The symmetric matrix WW' have non-negative eigenvalues and thus its largest eigenvalue cannot exceed its trace, this latter being equal to n . Then the smallest eigenvalue of $(WW')^{-1}$ cannot be smaller than $1/n$, and from the Courant–Fischer minimax theorem [32], the quadratic form $\mathbf{1}'(WW')^{-1}\mathbf{1}$ cannot be smaller than $(\mathbf{1}'\mathbf{1}/n) = 1$, so that $c^2 \leq 1$.

Remark 4: There are two particular situations: $c = 0$ and $c = 1$. The case $c=0$ arises if and only if the square matrix W is non-invertible: $W'\mathbf{1} = 0$. It is such that $u'v_i = 0$ for all $i = 1, \dots, n$, which corresponds to a cone which is a plane orthogonal to u and containing x_0 . The case $c=1$ arises if and only if all the n quantities $u'v_i = 1$, which means that $v_i = u$ for all $i = 1, \dots, n$: all points are aligned on the axis and the cone is degenerated.

3.3. Least squares best fitting cone

We consider the case $n \geq d$. In general we cannot have the n equalities $Wu - c\mathbf{1} = 0$ all satisfied together, but we can minimize $S = \|Wu - c\mathbf{1}\|^2$, i.e. we look for the values of (c, u) minimizing $S = (Wu - c\mathbf{1})'(Wu - c\mathbf{1})$.

Theorem 2. The optimal unit vector u is the eigenvector associated with the smallest eigenvalue λ_d of the inertia matrix $T = W'AW$ and the minimized sum of the squared distances of the n points to the surface of the cone is λ_d .

Proof. The solution of the optimization problem above should satisfy to $\text{grad}[S + L(1 - u'u)] = 0$, L being the Lagrangian associated with the constraint $u'u = 1$. It follows that $2nc - 2(1'Wu) = 0$ and that $2W'Wu - 2cW'\mathbf{1} - 2Lu = 0$.

We get $c = \mathbf{1}'Wu/n$, and then $L = u'W(I - \mathbf{1}\mathbf{1}'/n)Wu$, i.e. $L = u'Tu$. Then we get $S = u'Tu$ and $(I - uu')W'AWu = 0$, i.e. $PTu = 0$ and $Tu = Lu$. The latter equation is satisfied if and only if Tu is proportional to u , which means that u is an eigenvalue of T . Let $\lambda_1, \dots, \lambda_d$ be the eigenvalues of T sorted in decreasing order. From the Courant–Fischer minimax theorem, the minimum of S is $S^* = \lambda_d$ and the maximum of S is λ_1 .

Remark 5: When $n = d$, W is a square matrix so that AW is not of full rank because $1'A = 0$ and thus $1'AW = 0$. Then, $T = W'AW$, which can be written $T = (AW)'(AW)$ because $A^2 = A$, cannot be of full rank. We deduce that $\lambda_d = 0$ and we find again that $S=0$ and $Wu = c\mathbf{1}$.

Remark 6: For any n and discarding whether or not u is optimal, it is checked that we have indeed the optimal value $c^2 \leq 1$ as follows. We have $n^2c^2 = (\mathbf{1}'W)uu'(W'\mathbf{1})$, which is a quadratic form, which cannot exceed $(\mathbf{1}'W)(W'\mathbf{1})$ because the largest eigenvalue of uu' is 1. But $(\mathbf{1}'WW'\mathbf{1})$ cannot exceed $\mathbf{1}'\mathbf{1} = n$ times the largest eigenvalue of WW' , this latter being not greater than the trace of WW' , i.e. n . Thus $n^2c^2 \leq n^2$ and $c^2 \leq 1$. The sense of the eigenvector u is set to have a non-negative c value.

Remark 7: Another least squares method would be to minimize the sum of n squared distances of the points x_i to their orthogonal projection on the cone, rather than the sum of n squared distances of the v_i to their orthogonal projection on the cone. In this situation, a numerical minimizer is required.

