

Métodos para a Determinação da Estrutura Tridimensional de Proteínas e Nanopartículas

Gabriel P. Braun¹, Nelson Maculan²

UFRJ, Rio de Janeiro, RJ

Carlile Lavor³

Unicamp, Campinas, SP

A determinação da estrutura tridimensional de moléculas biológicas e nanopartículas é um problema de grande interesse em química, biologia e áreas relacionadas, já que possui grande impacto nas propriedades físico-químicas dessas substâncias [3]. Através da utilização de dados experimentais obtidos por técnicas como a ressonância magnética nuclear e o método da função de distribuição de pares, torna-se possível determinar estruturas tridimensionais de proteínas e nanoestruturas. Estes são exemplos de instâncias do problema de Geometria de Distâncias Não Associadas (uDGP), onde o objetivo principal é determinar as posições de pontos específicos com base em um conjunto de valores de distância que não foram previamente atribuídos a pares de pontos específicos:

Definição 1.1 (Unassigned Distance Geometry Problem, uDGP [1]). *Dado um inteiro $K > 0$, um conjunto de vértices V e uma lista de valores de distâncias d_1, d_2, \dots, d_m , encontrar uma função objetiva $g : \{1, \dots, m\} \rightarrow V \times V$ e uma função $x : V \rightarrow \mathbb{R}^K$ tal que $\forall \{i, j\} \in g(\{1, \dots, m\})$,*

$$\|x_i - x_j\| = \delta_{ij} \quad (1)$$

e

$$\delta_{ij} = d_{g^{-1}(i,j)} \quad (2)$$

em que $x_i = x(v_i)$, $x_j = x(v_j)$ e $\|x_i - x_j\|$ é a distância euclidiana entre as coordenadas x_i e x_j .

O problema pode ser formulado introduzindo uma variável binária $a_{i,j}^k$ tal que [1]:

$$a_{i,j}^k = 1 \iff \text{a distância } d_k \text{ está associada ao par } (i, j) \in V \times V \quad (3)$$

Em que $a_{i,j}^k$ está sujeito às restrições:

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^n a_{i,j}^k \leq 1 \quad k = 1, 2, \dots, m \quad (4)$$

$$\sum_{k=1}^m a_{i,j}^k = 1 \quad i = 1, \dots, n-1, \quad j = i+1, \dots, n \quad (5)$$

¹braun@cos.ufrj.br

²maculan@cos.ufrj.br

³clavor@unicamp.br

Apresentam-se, neste trabalho, novas formulações de programação matemática e uma abordagem heurística para resolver o uDGP. A formulação desenvolvida consiste em minimizar o erro absoluto na distância relativa à cada aresta y_k :

$$\min \sum_{k=1}^m y_k \quad (6)$$

sujeito às restrições 4, 5 e:

$$y_k \geq \alpha_k, \quad y_k \geq -\alpha_k \quad (7)$$

$$t_{i,j}^2 = \|x_i - x_j\|^2 \quad (8)$$

$$-(1 - a_{i,j}^k)D + t_{i,j} \leq z_{i,j}^k \leq t_{i,j} + (1 - a_{i,j}^k)D \quad (9)$$

$$-a_{i,j}^k D \leq z_{i,j}^k \leq a_{i,j}^k D \quad (10)$$

$$-(1 - a_{i,j}^k)D + (d_k + \alpha_k) \leq z_{i,j}^k \leq (d_k + \alpha_k) + (1 - a_{i,j}^k)D \quad (11)$$

em que $y_k, t_{ij}, z_{ijk} \geq 0$, $x_i, \alpha_k \in \mathbb{R}$, $a_{i,j}^k \in \{0, 1\}$ para $i = 1, \dots, n - 1$, $j = 1 + 1, \dots, n$, $k = 1, 2, \dots, m$, e $D = \max\{d_k\}$.

O modelo desenvolvido foi testado em pequenas instâncias de moléculas artificiais, que se assemelham à estrutura de proteínas [2] e em instâncias de *clusters* de Lennard-Jones, nanopartículas frequentemente utilizadas como *benchmark*. Além disso, foi desenvolvido um método heurístico que usa o modelo apresentado para a resolução iterativa de instâncias maiores.

Nossos resultados demonstram o desempenho superior dos modelos propostos em comparação com os métodos existentes documentados na literatura. Esses modelos têm um imenso potencial para uso prático, oferecendo soluções eficazes para a determinação de estruturas em aplicações envolvendo moléculas, nanopartículas e proteínas.

Referências

- [1] P. Duxbury, C. Lavor, L. Liberti e L. L. de Salles-Neto. “Unassigned distance geometry and molecular conformation problems”. Em: **Journal of Global Optimization** (2022), pp. 1–10.
- [2] C. Lavor. “On generating instances for the molecular distance geometry problem”. Em: **Global Optimization: from Theory to Implementation** (2006), pp. 405–414.
- [3] L. Liberti, C. Lavor, N. Maculan e A. Mucherino. “Euclidean distance geometry and applications”. Em: **SIAM review** 56.1 (2014), pp. 3–69.