Adv. Appl. Clifford Algebras 25 (2015), 925–942

0188-7009/040925-18
published online March 4, 2015
DOI 10.1007/s00006-015-0532-2

Advances in Applied Clifford Algebras

Clifford Algebra and the Discretizable Molecular Distance Geometry Problem

Carlile Lavor*, Rafael Alves, Weber Figueiredo, Antonio Petraglia and Nelson Maculan

Abstract. Nuclear Magnetic Resonance experiments can provide distances between pairs of atoms of a protein that are close enough and the problem is how to determine the 3D protein structure based on this partial distance information, called Molecular Distance Geometry Problem. It is possible to define an atomic order $1, ..., n$ and solve the problem iteratively using an exact method, called Branch-and-Prune (BP). The main step of BP algorithm is to solve a quadratic system to get the two possible positions for i, $i > 3$, in terms of the positions of $i-3, i-2, i-1$ and the distances $d_{i-1,i}, d_{i-2,i}, d_{i-3,i}$. Because of uncertainty in NMR data, some of the distances $d_{i-3,i}$ may not be precise and the main problem to apply BP is related to the difficulty of obtaining an analytical expression of the position of atom i in terms of the positions of the three previous ones and the corresponding distances. We present such expression and although it is similar to one already existing in the literature, based on polyspherical coordinates, a new proof is given, based on Clifford algebra, and we also explain how such expression can be useful in BP using a parameterization which depends on $d_{i-3,i}$. The results suggest that a master equation might exist, what is generally not believed by many researchers.

Keywords. Geometric algebra, distance geometry, protein conformation.

1. Introduction

The calculation of the three-dimensional structure of a biomolecule, which is associated to its chemical and biological properties, is a very important problem in computational chemistry, especially in the case of proteins [27]. This problem can be tackled experimentally, through Nuclear Magnetic Resonance (NMR) spectroscopy and X-ray crystallography [3], or theoretically, via

*Corresponding author.

molecular potential energy minimization [14, 16]. 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 [8, 12, 33].

We can model the problem 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 $d : E \to (0, \infty)$. We say that G is a weighted simple undirected graph when we associate values (distances) to the elements of E, if $\{u, v\} \in E$ then $u \neq v$, and $\{u, v\} = \{v, u\}$, respectively [11]. The question is how to find a function $x : V \to \mathbb{R}^3$ that associates each element of V with a point in $\mathbb{R}^3$ in such a way that the Euclidean distances between the points correspond to the values given by d.

Considering that the distance information provided by NMR experiments is precise, it is possible to define a vertex order $\mathbf{v}_1, ..., \mathbf{v}_n \in V$ and solve the problem iteratively using an exact method, called Branch-and-Prune (BP) [20]. The main step of BP algorithm is to solve a quadratic system to get the two possible positions for vertex $\mathbf{v}_i$, $i > 3$, by using the positions of $\mathbf{v}_{i-1}, \mathbf{v}_{i-2}, \mathbf{v}_{i-3}$ and the distances $d_{i-1,i}, d_{i-2,i}, d_{i-3,i}$.

Because of uncertainty in NMR data, some of the distances $d_{i-3,i}$ may not be precise and the main problem to apply BP is related to the difficulty of obtaining an analytical expression for the position of $\mathbf{v}_i$ in terms of the positions of the three previous vertices and the corresponding distances, where $d_{i-3,i}$ is represented by an interval of real numbers. Using Clifford algebra [1, 13, 23, 24], more specifically the results given in [5, 6], we present such expression. Although there is already a similar one in the literature, using polyspherical coordinates, the way we derive it is new and we also explain how it can be useful in BP algorithm using a parameterization which depends on $d_{i-3,i}$.

The rest of the paper is organized as follows. Section 2 introduces the distance geometry problem that we want to solve. In Section 3, we describe the matrix approach used to solve the problem. Sections 4 and 5 present the main contribution of this paper and some conclusions are given in Section 6.

2. Distance Geometry

The *Distance Geometry Problem* (DGP), in $\mathbb{R}^3$, can be formally defined as follows:

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

$$\forall \{u, v\} \in E, \ ||\mathbf{x}_u - \mathbf{x}_v|| = d_{u,v}, \tag{1}$$

where $\mathbf{x}_u = x(u)$, $\mathbf{x}_v = x(v)$, $d_{u,v} = d(\{u, v\})$, and $||\mathbf{x}_u - \mathbf{x}_v||$ is the Euclidean distance between $\mathbf{x}_u$ and $\mathbf{x}_v$. Notice that we have different notations for a vertex $v \in V$ and its position $\mathbf{x}_v \in \mathbb{R}^3$.

A very recent survey on distance geometry is given in [22] and an edited book with different applications can be found in [26].

A common approach to solve the DGP is to globally minimize the function

$$\sum_{\{u,v\}\in E} (||\mathbf{x}_u - \mathbf{x}_v||^2 - d_{u,v}^2)^2,$$

but solving such a problem is hard from a computational point of view [22]. In [15], some global optimization algorithms have been tested but none of them scale well to medium or large instances. A survey on different methods to the DGP is given in [21].

2.1. Discretizable Distance Geometry Problem

If there is a *trilateration vertex order* in a DGP graph, where every vertex v beyond the first four is adjacent to at least four predecessors u_1, u_2, u_3, u_4 (*i.e.* $\{u_1, v\}, \{u_2, v\}, \{u_3, v\}, \{u_4, v\} \in E$), it is possible to triangulate the position of each next vertex, implying a linear time algorithm to find the unique solution, up to rotations and translations [9].

In general, we do not have trilateration orders in protein graphs $G = (V, E, d)$, but using the information provided by NMR experiments and chemistry of proteins, it is possible to find vertex orders $\mathbf{v}_1, ..., \mathbf{v}_n \in V$ such that [25, 10]:

- For the first three vertices, there exist $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3 \in \mathbb{R}^3$ satisfying equations (1);
- Each vertex with rank greater than three is adjacent to at least three predecessors, *i.e.*

$$\forall i > 3, \exists\ j, k, l \text{ with } j < i, k < i, l < i : \{\mathbf{v}_j, \mathbf{v}_i\}, \{\mathbf{v}_k, \mathbf{v}_i\}, \{\mathbf{v}_l, \mathbf{v}_i\} \in E.$$

The class of DGP instances possessing these orders is called the *Discretizable Distance Geometry Problem* (DDGP), for which there is an exact method, called Branch-and-Prune (BP), for finding all solutions, up to rotations and translations [25] (in order to guarantee a finite number of solutions, the strict triangular inequalities related to the three adjacent predecessors of $\mathbf{v}_i$ must be satisfied).

