A distance geometry procedure using the Levenberg-Marquardt algorithm and with applications in biology ~~*but not only*~~

Antonio Mucherino

IRISA, University of Rennes 1

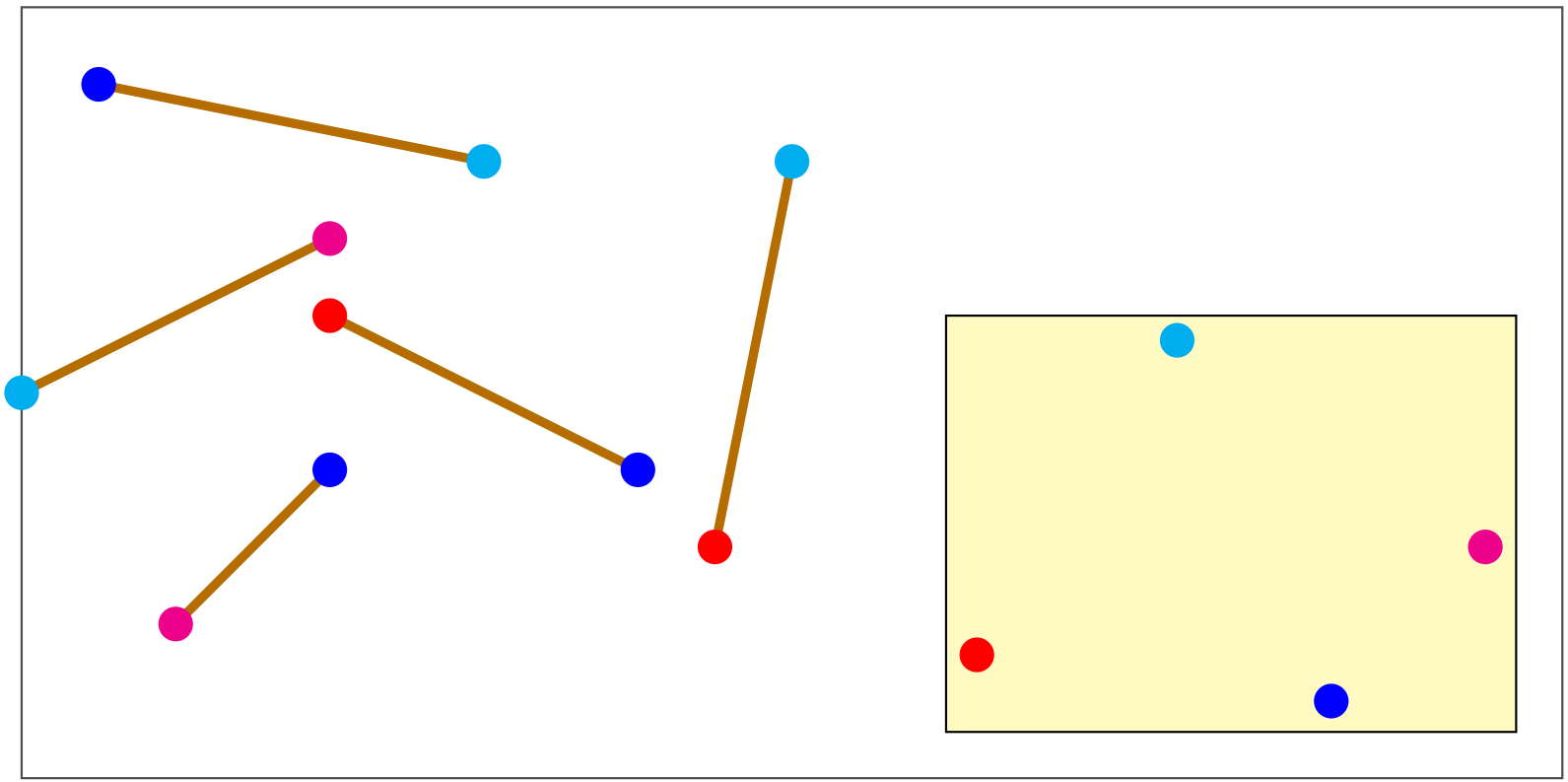
antonio.mucherino@irisa.fr

joint work with Douglas Gonçalves

IWBBIO22, Gran Canaria
June 28th 2022

The Distance Geometry Problem

Can you see how the sticks (left side) are able to define the positions for the points (in the yellow box, right side) ?



The Distance Geometry Problem

Let $G = (V, E, d)$ be a simple weighted undirected graph:

- V represents a set of objects
- E indicates which distances are known
- d provides the "distance information"

