Adv. Appl. Clifford Algebras 27 (2017), 439–452

0188-7009/010439-14
published online March 21, 2016
DOI 10.1007/s00006-016-0653-2

Advances in Applied Clifford Algebras

Geometric Algebra to Model Uncertainties in the Discretizable Molecular Distance Geometry Problem

Rafael Alves* and Carlile Lavor

Abstract. The discretizable molecular distance geometry problem (DMDGP) is related to the determination of 3D protein structure using distance information detected by nuclear magnetic resonance (NMR) experiments. The chemistry of proteins and the NMR distance information allow us to define an atomic order $v_1, \ldots, v_n$ such that the distances related to the pairs $\{v_{i-3}, v_i\}, \{v_{i-2}, v_i\}, \{v_{i-1}, v_i\}$, for $i > 3$, are available, which implies that the search space can be represented by a tree. A DMDGP solution can be represented by a path from the root to a leaf node of this tree, found by an exact method, called branch-and-prune (BP). Because of uncertainty in NMR data, some of the distances related to the pairs $\{v_{i-3}, v_i\}$ may not be precise values, being represented by intervals of real numbers $[\underline{d}_{i-3,i}, \overline{d}_{i-3,i}]$. In order to apply BP algorithm in this context, sample values from those intervals should be taken. The main problem of this approach is that if we sample many values, the search space increases drastically, and for small samples, no solution can be found. We explain how geometric algebra can be used to model uncertainties in the DMDGP, avoiding sample values from intervals $[\underline{d}_{i-3,i}, \overline{d}_{i-3,i}]$ and eliminating the heuristic characteristics of BP when dealing with interval distances.

Keywords. Conformal geometric algebra, Distance geometry, Branch and prune algorithm, 3D protein structure.

1. Distance Geometry and 3D Protein Structure

One of the important problems in computational biology is the calculation of the three-dimensional structure of a protein. Nuclear magnetic resonance (NMR) experiments can provide distances between pairs of atoms that are close enough and the problem is how to determine the 3D protein structure based on this partial distance information [4,7,23].

*Corresponding author.

Using a graph $G = (V, E, d)$, where V represents the set of atoms and E is the set of atom pairs for which a distance is available, defined by the function $d : E \to (0, \infty)$, the problem can be solved by finding a function $x : V \to \mathbb{R}^3$ that associates elements of V with points in $\mathbb{R}^3$ in such a way that the Euclidean distances between the points correspond to the values given by d. This is a *distance geometry problem* (DGP) in $\mathbb{R}^3$, formally defined as follows (recent literature about DGP can be found in [17–19]):

Definition 1. Given a simple undirected graph $G = (V, E, d)$ whose edges are weighted by $d : E \to (0, \infty)$, find a function $x \colon V \to \mathbb{R}^3$ such that

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

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

The information provided by NMR experiments and chemistry of proteins allow us to define a vertex orders $v_1, \ldots, v_n \in V$ such that (in [3], it is presented an analysis on computational complexity vertex orders in distance geometry):

- For the first three vertices, there exist $x_1, x_2, x_3 \in \mathbb{R}^3$ satisfying equations (1);
- Each vertex with rank greater than 3 is adjacent to three contiguous predecessors, i.e.

$$\forall i > 3, \{\{v_{i-3}, v_i\}, \{v_{i-2}, v_i\}, \{v_{i-1}, v_i\}\} \subset E.$$

The class of DGP instances possessing these orders is called the *discretizable molecular distance geometry problem* (DMDGP) [12,13], for which there is an exact method, called branch-and-prune (BP), for finding all solutions up to rotations and translations [16] (in order to guarantee a finite number of solutions, the strict triangular inequalities related to the three adjacent predecessors of v_i must be satisfied).

Because of uncertainty in NMR data [1,2,22], some of the distances related to the pairs $\{v_{i-3}, v_i\}$ may not be precise values. In [14], it is proposed an extension of BP algorithm, the interval BP (iBP), to manage the uncertainty in distance information, where the idea is to sample values from the intervals related to the pairs $\{v_{i-3}, v_i\}$. The main problem of this approach is that if we sample many values, the search space increases drastically, and for small samples, no solution can be found.

In [15], using geometric algebra, it was presented an analytical expression for the position of atom i in terms of the positions of the three previous ones and the corresponding distances, where $d_{i-3,i}$ is represented by an interval of real numbers $[\underline{d}_{i-3,i}, \overline{d}_{i-3,i}]$, which implies that the related positions for atom x_i are represented by an arc of a circle, instead of a point (see Sect. 3). This expression can be useful in iBP algorithm, as illustrated in [15], but it is assumed that $x_{i-1}, x_{i-2}, x_{i-3}$ are fixed, as a consequence of the sampling process.

This paper explains how conformal geometric algebra (CGA) can be used to model uncertainties in the DMDGP (related to the intervals

$[\underline{d}_{i-3,i}, \overline{d}_{i-3,i}]$), avoiding sample values from intervals $[\underline{d}_{i-3,i}, \overline{d}_{i-3,i}]$ and eliminating the heuristic characteristics of iBP.

In Sect. 2, we describe the classical approach used for solving the DMDGP. Section 3 presents the original contribution of this paper, explaining how CGA can model uncertainties in the DMDGP. Some conclusions and new research directions are given in Sect. 4.

