Carlile Lavor*, Rafael Alves, Michael Souza, and José Luis Aragón

NMR Protein Structure Calculation and Sphere Intersections

Abstract: Nuclear Magnetic Resonance (NMR) experiments can be used to calculate 3D protein structures and geometric properties of protein molecules allow us to solve the problem iteratively using a combinatorial method, called Branch-and-Prune (BP). The main step of BP algorithm is to intersect three spheres centered at the positions for atoms $i - 3, i - 2, i - 1$, with radii given by the atomic distances $d_{i-3,i}, d_{i-2,i}, d_{i-1,i}$, respectively, to obtain the position for atom i. Because of uncertainty in NMR data, some of the distances $d_{i-3,i}$ should be represented as interval distances $[\underline{d}_{i-3,i}, \overline{d}_{i-3,i}]$, where $\underline{d}_{i-3,i} \leq d_{i-3,i} \leq \overline{d}_{i-3,i}$. In the literature, an extension of the BP algorithm was proposed to deal with interval distances, where the idea is to sample values from $[\underline{d}_{i-3,i}, \overline{d}_{i-3,i}]$. We present a new method, based on conformal geometric algebra, to reduce the size of $[\underline{d}_{i-3,i}, \overline{d}_{i-3,i}]$, before the sampling process. We also compare it with another approach proposed in the literature.

Keywords: Molecular Distance Geometry Problem, Branch-and-Prune, Conformal Geometric Algebra, NMR data

MSC: 51K05, 15A66, 92E10

1 Introduction

Nuclear Magnetic Resonance (NMR) experiments provide short distance values between atoms of a protein molecule. The *Molecular Distance Geometry Problem* (MDGP) asks to realize the 3D protein structure using this partial distance information [4, 5, 31].

Precisely, the MDGP concerns with a graph $G = (V, E, d)$, where V is a set of vertices representing the atoms and E is a set of edges representing the atomic pairs for which a distance is available, given by the function $d : E \to (0, \infty)$. The problem amounts to find an embedding $x : V \to \mathbb{R}^3$ such that

$$||x_u - x_v|| = d_{u,v}, \forall \{u, v\} \in E, \tag{1}$$

where $x_u = x(u)$, $x_v = x(v)$, $d_{u,v} = d(\{u, v\})$, and $||x_u - x_v||$ is the Euclidean norm.

The classical approach to the MDGP is based on global optimization methods [29], where a MDGP solution is associated to the global minimizer of the problem

$$\min_{x_1,...,x_n \in \mathbb{R}^3} f(x_1, ..., x_n), \tag{2}$$

where f is a function $f : \mathbb{R}^{3n} \to [0, \infty)$ defined by

$$f(x_1, ..., x_n) = \sum_{\{u,v\} \in E} \left(||x_u - x_v||^2 - d_{u,v}^2 \right)^2.$$

*Corresponding Author: Carlile Lavor: University of Campinas (IMECC-UNICAMP), 13081-970, Campinas - SP, Brazil, E-mail: clavor@ime.unicamp.br
Rafael Alves: Federal University of ABC (CMCC-UFABC), 09606-070, Sao Bernardo - SP, Brazil, E-mail: alves.rafael@ufabc.edu.br
Michael Souza: Federal University of Ceará (UFC), 60440-900, Fortaleza - CE, Brazil, E-mail: souza.michael@gmail.com
José Luis Aragón: Centro de Física Aplicada y Tecnología Avanzada, Universidad Nacional Autónoma de México (UNAM), 76230, Quéretaro, Mexico, E-mail: jlaragon@unam.mx

Note that $x_1, \ldots, x_n \in \mathbb{R}^3$ is a MDGP solution if, and only if, $f(x_1, \ldots, x_n) = 0$.

We may furnish the set V of vertices with an ordering $V = \{v_1, \ldots, v_n\}$ [9, 15, 23, 25] so that the MDGP can be solved iteratively using a combinatorial method, namely the *Branch-and-Prune* (BP) method [8, 28]. In this situation, the MDGP is called the *Discretizable Molecular Distance Geometry Problem* (DMDGP) [19, 20], which can be stated as follows, where we use x_i instead of x_{v_i} and $d_{i,j}$ in place of d_{v_i,v_j}:

(DMDGP) Given a simple undirected graph $G = (V, E, d)$ in which the vertex set V is ordered as $V = \{v_1, \ldots, v_n\}$, whose edges are weighted by $d : E \to (0, \infty)$, subject to the following three constraints:

1. For the initial three vertices $v_1, v_2, v_3 \in V$, there exist points $x_1, x_2, x_3 \in \mathbb{R}^3$ satisfying equations (1);
2. For each v_i, with $i > 3$,
$$\{v_{i-3}, v_i\}, \{v_{i-2}, v_i\}, \{v_{i-1}, v_i\} \in E; \tag{3}$$
3. For each v_i, with $i > 3$,
$$d_{i-3,i-2} + d_{i-2,i-1} > d_{i-3,i-1}; \tag{4}$$

find a map $x : V \to \mathbb{R}^3$ such that
$$\forall \{v_i, v_j\} \in E, \ \|x_i - x_j\| = d_{i,j}.$$

Geometrically, the requirements (3) and (4) imply that, at each iteration of the BP algorithm, we intersect three spheres centered at the positions for vertices $v_{i-3}, v_{i-2}, v_{i-1}$ with radius $d_{i-3,i}, d_{i-2,i}, d_{i-1,i}$, respectively, resulting in two possible positions for v_i, $i > 3$. Distances $d_{i-1,i}$ and $d_{i-2,i}$ are considered precise values, and known *a priori*, since they are related to bond lengths and bond angles of a protein [20]. However, distances $d_{i-3,i}$ may be obtained from NMR experiments, and instead of being represented by real numbers, they should be given as interval distances $[\underline{d}_{i-3,i}, \overline{d}_{i-3,i}]$, where $\underline{d}_{i-3,i} \leq d_{i-3,i} \leq \overline{d}_{i-3,i}$. In this situation, we have the intersection of two spheres with a spherical shell, giving two arcs, instead of two points in $\mathbb{R}^3$.

