Unassigned Distance Geometry and Molecular Conformation Problems

Phil Duxbury · Carlile Lavor · Leo Liberti · Luiz Leduino de Salles-Neto

Abstract 3D protein structures and nanostructures can be obtained by exploiting distance information provided by experimental techniques, such as Nuclear Magnetic Resonance (NMR) and the Pair Distribution Function (PDF) method. These are examples of instances of the unassigned Distance Geometry Problem (uDGP), where the aim is to calculate the position of some points using a list of associated distance values not previoulsy assigned to the pair of points. We propose new mathematical programming formulations and a new heuristic to solve the uDGP related to molecular structure calculations. In aditton to theoretical results, computational experiments are also provided.

1 Introduction

Distance Geometry (DG) started when Menger characterized geometric concepts using the idea of distance [35]. In the majority of applications of DG, the input data consists of an incomplete list of distance values and the output

Michigan State University
Natural Science Building, 288 Farm Lane, East Lansing, USA.
E-mail: duxbury@pa.msu.edu

IMECC - Unicamp
Cidade Universitaria Zeferino Vaz, Campinas, Brazil.
E-mail: clavor@unicamp.br

CNRS LIX
Ecole Polytechnique, 91128 Palaiseau, France E-mail: liberti@lix.polytechnique.fr.

Federal University of Sao Paulo
Science and Technology Institute, Sao Jose dos Campos, Brazil. E-mail: luiz.leduino@unifesp.br.

is a set of positions in some Euclidean space realizing those given distances [6]. Other applications are given in [36, 39].

When distances are pre-assigned to pairs of objects, we have the *assigned Distance Geometry Problem* (aDGP), also called just DGP [28, 37], defined as follows:

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

$$\forall \{u, v\} \in E, \ ||x_i - x_j||_2 = \delta_{i,j}, \tag{1}$$

where $x_i = x(i)$, $x_j = x(j)$, and $\delta_{i,j} = \delta(\{i, j\})$.

Depending on the application, the embedding space can be very general, but for problems related to molecular geometry, we will fix it to $\mathbb{R}^3$. For example, 3D protein structures and nanostructures can be obtained by exploiting distance information between atom pairs provided by experimental techniques, such as Nuclear Magnetic Resonance (NMR) [28] and the Pair Distribution Function (PDF) method [15], respectively.

In general, in the context of molecular conformations, it is considered that the graph G is known *a priori*, but the information that is actually given by NMR experiments and PDF methods consists of a list of distance values that are only subsequently assigned to atom pairs [7].

In other words, while the distance is given, we do not know the two vertices having such a distance. That is, the associated graph is actually unknown and the only input is a vertex set and a list of distance values. This is the *unassigned Distance Geometry Problem* (uDGP), which has received much less attention than the aDGP [4, 6].

The formal definition of the uDGP is the following (since $\delta_{i,j} = \delta_{j,i}$, we will write $\{i, j\}$ instead of (i, j)):

Definition 2 Given a set of vertices V and a list of associated distance values $d_1, ..., d_m$, find an injective function $g : \{1, ..., m\} \to V \times V$ and a function $x \colon V \to \mathbb{R}^3$ such that, $\forall \{i, j\} \in g(\{1, ..., m\})$,

$$\delta_{i,j} = d_{g^{-1}(\{i,j\})} \tag{2}$$

and

$$||x_i - x_j||_2 = \delta_{i,j}.$$

Note that g is an assignment function that defines a set $E \subset V \times V$, the edges of the graph associated to the uDGP.

For historical notes and surveys on methods to solve DGP's, see [29] and [6, 7, 28], respectively. Recent books are given in [21, 22, 30].

In 1979, Saxe proved that the aDGP is NP-hard [40]. The uDGP is even more challenging because the graph itself and the graph realization both must be determined at the same time.

Although the uDGP is not new to mathematics [41], the literature focuses predominantly on one-dimensional problems motivated by DNA sequencing, often called partial digest problems [12].

For nanostructure calculations, there are two heuristics that have been proposed: TRIBOND [11,14] and LIGA [15]. Both methods are based on build-up approaches and suppose that sufficient distance constraints are available to ensure a unique solution at each step of the procedure.

In the context of molecular geometry, we propose mathematical programming formulations for the uDGP, one of the open problems in DG mentioned in [31]. In addition to theoretical results related to these formulations, we also propose a new heuristic for the problem (Section 2). Section 3 presents computational experiments and Section 4 concludes the paper with new research directions.

