An overview of distinct approaches for the molecular distance geometry problem

Carlile Lavor
State University of Campinas (IMECC-UNICAMP)
CP 6065, 13081-970, Campinas - SP, Brazil
clavor@ime.unicamp.br

Leo Liberti
LIX, École Polytechnique
F-91128, Palaiseau, France
liberti@lix.polytechnique.fr

Nelson Maculan
Federal University of Rio de Janeiro (COPPE-UFRJ)
CP 68511, 21945-970, Rio de Janeiro - RJ, Brazil
maculan@cos.ufrj.br

Abstract

We present a general overview of some of the most recent approaches for solving the molecular distance geometry problem, namely, the ABBIE algorithm, the DGSOL algorithm, d.c. optimization algorithms, the geometric build-up algorithm, and the BP algorithm.

1 Introduction

The determination of the three-dimensional structure of a molecule, especially in the protein folding framework, is one of the most important problems in computational biology. That structure is very important because it is associated to the chemical and biological properties of the molecule [7, 11, 46]. Basically, this problem can be tackled in two ways: experimentally, via nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography [8], or theoretically, through potential energy minimization [19].

The Molecular Distance Geometry Problem (MDGP) arises in NMR analysis. This experimental technique provides a set of inter-atomic distances d_{ij} for certain pairs of atoms (i, j) of a given protein [23, 24, 33, 57, 58]. The MDGP can be formulated as follows:

> Given a set S of atom pairs (i, j) on a set of m atoms and distances d_{ij} defined over S, find positions $x_1, \ldots, x_m \in \mathbb{R}^3$ of the atoms in the molecule such that
>
> $$||x_i - x_j|| = d_{ij} \quad \forall (i, j) \in S. \tag{1}$$

When the distances between all pairs of atoms of a molecule are given, a unique three-dimensional structure can be determined by a linear time algorithm [16]. However, because of errors in the given distances, a solution may not exist or may not be unique. In addition to this, because of the large scale of problems that arise in practice, the MDGP becomes

very hard to solve in general. Saxe [51] showed that the MDGP is NP-complete even in one spatial dimension.

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

$$f(x_1, \ldots, x_m) = \sum_{(i,j)\in S} (||x_i - x_j||^2 - d_{ij}^2)^2. \tag{2}$$

This function is everywhere infinitely differentiable and has an exponential number of local minimizers. Assuming that all the distances are correctly given, $x \in \mathbb{R}^{3m}$ solves the problem if and only if $f(x) = 0$.

Formulations (1) and (2) correspond to the exact MDGP. Since experimental errors may prevent solution existence (e.g. when the triangle inequality

$$d_{ij} \leq d_{ik} + d_{kj}$$

is violated for atoms i, j, k), we sometimes consider an ϵ-optimum solution of (1), i.e. a solution $x_1, \ldots, x_m$ satisfying

$$|||x_i - x_j|| - d_{ij}| \leq \epsilon \quad \forall (i,j) \in S. \tag{3}$$

Moré and Wu [41] showed that even obtaining such an ϵ-optimum solution is NP-hard for ϵ small enough.

In practice, it is often just possible to obtain lower and upper bounds on the distances [4]. Hence a more practical definition of the MDGP is to find positions $x_1, \ldots, x_m \in \mathbb{R}^3$ such that

$$l_{ij} \leq ||x_i - x_j|| \leq u_{ij} \quad \forall (i,j) \in S, \tag{4}$$

where l_{ij} and u_{ij} are lower and upper bounds on the distance constraints, respectively.

The MDGP is a particular case of a more general problem, called the distance geometry problem [6, 13, 14, 15, 54], which is intimately related to the Euclidean distance matrix completion problem [1, 28, 38].

Several methods have been developed to solve the MDGP, including the EMBED algorithm by Crippen and Havel [12, 25], the alternating projection algorithm by Glunt et al. [20, 21, 22], the multi-scaling algorithm by Trosset et al. [29, 52], a stochastic/perturbation algorithm by Zou, Bird, and Schnabel [59], variable neighborhood search-based algorithms by Liberti, Lavor, and Maculan [35, 39], the ABBIE algorithm by Hendrickson [26, 27], the DGSOL algorithm by Moré and Wu [41, 42, 43, 44, 45], the d.c. optimization algorithms by An and Tao [2, 3], the geometric build-up algorithm by Dong, Wu, and Wu [16, 17, 55], and the BP algorithm by Lavor, Liberti, and Maculan [37]. Two completely different approaches for solving the MDGP are given in [34] (based on quantum computation) and [53] (based on algebraic geometry).

