

Ricardo Emanuel Mueller

Relatório Final de Iniciação Científica

Blumenau

15 de Novembro de 2018

Ricardo Emanuel Mueller

Relatório Final de Iniciação Científica

Trabalho final de iniciação científica sob orientação do Prof. Dr. Felipe Delfini Caetano Fidalgo.

Universidade Federal de Santa Catarina

Curso de Licenciatura em Matemática

Programa de Graduação

Blumenau

15 de Novembro de 2018

Resumo

Este trabalho é uma apresentação de como foi o período de iniciação científica. O projeto intitulado “Um estudo de rotações como estratégia de adaptações do algoritmo Branch & Prune para resolução do Discretizable Molecular Distance Geometry Problem” tinha o objetivo de, em suma, estudar modos para produção de algoritmos capazes de resolver problemas relacionados à geometria de distâncias. Entretanto, para tal estudo, fazia-se necessária uma detalhada abordagem sobre os conteúdos que formam a base para o entendimento dos principais problemas. Assim, no intuito de facilitar a aprendizagem, o Professor Orientador optou por segmentar o trabalho em seu início. Neste sentido, o começo deste relatório terá ênfase no estudo de grafos. Em um segundo momento será discutido o Problema de Geometria de Distâncias Moleculares (PGDM). Estes assuntos estão diretamente ligados com disciplinas comuns em cursos de Graduação em Matemática. Portanto, este projeto foi de grande valia para o aluno de iniciação científica em questão.

Palavras-chaves: Algoritmo Branch & Prune. Grafos. Problema de Geometria de Distâncias Moleculares.

Sumário

1	INTRODUÇÃO	5
2	MATERIAIS E MÉTODOS	7
2.1	Teoria dos Grafos	7
2.2	Problema de Geometria de Distâncias Moleculares (PGDM)	11
3	RESULTADOS E DISCUSSÃO	15
4	CONSIDERAÇÕES FINAIS	17
	REFERÊNCIAS	19

1 Introdução

Partindo de uma técnica proposta por Kurt Wüthrich, sabe-se ser possível montar a estrutura 3D de uma molécula de proteína através de experimentos de Ressonância Magnética Nuclear (RMN), com base na distância entre os átomos. No entanto, reduzindo os átomos a pontos, matematicamente o problema reside em determinar as coordenadas de tais pontos. Assim, também pode-se considerar como arestas as distâncias entre os átomos. Logo, temos que encontrar as coordenadas em \mathbb{R}^3 dos pontos de um grafo ponderado.

A primeira parte do trabalho trata sobre grafos e possui (??) e (??) como principais referências. Vale destacar que no momento em que esta parte do projeto estava em andamento, outros alunos de iniciação científica eram responsáveis por fazer um estudo sobre quatérnios, uma vez que o professor orientador acreditava na eficácia da implementação de um método baseado em quatérnios.

Na segunda parte do trabalho é apresentado o “Problema de Geometria de Distâncias”, bem como são apresentados os pré-requisitos para compreensão. Esta seção teve (??) e (??) como principais referências.

2 Materiais e Métodos

No início da iniciação científica o professor orientador distribuiu materiais para a elaboração de seminários. Havia encontros semanais ou quinzenais para tais apresentações. O aluno autor deste relatório ficou responsável por apresentar uma introdução à “Teoria dos Grafos” aos demais pesquisadores.

2.1 Teoria dos Grafos

Grafos foi um assunto estudado por diversos matemáticos proeminentes. Pode-se dizer que em 1736 é que a teoria teve início, com base no artigo publicado por Leonhard Euler, sobre as 7 pontes de Königsberg. Existe uma íntima relação entre Grafos e algoritmos (podemos, por exemplo, representar um algoritmo por um grafo). Todavia, neste trabalho há um sentido a mais para se estudar isso, uma vez que moléculas podem ser representadas por grafos. Assim, partimos da seguinte definição:

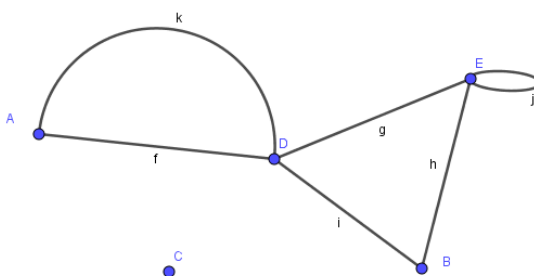
Definição 2.1.1. Um grafo G é um par ordenado da forma $(V(G), E(G))$, composto por um conjunto de vértices $V(G)$, de arestas $E(G)$ e uma função de incidência ψ_g que, por sua vez, associa a cada aresta de $V(G)$ um par não ordenado de vértices (nem sempre distintos) de $E(G)$. É dito que as arestas ligam os vértices, bem como denomina-se também tais vértices como extremidades desta aresta.

Exemplo 2.1.1. Vejamos um exemplo de grafo. Tome o grafo G_1 , de modo que:

$$V(G_1) = \{A, B, C, D, E\};$$

$$E(G_1) = \{f, g, h, i, j, k\}, \text{ com } f = \{A, D\}, g = \{D, E\}, h = \{B, E\}, i = \{B, D\}, j = \{E, E\} \text{ e } k = \{A, D\}.$$

Figura 1 – Grafo G_1



Fonte: Imagem Autoral.

Utilizando esta definição, pode-se denominar outras características das estruturas de um grafo. Assim, por exemplo, estudamos: *ordem*, *tamanho*, *incidência*, *adjacência*, *vizinhança*, *laço* e *elo*. Assim, os separamos por tipos, como grafo: *finito*, *nulo*, *simples*, *pseudografo*, *completo*, *conexo* e *bipartido*. Dentre estas, observe as seguintes definições:

Definição 2.1.2. Um grafo que não possui laços (arestas que possuem o mesmo vértice como extremidades), bem como não possui arestas paralelas (arestas cujas extremidades são os mesmo vértices) e onde quaisquer dois vértices são ligados por uma aresta é denominado grafo completo.

Definição 2.1.3. Grafo conexo é o nome dado para o tipo de grafo onde é possível estabelecer um caminho de qualquer vértice para qualquer outro vértice dele. Caso contrário, temos um grafo desconexo.

Definição 2.1.4. Chamamos de grafo bipartido o grafo onde pode-se particionar o conjunto de vértices em dois, de modo que cada aresta possua extremidades em ambos os conjuntos.

Exemplo 2.1.2. Exemplo de grafo bipartido:

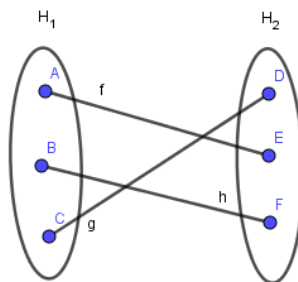
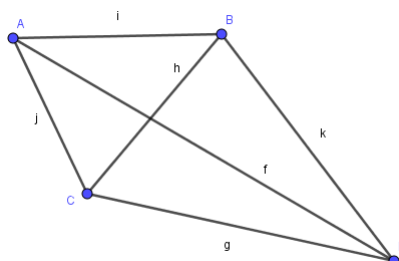


Figura 2 – Fonte: Imagem Autoral.

Exemplo 2.1.3. Exemplo de grafo conexo e completo:



Fonte: Imagem Autoral.

Em diversos momentos o professor orientador buscou trazer perguntas acerca dos assuntos tratados. Neste caso, perguntas que surgiram foram:

1. Todo grafo conexo pode ser bipartido?
2. Todo grafo bipartido é conexo?

A resposta é que nem todo grafo conexo pode ser bipartido, bem como nem todo grafo bipartido é conexo. Neste sentido, para verificar o primeiro item basta pegarmos qualquer grafo completo com 3 ou mais vértices e, uma vez que um dos conjuntos da bipartição terá mais de 2 vértices, com certeza teremos uma aresta cujas extremidades estarão num dos conjuntos da bipartição. Pelo Exemplo 2.1.2 podemos verificar que nem todo grafo bipartido é conexo, uma vez que partindo do vértice A, por exemplo, não podemos estabelecer um caminho até o vértice D.