2. Quadratic System and Matrix Approach

At each step of BP algorithm applied to a DMDGP instance, taking the Cartesian coordinates of the last three vertices (previously calculated), the position for vertex v_i, $i > 3$, is obtained by solving the quadratic system

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

which can result in up to two possible values for x_i, with probability one [18]. This recursive procedure defines a binary tree containing all possible positions for each vertex v_i on the respective layer. Each DMDGP solution can be represented as a path from the root to a leaf node of the tree. When there are other adjacent predecessors, one or both possible positions for v_i may be infeasible with respect to those additional distances. If both are infeasible, it is necessary to backtrack and repeat the procedure [13].

Using DMDGP orders, we can replace resolutions of quadratic systems by matrix multiplications (see below). Computational results presented in [13] demonstrate that the second approach guarantees more stability in BP algoritm.

A chain of n atoms of a molecule indexed by $1, \ldots, n$ can be described by *internal coordinates* [21], given by the *bond lengths* $d_{i-1,i}$ (the Euclidean distance between x_{i-1} and x_i), for $i = 2, \ldots, n$, the *bond angles* $\theta_{i-2,i}$ (the angle defined by the atoms $i-2, i-1, i$), for $i = 3, \ldots, n$, and the *torsion angles* $\omega_{i-3,i}$ (the angle between the normals through the planes defined by the atoms $i-3, i-2, i-1$ and $i-2, i-1, i$), for $i = 4, \ldots, n$ (see Fig. 1). Due to the properties of DMDGP orders, the values $d_{i-1,i}$, $\theta_{i-2,i}$, $\cos(\omega_{i-3,i})$ can be calculated using the distances between the atoms $i-3, i-2, i-1, i$, for $i = 4, \ldots, n$ [13]. Based on the previous determined positions for $x_1, x_2, x_3, \ldots, x_{i-1}$, BP algorithm can obtain the two possible values for $x_i = (x_{i_1}, x_{i_2}, x_{i_3})^T \in \mathbb{R}^3$ (related to the two values for $\sin(\omega_{i-3,i}) = \pm\sqrt{1 - \cos^2(\omega_{i-3,i})}$), using the following matrix multiplications [13]:

$$\begin{bmatrix} x_{i_1} \\ x_{i_2} \\ x_{i_3} \\ 1 \end{bmatrix} = B_1 B_2 \cdots B_i \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \quad \forall i = 1, \ldots, n,$$

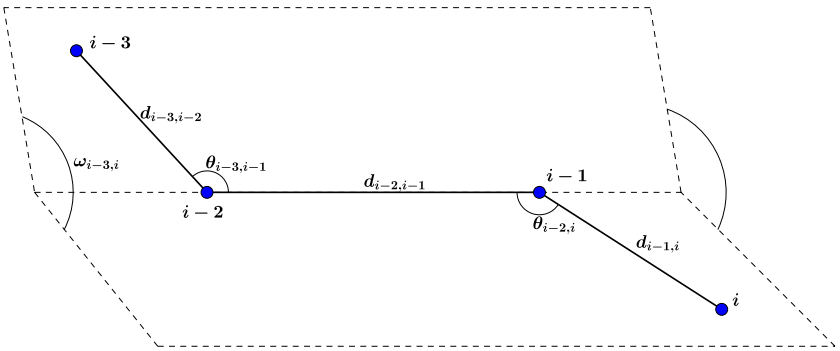

FIGURE 1. The internal coordinates of the atom i

where

$$B_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad B_2 = \begin{bmatrix} -1 & 0 & 0 & -d_{1,2} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$B_3 = \begin{bmatrix} -\cos\theta_{1,3} & -\sin\theta_{1,3} & 0 & -d_{2,3}\cos\theta_{1,3} \\ \sin\theta_{1,3} & -\cos\theta_{1,3} & 0 & d_{2,3}\sin\theta_{1,3} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

and

$$B_i = \begin{bmatrix} -\cos\theta_{i-2,i} & -\sin\theta_{i-2,i} & 0 & -d_{i-1,i}\cos\theta_{i-2,i} \\ \sin\theta_{i-2,i}\cos\omega_{i-3,i} & -\cos\theta_{i-2,i}\cos\omega_{i-3,i} & -\sin\omega_{i-3,i} & d_{i-1,i}\sin\theta_{i-2,i}\cos\omega_{i-3,i} \\ \sin\theta_{i-2,i}\sin\omega_{i-3,i} & -\cos\theta_{i-2,i}\sin\omega_{i-3,i} & \cos\omega_{i-3,i} & d_{i-1,i}\sin\theta_{i-2,i}\sin\omega_{i-3,i} \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

for $i = 4, \ldots, n$.

Using the above matrices, the first three atoms of the molecule can be fixed at positions

$$x_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \; x_2 = \begin{bmatrix} -d_{1,2} \\ 0 \\ 0 \end{bmatrix}, \; x_3 = \begin{bmatrix} -d_{1,2} + d_{2,3}\cos\theta_{1,3} \\ d_{2,3}\sin\theta_{1,3} \\ 0 \end{bmatrix}.$$

The distances $d_{i-1,i}$ and $d_{i-2,i}$ are related to the chemistry of proteins, considered as precise values, and the distances $d_{i-3,i}$, in general, are provided by NMR experiments. Because of uncertainty in NMR data, some of the distances $d_{i-3,i}$ may not be precise, being represented by intervals of real numbers $[\underline{d}_{i-3,i}, \overline{d}_{i-3,i}]$.