In [21], an extension of the BP algorithm was proposed to deal with interval distances, called *i*BP, where the idea is to sample values from $[\underline{d}_{i-3,i}, \overline{d}_{i-3,i}]$ [16]. Computational results presented in [10, 11, 32] confirm what it should be expected: sampling many values, the search space increases exponentially, and for small samples, a solution may not be found.

For a given vertex v_i, $i > 4$, if another distance $d_{j,i}$ ($j < i - 3$) is detected by NMR, another spherical shell must be considered. This new information can be used to reduce the size of the interval distance $[\underline{d}_{i-3,i}, \overline{d}_{i-3,i}]$, before applying the sampling process.

Computational results presented in [14, 26] confirm the improvement of *i*BP algorithm when such kind of interval reduction is implemented, before sampling values. Without interval reduction, it is necessary to select a distance value from the interval $[\underline{d}_{i-3,i}, \overline{d}_{i-3,i}]$ in order to calculate a position for vertice v_i. From positions for vertices v_{i-1}, v_{i-2}, v_i, we calculate a position for vertice v_{i+1}, making another selection from interval $[\underline{d}_{i-2,i+1}, \overline{d}_{i-2,i+1}]$, and so on. A DMDGP solution is obtained when such selections allow us to reach the last vertex of the DMDGP order such that all positions $x_1, \ldots, x_n$ satisfy the equations (1). The main cost of the *i*BP algorithm is related to backtracking in the search tree, when "wrong" distance values are selected. When interval distances are reduced, we also decrease the probability of selecting "wrong" distance values.

Using Conformal Geometric Algebra (CGA), we present a new way to make this reduction that simplifies the process considerably, compared to other approaches proposed in the literature.

2 Methods for reducing $[\underline{d}_{i-3,i}, \overline{d}_{i-3,i}]$

This section first describes a recent method proposed for reducing $[\underline{d}_{i-3,i}, \overline{d}_{i-3,i}]$ [26], which is an extension of the ideas presented in [14]. Then, we explain the new approach motivated by the results given in [6, 7].
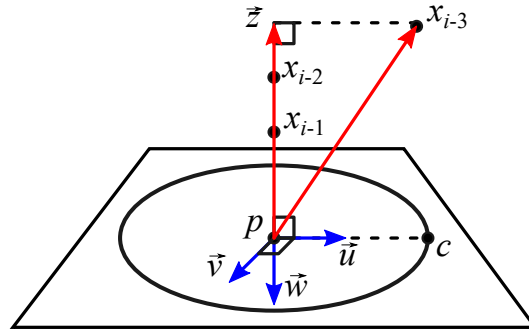
Figure 1: Parameterization of the circle with radius r.

2.1 Using polar coordinates

DMDGP definition allows us to fix positions for v_1, v_2, v_3. Requirements (3) and (4) define a system with up to two possible positions for v_i, $i > 3$, given by

$$
\begin{aligned}
||x_i - x_{i-1}|| &= d_{i-1,i}, \\
||x_i - x_{i-2}|| &= d_{i-2,i}, \\
||x_i - x_{i-3}|| &= d_{i-3,i},
\end{aligned}
\tag{5}
$$

where $x_{i-1}, x_{i-2}, x_{i-3} \in \mathbb{R}^3$ are previously calculated. When an additional edge $\{v_j, v_i\} \in E$ is given, $j < i - 3$, we have an extra equation to the system (5):

$$
||x_i - x_j|| = d_{j,i}.
\tag{6}
$$

If the points $x_j, x_{i-1}, x_{i-2}, x_{i-3} \in \mathbb{R}^3$ are not coplanar, we get a unique solution $x_i^\star$ for v_i, supposing $||x_i^\star - x_j|| = d_{j,i}$. However, both positions for v_i may not satisfy (6) and, in this case, must be pruned. Then, we have to consider the other possible positions for v_{i-1} and repeat the procedure until a DMDGP solution is found [20].

For all $i \geq 3$, the solution of the first two equations of the system (5) is a circle, as the result of intersection of two spheres, centered at $x_{i-1}, x_{i-2} \in \mathbb{R}^3$ with radii $d_{i-1,i}, d_{i-2,i} \in \mathbb{R}$, respectively.

Using a fixed point c on the circle, in the same plane defined by the points $x_{i-1}, x_{i-2}, x_{i-3}$ and nearest to x_{i-3}, and defining $\vec{w} = \frac{x_{i-1} - x_{i-2}}{||x_{i-1} - x_{i-2}||}$, $\vec{s} = proj_{\vec{w}}(c - x_{i-2})$, $p = x_{i-2} + \vec{s}$, and $\vec{z} = proj_{\vec{w}}(x_{i-3} - p)$, the circle can be described by (see Fig. 1)

$$
x(t) = p + r\cos(t)\vec{u} + r\sin(t)\vec{v},
\tag{7}
$$

for $t \in [-\pi, \pi]$, where

$$
r = ||c - p||,
$$

$$
\vec{u} = \frac{(x_{i-3} - p) - \vec{z}}{||(x_{i-3} - p) - \vec{z}||},
$$

and

$$
\vec{v} = \vec{u} \times \vec{w}.
$$

To check this, replace $x(t)$ in the first equation of (5),

$$
\begin{aligned}
||x(t) - x_{i-1}||^2 &= ||(p - x_{i-1}) + r\cos(t)\vec{u} + r\sin(t)\vec{v}||^2 \\
&= ||p - x_{i-1}||^2 + 2r(p - x_{i-1}) \cdot (\cos(t)\vec{u} + \sin(t)\vec{v}) \\
&\quad + ||r\cos(t)\vec{u} + r\sin(t)\vec{v}||^2,
\end{aligned}
$$

