

The Discretizable Molecular Distance Geometry Problem for Instances Containing Experimental Errors

C. Lavor* A. Mucherino† L. Liberti‡ N. Maculan§

January 31, 2009

The MOLECULAR DISTANCE GEOMETRY PROBLEM (MDGP) is the problem of finding the positions of the atoms of a molecular conformation when only some of the distances between such atoms are known. This problem can be also seen as the problem of finding an immersion in \mathbb{R}^3 of a given undirected and nonnegatively weighted graph $G = (V, E, d)$. In the graph, the set V of vertices represents the set of atoms of the molecule, the set E of edges contains the pairs of atoms whose relative distances are known, and the weights d correspond to the known distances. The most common approach to the MDGP is the formulation of a continuous global optimization, where the objective function measures the differences between the known distances d and the distances computed from the atoms of the possible solutions to the problem. The reader is referred to [1] for a survey on different approaches to the MDGP.

We introduced the DISCRETIZABLE MOLECULAR DISTANCE GEOMETRY PROBLEM (DMDGP) [2, 5], which consists of a certain subset of MDGP instances for which a discrete formulation can be supplied. The instances in the subset must satisfy two assumptions: all the distances between the atoms of each quadruplet $\{v_i, v_{i+1}, v_{i+2}, v_{i+3}\}$ in V must be known, and the atoms in each triplet $\{v_i, v_{i+1}, v_{i+2}\}$ in V cannot be aligned. If such assumptions are satisfied, then the MDGP can be formulated as a combinatorial optimization problem having 2^{n-3} possible solutions, where n is the number of atoms in V . All the possible atomic positions of the conformation can be placed on a binary tree, and this tree can be explored with the aim of finding solutions to the problem.

We proposed a Branch and Prune (BP) algorithm [2, 3, 4, 5] for the solution of the DMDGP. This algorithm is strongly based on the binary tree structure of the combinatorial problem. At each step of the algorithm, two possible atomic positions x_i and x'_i are computed for the i^{th} atom in V . The feasibility of the two computed positions is checked by exploiting the known distances d . Then, infeasible positions are rejected, so that entire branches of the binary tree

*Dept. of Applied Mathematics, State University of Campinas, Campinas-SP, Brazil.

†LIX, École Polytechnique, Palaiseau, France.

‡LIX, École Polytechnique, Palaiseau, France.

§Systems Engineering, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil.

are pruned. This pruning phase allows to reduce the dimension of the binary tree as soon as infeasible positions are discovered. In this way, an exhaustive search on all the other nodes of the tree is not computationally demanding.

In our computational experiences, we are considering instances that are generated from the known conformations of protein molecules and that simulate data obtained from experimental techniques, such as the Nuclear Magnetic Resonance (NMR). Experimental techniques can provide information about the distances between the hydrogens contained into a protein, and the detected distances are usually shorter than 6Å. In order to obtain an instance of the DMDGP and apply the BP algorithm, the two aforementioned assumptions must be satisfied. Unfortunately, they are not satisfied for any order given to the sequence of hydrogens, but particular orderings, if any, must be discovered in order to apply BP.

Moreover, the distances d obtained by experimental techniques are usually affected by errors: there can be uncertainty on the precise value of the distances, and a small percentage of wrong distances could be provided. Therefore, we are investigating possible extensions of the BP algorithm for considering instances whose distances are affected by experimental errors.

References

- [1] C. Lavor, L. Liberti, N. Maculan, *Molecular Distance Geometry Problem*, In: Encyclopedia of Optimization, C. Floudas, and P.M. Pardalos (Eds.), 2nd edition, Springer, NY, 2305–2311, 2009.
- [2] C. Lavor, L. Liberti, N. Maculan, *Discretizable Molecular Distance Geometry Problem*, Tech. Rep. q-bio.BM/0608012, arXiv, 2006.
- [3] C. Lavor, L. Liberti, A. Mucherino, and N. Maculan, *Recent Results on the Discretizable Molecular Distance Geometry Problem*, Proceedings of the conference ROADEF09, Nancy, France, Febraury 10/12, 2009.
- [4] C. Lavor, L. Liberti, A. Mucherino, and N. Maculan, *On a Discretizable Subclass of Instances of the Molecular Distance Geometry Problem*, Proceedings of the Conference SAC09, Honolulu, Hawaii, March 8/12, 2009.
- [5] L. Liberti, C. Lavor, N. Maculan, *A Branch-and-Prune Algorithm for the Molecular Distance Geometry Problem*, International Transactions in Operational Research **15** (1), 1–17, 2008.