None of those two approaches (quadratic systems or matrix multiplications) deal well with interval distances. In the matrix approach, the uncertainty in the value $d_{i-3,i}$ is related to the value $\cos\omega_{i-3,i}$ through the expression [15] (for $i = 4, ..., n$)

$$\cos\omega_{i-3,i} = \frac{2d^2_{i-2,i-1}(d^2_{i-3,i-2} + d^2_{i-2,i} - d^2_{i-3,i}) - (d_{i-3,i-2,i-1})(d_{i-2,i-1,i})}{\sqrt{4d^2_{i-3,i-2}d^2_{i-2,i-1} - (d^2_{i-3,i-2,i-1})}\sqrt{4d^2_{i-2,i-1}d^2_{i-2,i} - (d^2_{i-2,i-1,i})}},$$

where

$$d_{i-3,i-2,i-1} = d^2_{i-3,i-2} + d^2_{i-2,i-1} - d^2_{i-3,i-1},$$
$$d_{i-2,i-1,i} = d^2_{i-2,i-1} + d^2_{i-2,i} - d^2_{i-1,i}.$$

3. Conformal Geometric Algebra and Sphere Intersection

Geometrically, the solution of the system (2) is given by the intersection of three spheres. However, when the distance $d_{i-3,i}$ is represented by an interval, we have the intersection of two spheres with one spherical shell, giving two arcs, instead of two points (Fig. 2).

Using CGA, a null basis $\{e_0, e_\infty\}$ is added to the canonical basis in $\mathbb{R}^3$, $\{e_1, e_2, e_3\}$, and spheres can be represented by vectors in a five dimensional space: the conformal space [5,6,8,9,11,20]. The set $\{e_0, e_\infty\}$ is called a null basis because both vectors square to zero with respect to the geometric product. In the conformal space, e_0 represents its origin and e_∞ represents a point at infinity.

Considering $x_i \in \mathbb{R}^3$ and $r_i \in \mathbb{R}$ be the center and radius of a sphere in $\mathbb{R}^3$, respectively, Table 1 describes how to represent points, spheres, circles and point pairs in the conformal space, where "$\wedge$" indicates the outer product and the juxtaposition of vectors indicates the geometric product. Note that a circle is obtained from the intersection between two spheres and a point pair is the result of the intersection among three spheres.

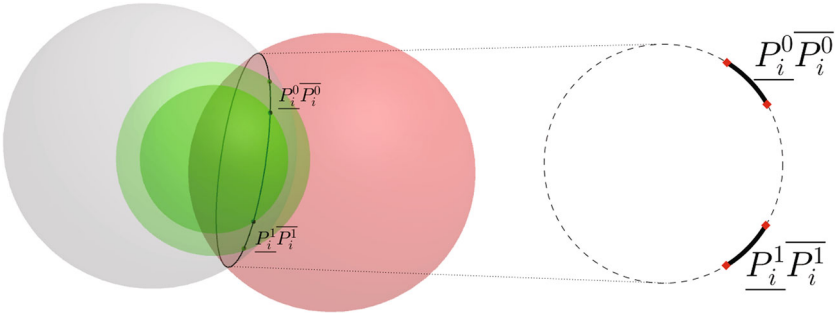

FIGURE 2. $\underline{P_i^0}\overline{P_i^0}$ and $\underline{P_i^1}\overline{P_i^1}$ are the arcs from the intersection of spheres

TABLE 1. Geometric elements represented in the conformal space

Element	Expression
Point	$X_i = x_i + \frac{1}{2}\|x_i\|^2 e_\infty + e_0$
Sphere	$S_i = X_i - \frac{1}{2} r_i^2 e_\infty$
Circle	$S_i \wedge S_j$
Point pair	$S_i \wedge S_j \wedge S_k$

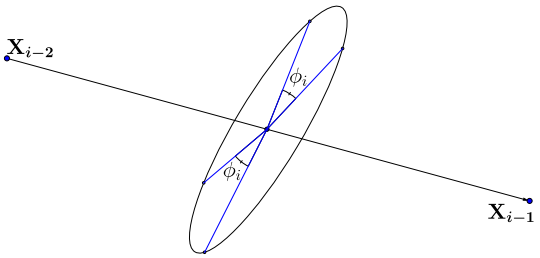

FIGURE 3. The rotation axis for the rotor R_i

From the hypothesis of the DMDGP, for $i = 4, \ldots, n$, we know two exact distances ($d_{i-2,i}$ and $d_{i-1,i}$) and one interval distance ($d_{i-3,i}$). Given points $x_{i-3}, x_{i-2}, x_{i-1}$ and distances $d_{i-3,i}, d_{i-2,i}, d_{i-1,i}$, where $d_{i-3,i} \in [\underline{d}_{i-3,i}, \overline{d}_{i-3,i}]$, we use Table 1 to calculate the point pairs

$$\underline{Pp}_i = \underline{S}_{i-3} \wedge S_{i-2} \wedge S_{i-1}, \tag{3}$$

$$\overline{Pp}_i = \overline{S}_{i-3} \wedge S_{i-2} \wedge S_{i-1}, \tag{4}$$

where underline and overline indicate the use of $\underline{d}_{i-3,i}$ and $\overline{d}_{i-3,i}$, respectively. Each point pair provides two extreme points, $\underline{P_i^0}, \underline{P_i^1}$ and $\overline{P_i^0}, \overline{P_i^1}$, one for each arc, $\underline{P_i^0}\overline{P_i^0}$ and $\underline{P_i^1}\overline{P_i^1}$ (Fig. 2).