Após este estudo, partiu-se para o estudo de *subgrafos*. Em especial, foram discutidos *caminhos* (tanto eulerianos quanto hamiltonianos), *cliques*, *ciclos*, *árvores*, *dígrafos* e, por fim, *grafos ponderados*. Aqui, é importante destacar a definição de *grafo ponderado*, pois podemos representar os esquemas de distância atômica com base neles.

Definição 2.1.5. *Chamamos de grafo ponderado o grafo que possui valores numéricos atribuídos as suas arestas.*

Após isto, seguiu-se para o estudo de matrizes de incidência e adjacência. Assim, considere as seguintes definições:

Definição 2.1.6. *As extremidades de uma aresta são denominadas incidentes à aresta, bem como esta aresta é incidente às extremidades.*

Definição 2.1.7. *Vértices distintos e incidentes à uma mesma aresta são denominados adjacentes. Nesse sentido, arestas que possuem um vértice em comum são adjacentes também.*

Definição 2.1.8. *Dois vértices distintos e adjacentes são denominados vizinhos. O conjunto de vértices vizinhos geralmente é denotado por $N_G(v)$.*

Assim, a partir de matrizes, é possível definir outras maneiras de representar um grafo.

Matriz de Incidência: Cada coluna representa os vértices incidentes a determinada aresta. Assim, cada linha representa as arestas incidentes a cada vértice. Contamos 1 para aresta incidente ao vértice (ou vice-versa) e 2 se tivermos um laço. Observe a matriz de

incidência de G_1 à seguir:

$$\begin{array}{cccccc} f & g & h & i & j & k \\ \left(\begin{array}{cccccc} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 2 & 0 \end{array} \right) & \begin{array}{l} A \\ B \\ C \\ D \\ E \end{array} \end{array}$$

Matriz de Adjacência: É uma tabela onde observa-se somente a relação de adjacência entre vértices. Caso um vértice seja adjacente a outro, bota-se número 1. No caso de um laço, bota-se 2. Observe a matriz de adjacência de G_1 :

$$\begin{array}{ccccc} A & B & C & D & E \\ \left(\begin{array}{ccccc} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 2 \end{array} \right) & \begin{array}{l} A \\ B \\ C \\ D \\ E \end{array} \end{array}$$

Por último, foram estudados alguns resultados sobre *grau de um vértice*. Assim, seguem outras definições, com a prova de 2 resultados na sequência.

Definição 2.1.9. *O grau de um vértice é o número de arestas incidentes a ele. Cada laço conta como duas arestas. O vértice com grau zero é chamado de vértice isolado.*

- $d_G(v)$: É o grau do vértice v . Se G é um grafo simples (sem laços ou arestas paralelas), então $d_G(v)$ denota o número de vizinhos de v em G .

Teorema 2.1.1. *Para todo grafo $G(V(G), E(G))$, com m arestas, vale:*

$$\sum_{v \in V} d_G(v) = 2m.$$

Demonstração: Considere a matriz de incidência M do grafo G . Para saber o grau de um determinado vértice basta olharmos para sua linha correspondente e somarmos os valores contidos na linha. Note que somar os valores presentes em todas as linhas é o mesmo que somar os valores presentes nas colunas. Como temos m colunas e cada aresta possui 2 vértices incidentes, segue que $\sum_{v \in V} d_G(v) = 2m$. \square

Teorema 2.1.2. *Em qualquer grafo, o número de vértices, cujo grau é ímpar, é par.*

Demonstração: Suponha que o número de vértices com grau ímpar seja ímpar. Note que a soma do grau de todos os vértices resulta em um número ímpar (resultado aritmético). Somando este resultado com o grau dos demais vértices, cujo grau é par, obtemos um número ímpar. Absurdo, pois $\sum_{v \in V} d(v) = 2m$. Portanto, o número de vértices, cujo grau é ímpar, é par. \square

