
On generating Instances for the Molecular Distance Geometry Problem

Carlile Lavor

Department of Applied Mathematics (IMECC-UNICAMP), State University of Campinas, CP 6065, 13081-970, Campinas-SP, Brazil carlile@ime.unicamp.br

Summary. The molecular distance geometry problem can be stated as the determination of the three-dimensional structure of a molecule using a set of distances between pairs of atoms. It can be formulated as a global minimization problem, where the main difficulty is the exponential increasing of local minimizers with the size of the molecule. The aim of this study is to generate new instances for the molecular distance geometry problem that can be used in order to test algorithms designed to solve it.

Key words: Molecular distance geometry problem, instance generation, NMR spectroscopy.

1 Introduction

The molecular distance geometry problem (MDGP) can be defined as the problem of finding Cartesian coordinates $x_1, \dots, x_N \in \mathbb{R}^3$ of the atoms of a molecule such that

$$\|x_i - x_j\| = d_{i,j} \quad ([i, j] \in S), \quad (1)$$

where S is the set of pairs of atoms $[i, j]$ whose Euclidean distances $d_{i,j}$ are known. If all distances are given, the problem can be solved in linear time [3]. Otherwise, the problem is NP-hard [8].

The distances $d_{i,j}$, in (1), can be obtained, for example, with nuclear magnetic resonance (NMR) data and with knowledge on bond lengths and bond angles of a molecule. Usually, NMR data only provide distances between certain close-range hydrogen atoms [1].

The MDGP can be formulated as a global minimization problem, where the objective function can be given by

$$f(x_1, \dots, x_N) = \sum_{[i,j] \in S} (\|x_i - x_j\|^2 - d_{i,j}^2)^2. \quad (2)$$

It may easily be seen that $x_1, \dots, x_N \in \mathbb{R}^3$ solve the problem if and only if $f(x_1, \dots, x_N) = 0$. For more details and other approaches for the MDGP see, for example, [3, 2, 4, 5, 6].

In [5], Moré and Wu consider the MDGP by applying the global continuation approach to obtain a global minimizer of the function (2). To test the method, they used instances based on a molecule positioned in a three-dimensional lattice.

The aim of this study is to generate new instances for the MDGP which are based on the parameters of a molecular potential energy function defined in [7].

In Section 2, we briefly describe Moré-Wu instances. Section 3 gives the model in which the new instances are based, explains how the instances are generated and presents some examples of the new instances. Section 4 ends with some conclusions.

2 Moré-Wu instances

The instances used in [5] are based on a molecule with s^3 atoms ($s = 1, 2, 3, \dots$) positioned in the three-dimensional lattice defined by

$$\{(i_1, i_2, i_3) \in \mathbb{R}^3 : 0 \leq i_k \leq s-1, k = 1, 2, 3\}.$$

There are two sets of instances. The first one has distances for both near and relatively distant atoms, while the second one only has distances for near atoms.

In the first set of instances, an order is defined for the atoms of the lattice by letting atom i be the atom at position (i_1, i_2, i_3) , where

$$i = 1 + i_1 + si_2 + s^2i_3,$$

and the set S , in (1), is defined by

$$S = \{[i, j] : |i - j| \leq s^2\}.$$

For example, for a molecule with 8 atoms ($s = 2$), the sequence of atoms is

$$\begin{aligned} x_1 &= (0, 0, 0), \\ x_2 &= (1, 0, 0), \\ x_3 &= (0, 1, 0), \\ x_4 &= (1, 1, 0), \\ x_5 &= (0, 0, 1), \\ x_6 &= (1, 0, 1), \\ x_7 &= (0, 1, 1), \\ x_8 &= (1, 1, 1), \end{aligned}$$

and the set S is given by

$$S = \{[1, 2], [1, 3], [1, 4], [1, 5], \\ [2, 1], [2, 3], [2, 4], [2, 5], [2, 6], \\ [3, 1], [3, 2], [3, 4], [3, 5], [3, 6], [3, 7], \\ [4, 1], [4, 2], [4, 3], [4, 5], [4, 6], [4, 7], [4, 8], \\ [5, 1], [5, 2], [5, 3], [5, 4], [5, 6], [5, 7], [5, 8], \\ [6, 2], [6, 3], [6, 4], [6, 5], [6, 7], [6, 8], \\ [7, 3], [7, 4], [7, 5], [7, 6], [7, 8]\}.$$