2.2. Discretizable Molecular Distance Geometry Problem

In fact, DGP graphs related to proteins have enough information to allow definition of vertex orders involving *contiguous* adjacent predecessors, which means that

$$\begin{aligned} \mathbf{v}_j &= \mathbf{v}_{i-3}, \\ \mathbf{v}_k &= \mathbf{v}_{i-2}, \\ \mathbf{v}_l &= \mathbf{v}_{i-1}, \end{aligned}$$

avoiding some numerical problems in DDGP [25]. The new class of DDGP instances where each three consecutive vertices $\{\mathbf{v}_{i-3}, \mathbf{v}_{i-2}, \mathbf{v}_{i-1}\}$ satisfy the strict triangular inequality and each four consecutive vertices $\{\mathbf{v}_{i-3}, \mathbf{v}_{i-2}, \mathbf{v}_{i-1}, \mathbf{v}_i\}$ is a clique (all the six distances involved are known),

for $i > 3$, is called the *Discretizable Molecular Distance Geometry Problem* (DMDGP), and the order itself is a DMDGP order [17, 18].

Thus, 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 $\mathbf{v}_i$, $i > 3$, is obtained by solving the quadratic system

$$\left\{ \begin{array}{l} ||\mathbf{x}_i - \mathbf{x}_{i-3}||^2 = d^2_{i-3,i}, \\ ||\mathbf{x}_i - \mathbf{x}_{i-2}||^2 = d^2_{i-2,i}, \\ ||\mathbf{x}_i - \mathbf{x}_{i-1}||^2 = d^2_{i-1,i}, \end{array} \right\} \tag{2}$$

which can result in up to two possible values for $\mathbf{x}_i$. When there are other adjacent predecessors, one or both possible positions for $\mathbf{v}_i$ may be infeasible with respect to those additional distances. If both are infeasible, it is necessary to backtrack and repeat the methodology [18].

Using DMDGP orders, we can replace resolutions of quadratic systems by matrix multiplications (see Section 3). Computational results presented in [18] demonstrate that the second approach guarantees more stability in BP algorithm. In [4], the authors present a detailed analysis on computational complexity vertex orders in distance geometry. In [7], DMDGP orders for all the side chains of proteins are given.

Because of uncertainty in NMR data [2, 31, 32], some of the distances related to the pairs $\{\mathbf{v}_{i-3}, \mathbf{v}_i\}$ may not be precise values. In [19], it is proposed an extension of BP algorithm, the interval BP (iBP), to manage the uncertainty in distance information, but none of those two approaches (quadratic systems or matrix multiplications) deal well with interval distances.

3. Matrix Algebra Approach

Considering a molecule as a chain of n atoms indexed by $1, ..., n$, in addition to its Cartesian coordinates $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^3$, it can also be described using the *internal coordinates* [28], given by the *bond lengths* d_i (the Euclidean distance between $\mathbf{x}_{i-1}$ and $\mathbf{x}_i$), for $i = 2, \ldots, n$, the *bond angles* θ_i (the angle defined by the atoms $i-2, i-1, i$), for $i = 3, \ldots, n$, and the *torsion angles* ω_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 Figure 1).

Due to the properties of DMDGP orders, the values d_i, θ_i, $\cos(\omega_i)$ can be obtained using the distances between the atoms $i-3, i-2, i-1, i$, for $i = 4, \ldots, n$ [18]. Based on the previous determined positions for $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, ..., \mathbf{x}_{i-1}$, BP algorithm calculates the two possible values for $\mathbf{x}_i = (x_i, y_i, z_i) \in \mathbb{R}^3$ (related to the two values for $\sin(\omega_i) = \pm\sqrt{1 - \cos^2(\omega_i)}$), for $i = 4, \ldots, n$, using the following matrix multiplications [18]:

$$\begin{bmatrix} x_i \\ y_i \\ z_i \\ 1 \end{bmatrix} = B_1 B_2 \ldots B_i \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \ i = 1, \ldots, n, \tag{3}$$

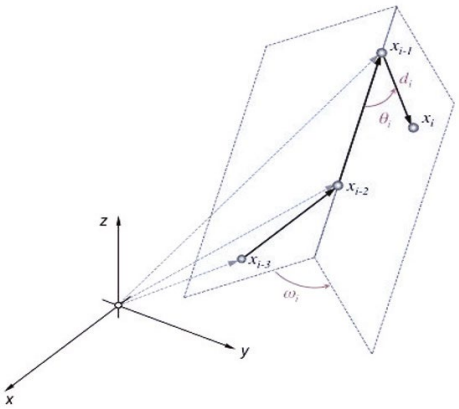

FIGURE 1. Representation of the internal coordinates of the atom i, for $i = 4, \ldots, n$: the distance d_i (the Euclidean distance between $\mathbf{x}_{i-1}$ and $\mathbf{x}_i$), the bond angle θ_i (the angle defined by the atoms $i-2$, $i-1$ and i) and the torsion angle ω_i (the angle between the normals through the planes defined by $i-3$, $i-2$, $i-1$ and $i-2$, $i-1$, i).

where

$$B_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad B_2 = \begin{bmatrix} -1 & 0 & 0 & -d_2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$B_3 = \begin{bmatrix} -\cos\theta_3 & -\sin\theta_3 & 0 & -d_3\cos\theta_3 \\ \sin\theta_3 & -\cos\theta_3 & 0 & d_3\sin\theta_3 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

and

$$B_i = \begin{bmatrix} -\cos\theta_i & -\sin\theta_i & 0 & -d_i\cos\theta_i \\ \sin\theta_i\cos\omega_i & -\cos\theta_i\cos\omega_i & -\sin\omega_i & d_i\sin\theta_i\cos\omega_i \\ \sin\theta_i\sin\omega_i & -\cos\theta_i\sin\omega_i & \cos\omega_i & d_i\sin\theta_i\sin\omega_i \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

In an equivalent way, matrices in $\mathbb{R}^{3\times 3}$ can also be used if the related translations and rotations are considered separately [30].