2 New Formulations for the uDGP

In this section, we present mathematical programming models for the uDGP, with associated theoretical results, and a heuristic to solve it.

2.1 Mathematical Programming Models

To take into account the assignment function g (Definition 2), we introduce binary variables $a_{i,j}^k$ such that

$$a_{i,j}^k = 1 \Leftrightarrow \text{distance } d_k \text{ is assigned to the pair } (i,j) \in V \times V.$$

Considering vertices $v_1, ..., v_n \in V$ and distance values $d_1, ..., d_m$ related to a uDGP instance, we propose the following model to the uDGP:

$$
\begin{aligned}
\min_{x_1,...,x_n, a_{i,j}^k} \quad & \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(\sum_{k=1}^{m} \left(a_{i,j}^k \left(||x_i - x_j||_2^2 - d_k^2 \right)^2 \right) \right) \\
& \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} a_{i,j}^k = 1, \ k = 1, ..., m, \\
\text{s.t.} \quad & \sum_{k=1}^{m} a_{i,j}^k \leq 1, \ i = 1, ..., n-1, \ j = i+1, ..., n, \\
& x_i \in \mathbb{R}^3, \ a_{i,j}^k \in \{0,1\}, \\
& k = 1, ..., m, \ i = 1, ..., n-1, \ j = i+1, ..., n.
\end{aligned}
\tag{3}
$$

Our first result states a relationship between a uDGP solution and a solution to model (3).

Theorem 1 *A pair (g, x) is a solution for a uDGP instance associated to a graph $G = (V, E)$, with $|V| = n$, $|E| = m$, $g : \{1, ..., m\} \to V \times V$, and $x : V \to \mathbb{R}^3$, if and only if (x, a) is a global optimal solution to (3).*

Proof If (g, x) is a solution for a given uDGP associated to a graph $G = (V, E)$, with $|V| = n$, $|E| = m$, $g : \{1, ..., m\} \to V \times V$, $x : V \to \mathbb{R}^3$, and distance values $d_1, ..., d_m$ related to binary variables $a_{i,j}^k$, such that

$$||x_i - x_j||_2 = \delta_{i,j},$$

where $\delta_{i,j} = d_{g^{-1}(\{i,j\})}$, then it is easy to see that (x, a) is a global optimal solution to (3).

Considering now that $(x_1, ..., x_n, a_{i,j}^k)$ is a global optimal solution to (3), for $i = 1, ..., n - 1$, $j = i + 1, ..., n$, and $k = 1, ..., m$, we have that

$$\sum_{k=1}^{m} \left(a_{i,j}^k \left(||x_i - x_j||_2^2 - d_k^2 \right)^2 \right) = 0$$

and, by constraints of problem (3), the values $a_{i,j}^k$ assign pairs $(i, j) \in V \times V$ such that

$$||x_i - x_j||_2 = d_k, \; k = 1, ..., m. \tag{4}$$

This implicitly defines a weighted graph $G = (V, E, d)$, $d : E \to \mathbb{R}$, with vertices and edges related to $x_1, ..., x_n$ and pairs (i, j), respectively, an injective function $g : \{1, ..., m\} \to V \times V$, such that

$$\delta_{i,j} = d_{g^{-1}(\{i,j\})},$$

and a realization of G, $x : V \to \mathbb{R}^3$, that satisfies (4). Thus, (g, x) is a solution of the uDGP associated to the distance values $d_1, ..., d_m$ and the graph $G = (V, E, d)$. □

In order to avoid the huge number of binary variables of the model (3) and inspired by the *Solid Isotropic Material with Penalization* (SIMP) method [5] and the ideas proposed in [34], we introduce a new formulation with only continuous variables:

$$\min_{t, x_1, ..., x_n, a_{i,j}^k} t - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(\sum_{k=1}^{m} \left(a_{i,j}^k \right)^2 \right)$$

$$\text{s.t.} \quad \begin{aligned} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(\sum_{k=1}^{m} \left(a_{i,j}^k \left(||x_i - x_j||_2^2 - d_k^2 \right)^2 \right) \right) &= t, \\ \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} a_{i,j}^k &= 1, \; k = 1, ..., m, \\ \sum_{k=1}^{m} a_{i,j}^k &\leq 1, \; i = 1, ..., n-1, \; j = i+1, ..., n, \\ t \geq 0, \; x_i \in \mathbb{R}^3, \; 0 \leq a_{i,j}^k &\leq 1, \\ k = 1, ..., m, \; i = 1, ..., n-1, \; j &= i+1, ..., n. \end{aligned} \tag{5}$$

The next result gives a relationship between a uDGP solution and a solution to model (5).