In the second set of instances, the set S is defined by

$$S = \{[i, j] : \|x_i - x_j\| \leq \sqrt{r}\},$$

where r is a parameter that defines the cutoff value. For example, for $r = 16$, we have

$$S = \{[1, 2], [1, 3], [1, 4], [1, 5], [1, 6], [1, 7], [1, 8], \\ [2, 1], [2, 3], [2, 4], [2, 5], [2, 6], [2, 7], [2, 8], \\ [3, 1], [3, 2], [3, 4], [3, 5], [3, 6], [3, 7], [3, 8], \\ [4, 1], [4, 2], [4, 3], [4, 5], [4, 6], [4, 7], [4, 8], \\ [5, 1], [5, 2], [5, 3], [5, 4], [5, 6], [5, 7], [5, 8], \\ [6, 1], [6, 2], [6, 3], [6, 4], [6, 5], [6, 7], [6, 8], \\ [7, 1], [7, 2], [7, 3], [7, 4], [7, 5], [7, 6], [7, 8]\}.$$

3 New instances

3.1 The background model

The new instances for the MDGP are based on the model proposed by Philips et al. [7]. This model considers a molecule as being a chain of N atoms with Cartesian coordinates given by $x_1, \dots, x_N \in \mathbb{R}^3$. For every pair of consecutive atoms i, j , let r_{ij} be the bond length which is the Euclidean distance between them. For every three consecutive atoms i, j, k , let θ_{ik} be the bond angle corresponding to the relative position of the third atom with respect to the line containing the previous two. Likewise, for every four consecutive atoms i, j, k, l , let ω_{il} be the angle, called the torsion angle, between the normals through the planes determined by the atoms i, j, k and j, k, l . The three-dimensional structure of a molecule is determined by minimizing the sum of the following terms:

$$\begin{aligned}
f_d &= \sum_{[i,j] \in M_1} c_{ij}^r (r_{ij} - r_{ij}^0)^2, \\
f_a &= \sum_{[i,j] \in M_2} c_{ij}^\theta (\theta_{ij} - \theta_{ij}^0)^2, \\
f_\omega &= \sum_{[i,j] \in M_3} c_{ij}^\omega (1 + \cos(n_{ij}\omega_{ij} - \omega_{ij}^0)), \\
f_v &= \sum_{[i,j] \in M_4} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right).
\end{aligned} \tag{3}$$

The terms f_d , f_a , f_ω are the potentials corresponding to bond lengths, bond angles, and torsion angles, respectively. The factor c_{ij}^r is the bond stretching force constant, c_{ij}^θ is the angle bending force constant, and c_{ij}^ω is the torsion force constant. The factors r_{ij}^0 and θ_{ij}^0 represent the equilibrium values for bond length and bond angle, respectively. The constant n_{ij} defines the number of minima involved and the constant ω_{ij}^0 is the phase angle that defines the position of the minima. The term f_v is the Lennard-Jones potential, where A_{ij} and B_{ij} are constants defined by each pair $[i, j]$ and r_{ij} is the Euclidean distance between atoms i and j . The sets M_1, M_2, M_3, M_4 are the sets of pairs of atoms separated by one covalent bond, two covalent bonds, three covalent bonds, and more than two covalent bonds, respectively.

In most conformation calculations, all bond lengths and bond angles are assumed to be fixed at their equilibrium values r_{ij}^0 and θ_{ij}^0 , respectively. Thus, the first three atoms in the chain can be fixed: the first one is fixed at $x_1 = (0, 0, 0)$, the second one is positioned at $x_2 = (-r_{12}, 0, 0)$, and the third one is fixed at $x_3 = (r_{23} \cos(\theta_{13}) - r_{12}, r_{23} \sin(\theta_{13}), 0)$. The fourth atom is determined by the torsion angle ω_{14} ; the fifth atom is determined by the torsion angles ω_{14} and ω_{25} ; the sixth atom is determined by the torsion angles ω_{14} , ω_{25} , and ω_{36} ; and so on.