2.2 Problema de Geometria de Distâncias Moleculares (PGDM)

Nesta etapa, encontros sem a presença do professor orientador foram recorrentes. O objetivo continuou sendo o mesmo, estudar os assuntos para apresentar os seminários. Entretanto, agora todos os alunos de iniciação científica estavam concentrados no estudo de (??). Além disso, quando necessário, o professor orientador pediu a implementação dos métodos sugeridos. Em diversos momentos surgiram dúvidas, comumente sanadas por revisões em (??) e (??).

Inicialmente, foram vistos aspectos históricos sobre o PGDM. Seguindo, começou-se a trabalhar com os problemas. As hipóteses presentes em (??) eram:

- Hipótese 1: as distâncias fornecidas pelos experimentos de RMN estão associados a pares de átomos conhecidos;
- Hipótese 2: todos os átomos da molécula da proteína cuja estrutura 3D queremos calcular são conhecidos;
- Hipótese 3: todos os átomos da molécula de proteína estão ligados a algum átomo, cuja distância é conhecida;
- Hipótese 4: existe uma ordem, dada *a priori*, entre os átomos da cadeia principal da proteína cuja estrutura 3D queremos calcular;
- Hipótese 5: as distâncias entre os átomos de uma molécula de proteína separados por duas ligações covalentes são conhecidas;
- Hipótese 6: as distâncias fornecidas pela RMN são representados por intervalos de números reais que contêm o valor correto associado.

Problema 2.2.1: Encontrar os pontos $x_i \in \mathbb{R}^3$, com $i = 1, \dots, n$, que resolvam a seguinte equação:

$$\|x_i - x_j\| = d_{ij}, \forall i, j \in E,$$

onde $E \subset \{1, \dots, n\} \times \{1, \dots, n\}$ e os d_{ij} são os valores de distâncias provenientes da Ressonância Magnética Nuclear (RMN).

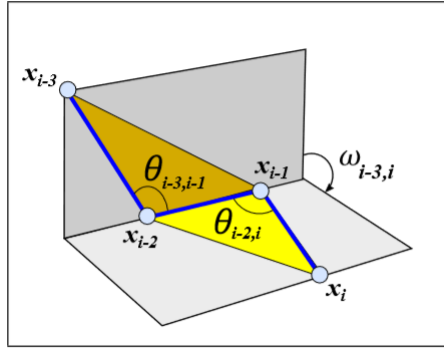
Perceba que pode-se reescrever este problema em termos de otimização. Assim, segue que dado:

$$f(x_1, \dots, x_n) = \sum_{(i,j) \in E} (\|x_i - x_j\| - d_{ij})^2.$$

Basta encontrar os $x_i \in \mathbb{R}^3$, $i = 1, \dots, n$, tais que $f(x_1, \dots, x_n) = 0$. Portanto, queremos

$$\min_{x_t \in \mathbb{R}^n} f(x_1, \dots, x_n).$$

Na sequência, foi estudado um pouco sobre a geometria das proteínas. E, a partir das hipóteses 3 e 5, viu-se a possibilidade de utilizarmos coordenadas internas para resolver o problema, ao invés das cartesianas. Conforme a figura abaixo (presente na página 12 de (??)), podemos representar uma proteína com base nos comprimentos das ligações covalentes $d_{i-1,i}$, ângulos planos θ_i e ângulos de torção ω_i .



Fonte: ALVES. LAVOR. SOUZA. MACULAN. (2017).

