Um estudo teórico-computacional da aplicação da Geometria de Distâncias no problema de conformação proteica

Guilherme Philippi¹ UFSC, Blumenau, SC Felipe Fidalgo² MAT/UFSC, Blumenau, SC

A Geometria de Distâncias originou-se dos esforços de Menger (1928), seguido por Blumenthal (1953), na caracterização de vários conceitos geométricos (como congruência e convexidade de conjuntos) em termos de distâncias [4]. Isso permitiu a formulação do Distance Geometry Problem (DGP), conhecido como problema fundamental de Geometria de Distâncias. Trata-se de um problema inverso, onde, dado um grafo simples, ponderado positivamente e não direcionado G = (V, E, d), e um inteiro k > 0, deseja-se encontrar uma função $x : V \to \mathbb{R}^k$ (dita realização de G em \mathbb{R}^k) tal que $\forall \{u, v\} \in E, ||x(u) - x(v)|| = d(\{u, v\})$.

Em particular, a restrição do DGP para k=3 é de interesse prático e é conhecido como $Molecular\ Distance\ Geometry\ Problem\ (MDGP)$ — carrega o termo $molecular\$ pois, mesmo que não seja exclusivo para essa aplicação [4], teve sua origem associada as estruturas moleculares. Supondo que o conjunto solução de um MDGP seja não vazio, sabe-se que ele é não enumerável ou finito [2]. A busca pela segunda possibilidade está associada ao conceito de $ordem\ nos\ vértices$ do grafo G do MDGP (a procura por essa ordem é caracterizado como $Discretization\ Vertex\ Order\ Problem$, ou, DVOP). Munido de tal ordem, o MDGP pode ser discretizado, gerando o problema principal desse trabalho, como segue formalmente definido [2]:

Discretizable Molecular Distance Geometry Problem (DMDGP): Dados um grafo simples, ponderado e não-direcionado G = (V, E, d), onde $d : E \to \mathbb{R}_+$, o subconjunto de vértices $U_0 = \{v_1, v_2, v_3\}$ e uma relação de ordem total $v_1, \ldots, v_{|V|} \in V$, que satisfaça as propriedades

1. U_0 é um 3-clique em G;

```
2. \forall v_i \in V tal que i > 3 nessa ordem: -U_i = \{v_i, v_{i-1}, v_{i-2}, v_{i-3}\} é um 4-clique em G; - vale a designal dade d_{i-3,i-1} < d_{i-3,i-2} + d_{i-2,i-1},
```

encontre uma realização $x: V \to \mathbb{R}^3$ tal que valha, $\forall \{u, v\} \in E, ||x(u) - x(v)|| = d(\{u, v\}).$

A ordenação no DMDGP garante, de fato, a finitude do conjunto solução do problema [2]. Além disso, ela organiza o espaço onde devemos fazer a busca por uma solução. Na verdade, a ordem induz uma estrutura de árvore binária no espaço de busca [1]. Isto é, a partir do quarto, sempre temos duas possibilidades para a realização de um vértice [4]. Analisando essa estrutura, propuseram um algorítimo chamado Branch-and-Prune (BP), que consiste em uma estratégia numérica recursiva para resolver o DMDGP eficientemente, utilizando uma busca combinatória no espaço de busca por soluções. Nele, realiza-se vértice por vértice do grafo, seguindo a ordenação definida pelo DMDGP, "podando" (descartando) todo sub-conjunto solução do problema onde

 $^{^1{\}rm guilherme.philippi@hotmail.com}$

²felipe.fidalgo@ufsc.br

uma realização x(v) não respeita ao menos uma das distâncias $d(\{u,v\})$ definidas pelo grafo, isto é, $||x(u) - x(v)|| \neq d(\{u,v\})$. Desde que esse algoritmo foi publicado, tem se verificado tanto sua beleza matemática quanto a sua eficiência numérica-computacional para resolver problemas em Geometria de Distâncias [1].

Especificamente, neste trabalho, buscou-se usar a Geometria de Distâncias como modelo na conformação molecular de proteínas, em conjunto com a interpretação de dados de Ressonância Magnética Nuclear (RMN) — experimentos que tem como resultado distâncias entre alguns átomos próximos na molécula. Esse é um problema inverso, onde possuímos uma série de distâncias e ângulos e queremos descobrir suas posições, logo, estamos interessados em uma função de realização no espaço \mathbb{R}^3 , descrita pelo MDGP. Mais do que isso, na verdade, estamos interessados em definir uma ordenação nos átomos da molécula (vértices do grafo) de forma a discretizar o problema.

Por sorte, essa ordenação já foi feita. Em [3], os autores propuseram o hand-crafted vertex order (HC), conforme esboça a Figura 1. Note que há uma estrutura que se repete. De fato, existe uma subestrutura nas proteínas chamada Cadeia Principal, que possui uma geometria bem conhecida (a partir de estudos usando cristalografia [1]). Com essas informações, podemos modelar o DMDGP utilizando os átomos da molécula como vértices e as distâncias conhecidas como arestas.

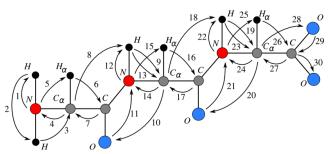


Figura 1: Ordenação HC [3].

Esse estudo teve como produto um software, chamado HCProt, que atua no processamento de dados retirados do repositório Worldwide Protein Data Bank afim de criar instâncias esparsas (usando a ordenação HC) que podem ser usadas como entrada para implementações do algoritmo BP. Automatiza-se, assim, a geração de instâncias DMDGP. Espera-se que este software possa auxiliar o processo de pesquisa dos autores da área. Fora desenvolvido, também, um algoritmo BP para realizar simulações computacionais, obtendo conformações validas de moléculas reais.

Agradecimentos

O presente trabalho foi realizado com o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq – Brasil. Agradeço a Família, ao Felipão e a quem prorrogou a submissão.

Referências

- [1] Fidalgo, F. Dividindo e conquistando com simetrias em geometria de distâncias. Tese de Doutorado, IMECC/UNICAMP, Campinas SP, 2015.
- [2] Lavor, C., Maculan, N., Souza, M. and Alves, R. Álgebra e Geometria no Cálculo de Estrutura Molecular, 31º Colóquio Brasileiro de Matemática. IMPA, Rio de Janeiro RJ, 2017.
- [3] Lavor, C., Liberti, L., Donald, B., Worley, B., Bardiaux, B., Malliavin, T. E. and Nilges, M. Minimal NMR distance information for rigidity of protein graphs, *Discrete Applied Mathematics*, Elsevier, 256:91–104, 2019. DOI:10.1016/j.dam.2018.03.071.
- [4] Liberti, L., Lavor, C., Maculan, N. and Mucherino, A. Euclidean distance geometry and applications, SIAM REVIEW, Society for Industrial and Applied Mathematics, volume 56, number 1, pages 3-69, 2014. DOI. 10.1137/120875909.