Once we have the starting and the ending point of an arc, we can define a rotor acting on that. In CGA, a rotor is defined by its rotation axis (or rotation plane) and rotation angle. Here, the rotation axis z_i is given by the centers of the spheres S_{i-2}, S_{i-1} (Fig. 3) and the rotation angle ϕ_i (in radians) is the angle corresponding to the arcs $\underline{P_i^0}\overline{P_i^0}$ and $\underline{P_i^1}\overline{P_i^1}$ (Fig. 2).

Defining the rotor R_i by

$$R_i = \cos\left(\frac{\lambda_i}{2}\right) + z_i^* \sin\left(\frac{\lambda_i}{2}\right), \quad 0 \leq \lambda_i \leq \phi_i,$$

where $z_i = X_{i-2} \wedge X_{i-1} \wedge e_\infty$ and z_i^* is the dual of z_i, we obtain

$$X_i^0(\lambda_i) = R_i \underline{P_i^0} \tilde{R}_i \text{ and } X_i^1(\lambda_i) = R_i \underline{P_i^1} \tilde{R}_i,$$

where $\tilde{R}$ is the reverse of R.

Using $X_i^0(\lambda_i)$ or $X_i^1(\lambda_i)$, we can describe the arc points obtained by the intersection of two spheres with a spherical shell. We remark that X_{i-2} and X_{i-1} do not need to be necessarily fixed points, as explained in the following.

3.1. An Example

Let us consider the same example (with seven atoms) presented in [15], where $d_{i-1,i} = 1$, for $i = 2, 3, 5, 6, 7$, $d_{3,4} = 2.3452$, $\theta_{i-2,i} = 120°$, for $i = 1, 4, 5, 6, 7$, $\theta_{2,4} = 77.69°$ and $\theta_{3,5} = 62.94°$. The distances $d_{i-3,i}$, for $i = 4, \ldots, 7$, are given by interval distances [see matrix D (8)]. Remember that the first three atoms can be fixed in $\mathbb{R}^3$ and the search begins at the fourth level of the BP tree (x^T is the transpose of x):

$$x_1 = (0, 0, 0)^T, \tag{5}$$

$$x_2 = (-1, 0, 0)^T, \tag{6}$$

$$x_3 = (-1.5, 0.866025, 0)^T. \tag{7}$$

Remark in the example from [15], the correct values for $d_{3,4}$, $\theta_{2,4}$ and $\theta_{3,5}$ are those shown above (all the results are correct, since it were used those values). The calculations in the conformal space were done using the software Gaalop [10].

The matrix below gives all the known distances for our example (we have precise and interval distances):

$$D = \begin{bmatrix} 0 & 1 & 1.73205 & [1.75, 2] & * & * & * \\ 1 & 0 & 1 & 2.3452 & [2.3, 2.5] & * & * \\ 1.73205 & 1 & 0 & 2.3452 & 2.09165 & [1.9, 2.3] & * \\ [1.75, 2] & 2.3452 & 2.3452 & 0 & 1 & 1.73205 & [2.2, 2.5] \\ * & [2.3, 2.5] & 2.09165 & 1 & 0 & 1 & 1.73205 \\ * & * & [1.9, 2.3] & 1.73205 & 1 & 0 & 1 \\ * & * & * & [2.2, 2.5] & 1.73205 & 1 & 0 \end{bmatrix}. \tag{8}$$

Atom x_4

The pair $\{x_2, d_{2,4}\}$ defines the sphere $S_{2,4}$ (center at x_2 with radius $d_{2,4}$) and $\{x_3, d_{3,4}\}$ defines the sphere $S_{3,4}$ (center at x_3 with radius $d_{3,4}$). The intersection $S_{2,4} \wedge S_{3,4}$ gives the circle C_4, where x_4 lies. For each extreme of the interval distance $d_{1,4} \in [1.75, 2.2]$, we have two pairs involving x_1, giving spheres $\underline{S}_{1,4}$ and $\overline{S}_{1,4}$. The respective point pairs are determined by

$$\underline{Pp}_4 = \underline{S}_{1,4} \wedge S_{2,4} \wedge S_{3,4}, \tag{9}$$

$$\overline{Pp}_4 = \overline{S}_{1,4} \wedge S_{2,4} \wedge S_{3,4}, \tag{10}$$

implying that

$$\underline{Pp}_4 = 1.33e_{12} \wedge e_\infty - 0.866e_{12} \wedge e_0 - 1.36e_1 \wedge E + 0.622e_2 \wedge E, \tag{11}$$

$$\overline{Pp}_4 = 1.73e_{12} \wedge e_\infty - 0.866e_{12} \wedge e_0 - 1.12e_1 \wedge E + 0.216e_2 \wedge E, \tag{12}$$

where $e_{12} = e_1 e_2$ and $E = e_\infty \wedge e_0$.