In that model, the bond lengths and bond angles are set to $r_{ij} = 1.526 \text{ \AA}$ (for all $[i, j] \in M_1$) and $\theta_{ij} = 109.5^\circ$ (for all $[i, j] \in M_2$), respectively. Also, $c_{ij}^\omega = 1$, $n_{ij} = 3$, and $\omega_{ij}^0 = 0$ (for all $[i, j] \in M_3$), providing three “preferred” torsion angles at 60° , 180° , and 300° . Using these parameters, we can generate distances between pairs of atoms and obtain instances for the MDGP.

3.2 Generation of instances

Considering bond lengths r_{ij} and bond angles θ_{ij} fixed ($r_{ij} = 1.526 \text{ \AA}$, $\forall [i, j] \in M_1$; $\theta_{ij} = 109.5^\circ$, $\forall [i, j] \in M_2$), the three-dimensional structure of a molecule can be completely determined by its torsion angles. In [7], the computational results were obtained for problems with 4 through 30 atoms. For a molecule with 18 atoms, for example, 15 torsion angles were obtained:

$$\begin{aligned}
&181^\circ, 176^\circ, 293^\circ, 292^\circ, 165^\circ, 294^\circ, 193^\circ, \\
&166^\circ, 61^\circ, 197^\circ, 65^\circ, 66^\circ, 193^\circ, 67^\circ, 181^\circ.
\end{aligned} \tag{4}$$

Note that these values can be viewed as perturbations of the “preferred” torsion angles 60° , 180° , and 300° , cited above. Thus, based on the model described in the previous subsection, we can generate torsion angles for a molecule, for example, selecting one value ω from the set

$$\{60^\circ, 180^\circ, 300^\circ\} \quad (5)$$

and another one from the set

$$\{\omega + i : i = -15^\circ, \dots, 15^\circ\}. \quad (6)$$

Both of these selections are random.

To generate distances in order to define the set S , in (1), we first obtain Cartesian coordinates for each atom of the chain $(x_{n_1}, x_{n_2}, x_{n_3})$, using the following matrices [7] (note that the first three atoms of the chain are fixed):

$$\begin{bmatrix} x_{n_1} \\ x_{n_2} \\ x_{n_3} \\ 1 \end{bmatrix} = B_1 B_2 \dots B_n \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad (n = 1, \dots, N),$$

where

$$B_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad B_2 = \begin{bmatrix} -1 & 0 & 0 & -r_{12} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$B_3 = \begin{bmatrix} -\cos \theta_{13} & -\sin \theta_{13} & 0 & -r_{23} \cos \theta_{13} \\ \sin \theta_{13} & -\cos \theta_{13} & 0 & r_{23} \sin \theta_{13} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

and $B_i =$

$$\begin{bmatrix} -\cos \theta_{(i-2)i} & -\sin \theta_{(i-2)i} & 0 & -r_{(i-1)i} \cos \theta_{(i-2)i} \\ \sin \theta_{(i-2)i} \cos \omega_{(i-3)i} & -\cos \theta_{(i-2)i} \cos \omega_{(i-3)i} & -\sin \omega_{(i-3)i} & r_{(i-1)i} \sin \theta_{(i-2)i} \cos \omega_{(i-3)i} \\ \sin \theta_{(i-2)i} \sin \omega_{(i-3)i} & -\cos \theta_{(i-2)i} \sin \omega_{(i-3)i} & \cos \omega_{(i-3)i} & r_{(i-1)i} \sin \theta_{(i-2)i} \sin \omega_{(i-3)i} \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

for $i = 4, \dots, N$ ($r_{ij} = 1.526 \text{ \AA}$, $\forall [i, j] \in M_1$; $\theta_{ij} = 109.5^\circ$, $\forall [i, j] \in M_2$; and ω_{ij} , $\forall [i, j] \in M_3$, is obtained according to the rule explained above).

Recall that the problem is to determine the Cartesian coordinates for atoms of a molecule using only a subset of all distances between them. With a prescribed cutoff value d , we can generate an instance for the MDGP defining the set S by

$$S = \{[i, j] : \|x_i - x_j\| \leq d\}.$$

3.3 Examples

Usually, the distance data obtained from NMR experiments are less than or equal to 5 Å [1]. For each example in this subsection, we will select distances with a cutoff value equal to 4 Å.