For $i = 5, ..., n$, equation (3) can also be written as

$$\begin{bmatrix} x_i \\ y_i \\ z_i \\ 1 \end{bmatrix} = Q_{i-1} B_i \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} d_i \left[-q_{11}^{i-1} \cos\theta_i + \sin\theta_i \left(q_{12}^{i-1} \cos\omega_i + q_{13}^{i-1} \sin\omega_i \right) \right] + q_{14}^{i-1} \\ d_i \left[-q_{21}^{i-1} \cos\theta_i + \sin\theta_i \left(q_{22}^{i-1} \cos\omega_i + q_{23}^{i-1} \sin\omega_i \right) \right] + q_{24}^{i-1} \\ d_i \left[-q_{31}^{i-1} \cos\theta_i + \sin\theta_i \left(q_{32}^{i-1} \cos\omega_i + q_{33}^{i-1} \sin\omega_i \right) \right] + q_{34}^{i-1} \\ 1 \end{bmatrix},$$

where matrix $Q_{i-1} = B_1 \dots B_{i-1}$ is given by

$$Q_{i-1} = \begin{bmatrix} q_{11}^{i-1} & q_{12}^{i-1} & q_{13}^{i-1} & q_{14}^{i-1} \\ q_{21}^{i-1} & q_{22}^{i-1} & q_{23}^{i-1} & q_{24}^{i-1} \\ q_{31}^{i-1} & q_{32}^{i-1} & q_{33}^{i-1} & q_{34}^{i-1} \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

This means that to obtain the Cartesian coordinates $\mathbf{x}_i$, for $i = 5, ..., n$, in addition to the internal coordinates d_i, θ_i, ω_i of atom i, we have to use the elements of the matrix Q_{i-1}, which depends on all the internal coordinates of atoms $i-1, i-2, \dots, 2, 1$.

Another possibility is to use a new reference system defined by $\mathbf{x}_{i-1}, \mathbf{x}_{i-2}, \mathbf{x}_{i-3}$, but additional translations and rotations must be considered in order to calculate the Cartesian coordinates $\mathbf{x}_i$ in terms of the initial reference system $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$.

4. Clifford Algebra Approach

An alternative method for the classical matrix algebra approach was proposed in [5, 6], based on the Clifford algebra (also known as geometric algebra). With this new method, we can efficiently generate chain segment rotations in a recursive way and obtain an analytical expression for the Cartesian coordinates $\mathbf{x}_i$, using only $\mathbf{x}_{i-1}, \mathbf{x}_{i-2}, \mathbf{x}_{i-3}$ and the internal coordinates d_i, θ_i, ω_i of atom i, for $i = 4, \dots, n$.

4.1. Canonical form rotor equation

First, we rederive the results presented in [5, 6]. Using the matrices (3), the first three atoms of the molecule can be fixed at positions

$$\mathbf{x}_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \ \mathbf{x}_2 = \begin{bmatrix} -d_2 \\ 0 \\ 0 \end{bmatrix}, \ \mathbf{x}_3 = \begin{bmatrix} -d_2 + d_3 \cos\theta_3 \\ d_3 \sin\theta_3 \\ 0 \end{bmatrix}.$$

Let the vector $\mathbf{r}_i$, for $i = 3, ..., n$, be defined by

$$\mathbf{r}_i = \mathbf{x}_i - \mathbf{x}_{i-1}.$$

Now, for $i = 4, ..., n$, let the vector $\mathbf{r}_i^\theta$ and the rotor R_i^θ be defined by (see Figure 2 and Figure 3)

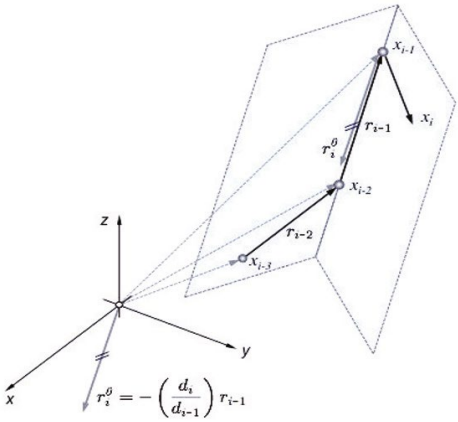

FIGURE 2. Definition of the vector $\mathbf{r}_i^\theta$ in the opposite sense of $\mathbf{r}_{i-1} = \mathbf{x}_{i-1} - \mathbf{x}_{i-2}$.

$$\mathbf{r}_i^\theta = -\left(\frac{d_i}{d_{i-1}}\right)\mathbf{r}_{i-1},$$

and

$$R_i^\theta = \cos\left(\frac{\theta_i}{2}\right) + \hat{B}_i^\theta \sin\left(\frac{\theta_i}{2}\right),$$

where $\hat{B}_i^\theta$ is the normalized bivector $B_i^\theta = \mathbf{r}_{i-1} \wedge \mathbf{r}_{i-2}$. Using $\mathbf{r}_i^\theta$ and R_i^θ, for $i = 4, ..., n$, let the vector $\mathbf{r}_i^\omega$ be defined by (see Figure 3)

$$\mathbf{r}_i^\omega = \mathbf{r}_i^\theta \left(R_i^\theta\right)^2.$$

For $i = 4, ..., n$, let the rotor R_i^ω be defined by (see Figure 4)

$$R_i^\omega = \cos\left(\frac{\omega_i}{2}\right) + \hat{B}_i^\omega \sin\left(\frac{\omega_i}{2}\right),$$

where $\hat{B}_i^\omega$ is the normalized bivector B_i^ω, orthogonal to the vector $\mathbf{r}_{i-1}$ and given by

$$B_i^\omega = \mathbf{r}_{i-1} \cdot I,$$

where $I = \mathbf{e}_1 \wedge \mathbf{e}_2 \wedge \mathbf{e}_3$. Using $\mathbf{r}_i^\omega$ and R_i^ω, for $i = 4, ..., n$, let the vector $\mathbf{r}_i$ be defined by (see Figure 4)

$$\mathbf{r}_i = R_i^\omega \mathbf{r}_i^\omega \tilde{R}_i^\omega.$$

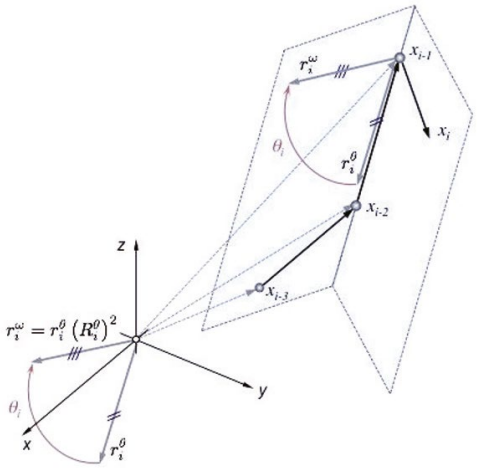

FIGURE 3. Action of the rotor R_i^θ: it rotates the vector $\mathbf{r}_i^\theta$, by the angle θ_i, to obtain $\mathbf{r}_i^\omega$.