The wireless network sensor positioning problem is closely related to the MDGP, the main difference being the presence of fixed anchor points with known positions: results derived for this problem can often be applied to the MDGP. Amongst the most notable, [18] shows that the MDGP associated to a trilateration graph (a graph with an order on the vertices such that each vertex is adjacent to the preceding 4 vertices) can be solved

in polynomial time; [40] provides a detailed study of Semi Definite Programming (SDP) relaxations applied to distance geometry problems.

The aim of this paper is to present a general overview of some of the most recent approaches for solving the MDGP, namely, the ABBIE algorithm, the DGSOL algorithm, d.c. optimization algorithms, the geometric build-up algorithm, and the BP algorithm.

2 ABBIE algorithm

In [26, 27], Hendrickson describes an approach to the exact MDGP that replaces a large optimization problem, given by (2), by a sequence of smaller ones. He exploits some combinatorial structure inherent in the MDGP, which allows him to develop a divide-and-conquer algorithm based on a graph-theoretic viewpoint.

If the atoms and the distances are considered as nodes and edges of a graph, respectively, the MDGP can be described by a distance graph and the solution to the problem is an embedding of the distance graph in an Euclidean space. When some of the atoms can be moved without violating any distance constraints, there may be many embeddings. The graph is then called flexible or otherwise rigid.

If the graph is rigid or does not have partial reflections, for example, then the graph has a unique embedding. These necessary conditions can be used to find subgraphs that have unique embeddings. The problem can then be solved by decomposing the graph into such subgraphs, in which the minimization problems associated to the function (2) are solved. The solutions found for the subgraphs can then be combined into a solution for the whole graph.

This approach to the MDGP has been implemented in a code named ABBIE and tested on simulated data provided by the bovine pancreatic ribonuclease A, a typical small protein consisting of 124 amino acids, whose three-dimensional structure is known [47]. The data set consists of all distances between pairs of atoms in the same amino acid, along with 1167 additional distances corresponding to pairs of hydrogen atoms that were within 3.5 Å of each other. It was used fragments of the protein consisting of the first $20, 40, 60, 80$ and 100 amino acids as well as the full protein, with two sets of distance constraints for each size corresponding to the largest unique subgraphs and the reduced graphs. These problems have from 63 up to 777 atoms.

3 DGSOL algorithm

In [43], Moré and Wu formulated the exact MDGP in terms of finding the global minimum of a similar function to (2),

$$f(x_1, \ldots, x_m) = \sum_{(i,j)\in S} w_{ij}(||x_i - x_j||^2 - d_{ij}^2)^2, \tag{5}$$

where w_{ij} are positive weights (in numerical results $w_{ij} = 1$ was used).

Following the ideas described in [56], Moré and Wu proposed an algorithm, called DGSOL, based on a continuation approach for global optimization. The idea is gradually transform the function (5) into a smoother function with fewer local minimizers, where an

optimization algorithm is then applied to the transformed function, tracing their minimizers back to the original function. For other works based on continuation approach, see [9, 10, 30, 31, 32, 49].

The transformed function $\langle f\rangle_\lambda$, called the Gaussian transform, of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined by

$$\langle f\rangle_\lambda(x) = \frac{1}{\pi^{n/2}\lambda^n}\int_{\mathbb{R}^n} f(y)\exp\left(-\frac{||y-x||^2}{\lambda^2}\right)dy, \tag{6}$$

where the parameter λ controls the degree of smoothing. The value $\langle f\rangle_\lambda(x)$ is a weighted average of $f(x)$ in a neighborhood of x, where the size of the neighborhood decreases as λ decreases: as $\lambda \to 0$, the average is carried out on the singleton set $\{x\}$, thus recovering the original function in the limit. Smoother functions are obtained as λ increases.