Definition

The **DGP** in dimension K.
Determine whether there exists a realization

$$x : V \longrightarrow \mathbb{R}^K$$

of G in $\mathbb{R}^K$ such that, for all edges $\{u, v\} \in E$,

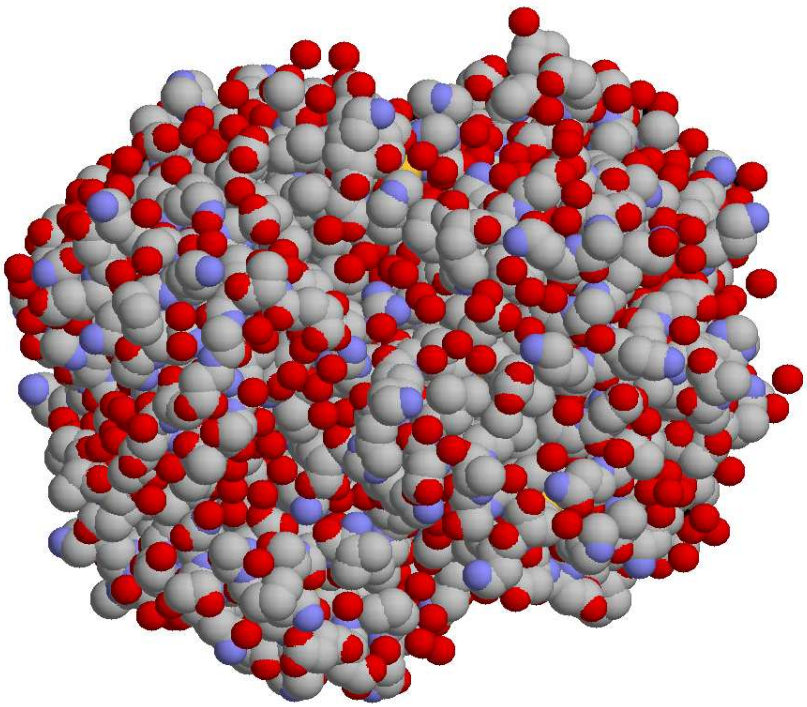
$$||x_u - x_v|| = d(u, v).$$

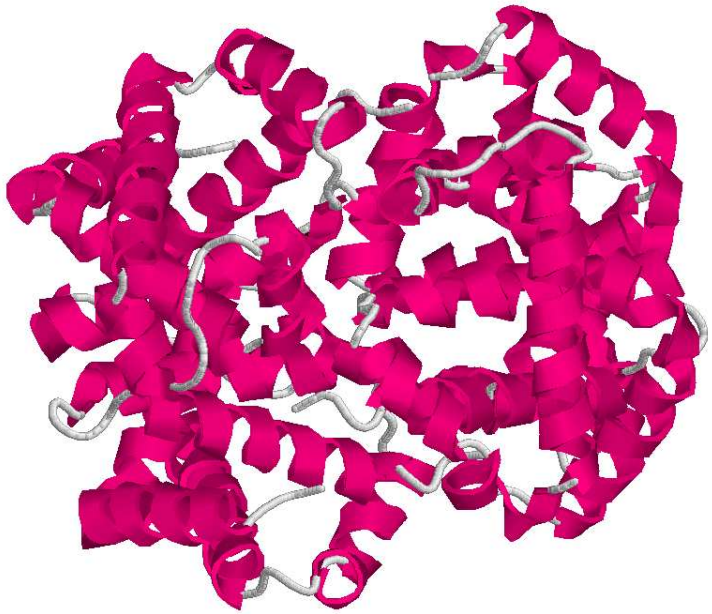
The **DGP** is NP-hard[1].

[1] J. Saxe, Embeddability of Weighted Graphs in k-Space is Strongly NP-hard, Proceedings of 17th Allerton Conference in Communications, Control and Computing, 480–489, 1979.

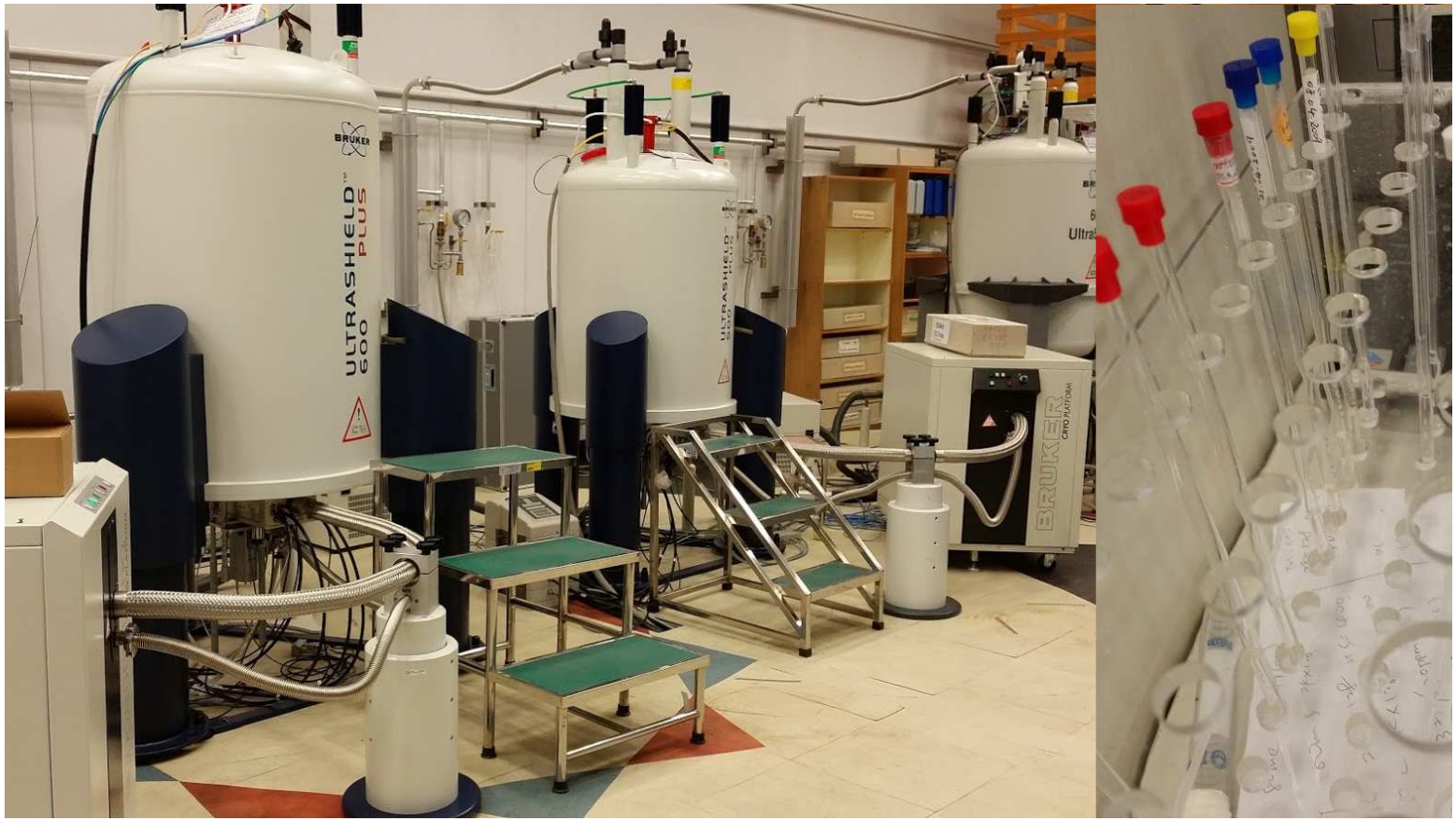
NMR protein structure determination

The hemoglobin:





NMR protein structure determination



Academia Sinica, Taiwan, June 2017.

