

Rigidez em Grafos de Proteínas

Carlile Lavor

Rigidez em Grafos de Proteínas

Rigidez em Grafos de Proteínas

Copyright © 2019 Carlile Lavor.
Publicado no Brasil / Published in Brazil.

ISBN 978-85-244-0482-5

MSC (2010) Primary: 68R10, Secondary: 05C62, 05C85, 92E10

Coordenação Geral

Nancy Garcia

Produção Books in Bytes

Capa Sergio Vaz

Realização da Editora do IMPA

IMPA

Estrada Dona Castorina, 110

Jardim Botânico

22460-320 Rio de Janeiro RJ

Telefones: (21) 2529-5005

2529-5276

www.impa.br

editora@impa.br

Conteúdo

Prefácio	viii
Introdução	1
1 Geometria de Distâncias (GD)	3
1.1 Matriz de Distâncias	3
1.2 Combinatória do Problema	5
2 Geometria de Proteínas	10
2.1 A Molécula de Proteína e a RMN	10
2.2 Coordenadas Cartesianas e Internas	11
3 Grafos de Proteínas	15
3.1 O Problema Fundamental da GD: o DGP	15
3.2 Rigidez de Grafos	17
3.3 Ordenação de Vértices no DGP	19
3.4 Algoritmo Branch & Prune (BP)	22
3.5 Ordenação de Vértices no DMDGP	28
4 Repetição de Vértices no DMDGP	37
4.1 Reordenação no DMDGP	37
4.2 A Reordenação hc (hand-crafted)	41

5 Geometria Molecular: do 3D ao 5D	44
5.1 Esferas e o DMDGP	44
5.2 O Modelo Conforme	46
6 Conclusão	53
Bibliografia	56

À Jaqueline, esposa e companheira de todas as horas.

Prefácio

Esse texto é uma extensão de outros dois [Lavor e Liberti \(2014\)](#) e [Lavor, Maculan et al. \(2017\)](#), apresentados como minicursos no Congresso Nacional de Matemática Aplicada e Computacional, em 2014, e no Colóquio Brasileiro de Matemática, em 2017, respectivamente. Além de conter novidades, várias melhorias e correções foram realizadas.

Iniciamos com a descrição do problema a ser estudado: usando dados de distâncias entre átomos próximos de uma dada proteína, provenientes de experimentos de Ressonância Magnética Nuclear (RMN), determinar a posição no espaço de todos os átomos da molécula. Também mencionamos duas maneiras de representar matematicamente o problema: solução de sistemas quadráticos e minimização de funções.

No Capítulo 1, começamos com alguns aspectos históricos importantes da teoria que fundamenta nosso problema, a Geometria de Distâncias, e introduzimos, por meio de exemplos, a abordagem combinatória que será adotada a partir de então.

No Capítulo 2, discutimos conceitos básicos sobre moléculas de proteínas, bem como noções fundamentais sobre a técnica de RMN. Em seguida, exibimos as duas maneiras clássicas de representar a estrutura 3D de uma proteína: coordenadas cartesianas e coordenadas internas.

O Capítulo 3 é o capítulo mais denso. Baseado no conceito de grafo, definimos o problema fundamental da Geometria de Distâncias, o *Distance Geometry Problem* (DGP), introduzimos o conceito de rigidez de grafos e de ordenação em seus vértices, e descrevemos uma classe do DGP, o *Discretizable Molecular DGP*

(DMDGP), definida para modelar o cálculo da estrutura 3D de proteínas usando dados de RMN. O algoritmo *Branch & Prune* (BP), para resolver o DMDGP, também é apresentado.

O Capítulo 4 discute as adaptações necessárias no DMDGP para que as incertezas dos experimentos de RMN possam ser levadas em conta na solução do problema com dados reais.

O último capítulo trata de um assunto em pleno desenvolvimento. Motivado pela interpretação geométrica do BP, propomos um novo modelo para representar a molécula de proteína, não mais usando as coordenadas homogêneas do \mathbb{R}^4 , mas o modelo conforme, definido em \mathbb{R}^5 .

Finalmente, agradecemos a oportunidade dada pela organização do 32.º Colóquio Brasileiro de Matemática, o suporte financeiro do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). Em particular, minha gratidão ao futuro Dr. Marcelo Carielo, pelas valiosas sugestões de melhoria do texto, e ao Prof. Paulo Ney de Souza, pelo competente e caprichado trabalho de edição.

Campinas, 10 de maio de 2019.
Carlile Lavor

Introdução

A estrutura 3D de uma molécula de proteína está associada a inúmeros processos biológicos, como divisão celular, invasão de células hospedeiras por patógenos, comunicação entre células, etc. A determinação de estruturas de proteínas pode ser considerada como a ponte entre a descrição de processos biológicos celulares e o mundo da físico-química [Donald \(2011\)](#).

A cristalografia de raios X foi o primeiro método que permitiu a determinação da estrutura proteica, mas removendo a molécula do meio aquoso. A partir da década de 1990, o desenvolvimento da Ressonância Magnética Nuclear (RMN) permitiu o estudo da estrutura proteica em solução, enfatizando sua dinâmica interna, cuja flexibilidade foi reconhecida como um fator fundamental em muitos processos biológicos [Güntert \(1998\)](#).

Diferentemente da cristalografia, que determina as posições de cada átomo da proteína, os experimentos de RMN são utilizados para determinar distâncias euclidianas entre átomos próximos de hidrogênio e o problema é determinar sua estrutura 3D com base nessa informação de distâncias. A técnica foi proposta pelo químico Kurt Wüthrich, cujo trabalho resultou no Nobel de Química, em 2002 [Wüthrich \(1986\)](#).

Assumiremos que cada distância dada $d_{i,j}$ automaticamente “revela” o par de átomos i, j associado e que todos os átomos que formam a proteína são conhecidos *a priori* (quando apenas uma lista de valores de distâncias é fornecida, onde não se sabe a que par de átomos se referem, o problema se torna ainda mais complicado [Billinge et al. \(2018\)](#)).

Enunciamos, então, a primeira formulação matemática para o problema (em

todo o texto, consideraremos apenas a norma euclidiana):

Definição 1. *Determinar os pontos $x_i \in \mathbb{R}^3$, $i = 1, \dots, n$ (n é o número de átomos da molécula), satisfazendo as equações*

$$\|x_i - x_j\| = d_{i,j}, \quad \forall (i, j) \in E, \quad (1)$$

onde $E \subset \{1, \dots, n\} \times \{1, \dots, n\}$ e $d_{i,j}$ são os valores de distâncias dados como entrada do problema.

A fim de considerar erros de medição e dados com ruído, as igualdades acima podem ser substituídas por desigualdades, levando a várias outras formulações [Cassioli, Bordiaux et al. \(2015\)](#), [T. M. Costa et al. \(2017\)](#), [D’Ambrosio et al. \(2017\)](#), [Gonçalves, Mucherino et al. \(2017\)](#), [Neto, Lavor e Lodwick \(s.d.\)](#), [Souza, Lavor et al. \(2013\)](#), [Souza, Xavier et al. \(2011\)](#) e [Worley et al. \(2018\)](#).

Como existem evidências [Bajaj \(1988\)](#) de que não seja possível obter uma fórmula “fechada” para a solução do sistema (1), podemos tentar resolvê-lo numericamente. Entretanto, o sistema tem mais variáveis do que equações, trazendo dificuldades também para a abordagem numérica, além de outros fatores [Liberti, Lavor, Maculan e Mucherino \(2014\)](#).

A maneira clássica de considerar o problema é por meio de métodos de otimização contínua [Liberti, Lavor, Maculan e Mucherino \(2014\)](#), definindo uma única expressão com todas as equações do sistema (1), dada por

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

Resolver o problema é o mesmo que encontrar valores $x_i \in \mathbb{R}^3$, $i = 1, \dots, n$, tais que $f(x_1, \dots, x_n) = 0$. Como zero é o menor valor possível para f , queremos minimizar a função $f: \mathbb{R}^n \rightarrow \mathbb{R}$ (2).

A função f acima tem vários mínimos locais e o que se deseja é encontrar o menor deles: o mínimo global. Em otimização, distinguir um mínimo local de um global é um problema delicado [Floudas e Gounaris \(2009\)](#). Além disso, a quantidade de mínimos locais cresce exponencialmente com a quantidade de átomos da molécula, tornando o problema ainda mais complicado [Liberti, Lavor, Maculan e Mucherino \(2014\)](#).

No capítulo seguinte, apresentaremos a teoria matemática que trata do problema em questão: a Geometria de Distâncias.

I

Geometria de Distâncias (GD)

1.1 Matriz de Distâncias

Geometria de Distâncias (GD) é o estudo da geometria baseada não mais em pontos, retas e planos, mas no conceito de distância. Podemos datar suas origens com Heron de Alexandria, por volta de 50 d.C., que calculou a área de um triângulo usando o comprimento de seus lados. Com um salto de quase dois mil anos, em 1842, Arthur Cayley considerou um problema parecido, com o tetraedro [Liberti e Lavor \(2016\)](#).

Em 1928, sob a influência de Hilbert, Karl Menger, então um jovem professor na Universidade de Viena, propôs uma nova axiomatização para espaços métricos usando o conceito de distância e o que é hoje conhecido como determinante de Cayley–Menger, generalizando o resultado de Heron, [Menger \(1928\)](#).

Durante o famoso Círculo de Viena, Menger organizou uma série de seminários, onde Gödel, em 1933, apresentou resultados sobre GD na esfera e em superfícies [Liberti, Swirszc e Lavor \(2016\)](#). Logo depois, em 1935, Isaac [Schoenberg \(1935\)](#) demonstrou a equivalência entre matrizes de distâncias euclidianas e matrizes de Gram. Com os resultados de [Blumenthal \(1953\)](#), em 1953, a GD se torna, finalmente, uma nova área do conhecimento.

A grande questão era encontrar condições necessárias e suficientes para decidir

se uma dada matriz é uma matriz de distâncias euclidianas (MDE), ou seja, uma matriz simétrica tal que existe um número inteiro positivo K e um conjunto de pontos em \mathbb{R}^K , onde as distâncias euclidianas entre esses pontos são iguais às entradas da matriz.

Tais condições podem ser obtidas com o determinante de Cayley–Menger, definido sobre o K -simplex, que é a generalização do conceito de segmento (em \mathbb{R}^1), triângulo (em \mathbb{R}^2) e tetraedro (em \mathbb{R}^3) [Sippl e Scheraga \(1986\)](#).

Dado um conjunto $S = \{p_0, p_1, \dots, p_K\}$ de $K + 1$ pontos em \mathbb{R}^K , o volume do K -simplex (que estende a noção de comprimento, área e volume) definido pelos pontos de S é dado por

$$\Delta_K(S) = \sqrt{\frac{(-1)^{K+1}}{2^K (K!)^2} CM(S)},$$

onde $CM(S)$ é o determinante de Cayley–Menger,

$$CM(S) = \begin{vmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & d_{0,1}^2 & \cdots & d_{0,K}^2 \\ 1 & d_{0,1}^2 & 0 & \cdots & d_{1,K}^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & d_{0,K}^2 & d_{1,K}^2 & \cdots & 0 \end{vmatrix},$$

com $d_{i,j} = \|p_i - p_j\|$, $\forall i, j \in \{0, \dots, K\}$.

Exercício 2. Usando $CM(\{p_0, p_1, p_2\})$, com $p_0, p_1, p_2 \in \mathbb{R}^2$, obtenha a fórmula de Heron: $\Delta_2(\{p_0, p_1, p_2\})$.

Dados um conjunto de pontos em \mathbb{R}^K , calcular as distâncias correspondentes é trivial, mas nosso interesse é o problema inverso: queremos obter os pontos, dadas as distâncias entre eles.