and note that $p - x_{i-1}$ is orthogonal to $\vec{u}$ and $\vec{v}$:

$$||x(t) - x_{i-1}||^2 = ||p - x_{i-1}||^2 + r^2 = d_{i-1,i}^2.$$

A similar argument gives

$$||x(t) - x_{i-2}||^2 = ||p - x_{i-2}||^2 + r^2 = d_{i-2,i}^2.$$

From (7) and by the fact that $p - x_{i-3}$ is orthogonal to $\vec{v}$, we obtain

$$||x(t) - x_{i-3}||^2 = d_{i-3,i}^2 \Leftrightarrow$$
$$||p - x_{i-3}||^2 + 2r\cos(t)(\vec{u} \cdot (p - x_{i-3})) + r^2 = d_{i-3,i}^2 \Leftrightarrow$$
$$\cos(t) = \frac{d_{i-3,i}^2 - r^2 - ||p - x_{i-3}||^2}{2r(\vec{u} \cdot (p - x_{i-3}))},$$

implying that the solution for

$$||x(t) - x_{i-3}|| = d_{i-3,i}$$

is given by

$$\cos(t) = \frac{d_{i-3,i}^2 - r^2 - ||p - x_{i-3}||^2}{2r\vec{u} \cdot (p - x_{i-3})}, \tag{8}$$

for $t \in [-\pi, \pi]$.

Due to NMR uncertainties, the distance $d_{i-3,i}$ can be an interval $[\underline{d}_{i-3,i}, \overline{d}_{i-3,i}]$. Assuming that $\underline{d}_{i-3,i} \geq ||x(0) - x_{i-3}||$ (the minimum distance between the circle and the point x_{i-3}) and $\overline{d}_{i-3,i} \leq ||x(\pi) - x_{i-3}||$ (the maximum distance between the circle and the point x_{i-3}), and using expression (8), we obtain that the solution for

$$\underline{d}_{i-3,i} \leq ||x(t) - x_{i-3}|| \leq \overline{d}_{i-3,i}$$

is given by

$$x(t) = p + r\cos(t)\vec{u} + r\sin(t)\vec{v}, \ t \in [-\beta, -\alpha] \cup [\alpha, \beta],$$

for $\alpha, \beta \in [0, \pi]$ such that

$$\alpha = \arccos\left(\frac{(\underline{d}_{i-3,i})^2 - r^2 - ||p - x_{i-3}||^2}{2r(\vec{u} \cdot (p - x_{i-3}))}\right),$$
$$\beta = \arccos\left(\frac{(\overline{d}_{i-3,i})^2 - r^2 - ||p - x_{i-3}||^2}{2r(\vec{u} \cdot (p - x_{i-3}))}\right).$$

From this point, *i*BP algorithm samples values from $[\underline{d}_{i-3,i}, \overline{d}_{i-3,i}]$ for solving system (5) and continues the search. However, if there is a pair of atoms $\{v_j, v_i\}$ $(j < i - 3)$ that is close enough, NMR experiments provide other interval distance $[\underline{d}_{j,i}, \overline{d}_{j,i}]$ that should be used before sampling values from $[\underline{d}_{i-3,i}, \overline{d}_{i-3,i}]$, in order to try to reduce the size of $[\underline{d}_{i-3,i}, \overline{d}_{i-3,i}]$.

Using the point x_j instead of x_{i-3}, another parameterization of the circle defined by the first two equations of the system (5) must be defined, for $t \in [-\pi, \pi]$,

$$x'(t) = p + r\cos(t)\vec{u_j} + r\sin(t)\vec{v_j},$$

with $\{\vec{u}, \vec{v}\}$ replaced by $\{\vec{u_j}, \vec{v_j}\}$, where $\vec{z_j} = proj_{\vec{w}}(x_j - p)$, $\vec{u_j} = \frac{(x_j-p)-\vec{z_j}}{||(x_j-p)-\vec{z_j}||}$, and $\vec{v_j} = \vec{u_j} \times \vec{w}$.

To describe the solution of

$$\underline{d}_{j,i} \leq ||x'(t) - x_j|| \leq \overline{d}_{j,i},$$

in terms of the first parameterization, it is necessary to obtain the coordinates of $x'(0) - p$ in terms of $\vec{u}$ and $\vec{v}$ [26], which results in (see Fig. 2)
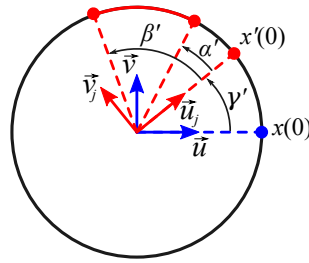
Figure 2: Two parameterizations for the circle.

$$t \in [\gamma - \beta', \gamma - \alpha'] \cup [\gamma + \alpha', \gamma + \beta'],$$

where

$$\gamma = \arctan 2 \left((x'(0) - p) \cdot \vec{v}, (x'(0) - p) \cdot \vec{u} \right),$$

$$\alpha' = \arccos \left(\frac{(\underline{d}_{j,i})^2 - r^2 - ||p - x_j||^2}{2r(\vec{u_j} \cdot (p - x_j))} \right),$$

$$\beta' = \arccos \left(\frac{(\overline{d}_{j,i})^2 - r^2 - ||p - x_j||^2}{2r(\vec{u_j} \cdot (p - x_j))} \right).$$

Finally, the solution of the system

$$||x_i - x_{i-1}|| = d_{i-1,i},$$

$$||x_i - x_{i-2}|| = d_{i-2,i},$$

$$\underline{d}_{i-3,i} \leq ||x_i - x_{i-3}|| \leq \overline{d}_{i-3,i}$$

$$\underline{d}_{j,i} \leq ||x_i - x_j|| \leq \overline{d}_{j,i},$$

for $j < i - 3$, $x_{i-1}, x_{i-2}, x_{i-3}, x_j \in \mathbb{R}^3$, and $d_{i-1,i}, d_{i-2,i}, \underline{d}_{i-3,i}, \overline{d}_{i-3,i}, \underline{d}_{j,i}, \overline{d}_{j,i} \in [0, \infty)$, is given by

$$x(t) = p + r \cos(t) \vec{u} + r \sin(t) \vec{v},$$

where

$$t \in ([-\beta, -\alpha] \cup [\alpha, \beta]) \text{ and } t \in ([\gamma - \beta', \gamma - \alpha'] \cup [\gamma + \alpha', \gamma + \beta']).$$

The next subsection describes a new model for the 3D space, where spheres are basic objects like points and planes. This model also provides a way to intersect spheres by defining a product among them.