DGP in structural biology

The DGP is generally reformulated as a *continuous global optimization* problem:

$$\min g(X), \quad \text{where } g(X) = \sum_{\{u,v\}} (||x_u - x_v||^2 - d(u,v)^2)^2$$

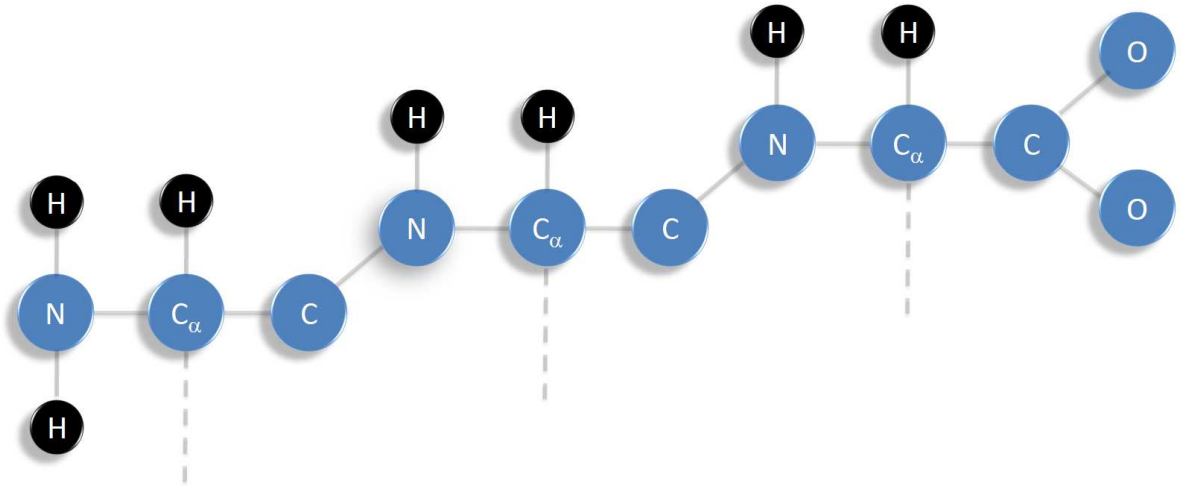
The majority of available NMR models of protein conformations were obtained by solving the corresponding DGP by the meta-heuristic Simulated Annealing (SA).

Some disadvantages:

- SA is not the best meta-heuristic search
- there are no guarantees to converge to the optimal solution(s)
- even with multi-start techniques, there is no hope to identify all optimal solutions

NMR protein structure determination

The chemical composition of a protein backbone:



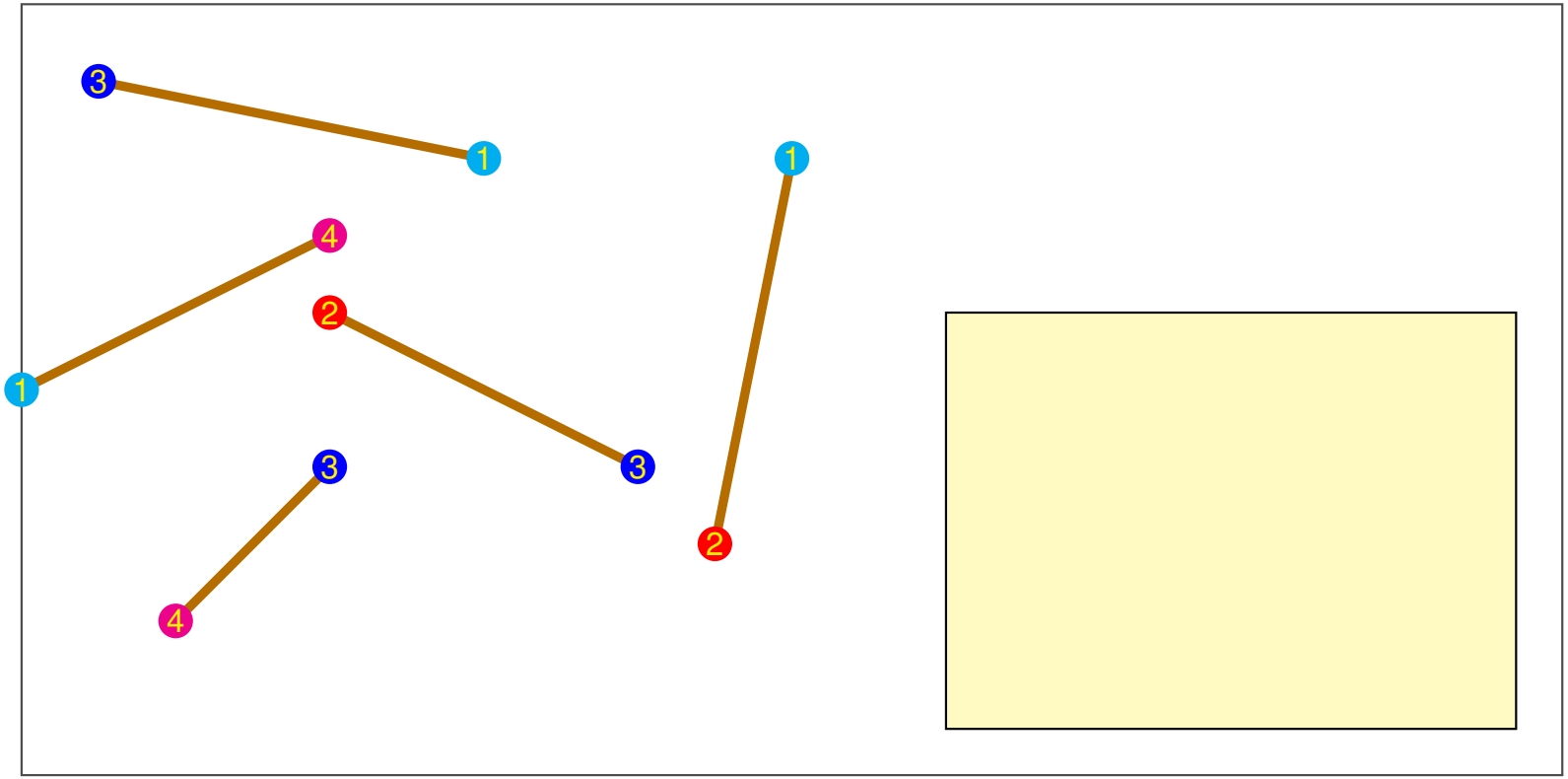
Exact distances can be derived from the chemical composition of the molecule

Interval distances are obtained through the NMR experiments:

- mostly distances between *hydrogens*
- *short* distances ($\leq$ 5Å)
- *large* range for intervals (about $2-3$Å $\Rightarrow$ uncertainty on the distance values)