Theorem 2 *A pair (g, x) is a solution for a feasible uDGP instance associated to a graph $G = (V, E)$, with $|V| = n$, $|E| = m$, $g : \{1, ..., m\} \to V \times V$, and $x : V \to \mathbb{R}^3$, if and only if (t, x, a) is a global optimal solution to (5) with globally optimal objective function value equal to $-m$.*

Proof If (g, x) is a solution for a given uDGP associated to a graph $G = (V, E)$, with $|V| = n$, $|E| = m$, $g : \{1, ..., m\} \to V \times V$, $x : V \to \mathbb{R}^3$, and distance values $d_1, ..., d_m$ related to binary variables $a_{i,j}^k$, such that

$$\|x_i - x_j\| = \delta_{i,j},$$

where $\delta_{i,j} = d_{g^{-1}(\{i,j\})}$, we obtain, from Theorem 3, that (x, a) is a global optimal solution to (3). Considering this solution in model (5), we have $a_{i,j}^k \in \{0, 1\}$ and

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(\sum_{k=1}^{m} \left(a_{i,j}^k \left(\|x_i - x_j\|^2 - d_k^2 \right)^2 \right) \right) = 0,$$

which implies that $t = 0$ and that $(0, x, a)$ is also a global optimum solution to model (5), with globally optimal objective function value equal to $-m$.

Let us consider the other direction of the theorem. For a global optimal solution (t, x, a) of the model (5), if there exist positive integers l_1, l_2, l_3 with $l_1 \le n - 1$, $l_2 \le n$, $l_3 \le m$, such that

$$0 < a_{l_1, l_2}^{l_3} < 1,$$

then

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(\sum_{k=1}^{m} \left(a_{i,j}^k \right)^2 \right) < \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(\sum_{k=1}^{m} a_{i,j}^k \right). \tag{6}$$

Since we are considering a feasible uDGP, for $k = 1, ..., m$,

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} a_{i,j}^k = 1 \Rightarrow \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(\sum_{k=1}^{m} a_{i,j}^k \right) = m,$$

implying that, from (6) and $t \ge 0$,

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(\sum_{k=1}^{m} \left(a_{i,j}^k \right)^2 \right) < m \Rightarrow t - \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left(\sum_{k=1}^{m} \left(a_{i,j}^k \right)^2 \right) > t - m,$$

which is a contradiction, because we already know that $-m$ is the optimal value for model (5). Thus,

$$a_{i,j}^k \in \{0, 1\},$$

for $i = 1, ..., n - 1$, $j = i + 1, ..., n$, and $k = 1, ..., m$. From the constraints of problem (5), the values $a_{i,j}^k$ assign pairs $(i, j) \in V \times V$ such that

$$\|x_i - x_j\| = d_k, \ k = 1, ..., m. \tag{7}$$

This implicitly defines a weighted graph $G = (V, E, \delta)$, $\delta : E \to \mathbb{R}$, with vertices and edges related to $x_1, ..., x_n$ and pairs (i, j), respectively, an injective function $g : \{1, ..., m\} \to V \times V$, such that

$$\delta_{i,j} = d_{g^{-1}(\{i,j\})},$$

and a realization of G, $x : V \to \mathbb{R}^3$, that satisfies (7). Thus, (g, x) is a solution of the uDGP associated to the distance values $d_1, ..., d_m$ and the graph $G = (V, E, d)$.   □

From the proof above, note that model (5) also provides a "certificate" of infeasibility of the uDGP instance if the globally optimal objective function value is strictly greater than $-m$.

2.2 A Heuristic Approach

Model (5) can solve larger instances, compared to model (3), but to solve instances with hundreds of atoms, we propose a new heuristic inspired by the TRIBOND method [14] and model (5).