Using torsion angles defined by (4), from a total of 153 pairs of atoms, 67 pairs were selected (to represent the set S , we consider $[i, j]$ and $[j, i]$ associated to the same pair of atoms):

$$\begin{aligned}
 S = \{ & [1, 2], [1, 3], [1, 4], [1, 14], [1, 15], [1, 16], [1, 17], [1, 18], \\
 & [2, 1], [2, 3], [2, 4], [2, 5], [2, 15], [2, 16], [2, 17], [2, 18], \\
 & [3, 1], [3, 2], [3, 4], [3, 5], [3, 6], [3, 7], [3, 15], [3, 16], [3, 17], [3, 18], \\
 & [4, 1], [4, 2], [4, 3], [4, 5], [4, 6], [4, 7], [4, 17], [4, 18], \\
 & [5, 2], [5, 3], [5, 4], [5, 6], [5, 7], [5, 8], \\
 & [6, 3], [6, 4], [6, 5], [6, 7], [6, 8], [6, 9], \\
 & [7, 3], [7, 4], [7, 5], [7, 6], [7, 8], [7, 9], [7, 10], \\
 & [8, 5], [8, 6], [8, 7], [8, 9], [8, 10], [8, 11], \\
 & [9, 6], [9, 7], [9, 8], [9, 10], [9, 11], [9, 12], [9, 15], [9, 17], \\
 & [10, 7], [10, 8], [10, 9], [10, 11], [10, 12], [10, 13], \\
 & [11, 8], [11, 9], [11, 10], [11, 12], [11, 13], [11, 14], [11, 15], \\
 & [12, 9], [12, 10], [12, 11], [12, 13], [12, 14], [12, 15], \\
 & [13, 10], [13, 11], [13, 12], [13, 14], [13, 15], [13, 16], \\
 & [14, 1], [14, 11], [14, 12], [14, 13], [14, 15], [14, 16], [14, 17], \\
 & [15, 1], [15, 2], [15, 3], [15, 9], [15, 11], [15, 12], [15, 13], [15, 14], [15, 16], [15, 17], [15, 18], \\
 & [16, 1], [16, 2], [16, 3], [16, 13], [16, 14], [16, 15], [16, 17], [16, 18], \\
 & [17, 1], [17, 2], [17, 3], [17, 4], [17, 9], [17, 14], [17, 15], [17, 16], [17, 18], \\
 & [18, 1], [18, 2], [18, 3], [18, 4], [18, 15], [18, 16], [18, 17] \quad \}.
 \end{aligned}$$

We can obtain a better representation of selected pairs of atoms by defining a matrix $A \in \{0, 1\}^{N \times N}$ such that

$$A_{ij} = \begin{cases} 1, & \text{if } [i, j] \in S \\ 0, & \text{if } [i, j] \notin S \end{cases}.$$

We define $A_{ii} = i$ to indicate, at the i -th row, which atoms are close to atom i according to the prescribed cutoff value. For example, the matrix A associated to the set S above is:

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 3 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 4 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 5 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 6 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 7 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 8 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 9 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 2 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 3 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 4 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 5 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 6 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 7 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 8 \end{bmatrix}.$$

Note that, using this matrix, it is easy to see that in addition to the 4-th atom being close to the atoms 1, 2, 3 and 5, 6, 7, it is also close to the last two atoms of the chain. The three-dimensional structure of the molecule defined by the torsion angles (4) is given in Figure 1.

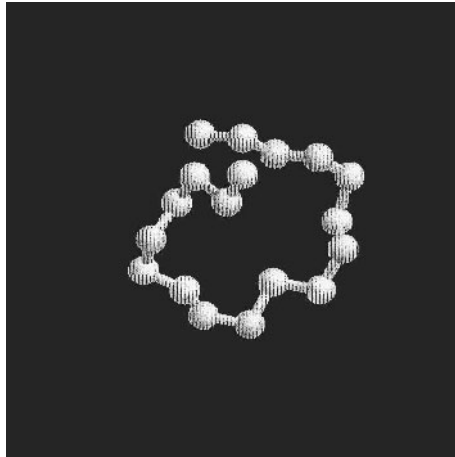


Fig. 1. Molecule associated to the matrix A .