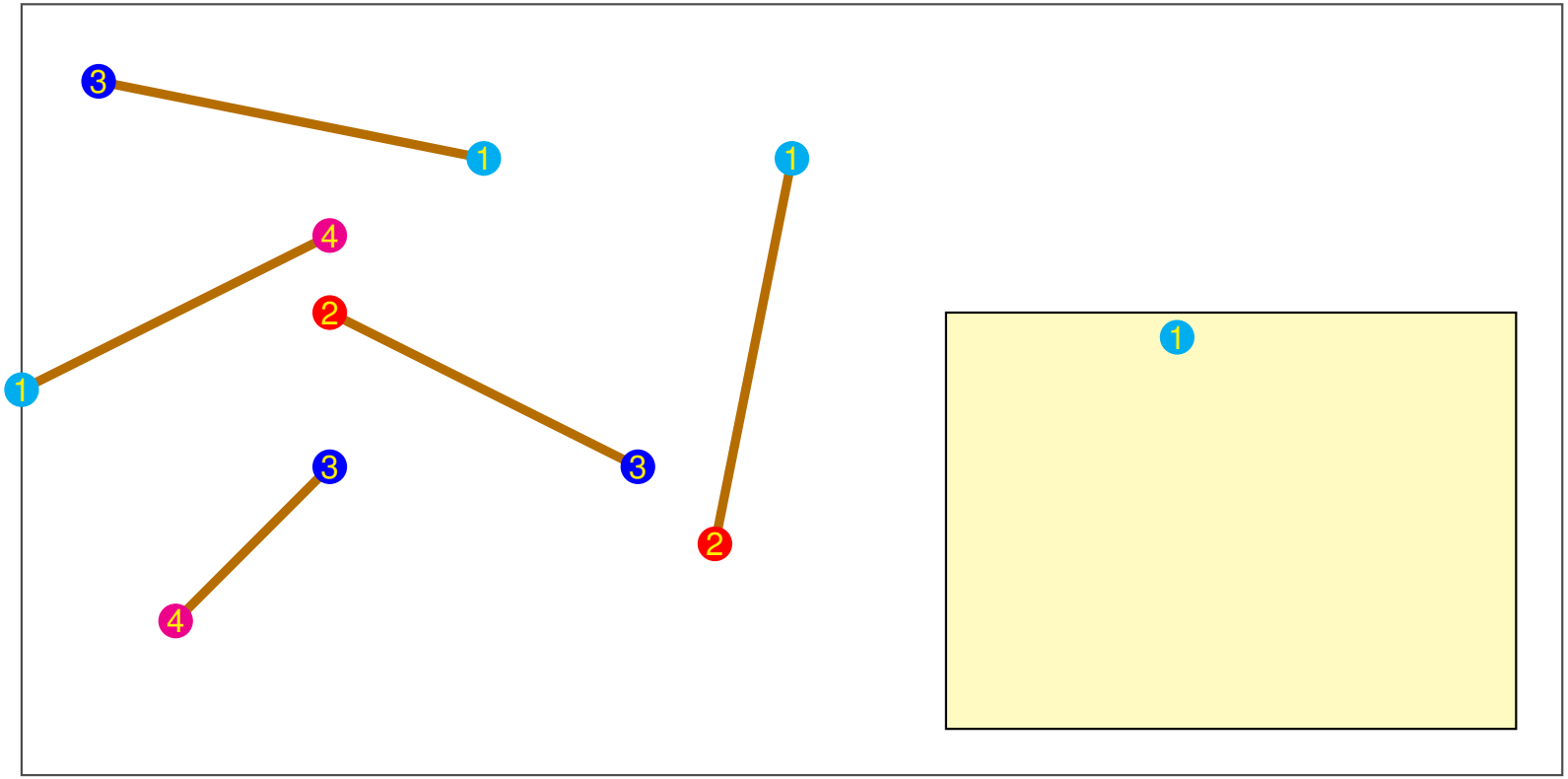
Discretizable DGP

What about assigning a numerical label to every vertex, and to build up the solutions step by step?