First, we need to find a "core" (positions in $\mathbb{R}^3$ for five vertices with ten associated distances provided from the list of distance values), solving model (5) considering just five points, and then increase its size by adding one vertex position at a time solving a modification of model (5), where four random points (already fixed) are used in order to find the next position:

1. Find a core $x_1, ..., x_5 \in \mathbb{R}^3$ solving the problem

$$
\begin{aligned}
\min_{t, x_1, ..., x_5, a_{i,j}^k} \quad & t - \sum_{i=1}^{4} \sum_{j=i+1}^{5} \left(\sum_{k=1}^{m} \left(a_{i,j}^k \right)^2 \right) \\
& \sum_{i=1}^{4} \sum_{j=i+1}^{5} \left(\sum_{k=1}^{m} \left(a_{i,j}^k \left(\|x_i - x_j\|_2^2 - d_k^2 \right)^2 \right) \right) = t, \\
\text{s.t.} \quad & \sum_{i=1}^{4} \sum_{j=i+1}^{5} a_{i,j}^k = 1, \ k = 1, ..., m, \\
& \sum_{k=1}^{m} a_{i,j}^k \leq 1, \ i = 1, ..., 4, \ j = i+1, ..., 5, \\
& t \geq 0, \ x_1, ..., x_5 \in \mathbb{R}^3, \ 0 \leq a_{i,j}^k \leq 1, \ k = 1, ..., m.
\end{aligned}
\tag{8}
$$

2. For $i = 6, ..., n$, solve the problem

$$
\begin{aligned}
\min_{t, x_i, a_{i,j}^k} \quad & t - \sum_{k=1}^{m_i} \left(\sum_{j \in J} \left(a_{i,j}^k \right)^2 \right) \\
& \sum_{j \in J} \sum_{k=1}^{m_i} \left(a_{i,j}^k \left(\|x_i - y_j\|_2^2 - d_k^2 \right)^2 \right) = t \\
\text{s.t.} \quad & \sum_{j \in J} a_{i,j}^k = 1, \quad k = 1, ..., m_i, \\
& \sum_{k=1}^{m_i} a_{i,j}^k \leq 1, \ j \in J, \\
& t \geq 0, \ x_i \in \mathbb{R}^3, \ 0 \leq a_{i,j}^k \leq 1, \ k = 1, ..., m_i, \ j \in J,
\end{aligned}
\tag{9}
$$

where $x_i \in \mathbb{R}^3$ is the position to be determined, J is a random set with four indices related to already fixed points $y_j \in \mathbb{R}^3$, $j \in J \subset \{1, ..., i-1\}$, and m_i is the number of available distances.

3. If a set of compatible distances cannot be found for some $i = 6, ..., n$, find a new core (go to Step 1) and restart.

The importance of a core in Step 1 is to allow, with high probability [14], to start correctly the reconstruction of the molecular structure. After finding a core, the geometric idea of Step 2 is to intersect fours spheres [32] (centered at points y_j), which gives one point if there are consistent distance values (radii of the spheres) from the list of distances.

3 Computational Results

We generate uDGP instances in the following way. We consider a sequence of covalently connected atoms indexed by $1, ..., n$. The 3D structure of the instance can be defined in terms of the lengths of the covalent bonds $d_{1,2}, \ldots, d_{n-1,n}$, covalent angles $\theta_{1,3}, \ldots, \theta_{n-2,n}$ (formed by three consecutive atoms), and torsion angles $\omega_{1,4}, \ldots, \omega_{n-3,n}$ (formed by four consecutive atoms).

By fixing the lengths of the covalent bonds ($d_{i-1,i} = 1.0$) and the values of the covalent angles ($\theta_{i-2,i} = 2.0$ radians), a 3D molecular structure is determined by the torsion angles $\omega_{1,4}, \ldots, \omega_{n-3,n} \in [0, 2\pi]$, randomly chosen from the set $\{\frac{\pi}{3}, \pi, \frac{5\pi}{3}\}$. More details about instance generation are given in [16, 20].

Differently from TRIBOND and LIGA methods, instances with incomplete list of distances, i.e. $m < \frac{n(n-1)}{2}$, can be easier to solve by the proposed approach, since there will be many global optimum solutions for the model (5). Thus, in order to guarantee a unique solution, we consider instances with all the distances $d_1, ..., d_m$, where $m = \frac{n(n-1)}{2}$.

We used the software AMPL with the solver Baron 17.4.1 on a Lenovo notebook, with 6 MB RAM and intel celeron 1.6 GHz, to solve the three models proposed: M1 (3), M2 (5), and M3 (9).