$\tilde{R}_i^\omega$ is called the reverse of R_i^ω. It is obtained from the reversion of the order in the outer products among basis vectors. In our case, we can derive $\tilde{R}_i^\omega$ from

$$\tilde{B}_i^\omega = \mathbf{r}_{i-1} \cdot \tilde{I},$$

where $\tilde{I} = e_3 \wedge e_2 \wedge e_1$. For $i = 4, ..., n$, we can then write (see Figure 5)

$$\mathbf{x}_i = \mathbf{x}_{i-1} + \mathbf{r}_i.$$

4.2. Analytical form rotor equation

Using the results of the previous section, we have that

$$\mathbf{x}_i = \mathbf{x}_{i-1} + ABC,$$

where

$$\begin{aligned} A &= \left(\cos\left(\frac{\omega_i}{2}\right) + \frac{\mathbf{r}_{i-1}\tilde{I}}{||\mathbf{r}_{i-1}||}\sin\left(\frac{\omega_i}{2}\right)\right)\left(\frac{d_i}{d_{i-1}}\mathbf{r}_{i-1}\right), \\ B &= \left(\cos\left(\frac{\theta_i}{2}\right) + \frac{\mathbf{r}_{i-1} \wedge \mathbf{r}_{i-2}}{||\mathbf{r}_{i-1} \wedge \mathbf{r}_{i-2}||}\sin\left(\frac{\theta_i}{2}\right)\right)^2, \\ C &= \left(\cos\left(\frac{\omega_i}{2}\right) + \frac{\mathbf{r}_{i-1}I}{||\mathbf{r}_{i-1}||}\sin\left(\frac{\omega_i}{2}\right)\right), \end{aligned}$$

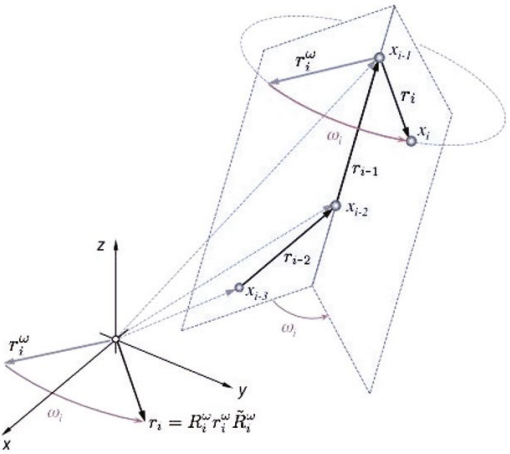

FIGURE 4. Action of the rotor R_i^ω: it rotates the vector $\mathbf{r}_i^\omega$, by the angle ω_i, to obtain the vector $\mathbf{r}_i$.

and $\tilde{I} = \mathbf{e}_3 \wedge \mathbf{e}_2 \wedge \mathbf{e}_1$. Since B can be rewritten as

$$B = \left(\cos(\theta_i) + \frac{\mathbf{r}_{i-1} \wedge \mathbf{r}_{i-2}}{||\mathbf{r}_{i-1} \wedge \mathbf{r}_{i-2}||} \sin(\theta_i)\right),$$

we obtain that

$$\mathbf{x}_i = \mathbf{x}_{i-1} - \frac{d_{i,i-1}}{d_{i-1,i-2}} \left(\cos\left(\frac{\omega_i}{2}\right) \mathbf{r}_{i-1} - ||\mathbf{r}_{i-1}|| \sin\left(\frac{\omega_i}{2}\right) I\right) PQ,$$

where

$$\begin{aligned} P &= \left(\cos(\theta_i) + \frac{\mathbf{r}_{i-1} \wedge \mathbf{r}_{i-2}}{||\mathbf{r}_{i-1} \wedge \mathbf{r}_{i-2}||} \sin(\theta_i)\right), \\ Q &= \left(\cos\left(\frac{\omega_i}{2}\right) + \frac{\mathbf{r}_{i-1} I}{||\mathbf{r}_{i-1}||} \sin\left(\frac{\omega_i}{2}\right)\right). \end{aligned}$$

As a result of the product PQ, we get

$$\begin{aligned} PQ &= \left(\cos(\theta_i) + \frac{\mathbf{r}_{i-1} \wedge \mathbf{r}_{i-2}}{||\mathbf{r}_{i-1} \wedge \mathbf{r}_{i-2}||} \sin(\theta_i)\right) \left(\cos\left(\frac{\omega_i}{2}\right) + \frac{\mathbf{r}_{i-1} I}{||\mathbf{r}_{i-1}||} \sin\left(\frac{\omega_i}{2}\right)\right) \\ &= \cos(\theta_i) \cos\left(\frac{\omega_i}{2}\right) + \cos(\theta_i) \sin\left(\frac{\omega_i}{2}\right) \frac{\mathbf{r}_{i-1} I}{||\mathbf{r}_{i-1}||} \\ &\quad + \sin(\theta_i) \cos\left(\frac{\omega_i}{2}\right) \frac{\mathbf{r}_{i-1} \wedge \mathbf{r}_{i-2}}{||\mathbf{r}_{i-1} \wedge \mathbf{r}_{i-2}||} \\ &\quad + \sin(\theta_i) \sin\left(\frac{\omega_i}{2}\right) \frac{(\mathbf{r}_{i-1} \wedge \mathbf{r}_{i-2}) \mathbf{r}_{i-1} I}{||\mathbf{r}_{i-1}||||\mathbf{r}_{i-1} \wedge \mathbf{r}_{i-2}||}, \end{aligned}$$