2.2 A conformal geometric algebra approach

The Euclidean space $\mathbb{R}^3$ can be represented by the *conformal model* [27], adding two extra dimensions e_0 and e_∞, where a point $x \in \mathbb{R}^3$ is represented in $\mathbb{R}^5$ by

$$X = x + e_0 + \frac{1}{2}||x||^2 e_\infty,$$

with a new metric defined by

$$e_i \cdot e_j = \delta_{ij}, \quad i, j = 1, 2, 3,$$

$$e_0 \cdot e_i = e_\infty \cdot e_i = 0, \quad i = 1, 2, 3,$$

$$e_0 \cdot e_0 = e_\infty \cdot e_\infty = 0,$$

$$e_0 \cdot e_\infty = -1,$$

where the usual Euclidean metric still holds for e_1, e_2, e_3.

An interesting property of the conformal model is that the inner product $X \cdot Y$ ($X, Y \in \mathbb{R}^5$) is the squared Euclidean distance between $x, y \in \mathbb{R}^3$, up to a constant factor:

$$
\begin{aligned}
X \cdot Y &= \left(x + e_0 + \frac{1}{2} ||x||^2 e_\infty \right) \cdot \left(y + e_0 + \frac{1}{2} ||y||^2 e_\infty \right) \\
&= x \cdot y - \left(\frac{1}{2} ||x||^2 + \frac{1}{2} ||y||^2 \right) \\
&= -\frac{1}{2} ||x - y||^2.
\end{aligned}
$$

From this result, a sphere in $\mathbb{R}^3$ is encoded as a vector $S \in \mathbb{R}^5$ [12], given by

$$
S = C - \frac{r^2}{2} e_\infty,
$$

where C is the conformal representation of the sphere center $c \in \mathbb{R}^3$ and $r \in \mathbb{R}$ is its radius. To see this, we use

$$
X \cdot e_\infty = \left(x + e_0 + \frac{1}{2} ||x||^2 e_\infty \right) \cdot e_\infty = -1
$$

to get

$$
\begin{aligned}
X \cdot S &= X \cdot \left(C - \frac{r^2}{2} e_\infty \right) \\
&= X \cdot C - \frac{r^2}{2} (X \cdot e_\infty) \\
&= -\frac{1}{2} ||x - c||^2 + \frac{r^2}{2},
\end{aligned}
$$

which implies that

$$
X \cdot S = 0 \Leftrightarrow ||x - c||^2 = r^2.
$$

A spherical shell with center $c \in \mathbb{R}^3$ and radii $\underline{r}, \bar{r} \in \mathbb{R}$, $0 < \underline{r} < \bar{r}$, is simply given by

$$
S = C - \frac{r^2}{2} e_\infty, \ r \in [\underline{r}, \bar{r}],
$$

where $C \in \mathbb{R}^5$ is the conformal representation of $c \in \mathbb{R}^3$.

Sphere intersections can also be encoded in the conformal model if a more general product (associative and distributive), called *geometric product* [18], is introduced by

$$
\begin{aligned}
e_i e_j + e_j e_i &= 2\delta_{ij}, \\
e_0 e_i &= -e_i e_0, \\
e_\infty e_i &= -e_i e_\infty, \\
e_0^2 &= e_\infty^2 = 0, \\
e_0 e_\infty + e_\infty e_0 &= -2,
\end{aligned}
$$

for $i, j = 1, 2, 3$.

From the geometric product, the inner product defined above can be given by

$$
a \cdot b = \frac{1}{2}(ab + ba)
$$

and another product, called *outer product* [18], is defined by

$$
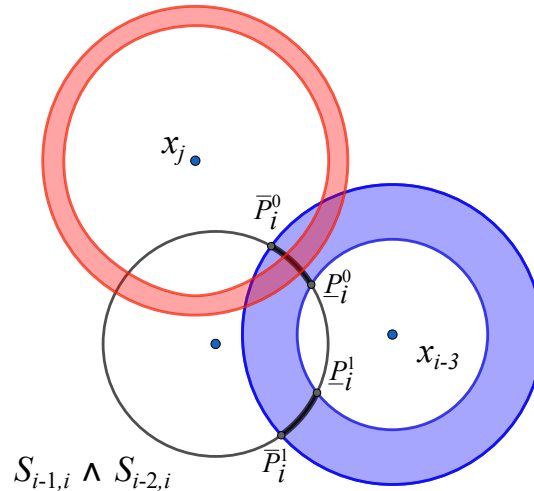a \wedge b = \frac{1}{2}(ab - ba),
$$

Figure 3: Solution of equations and inequalities.

for $a, b \in \mathbb{R}^5$.

The intersection of distinct spheres is given by the outer product of their vector representations. For example, the intersection between two spheres, given in the conformal model by $\sigma_1, \sigma_2 \in \mathbb{R}^5$, is the circle given by [12]

$$\sigma = \sigma_1 \wedge \sigma_2. \tag{9}$$

It is important to emphasize that in the general case, the result of the intersection of spheres in $\mathbb{R}^3$ using (9) may result in an element geometrically interpreted as an imaginary circle when no finite point is shared by the spheres, a tangent plane (or point circle with attitude) when the spheres share a single finite point or a real circle. But due to molecular structure restrictions, only point circles and real circles are expected as a result of the intersections.