This approach to the MDGP has been implemented and tested on two artificial models of problems, where the molecule has $m = s^3$ atoms located in the three-dimensional lattice

$$\{(i_1, i_2, i_3) : 0 \le i_1 < s, 0 \le i_2 < s, 0 \le i_3 < s\}$$

for an integer $s \ge 1$. In numerical results, it was considered $m = 27, 64, 125, 216$.

In the first model, the ordering for the atoms is specified by letting i be the atom at the position (i_1, i_2, i_3),

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

and the set of atom pairs whose distances are known, S, is given by

$$S = \{(i, j) : |i - j| \le r\}, \tag{7}$$

where $r = s^2$. In the second model, the set S is specified by

$$S = \{(i, j) : ||x_i - x_j|| \le \sqrt{r}\}, \tag{8}$$

where $x_i = (i_1, i_2, i_3)$ and $r = s^2$. For both models, s is considered in the interval $3 \le s \le 6$.

In (7), S includes all nearby atoms, while in (8), S includes some of nearby atoms and some relatively distant atoms.

It was showed that the DGSOL algorithm usually finds a solution from any given starting point, whereas the local minimization algorithm used in the multistart methods is unreliable as a method for determining global solutions. It was also showed that the continuation approach determines a global solution with less computational effort that is required by the multistart approach.

4 D.C. optimization algorithms

In [2, 3], An and Tao proposed an approach for solving the exact MDGP, based on the d.c. (difference of convex functions) optimization algorithms. They worked in $\mathcal{M}_{m,3}(\mathbb{R})$, the space of real matrices of order $m \times 3$, where for $X \in \mathcal{M}_{m,3}(\mathbb{R})$, X_i (resp., X^i) is its ith row (resp., ith column). By identifying a set of positions of atoms $x_1, \ldots, x_m$ with the matrix X, $X_i^T = x_i$ for $i = 1, ..., m$, they expressed the MDGP by

$$0 = \min\left\{\sigma(X) := \frac{1}{2}\sum_{(i,j)\in S, i<j} w_{ij}\theta_{ij}(X) : X \in \mathcal{M}_{m,3}(\mathbb{R})\right\}, \tag{9}$$

where $w_{ij} > 0$ for $i \neq j$ and $w_{ii} = 0$ for all i. The pairwise potential $\theta_{ij} : \mathcal{M}_{m,3}(\mathbb{R}) \to \mathbb{R}$ is defined for problem (1) by either

$$\theta_{ij}(X) = \left(d_{ij}^2 - ||X_i^T - X_j^T||^2\right)^2 \tag{10}$$

or

$$\theta_{ij}(X) = \left(d_{ij} - ||X_i^T - X_j^T||\right)^2, \tag{11}$$

and for problem (4) by

$$\theta_{ij}(X) = \min{}^2\left\{\frac{||X_i^T - X_j^T||^2 - l_{ij}^2}{l_{ij}^2}, 0\right\} + \max{}^2\left\{\frac{||X_i^T - X_j^T||^2 - u_{ij}^2}{u_{ij}^2}, 0\right\}. \tag{12}$$

Similarly to (2), X is a solution if and only if it is a global minimizer of problem (9) and $\sigma(X) = 0$.

While the problem (9) with θ_{ij} given by (11) or (12) is a nondifferentiable optimization problem, it is a d.c. optimization problem.

An and Tao demonstrated that the d.c. algorithms can be adapted for developing efficient algorithms for solving large-scale exact MDGPs. They proposed various versions of d.c. algorithms that are based on different formulations for the problem. Due its local character, the global optimality cannot be guaranteed for a general d.c. problem. However, the fact that the global optimality can be obtained with a suitable starting point motivated them to investigate a technique for computing good starting points for the d.c. algorithms in the solution of (9), with θ_{ij} defined by (11).

The algorithms have been tested on three sets of data: the artificial data from Moré and Wu [43] (with up to 4096 atoms), 16 proteins in the PDB [5] (from 146 up to 4189 atoms), and the data from Hendrickson [27] (from 63 up to 777 atoms). Using these data, they showed that the d.c. algorithms can efficiently solve large-scale exact MDGPs.