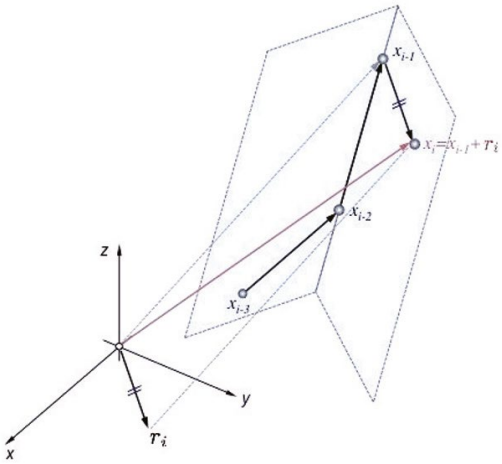

FIGURE 5. From $\mathbf{r}_i$, we obtain $\mathbf{x}_i = \mathbf{x}_{i-1} + \mathbf{r}_i$, the position vector for the atom i.

which means that

$$\mathbf{x}_i = \mathbf{x}_{i-1} - \frac{d_i}{d_{i-1}} \left(R + S + T + U \right),$$

where

$$\begin{aligned}
R &= \cos^2\left(\frac{\omega_i}{2}\right) \cos\left(\theta_i\right) \mathbf{r}_{i-1} + \cos\left(\frac{\omega_i}{2}\right) \sin\left(\frac{\omega_i}{2}\right) \cos\left(\theta_i\right) ||\mathbf{r}_{i-1}|| I, \\
S &= \cos^2\left(\frac{\omega_i}{2}\right) \sin\left(\theta_i\right) \frac{\mathbf{r}_{i-1} \left(\mathbf{r}_{i-1} \wedge \mathbf{r}_{i-2}\right)}{||\mathbf{r}_{i-1} \wedge \mathbf{r}_{i-2}||} \\
&\quad + \cos\left(\frac{\omega_i}{2}\right) \sin\left(\frac{\omega_i}{2}\right) \sin\left(\theta_i\right) \frac{\mathbf{r}_{i-1} \left(\mathbf{r}_{i-1} \wedge \mathbf{r}_{i-2}\right) \mathbf{r}_{i-1} I}{||\mathbf{r}_{i-1} \wedge \mathbf{r}_{i-2}|| ||\mathbf{r}_{i-1}||}, \\
T &= -\cos\left(\frac{\omega_i}{2}\right) \sin\left(\frac{\omega_i}{2}\right) \cos\left(\theta_i\right) ||\mathbf{r}_{i-1}|| I \\
&\quad + \sin^2\left(\frac{\omega_i}{2}\right) \cos\left(\theta_i\right) \mathbf{r}_{i-1}, \\
U &= -\sin\left(\frac{\omega_i}{2}\right) \cos\left(\frac{\omega_i}{2}\right) \sin\left(\theta_i\right) \frac{||\mathbf{r}_{i-1}|| I \left(\mathbf{r}_{i-1} \wedge \mathbf{r}_{i-2}\right)}{||\mathbf{r}_{i-1} \wedge \mathbf{r}_{i-2}||} \\
&\quad - \sin^2\left(\frac{\omega_i}{2}\right) \sin\left(\theta_i\right) \frac{I \left(\mathbf{r}_{i-1} \wedge \mathbf{r}_{i-2}\right) \mathbf{r}_{i-1} I}{||\mathbf{r}_{i-1} \wedge \mathbf{r}_{i-2}||}.
\end{aligned}$$

Using the properties

$$\begin{aligned} I\mathbf{r}_{i-1}I &= -\mathbf{r}_{i-1}, \\ I(\mathbf{r}_{i-1}\wedge\mathbf{r}_{i-2})\mathbf{r}_{i-1}I &= -(\mathbf{r}_{i-1}\wedge\mathbf{r}_{i-2})\mathbf{r}_{i-1} = \mathbf{r}_{i-1}(\mathbf{r}_{i-1}\wedge\mathbf{r}_{i-2}), \\ \mathbf{r}_{i-1}(\mathbf{r}_{i-1}\wedge\mathbf{r}_{i-2})\mathbf{r}_{i-1}I &= -||\mathbf{r}_{i-1}||^2(\mathbf{r}_{i-1}\wedge\mathbf{r}_{i-2})I \\ &= -||\mathbf{r}_{i-1}||^2 I(\mathbf{r}_{i-1}\wedge\mathbf{r}_{i-2}), \end{aligned}$$

we get

$$\mathbf{x}_i = \mathbf{x}_{i-1} - \frac{d_i}{d_{i-1}}\left(V - W\right),$$

where

$$\begin{aligned} V &= \cos(\theta_i)\mathbf{r}_{i-1} + \frac{\cos(\omega_i)\sin(\theta_i)}{||\mathbf{r}_{i-1}\wedge\mathbf{r}_{i-2}||}\mathbf{r}_{i-1}(\mathbf{r}_{i-1}\wedge\mathbf{r}_{i-2}), \\ W &= \frac{||\mathbf{r}_{i-1}||\sin(\omega_i)\sin(\theta_i)}{||\mathbf{r}_{i-1}\wedge\mathbf{r}_{i-2}||} I(\mathbf{r}_{i-1}\wedge\mathbf{r}_{i-2}), \end{aligned}$$

which implies that

$$\mathbf{x}_i = -\left(\frac{d_i}{||\mathbf{x}_{i-1}-\mathbf{x}_{i-2}||}\right)(X + YZ),$$

where

$$\begin{aligned} X &= \mathbf{x}_{i-1}\left(\cos(\theta_i) - \frac{||\mathbf{x}_{i-1}-\mathbf{x}_{i-2}||}{d_i}\right) - \mathbf{x}_{i-2}\cos(\theta_i), \\ Y &= (\cos(\omega_i)\sin(\theta_i)(\mathbf{x}_{i-1}-\mathbf{x}_{i-2}) - \sin(\omega_i)\sin(\theta_i)||\mathbf{x}_{i-1}-\mathbf{x}_{i-2}||I), \\ Z &= \left(\frac{(\mathbf{x}_{i-1}-\mathbf{x}_{i-2})\wedge(\mathbf{x}_{i-2}-\mathbf{x}_{i-3})}{||(\mathbf{x}_{i-1}-\mathbf{x}_{i-2})\wedge(\mathbf{x}_{i-2}-\mathbf{x}_{i-3})||}\right). \end{aligned}$$