For all values of n, we generated 5 random instances according to the procedure described above.

For models M1 and M2, we stopped with $n = 5$ and $n = 10$, respectively, because no solution was found considering 3000 minutes as the limit time (see Table 1).

For each n, Table 2 shows the average of the computational time, in minutes, necessary to solve all the 5 instances randomly generated.

From Tables 1 and 2, we notice that the proposed heuristic for solving problem (5) finds the global optimum solutions, in all the cases, in a reasonable time. This means that, during the execution of the heuristic, for a core found in Step 1, problem (9) was solved in Step 2, for all $i = 6, ..., n$.

Table 1 Number of solutions found with time limit = 3000 minutes

n	M1	M2	M3
5	5	5	5
10	0	3	5
20	0	0	5
50	0	0	5
100	0	0	5
200	0	0	5
300	0	0	5
400	0	0	5

Table 2 Computational time in minutes

Number of atoms	M1	M2	M3
5	0.97	0.72	0.72
10	-	2850.03	3.99
20	-	-	9.39
50	-	-	60.52
100	-	-	114.57
200	-	-	256.24
300	-	-	427.87
400	-	-	792.63

4 Conclusions

TRIBOND and LIGA are the first generation methods to solve the uDGP applied to molecular conformation problems and computational results presented in this paper point to different approaches that could be starting points for new research directions.

We are particularly interested in such kind of problems related to protein structure calculations using distance information given by NMR experiments, called the Molecular DGP (MDGP) [28]. Many methods applied to the MDGP suppose that the available distances are pre-assigned to the pairs of atoms. However, as we mentioned in the Introduction, the data that is actually provided by NMR consists of just a list of distances.

The geometry of proteins allows us to define vertex orders $v_1, ..., v_n$ on the associated graph $G = (V, E)$ [8, 25] such that

1. The first four vertices can be fixed in $\mathbb{R}^3$, since they define a clique;
2. Each vertex with rank greater than four is adjacent to at least two contiguous predecessors, *i.e.*

$$\forall i > 4, \{v_{i-2}, v_i\}, \{v_{i-1}, v_i\} \in E.$$

Property 1 can help to define the core (instead of solving problem (8)) and property 2 can be useful in the definition of the set J in the model (9).

In [23], the authors propose a new vertex order for protein molecules where pairs $\{v_{i-3}, v_i\}$ can also be included in the set of edges of the associated graph. When distances $d_{i-3,i}$ are also known, the search space of the problem can be discretized and if the assignment function g is defined in advance, we have the so-called Discretizable MDGP (DMDGP) [17,18], allowing the application of a combinatorial method, called *Branch-and-Prune* (BP) [27].

As mentioned in the recent survey on DGP's [7], experiences with TRI-BOND and LIGA, together with recent results on BP methods for DGP's [13, 33, 26], emphasize the importance of vertex orders in molecular reconstruction from distance information.

Our main research direction now is to consider protein vertex orders in the models proposed in this work to deal with uncertainties in the NMR distance information, already discussed in many approaches to the aDGP [1–3, 9, 10, 19, 24, 38, 42].

The proposed models can also incorporate uncertainties representing distance values as interval distances $[\underline{d}_k, \overline{d}_k]$, $0 < \underline{d}_k \leq \overline{d}_k$, where precise distance values d_k are replaced by $\underline{d}_k + \lambda_k w_k$, with $0 \leq \lambda_k \leq 1$ and $w_k = \overline{d}_k - \underline{d}_k$ [38], implying that λ_k would be new variables and w_k new input data.

Acknowledgements

References

1. 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.
2. 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.
3. A. Baez-Sanchez and C. Lavor, On the estimation of unknown distances for a class of Euclidean distance matrix completion problems with interval data, Linear Algebra and its Applications, 592:287-305, 2020.
4. P. Bartmeyer and C. Lyra, A new quadratic relaxation for binary variables applied to the distance geometry problem, *Structural and Multidisciplinary Optimization*, 62:2197-2201, 2020.
5. M. Bendsoe and O. Sigmund, Topology Optimization: Theory, Methods and Applications, Springer, New York 2003.
6. 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.
7. S. Billinge, P. Duxbury, D. Gonçalves, C. Lavor, and A. Mucherino, Recent results on assigned and unassigned distance geometry with applications to proteinmolecules and nanostructures, *Annals of Operations Research*, 271:161–203, 2018.
8. A. Cassioli, O. Gunluk, C. Lavor, and L. Liberti, Discretization vertex orders in distance geometry, *Discrete Applied Mathematics*, 197:27-41, 2015.
9. T. Costa, H. Bouwmeester, W. Lodwick, and C. Lavor, Calculating the possible conformations arising from uncertainty in the molecular distance geometry problem using constraint interval analysis, *Information Sciences*, 415-416:41-52, 2017.