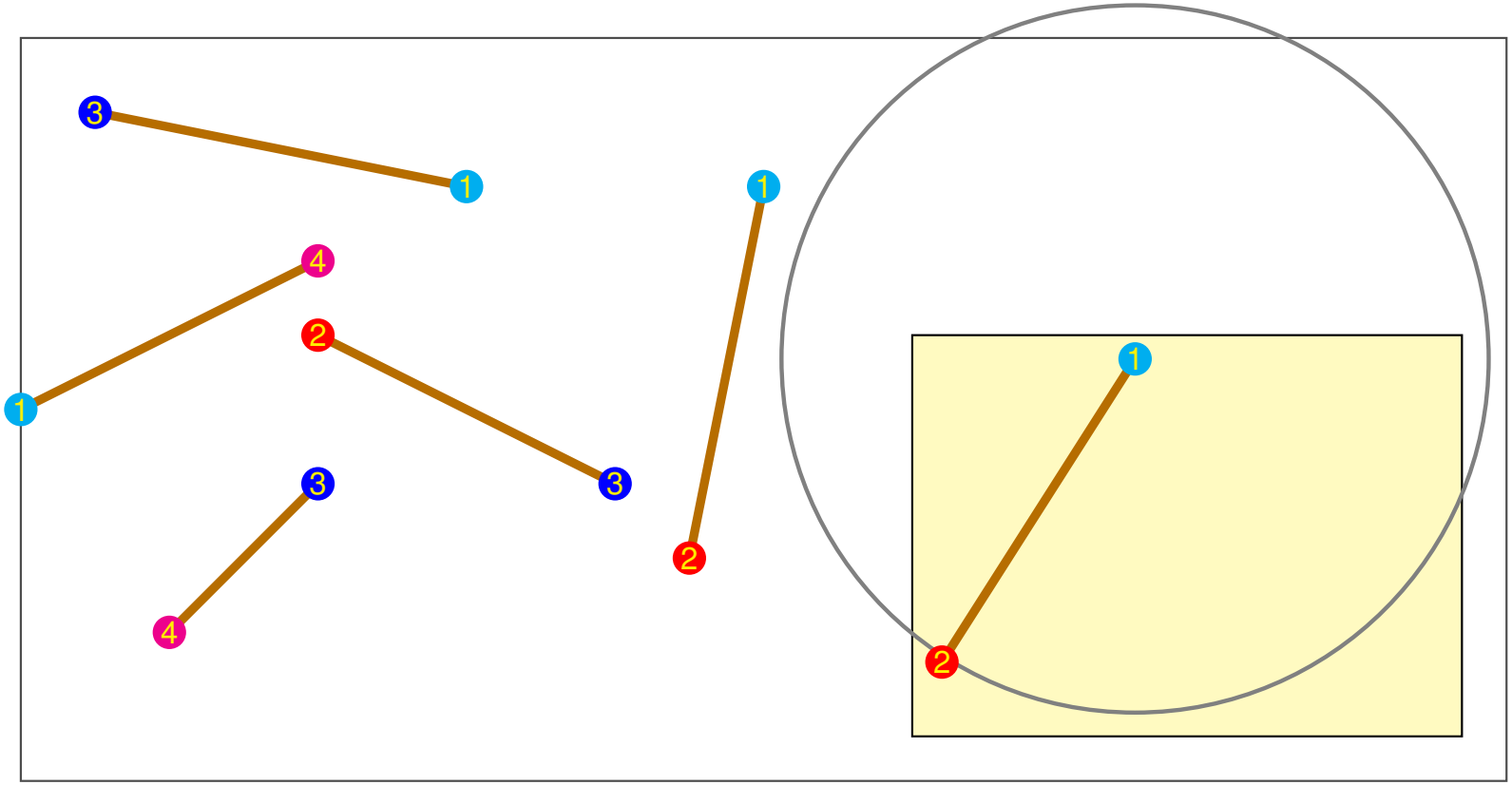
Discretizable DGP

What about assigning a numerical label to every vertex, and to build up the solutions step by step?



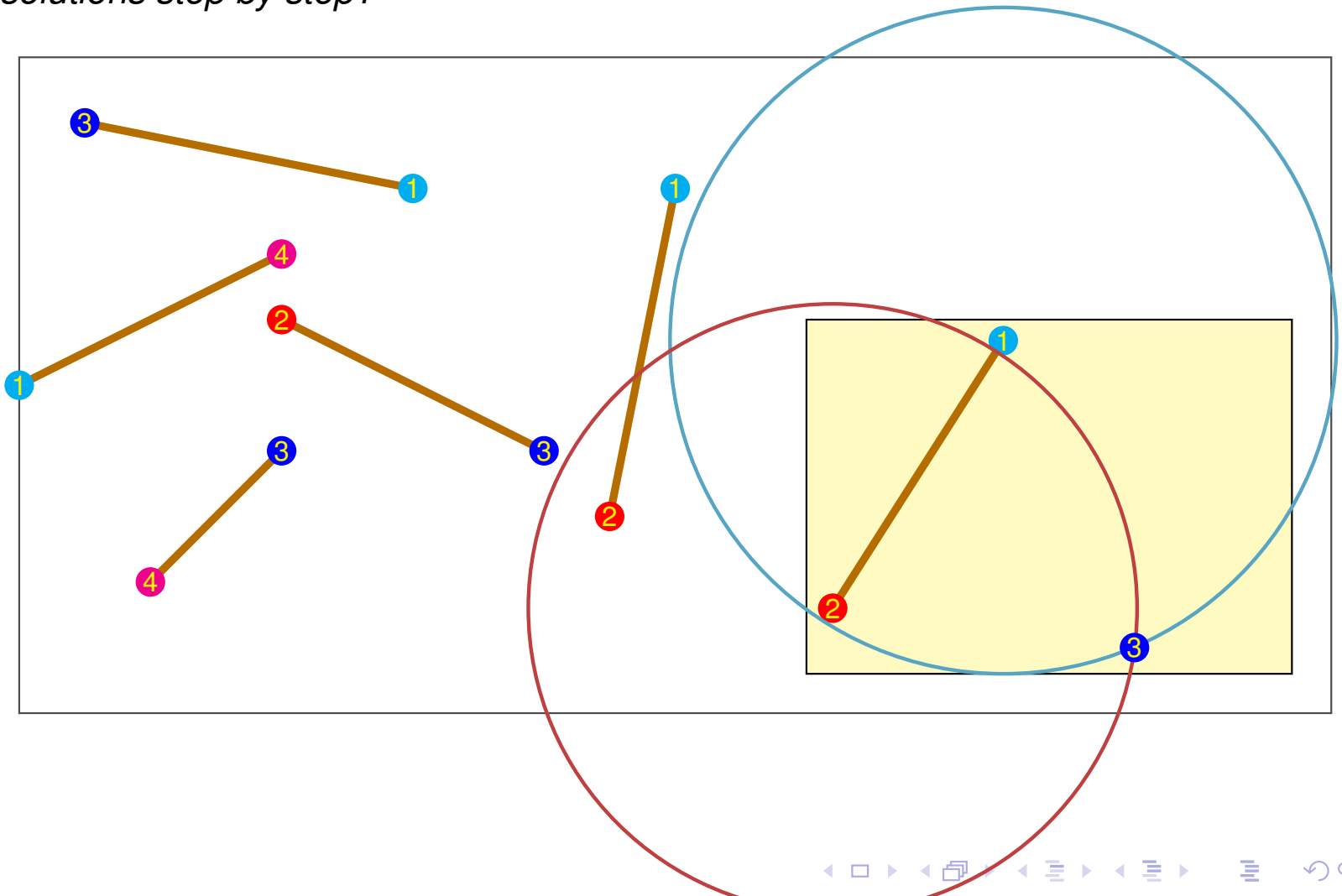
Discretizable DGP

What about assigning a numerical label to every vertex, and to build up the solutions step by step?