3.4. Conicity index

Definition 3. The conicity index is $\kappa = d \cdot S^*/n$.

Theorem 3. κ takes values in $[0;1]$.

Proof. We look for the upper bound of the minimal $S^* = \lambda_d$ given d and n . The maximal value for λ_d is λ_1 . In this situation all the d eigenvalues of T are equal to $\text{Tr}(T)/d$, i.e. $d \cdot S^* = \text{Tr}(W'W) - \text{Tr}(W'\mathbf{1}'W)/n$. Thus $d \cdot S^* = n(1 - \bar{v}'\bar{v})$, where the mean of the n unit vectors is $\bar{v} = W'\mathbf{1}/n$. Since $\|\bar{v}\| \leq 1$, $S^* \in [0, n/d]$.

The value $\kappa=0$ indicates that all points lie at the surface of the cone. The value $\kappa=1$ indicates the worst possible fitting by the cone. It is reached when T is proportional to the identity matrix I and when $\bar{v} = 0$. This extreme value is reached, e.g., for the vertices of a regular d -simplex with the apex at its center [33].

3.5. Smallest enclosing cone

Having a fixed apex x_0 and n input data points x_1, \dots, x_n , we propose in the case $d=3$ a minimal enclosing cone algorithm in the sense of a minimal angle α . Such a cone has 3 free parameters: the unit vector u and the angle α . A point x_i is enclosed in the cone if $\cos(x_i - x_0, u) \geq \cos \alpha$.

The smallest enclosing cone is sought among the minimal cones circumscribed to successively $k = 1, 2$, and 3 points and containing the $n-k$ other points.

The trivial case $k=1$ corresponds to n points aligned with x_0 : if it is the case the algorithm stops.

The case $k=2$ is solved via enumerating the $n(n-1)/2$ pairs of input points and computing for each pair its minimal circumscribed cone (the minimal cone circumscribed to 2 points x_i and x_j is such that its axis is bisecting the angle $x_1 - x_0 - x_2$). If the latter encloses all points, it is retained. If there is one or several retained cones, the one with the smallest angle is the solution and the algorithm stops.

If not, we enumerate the $n(n-1)(n-2)/6$ triplets of input points and we compute for each triplet its circumscribed cone as shown in Section 3.2. The one with

the smallest angle is retained and the algorithm terminates.

Remark 8: For some applications the minimal enclosing half cone of fixed apex x_0 needs to be considered. If it happens that the algorithm above does not output such a half cone, another algorithm is required, based on circumscribed half cones. If we add at each step of the above algorithm the constraint that all n points are enclosed in the same half cone, either a valid optimal half cone is returned, or no half cone is found. When x_0 is an extreme point of the convex hull of $\{x_0, x_1, \dots, x_n\}$, there is at least one enclosing half cone because a half cone is a convex set, and therefore the minimal enclosing half cone necessarily exists. When x_0 is in the interior of the convex hull of $\{x_0, x_1, \dots, x_n\}$, or equivalently, x_0 is interior to the convex hull of $\{x_1, \dots, x_n\}$, no enclosing half cone exists.

The convex hull of a finite set of points can be computed by standard methods such as the beneath-beyond method [34]. It is pointed out that, for large n values, the computation of the smallest enclosing half cone can be much faster if it is applied to the vertices of the convex hull of the n points rather than to the n points.

Remark 9: When a finite-volume revolution cone is needed, it is proposed to orthogonally project the n points x_i on the axis, then to close the conic solid by two circular disks orthogonal to the axis and intersecting it at the two extreme projected points. If a half cone was considered, only one disk is needed.

Acknowledgements

The author is grateful to one of the reviewers for his/her encouraging comments and for a pertinent suggestion.

