

Uma proposta de adaptação do algoritmo Branch-and-Prune usando a interseção de quatro esferas para o Problema de Geometria de Distâncias Moleculares

Clarice de Souza Santos¹, Rosiane de Freitas Rodrigues^{1,2}, Kelson Mota³

¹Programa de Pós-graduação em Informática - PPGI

²Instituto de Computação - IComp, UFAM

³Programa de Pós-graduação em Química - PPGQ
Departamento de Química, UFAM

{rosiane, clarice}@icomp.ufam.edu.br, kelsonmota@ufam.edu.br

Resumo. O problema de determinar a estrutura tridimensional de uma molécula, posicionando no espaço todos os átomos que a compõem, é denominado de Problema de Geometria de Distâncias Moleculares. Determinar tal estrutura geométrica partindo-se de um conjunto arbitrário incompleto de distâncias é um problema NP-difícil computacionalmente, onde conseguir uma solução viável em um tempo de execução razoável é um interessante desafio matemático e computacional. Neste trabalho, está sendo proposto uma adaptação do algoritmo de Branch-and-Prune, considerando o cálculo da interseção de quatro esferas através de sistema de equações lineares de distâncias Euclidianas interatômicas e sistema de matrizes de coordenadas internas da molécula, envolvendo ângulos de ligação e torção. Este é um resultado preliminar de uma pesquisa em andamento, ainda em estágio inicial, mas com promissores resultados, onde, a estrutura 3D parcial de proteínas da base PDB pôde ser determinada em tempo linear.

1. Introdução

Muitas pesquisas em biologia tem como foco as propriedades e atividades das proteínas, que são moléculas fundamentais de sistemas biológicos constituídas por uma cadeia linear de aminoácidos. A função de uma proteína está relacionada tanto a sua composição atômica quanto a sua estrutura geométrica [Dong and Wu 2002], cuja determinação apresenta grandes desafios de pesquisa, uma vez que ainda não se consegue determinar a estrutura tridimensional de uma molécula precisamente. As tentativas para isto envolvem a determinação de modelos teóricos envolvendo a minimização da energia potencial ou por simulação de dinâmica molecular, ou experimentalmente, através de Cristalografia de Raio-X ou de Ressonância Magnética Nuclear (RMN) [Silva and Lavor 2008].

O Problema de Geometria de Distâncias Moleculares - PGDM (do inglês, *Molecular Distance Geometry Problem* - MDGP) pode ser definido como a obtenção de uma configuração tridimensional para a molécula respeitando-se as distâncias euclidianas conhecidas, ou seja, considerando uma molécula formada por n átomos a_1, a_2, \dots, a_n para a qual são conhecidas um conjunto de distâncias d_{ij} entre pares de átomos a_i e a_j , procura-se encontrar as posições x_1, x_2, \dots, x_n para os átomos respeitando as distâncias conhecidas [Lavor et al. 2011].

As distâncias entre pares de átomos, no conjunto de distâncias, podem ser valores exatos ou limites inferiores e superiores. No primeiro caso, o problema é dito ser de distâncias exatas e no segundo, ser de distâncias inexatas ou faixas de distância. Estes dois tipos podem conter as distâncias entre todos os átomos ou somente entre alguns deles e sendo assim o problema pode ser classificado dependendo do conjunto de distâncias em duas formas: conjunto completo de distâncias - todas as distâncias entre quaisquer pares de átomos são conhecidas, e isto resulta em um problema polinomial [Dong and Wu 2002]; conjunto arbitrário de distâncias - são conhecidas somente as distâncias entre alguns átomos da molécula, neste caso o problema é NP-difícil [Saxe 1979].

2. Materiais e Métodos

Existem vários algoritmos para resolver o PGDM, utilizando em sua maioria uma das duas seguintes técnicas matemáticas para resolver tal problema: resolução de sistemas lineares formados a partir das equações das distâncias euclidianas interatômicas; resolução do sistema de matrizes formados a partir dos dados de coordenadas internas da molécula. A grande maioria dos algoritmos resolve a versão contínua clássica do problema tal como o *Geometric Build-up* (GBU) [Dong and Wu 2003], entretanto Lavor et al. propuseram um modelo discreto do problema, o *Discretizable Molecular Distance Geometry Problem* (DMDGP) e também um método de resolução denominado *Branch and Prune* (BP) [Lavor et al. 2011].