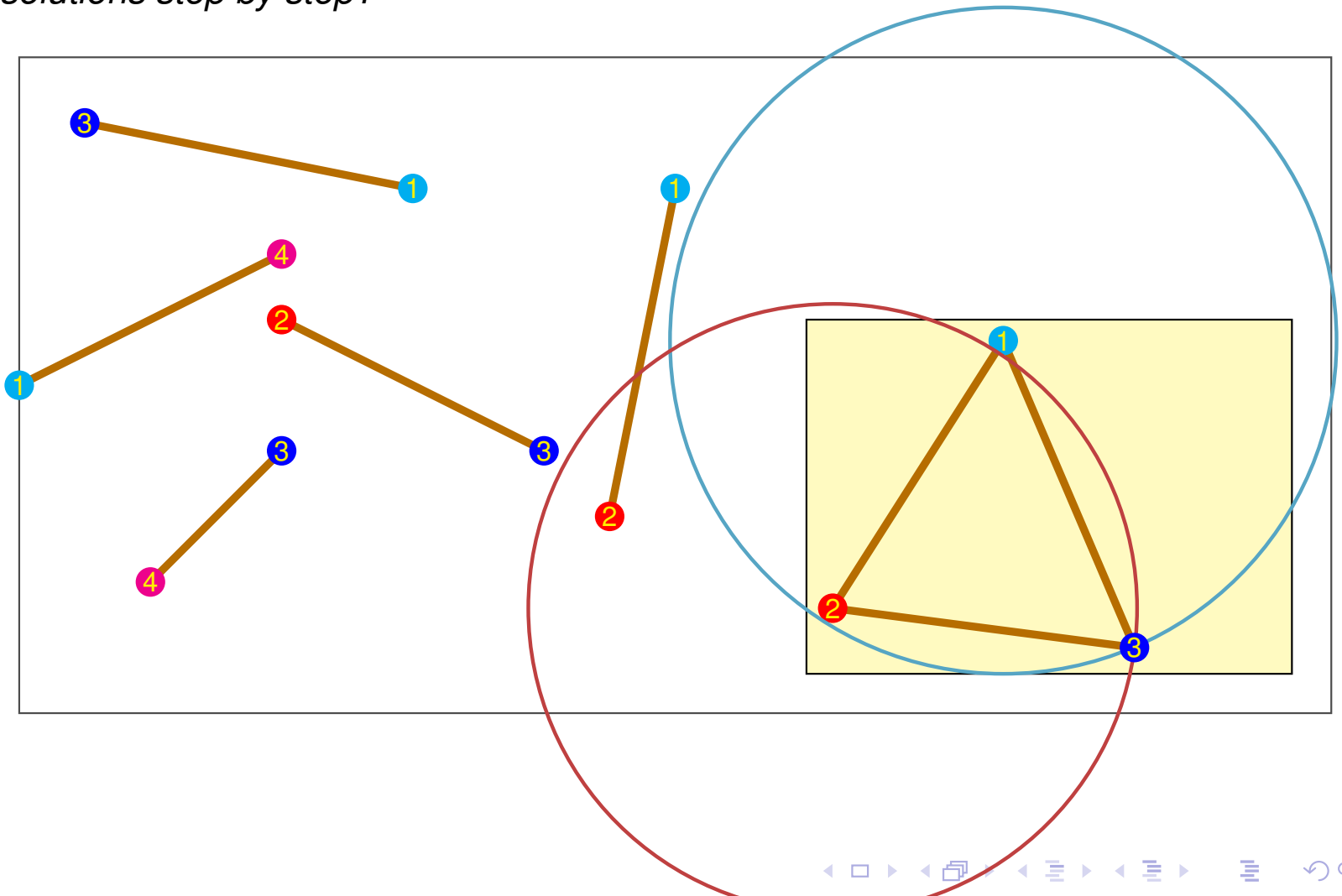
Discretizable DGP

What about assigning a numerical label to every vertex, and to build up the solutions step by step?



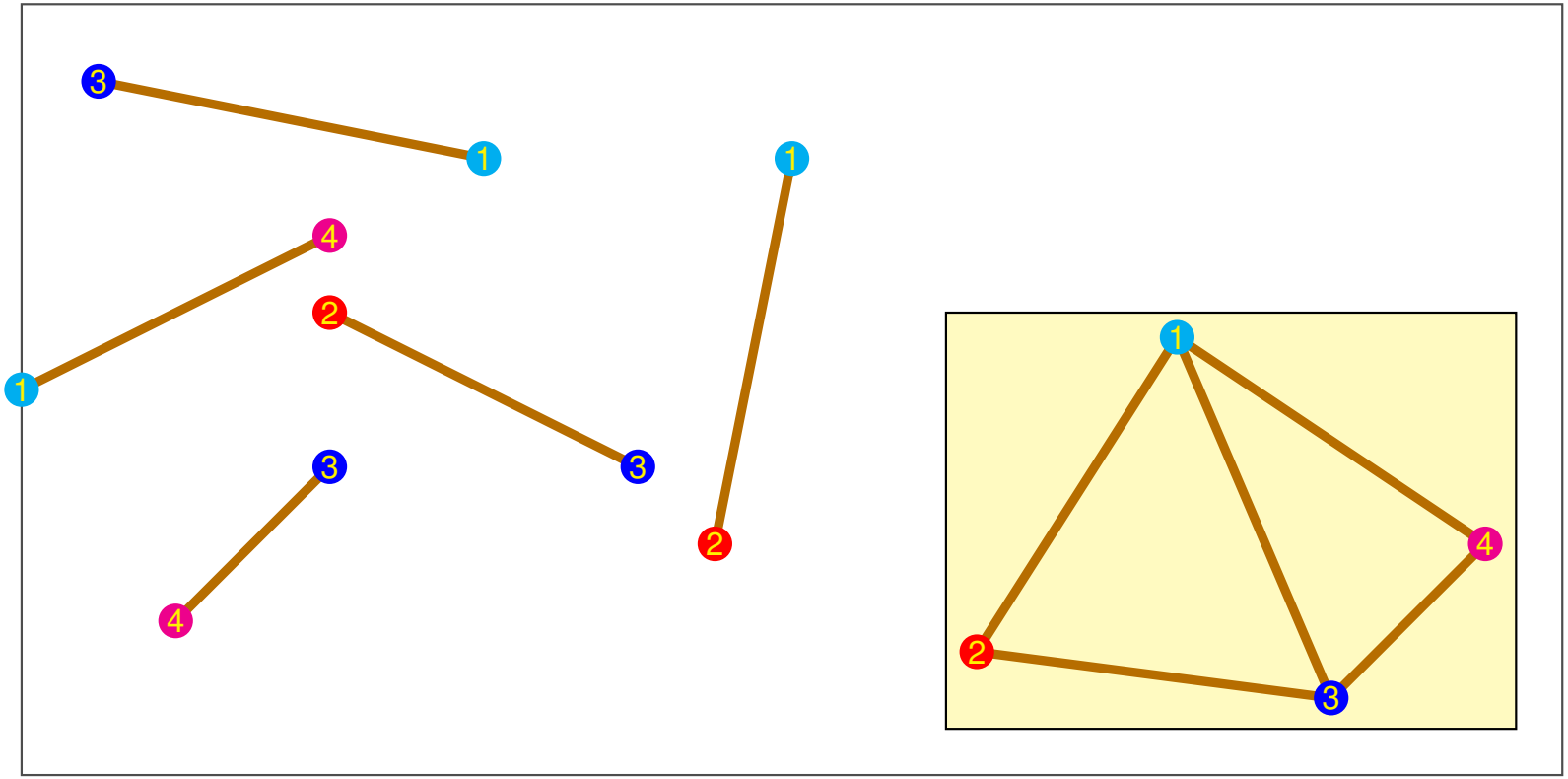
Discretizable DGP

What about assigning a numerical label to every vertex, and to build up the solutions step by step?



Discretizable DGP

What about assigning a numerical label to every vertex, and to build up the solutions step by step?