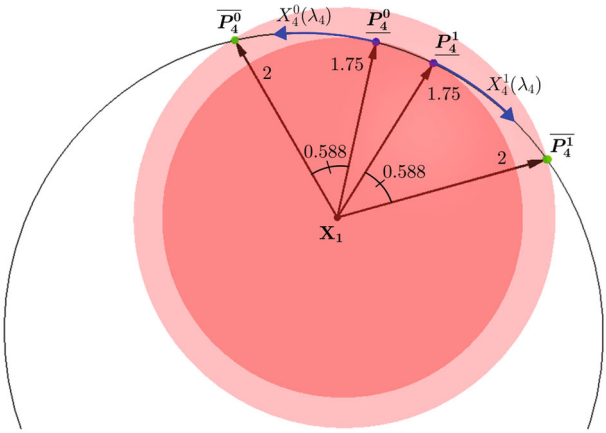

FIGURE 4. The four points on C_4 creating the two arcs for X_4

Using the Formula (13), we extract the points $\underline{P_4^0}$ and $\underline{P_4^1}$, from $\underline{Pp_4}$, and $\overline{P_4^0}$ and $\overline{P_4^1}$, from $\overline{Pp_4}$:

$$P_i^0 = \frac{Pp_i^* - \sqrt{(Pp_i^*)^2}}{-e_\infty \rfloor Pp_i^*} \text{ and } P_i^1 = \frac{Pp_i^* + \sqrt{(Pp_i^*)^2}}{-e_\infty \rfloor Pp_i^*}; \tag{13}$$

$$\underline{P_4^0} = 0.719e_1 + 1.57e_2 - 0.287e_3 + 1.53e_\infty + 1e_0, \tag{14}$$

$$\underline{P_4^1} = 0.719e_1 + 1.57e_2 + 0.287e_3 + 1.53e_\infty + 1e_0, \tag{15}$$

$$\overline{P_4^0} = 0.25e_1 + 1.3e_2 - 1.5e_3 + 2e_\infty + 1e_0, \tag{16}$$

$$\overline{P_4^1} = 0.25e_1 + 1.3e_2 + 1.5e_3 + 2e_\infty + 1e_0, \tag{17}$$

and calculate the angle ϕ_4 corresponding to the arcs $\underline{P_4^0}\overline{P_4^0}$ and $\underline{P_4^1}\overline{P_4^1}$:

$$\phi_4 = 0.588.$$

In (13), the symbol $\rfloor$ represents a left contraction, which is an extension of the inner product for vectors. More details on the contractions and on this formula are found in [5].

The points X_2 and X_3 define the rotation axis for the rotor R_4 , giving by

$$R_4 = \cos(\tfrac{\lambda_4}{2}) + z_4^* \sin(\tfrac{\lambda_4}{2}), \quad 0 \leq \lambda_4 \leq 0.588, \tag{18}$$

where $z_4 = X_2 \wedge X_3 \wedge e_\infty$, and the two possible arcs are the following (Fig. 4):

$$X_4^0(\lambda_4) = R_4 \underline{P_4^0} \tilde{R}_4 \text{ and } X_4^1(\lambda_4) = \tilde{R}_4 \underline{P_4^1} R_4. \tag{19}$$

The position $x_4 = (0.625, 1.51554, -0.75)^T$, from the example given in [15], is obtained by $X_4^0(0.208)$. However, we can continue the search without arc sampling, considering X_4^0 as a function of $\lambda_4 \in [0, 0.588]$.

The rotor R_4, given as a sum of a scalar and a bivector in (18), can be rewritten as

$$R_4 = a_0 + a_1 e_{12} + a_2 e_{13} + a_3 e_{23} + a_4 e_1 \wedge e_\infty + a_5 e_2 \wedge e_\infty + a_6 e_3 \wedge e_\infty,$$

TABLE 2. Table of multiplication ($X\tilde{B}$)

$\boldsymbol{e}_1$	$\boldsymbol{e}_2$	$\boldsymbol{e}_3$	$\boldsymbol{e}_\infty$	$\boldsymbol{e}_0$	
$-e_2$	e_1	$-e_{123}$	$-e_{12}\wedge e_\infty$	$-e_{12}\wedge e_0$	$\boldsymbol{e}_{21}$
$-e_3$	e_{123}	e_1	$-e_{13}\wedge e_\infty$	$-e_{13}\wedge e_0$	$\boldsymbol{e}_{31}$
$-e_{123}$	$-e_3$	e_2	$-e_{23}\wedge e_\infty$	$-e_{23}\wedge e_0$	$\boldsymbol{e}_{32}$
$-e_\infty$	$e_{12}\wedge e_\infty$	$e_{13}\wedge e_\infty$	0	$-e_1-e_1\wedge E$	$\boldsymbol{e}_\infty\wedge\boldsymbol{e}_1$
$-e_{12}\wedge e_\infty$	$-e_\infty$	$e_{23}\wedge e_\infty$	0	$-e_2-e_2\wedge E$	$\boldsymbol{e}_\infty\wedge\boldsymbol{e}_2$
$-e_{13}\wedge e_\infty$	$-e_{23}\wedge e_\infty$	$-e_\infty$	0	$-e_3-e_3\wedge E$	$\boldsymbol{e}_\infty\wedge\boldsymbol{e}_3$

TABLE 3. Table of multiplication (BX)