Let us analyze geometrically the solution of the system

$$\|x_i - x_{i-1}\| = d_{i-1,i},$$

$$\|x_i - x_{i-2}\| = d_{i-2,i}, \tag{10}$$

$$\underline{d}_{i-3,i} \leq \|x_i - x_{i-3}\| \leq \overline{d}_{i-3,i}$$

$$\underline{d}_{j,i} \leq \|x_i - x_j\| \leq \overline{d}_{j,i},$$

for $j < i - 3$, $x_{i-1}, x_{i-2}, x_{i-3}, x_j \in \mathbb{R}^3$, and $d_{i-1,i}, d_{i-2,i}, \underline{d}_{i-3,i}, \overline{d}_{i-3,i}, \underline{d}_{j,i}, \overline{d}_{j,i} \in [0, \infty)$.

From Fig. 3, we see that this solution is a subset of the union of two arcs of a circle defined by

$$x(t) = p + r\cos(t)\vec{u} + r\sin(t)\vec{v},$$

where

$$t \in \left([-\beta, -\alpha] \cup [\alpha, \beta]\right) \text{ and } t \in \left([\gamma - \beta', \gamma - \alpha'] \cup [\gamma + \alpha', \gamma + \beta']\right),$$

as explained in subsection 2.1.

Let us denote by $P_i^0 \overline{P_i^0}$ and $P_i^1 \overline{P_i^1}$ the arcs obtained from the intersection of spheres $S_{i-1,i}$, $S_{i-2,i}$ with the spherical shell $S_{i-3,i}$ (see Fig. 3), given by

$$S_{i-1,i} = X_{i-1} - \frac{d_{i-1,i}^2}{2}e_\infty,$$

$$S_{i-2,i} = X_{i-2} - \frac{d_{i-2,i}^2}{2}e_\infty,$$

$$S_{i-3,i} = X_{i-3} - \frac{d_{i-3,i}^2}{2}e_\infty, \ d_{i-3,i} \in [\underline{d}_{i-3,i}, \overline{d}_{i-3,i}],$$
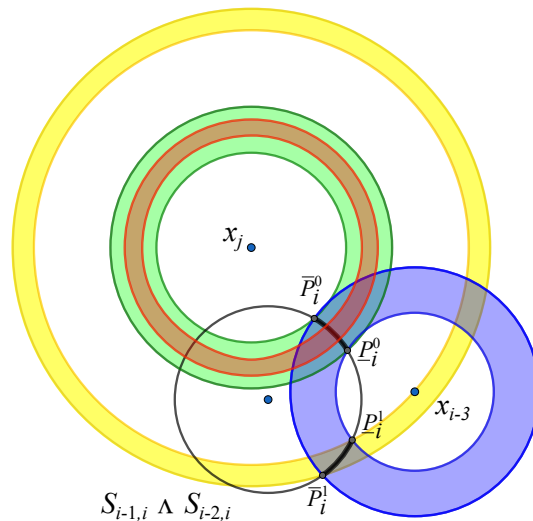
Figure 4: Auxiliary spherical shells.

where $X_{i-1}, X_{i-2}, X_{i-3}$ are the conformal representation of the points $x_{i-1}, x_{i-2}, x_{i-3} \in \mathbb{R}^3$ and $d_{i-1,i}, d_{i-2,i}, \underline{d}_{i-3,i}, \overline{d}_{i-3,i}$ are the respective radii.

Motivated by the geometry of the problem (see Fig. 4), we define two more spherical shells, with the same center x_j, and with interval radii given by the distances between x_j and $\underline{P_i^0}, \overline{P_i^0}$ and the distances between x_j and $\underline{P_i^1}, \overline{P_i^1}$:

$$S_0 = X_j - \frac{r_0^2}{2} e_\infty, \, r_0 \in [\underline{r_0}, \overline{r_0}],\tag{11}$$

with

$$
\begin{aligned}
\underline{r_0} &= \min\left\{\sqrt{-2X_j \cdot \underline{P_i^0}}, \sqrt{-2X_j \cdot \overline{P_i^0}}\right\}, \\
\overline{r_0} &= \max\left\{\sqrt{-2X_j \cdot \underline{P_i^0}}, \sqrt{-2X_j \cdot \overline{P_i^0}}\right\},
\end{aligned}
$$

and

$$S_1 = X_j - \frac{r_1^2}{2} e_\infty, \, r_1 \in [\underline{r_1}, \overline{r_1}],\tag{12}$$

with

$$
\begin{aligned}
\underline{r_1} &= \min\left\{\sqrt{-2X_j \cdot \underline{P_i^1}}, \sqrt{-2X_j \cdot \overline{P_i^1}}\right\}, \\
\overline{r_1} &= \max\left\{\sqrt{-2X_j \cdot \underline{P_i^1}}, \sqrt{-2X_j \cdot \overline{P_i^1}}\right\}.
\end{aligned}
$$

These new spherical shells are of the essence of the problem, which implies that the solution of the system (10) is given by (see Fig. 5)

$$[S_{i-1,i} \wedge S_{i-2,i} \wedge S_{j,i}^0] \cup [S_{i-1,i} \wedge S_{i-2,i} \wedge S_{j,i}^1],\tag{13}$$

with

$$S_{j,i}^0 = X_j - \frac{t_0^2}{2} e_\infty \text{ and } S_{j,i}^1 = X_j - \frac{t_1^2}{2} e_\infty,$$

$$t_0 \in [\max\{\underline{r_0}, \underline{d}_{j,i}\}, \min\{\overline{r_0}, \overline{d}_{j,i}\}] \text{ and } t_1 \in [\max\{\underline{r_1}, \underline{d}_{j,i}\}, \min\{\overline{r_1}, \overline{d}_{j,i}\}],$$

where $[\underline{d}_{j,i}, \overline{d}_{j,i}]$ is the interval radius associated to the pair $\{v_j, v_i\}, j < i - 3$.