Usando determinantes de Cayley–Menger, é possível estabelecer condições necessárias e suficientes para decidir se uma dada matriz é uma MDE [Sippl e Scheraga \(1986\)](#). Geometricamente, esse resultado afirma que para uma matriz ser uma MDE em \mathbb{R}^K , mas não em \mathbb{R}^{K-1} , deve haver um K -simplex S em \mathbb{R}^K , com $\Delta_K(S) \neq 0$, e qualquer $(K + 1)$ -simplex e $(K + 2)$ -simplex, contendo S como uma das faces, deve estar em \mathbb{R}^K .

Em [Schoenberg \(1935\)](#), é definida uma bijeção entre matrizes de distâncias euclidianas e matrizes positivas semidefinidas (MPS), implicando que determinar se uma matriz é MDE ou MPS é, essencialmente, o mesmo problema.

Apesar de estarmos interessados em um problema com o K fixado ($K = 3$), não dispomos de todas as distâncias exigidas na caracterização dada pelos determinantes de Cayley–Menger. Apenas em 1978, [Yemini \(1978\)](#), pela primeira vez, considera um problema em GD com um conjunto “incompleto” de distâncias.

Quando todas as distâncias são conhecidas e a dimensão K é fixada (ver [Alencar, Lavor e Liberti \(2019\)](#), para K não fixado), podemos facilmente resolver o problema usando um outro ponto de vista, adotado a partir de agora, e que pode ser utilizado também quando não são conhecidas todas as distâncias [Liberti, Lavor, Mucherino et al. \(2011\)](#).

Além da teoria matemática associada à GD, o interesse por esse tema de pesquisa explica-se pela riqueza e variedade de suas aplicações, não apenas em geometria molecular [Crippen e Havel \(1988\)](#). Existem aplicações em muitos campos da ciência e engenharia, incluindo protocolos de sincronização de relógios, localização de redes de sensores, astronomia, robótica e nanoestruturas [Anderson et al. \(2009\)](#), [Billinge et al. \(2016\)](#), [Lavor, Liberti, Lodwick et al. \(2017\)](#), [Liberti e Lavor \(2017\)](#), [Moreira et al. \(2018\)](#), [Mucherino, Lavor, Liberti e Maculan \(2013\)](#), [Nielsen e Roth \(1999\)](#) e [Santiago et al. \(2018\)](#).

1.2 Combinatória do Problema

Para ganharmos alguma familiaridade com a nova abordagem, consideremos o problema de calcular as posições de quatro pontos u, v, r, s no plano, dadas por $x_u, x_v, x_r, x_s \in \mathbb{R}^2$, conhecendo todas as distâncias envolvidas: $d_{u,v}, d_{u,r}, d_{u,s}, d_{v,r}, d_{v,s}, d_{r,s}$. Para excluir soluções obtidas por translações e rotações, podemos fixar, por exemplo, u, v, r , tal que

$$\begin{aligned}\|x_u - x_v\| &= d_{u,v}, \\ \|x_u - x_r\| &= d_{u,r}, \\ \|x_v - x_r\| &= d_{v,r}.\end{aligned}$$

Exercício 3. Explique como “fixar” $x_u, x_v, x_r \in \mathbb{R}^2$, dadas $d_{u,v}, d_{u,r}, d_{v,r}$.

Dados x_u, x_v, x_r , podemos obter $x_s \in \mathbb{R}^2$ definindo o sistema abaixo:

$$\begin{aligned}\|x_s - x_u\| &= d_{u,s}, \\ \|x_s - x_v\| &= d_{v,s}, \\ \|x_s - x_r\| &= d_{r,s}.\end{aligned}$$

Elevando ao quadrado os termos das igualdades acima, obtemos

$$\begin{aligned}\|x_s\|^2 - 2(x_s \cdot x_u) + \|x_u\|^2 &= d_{u,s}^2, \\ \|x_s\|^2 - 2(x_s \cdot x_v) + \|x_v\|^2 &= d_{v,s}^2, \\ \|x_s\|^2 - 2(x_s \cdot x_r) + \|x_r\|^2 &= d_{r,s}^2,\end{aligned}$$

onde $x_i \cdot x_j$ é o produto interno usual. Subtraindo agora a primeira equação das outras duas, ficamos com apenas duas equações,

$$\begin{aligned}2(x_v - x_u) \cdot x_s &= \|x_v\|^2 - \|x_u\|^2 + d_{u,s}^2 - d_{v,s}^2, \\ 2(x_r - x_u) \cdot x_s &= \|x_r\|^2 - \|x_u\|^2 + d_{u,s}^2 - d_{r,s}^2,\end{aligned}$$

que podem ser reescritas como

$$Ax = b,$$

onde

$$\begin{aligned}A &= 2 \begin{bmatrix} x_{v1} - x_{u1} & x_{v2} - x_{u2} \\ x_{r1} - x_{u1} & x_{r2} - x_{u2} \end{bmatrix}, \\ b &= \begin{bmatrix} \|x_v\|^2 - \|x_u\|^2 + d_{u,s}^2 - d_{v,s}^2 \\ \|x_r\|^2 - \|x_u\|^2 + d_{u,s}^2 - d_{r,s}^2 \end{bmatrix}, \\ x_u &= \begin{bmatrix} x_{u1} \\ x_{u2} \end{bmatrix}, \quad x_v = \begin{bmatrix} x_{v1} \\ x_{v2} \end{bmatrix}, \quad x_r = \begin{bmatrix} x_{r1} \\ x_{r2} \end{bmatrix},\end{aligned}$$

e

$$x = \begin{bmatrix} x_{s1} \\ x_{s2} \end{bmatrix}.$$

Supondo que a matriz A seja inversível, temos uma única solução x^* , dada por

$$x^* = A^{-1}b. \tag{1.1}$$

O exemplo acima sugere que, dispondo de todas as distâncias, o problema pode ser facilmente resolvido, em tempo proporcional à quantidade de pontos, por meio de uma sequência de sistemas lineares, cujas matrizes associadas sejam inversíveis.

Exercício 4. *No exemplo acima, se existe A^{-1} , a solução é obtida para quaisquer valores $d_{u,s}$, $d_{v,s}$, $d_{r,s}$? Qual é a relação desse fato com a solução do sistema quadrático original?*

Exercício 5. *No exemplo acima, caso não exista a inversa de A , outros três pontos iniciais podem ser fixados. Se nenhuma dessas possibilidades garantir a inversa de A , podemos afirmar que o problema não tem solução?*

Já sabemos que “faltam” distâncias no problema que desejamos resolver. O procedimento proposto acima encontraria, então, apenas uma parte da solução, mas admitindo que todos os valores dados são, de fato, distâncias corretas entre átomos de uma molécula, o problema obviamente tem solução. Mas como encontrá-la?

A abordagem via otimização seria uma saída, mas a solução é o minimizador global da função (2), que pode ser muito difícil de ser obtido, dificuldade confirmada por um resultado teórico que classifica o problema (1) como NP-difícil [Saxe \(1979\)](#). De uma maneira simplificada, isso significa que resolver o problema pode exigir um custo exponencial em função da quantidade de pontos que desejamos localizar.

Exercício 6. *Ainda sobre o exemplo acima, que condições podemos impor sobre os pontos $x_u, x_v, x_r \in \mathbb{R}^2$ para garantir que o sistema linear associado tenha solução?*

Consideremos o mesmo problema anterior, mas imaginando que os pontos estejam no espaço 3D. Fixando também u, v, r , obtemos o mesmo sistema quadrático:

$$\begin{aligned} \|x_s - x_u\|^2 &= d_{u,s}^2, \\ \|x_s - x_v\|^2 &= d_{v,s}^2, \\ \|x_s - x_r\|^2 &= d_{r,s}^2. \end{aligned} \tag{1.2}$$

Novamente, subtraímos a primeira equação das outras duas:

$$\begin{aligned} 2(x_v - x_u) \cdot x_s &= \|x_v\|^2 - \|x_u\|^2 + d_{u,s}^2 - d_{v,s}^2, \\ 2(x_r - x_u) \cdot x_s &= \|x_r\|^2 - \|x_u\|^2 + d_{u,s}^2 - d_{r,s}^2. \end{aligned}$$

Aparentemente, nenhuma diferença com o problema anterior. Entretanto, escrevendo explicitamente o sistema linear associado,

$$\begin{bmatrix} x_{v1} - x_{u1} & x_{v2} - x_{u2} & x_{v3} - x_{u3} \\ x_{r1} - x_{u1} & x_{r2} - x_{u2} & x_{r3} - x_{u3} \end{bmatrix} \begin{bmatrix} x_{s1} \\ x_{s2} \\ x_{s3} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \|x_v\|^2 - \|x_u\|^2 + d_{u,s}^2 - d_{v,s}^2 \\ \|x_r\|^2 - \|x_u\|^2 + d_{u,s}^2 - d_{r,s}^2 \end{bmatrix},$$

obtemos uma matriz 2×3 e não mais uma matriz quadrada.

Reescrevendo o sistema,

$$\begin{bmatrix} x_{v1} - x_{u1} & x_{v2} - x_{u2} \\ x_{r1} - x_{u1} & x_{r2} - x_{u2} \end{bmatrix} \begin{bmatrix} x_{s1} \\ x_{s2} \end{bmatrix} + \begin{bmatrix} x_{v3} - x_{u3} \\ x_{r3} - x_{u3} \end{bmatrix} [x_{s3}] = \frac{1}{2} \begin{bmatrix} \|x_v\|^2 - \|x_u\|^2 + d_{u,s}^2 - d_{v,s}^2 \\ \|x_r\|^2 - \|x_u\|^2 + d_{u,s}^2 - d_{r,s}^2 \end{bmatrix}$$

e supondo que a matriz $\begin{bmatrix} x_{v1} - x_{u1} & x_{v2} - x_{u2} \\ x_{r1} - x_{u1} & x_{r2} - x_{u2} \end{bmatrix}$ seja inversível, obtemos:

$$\begin{bmatrix} x_{s1} \\ x_{s2} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} x_{v1} - x_{u1} & x_{v2} - x_{u2} \\ x_{r1} - x_{u1} & x_{r2} - x_{u2} \end{bmatrix}^{-1} \begin{bmatrix} \|x_v\|^2 - \|x_u\|^2 + d_{u,s}^2 - d_{v,s}^2 \\ \|x_r\|^2 - \|x_u\|^2 + d_{u,s}^2 - d_{r,s}^2 \end{bmatrix} \quad (1.3)$$

$$- \begin{bmatrix} x_{v1} - x_{u1} & x_{v2} - x_{u2} \\ x_{r1} - x_{u1} & x_{r2} - x_{u2} \end{bmatrix}^{-1} \begin{bmatrix} x_{v3} - x_{u3} \\ x_{r3} - x_{u3} \end{bmatrix} [x_{s3}].$$

Ou seja, não temos mais uma única solução. Para cada $x_{s3} \in \mathbb{R}$, obtemos valores para x_{s1} e x_{s2} .

Devemos, então, retornar ao sistema quadrático (1.2), escolher uma das equações (por exemplo, $\|x_s - x_u\|^2 = d_{u,s}^2$) e resolvê-la usando a solução do sistema linear acima.

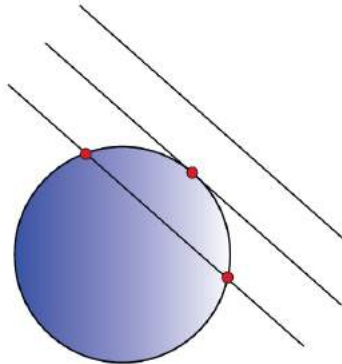


Figura 1.1: Interseção de uma esfera e uma reta.