Foram implementados os algoritmos BP, para o caso onde se tem previamente apenas um subconjunto arbitrário de distâncias e foi proposta uma adaptação do BP com interseção de quatro esferas. Os algoritmos foram implementados na linguagem de programação C/C++ e o ambiente de teste utilizado envolveu o sistema operacional Ubuntu 11.10, em máquina com processador Intel® Core™ 2 Duo CPU 2.93GHz e memória de 2GB.

O BP original usa a interseção de três esferas para calcular a possível posição do átomo seguinte, gerando duas possibilidades. Como a interseção de quatro esferas é somente um ponto, se for possível encontrar uma distância entre o átomo atual e um já fixado diferente dos três anteriores é possível gerar uma quarta esfera e usar a interseção das quatro como a posição única do átomo.

O algoritmo, apresentado em Algoritmo 1, fixa os quatro primeiros átomos como o BP, mas a partir do quinto átomo x_i , para $i \geq 5$, além de utilizar os três átomos anteriores x_{i-1} , x_{i-2} e x_{i-3} , procura um quarto átomo já fixado x_v que tenha distância conhecida para o átomo x_i , para ser o centro da quarta esfera. Quando a quarta esfera é encontrada, o sistema de equações da interseção de esferas é usado para encontrar a coordenada do átomo. Sempre que for possível encontrar a quarta esfera, a coordenada do átomo é determinada de forma única, assim grande parte da estrutura é conseguida de forma linear como mostrado na figura 1.

3. Resultados

Para testar a acurácia dos valores obtidos pelos algoritmos BP original e BP com quatro esferas, os resultados obtidos foram comparados com as saídas do software *MD-Jeep*,

Algoritmo 1: Algoritmo BP com quatro esferas

```
1 Fixar os átomos  $x_1, x_2, x_3$  e  $x_4$  como o BP original
2 para  $i = 5$  até  $n$  faça
3   | Procurar um átomo  $x_v | v < i - 3$  com distância conhecida para o átomo  $i$ 
4   | Fixar o átomo  $i$  usando os átomos  $x_{i-1}, x_{i-2}, x_{i-3}$  e  $x_v$ 
5 fim para
```

que é um *Branch and Prune* implementado por [Mucherino et al.], para as mesmas entradas. Os experimentos computacionais foram feitos utilizando instâncias do *Protein Data Bank* [PDB 2013] e também o fragmento dA1rC2rG3dC4T5rC6rG7dA8rG9, gerado por processos de mecânica quântica usando o programa Hyperchem.

O algoritmo BP original implementado gerou as mesmas coordenadas que o MD-Jeep para todas as instâncias de testes e foi em alguns casos mais rápido devido a otimizações propostas sobre as estruturas de dados originais.

O método BP com quatro esferas gerou as mesmas coordenadas que o *MD-Jeep* para todas as instâncias de teste em tempo linear em relação ao número de átomos. Contudo, o algoritmo ainda não pode ser usado para determinar a completa estrutura molecular. Isto acontece porque na resolução de sistemas lineares ocorre um acúmulo de erros na determinação da posição, o que impede a aplicação desta técnica para determinar a estrutura da molécula inteira. Uma comparação entre os passos para encontrar as coordenadas dos átomos da molécula de feromônio contida na entrada 2erl é mostrada na Figura 1, onde o algoritmo BP com quatro esferas conseguiu fixar linearmente somente 30 dos 120 átomos da molécula.

A aplicação de métodos que minimizem o acúmulo de erros estão em desenvolvimento para o algoritmo proposto. Também está sendo elaborada uma versão adaptada do método BP envolvendo o cálculo da interseção de quatro esferas através da resolução do sistema de matrizes de coordenadas internas das proteínas, usando ângulos de ligação e torção. Afim de testar a versão adaptada, planeja-se usar instâncias obtidas no laboratório de físico-química da UFAM, de onde se obtém as coordenadas internas de proteínas através de técnicas de mecânica quântica e uso do programa Hyperchem.

