# Searching for Optimal Orders for Discretized Distance Geometry

D.S. Gonçalves, [a] J. Nicolas, [b] A. Mucherino [c]

[a] *Federal University of Santa Catarina, Florianópolis, Brazil*
douglas@mtm.ufsc.br

[b] *INRIA Rennes Bretagne Atlantique, Rennes, France*
jnicolas@inria.fr

[c] *IRISA, University of Rennes 1, Rennes, France*
antonio.mucherino@irisa.fr

*Key words:* discretization orders, constraint programming, protein backbones

---

The Distance Geometry Problem (DGP) [5] asks whether a simple weighted undirected graph $G = (V, E, d)$ can be embedded in a $K$-dimensional space so that the distances between vertex pairs, which are given by the edge weights, are preserved. DGP has various applications; one of the most important applications arises in biology, where vertices represent atoms of a given molecule, and distances can be obtained by studying the chemical structure of the molecule, as well as by performing NMR experiments [6]. In this application, the dimension of the space is fixed to $K = 3$.

A discretization order is an order on the vertices of $G$ (the atoms of a molecule) that allows to discretize the set of possible embeddings of $G$, that is a priori continuous [3]. Such an order ensures that the first $K$ vertices of $G$ can be embedded uniquely, while all others only admit a finite number of positions in the $K$-dimensional space. Since proteins are very important molecules in biology, suitable discretization orders have been identified for their backbones [4,7]. More recently, however, we found out that, among all possible discretization orders that might exist, there are a few that can bring to the definition of smaller discrete search spaces, which still contain all solutions to the problem.

The search space of a discretizable DGP instance is a tree where, layer by layer, vertex positions for the same $v \in V$ are contained. For all $v > K$ in the given ordering, the discretization is ensured by the existence of $K$ reference vertices, preceding $v$ in the ordering, and for which the relative distances are known. In the biological application, this assumption allows for computing candidate positions for $v$ by intersecting spheres and spherical shells in the

three-dimensional space. The former geometrical object derives from known precise distances, whereas the latter derives from imprecise ones, which are represented by real-valued intervals. If the three reference vertices are related to precise distances, then, for a given triplet of positions for the reference vertices, there are only two candidate positions for $v$, obtained by intersecting three spheres. Otherwise, if we allow only one of these three distances to be imprecise, then this intersection gives 2 disjoint curves, from which we can sample feasible positions [2].

The aim of this work is to find optimized discretization orders, i.e. orders that allow for constructing discrete search spaces that have the minimum number of branches, while all solutions to the problem are preserved. In order to do so, we are working on a constrained multi-objective optimization problem [1], where we have the following list of objectives to be optimized:

- maximization of precise distances used in the discretization (i.e. in the intersections);
- maximization of the use of distances $(H^i_\alpha, H^{i+1})$ in the discretization, where $H^i_\alpha$ is the hydrogen bonded to the Carbon $C_\alpha$ of the amino acid $i$ of the protein, and $H^{i+1}$ is the hydrogen bonded to the Nitrogen of the amino acid $i+1$;
- maximization of the use of pruning distances in upper levels of the tree (closer to the root);
- minimization of the use of interval distances (in the discretization) in lower levels of the tree.

Our main aim is to identify optimal discretization orders for particular classes of instances, such as all instances related to protein backbones. This work represents a little step ahead for the solution of large NMR instances, where the search domain is generally huge in size, and its complete exploration is currently not yet possible. Critical is therefore the minimization of the size of the search domain.

## References


[1] M. Gelfond, *Handbook of Knowledge Representation*, Elsevier, 2007.

[2] D.S. Gonçalves, A. Mucherino, C. Lavor, *An Adaptive Branching Scheme for the Branch & Prune Algorithm applied to Distance Geometry*, to appear in IEEE Conference Proceedings, Federated Conference on Computer Science and Information Systems (FedCSIS14), Workshop on Computational Optimization (WCO14), Warsaw, Poland, September 2014.

[3] C. Lavor, J. Lee, A. Lee-St.John, L. Liberti, A. Mucherino, M. Sviridenko, *Discretization Orders for Distance Geometry Problems*, Optimization Letters **6**(4), 783–796, 2012.

[4] C. Lavor, L. Liberti, A. Mucherino, *The interval Branch-and-Prune Algorithm for the Discretizable Molecular Distance Geometry Problem with Inexact Distances*, Journal of Global Optimization **56**(3), 855–871, 2013.

[5] L. Liberti, C. Lavor, N. Maculan, A. Mucherino, *Euclidean Distance Geometry and Applications*, SIAM Review **56**(1), 3–69, 2014.

[6] T.E. Malliavin, A. Mucherino, M. Nilges, *Distance Geometry in Structural Biology: New Perspectives*. In: "Distance Geometry: Theory, Methods and Applications", A. Mucherino, C. Lavor, L. Liberti, N. Maculan (Eds.), Springer, 329-350, 2013.

[7] A. Mucherino, *On the Identification of Discretization Orders for Distance Geometry with Intervals*, Lecture Notes in Computer Science 8085, F. Nielsen and F. Barbaresco (Eds.), Proceedings of Geometric Science of Information (GSI13), Paris, France, 231–238, 2013.