Geometricamente, temos a interseção entre uma reta, dada por (1.3), e uma esfera, dada por

$$\|x_s - x_u\| = d_{u,s},$$

resultando em 3 possibilidades (Fig. 1.1): conjunto vazio (a reta não intercepta a esfera), apenas um ponto (a reta é tangente à esfera), dois pontos (a reta é secante à esfera).

O problema que acabamos de considerar, apenas com quatro pontos, ilustra dois aspectos importantes da abordagem combinatória que iremos investigar:

- A relação existente entre a quantidade de soluções e a dimensão do problema;
- A interpretação geométrica do problema.

Exercício 7. *Compare os dois exemplos apresentados, levando em conta os aspectos destacados acima.*

2

Geometria de Proteínas

2.1 A Molécula de Proteína e a RMN

Uma proteína é formada pela combinação de 20 moléculas menores, chamadas aminoácidos, que estão quimicamente ligadas umas às outras.

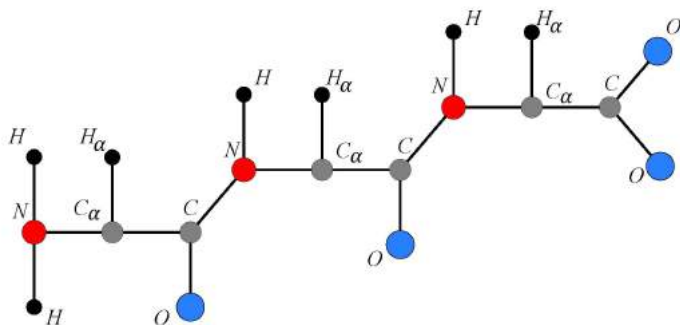


Figura 2.1: Cadeia principal de uma proteína.

A *cadeia principal* de uma proteína é definida por uma sequência de três áto-

mos, N, C_α, C , onde o átomo C_α está ligado a outro grupo de átomos, chamado de *cadeia lateral*, que distingue um aminoácido de outro (Fig. 2.1). Os átomos H, H_α, O , ligados a N, C_α, C , respectivamente, terão sua importância explicada mais adiante.

A cadeia principal é o esqueleto da proteína, que já dá uma boa ideia de sua estrutura 3D. Nesse texto, iremos nos restringir, essencialmente, à cadeia principal. Mais detalhes sobre cadeias laterais podem ser obtidos em [V. Costa et al. \(2014\)](#) e [Sallaume et al. \(2013\)](#).

Obviamente, o átomo é um objeto físico extremamente complexo, mas em cálculos de geometria de proteínas (estruturas muito maiores do que um átomo), é “padrão” assumir as seguintes hipóteses [Gibson e Scheraga \(1997\)](#):

- Cada átomo é representado como um ponto do \mathbb{R}^3 ;
- Os comprimentos das ligações covalentes entre quaisquer pares de átomos são conhecidos *a priori*;
- Os ângulos definidos por três átomos ligados consecutivamente também são dados *a priori*.

Em experimentos de Ressonância Magnética Nuclear (RMN), a proteína é submetida a um campo magnético, que induz um alinhamento do momento magnético dos átomos de hidrogênio. Em geral, se dois desses átomos estiverem separados por até 5 \AA ($1 \text{ \AA} = 10^{-10} \text{ m}$), é possível estimar sua distância relativa, mas devido à dinâmica da proteína em meio aquoso, ao ruído experimental e à influência de átomos vizinhos, apenas limites superiores são obtidos e os correspondentes limites inferiores são dados pela soma dos raios de van der Waals dos átomos envolvidos [Havel e Wüthrich \(1984\)](#). Isso significa que, quando dois átomos de hidrogênio estão próximos o suficiente, temos uma aproximação da distância entre eles dada por um intervalo de números reais, que chamaremos de *distância intervalar*. Isto é, os dados fornecidos pela RMN serão representados por $[\underline{d}_{i,j}, \bar{d}_{i,j}]$, onde o valor “correto” $d_{i,j}$ da distância associada é tal que

$$0 < \underline{d}_{i,j} \leq d_{i,j} \leq \bar{d}_{i,j}.$$

2.2 Coordenadas Cartesianas e Internas

Existe uma outra maneira de representar a estrutura 3D de uma proteína, com algumas vantagens em relação às *coordenadas cartesianas* $x_1, \dots, x_n \in \mathbb{R}^3$, por meio das *coordenadas internas* [Thompson \(1967\)](#).

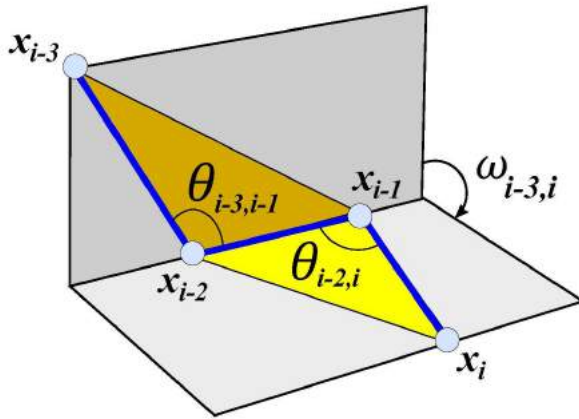


Figura 2.2: Coordenadas cartesianas e internas.

As coordenadas internas são definidas *pelos comprimentos das ligações covalentes* $d_{1,2}, \dots, d_{n-1,n}$, *pelos ângulos planos* $\theta_{1,3}, \dots, \theta_{n-2,n}$ (formados por três átomos consecutivos) e *pelos ângulos de torção* $\omega_{1,4}, \dots, \omega_{n-3,n}$ (formados por quatro átomos consecutivos) (Fig. 2.2). Cada $\omega_{i-3,i}$ é o ângulo entre os vetores normais dos planos definidos pelos átomos $i-3, i-2, i-1$ e $i-2, i-1, i$, respectivamente.

Para calcular os valores associados às distâncias entre os átomos, precisamos determinar as coordenadas cartesianas a partir das coordenadas internas, que podem ser obtidas por meio das matrizes abaixo, [Lavor \(2006\)](#).

Considerando que as coordenadas do ponto $x_i \in \mathbb{R}^3$, $i = 1, \dots, n$, são dadas por (x_{i1}, x_{i2}, x_{i3}) , temos:

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

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

$$B_i = \begin{bmatrix} -\cos \theta_{i-2,i} & -\sin \theta_{i-2,i} & 0 & -d_{i-1,i} \cos \theta_{i-2,i} \\ \sin \theta_{i-2,i} \cos \omega_{i-3,i} & -\cos \theta_{i-2,i} \cos \omega_{i-3,i} & -\sin \omega_{i-3,i} & d_{i-1,i} \sin \theta_{i-2,i} \cos \omega_{i-3,i} \\ \sin \theta_{i-2,i} \sin \omega_{i-3,i} & -\cos \theta_{i-2,i} \sin \omega_{i-3,i} & \cos \omega_{i-3,i} & d_{i-1,i} \sin \theta_{i-2,i} \sin \omega_{i-3,i} \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2.1)$$

Na última matriz, $i = 4, \dots, n$.

Usando as matrizes acima e fixando os valores $d_{1,2}$, $d_{2,3}$, $\theta_{1,3}$, as posições dos três primeiros átomos $x_1, x_2, x_3 \in \mathbb{R}^3$ são dadas por

$$x_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix},$$

$$x_2 = \begin{bmatrix} -d_{1,2} \\ 0 \\ 0 \end{bmatrix},$$

$$x_3 = \begin{bmatrix} -d_{1,2} + d_{2,3} \cos \theta_{1,3} \\ d_{2,3} \sin \theta_{1,3} \\ 0 \end{bmatrix}.$$

Isso significa que esses átomos podem ser fixados, excluindo estruturas obtidas por meio de rotações e translações a partir de uma dada estrutura.

Exercício 8. *Interprete geometricamente a “fixação” dada acima pelas matrizes B_1 , B_2 e B_3 .*

Nas matrizes acima, os pontos em \mathbb{R}^3 estão sendo representados em \mathbb{R}^4 , por meio das *coordenadas homogêneas*, muito úteis em computação gráfica [Fernandes, Lavor e Oliveira Neto \(2017\)](#).

Exercício 9. *Que vantagens as coordenadas homogêneas podem oferecer em geometria molecular?*

Considerando os comprimentos das ligações covalentes e os ângulos planos definidos como valores fixos e conhecidos, a estrutura 3D de uma molécula fica completamente determinada, a menos de rotações e translações, pelos valores dos

ângulos de torção $\omega_{1,4}, \dots, \omega_{n-3,n}$, cada um podendo variar no intervalo $[0, 2\pi]$. Além dessa vantagem, comparadas com as coordenadas cartesianas, as coordenadas internas também podem ser úteis na geração de instâncias em GD.

Utilizando valores relacionados a dados reais de moléculas de proteínas, podemos fixar os comprimentos das ligações covalentes ($d_{i-1,i} = 1,526 \text{ \AA}$) e os valores dos ângulos planos ($\theta_{i-2,i} = 1,91$ radianos).

Escolhendo aleatoriamente valores para $\omega_{i-3,i} \in [0, 2\pi]$, $i = 4, \dots, n$, e aplicando as matrizes acima para calcular as coordenadas cartesianas de todos os átomos, selecionamos os pares de pontos (i, j) cujas distâncias euclidianas $d_{i,j}$ sejam menores do que um determinado valor (para simular instâncias associadas ao problema em questão, podemos escolher pares de pontos tais que $d_{i,j} \leq 5 \text{ \AA}$ [Lavor \(2006\)](#)). Desse modo, a partir de uma estrutura conhecida, dada pelos valores $\omega_{1,4}, \dots, \omega_{n-3,n}$, construímos o problema inverso, definido pelas distâncias $d_{i,j}$ associadas.

Exercício 10. *De que maneira podemos escolher os valores $\omega_{i-3,i} \in [0, 2\pi]$, $i = 4, \dots, n$, a fim de gerar estruturas mais compactas ou esticadas?*

Exercício 11. *Considerando duas instâncias, uma mais compacta e outra mais esticada, qual delas é mais difícil de resolver?*

3

Grafos de Proteínas

3.1 O Problema Fundamental da GD: o DGP

Para facilitar a consulta à literatura sobre GD, adotaremos as siglas em inglês dos problemas que iremos definir a partir de agora.

O problema fundamental da GD, o *Distance Geometry Problem* (DGP), é determinar um conjunto de pontos, em um dado espaço geométrico, cujas distâncias entre alguns deles são conhecidas. O DGP pode ser definido de maneira genérica, considerando o espaço geométrico e o conceito de distância associados de maneira bem abstrata [M. M. Deza e E. Deza \(2009\)](#) e [Liberti, Lavor, Maculan e Mucherino \(2014\)](#). Entretanto, daremos uma definição ligada à geometria molecular.

Introduziremos agora o conceito de *grafo*, um objeto matemático essencial para explorar as propriedades do problema em questão. A *Teoria dos Grafos*, [Harary \(1969\)](#) é uma área de pesquisa da matemática muito rica, tanto do ponto de vista teórico quanto prático, e com uma importância cada vez maior no mundo moderno (basta citar as chamadas *redes complexas* e a era *big data*).

Um grafo $G = (V, E)$ é definido por dois conjuntos: os *vértices* V e as *arestas* E , definidas por pares de vértices. Note que a definição não exige nenhuma estrutura sobre o conjunto V , que pode ser formado por qualquer coleção de objetos. De certa forma, é essa flexibilidade que gera o poder desse objeto matemático

para modelar uma quantidade imensa de problemas, dentro e fora da matemática.