References

- [1] C.A. Tolman, J. Am. Chem. Soc. 92 (1970) 2956.
- [2] A. Immirzi, A. Musco, Inorg. Chim. Acta 25 (1977) L41.
- [3] C.A. Tolman, W.C. Seidel, L.W. Gosser, J. Am. Chem. Soc. 96 (1974) 53.
- [4] C.A. Tolman, Chem. Rev. 77 (1977) 313.
- [5] T.L. Brown, K.J. Lee, Coord. Chem. Rev. 128 (1993) 89.
- [6] N.J. Coville, K. du Plooy, W. Pickl, Coord. Chem. Rev. 116 (1992) 1.
- [7] N.J. Coville, M.S. Loonat, D. White, L. Carlton, Organometallics 11 (1992) 1082.
- [8] D. White, L. Carlton, N.J. Coville, J. Organomet. Chem. 440 (1992) 15.
- [9] D. White, N.J. Coville, Adv. Organomet. Chem. 36 (1994) 95.
- [10] J.A. Bilbrey, A.H. Kazez, J. Locklin, W.D. Allen, J. Comput. Chem. 34 (2013) 1189.
- [11] J.A. Bilbrey, W.D. Allen, Ann. Rep. Comp. Chem. 9 (2013) 3.
- [12] J.A. Bilbrey, Doctoral thesis dissertation, sections 2 & 3, Georgia, Athens, 2014.
- [13] D. White, B. Craig Taverner, P.G.L. Leach, N.J. Coville, J. Comput. Chem. 14 (1993) 1042.
- [14] I.A. Guzei, M. Wendt, Dalton Trans. 33 (2006) 3991.
- [15] M. Petitjean, J. Comput. Chem. 15 (1994) 507.
- [16] A. Gavezotti, J. Am. Chem. Soc. 105 (1983) 5220.
- [17] A. Bondi, J. Phys. Chem. 68 (1964) 441.
- [18] M. Petitjean, in: A. Mucherino, C. Lavor, L. Liberti, N. Maculan (Eds.), Distance geometry: theory, methods, and applications. Chap. 4, Springer, 2013, pp. 61–83.
- [19] B. Craig Taverner, J. Comput. Chem. 17 (1996) 1612.
- [20] J.A. Bilbrey, A.H. Kazez, J. Locklin, W.D. Allen, J. Chem. Theory Comput. 9 (2013) 5734.
- [21] J.T. DeSanto, J.A. Mosbo, B.N. Storhoff, P.L. Bock, R.E. Bloss, Inorg. Chem. 19 (1980) 3086.
- [22] M. Chin, G.L. Durst, S.R. Head, P.L. Bock, J.A. Mosbo, J. Organomet. Chem. 470 (1994) 73.

- [23] D. White, B. Craig Taverner, N.J. Coville, P.W. Wade, J. Organomet. Chem. 495 (1995) 41.
- [24] T.E. Müller, D.M.P. Mingos, *Transit. Met. Chem.* 20 (1995) 533.
- [25] D.M.P. Mingos, in: G.J. Leigh, N. Winterton (Eds.), *Modern coordination chemistry: the legacy of Joseph Chatt*. Section 3, RSC, Cambridge, UK, 2002, pp. 69–78.
- [26] F.H. Allen, *Acta Cryst.* B58 (2002) 380.
- [27] L. Benkaidali, F. André, B. Maouche, P. Siregar, M. Benyettou, F. Maurel, M. Petitjean, *Bioinformatics* 30 (2014) 792.
- [28] M. Petitjean, *Appl. Algebra Engrg. Comm. Comput.* 23 (2012) 151.
- [29] M. Petitjean, *J. Math. Phys.* 40 (1999) 4587.
- [30] M. Petitjean, *J. Math. Phys.* 43 (2002) 4147.
- [31] M. Petitjean, *Comput. Chem.* 22 (1998) 463.
- [32] P. Lascaux, R. Théodor, *Analyse numérique appliquée à l'art de l'ingénieur*, tome 1, section 1.4.2, Masson, Paris, 1986.
- [33] M. Petitjean, *J. Math. Chem.* 22 (1997) 185.
- [34] H. Edelsbrunner, *Algorithms in Combinatorial Geometry*, in: W. Brauer, G. Rozenberg, A. Salomaa (Eds.), Section 8.4, Springer-Verlag, Berlin, 1987, p. 147.