Now we briefly describe three new examples, also with 18 atoms. Using the schema given in subsection 3.2, 10 instances were randomly generated. We consider one with 78 selected pairs of atoms (intermediate number of 10 runs), another with 56 selected pairs of atoms (minimum number of 10 runs),

and the last one with 113 selected pairs of atoms (maximum number of 10 runs). Below, we present the matrix representations A_{int} , A_{min} , and A_{max} , for these three cases. For the first example, we also give the three-dimensional structure of the corresponding molecule (Figure 2).

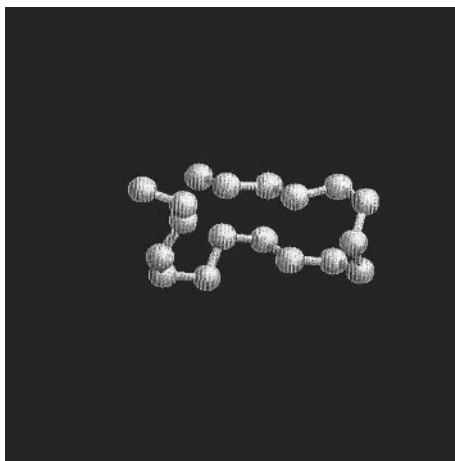


Fig. 2. Molecule associated to the matrix A_{int} .

Instance with 78 selected pairs:

$$A_{int} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 2 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 3 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 4 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 5 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 6 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 7 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 8 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 9 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

Instance with 56 selected pairs:

$$A_{\min} = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 3 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 4 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 5 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 6 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 7 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 8 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 9 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 10 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 11 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 12 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 13 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 14 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 15 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 16 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 17 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 18 \end{bmatrix}.$$

Instance with 113 selected pairs:

$$A_{\max} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 3 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 4 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 5 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 6 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 & 7 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 8 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 9 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 10 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 11 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 12 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 13 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 14 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 15 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 16 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 17 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 18 \end{bmatrix}.$$

4 Conclusion

This study presented a new way to generate instances for the molecular distance geometry problem which is based on the parameters of a molecular potential energy function given in [7].

In [5], Moré and Wu consider the molecular distance geometry problem by applying the global continuation approach where they used instances based on a molecule positioned in a three-dimensional lattice.

The instances generated by the method proposed here have a more flexible geometric conformation, making the corresponding molecules more “realistic”. In the examples of subsection 3.3, we have seen that we can obtain sets of instances with a distinct number of selected pairs of atoms, trying to capture different features in distance data from real problems.

Based on the idea presented here, we can generate many other instances for the molecular distance geometry problem. For example, we can use a potential energy function different from the function (3) and, manipulating its parameters, we can adjust the values of sets (5) and (6) in order to generate instances with distinct characteristics.

Acknowledgments

The author would like to thank FAPERJ and CNPq for their support.

References

1. Creighton, T.E. (1993). *Proteins: Structures and Molecular Properties*. W.H. Freeman and Company, New York.
2. Crippen, G.M. and Havel, T.F. (1988). *Distance Geometry and Molecular Conformation*. John Wiley & Sons, New York.
3. Dong, Q. and Wu, Z. (2002). A linear-time algorithm for solving the molecular distance geometry problem with exact inter-atomic distances. *Journal of Global Optimization*, 22:365-375.
4. Hendrickson, B.A. (1995). The molecule problem: exploiting structure in global optimization. *SIAM Journal on Optimization*, 5:835-857.
5. Moré, J.J. and Wu, Z. (1997). Global continuation for distance geometry problems. *SIAM Journal on Optimization*, 7:814-836.
6. Moré, J.J. and Wu, Z. (1999). Distance geometry optimization for protein structures. *Journal of Global Optimization*, 15:219-234.
7. Phillips, A.T., Rosen, J.B., and Walke, V.H. (1996). Molecular structure determination by convex underestimation of local energy minima. In *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, Volume 23, pages 181-198, Providence, American Mathematical Society.
8. Saxe, J.B. (1979). *Embeddability of weighted graphs in k -space is strongly NP-hard*. In Proc. 17th Allerton Conference in Communications, Control, and Computing, Monticello, IL, pages 480-489.