The intersection of two circles in 2D gives (in general) two points.

Discretizable DGP

Definition

Given a graph $G = (V, E, d)$, there must exist[2] a vertex order[3] on V such that

(A1) $G[\{1, 2, \ldots, K\}]$ is a clique;

(A2) $\forall v \in V : v > K, \quad \exists\, u_1, u_2, \ldots, u_K :$

$$\begin{cases} u_1 < v,\ u_2 < v,\ \ldots\ u_K < v, \\ \{\{u_1, v\}, \{u_2, v\}, \ldots, \{u_K, v\}\} \subset E, \\ \mathcal{V}(S[u_1, u_2, \ldots, u_K]) \neq 0, \end{cases}$$

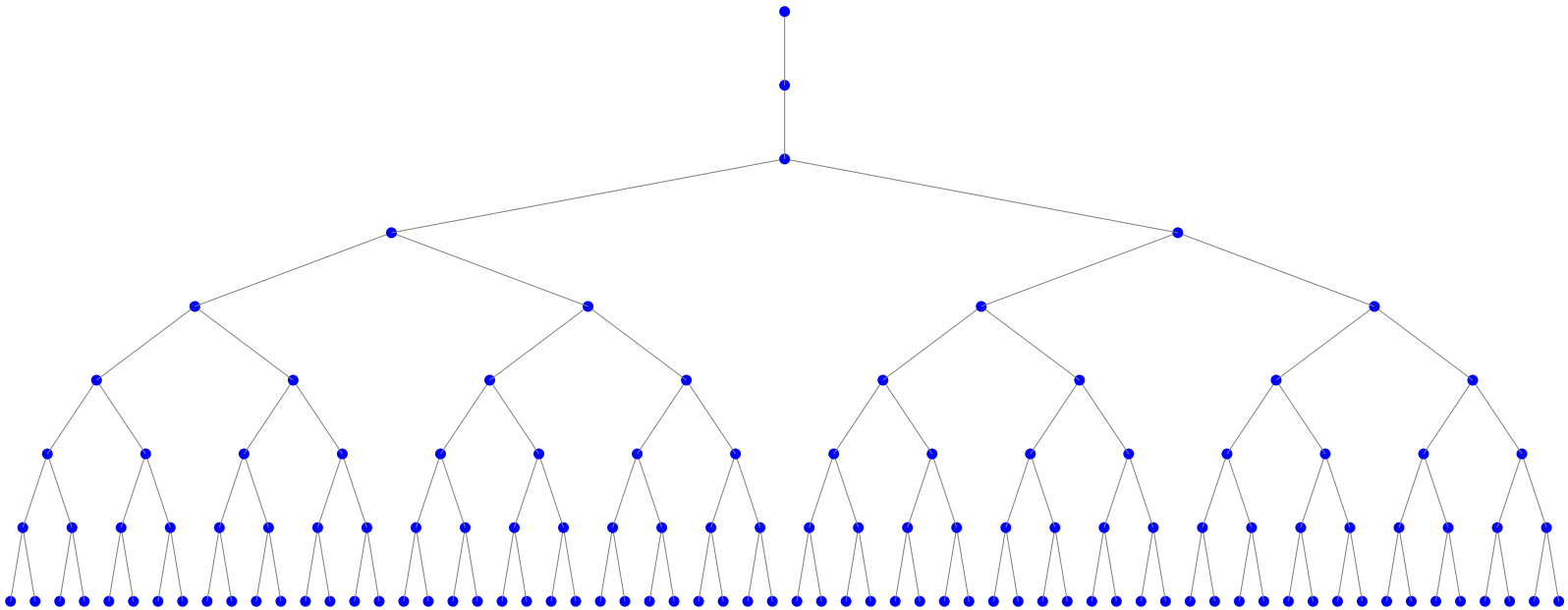
where $S[\ldots]$ is the simplex defined by u_1, u_2, $\ldots$, u_K, and $\mathcal{V}$ is the volume of its argument.

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

[3] J. Omer, A. Mucherino, The Referenced Vertex Ordering Problem: Theory, Applications and Solution Methods, Open Journal of Mathematical Optimization **2**, article no. 6, 29 pages, 2021.

The Branch & Prune algorithm

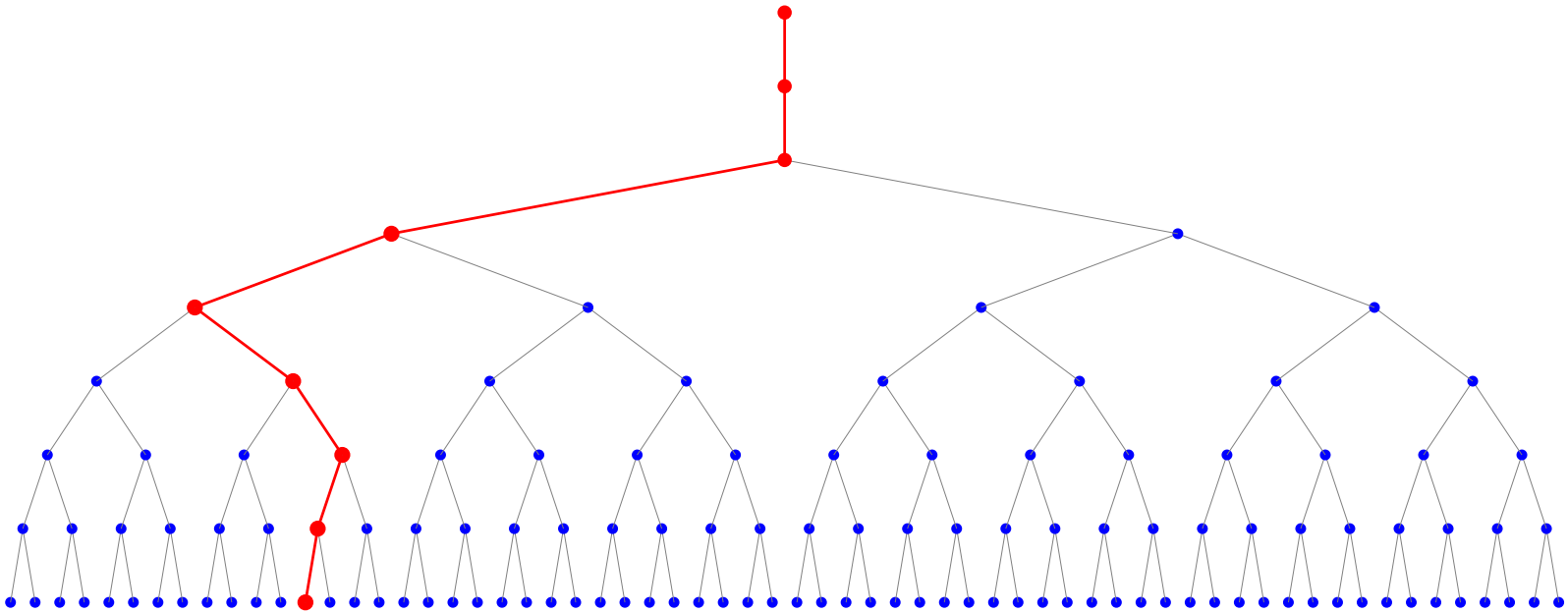
The Branch & Prune (BP) algorithm[4] is based on the idea of branching over all possible positions for each vertex, and of pruning tree branches by exploiting distances not used in the discretization process (*pruning distances*).