The first part of the union in (13) is a subset of the arc $\underline{P_i^0}\overline{P_i^0}$, and the second is a subset of the arc $\underline{P_i^1}\overline{P_i^1}$ (see Fig. 5).

Figure 5: Solution of the problem using CGA.

The intervals $[\max\{\underline{r_0}, \underline{d}_{j,i}\}, \min\{\overline{r_0}, \overline{d}_{j,i}\}]$ and $[\max\{\underline{r_1}, \underline{d}_{j,i}\}, \min\{\overline{r_1}, \overline{d}_{j,i}\}]$ are obtained using the following property:

$$t \in [\underline{a}, \overline{a}] \cap [\underline{b}, \overline{b}] \Leftrightarrow \max\{\underline{a}, \underline{b}\} \le t \le \min\{\overline{a}, \overline{b}\}.$$

If one of these intervals is empty, we simply remove it from the calculation (see example in the next section).

The CGA approach has, at least, two advantages. The first one is related to the fact that the geometric interpretation of the problem that must be solved, given by

$$||x_i - x_{i-1}|| = d_{i-1,i},$$

$$||x_i - x_{i-2}|| = d_{i-2,i}, \tag{14}$$

$$\underline{d}_{i-3,i} \le ||x_i - x_{i-3}|| \le \overline{d}_{i-3,i}$$

$$\underline{d}_{j,i} \le ||x_i - x_j|| \le \overline{d}_{j,i},$$

for $j < i-3$, $x_{i-1}, x_{i-2}, x_{i-3}, x_j \in \mathbb{R}^3$, and $d_{i-1,i}, d_{i-2,i}, \underline{d}_{i-3,i}, \overline{d}_{i-3,i}, \underline{d}_{j,i}, \overline{d}_{j,i} \in [0, \infty)$, can be described in the language of the CGA, which allowed a better "view" of the problem, in addition to solve it just comparing distance values. The second advantage is based on the possibility to solve problems in higher dimensions, where sphere intersections are also involved [1].

3 Example

To illustrate the difference between the two approaches, let us consider a DMDGP instance defined by the graph $G = (V, E, d)$, given by

$$
\begin{aligned}
V &= \{v_1, v_2, v_3, v_4, v_5\}, \\
E &= \{\{v_1, v_2\}, \{v_1, v_3\}, \{v_1, v_4\}, \{v_1, v_5\}, \\
&\quad \{v_2, v_3\}, \{v_2, v_4\}, \{v_2, v_5\}, \\
&\quad \{v_3, v_4\}, \{v_3, v_5\} \\
&\quad \{v_4, v_5\}\},
\end{aligned}
$$

and

$$d_{1,2} = \sqrt{2}, \quad d_{1,3} = \sqrt{2}, \quad d_{1,4} = 1.0, \quad d_{1,5} \in [0.8, 0.9],$$
$$d_{2,3} = \sqrt{2}, \quad d_{2,4} = 1.0, \quad d_{2,5} \in [0.7, 1.0],$$
$$d_{3,4} = 1.0, \quad d_{3,5} = 1.0,$$
$$d_{4,5} = 1.0.$$

The first four vertices can be fixed at

$$x_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \; x_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \; x_3 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \; x_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Using the polar coordinates approach, we first calculate

$$c = (0.5, 0.8660254, 0.0)^T,$$

$$\vec{w} = \frac{x_4 - x_3}{||x_4 - x_3||} = (-1, 0, 0)^T,$$

$$\vec{s} = proj_{\vec{w}}(c - x_3) = (-0.5, 0, 0)^T,$$

$$p = x_3 + \vec{s} = (0.5, 0, 0)^T,$$

$$\vec{z} = proj_{\vec{w}}(x_2 - p) = (-0.5, 0, 0)^T,$$

$$r = ||c - p|| = 0.8660254,$$

$$\vec{u} = \frac{(x_2 - p) - \vec{z}}{||(x_2 - p) - \vec{z}||} = (0, 1, 0)^T,$$

$$\vec{v} = \vec{u} \times \vec{w} = (0, 0, -1)^T,$$

which implies that the circle defined by the system

$$||x_5 - x_4|| = d_{4,5}$$

$$||x_5 - x_3|| = d_{3,5} \tag{15}$$

can be described by

$$x(t) = p + r\cos(t)\vec{u} + r\sin(t)\vec{v}, \tag{16}$$

for $t \in [-\pi, \pi]$, and that the solution of

$$\underline{d}_{2,5} \leq ||x(t) - x_2|| \leq \overline{d}_{2,5}$$

is given by

$$t \in [-\beta, -\alpha] \cup [\alpha, \beta],$$

where

$$\alpha = \arccos\left(\frac{(\underline{d}_{2,5})^2 - r^2 - ||p - x_2||^2}{2r(\vec{u} \cdot (p - x_2))}\right) = 0.511934,$$

$$\beta = \arccos\left(\frac{(\overline{d}_{2,5})^2 - r^2 - ||p - x_2||^2}{2r(\vec{u} \cdot (p - x_2))}\right) = 0.955317.$$