5 Geometric build-up algorithm

In [17], Dong and Wu proposed the solution of the exact MDGP by an algorithm, called the geometric build-up algorithm, based on a geometric relationship between coordinates and distances associated to the atoms of a molecule. It is assumed that it is possible to determine the coordinates of at least four atoms, which are marked as fixed; the remaining ones are non-fixed. The coordinates of a non-fixed atom a can be calculated by using the coordinates of four non-coplanar fixed atoms such that the distances between any of these four atoms and the atom a are known. If such four atoms are found, the atom a changes its status to fixed. More specifically, let b_1, b_2, b_3, b_4 be the four fixed atoms whose Cartesian coordinates are already known. Now suppose that the Euclidean distances among the atom a and the atoms b_1, b_2, b_3, b_4, namely d_{a,b_i}, for $i = 1, 2, 3, 4$, are known. That is,

$$\begin{aligned} ||a - b_1|| &= d_{a,b_1}, \\ ||a - b_2|| &= d_{a,b_2}, \\ ||a - b_3|| &= d_{a,b_3}, \\ ||a - b_4|| &= d_{a,b_4}. \end{aligned}$$

Squaring both sides of these equations, we have:

$$\begin{aligned} ||a||^2 - 2a^T b_1 + ||b_1||^2 &= d^2_{a,b_1}, \\ ||a||^2 - 2a^T b_2 + ||b_2||^2 &= d^2_{a,b_2}, \\ ||a||^2 - 2a^T b_3 + ||b_3||^2 &= d^2_{a,b_3}, \\ ||a||^2 - 2a^T b_4 + ||b_4||^2 &= d^2_{a,b_4}. \end{aligned}$$

By subtracting one of these equations from the others, it is obtained a linear system that can be used to determine the coordinates of the atom a. For example, subtracting the first equation from the others, we obtain

$$Ax = b, \tag{13}$$

where

$$A = -2 \begin{bmatrix} (b_1 - b_2)^T \\ (b_1 - b_3)^T \\ (b_1 - b_3)^T \end{bmatrix},$$

$$x = a,$$

and

$$b = \begin{bmatrix} \left(d^2_{a,b_1} - d^2_{a,b_2}\right) - \left(||b_1||^2 - ||b_2||^2\right) \\ \left(d^2_{a,b_1} - d^2_{a,b_3}\right) - \left(||b_1||^2 - ||b_3||^2\right) \\ \left(d^2_{a,b_1} - d^2_{a,b_4}\right) - \left(||b_1||^2 - ||b_4||^2\right) \end{bmatrix}.$$

Since b_1, b_2, b_3, b_4 are non-coplanar atoms, the system (13) has a unique solution.

If the exact distances between all pairs of atoms are given, this approach can determine the coordinates of all atoms of the molecule in linear time [16].

Dong and Wu implemented such an algorithm, but they verified that it is very sensitive to the numerical errors introduced in calculating the coordinates of the atoms. In [55], Wu and Wu proposed the updated geometric build-up algorithm showing that, in this algorithm, the accumulation of the errors in calculating the coordinates of the atoms can be controlled and prevented. They have been tested the algorithm with a set of problems generated using the known structures of 10 proteins downloaded from the PDB data bank [5], with problems from 404 up to 4201 atoms.

6 BP algorithm

In [37], Lavor, Liberti, and Maculan propose an algorithm, called branch-and-prune (BP), based on a discrete formulation of the exact MDGP. They observe that the particular structures of proteins makes it possible to formulate the MDGP applied to protein backbones as a discrete search problem. They formalize this by introducing the discretizable molecular distance geometry problem (DMDGP), which consists of a certain subset of MDGP instances (to which most protein backbones belong) for which a discrete formulation can be supplied. This approach requires that the bond lengths and angles, as well as the distances between atoms separated by three consecutive bond lengths are known.

In order to describe a backbone of a protein with m atoms, in addition to the bond lengths $d_{i-1,i}$, for $i = 2, \ldots, m$, and the bond angles $\theta_{i-2,i}$, for $i = 3, \ldots, m$, it is necessary to consider the torsion angles $\omega_{i-3,i}$, for $i = 4, \ldots, m$, which are the angles between the normals through the planes defined by the atoms $i-3, i-2, i-1$ and $i-2, i-1, i$.