	$\boldsymbol{e}_1$	$\boldsymbol{e}_2$	$\boldsymbol{e}_3$	$\boldsymbol{e}_\infty$	$\boldsymbol{e}_0$
$\boldsymbol{e}_{12}$	$-e_2$	e_1	e_{123}	$e_{12}\wedge e_\infty$	$e_{12}\wedge e_0$
$\boldsymbol{e}_{13}$	$-e_3$	$-e_{123}$	e_1	$e_{13}\wedge e_\infty$	$e_{13}\wedge e_0$
$\boldsymbol{e}_{23}$	e_{123}	$-e_3$	e_2	$e_{23}\wedge e_\infty$	$e_{23}\wedge e_0$
$\boldsymbol{e}_1\wedge\boldsymbol{e}_\infty$	$-e_\infty$	$-e_{12}\wedge e_\infty$	$-e_{13}\wedge e_\infty$	0	$-e_1+e_1\wedge E$
$\boldsymbol{e}_2\wedge\boldsymbol{e}_\infty$	$e_{12}\wedge e_\infty$	$-e_\infty$	$-e_{23}\wedge e_\infty$	0	$-e_2+e_2\wedge E$
$\boldsymbol{e}_3\wedge\boldsymbol{e}_\infty$	$e_{13}\wedge e_\infty$	$e_{23}\wedge e_\infty$	$-e_\infty$	0	$-e_3+e_3\wedge E$

TABLE 4. Table of multiplication ($BX\tilde{B}$)

	$\boldsymbol{e}_1$	$\boldsymbol{e}_2$	$\boldsymbol{e}_3$	$\boldsymbol{e}_\infty$	$\boldsymbol{e}_0$	
$\boldsymbol{e}_{12}$	$-e_1$	$-e_2$	e_3	e_∞	e_0	$\boldsymbol{e}_{21}$
$\boldsymbol{e}_{13}$	$-e_1$	e_2	$-e_3$	e_∞	e_0	$\boldsymbol{e}_{31}$
$\boldsymbol{e}_{23}$	e_1	$-e_2$	$-e_3$	e_∞	e_0	$\boldsymbol{e}_{32}$
$\boldsymbol{e}_1\wedge\boldsymbol{e}_\infty$	0	0	0	0	$2e_\infty$	$\boldsymbol{e}_\infty\wedge\boldsymbol{e}_1$
$\boldsymbol{e}_2\wedge\boldsymbol{e}_\infty$	0	0	0	0	$2e_\infty$	$\boldsymbol{e}_\infty\wedge\boldsymbol{e}_2$
$\boldsymbol{e}_3\wedge\boldsymbol{e}_\infty$	0	0	0	0	$2e_\infty$	$\boldsymbol{e}_\infty\wedge\boldsymbol{e}_3$

where $a_i\in\mathbb{R}$, $i=0,1,...,6$. Considering $R=b+B$ (b is a scalar and B is a bivector), the rotation $X^{'}=RX\tilde{R}$ can be written as

$$X^{'}=bXb+bX\tilde{B}+BXb+BX\tilde{B},$$

where Tables 2, 3 and 4 show the multiplication rules for basis elements in $bX\tilde{B}, BXb, BX\tilde{B}$, respectively (the first term in $X^{'}$ is just a scalar-vector product b^2X).

From Tables 2, 3 and 4, and from (18) and (19), we obtain

$$R_4=\cos\left(\tfrac{\lambda_4}{2}\right)+\sin\left(\tfrac{\lambda_4}{2}\right)(0.866e_{13}+0.5e_{23}+0.866e_3\wedge e_\infty) \tag{20}$$

and

$$\begin{aligned}X_4^0(\lambda_4)=&(0.719c^2-0.496cs-3.22s^2)e_1+(1.57c^2-0.286cs-0.703s^2)e_2\\&+(-0.286c^2-4.55cs+0.286s^2)e_3+(1.53c^2+0.496cs+5.47s^2)e_\infty\\&+(c^2+s^2)e_0,\end{aligned}$$

where $c = \cos(\frac{\lambda_4}{2})$ and $s = \sin(\frac{\lambda_4}{2})$. That is,

$$x_4(\lambda_4) = \begin{bmatrix} 0.719\cos^2(\frac{\lambda_4}{2}) - 0.496\cos(\frac{\lambda_4}{2})\sin(\frac{\lambda_4}{2}) - 3.22\sin^2(\frac{\lambda_4}{2}) \\ 1.57\cos^2(\frac{\lambda_4}{2}) - 0.286\cos(\frac{\lambda_4}{2})\sin(\frac{\lambda_4}{2}) - 0.703\sin^2(\frac{\lambda_4}{2}) \\ -0.286\cos^2(\frac{\lambda_4}{2}) - 4.55\cos(\frac{\lambda_4}{2})\sin(\frac{\lambda_4}{2}) + 0.286\sin^2(\frac{\lambda_4}{2}) \end{bmatrix},$$

for $\lambda_4 \in [0, 0.588]$.

Atoms x_5 and x_6

In order to determine x_5, we have to consider the three predecessors x_2, x_3, x_4, whose distances to x_5 are given in the distance matrix D (8). But now, the sphere $S_{4,5}$ has a "moving" center, which causes a change in the position of the circle C_5, given by the intersection $S_{3,5} \wedge S_{4,5}$. The rotation axis for R_5, defined by $X_3 \wedge X_4^0(\lambda_4) \wedge e_\infty$, also changes when λ_4 varies. However, the angle ϕ_5 corresponding to the arcs in C_5 does not depend on λ_4. The position X_5 depends on "local" rotation given by R_5, through the axis determined by the "global" change caused by R_4.