For the interval distance $d_{1,5}$, another parameterization is obtained as

$$x'(t) = p + r\cos(t)\vec{u_j} + r\sin(t)\vec{v_j},$$

where

$$\vec{u_j} = (0, 0, 1)^T,$$
$$\vec{v_j} = (0, 1, 0)^T,$$

implying that the solution of the system

$$||x_5 - x_4|| = d_{4,5}$$

$$||x_5 - x_3|| = d_{3,5} \tag{17}$$
$$\underline{d}_{2,5} \leq ||x_5 - x_2|| \leq \overline{d}_{2,5}$$
$$\underline{d}_{1,5} \leq ||x_5 - x_1|| \leq \overline{d}_{1,5}$$

is given by

$$x(t) = p + r\cos(t)\vec{u} + r\sin(t)\vec{v},$$

for

$$t \in ([\gamma - \beta', \gamma - \alpha'] \cup [\gamma + \alpha', \gamma + \beta']) \cap ([-\beta, -\alpha] \cup [\alpha, \beta]) = [-0.903017, -0.757421],$$

where

$$\gamma = \arctan 2\left((x'(0) - p) \cdot \vec{v}, (x'(0) - p) \cdot \vec{u}\right) = -1.5708,$$

$$\alpha' = \arccos\left(\frac{(\underline{d}_{1,5})^2 - r^2 - ||p - x_1||^2}{2r(\vec{u_j} \cdot (p - x_1))}\right) = 0.667783,$$

$$\beta' = \arccos\left(\frac{(\overline{d}_{1,5})^2 - r^2 - ||p - x_1||^2}{2r(\vec{u_j} \cdot (p - x_1))}\right) = 0.813379.$$

Now, let us see how to solve the example using the CGA approach.

From the intersections of spheres $\underline{S}_{2,5} \wedge S_{3,5} \wedge S_{4,5}$ and $\overline{S}_{2,5} \wedge S_{3,5} \wedge S_{4,5}$, with radii given by $\underline{d}_{2,5}, d_{3,5}, d_{4,5}, \overline{d}_{2,5}$, respectively, we obtain the arcs $\underline{P_5^0}\,\overline{P_5^0}$ and $\underline{P_5^1}\,\overline{P_5^1}$, defined by the points

$$\underline{P_5^0} = e_0 + 0.5e_1 + 0.755e_2 - 0.4242e_3 + 0.5e_\infty,$$
$$\underline{P_5^1} = e_0 + 0.5e_1 + 0.755e_2 + 0.4242e_3 + 0.5e_\infty,$$
$$\overline{P_5^0} = e_0 + 0.5e_1 + 0.5e_2 - 0.7071e_3 + 0.5e_\infty,$$
$$\overline{P_5^1} = e_0 + 0.5e_1 + 0.5e_2 + 0.7071e_3 + 0.5e_\infty.$$

The radii of the spherical shells S_0 (11) and S_1 (12), centered at X_1, are

$$\underline{r_0} = \min\left\{\sqrt{-2X_1 \cdot \underline{P_5^0}}, \sqrt{-2X_1 \cdot \overline{P_5^0}}\right\} = \min\{1.69, 1.07\} = 1.07$$

$$\overline{r_0} = \max\left\{\sqrt{-2X_1 \cdot \underline{P_5^0}}, \sqrt{-2X_1 \cdot \overline{P_5^0}}\right\} = \max\{1.69, 1.07\} = 1.69$$

and

$$\underline{r_1} = \min\left\{\sqrt{-2X_1 \cdot \underline{P_5^1}}, \sqrt{-2X_1 \cdot \overline{P_5^1}}\right\} = \min\{1.85, 0.765\} = 0.765$$

$$\overline{r_1} = \max\left\{\sqrt{-2X_1 \cdot \underline{P_5^1}}, \sqrt{-2X_1 \cdot \overline{P_5^1}}\right\} = \max\{1.85, 0.765\} = 1.85,$$

implying that

$$t_0 \in [\max\{\underline{r_0}, \underline{d}_{1,5}\}, \min\{\overline{r_0}, \overline{d}_{1,5}\}] = \emptyset$$

and

$$t_1 \in [\max\{\underline{r_1}, \underline{d}_{1,5}\}, \min\{\overline{r_1}, \overline{d}_{1,5}\}] = [0.8, \ 0.9].$$

Hence, the solution of the system

$$\|x_5 - x_4\| = d_{4,5}$$

$$\|x_5 - x_3\| = d_{3,5} \tag{18}$$

$$\underline{d}_{2,5} \le \|x_5 - x_2\| \le \overline{d}_{2,5}$$

$$\underline{d}_{1,5} \le \|x_5 - x_1\| \le \overline{d}_{1,5}$$

is given by

$$S_{4,5} \wedge S_{3,5} \wedge S^1_{1,5},$$

where

$$S_{4,5} = X_4 - \frac{d^2_{4,5}}{2}e_\infty, \ d_{4,5} = 1,$$

$$S_{3,5} = X_3 - \frac{d^2_{3,5}}{2}e_\infty, \ d_{3,5} = 1,$$

$$S^1_{1,5} = X_1 - \frac{t^2_1}{2}e_\infty, \ t_0 \in [0.8, 0.9].$$

The superscript 1 in $S^1_{1,5}$ indicates that the solution is a subset of the arc $\overline{P^1_5 P^1_5}$.

Doing the calculations, the size of the interval associated to the distance $\overline{d_{2,5}}$ changed from $[0.7, 1.0]$ to $[0.861, 0.963]$.

All the calculations were done using GAALOP [17].

4 Conclusions

NMR experiments do not provide precise distances between atoms in a protein molecule and dealing with interval distances is a big challenge for DMDGP solution methods [30].

Based on Conformal Geometric Algebra (CGA), we present a new approach that allows us to incorporate the geometry involved when uncertainties must be taken into account, in addition to simplify the understanding of the problem.

Other results that also apply CGA in the DMDGP can be found in [2, 3, 13, 22, 24].

Conflict of interest: Authors state no conflict of interest.

References

[1] J. Alencar, C. Lavor, and L. Liberti, Realizing Euclidean distance matrices by sphere intersection, *Discrete Applied Mathematics*, 256:5-10, 2019.

[2] R. Alves and C. Lavor, Geometric algebra to model uncertainties in the discretizable molecular distance geometry problem, *Advances in Applied Clifford Algebra*, 27:439-452, 2017.

[3]   R. Alves, C. Lavor, C. Souza, and M. Souza, Clifford algebra and discretizable distance geometry, *Mathematical Methods in the Applied Sciences*, 41:3999-4346, 2018.