Em nosso problema, os vértices estão relacionados aos átomos de uma molécula de proteína e as arestas aos pares de átomos cujas distâncias são conhecidas. Quando associamos valores numéricos às arestas de um grafo, dizemos que o grafo é *ponderado* nas arestas. Isto é, temos uma função $d : E \rightarrow \mathbb{R}$, que acrescentamos na definição do grafo: $G = (V, E, d)$.

Dizemos que o grafo $G = (V, E)$ é *simples*, quando

$$\{a, b\} \in E \Rightarrow a \neq b;$$

é *conexo*, quando não é possível separar o conjunto de seus vértices V em dois outros, $V = A \cup B$, de tal maneira que não exista nenhuma aresta $\{a, b\} \in E$, com $a \in A$ e $b \in B$; e é *completo*, quando E contém todos os pares possíveis.

Uma *clique* em $G = (V, E)$ é um outro grafo $G' = (V', E')$, onde $V' \subset V$, $E' \subset E$ e G' é completo.

Usando grafos, definimos formalmente o DGP:

Definição 12. *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*

$$\forall \{u, v\} \in E, \|x(u) - x(v)\| = d(\{u, v\}). \quad (3.1)$$

Resolver o problema é associar a cada vértice de G um único ponto em \mathbb{R}^3 , satisfazendo as equações (3.1). Ao posicionarmos os pares de vértices $\{u, v\} \in E$ em \mathbb{R}^3 , temos que “acertar” a distância calculada $\|x(u) - x(v)\|$ com o valor dado $d(\{u, v\})$.

Note que os dados de entrada do DGP estão todos “concentrados” no grafo G e a saída do problema é justamente a função $x : V \rightarrow \mathbb{R}^3$. Observe, também, que apesar dos vértices estarem associados aos átomos da proteína, eles são entes *abstratos*, enquanto que os átomos são objetos do espaço 3D, cujas posições são dadas pela função $x : V \rightarrow \mathbb{R}^3$.

A utilização do conceito de grafo simplifica a representação dos dados de entrada, mas a principal vantagem está em poder explorar sua estrutura para compreender melhor o problema.

A ideia de usar grafos para representar moléculas coincide com sua própria história, pois a origem da palavra grafo está relacionada a essa representação [Brown \(1864\)](#). Provavelmente, essa seja a relação mais profunda entre a química e a matemática discreta [Sylvester \(1877\)](#).

Assumindo que os dados de entrada estão corretos e precisos, o conjunto X de soluções de um DGP produzirá todas as estruturas 3D da proteína que são compatíveis com as distâncias dadas. Obviamente, qualquer $x \in X$ pode ser transladado ou rotacionado em \mathbb{R}^3 , implicando que o conjunto de soluções não é apenas infinito, mas incontável. Entretanto, excluindo o efeito de translações e rotações, a cardinalidade de X depende apenas da estrutura do grafo associado $G = (V, E, d)$. Já vimos que, se o conjunto de arestas E contiver todos os pares possíveis de V , existe apenas uma solução que pode ser encontrada em tempo linear.

Usando geometria algébrica, é possível provar que há apenas duas possibilidades em relação à cardinalidade do conjunto de soluções X (a menos de rotações e translações): finita ou incontável, supondo que $X \neq \emptyset$ [Benedetti e Risler \(1990\)](#). Este resultado está fortemente ligado à rigidez do grafo associado [Graver, B. Servatius e H. Servatius \(1993\)](#) (ver próxima seção).

Por exemplo, se o grafo for rígido, o conjunto de soluções é finito (a menos de translações e rotações). Neste caso, uma busca combinatória pode ser mais apropriada do que uma pesquisa “contínua”, porque além da precisão e eficiência dos métodos combinatórios, essa abordagem pode produzir mais informações sobre a cardinalidade e a estrutura do conjunto de soluções [Lavor, Liberti, Maculan et al. \(2012a\)](#) e [Liberti, Masson et al. \(2014\)](#).

3.2 Rigidez de Grafos

Dados um grafo $G = (V, E, d)$ de um DGP e uma função $x: V \rightarrow \mathbb{R}^3$ satisfazendo todas as equações (3.1), dizemos que o par (G, x) é uma *realização* de G (para simplificar a notação, usaremos x_u, x_v , no lugar de $x(u), x(v)$, e $d_{u,v}$, no lugar de $d(\{u, v\})$).

Iniciaremos com a definição de rigidez de uma realização [Jackson e Jordán \(2005\)](#), baseada nos conceitos de duas relações: isometria e congruência.

Dois realizações (G, x) and (G, y) são *isométricas*, denotadas por $(G, x) \sim (G, y)$, se

$$\forall \{u, v\} \in E, \|x_u - x_v\| = \|y_u - y_v\|,$$

e *congruentes*, denotadas por $(G, x) \equiv (G, y)$, se

$$\forall u \neq v \in V, \|x_u - x_v\| = \|y_u - y_v\|.$$

Exercício 13. Prove que congruência implica isometria, mas o inverso é falso.

Note que duas realizações são congruentes apenas quando todos os pares de vértices de V tiverem as mesmas distâncias correspondentes, e não apenas os pares em E . Pode-se demonstrar que, se duas realizações são congruentes, uma delas pode ser obtida a partir da outra por composições de translações, rotações e reflexões [Blumenthal \(1953\)](#).

Definição 14. *Dado um grafo $G = (V, E, d)$ de um DGP, (G, x) é uma realização rígida se, para qualquer outra realização (G, y) ,*

$$(G, x) \sim (G, y) \Rightarrow (G, x) \equiv (G, y).$$

Geometricamente, isso significa que uma realização é rígida se não tiver deformações contínuas (sem deixar de satisfazer as equações do DGP associado), além da composição de translações, rotações e reflexões. Dito de outra maneira: em uma realização rígida, o único modo de mover continuamente um ponto, de modo que todas as distâncias sejam preservadas e não apenas as distâncias dadas pelas arestas, é mover todos os pontos.

Para definir rigidez em grafos, precisamos do conceito de *rigidez infinitesimal* de uma realização [Tay e Whiteley \(1985\)](#).

Definição 15. *Dados um grafo $G = (V, E, d)$ de um DGP, uma realização (G, x) , $|V| = n$ e $|E| = m$, considere o sistema linear*

$$R\lambda = 0,$$

onde $R \in \mathbb{R}^{m \times 3n}$ e cada linha $\{u, v\}$ de R ($\{u, v\} \in E$) tenha exatamente 6 entradas diferentes de zero dadas por

$$x_{ui} - x_{vi}$$

e

$$x_{vi} - x_{ui},$$

para $i = 1, 2, 3$. A realização é infinitesimalmente rígida se as únicas soluções de $R\lambda = 0$ forem translações ou rotações.

Em [Gluck \(1975\)](#), prova-se que rigidez infinitesimal implica rigidez, induzindo a definição a seguir [Hendrickson \(1992\)](#).

Definição 16. *Um grafo $G = (V, E, d)$ de um DGP é rígido se existir uma realização (G, x) infinitesimalmente rígida.*

Pode-se também definir rigidez de grafos independentemente do conceito de realização, chamada *rigidez genérica* Connelly (2005), que não trataremos nesse texto.

Uma caracterização de grafos rígidos no plano foi descrita em Laman (1970), mas nenhuma caracterização completa é conhecida para o espaço 3D. Um método heurístico foi introduzido em Sitharam e Zhou (2004) e conjecturas podem ser encontradas em Jackson e Jordán (2008).

Definição 17. *Um grafo $G = (V, E, d)$ de um DGP é globalmente rígido se existir uma única realização, a menos de congruências.*

Baseado em condições necessárias e suficientes para a rigidez global em \mathbb{R}^2 , Hendrickson (1992) conjecturou que as mesmas condições seriam suficientes para \mathbb{R}^3 , mas o resultado foi refutado em Connelly (2005). Algumas propriedades de grafos que garantem a rigidez global em \mathbb{R}^2 e \mathbb{R}^3 são dadas em Anderson et al. (2009).

Dada essa “dificuldade” em caracterizar a rigidez em \mathbb{R}^3 , a partir de agora, apresentaremos uma condição suficiente (mas não necessária) para a rigidez de grafos do DGP, baseada na ordenação de seus vértices.

3.3 Ordenação de Vértices no DGP

A ordenação de vértices é empregada na resolução de muitos problemas modelados por grafos Bodlaender et al. (2012) e Mueller, Martin e Lumsdaine (2007), mas a ideia de explorar tal fato para investigar a rigidez apareceu pela primeira vez em Henneberg (1886).

Em exemplos anteriores, vimos que, se cada vértice for *adjacente* a pelo menos quatro outros previamente fixados (*i.e.*, se existirem arestas entre eles), é possível calcular unicamente a posição de cada vértice, implicando na resolução do problema em tempo linear.

Dada uma ordenação nos vértices de um DGP, a quantidade de predecessores adjacentes a cada vértice que se deseja realizar é um fator crítico: sendo menor do que três, o número de soluções do DGP pode ser incontável; a partir de quatro, a solução é única, encontrada rapidamente Liberti, Lavor, Maculan e Mucherino (2014). Ou seja, o número de predecessores adjacentes, em uma dada ordenação dos vértices de um DGP, está relacionado à cardinalidade do conjunto de soluções e, também, ao esforço computacional necessário para resolver o problema.

Exercício 18. *Faça um desenho no plano para ilustrar o resultado acima.*

Em geral, não é possível encontrar ordenações de vértices em grafos de proteínas $G = (V, E, d)$ que permitam a solução do problema em tempo linear [Lavor, Liberti, Maculan et al. \(2012b\)](#). Entretanto, utilizando as informações fornecidas por experimentos de RMN e a geometria de proteínas, podemos tentar encontrar ordenações $v_1, \dots, v_n \in V$, tais que:

- Os primeiros três vértices formam uma clique:

$$\{v_1, v_2\}, \{v_1, v_3\}, \{v_2, v_3\} \in E;$$

- A partir do quarto vértice, todos são adjacentes a pelo menos três predecessores:

$$\forall i > 3, \exists j, k, l \text{ com } j < i, k < i, l < i : \{v_j, v_i\}, \{v_k, v_i\}, \{v_l, v_i\} \in E.$$

A classe de instâncias do DGP que possui essa ordenação, em que a clique inicial tem uma realização em \mathbb{R}^3 e as desigualdades triangulares relacionadas aos predecessores adjacentes a v_i ($i > 3$) são satisfeitas estritamente (*i.e.*, $d_{v_j, v_k} + d_{v_k, v_l} > d_{v_j, v_l}$), é chamada de *Discretizable DGP* (DDGP) e as ordenações associadas são ditas *ordenações DDGP* [Gonçalves e Mucherino \(2014\)](#) e [Mucherino, Lavor e Liberti \(2012b\)](#).

A clique inicial garante que o conjunto de soluções conterá apenas soluções incongruentes e a desigualdade triangular estrita impede uma quantidade incontável de soluções [Mucherino, Lavor e Liberti \(2012b\)](#). Neste mesmo artigo, foi demonstrado que o grafo de qualquer instância do DDGP é rígido. Também, um método combinatório [Liberti, Lavor e Maculan \(2008\)](#), chamado *Branch and Prune* (BP), foi apresentado para encontrar todas as soluções incongruentes do problema. O algoritmo BP pode ser exponencial no pior caso [dos Santos Carvalho, Lavor e Protti \(2008\)](#), o que é consistente com o fato do DDGP também ser um problema NP-difícil [Mucherino, Lavor e Liberti \(2012b\)](#).

