

Paralelismo *Dataflow* em Geometria de Distâncias Moleculares Intervalares

Sandro P. Vilela¹

Programa de Engenharia de Sistemas e Computação, COPPE, UFRJ, Rio de Janeiro, RJ

Leandro A. J. Marzulo²

Departamento de Informática e Ciência da Computação, IME, UERJ, Rio de Janeiro, RJ

Felipe M. G. França³

Programa de Engenharia de Sistemas e Computação, COPPE, UFRJ, Rio de Janeiro, RJ

1 Introdução

O Problema de Geometria de Distâncias (*DGP*, *Distance Geometry Problem*) tem sido objeto de muita pesquisa, possui várias aplicações, tais como na localização em redes de sensores, no reconhecimento de imagens, nas predições de estruturas moleculares, entre outras, [1]. O *DGP* consiste na determinação das coordenadas de um determinado conjunto de pontos, a partir de algumas distâncias conhecidas entre os pares desses pontos [5]. Quando as distâncias entre alguns destes são dadas por intervalos, temos a versão intervalar do problema. Para o *DGP* relacionado a proteínas e empregando-se distâncias intervalares têm-se o *Interval Discretizable Molecular Distance Geometry Problem (iDMDGP)*. Neste trabalho iremos tratar da aplicação do *iDMDGP* para a reconstrução de estruturas tridimensionais de proteínas, através de dados (intervalares) que simulam medidas experimentais obtidas por Ressonância Magnética Nuclear. Empregaremos, para isto, uma abordagem paralela para o *iDMDGP*, baseada no modelo de execução *Dataflow*, o qual, tem se mostrado uma boa opção para a programação paralela, [4].

2 Metodologia

De uma maneira breve, o *iDMDGP* pode ser formulado como: Encontre as posições $x_1, \dots, x_n \in R^3$ dos átomos da molécula, tais que

$$d_{inf} \leq |x_i - x_j| \leq d_{sup} \quad (i, j) \in S, \quad (1)$$

¹sandro@cos.ufrj

²leandro@ime.uerj.br

³felipe@cos.ufrj

onde d_{inf} e d_{sup} correspondem, respectivamente, ao limite inferior e superior para a distância entre os átomos i e j . Como metodologia para encontrar as soluções (as posições atômicas) da equação 1, empregaremos um algoritmo *Branch & Prune*, [2], o qual, explorará o espaço de busca (topologia em árvore), de maneira recursiva, verificando a cada novo ramo adicionado, se a solução (posição atômica) encontrada satisfaz as condições impostas pelos dados experimentais, *pruning phase*. Ao se descobrir posições inviáveis reduz-se drasticamente o espaço de busca, no pior cenário, este pode crescer exponencialmente. O nosso principal objetivo consiste em reduzir o tempo de processamento desta busca através de uma abordagem de paralelismo por *Dataflow*. Para isto, vamos empregar a biblioteca *Dataflow Sucuri*, em Python, como ferramenta para implementar o paralelismo. Para a geração dos dados, que simulam a saída de experimentos de Ressonância Magnética Nuclear, os quais serão fornecidos como entrada ao algoritmo *Branch & Prune*, criamos um programa baseado na geometria das proteínas e principalmente nas instruções apresentadas em [3].

3 Conclusão

A abordagem proposta visa paralelizar o iDMDGP utilizando uma abordagem *Dataflow* por meio da biblioteca Sucuri. Esperando-se assim obter uma redução significativa no tempo total de processamento computacional. Resultados preliminares serão apresentados.

Referências

- [1] C. Lavor, L. Liberti, A. Mucherino, *Recent advances on the discretizable molecular distance geometry problem*, *European Journal of Operational Research*, 219: 698-706, 2012.
- [2] C. Lavor, L. Liberti, N. Maculan, A. Mucherino, *The Discretizable Molecular Distance Geometry Problem*. *Computational Optimization and Applications* 52, 115-146, 2012.
- [3] C. Lavor, N. Maculan, M. Souza e R. Alves, *Álgebra e Geometria no Cálculo de Estrutura Molecular*. 31º Colóquio Brasileiro de Matemática, 2017.
- [4] L. A. J. Marzulo, T. A. O. Alves, B. F. Goldstein, F. M. G. França, *Exploiting Parallelism in Linear Algebra Kernels through Dataflow Execution*. *International Symposium on Computer Architecture and High Performance Computing Workshop (SBAC-PADW)*, 2015.
- [5] A. Mucherino, C. Lavor, L. Liberti, N. Maculan, *Distance Geometry. Theory, Methods, and Applications*. Springer New York Heidelberg Dordrecht London. ISBN: 978-1-4614-5127-3, doi: 10.1007/978-1-4614-5128-0, 2013.