[4]   S. Billinge, P. Duxbury, D. Gonçalves, C. Lavor, and A. Mucherino, Assigned and unassigned distance geometry: applications to biological molecules and nanostructures, *4OR*, 14:337-376, 2016.

[5]   S. Billinge, P. Duxbury, D. Gonçalves, C. Lavor, and A. Mucherino, Recent results on assigned and unassigned distance geometry with applications to protein molecules and nanostructures, *Annals of Operations Research*, 271:161–203, 2018.

[6]   M. Carielo and L. Fernandes, Using conformal geometric algebra to determine the structure of proteins with uncertainties, Early Proceedings of the AGACSE 2018, 197-198, 2018.

[7]   M. Carielo, Interseção de Esferas e Álgebra Geométrica, PhD thesis, University of Campinas, 2019.

[8]   R. Carvalho, C. Lavor, and F. Protti, Extending the geometric build-up algorithm for the molecular distance geometry problem, *Information Processing Letters*, 108 (2008), 234-237.

[9]   A. Cassioli, O. Gunluk, C. Lavor, and L. Liberti, Discretization vertex orders in distance geometry, *Discrete Applied Mathematics*, 197:27-41, 2015.

[10] A. Cassioli, B. Bordeaux G. Bouvier, A. Mucherino, R. Alves, L. Liberti, M. Nilges, C. Lavor, and T. Malliavin, An algorithm to enumerate all possible protein conformations verifying a set of distance constraints, *BMC Bioinformatics*, 16:16–23, 2015.

[11] C. Dambrosio, V. Ky, C. Lavor, L. Liberti, N. Maculan, New error measures and methods for realizing protein graphs from distance data, *Discrete and Computational Geometry*, 57:371-418, 2017.

[12] L. Dorst, D. Fontijne, and S. Mann, Geometric Algebra for Computer Science: An Object-Oriented Approach to Geometry, Morgan Kaufman, 2007.

[13] L. Dorst, Boolean combination of circular arcs using orthogonal spheres, *Advances in Applied Clifford Algebra*, 29:1-21, 2019.

[14] D. Gonçalves, A. Mucherino, and C. Lavor, An adaptive branching scheme for the branch & prune algorithm applied to distance geometry, Proceedings of the 2014 Federated Conference on Computer Science and Information Systems, 457-463, 2014.

[15] D. Gonçalves and A. Mucherino, Discretization orders and efficient computation of Cartesian coordinates for distance geometry, *Optimization Letters*, 8:2111-2125, 2014.

[16] D. Gonçalves, A. Mucherino, C. Lavor, and L. Liberti, Recent advances on the interval distance geometry problem, *Journal of Global Optimization*, 69:525-545, 2017.

[17] D. Hildenbrand, *Foundations of Geometric Algebra Computing*, Springer, Berlin, 2012.

[18] K. Kanatani, Understanding Geometric Algebra: Hamilton, Grassmann, and Clifford for computer vision and graphics, CRC Press, 2015.

[19] C. Lavor, L. Liberti, N. Maculan and A. Mucherino, Recent advances on the discretizable molecular distance geometry problem, *European Journal of Operational Research*, 219:698-706, 2012.

[20] C. Lavor, L. Liberti, N. Maculan and A. Mucherino, The discretizable molecular distance geometry problem, *Computational Optimization and Applications*, 52:115-146, 2012.

[21] C. Lavor, L. Liberti, and A. Mucherino, The interval BP algorithm for the discretizable molecular distance geometry problem with interval data, *Journal of Global Optimization*, 56:855-871, 2013.

[22] C. Lavor, R. Alves, W. Figueiredo, A. Petraglia, and N. Maculan, Clifford algebra and the discretizable molecular distance geometry problem, *Advances in Applied Clifford Algebra*, 25:925-942, 2015.

[23] C. Lavor, L. Liberti, B. Donald, B. Worley, B. Bardiaux, T. Malliavin, and M. Nilges, Minimal NMR distance information for rigidity of protein graphs, *Discrete Applied Mathematics*, 256:91-104, 2019.

[24] C. Lavor and R. Alves, Oriented conformal geometric algebra and the molecular distance geometry problem, *Advances in Applied Clifford Algebra*, 29:1-19, 2019.

[25] C. Lavor, M. Souza, L. Mariano, and L. Liberti, On the polinomiality of finding ^{K}DMDGP re-orders, *Discrete Applied Mathematics*, 267:190-194, 2019.

[26] C. Lavor, M. Souza, L. Carvalho, D. Gonçalves, and A. Mucherino, Improving the sampling process in the interval branch-and-prune algorithm for the discretizable molecular distance geometry, *Applied Mathematics and Computation*, accepted.

[27] H. Li, D. Hestenes, and A. Rockwood, Generalized homogeneous coordinates for computational geometry, in Geometric Computing with Clifford Algebra, G. Sommer (ed.), Springer, pp. 25-58, 2001.

[28] L. Liberti, C. Lavor, N. Maculan, A branch-and-prune algorithm for the molecular distance geometry problem, International Transactions in Operational Research, 15:1-17, 2008.

[29] L. Liberti, C. Lavor, N. Maculan, and A. Mucherino, Euclidean distance geometry and applications, *SIAM Review*, 56:3–69, 2014.

[30] L. Liberti and C. Lavor, Open research areas in distance geometry, in Open Problems in Optimization and Data Analysis, A. Migalas, P. Pardalos (eds.), Springer, 183-223, 2018.

[31] A. Mucherino, C. Lavor, L. Liberti, and N. Maculan, eds., Distance Geometry: Theory, Methods, and Applications, Springer, New York (2013).

[32] B. Worley, F. Delhommel, F. Cordier, T. Malliavin, B. Bardiaux, N. Wolff, M. Nilges, C. Lavor, L. Liberti, Tuning interval branch-and-prune for protein structure determination, *Journal of Global Optimization*, 72:109-127, 2018.