Em uma ordenação DDGP, o quarto vértice v_4 pode ser realizado resolvendo o sistema quadrático abaixo (para simplificar a notação, usaremos x_i e $d_{i,j}$, ao invés de x_{v_i} e d_{v_i, v_j} , respectivamente),

$$\begin{aligned} \|x_4 - x_1\|^2 &= d_{1,4}^2, \\ \|x_4 - x_2\|^2 &= d_{2,4}^2, \\ \|x_4 - x_3\|^2 &= d_{3,4}^2, \end{aligned}$$

resultando em até duas posições possíveis para v_4 em \mathbb{R}^3 [Mucherino, Lavor e Liberti \(2012b\)](#). Usando a mesma estratégia, para cada posição já determinada para v_4 , obtemos outras duas posições para v_5 , e assim por diante. Por causa da rigidez do grafo do DDGP, temos um número finito de 2^{n-3} possíveis soluções.

Se tivermos uma informação de distância “extra”, $\{v_r, v_i\} \in E$ com $r < i$, podemos adicionar mais uma equação ao sistema relacionado a v_i , resultando em

$$\|x_i - x_j\| = d_{j,i},$$

$$\|x_i - x_k\| = d_{k,i},$$

$$\|x_i - x_l\| = d_{l,i},$$

$$\|x_i - x_r\| = d_{r,i}.$$

Elevando ao quadrado ambos os lados dessas equações, obtemos

$$\|x_i\|^2 - 2(x_i \cdot x_j) + \|x_j\|^2 = d_{j,i}^2,$$

$$\|x_i\|^2 - 2(x_i \cdot x_k) + \|x_k\|^2 = d_{k,i}^2,$$

$$\|x_i\|^2 - 2(x_i \cdot x_l) + \|x_l\|^2 = d_{l,i}^2,$$

$$\|x_i\|^2 - 2(x_i \cdot x_r) + \|x_r\|^2 = d_{r,i}^2.$$

Subtraindo agora uma dessas equações das outras, eliminamos o termo $\|x_i\|^2$ e obtemos um sistema linear na variável $x_i \in \mathbb{R}^3$. Se os pontos x_j, x_k, x_l, x_r não estão no mesmo plano, temos uma solução única x_i^* para v_i , supondo $\|x_i^* - x_r\| = d_{r,i}$.

Quando há outros predecessores adjacentes a v_i , além de v_j, v_k, v_l , uma ou ambas as posições possíveis para v_i podem ser inviáveis com relação a essas distâncias adicionais. Se ambas forem inviáveis, é necessário “retroceder” e repetir o procedimento com uma posição diferente para os vértices anteriores [Mucherino, Lavor e Liberti \(2012b\)](#).

Exercício 19. *Explique o motivo de exigir que os pontos x_j, x_k, x_l, x_r não estejam no mesmo plano.*

A ordenação DDGP “organiza” o espaço de busca em uma *árvore binária* e as distâncias “adicionais” podem ser usadas para reduzir o espaço de busca, podendo posições inviáveis na árvore.

A árvore começa com as três posições fixadas para a clique inicial, x_1, x_2, x_3 . No nível $i > 3$, a árvore contém todas as posições possíveis (2^{i-3}) para o vértice v_i , se não ocorrer nenhuma poda. A busca termina quando um caminho da raiz

($i = 1$) da árvore para um nó folha ($i = n$) é encontrado, de tal maneira que todas as posições relativas aos vértices no caminho satisfaçam as equações que definem o DDGP.

Considerando dados de entrada precisos, o desempenho desse procedimento é impressionante do ponto de vista da eficiência e confiabilidade [Lavor, Liberti, Maculan et al. \(2012b\)](#). Embora o DDGP seja um problema NP-difícil, uma ordenação DDGP pode ser encontrada em tempo polinomial [Lavor, Lee et al. \(2012\)](#).

Na definição do DDGP, a única exigência sobre os predecessores adjacentes a v_i ($i > 3$) é que a desigualdade triangular estrita associada deva ser satisfeita. No entanto, dependendo da instância, se as distâncias $d_{j,i}$, $d_{k,i}$, $d_{l,i}$ não forem bem “dimensionadas”, a incidência de erros numéricos na resolução do sistema quadrático relacionado pode aumentar, impedindo que uma solução seja encontrada [Mucherino, Lavor e Liberti \(2012b\)](#). Além disso, quando os vértices $\{v_j, v_k, v_l\}$ não formam uma clique, o sistema quadrático associado pode não ter solução.

Exercício 20. *Explique a frase acima: “quando os vértices $\{v_j, v_k, v_l\}$ não formam um clique, o sistema quadrático associado pode não ter solução”.*

Grafos de proteínas com dados de experimentos de RMN têm informações suficientes para permitir a definição de ordenações de vértices envolvendo predecessores imediatamente adjacentes, evitando os problemas mencionados acima sobre o DDGP.

Antes de tratar essa nova ordenação, vejamos o algoritmo BP com um pouco mais de detalhe.

3.4 Algoritmo Branch & Prune (BP)

Como dissemos, o algoritmo para encontrar uma solução do DDGP é chamado de *Branch and Prune* (BP) [Liberti, Lavor e Maculan \(2008\)](#). Essencialmente, a ideia é “realizar” vértice por vértice, seguindo a ordenação dada.

Seja um DGP dado pelo grafo $G = (V, E)$ abaixo, sem considerar os pesos

nas arestas:

$$\begin{aligned}
 V &= \{p, q, r, s, t, u, v\}, \\
 E &= \{\{p, q\}, \{p, r\}, \{p, s\}, \{p, u\}, \{p, v\}, \\
 &\quad \{q, r\}, \{q, s\}, \{q, t\}, \{q, u\}, \{q, v\}, \\
 &\quad \{r, s\}, \{r, t\}, \{r, v\}, \\
 &\quad \{s, t\}, \{s, v\}, \\
 &\quad \{t, u\}, \{t, v\}, \\
 &\quad \{u, v\}\}.
 \end{aligned}$$

Considerando a ordenação

$$V = \{r, q, t, s, v, u, p\},$$

$\{r, q, t\}$ é uma clique e

$$\begin{aligned}
 \{\{r, s\}, \{q, s\}, \{t, s\}\} &\subset E, \\
 \{\{r, v\}, \{q, v\}, \{t, v\}, \{s, v\}\} &\subset E, \\
 \{\{q, u\}, \{t, u\}, \{v, u\}\} &\subset E, \\
 \{\{r, p\}, \{q, p\}, \{s, p\}, \{v, p\}, \{u, p\}\} &\subset E.
 \end{aligned}$$

Denotando os vértices ordenados por

$$v_1 = r, v_2 = q, v_3 = t, v_4 = s, v_5 = v, v_6 = u, v_7 = p,$$

temos

$$\begin{aligned}
 V &= \{v_1, v_2, v_3, v_4, v_5, v_6, v_7\}, \\
 E &= \{\{v_1, v_2\}, \{v_1, v_3\}, \{v_1, v_4\}, \{v_1, v_5\}, \{v_1, v_7\}, \\
 &\quad \{v_2, v_3\}, \{v_2, v_4\}, \{v_2, v_5\}, \{v_2, v_6\}, \{v_2, v_7\}, \\
 &\quad \{v_3, v_4\}, \{v_3, v_5\}, \{v_3, v_6\}, \\
 &\quad \{v_4, v_5\}, \{v_4, v_7\}, \\
 &\quad \{v_5, v_6\}, \{v_5, v_7\}, \\
 &\quad \{v_6, v_7\}\}.
 \end{aligned}$$

Para facilitar a construção dos sistemas quadráticos, consideremos os vértices anteriores a v_4, v_5, v_6, v_7 :

$$\begin{aligned} \{\{v_1, v_4\}, \{v_2, v_4\}, \{v_3, v_4\}\} &\subset E, \\ \{\{v_1, v_5\}, \{v_2, v_5\}, \{v_3, v_5\}, \{v_4, v_5\}\} &\subset E, \\ \{\{v_2, v_6\}, \{v_3, v_6\}, \{v_5, v_6\}\} &\subset E, \\ \{\{v_1, v_7\}, \{v_2, v_7\}, \{v_4, v_7\}, \{v_5, v_7\}, \{v_6, v_7\}\} &\subset E. \end{aligned}$$

Como fizemos na seção anterior, usaremos x_i e $d_{i,j}$, ao invés de x_{v_i} e d_{v_i,v_j} , respectivamente.

Vamos supor que v_1, v_2, v_3 já estejam fixados em \mathbb{R}^3 . Para fixar v_4 , temos

$$\{\{v_1, v_4\}, \{v_2, v_4\}, \{v_3, v_4\}\} \subset E$$

e, montando o sistema quadrático associado a $x_4 \in \mathbb{R}^3$, como única variável, obtemos:

$$\begin{aligned} \|x_4 - x_1\|^2 &= d_{1,4}^2, \\ \|x_4 - x_2\|^2 &= d_{2,4}^2, \\ \|x_4 - x_3\|^2 &= d_{3,4}^2. \end{aligned}$$

Abrindo as contas e subtraindo, por exemplo, a primeira equação das outras duas, obtemos um sistema linear

$$Ax_4 = b,$$

onde $A \in \mathbb{R}^{2 \times 3}$ e $b \in \mathbb{R}^2$. Como A tem posto completo (por hipótese), obtemos x_4 em função de um parâmetro $\lambda \in \mathbb{R}$, denotado por $x_4(\lambda)$. Substituindo $x_4(\lambda)$ em uma das equações do sistema quadrático acima, por exemplo, na equação $\|x_4 - x_1\|^2 = d_{1,4}^2$, obtemos a seguinte equação do segundo grau em λ :

$$\|x_4(\lambda)\|^2 - 2x_4(\lambda) \cdot x_1 + \|x_1\|^2 - d_{1,4}^2 = 0.$$

Com os valores de λ , obtemos os possíveis valores para x_4 .

Para cada posição de v_4 (x_4^0 e x_4^1), fazendo o mesmo procedimento acima, podemos também obter duas posições para v_5 . Escolhendo x_4^0 , temos o seguinte sistema quadrático:

$$\begin{aligned} \|x_5 - x_2\|^2 &= d_{25}^2, \\ \|x_5 - x_3\|^2 &= d_{35}^2, \\ \|x_5 - x_4^0\|^2 &= d_{45}^2. \end{aligned}$$

Entretanto, temos também $\{v_1, v_5\} \in E$, que devemos usar para testar cada uma das possíveis posições obtidas para v_5 (x_5^0 e x_5^1). Isto é,

$$\|x_5^0 - x_1\| = d_{1,5}$$

ou

$$\|x_5^1 - x_1\| = d_{1,5}?$$

Exercício 21. *As duas equações acima podem ser satisfeitas ao mesmo tempo?*

Supondo que os pontos x_1, x_2, x_3, x_4^0 não sejam coplanares, apenas uma das equações deve ser satisfeita, digamos a primeira. Descartemos, então, x_5^1 e consideremos x_5^0 .

Passemos para o vértice v_6 . Como temos apenas $\{\{v_2, v_6\}, \{v_3, v_6\}, \{v_5, v_6\}\} \subset E$, montamos o sistema quadrático

$$\|x_6 - x_2\|^2 = d_{2,6}^2,$$

$$\|x_6 - x_3\|^2 = d_{3,6}^2,$$

$$\|x_6 - x_5^0\|^2 = d_{5,6}^2,$$

e obtemos as duas soluções possíveis: x_6^0 e x_6^1 . Neste momento, nada pode ser dito sobre a viabilidade dos pontos, pois não temos mais vértices anteriores ligados a v_6 . Para continuar, temos então que fazer uma escolha (por exemplo, x_6^0), mas ao mesmo tempo, lembrar que há outra possibilidade.