Assim, considerando $x_i \in \mathbb{R}^3$, $i = 1, \dots, n$, dadas por (x_{i1}, x_{i2}, x_{i3}) , podemos utilizar o seguinte método para determinar as coordenadas:

$$\begin{bmatrix} x_{i1} \\ x_{i2} \\ x_{i3} \\ 1 \end{bmatrix} = B_1 B_2 \cdots B_i \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix},$$

onde

$$B_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad B_2 = \begin{bmatrix} -1 & 0 & 0 & -d_{1,2} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$B_3 = \begin{bmatrix} -\cos(\theta_{1,3}) & -\sin(\theta_{1,3}) & 0 & -d_{2,3} \cos(\theta_{1,3}) \\ \sin(\theta_{1,3}) & -\cos(\theta_{1,3}) & 0 & d_{2,3} \sin(\theta_{1,3}) \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

e

$$B_i = \begin{bmatrix} -\cos(\theta_{i-2,i}) & -\sin(\theta_{i-2,i}) & 0 & -d_{i-1,i} \cos(\theta_{i-2,i}) \\ \sin(\theta_{i-2,i}) \cos(\omega_{i-3,i}) & -\cos(\theta_{i-2,i}) \cos(\omega_{i-3,i}) & -\sin(\omega_{i-3,i}) & d_{i-1,i} \sin(\theta_{i-2,i}) \cos(\omega_{i-3,i}) \\ \sin(\theta_{i-2,i}) \sin(\omega_{i-3,i}) & -\cos(\theta_{i-2,i}) \sin(\omega_{i-3,i}) & \cos(\omega_{i-3,i}) & d_{i-1,i} \sin(\theta_{i-2,i}) \sin(\omega_{i-3,i}) \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

com $i = 4, \dots, n$ para a última matriz.

Durante esta parte da iniciação científica, o aluno autor deste texto chegou num resultado interessante e que ajuda na compreensão geométrica de como o método funciona. Pode-se escrever a matriz B_i do seguinte modo:

$$B_i = R_x(w_{i-3,i}) \cdot R_x(\pi) \cdot R_z(\theta_{i-2,i}) \cdot R_y(\pi) \cdot T_x(d_{i-1,i}),$$

onde $R_\alpha(\gamma)$ representa a matriz de rotação que rotaciona um vetor por um ângulo γ em torno do eixo α , a partir da regra da mão direita. Além disso, $T_\beta(\lambda)$ representa uma translação de “tamanho” λ em relação ao eixo β .

Após isto, partiu-se para o estudo mais aprofundado sobre geometria de distâncias. Nesta parte, viu-se a relação entre grafos e o Problema de Geometria de Distâncias Moleculares (PGDM), fato evidenciado pela seguinte definição (presente na página 21 de (??)):

Definição 2.2.1. *Dado um grafo simples $G=(V,E,d)$, conexo e ponderado nas arestas por $d : E \rightarrow (0, +\infty)$, encontre uma função $x : V \rightarrow \mathbb{R}^3$ tal que:*

$$\forall \{u,v\} \in E, \|x(u) - x(v)\| = d(u,v).$$

Assim, foi possível discutir o número de soluções PGDM e a complexidade computacional envolvida.

Por fim, este estudo deu a base necessária para a compreensão do algoritmo de Branch & Prune, bem como permitiu que fossem feitas implementações computacionais.

3 Resultados e Discussão

Em matemática é geralmente difícil para um aluno de graduação produzir conteúdo inédito. Assim, supõe-se que os maiores ganhos foram em experiência na atividade de pesquisa.

O fato de existirem 5 membros na pesquisa fez com que vários modos de pensar entrassem em conflito. Isto gerou um engrandecimento sobre a ideia de matemática que cada um tinha, além de permitir a produção de novos pontos de vista acerca dos principais tópicos.

4 Considerações Finais

Este projeto de iniciação científica exigiu muitos conhecimentos sobre Geometria Analítica e Álgebra Linear. Neste sentido, além de estar trabalhando com algo novo, foi uma oportunidade de consolidar o que já havia visto durante a graduação. Além disso, como o discente pretende seguir carreira acadêmica, foi um ótimo momento para estar em contato com o trabalho de um pesquisador.

Foi um período positivo em vários outros aspectos também. Amizades foram feitas e aprendeu-se a conversar sobre matemática.

Referências