It is known that [48], given all the bond lengths $d_{1,2}, \ldots, d_{m-1,m}$, bond angles $\theta_{13}, \ldots, \theta_{m-2,m}$, and torsion angles $\omega_{1,4}, \ldots, \omega_{m-3,m}$ of a molecule with m atoms, the Cartesian coordinates $(x_{i_1}, x_{i_2}, x_{i_3})$ for each atom i in the molecule can be obtained using the following formulae:

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

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 $i = 4, ..., m$.

Since all the bond lengths and bond angles are assumed to be given in the instance, the Cartesian coordinates of all atoms of a molecule can be completely determined by using the values of $\cos\omega_{i-3,i}$ and $\sin\omega_{i-3,i}$, for $i = 4, ..., m$.

For instances of the DMDGP class, for all $i = 4, ..., m$, the value of $\cos\omega_{i-3,i}$ can be computed by the formula

$$\cos\omega_{i-3,i} = \frac{d_{i-3,i-2}^2 + d_{i-2,i}^2 - 2d_{i-3,i-2}d_{i-2,i}\cos\theta_{i-2,i}\cos\theta_{i-1,i+1} - d_{i-3,i}^2}{2d_{i-3,i-2}d_{i-2,i}\sin\theta_{i-2,i}\sin\theta_{i-1,i+1}}, \qquad (14)$$

which is just a rearrangement of the cosine law for torsion angles [50] (p. 278), and all the values in the expression (14) are given in the instance. This allows to express the position of the i-th atom in terms of the preceding three, giving 2^{m-3} possible conformations, which characterizes the discretization of the problem.

The idea of the BP algorithm is that at each step the i-th atom can be placed in two possible positions. However, either of both of these positions may be infeasible with respect to some constraints. The search is branched on all atomic positions which are feasible with respect to all constraints; by contrast, if a position is not feasible the search scope is pruned.

The algorithm has been tested on the artificial data from Moré and Wu [43] (with up to 216 atoms) and on the artificial data from Lavor [36] (a selection from 10 up to 100 atoms).

7 Conclusion

This paper surveys some of the methods to solve the Molecular Distance Geometry Problem, with particular reference to five existing algorithms: ABBIE, DGSOL, the DCA approach, the geometric build-up algorithm and the BP algorithm.

Acknowledgments

The authors would like to thank CNPq and FAPESP for their financial support.

References

[1] Alfakih, A.Y., Khandani A., and Wolkowicz, H., Solving Euclidean distance matrix completion problems via semidefinite programming, *Computational Optimization and Applications* **12** (1999), 13-30.

[2] An, L.T.H., Solving large-scale molecular distance geometry problems by a smoothing technique via the Gaussian transform and d.c. programming, *Journal of Global Optimization* **27**(2003), 375-397.

[3] An, L.T.H., and Tao, P.D., Large-scale molecular optimization from distance matrices by a d.c. optimization approach, *SIAM Journal on Optimization* **14**(2003), 77-114.

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

[5] Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E., The protein data bank, *Nucleic Acids Research* **28**(2000), 235-242.

[6] Blumenthal, L.M., *Theory and Applications of Distance Geometry*, Oxford University Press, London, 1953.

[7] Brooks III, C.L., Karplus, M., Pettitt, B.M., *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*, Wiley, New York, 1988.

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

[9] Coleman, T.F., Shalloway, D., and Wu, Z., Isotropic effective energy simulated annealing searches for low energy molecular cluster states, *Computational Optimization and Applications* **2** (1993), 145-170.

[10] Coleman, T.F., Shalloway, D., and Wu, Z., A parallel build-up algorithm for global energy minimizations of molecular clusters using effective energy simulated annealing, *Journal of Global Optimization* **4**(1994), 171-185.

[11] Creighton, T.E., *Proteins: Structures and Molecular Properties*, Freeman and Company, New York, 1993.

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

[13] Dattorro, J., *Convex Optimization and Euclidean Distance Geometry*, Meboo Publishing USA, Palo Alto, 2005.

[14] De Leeuw, J., Applications of convex analysis to multidimensional scaling, in Barra, J.R., Brodeau, F., Romier, G., and van Cutsem, B., (eds.), *Recent Developments in Statistics*, pp. 133-145, North-Holland, Amsterdam, 1977.