Vamos ao vértice seguinte v_7 . Considerando o sistema

$$\|x_7 - x_4^0\|^2 = d_{4,7}^2,$$

$$\|x_7 - x_5^0\|^2 = d_{5,7}^2,$$

$$\|x_7 - x_6^0\|^2 = d_{6,7}^2,$$

obtemos mais duas soluções possíveis: x_7^0 e x_7^1 . Neste caso, existem duas arestas “adicionais”, $\{v_1, v_7\} \in E$ e $\{v_2, v_7\} \in E$, que devem ser utilizadas para testar cada uma das possíveis posições para v_7 . Ou seja,

$$\|x_7^0 - x_1\| = d_{1,7} \text{ ou } \|x_7^1 - x_1\| = d_{1,7}$$

e

$$\|x_7^0 - x_2\| = d_{2,7} \text{ ou } \|x_7^1 - x_2\| = d_{2,7}.$$

Novamente, supondo que os pontos x_1, x_4^0, x_5^0, x_6^0 e x_2, x_4^0, x_5^0, x_6^0 não sejam coplanares, temos duas possibilidades:

- Nenhuma das possíveis soluções é viável;
- Apenas uma das possíveis soluções é viável.

Exercício 22. *Explique as duas possibilidades dadas acima.*

O primeiro caso pode ocorrer, caso a seleção que fizemos para v_6 tenha sido “errada”. Então, temos que “retroceder”, escolher x_6^1 e repetir o processo. Quando temos mais de uma aresta adicional, devemos testá-las, uma por uma. Basta uma equação não ser satisfeita para inviabilizar a posição em questão.

No segundo caso, temos uma situação parecida com a do vértice v_5 , mas com duas arestas adicionais, no lugar de apenas uma: $\{v_1, v_7\} \in E$ e $\{v_2, v_7\} \in E$ (teoricamente, podemos usar qualquer uma dessas arestas para a escolha entre x_7^0 e x_7^1).

Vamos supor que tenha ocorrido o primeiro caso (x_7^0 e x_7^1 são inviáveis). Então, retornamos, escolhemos x_6^1 e resolvemos um novo sistema quadrático, obtendo y_7^0 e y_7^1 . Como temos $\{v_1, v_7\} \in E$ (supondo que os pontos $\{x_1, x_4, x_5^0, x_6^1\}$ não sejam coplanares), apenas uma das posições é válida, digamos y_7^1 . Portanto, finalmente, encontramos uma solução do problema:

$$x_1, x_2, x_3, x_4^0, x_5^0, x_6^1, y_7^1.$$

Exercício 23. *Como garantir que y_7^0 ou y_7^1 irá satisfazer as restrições $\{v_1, v_7\} \in E$ e $\{v_2, v_7\} \in E$?*

No exemplo acima, caso $\{v_1, v_6\} \in E$ ou $\{v_4, v_6\} \in E$, não teríamos mais de uma opção para v_6 e a solução seria encontrada rapidamente, sem a necessidade de “retornos” na árvore do BP.

Como pode ser percebido, decidimos fazer uma “busca pela esquerda”, isto é, quando não temos arestas adicionais, optamos por escolher a solução x_i^0 do sistema quadrático associado. Outras opções podem ser definidas [Fidalgo et al. \(2018\)](#) e [Nucci, Nogueira e Lavor \(2013\)](#), mas temos que ter o cuidado para não nos “perdermos”, principalmente na hora de fazer um retorno na árvore.

Baseado no exemplo anterior, vimos também que, essencialmente, o procedimento de resolução do problema pode ser definido como uma sequência de sistemas quadráticos e testes de viabilidade (quando existem arestas adicionais).

Com esses dois subproblemas em mente, podemos dividir as arestas E de um DDGP, com ordenação v_1, \dots, v_n , em dois conjuntos disjuntos,

$$E = E_d \cup E_p,$$

onde

$$E_d = \{\{a_4, v_4\}, \{b_4, v_4\}, \{c_4, v_4\}, \dots, \{a_n, v_n\}, \{b_n, v_n\}, \{c_n, v_n\}\}$$

é o conjunto das *arestas de discretização* e

$$E_p = E - E_d$$

é o conjunto das *arestas de poda*.

As arestas de discretização “moldam” o espaço de busca como uma árvore binária e as arestas de poda indicam quando é possível fazer uma poda na árvore para reduzir o espaço de busca.

Para resolver o problema, devemos tentar descer pela árvore, da raiz até o último nível, passando por todos os testes de viabilidade, onde a solução será dada pelo caminho percorrido entre o primeiro e o último nível da árvore, considerando apenas os nós viáveis. Em alguns casos, temos que retroceder na árvore para refazer o caminho, o que pode aumentar (exponencialmente) o custo computacional da busca.

Exercício 24. *Em que situação pode se dar esse aumento exponencial?*

Podemos descer pela árvore até o último nível, sem nenhum retorno, em dois casos extremos:

- $E_p = \emptyset$,
- $\forall v_i, i = 5, \dots, n$, existem exatamente 4 vértices (que geram pontos não coplanares) anteriores a v_i , a_i, b_i, c_i, d_i , com $\{\{a_i, v_i\}, \{b_i, v_i\}, \{c_i, v_i\}, \{d_i, v_i\}\} \subset E$.

Exercício 25. *Prove que, nesses dois casos, uma solução é encontrada rapidamente, com um custo proporcional a n .*

No primeiro caso, basta escolher qualquer uma das soluções dos sistemas quadráticos

$$\|x_{v_i} - x_{a_i}\|^2 = d_{a_i, v_i}^2,$$

$$\|x_{v_i} - x_{b_i}\|^2 = d_{b_i, v_i}^2,$$

$$\|x_{v_i} - x_{c_i}\|^2 = d_{c_i, v_i}^2,$$

pois não há nenhum teste de viabilidade a ser feito.

No segundo caso, basta resolver os sistemas quadráticos

$$\begin{aligned} \|x_{v_i} - x_{a_i}\|^2 &= d_{a_i, v_i}^2, \\ \|x_{v_i} - x_{b_i}\|^2 &= d_{b_i, v_i}^2, \\ \|x_{v_i} - x_{c_i}\|^2 &= d_{c_i, v_i}^2, \\ \|x_{v_i} - x_{d_i}\|^2 &= d_{d_i, v_i}^2, \end{aligned}$$

com solução única, também com nenhum teste de viabilidade a ser feito (o único teste a ser realizado foi incluído no sistema).

Exercício 26. *Observando que as arestas de discretização E_d definem os sistemas quadráticos e as arestas de poda E_p definem os testes de viabilidade, exiba os principais passos do algoritmo BP.*

3.5 Ordenação de Vértices no DMDGP

Voltemos agora à ordenação dos vértices do DDGP, onde é dada a clique inicial $\{v_1, v_2, v_3\}$ e, para todo $v_i, i = 4, \dots, n$, existem (pelo menos) três vértices anteriores a_i, b_i, c_i , com

$$\{\{a_i, v_i\}, \{b_i, v_i\}, \{c_i, v_i\}\} \subset E,$$

e

$$d_{a_i, b_i} + d_{b_i, c_i} > d_{a_i, c_i}.$$

Admitindo que não há erro algum nas distâncias relacionadas às arestas de discretização E_d , as desigualdades triangulares são satisfeitas, mas não necessariamente de maneira estrita (já sabemos que a desigualdade estrita é necessária para evitar uma quantidade não enumerável de soluções).

O que pode dar problema, na prática, não é o fato de exigirmos $>$, no lugar de \geq . Dependendo da instância do DDGP, pode faltar alguma distância envolvendo os vértices a_i, b_i, c_i fazendo com que o sistema

$$\begin{aligned} \|x_{v_i} - x_{a_i}\|^2 &= d_{a_i, v_i}^2 \\ \|x_{v_i} - x_{b_i}\|^2 &= d_{b_i, v_i}^2 \\ \|x_{v_i} - x_{c_i}\|^2 &= d_{c_i, v_i}^2 \end{aligned}$$

não tenha solução em \mathbb{R}^3 . Uma saída seria exigir que, para todo $v_i, i = 4, \dots, n$, as três distâncias entre os vértices a_i, b_i, c_i sejam conhecidas. Além disso, podemos assumir também que os vértices a_i, b_i, c_i sejam imediatamente anteriores a v_i , o que pode, de fato, acontecer em várias aplicações [Lavor, Liberti e Mucherino \(2013\)](#).

Estamos interessados em encontrar uma ordenação em que os vértices utilizados na construção de cada sistema quadrático formem uma clique e sejam imediatamente anteriores ao vértice que se deseja realizar.

Voltemos ao mesmo problema da seção anterior, com grafo $G = (V, E)$ dado por

$$\begin{aligned} V &= \{p, q, r, s, t, u, v\}, \\ E &= \{\{p, q\}, \{p, r\}, \{p, s\}, \{p, u\}, \{p, v\}, \\ &\quad \{q, r\}, \{q, s\}, \{q, t\}, \{q, u\}, \{q, v\}, \\ &\quad \{r, s\}, \{r, t\}, \{r, v\}, \\ &\quad \{s, t\}, \{s, v\}, \\ &\quad \{t, u\}, \{t, v\}, \\ &\quad \{v, u\}\}, \end{aligned}$$

considerando uma nova ordenação

$$V = \{p, u, v, q, t, r, s\}.$$

Para facilitar, denotemos

$$u_1 = p, u_2 = u, u_3 = v, u_4 = q, u_5 = t, u_6 = r, u_7 = s,$$

onde usamos u_i , no lugar de v_i , apenas para enfatizar que estamos utilizando uma ordenação diferente da anterior.

Podemos verificar que também temos uma ordenação DDGP, pois além da clique inicial $\{u_1, u_2, u_3\}$, temos também as seguintes cliques:

$$\begin{aligned} &\{u_1, u_2, u_3, u_4\}, \\ &\{u_2, u_3, u_4, u_5\}, \\ &\{u_3, u_4, u_5, u_6\}, \\ &\{u_4, u_5, u_6, u_7\}. \end{aligned}$$

Para acompanharmos melhor o “andamento” do BP, destacamos:

$$\begin{aligned} \{\{u_1, u_4\}, \{u_2, u_4\}, \{u_3, u_4\}\} &\subset E, \\ \{\{u_2, u_5\}, \{u_3, u_5\}, \{u_4, u_5\}\} &\subset E, \\ \{\{u_1, u_6\}, \{u_3, u_6\}, \{u_4, u_6\}, \{u_5, u_6\}\} &\subset E, \\ \{\{u_1, u_7\}, \{u_3, u_7\}, \{u_4, u_7\}, \{u_5, u_7\}, \{u_6, u_7\}\} &\subset E. \end{aligned}$$

Para fixar u_4 , montamos o sistema quadrático abaixo:

$$\begin{aligned} \|x_4 - x_1\|^2 &= d_{1,4}^2, \\ \|x_4 - x_2\|^2 &= d_{2,4}^2, \\ \|x_4 - x_3\|^2 &= d_{3,4}^2. \end{aligned}$$

Escolhendo uma das duas possibilidades para posicionar u_4 (digamos x_4^0), obtemos o sistema quadrático seguinte para posicionar u_5 :

$$\begin{aligned} \|x_5 - x_2\|^2 &= d_{2,5}^2, \\ \|x_5 - x_3\|^2 &= d_{3,5}^2, \\ \|x_5 - x_4^0\|^2 &= d_{4,5}^2. \end{aligned}$$

Novamente, escolhemos uma das duas possibilidades para posicionar u_5 (digamos x_5^0) e obtemos um novo sistema quadrático:

$$\begin{aligned} \|x_6 - x_3\|^2 &= d_{3,6}^2, \\ \|x_6 - x_4^0\|^2 &= d_{4,6}^2, \\ \|x_6 - x_5^0\|^2 &= d_{5,6}^2. \end{aligned}$$