Now, we can use the properties

$$I(a\wedge b) = I(a\times b)I = -(a\times b)$$

and

$$a(a\wedge b) = a\cdot(a\wedge b) = (a\cdot a)b - a(b\cdot a) = ||a||^2 b - a(b\cdot a)$$

to eliminate the exterior and geometric products from Clifford algebra and, finally, obtain that

$$\mathbf{x}_i = -\left(\frac{d_i}{||\mathbf{x}_{i-1}-\mathbf{x}_{i-2}||}\right)(\alpha + \beta\,(\gamma + \delta)), \tag{4}$$

where $\beta \in \mathbb{R}$ and $\alpha, \gamma, \delta \in \mathbb{R}^3$, given by

$$\begin{aligned} \alpha &= \left(\cos(\theta_i) - \frac{||\mathbf{x}_{i-1}-\mathbf{x}_{i-2}||}{d_i}\right)\mathbf{x}_{i-1} - \cos(\theta_i)\mathbf{x}_{i-2}, \\ \beta &= \frac{\sin(\theta_i)}{||(\mathbf{x}_{i-1}-\mathbf{x}_{i-2})\times(\mathbf{x}_{i-2}-\mathbf{x}_{i-3})||}, \\ \gamma &= \cos(\omega_i)\left(||\mathbf{x}_{i-1}-\mathbf{x}_{i-2}||^2(\mathbf{x}_{i-2}-\mathbf{x}_{i-3})\right. \\ &\quad \left. -\,(\mathbf{x}_{i-1}-\mathbf{x}_{i-2})\cdot(\mathbf{x}_{i-2}-\mathbf{x}_{i-3})(\mathbf{x}_{i-1}-\mathbf{x}_{i-2})\right), \\ \delta &= \sin(\omega_i)||\mathbf{x}_{i-1}-\mathbf{x}_{i-2}||\left((\mathbf{x}_{i-1}-\mathbf{x}_{i-2})\times(\mathbf{x}_{i-2}-\mathbf{x}_{i-3})\right). \end{aligned}$$

5. Application to Distance Geometry

In [28], using the concept of polyspherical coordinates and based on a different approach (with practically no use of Clifford algebra), the authors present a similar expression to (4), but no connection to distance geometry is provided. In order to explain this connection, we need first to describe $\cos(\omega_i)$ in terms of the distances between the atoms $i-3, i-2, i-1, i$.

5.1. Angles and distances in a trihedron

A *trihedron* is a triple of vectors in $\mathbb{R}^3$ with a common vertex S yielding three faces (see Figure 6). The *cosine law for a trihedron* [29] states a relation between the cosine and sine of angles formed by its vectors, called *face angles*, and the cosine of one of the torsion angles defined by its faces. Let α, β, γ be the three face angles and ω the torsion angle opposite to γ. Then, the cosine law for a trihedron is given by

$$\cos\gamma = \cos\alpha\cos\beta + \sin\alpha\sin\beta\cos\omega. \tag{5}$$

Since $\alpha, \beta \in [0, \pi]$, we can uniquely determine $\sin\alpha$ and $\sin\beta$ through $\sin\alpha = \sqrt{1-\cos^2\alpha}$ and $\sin\beta = \sqrt{1-\cos^2\beta}$. Moreover, if α and β are different from 0 and π, we can rearrange equation (5) and obtain the following expression for the cosine of the torsion angle ω:

$$\cos\omega = \frac{\cos\gamma - \cos\alpha\cos\beta}{\sqrt{1-\cos^2\alpha}\sqrt{1-\cos^2\beta}}.$$

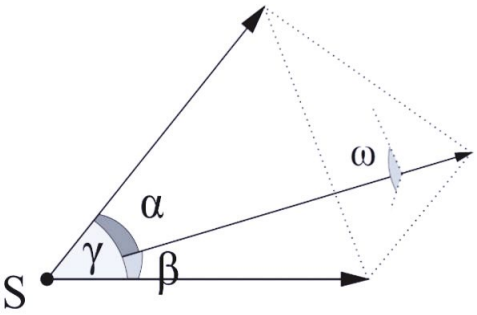

FIGURE 6. Face angles and a torsion angle of a trihedron.

Now, consider a quadruplet of consecutive atoms $i-3, i-2, i-1, i$ of a molecule. In order to consider the torsion angle ω_i, given by the normals through the planes defined by $i-3, i-2, i-1$, and $i-2, i-1, i$, we can choose the atom $i-2$ to be the common vertex (see Figure 7). Using the classical

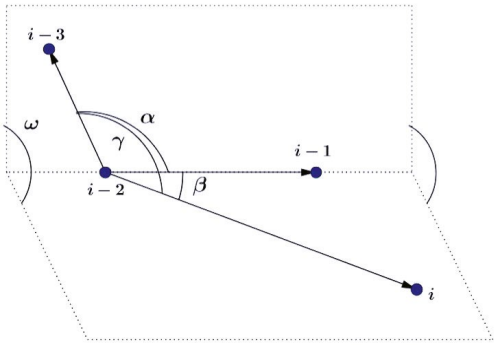

FIGURE 7. A trihedron defined by atoms $i-3, i-2, i-1, i$.

cosine law, we can write $\cos\alpha$, $\cos\beta$, $\cos\gamma$ in terms of the distances between the atoms $i-3, i-2, i-1, i$ and obtain that

$$\cos(\omega_i) = \frac{2d^2_{i-2,i-1}(d^2_{i-3,i-2} + d^2_{i-2,i} - d^2_{i-3,i}) - (d_{i-3,i-2,i-1})(d_{i-2,i-1,i})}{\sqrt{4d^2_{i-3,i-2}d^2_{i-2,i-1} - (d^2_{i-3,i-2,i-1})}\sqrt{4d^2_{i-2,i-1}d^2_{i-2,i} - (d^2_{i-2,i-1,i})}}, \tag{6}$$

where

$$\begin{aligned} d_{i-3,i-2,i-1} &= d^2_{i-3,i-2} + d^2_{i-2,i-1} - d^2_{i-3,i-1}, \\ d_{i-2,i-1,i} &= d^2_{i-2,i-1} + d^2_{i-2,i} - d^2_{i-1,i}. \end{aligned}$$

5.2. The cosine of a torsion angle as a parameter

As we have already mentioned, the main step of BP algorithm is to determine the two possible positions for atom i, $i > 3$, in terms of the positions of atoms $i-3, i-2, i-1$ and the distances $d_{i-1,i}, d_{i-2,i}, d_{i-3,i}$.

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. However, because of uncertainty in NMR data, some of the distances $d_{i-3,i}$ may not be precise (they can be represented by an interval of real numbers) and the main problem to apply BP is related to the difficulty of obtaining an analytical expression for the position of atom i ($\mathbf{x}_i$) in terms of the positions of the three previous ones and the corresponding distances [19].