[15] De Leeuw, J., Convergence of the majorization method for multidimensional scaling, *Journal of Classification* **5** (1988), 163-180.

[16] Dong, Q., and Wu, Z., A linear-time algorithm for solving the molecular distance geometry problem with exact inter-atomic distances, *Journal of Global Optimization* **22**(2002), 365-375.

[17] Dong, Q., and Wu, Z., A geometric build-up algorithm for solving the molecular distance geometry problem with sparse distance data, *Journal of Global Optimization* **26** (2003) 321–333.

[18] Eren, T., Goldenberg, D.K., Whiteley, W., Yang, Y.R., Morse, A.S., Anderson, B.D.O., and Belhumeur, P.N., Rigidity, computation, and randomization in network localization, in *Proc. of IEEE Infocom* (2004), 2673-2684.

[19] Floudas, C.A., and Pardalos, P.M. (eds.), Optimization in Computational Chemistry and Molecular Biology, *Nonconvex Optimization and its Applications*, vol. 40, Kluwer Academic Publishers, The Netherlands, 2000.

[20] Glunt, W., Hayden, T.L., Hong, S., and Wells, J., An alternating projection algorithm for computing the nearest euclidean distance matrix, *SIAM Journal on Matrix Analysis and Applications* **11** (1990), 589-600.

[21] Glunt, W., Hayden, T.L., and Raydan, M., Molecular conformations from distance matrices, *Journal of Computational Chemistry* **14** (1993), 114-120.

[22] Glunt, W., Hayden, T.L., and Raydan, M., Preconditioners for distance matrix algorithms, *Journal of Computational Chemistry* **15** (1994), 227-232.

[23] Gunther, H., *NMR Spectroscopy: Basic Principles, Concepts, and Applications in Chemistry*, Wiley, New York, 1995.

[24] Havel, T.F., An evaluation of computational strategies for use in the determination of protein structure from distance geometry constraints obtained by nuclear magnetic resonance, *Progress in Biophysics and Molecular Biology* **56** (1991), 43-78.

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

[26] Hendrickson, B.A., The Molecule Problem: Determining Conformation from Pairwise Distances, Ph.D. thesis, Cornell University, Ithaca, NY, 1991.

[27] Hendrickson, B.A. , The molecule problem: exploiting structure in global optimization, *SIAM Journal on Optimization* **5** (1995), 835-857.

[28] Huang, H.X., Liang, Z.A., Pardalos, P.M., Some properties for the euclidean distance matrix and positive semidefinite matrix completion problems, *Journal of Global Optimization* **25** (2003) 3–21.

[29] Kearsley, A.J., Tapia, R.A., and Trosset, M.W., The solution of the metric STRESS and SSTRESS problems in multidimensional scaling by Newton's method, *Computational Statistics* **13** (1998), 369-396.

[30] Kostrowicki, J., and Piela, L., Diffusion equation method of global minimization: performance for standard functions, *Journal of Optimization Theory and Applications* **69** (1991), 269-284.

[31] Kostrowicki, J., Piela, L., Cherayil, B.J., and Scheraga, H.A., Performance of the diffusion equation method in searches for optimum structures of clusters of Lennard-Jones atoms, *Journal of Physical Chemistry* **95** (1991), 4113-4119.

[32] Kostrowicki, J., and Scheraga, H.A., Application of the diffusion equation method for global optimization to oligopeptides, *Journal of Physical Chemistry* **96** (1992), 7442-7449.

[33] Kuntz, I.D., Thomason, J.F., and Oshiro, C.M., Distance geometry, in Oppenheimer, N.J., and James, T.L. (eds.), *Methods in Enzymology*, vol. 177, pp. 159-204, Academic Press, New York, 1993.

[34] Lavor, C., Liberti, L., and Maculan, N., Grover's algorithm applied to the molecular distance geometry problem, in *Proc. of VII Brazilian Congress of Neural Networks* (2005).

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

[36] Lavor, C., On generating instances for the molecular distance geometry problem, in Liberti, L., and Maculan, N. (eds.), *Global Optimization: from Theory to implementation*, Springer, Berlin, 2006.