Temos também $\{u_1, u_6\} \in E$, que usamos para testar cada uma das possíveis posições obtidas para u_6 (x_6^0 e x_6^1). Isto é,

$$\|x_6^0 - x_1\| = d_{1,6}$$

ou

$$\|x_6^1 - x_1\| = d_{1,6}?$$

Pode ser que nenhuma das equações seja satisfeita, implicando que fizemos uma escolha “errada” para u_5 (vamos supor que este foi o caso). Temos que retornar e repetir o procedimento, utilizando x_5^1 . O novo sistema quadrático é, então:

$$\|x_6 - x_3\|^2 = d_{3,6}^2,$$

$$\|x_6 - x_4^0\|^2 = d_{4,6}^2,$$

$$\|x_6 - x_5^1\|^2 = d_{5,6}^2.$$

Utilizamos novamente $\{u_1, u_6\} \in E$ para testar cada uma das novas possíveis posições obtidas para u_6 (y_6^0 e y_6^1). Ou seja,

$$\|y_6^0 - x_1\| = d_{1,6}$$

ou

$$\|y_6^1 - x_1\| = d_{1,6}?$$

Supondo que os pontos $\{x_1, x_3, x_4^0, x_5^1\}$ não sejam coplanares, apenas uma das equações deve ser satisfeita, digamos a primeira. Descartamos, então, y_6^1 e consideramos y_6^0 . O novo sistema quadrático é

$$\|x_7 - x_4^0\|^2 = d_{4,7}^2,$$

$$\|x_7 - x_5^1\|^2 = d_{5,7}^2,$$

$$\|x_7 - y_6^0\|^2 = d_{6,7}^2.$$

Usando $\{u_1, u_7\} \in E$ e $\{u_3, u_7\} \in E$ para testar cada uma das novas possíveis posições obtidas para u_7 (x_7^0 e x_7^1), temos:

$$\|x_7^0 - x_1\| = d_{1,7} \text{ ou } \|x_7^1 - x_1\| = d_{1,7}$$

e

$$\|x_7^0 - x_3\| = d_{3,7} \text{ ou } \|x_7^1 - x_3\| = d_{3,7}.$$

Vamos supor que x_7^1 seja o ponto selecionado. Portanto, chegamos à solução do problema:

$$x_1, x_2, x_3, x_4^0, x_5^1, y_6^0, x_7^1.$$

Exercício 27. Além do fato de garantir que todos os sistemas quadráticos tenham solução, qual foi a vantagem da nova ordenação

$$\{p, u, v, q, t, r, s\}$$

em comparação à ordenação anterior

$$\{r, q, t, s, v, u, p\}?$$

Observe que, em uma ordenação DDGP v_1, \dots, v_n , onde o conjunto $\{v_{i-3}, v_{i-2}, v_{i-1}, v_i\}$ é uma clique, para $i = 4, \dots, n$, todas as distâncias entre vértices consecutivos,

$$d_{1,2}, \dots, d_{n-1,n},$$

bem como todos os ângulos planos formados por três vértices consecutivos,

$$\theta_{1,3}, \dots, \theta_{n-2,n},$$

são dados *a priori*. Os valores $d_{1,2}, \dots, d_{n-1,n}$ são, obviamente, obtidos pela própria definição da ordenação, e os valores $\theta_{1,3}, \dots, \theta_{n-2,n}$ podem ser calculados pela Lei dos Cossenos.

A primeira consequência importante é que, usando coordenadas internas (ver capítulo 2), a estrutura 3D que estamos procurando pode ser determinada pelos ângulos de torção

$$\omega_{1,4}, \dots, \omega_{n-3,n},$$

definidos por quatro vértices consecutivos (recordemos que o ângulo de torção $\omega_{i-3,i}$ é o ângulo entre os vetores normais dos planos associados aos vértices $v_{i-3}, v_{i-2}, v_{i-1}$ e v_{i-2}, v_{i-1}, v_i , respectivamente).

A segunda implicação, ainda mais importante, é que podemos substituir a resolução dos sistemas quadráticos por algo bem mais simples e estável numericamente.

Com as distâncias dadas pela clique $\{v_{i-3}, v_{i-2}, v_{i-1}, v_i\}$, os valores dos cossenos dos ângulos de torção também podem ser calculados previamente [Lavor, Alves et al. \(2015\)](#), usando a fórmula

$$\cos(\omega_{i-3,i}) = \frac{2d_{i-2,i-1}^2 \left(d_{i-3,i-2}^2 + d_{i-2,i}^2 - d_{i-3,i}^2 \right) - (d_{i-3,i-2,i-1}) (d_{i-2,i-1,i})}{\sqrt{4d_{i-3,i-2}^2 d_{i-2,i-1}^2 - (d_{i-3,i-2,i-1})^2} \sqrt{4d_{i-2,i-1}^2 d_{i-2,i}^2 - (d_{i-2,i-1,i})^2}} \quad (3.2)$$

onde

$$\begin{aligned} d_{i-3,i-2,i-1} &= d_{i-3,i-2}^2 + d_{i-2,i-1}^2 - d_{i-3,i-1}^2, \\ d_{i-2,i-1,i} &= d_{i-2,i-1}^2 + d_{i-2,i}^2 - d_{i-1,i}^2, \end{aligned}$$

para $i = 4, \dots, n$.

Estamos prontos para definir um novo problema: o *Discretizable Molecular DGP* (DMDGP):

Definição 28. Considere o grafo $G = (V, E, d)$ de um DGP e uma ordenação em V , denotada por v_1, \dots, v_n , tal que

- existe uma realização para v_1, v_2, v_3 ;
- para todo v_i , $i = 4, \dots, n$, existem três vértices adjacentes, e imediatamente anteriores,

$$\{\{v_{i-3}, v_i\}, \{v_{i-2}, v_i\}, \{v_{i-1}, v_i\}\} \subset E,$$

onde

$$d_{i-3, i-2} + d_{i-2, i-1} > d_{i-3, i-1}.$$

Encontre uma função $x: V \rightarrow \mathbb{R}^3$, tal que

$$\forall \{v_i, v_j\} \in E, \quad \|x_i - x_j\| = d_{i,j}. \quad (3.3)$$

As informações de distâncias das cliques definidas por

$$\{v_{i-3}, v_{i-2}, v_{i-1}, v_i\}$$

de um grafo $G = (V, E, d)$ de um DMDGP, com a ordenação v_1, \dots, v_n , permite que os valores abaixo da estrutura 3D que desejamos obter sejam previamente calculados:

- $d_{1,2}, \dots, d_{n-1,n}$;
- $\theta_{1,3}, \dots, \theta_{n-2,n}$;
- $\cos(\omega_{1,4}), \dots, \cos(\omega_{n-3,n})$.

Isso significa que os dois valores possíveis, para todos os ângulos de torção $\omega_{i-3,i}$, $i = 4, \dots, n$, podem ser conhecidos antes da resolução do DMDGP.

Exercício 29. Usando as matrizes do capítulo 2 para obter as coordenadas cartesianas das duas possíveis soluções para um dado vértice v_i (fixadas as posições de v_{i-3} , v_{i-2} , v_{i-1}), explique como realizar uma poda na árvore do BP, tendo uma distância adicional $d_{j,i}$ de um DMDGP, para $j < i - 3$.

É importante mencionar que encontrar uma ordenação DMDGP pode ser complicado, pois é um problema NP-difícil [Cassioli, Günlük et al. \(2015\)](#). Ou seja, escapamos da resolução dos sistemas quadráticos, mas aumentamos exponencialmente o custo para encontrar a nova ordenação (ver próximo capítulo).

Em qualquer instância do DMDGP, as duas posições para v_4 ($x_4^0, x_4^1 \in \mathbb{R}^3$) podem ser consideradas, já que não há arestas adicionais para inviabilizar uma delas. Isso implica que, para qualquer solução encontrada na subárvore esquerda, tendo o nó x_4^0 como raiz, existe uma outra simétrica em relação ao plano definido por x_1, x_2, x_3 [Lavor, Liberti, Maculan et al. \(2012b\)](#). Uma consequência imediata desse fato é que a cardinalidade do conjunto solução é um número par. Entretanto, desde os primeiros resultados computacionais obtidos para o DMDGP [Liberti, Lavor e Maculan \(2008\)](#), a quantidade de soluções era sempre potência de dois. Apenas recentemente, usando *Teoria de Grupos*, conseguimos apresentar uma demonstração matemática para esse fato [Liberti, Masson et al. \(2014\)](#), relacionada a uma *simetria* envolvendo as soluções do DMDGP.

Para dar uma ideia da importância desse resultado, consideremos o conjunto S definido abaixo, para um grafo $G = (V, E, d)$ de um DMDGP:

$$S = \{v \in V : \exists \{u, w\} \in E \text{ tal que } u + 3 < v \leq w\}.$$

Para simplificar a notação, $u + 3$ é o terceiro vértice depois de u , e $u - 3$ é o terceiro vértice anterior a u .

Vamos identificar quem são os elementos de S , considerando a ordenação

$$v_1, \dots, v_n \in V.$$

O primeiro candidato é v_4 , que estará em S se não existir nenhuma aresta $\{u, w\} \in E$ tal que $u + 3 < v \leq w$. Se existir algum $u \in V$ com tal característica, teremos $u < v_4 - 3$, o que é impossível, pois $v_4 - 3 = v_1$. Ou seja, $v_4 \in S$, para qualquer DMDGP.

Vejam os que acontece com v_5 (supondo que o DMDGP tenha solução):

- Se não existir $\{u, v_5\} \in E$, para $u < v_5 - 3$, temos que considerar dois casos:
 - Se não existir $\{u, w\} \in E$, tal que $u + 3 < v_5 < w$, então $v_5 \in S$.
 - Caso exista $\{u, w\} \in E$, tal que $u + 3 < v_5 < w$, então $v_5 \notin S$.

Exercício 30. *Mostre que, se existir $\{u, v_5\} \in E$, tal que $u < v_5 - 3$, teremos $\{v_1, v_5\} \in E$, o que implica que apenas uma das possibilidades para v_5 é viável: ou x_5^0 ou x_5^1 .*

Exercício 31. *Para qualquer DMDGP, o conjunto S pode ser obtido antes mesmo de aplicar o BP?*

A importância do conjunto S é que ele identifica outros *planos de simetria*, próprios de cada DMDGP (como vimos acima, os vértices $\{v_1, v_2, v_3\}$ definem um plano de simetria válido para qualquer DMDGP).

Por exemplo, se $v_5 \in S$, as duas posições para v_5 são viáveis (x_5^0 e x_5^1) e, ao mesmo tempo, fazem parte de soluções do problema, pois não há aresta adicional para tornar tais posições inviáveis. Na verdade, pode ser demonstrado que os pontos x_2, x_3, x_4 definem um novo plano de simetria [Liberti, Masson et al. \(2014\)](#). Isso significa que, se existe uma solução iniciando com $x_1, x_2, x_3, x_4^0, x_5^0$ ou $x_1, x_2, x_3, x_4^1, x_5^0$, teremos uma outra, simétrica em relação ao plano definido por x_2, x_3, x_4 .

Consideremos a mesma instância DMDGP dada anteriormente, voltando a usar a notação v_i :

$$\begin{aligned} V &= \{v_1, v_2, v_3, v_4, v_5, v_6, v_7\}, \\ E &= \{\{v_1, v_2\}, \{v_1, v_3\}, \{v_1, v_4\}, \{v_1, v_6\}, \{v_1, v_7\}, \\ &\quad \{v_2, v_3\}, \{v_2, v_4\}, \{v_2, v_5\}, \\ &\quad \{v_3, v_4\}, \{v_3, v_5\}, \{v_3, v_6\}, \{v_3, v_7\}, \\ &\quad \{v_4, v_5\}, \{v_4, v_6\}, \{v_4, v_7\}, \\ &\quad \{v_5, v_6\}, \{v_5, v_7\}, \\ &\quad \{v_6, v_7\}\}. \end{aligned}$$

Como $\{v_1, v_7\} \in E$, $S = \{v_4\}$, implicando em apenas um plano de simetria, definido por x_1, x_2, x_3 , comum a qualquer DMDGP.