To understand how R_4 acts on X_5, let $\lambda_4 = 0$, implying that $X_4^0(0) = \underline{P_4^0}$, $R_4 = 1$ (the identity transformation), and $z_5 = X_3 \wedge \underline{P_4^0} \wedge e_\infty$. Using $\underline{P_4^0}$ as the center of $S_{4,5}$, we get the points $\underline{P_5^0}, \underline{P_5^1}$ and $\overline{P_5^0}, \overline{P_5^1}$ (the first two for the lower bound of $d_{2,5}$ and the others for the upper bound).

Considering the arc $\underline{P_5^0}\overline{P_5^0}$, we can see that the rotor R_4 determines the rotation axis for R_5 and also the position of $\underline{P_5^0}\overline{P_5^0}$. This means that we can describe the whole set of possible positions for X_5^0, without expliciting the possible positions for X_4^0. In fact, we can fix any position for X_4^0 (we choose $X_4^0(0)$). Algebraically, we have:

$$z_5 = R_4(X_3 \wedge \underline{P_4^0} \wedge e_\infty)\tilde{R}_4, \tag{21}$$

$$R_5 = \cos(\tfrac{\lambda_5}{2}) + z_5^* \sin(\tfrac{\lambda_5}{2}), \quad 0 \le \lambda_5 \le \phi_5, \tag{22}$$

$$X_5^0(\lambda_4, \lambda_5) = R_5 R_4 \underline{P_5^0} \tilde{R}_4 \tilde{R}_5. \tag{23}$$

When $\lambda_4 = 0$, X_5^0 suffers only the local transformation given by R_5.

In our example, we have $\phi_5 = 0.88$ and, for $\lambda_4 = 0$, $\underline{P_5^0} = (0.0895, 1.73, -1.05)^T$. In the example from [15], another point is selected from the arc $\underline{P_5^0}\overline{P_5^0}$, given by $X_5^0(0.208, 0.229) = (-0.156, 1.79, -1.31)^T$.

Positioning the sixth atom is analogous to the fifth one. Thus, we go straight to the seventh atom, which illustrates the general case. We mention that the selected point used in the example from [15] is $X_6^1(0.208, 0.229, 0.334) = (-0.664, 1.096, -1.83)^T$, for $\phi_6 = 0.483$.

So far, the path we are following is shown in Fig. 5.

Atom x_7

For x_7, none of the three predecessors are fixed. The rotation axis for R_7 is defined by X_5 and X_6, which depends on R_4, in addition to R_5 and R_6. Fixing $\lambda_4 = \lambda_5 = \lambda_6 = 0$ and choosing X_4^0, X_5^0, X_6^1, we can calculate the values for $\underline{P_4^0}$, $\underline{P_5^0}, \underline{P_6^1}$:

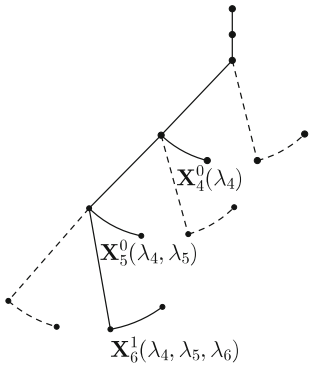

FIGURE 5. Path chosen in the tree to follow the search

$$X_4^0(0) = \underline{P_4^0} = 0.719e_1 + 1.57e_2 - 0.287e_3 + 1.53e_\infty + e_0,$$
$$X_5^0(0,0) = \underline{P_5^0} = 0.089e_1 + 1.73e_2 - 1.05e_3 + 2.06e_\infty + e_0,$$
$$X_6^1(0,0,0) = \underline{P_6^1} = -0.027e_1 + 1.04e_2 - 1.76e_3 + 2.09e_\infty + e_0.$$

Using the interval distance $d_{4,7} \in [2.2, 2.5]$, we get the point pairs related to X_7, the associated angle $\phi_7 = 0.116$, and $\underline{P_7^0}$:

$$X_7^0(0,0,0,0) = \underline{P_7^0} = -0.412e_1 + 0.148e_2 - 1.53e_3 + 1.26e_\infty + e_0.$$

With the values above, we can obtain all of the rotation axis and their corresponding rotors as follows. For $i = 4, \ldots, 7$, we have

$$R_i = \cos\left(\frac{\lambda_i}{2}\right) + z_i^* \sin\left(\frac{\lambda_i}{2}\right), \quad 0 \leq \lambda_i \leq \phi_i,$$

where

$$z_4 = X_2 \wedge X_3 \wedge e_\infty,$$
$$z_5 = R_4(X_3 \wedge \underline{P_4^0} \wedge e_\infty)\tilde{R}_4,$$
$$z_6 = R_5 R_4(\underline{P_4^0} \wedge \underline{P_5^0} \wedge e_\infty)\tilde{R}_4\tilde{R}_5,$$
$$z_7 = R_6 R_5 R_4(\underline{P_5^0} \wedge \underline{P_6^1} \wedge e_\infty)\tilde{R}_4\tilde{R}_5\tilde{R}_6.$$

Finally, we obtain X_7^0 in terms of $\lambda_4, \lambda_5, \lambda_6, \lambda_7$, given by

$$X_7^0(\lambda_4, \lambda_5, \lambda_6, \lambda_7) = R_7 R_6 R_5 R_4 \underline{P_7^0} \tilde{R}_4\tilde{R}_5\tilde{R}_6\tilde{R}_7.$$

Figure 6 shows the path we followed in our example. For any atom x_i, $i > 7$, the situation is similar to x_7.