10. 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.

11. P. Duxbury, L. Granlund, S. Gujarathi, P. Juhas, and S. Billinge, The unassigned distance geometry problem, *Discrete Applied Mathematics*, 204:117-132, 2016.

12. L. Fontoura, R. Martinelli, M. Poggi, and T. Vidal, The minimum distance superset problem: formulations and algorithms, *Journal of Global Optimization*, 72:27–53, 2018.

13. 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.

14. S. Gujarathi, C. Farrow, C. Glosser, L. Granlund, and P. Duxbury, Ab-initio reconstruction of complex Euclidean networks in two dimensions, *Physical Review E*, 89, 053311, 2014.

15. P. Juhás, D. Cherba, P. Duxbury, W. Punch, and S. Billinge, Ab initio determination of solid-state nanostructure, *Nature*, 440:655-658, 2006.

16. C. Lavor, On generating instances for the molecular distance geometry problem, in *Global Optimization: From Theory to Implementation*, L. Liberti and N. Maculan (eds.), Springer, Berlin, 405-414, 2006.

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

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

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*, 56:855-871, 2013.

20. 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.

21. C. Lavor, L. Liberti, W. Lodwick, and T. Mendonça da Costa, An Introduction to Distance Geometry applied to Molecular Geometry, SpringerBriefs, Springer, New York (2017).

22. C. Lavor, S. Xambó-Descamps, I. Zaplana, A Geometric Algebra Invitation to Space-Time Physics, Robotics and Molecular Geometry, SpringerBriefs, Springer, New York (2018).

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. Mariano, D. Gonçalves, and A. Mucherino, Improving the sampling process in the interval Branch-and-Prune algorithm for the discretizable molecular distance geometry problem, *Applied Mathematics and Computation*, 389:125586, 2021.

27. 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.

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

29. L. Liberti and C. Lavor, Six mathematical gems from the history of distance geometry, *International Transactions in Operational Research*, 23:897–920, 2016.

30. L. Liberti and C. Lavor, Euclidean Distance Geometry: An Introduction, Springer, New York (2017).

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

32. D. Maioli, C. Lavor, and D. Gonçalves, A note on computing the intersection of spheres in $\mathbb{R}^n$, *ANZIAM Journal*, 59:271-279, 2017.

33. T. Malliavin, A. Mucherino, C. Lavor, and L. Liberti, Systematic exploration of protein conformational space using a distance geometry approach, *Journal of Chemical Information and Modeling*, 59:4486-4503, 2019.
34. J.M. Martínez, A note on the theoretical convergence properties of the SIMP method, *Structural and Multidisciplinary Optimization*, 29:319-323, 2005.
35. K. Menger, Untersuchungen uber allgemeine Metrik, *Mathematische Annalen*, 100:75-163, 1928.
36. N. Moreira, L. Duarte, C. Lavor, and C. Torezzan, A novel low-rank matrix completion approach to estimate missing entries in Euclidean distance matrix, *Computational and Applied Mathematics*, 37:4989-4999, 2018.
37. A. Mucherino, C. Lavor, L. Liberti, and N. Maculan, eds., Distance Geometry: Theory, Methods, and Applications, Springer, New York (2013).
38. L. Salles Neto, C. Lavor, W. Lodwick, A constrained interval approach to the generalized distance geometry problem, *Optimization Letters,* 14:483-492, 2020.
39. C. Santiago, C. Lavor, S. Monteiro, A. Kroner-Martins, A new algorithm for the small-field astrometric point-pattern matching problem, *Journal of Global Optimization*, 72:55-70, 2018.
40. J. Saxe, Embeddability of weighted graphs in k-space is strongly np-hard, Proc. of the 17th Allerton Conference in Communications, Control and Computing, (1979), 480-489.
41. S. Skiena, W. Smith, and P. Lemke, Reconstructing sets from interpoint distances, Proc. of the Sixth ACM Symposium on Computational Geometry, (1990), 332–339.
42. 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.