[4] C. Lavor, L. Liberti, N. Maculan, A. Mucherino, The Discretizable Molecular Distance Geometry Problem, Computational Optimization and Applications **52**, 115–146, 2012.

The Branch & Prune algorithm

The Branch & Prune (BP) algorithm[4] is based on the idea of branching over all possible positions for each vertex, and of pruning tree branches by exploiting distances not used in the discretization process (*pruning distances*).

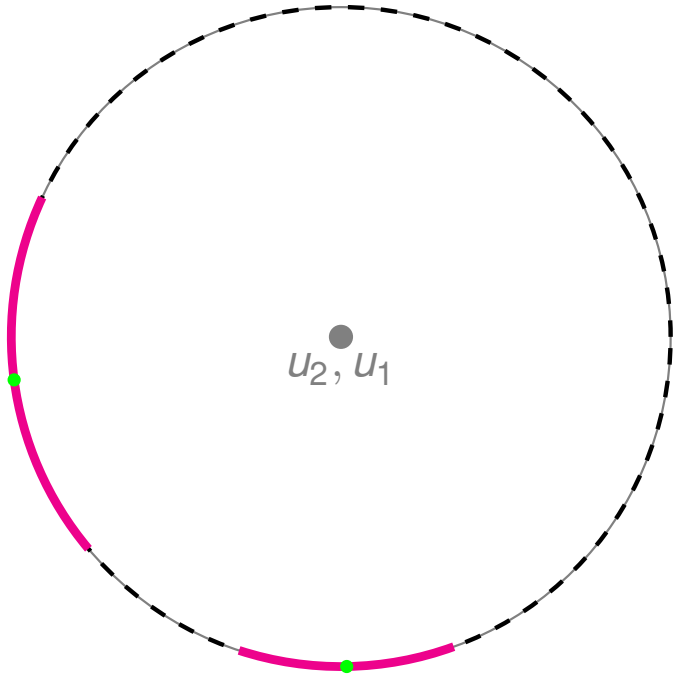


The DDGP is NP-hard $\Longrightarrow$ Exponential worst-case complexity!

[4] C. Lavor, L. Liberti, N. Maculan, A. Mucherino, The Discretizable Molecular Distance Geometry Problem, Computational Optimization and Applications **52**, 115–146, 2012.

A coarse-grained representation

We consider a coarse-grained representation of the search domain.



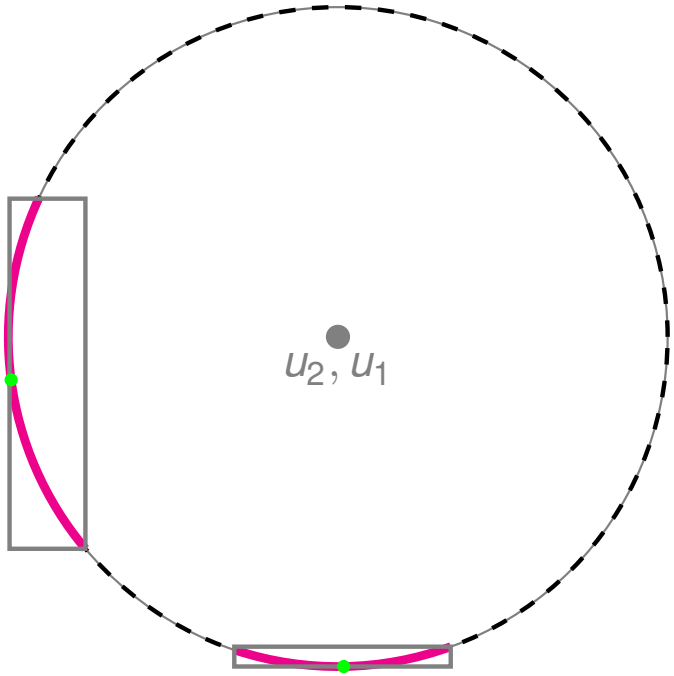
In such a case, a pair (position,box) is associated to a new node of the tree.

The *position* in the *box* can subsequently be refined by *local optimization*.[5]

[5] A. Mucherino, J-H. Lin, D.S. Gonçalves, A Coarse-Grained Representation for Discretizable Distance Geometry with Interval Data, Lecture Notes in Computer Science **11465**, Lecture Notes in Bioinformatics series, I. Rojas et al (Eds.), Proceedings of the 7th International Work-Conference on Bioinformatics and Biomedical Engineering (IWBBIO19), Part I, Granada, Spain, 3–13, 2019.

A coarse-grained representation

We consider a coarse-grained representation of the search domain.



In such a case, a pair (position,box) is associated to a new node of the tree.

The *position* in the *box* can subsequently be refined by *local optimization*.[5]

[5] A. Mucherino, J-H. Lin, D.S. Gonçalves, A Coarse-Grained Representation for Discretizable Distance Geometry with Interval Data, Lecture Notes in Computer Science **11465**, Lecture Notes in Bioinformatics series, I. Rojas et al (Eds.), Proceedings of the 7th International Work-Conference on Bioinformatics and Biomedical Engineering (IWBBIO19), Part I, Granada, Spain, 3–13, 2019.

Local refinement by local optimization

But how to perform the local optimization step?

- spectral projected gradient
- Gauss-Netwon second order method
- Levenberg-Marquardt method[6]

Preliminary experiments are showing that the use of local optimization is very effective in correcting the approximation errors that can propagate along the search domain.

[6] This paper!

Thanks!

antonio.mucherino@irisa.fr

PROJET FINANCÉ PAR L'ANR - PROJECT FUNDED BY THE ANR
ANR