4. Conclusion

Nuclear magnetic resonance experiments provide distance information that can be used to determine 3D protein structures. This problem can be defined as a DMDGP, where some of the distances are represented by interval of real numbers, due to the uncertainties in NMR information.

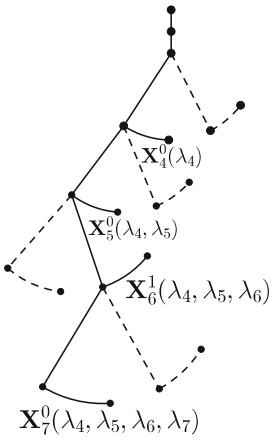

FIGURE 6. Path chosen in the tree to follow the search

In [15], it was proposed a way to deal with interval distances that can help to reduce the search space of the problem, but sample points still need to be selected in order to apply the BP algorithm [14].

The contribution of this paper is the application of CGA to model DMDGP with interval distances, avoiding sampling process and eliminating the heuristic characteristics of BP in this new scenario.

The next challenge is to combine the results of this paper with the ideas presented in [15], in order to define a new algorithm that can be able to choose which paths in the "interval" BP tree should be taken to find "interval" DMDGP solutions.

Acknowledgments

The authors would like to thank the Brazilian research agencies CNPq, CAPES, and FAPESP for their financial support, and the anonymous referees that made important comments and valuable suggestions to this work.

References

[1] Berger, B., Kleinberg, J., Leighton, T.: Reconstructing a three-dimensional model with arbitrary errors. J. ACM **46**, 212–235 (1999)

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

[3] Cassioli, A., Gunluk, O., Lavor, C., Liberti, L.: Discretization vertex orders in distance geometry. Discrete Appl. Math. **197**, 27–41 (2015)

[4] Crippen, G., Havel, T.: Distance Geometry and Molecular Conformation. Wiley, New York (1988)

[5] Dorst, L., Fontijne, D., Mann, S.: Geometric Algebra for Computer Science: An Object-Oriented Approach to Geometry. The Morgan Kaufmann Series in Computer Graphics (2007)

[6] Franchini, S., Vassalo, G., Sorbello, F.: A brief introduction to Clifford algebra. Technical report no. 2/2010. Dipartimento di Ingegneria Informatica, Università degli Studi di Palermo (2010)

[7] Havel, T.: Distance geometry. In: Grant, D., Harris, R. (eds.) Encyclopedia of Nuclear Magnetic Resonance, pp. 1701–1710. Wiley, New York (1995)

[8] Hestenes, D.: Old wine in new bottles: a new algebraic framework for computational geometry. In: Advances in Geometric Algebra with Applications in Science and Engineering, pp. 1–14 (2001)

[9] Hildenbrand, D.: Foundations of Geometric Algebra Computing. Springer, Berlin (2012)

[10] Hildenbrand, D.: Home page of Gaalop. http://www.gaalop.de

[11] Hildenbrand, D., Fontijne, D., Perwass, C., Dorst, L.: Geometric algebra and its application to computer graphics. Tutorial 3. In: Proceedings of Eurographics (2004)

[12] Lavor, C., Liberti, L., Maculan, N., Mucherino, A.: Recent advances on the discretizable molecular distance geometry problem. Eur. J. Oper. Res. **219**, 698–706 (2012)

[13] Lavor, C., Liberti, L., Maculan, N., Mucherino, A.: The discretizable molecular distance geometry problem. Comput. Optim. Appl. **52**, 115–146 (2012)

[14] Lavor, C., Liberti, L., Mucherino, A.: The interval BP algorithm for the discretizable molecular distance geometry problem with interval data. J. Global Optim. **56**, 855–871 (2013)

[15] Lavor, C., Alves, R., Figueiredo, W., Petraglia, A., Maculan, N.: Clifford algebra and the discretizable molecular distance geometry problem. Adv. Appl. Clifford Algebra **25**, 925–942 (2015)

[16] Liberti, L., Lavor, C., Maculan, N.: A branch-and-prune algorithm for the molecular distance geometry problem. Int. Trans. Oper. Res. **15**, 1–17 (2008)

[17] Liberti, L., Lavor, C., Mucherino, A., Maculan, N.: Molecular distance geometry methods: from continuous to discrete. Int. Trans. Oper. Res. **18**, 33–51 (2010)

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

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

[20] Perwass, C.: Geometric Algebra with Applications in Engineering. Springer (2009)

[21] Pesonen, J., Henriksson, O.: Polymer conformations in internal (polyspherical) coordinates. J. Comput. Chem. **31**, 1874–1881 (2009)

[22] Souza, M., Lavor, C., Muritiba, A., Maculan, N.: Solving the molecular distance geometry problem with inaccurate distance data. BMC Bioinform. **14**, S71–S76 (2013)

[23] Wütrich, K.: Protein structure determination in solution by nuclear magnetic resonance spectroscopy. Science **243**, 45–50 (1989)

R. Alves and C. Lavor
University of Campinas (IMECC-UNICAMP)
Campinas
SP
13081-970
Brazil
e-mail: rafaelsoalves@uol.com.br

C. Lavor
e-mail: clavor@ime.unicamp.br

Received: January 19, 2016.
Accepted: March 4, 2016.