4. Conclusões e Perspectivas

Este trabalho abordou o Problema de Geometria de Distâncias Moleculares (PGDM), no qual se tenta resolver a seguinte questão: dado um grafo $G = (V, E)$, ponderado e não-direcionado, existe uma imersão válida de G em \mathbb{R}^3 ? Este questionamento tem sido fonte de intensas pesquisas e neste trabalho foi apresentada uma adaptação do algoritmo da literatura *Branch and Prune*.

A adaptação proposta do método *Branch-and-Prune* foi realizada através da abordagem da resolução do sistema de equações lineares de distâncias euclidianas interatômicas. Propõe-se uma abordagem através da resolução das matrizes do sistema de coordenadas internas da molécula, obtidas através dos ângulos de ligação e torção. O BP adaptado está em refinamento, mas, já pôde ser demonstrado que em alguns casos, os dados necessários para o cálculo do quinto átomo, partindo-se de quatro esferas, pode ser

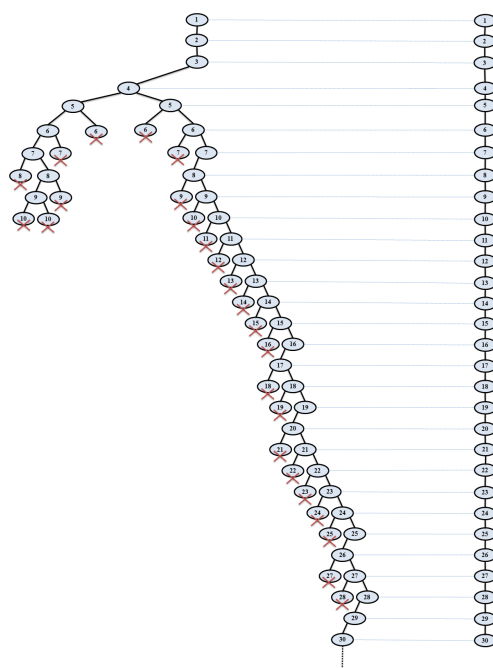


Figure 1. Etapas para encontrar as coordenadas dos átomos da molécula de feromônio contida na entrada 2erl. Na esquerda os passos da execução no Branch and Prune original e na direita a execução no BP com quatro esferas.

gerado experimentalmente por técnicas de mecânica quântica e, então, pode-se determinar a estrutura tridimensional da molécula em questão, em tempo linear. Na modificação proposta do algoritmo BP, apenas a estrutura parcial das proteínas da base PDB testadas puderam ser obtidas. Este resultado é significativo uma vez que para o caso geral, o problema é NP-difícil.

Referências

- Dong, Q. and Wu, Z. (2002). A linear-time algorithm for solving the molecular distance geometry problem with exact inter-atomic distances. *Journal of Global Optimization*, 22:365–375.
- Dong, Q. and Wu, Z. (2003). A geometric build-up algorithm for solving the molecular distance geometry problem with sparse distance data. *Journal of Global Optimization*.
- Lavor, C., Liberti, L., Maculan, N., and Mucherino, A. (2011). The discretizable molecular distance geometry problem. *Comput Optim Appl*.
- Mucherino, A., Liberti, L., Lavor, C., and Maculan, N. MD-jeep: a software tool for Distance Geometry. <http://www.antoniomucherino.it/en/mdjeep.php>.
- PDB (2013). Protein data bank. Disponível em <http://www.rcsb.org/pdb/home/home.do>. Acessado em fevereiro de 2013.
- Saxe, J. (1979). Embeddability of weighted graphs in k-space is strongly np-hard. *Proceedings of 17th Allerton Conference in Communications, Control and Computing*, pages 480–489.
- Silva, W. and Lavor, C. and Ochiand, L. S. (2008). Cálculo de estruturas de proteínas. *Simpósio Brasileiro de Pesquisa Operacional*.