The integration of Clifford Algebra in the *i*BP algorithm for the DMDGP

Rafael Alves[1], Andrea Cassioli[2], Antonio Mucherino[3], Carlile Lavor[1], Leo Liberti[3]

[1] IMECC-UNICAMP, Campinas-SP, Brazil.
Email: rafaelsoalves@uol.com.br,clavor@ime.unicamp.br
[2] LIX, École Polytechnique, Palaiseau, France. Email: {cassioli,liberti}@lix.polytechnique.fr
[3] IRISA, University of Rennes 1, Rennes, France. Email: antonio.mucherino@irisa.fr

Finding protein conformations is a fundamental problem in biology. The conformation of such molecules can help understanding the function they perform in living organisms, and bring to important discoveries in biology, as well as in biomedicine.

The Molecular Distance Geometry Problem (MDGP) consists in finding the suitable conformations for a certain molecule which satisfy a set of constraints based on some distances between pairs of its atoms, provided by Nuclear Magnetic Resonance (NMR) experiments [1, 5, 6]. When the distance information is given through a list of lower and upper bounds on the distances, the problem is also referred to as *interval* MDGP (*i*MDGP) [5]. The *i*MDGP, by its nature, is a constraint satisfaction problem, which is NP-hard [8]. Over the years, its solution has been attempted by formulating global optimization problems in continuous spaces [6], where a penalty objective function is generally employed in order to measure the *satisfaction* of the distance constraints for given molecular conformations. More recently, a new class of *i*MDGP instances has been introduced, where the search domain can be reduced to a discrete space having the structure of a tree [2]. We refer to this class of problems as the *interval* Discretizable MDGP (*i*DMDGP).

We developed an exact algorithm named *interval* Branch & Prune (*i*BP) [3, 4] which is potentially able to discover all solutions for a given *i*DMDGP instance. The basic idea is to build a tree of possible atomic positions for the atoms forming the molecule, and to explore such a tree in an efficient way. Branch by branch, new atomic positions are computed and added to the tree, while their feasibility is checked, and, in case of infeasibilities, branches are removed from the tree and not explored. The pruning phase of *i*BP allows to focus the searches on the branches of the tree where there are solutions to the problem.

The tree of atomic positions, which is the search space for *i*BP, is constructed by using discretization distances. When all of them are exact, each node of the tree refers to one atomic position. When one of such discretization distances is instead represented by an interval, a curve in the three-dimensional space can be associated to the tree node. In fact, the intersection among two spheres (related to the two exact distances) and one spherical shell (related to the inexact distance) provides, with probability 1, two disjoint curves (see Figure 1). In this situation, in order to guarantee the discretization, each curve can be replaced by a discrete set of points selected from the two curves, so that only a prederminited number of possible atomic positions is considered [3].

While this strategy was proved to work in practice for relatively small-sized instances, we could observe that the role played by the "predetermined number of pos-

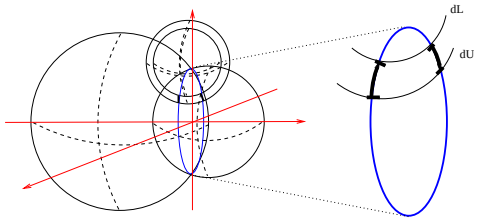

Fig. 1. The intersection among two spheres and one spherical shell

.

sible atomic positions" is extremely important. When too small, it may bring to the identification of an empty set of solutions, because the whole tree is pruned. On the other side, when it is too large, it could lead to a huge number of possible solutions, where many pairs of conformations are very similar to each other.

For this reason, we are working on a methodology for considering the curves defined by the intersections (all the points they contain, not only a discrete subset). To this purpose, we are studying a way to apply Clifford Algebra [7, 9] for managing such curves by their equations. This will help the *i*BP algorithm to identify a complete set of representative solutions for all instances of the *i*DMDGP, as well as to study the flexibility of the corresponding molecular conformations.

Acknowledgments

The authors wish to thank CNPq, the French Embassy in São Paulo, and UNICAMP.

References

1. G.M. Crippen and T.F. Havel, *Distance Geometry and Molecular Conformation*, John Wiley & Sons, New York, 1988.
2. C. Lavor, L. Liberti, N. Maculan, A. Mucherino, *Recent Advances on the Discretizable Molecular Distance Geometry Problem*, European Journal of Operational Research **219**, 698–706, 2012.
3. C. Lavor, L. Liberti, A. Mucherino, *The interval Branch-and-Prune Algorithm for the Discretizable Molecular Distance Geometry Problem with Inexact Distances*, to appear in Journal of Global Optimization, 2012.
4. 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.
5. L. Liberti, C. Lavor, N. Maculan, A. Mucherino, *Euclidean distance geometry and applications*, Tech. Rep. 1205.0349v1 [q-bio.QM], arXiv, 2012.
6. L. Liberti, C. Lavor, A. Mucherino, N. Maculan, *Molecular Distance Geometry Methods: from Continuous to Discrete*, International Transactions in Operational Research **18**, 33–51, 2010.
7. P. Lounesto, *Clifford Algebra and Spinnors*, Cambridge University Press, Cambridge, 1997.
8. J.B. Saxe, *Embeddability of Weighted Graphs in k-Space is Strongly NP-hard*, Proceedings of 17^{th} Allerton Conference in Communications, Control and Computing, 480–489, 1979.
9. G. Sommer, *Geometric Computing with Clifford Algebras*, Springer, New York, 2001.