Since $\cos(\omega_i)$ can be calculated in terms of the distances between the atoms $i-3, i-2, i-1, i$, given by expression (6), for each interval related to the distance $d_{i-3,i}$, denoted by $[d^{\min}_{i-3,i}, d^{\max}_{i-3,i}]$ (see Figure 8), we obtain an associated interval for $\cos(\omega_i)$. It is important to notice that, by the definition

of the DMDGP, all of the intervals for $\cos(\omega_i)$, for $i = 4, \ldots, n$, can be previously calculated for each DMDGP instance. This means that, in the expression (6), only the value for $d_{i-3,i}$ is given as an interval and all the other distances are precise values.

Denoting $\cos(\omega_i)$ by $t \in [-1, 1]$, we can parameterize the position of atom i using expression (4), for $i = 4, \ldots, n$, given by

$$\mathbf{x}_i(t) = \mathbf{p}_1 + t\mathbf{p}_2 \pm \sqrt{1-t^2}\mathbf{p}_3, \tag{7}$$

where $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$ are vectors in $\mathbb{R}^3$ given in terms of d_i, θ_i (given as part of the input of the problem) and $\mathbf{x}_{i-1}, \mathbf{x}_{i-2}, \mathbf{x}_{i-3}$ (previously calculated by BP):

$$\begin{aligned}
\mathbf{p}_1 &= -\left(\frac{d_i}{||\mathbf{x}_{i-1} - \mathbf{x}_{i-2}||}\right) \\
&\quad \left(\left(\cos(\theta_i) - \frac{||\mathbf{x}_{i-1} - \mathbf{x}_{i-2}||}{d_i}\right)\mathbf{x}_{i-1} - \cos(\theta_i)\mathbf{x}_{i-2}\right), \\
\mathbf{p}_2 &= -\left(\frac{d_i}{||\mathbf{x}_{i-1} - \mathbf{x}_{i-2}||}\right)\left(\frac{\sin(\theta_i)}{||(\mathbf{x}_{i-1} - \mathbf{x}_{i-2}) \times (\mathbf{x}_{i-2} - \mathbf{x}_{i-3})||}\right) \\
&\quad \left(||\mathbf{x}_{i-1} - \mathbf{x}_{i-2}||^2(\mathbf{x}_{i-2} - \mathbf{x}_{i-3})\right. \\
&\quad \left.-(\mathbf{x}_{i-1} - \mathbf{x}_{i-2}) \cdot (\mathbf{x}_{i-2} - \mathbf{x}_{i-3})(\mathbf{x}_{i-1} - \mathbf{x}_{i-2})\right), \\
\mathbf{p}_3 &= -\left(\frac{d_i}{||\mathbf{x}_{i-1} - \mathbf{x}_{i-2}||}\right)\left(\frac{\sin(\theta_i)}{||(\mathbf{x}_{i-1} - \mathbf{x}_{i-2}) \times (\mathbf{x}_{i-2} - \mathbf{x}_{i-3})||}\right) \\
&\quad ||\mathbf{x}_{i-1} - \mathbf{x}_{i-2}|| \left((\mathbf{x}_{i-1} - \mathbf{x}_{i-2}) \times (\mathbf{x}_{i-2} - \mathbf{x}_{i-3})\right).
\end{aligned}$$

By using expression (7), for each value of t in $[-1, 1]$, we obtain the two solutions of the system (2), and for each subinterval of $[-1, 1]$, we obtain the two related arcs (see Figure 8).

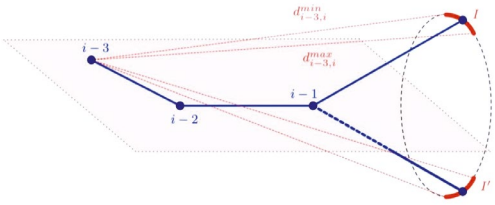

FIGURE 8. Arcs associated to the distance $d_{i-3,i}$. For each value of $d_{i-3,i} \in \left[d^{min}_{i-3,i}, d^{max}_{i-3,i}\right]$, we obtain two possible positions for the atom i, one on each of the arcs I and I'.

To the best of our knowledge, in the DMDGP literature (see the very recent survey [22]), this is the first time that an analytical solution to

the system (2) is given, considering that $d_{i-3,i}$ is represented by an interval of real numbers, in the context of distance geometry.

5.3. An illustrative example

In order to illustrate how the expression (7) can be useful in the DMDGP, consider that we have the following points ($d_i = 1$ and $\theta_i = 120°$, for $i = 2, ..., 7$),

$$\mathbf{x}_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \ \mathbf{x}_2 = \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix}, \ \mathbf{x}_3 = \begin{bmatrix} -1.5 \\ 0.866025 \\ 0 \end{bmatrix},$$

$$\mathbf{x}_4 = \begin{bmatrix} 0.625 \\ 1.51554 \\ -0.75 \end{bmatrix}, \ \mathbf{x}_5 = \begin{bmatrix} -0.15625 \\ 1.78618 \\ -1.3125 \end{bmatrix}, \ \mathbf{x}_6 = \begin{bmatrix} -0.664063 \\ 1.09606 \\ -1.82813 \end{bmatrix},$$

and that the problem is to describe the two possible arcs for $\mathbf{x}_7$ related to the interval distance $2.2 \leq d_{4,7} \leq 2.5$. Using expression (6) and the associated distances, we get

$$t = \cos\omega_7 = \frac{11 - 2d_{4,7}^2}{3}. \tag{8}$$

This implies that, for $2.2 \leq d_{4,7} \leq 2.5$, we have $-0.5 \leq t \leq 0.44$. From expression (7), we can obtain the two arcs related to $\mathbf{x}_7$.

Let us also consider that NMR experiments provide the interval distance $2.55 \leq d_{1,7} \leq 2.85$. Using again expression (6), but for the trihedron defined by $\mathbf{x}_1, \mathbf{x}_5, \mathbf{x}_6, \mathbf{x}_7$, we get

$$-0.43 \leq t' \leq 0.001.$$

From expression (7), we can obtain the four points that define the two arcs associated to the interval distance $d_{1,7}$. Then, calculating the distances between $\mathbf{x}_4$ and these four points, we have

$$2.51 \leq ||\mathbf{x}_7 - \mathbf{x}_4|| \leq 2.60 \tag{9}$$

and

$$2.16 \leq ||\mathbf{x}_7 - \mathbf{x}_4|| \leq 2.30. \tag{10}$$

Since the given interval for $d_{4,7}$ is $[2.2, 2.5]$, the interval (9) can be discarded. However, from the interval (10), we have a new upper bound for $d_{4,7}$: 2.3. From (8), we obtain

$$0.13 \leq t \leq 0.44.$$

This interval reduction, from $[-0.5, 0.44]$ to $[0.13, 0.44]$, can be very useful in the iBP (the interval BP algorithm), since the discretization process will be done in a smaller arc. In addition to this, the sample points that must be taken from the new arc can be obtained in a faster way, comparing to resolutions of quadratic systems or matrix multiplication.