[37] Lavor, C., Liberti, L. and Maculan, N., The discretizable molecular distance geometry problem, in preparation.

[38] Laurent, M., Cuts, matrix completions and a graph rigidity, *Mathematical Programming* **79** (1997), 255-283.

[39] Liberti, L., Lavor, C., and Maculan, N., Double VNS for the molecular distance geometry problem, in *Proc. of MECVNS Conference* (2005).

[40] Man-cho So, A. and Ye, Y., The theory of semidefinite programming for sensor network location, *Mathematical Programming*, Online First, DOI:10.1007/s10107-006-0040-1.

[41] Moré, J.J., and Wu, Z., ε-Optimal solutions to distance geometry problems via global continuation, in Pardalos, P.M., Shalloway, D., Xue, G. (eds.), *Global Minimization of Non-Convex Energy Functions: Molecular Conformation and Protein Folding*, pp. 151-168, American Mathematical Society, Providence, IR, 1996.

[42] Moré, J.J., and Wu, Z., Smoothing techniques for macromolecular global optimization, in Di Pillo, G., and Gianessi, F. (eds.), *Nonlinear Optimization and Applications*, pp. 297-312, Plenum Press, New York, 1996.

[43] Moré, J.J., and Wu, Z., Global continuation for distance geometry problems, *SIAM Journal on Optimization* **7** (1997), 814-836.

[44] Moré, J.J., and Wu, Z., Issues in large scale global molecular optimization, in Biegler, L., Coleman, T., Conn, A., and Santosa, F. (eds.), *Large Scale Optimization with Applications*, pp. 99-122, Springer, Berlin, 1997.

[45] Moré, J.J., and Wu, Z., Distance geometry optimization for protein structures, *Journal of Global Optimization* **15** (1999), 219-234.

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

[47] Palmer, K.A., and Scheraga, H.A., Standard-geometry chains fitted to X-ray derived structures: validation of the rigid-geometry approximation. II. Systematic searches for short loops in proteins: applications to bovine pancreatic ribonuclease A and human lysozyme, *Journal of Computational Chemistry* **13** (1992), 329-350.

[48] Phillips, A.T., Rosen, J.B., and Walke, V.H., Molecular structure determination by convex underestimation of local energy minima, in Pardalos, P.M., Shalloway, D., Xue, G. (eds.), *Global Minimization of Non-Convex Energy Functions: Molecular Conformation and Protein Folding*, pp. 181-198, American Mathematical Society, Providence, IR, 1996.

[49] Piela, L., Kostrowicki, J., and Scheraga, H.A., The multiple-minima problem in the conformational analysis of molecules: deformation of the protein energy hypersurface by the diffusion equation method, *Journal of Physical Chemistry* **93**(1989), 3339-3346.

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

[51] Saxe, J.B., Embeddability of weighted graphs in k-space is strongly NP-hard, in *Proc. of 17th Allerton Conference in Communications, Control, and Computing* (1979), 480-489.

[52] Trosset, M., Applications of multidimensional scaling to molecular conformation, *Computing Science and Statistics* **29**(1998), 148-152.

[53] Wang, L., Mettu, R.R., and Donald, B.R., An algebraic geometry approach to protein structure determination from NMR data, in *Proc. of the 2005 IEEE Computational Systems Bioinformatics Conference* (2005).

[54] Wang, Z., Zheng, S., Boyd, S., and Ye, Y., Further relaxations of the SDP approach to sensor network localization, http://www.stanford.edu/~yyye/relaxationsdp6.pdf, 2006.

[55] Wu, D., and Wu, Z. An updated geometric build-up algorithm for solving the molecular distance geometry problem with sparse distance data, *Journal of Global Optimization*, to appear.

[56] Wu, Z., The effective energy transformation scheme as a special continuation approach to global optimization with application to molecular conformation, *SIAM Journal on Optimization* **6** (1996), 748-768.

[57] Wütrich, K., The development of nuclear magnetic resonance spectroscopy as a technique for protein structure determination, *Accounts of Chemical Research* **22** (1989), 36-44.

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

[59] Zou, Z., Bird, R.H., and Schnabel, R.B., A stochastic/perturbation global optimization algorithm for distance geometry problems, *Journal of Global Optimization* **11** (1997), 91-105.