6. Conclusion

The DMDGP is a distance geometry problem related to protein structure determination using NMR data, where it is possible to define an atomic ordering $1, ..., n$ in order to solve the problem iteratively using an exact method, called Branch-and-Prune (BP). The main step of BP algorithm is to determine the two possible positions for atom i, $i > 3$, in terms of the positions of atoms $i-3, i-2, i-1$ and the distances $d_{i-1,i}, d_{i-2,i}, d_{i-3,i}$.

In [19], it was proposed an extension of BP algorithm (iBP) to manage the uncertainty in NMR data related to some of the distances $d_{i-3,i}$, but none of the two common approaches, based on quadratic systems or matrix multiplications, deal well with interval distances.

We 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. Although this analytical expression is similar to one already existing in the literature, based on polyspherical coordinates, a new proof is given, based on Clifford algebra, indicating that a master equation might exist. We also explained how such expression can be useful in iBP algorithm using a parameterization which depends on $d_{i-3,i}$.

Acknowledgments

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

References

[1] R. Alves and C. Lavor, *Clifford algebra applied to Grover's algorithm*. Advances in Applied Clifford Algebra **20** (2010), 477-488.

[2] B. Berger, J. Kleinberg, and T. Leighton, *Reconstructing a three-dimensional model with arbitrary errors*. Journal of the ACM **46** (1999), 212-235.

[3] A. Brünger and M. Nilges, *Computational challenges for macromolecular structure determination by X-ray crystallography and solution NMR-spectroscopy*. Quarterly Reviews of Biophysics **26** (1993), 49-125.

[4] A. Cassioli, O. Gunluk, C. Lavor, and L. Liberti, *Discretization vertex orders in distance geometry*. IBM Research Report RC25434, 2013 (accepted in Discrete Applied Mathematics).

[5] P. Chys, *Application of geometric algebra for the description of polymer conformations*. Journal of Chemical Physics **128** (2008), 104107(1)-104107(12).

[6] P. Chys and P. Chacón, *Spinor product computations for protein conformations*. Journal of Computational Chemistry **33** (2012), 1717-1729.

[7] V. Costa, A. Mucherino, C. Lavor, A. Cassioli, L. Carvalho, and N. Maculan, *Discretization orders for protein side chains*. Journal of Global Optimization, **60** (2014), 333-349.

[8] G. Crippen and T. Havel, *Distance Geometry and Molecular Conformation*. Wiley, New York, 1988.

[9] T. Eren, D. Goldenberg, W. Whiteley, Y. Yang, A. Morse, B. Anderson, and P. Belhumeur, *Rigidity, computation, and randomization in network localization*. IEEE Infocom Proc. **4** (2004), pp. 2673–2684.

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

[11] F. Harary, *Graph Theory*. Addison-Wesley, Reading, 1994.

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

[13] D. Hestenes, *New Foundations for Classical Mechanics*. Kluwer, Boston, 1999.

[14] C. Lavor and N. Maculan, *A function to test methods applied to global minimization of potential energy of molecules*. Numerical Algorithms **35** (2004), 287-300.

[15] C. Lavor, L. Liberti, and N. Maculan, *Computational experience with the molecular distance geometry problem*. In Global Optimization: Scientific and Engineering Case Studies, J. Pintér, ed., Springer, Berlin (2006), pp. 213–225.

[16] C. Lavor, *Analytic evaluation of the gradient and Hessian of molecular potential energy functions*. Physica D **227** (2007), 135-141.

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

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

[19] 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** (2013), 855-871.

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

[21] L. Liberti, C. Lavor, A. Mucherino, and N. Maculan, *Molecular distance geometry methods: from continuous to discrete*. International Transactions in Operational Research **18** (2010), 33-51.

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

[23] P. Lounesto, *Clifford Algebras and Spinnors*. Cambridge University Press, Cambridge, 1997.

[24] N. Melo and C. Lavor, *A Clifford algebra of signature (n,3n) and the density operators of quantum information theory*. Advances in Applied Clifford Algebra **23** (2013), 143-152.

[25] A. Mucherino, C. Lavor, and L. Liberti, *The discretizable distance geometry problem*. Optimization Letters **6** (2012), 1671–1686.

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

[27] A. Neumaier, *Molecular modeling of proteins and mathematical prediction of protein structure*. SIAM Review **39** (1997), 407-460.

[28] J. Pesonen and O. Henriksson, *Polymer conformations in internal (polyspherical) coordinates*. Journal of Computational Chemistry **31** (2009), 1874-1881.

[29] A. Pogorelov, *Geometry*. Mir Publishers, Moscow, 1987.

[30] C. Seok and E. Coutsias, *Efficiency of rotational operators for geometric manipulation of chain molecules*. Bulletin of the Korean Chemical Society **28** (2007), 1705-1708.

[31] M. Souza, A. Xavier, C. Lavor, and N. Maculan, *Hyperbolic smoothing and penalty techniques applied to molecular structure determination*. Operations Research Letters **39** (2011), 461–465.

[32] M. Souza, C. Lavor, A. Muritiba, and N. Maculan, *Solving the molecular distance geometry problem with inaccurate distance data*. BMC Bioinformatics **14** (2013), S71-S76.

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

Carlile Lavor and Rafael Alves
University of Campinas (IMECC-UNICAMP)
13081-970, Campinas - SP, Brazil
e-mail: clavor@ime.unicamp.br
rafaelsoalves@uol.com.br

Weber Figueiredo
CEFET - RJ
20271-110, Rio de Janeiro - RJ, Brazil
e-mail: weber@globo.com

Antonio Petraglia and Nelson Maculan
Federal University of Rio de Janeiro (COPPE-UFRJ)
21945-970, Rio de Janeiro - RJ, Brazil
e-mail: petra@pads.ufrj.br
maculan@cos.ufrj.br

Received: March 17, 2014.
Accepted: November 7, 2014.