Exercício 32. *Para o exemplo acima, mostre que existem apenas duas soluções.*

A partir da solução já obtida pelo BP,

$$x_1, x_2, x_3, x_4^0, x_5^1, x_6^0, x_7^1,$$

temos uma outra, simétrica ao plano definido por $\{x_1, x_2, x_3\}$.

Supondo que $\{v_1, v_7\} \notin E$ e $\{v_3, v_7\} \notin E$, o conjunto S incluiria mais um elemento, gerando um outro plano de simetria, definido por $\{v_4, v_5, v_6\}$.

Exercício 33. *Verifique que, removendo $\{v_1, v_7\}$ e $\{v_3, v_7\}$ das arestas do exemplo acima, $S = \{v_4, v_7\}$.*

Considerando, então, a nova instância dada por

$$\begin{aligned}
 V &= \{v_1, v_2, v_3, v_4, v_5, v_6, v_7\}, \\
 E &= \{\{v_1, v_2\}, \{v_1, v_3\}, \{v_1, v_4\}, \{v_1, v_6\}, \\
 &\quad \{v_2, v_3\}, \{v_2, v_4\}, \{v_2, v_5\}, \\
 &\quad \{v_3, v_4\}, \{v_3, v_5\}, \{v_3, v_6\}, \\
 &\quad \{v_4, v_5\}, \{v_4, v_6\}, \{v_4, v_7\}, \\
 &\quad \{v_5, v_6\}, \{v_5, v_7\}, \\
 &\quad \{v_6, v_7\}\},
 \end{aligned}$$

vamos representar a primeira solução por uma sequência de zeros e uns, denotando as três primeiras posições por 0, 0, 0:

$$s_1 = (0, 0, 0, 0, 1, 0, 0).$$

Como sabemos que temos uma simetria no vértice v_7 , uma outra solução é dada por

$$s_2 = (0, 0, 0, 0, 1, 0, 1).$$

Por conta da simetria no vértice v_4 , obtemos mais duas soluções, a partir das anteriores, dadas por

$$s_3 = (0, 0, 0, 1, 0, 1, 1)$$

e

$$s_4 = (0, 0, 0, 1, 0, 1, 0).$$

O estudo das simetrias do DMDGP foi motivado por uma questão puramente teórica, relacionada à quantidade de soluções obtidas nos resultados computacionais, que sempre resultava como uma potência de dois. Além da importância teórica, as simetrias do DMDGP têm duas consequências práticas de grande importância:

- Usando apenas os dados de entrada do DMDGP, podemos calcular, *a priori*, a cardinalidade do conjunto solução, dada por $2^{|S|}$ [Liberti, Lavor, Alencar et al. \(2013\)](#);
- Para encontrar todas as soluções de um DMDGP, basta usar o BP, ou qualquer outro método, para encontrar uma única solução, pois todas as outras podem ser obtidas, em tempo linear, usando as simetrias associadas [Mucherino, Lavor e Liberti \(2012a\)](#).

4

Repetição de Vértices no DMDGP

4.1 Reordenação no DMDGP

Como estamos interessados em determinar a estrutura 3D da cadeia principal de uma proteína, a sequência de átomos $N^i, C_\alpha^i, C^i, i = 1, \dots, p$ (p é o número de aminoácidos), seria a opção natural para tentar definir uma ordenação DMDGP. No entanto, essa ordenação não fornece todas as distâncias $d_{i-3,i}$ necessárias para definir um DMDGP.

Lembrando que os experimentos de RMN fornecem distâncias entre átomos de hidrogênio que estão a menos de 5 Å de distância um dos outros, uma ordenação envolvendo apenas hidrogênios foi proposta em [Lavor, Mucherino et al. \(2011\)](#), mas considerando distâncias precisas entre eles. Uma tentativa de superar essa limitação foi incluir os átomos da cadeia principal [Lavor, Liberti e Mucherino \(2013\)](#), permitindo a repetição de vértices.

A ideia é permitir a repetição, de modo que pelo menos três predecessores adjacentes a cada vértice na ordenação possam também ser contíguos. Essa ordenação, com repetição de alguns vértices, é chamada de *reordenação*, definida abaixo.

Inicialmente, devemos particionar o conjunto de arestas E do grafo $G = (V, E, d)$,

$$E = E' \cup E'',$$

onde

$$\{u, v\} \in E' \Leftrightarrow d_{u,v} \in (0, \infty)$$

e

$$\{u, v\} \in E'' \Leftrightarrow d_{u,v} = [\underline{d}_{u,v}, \bar{d}_{u,v}],$$

com $0 < \underline{d}_{u,v} < \bar{d}_{u,v}$.

Observe que a função d é agora mais geral, incluindo as distâncias intervalares associadas às incertezas dos dados de RMN. Como veremos em seguida, E' representa pares de átomos separados por uma ou duas ligações covalentes (valores precisos de distâncias), e E'' representa pares de átomos de hidrogênio cujas distâncias são fornecidas por RMN (distâncias intervalares).

Definição 34. *Uma reordenação nos vértices do grafo $G = (V, E, d)$ de um DMDGP é uma sequência $r: \mathbb{N} \rightarrow V \cup \{0\}$, com comprimento $|r| \in \mathbb{N}$ ($r_i = r(i) = 0$, para todo $i > |r|$), tal que*

1. $\{r_1, r_2\}, \{r_1, r_3\}, \{r_2, r_3\} \in E'$;
2. $\forall i \in \{4, \dots, |r|\}, \{r_{i-1}, r_i\}, \{r_{i-2}, r_i\} \in E'$;
3. $\forall i \in \{4, \dots, |r|\}, \{r_{i-3}, r_i\} \in E' \cup E''$ ou $r_{i-3} = r_i$.

A primeira propriedade diz que $d_{r_1, r_2}, d_{r_1, r_3}, d_{r_2, r_3} \in (0, \infty)$ e a segunda diz que $d_{r_{i-1}, r_i}, d_{r_{i-2}, r_i} \in (0, \infty)$, $i = 4, \dots, |r|$. Ou seja, todas essas distâncias são valores precisos e maiores do que zero.

Na terceira propriedade, existem três possibilidades para d_{r_{i-3}, r_i} , $i = 4, \dots, |r|$:

- $d_{r_{i-3}, r_i} \in (0, \infty)$, quando o par $\{r_{i-3}, r_i\}$ está relacionado a átomos separados por uma ou duas ligações covalentes;
- $d_{r_{i-3}, r_i} = [\underline{d}_{r_{i-3}, r_i}, \bar{d}_{r_{i-3}, r_i}]$, com $0 < \underline{d}_{r_{i-3}, r_i} < \bar{d}_{r_{i-3}, r_i}$, quando $\{r_{i-3}, r_i\}$ é um par de átomos de hidrogênio cuja distância foi detectada pela RMN;
- $d_{r_{i-3}, r_i} = 0$, quando há uma repetição na ordenação ($r_{i-3} = r_i$).

Quando $r_i = r_j$, para algum $i \neq j$ ($r_{i-3} = r_i$ é um caso particular), teremos $d_{r_i, r_j} = 0$.

Exercício 35. *Durante a execução do BP, o que acontece quando $d_{r_i, r_j} = 0$, para algum $i \neq j$?*

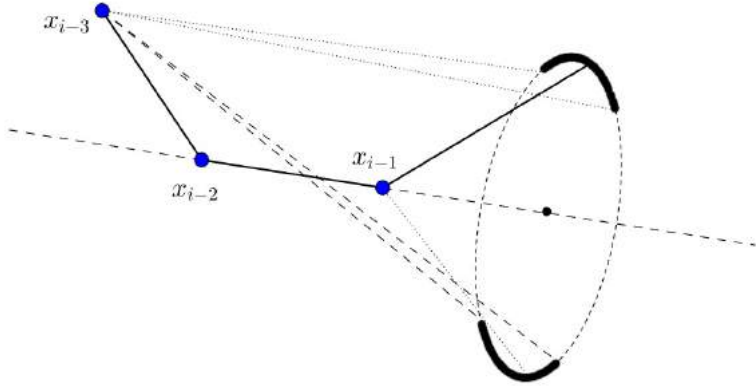


Figura 4.1: Distância intervalar e arcos correspondentes.

Se a repetição de vértices for usada de maneira inadequada, poderemos ter um triângulo com um lado de comprimento zero, implicando numa infinidade de possíveis posições para um dado átomo. Assim, para manter a natureza combinatoria do problema, a repetição de vértices é permitida apenas entre pares $\{r_i, r_j\}$ com $|i - j| \geq 3$.

Exercício 36. *Com a repetição ocorrendo apenas entre pares $\{r_i, r_j\}$ com $|i - j| \geq 3$, não há ramificação no nível $\max\{i, j\}$ da árvore de busca correspondente. Como a repetição de vértices poderia ser utilizada para controlar possíveis erros numéricos durante a execução do BP?*

Para entender o que acontece quando $\{r_{i-3}, r_i\} \in E''$, podemos reescrever a expressão (3.2) como

$$\cos(\omega_{i-3,i}) = \frac{a + b d_{i-3,i}^2}{c},$$

onde $a, b, c \in \mathbb{R}$ e $d_{i-3,i} \in [\underline{d}_{r_{i-3},r_i}, \bar{d}_{r_{i-3},r_i}]$.

Exercício 37. *Como podemos garantir que $a, b, c \in \mathbb{R}$?*

Considerando $\omega_{i-3,i} = 0$ e $\omega_{i-3,i} = 2\pi$, obtemos o menor valor para $\underline{d}_{r_{i-3},r_i}$ (d_{r_{i-3},r_i}^{\min}) e o maior valor para \bar{d}_{r_{i-3},r_i} (d_{r_{i-3},r_i}^{\max}), respectivamente, implicando em

$$[\underline{d}_{r_{i-3},r_i}, \bar{d}_{r_{i-3},r_i}] \subset [d_{r_{i-3},r_i}^{\min}, d_{r_{i-3},r_i}^{\max}].$$

Quando $d_{i-3,i}$ é um valor preciso ($d_{i-3,i} \in \mathbb{R}$), com $d_{r_{i-3},r_i}^{\min} < d_{i-3,i} < d_{r_{i-3},r_i}^{\max}$, obtemos dois valores possíveis para $\omega_{i-3,i}$, associados a duas posições em \mathbb{R}^3 para r_i . Entretanto, quando $d_{i-3,i} = [d_{r_{i-3},r_i}, \bar{d}_{r_{i-3},r_i}]$, com $d_{r_{i-3},r_i}^{\min} < d_{r_{i-3},r_i} < \bar{d}_{r_{i-3},r_i} < d_{r_{i-3},r_i}^{\max}$, obtemos dois intervalos para $\omega_{i-3,i}$, relacionados a dois arcos em \mathbb{R}^3 para r_i (Fig. 4.1).

Exercício 38. *Mostre que uma reordenação corresponde a uma ordenação DMDGP, onde alguns dos pares $\{r_i, r_j\}$, com $|i - j| \geq 3$, podem não corresponder a distâncias precisas.*

O conceito de reordenação foi um passo fundamental na aplicação do DMDGP como um modelo matemático para problemas relacionados à determinação de estruturas de proteínas 3D usando dados de RMN.

Usando a ideia de reordenação, os autores de [Lavor, Liberti e Mucherino \(2013\)](#) desenvolveram uma extensão do algoritmo BP, chamada *iBP*, onde alguns valores são selecionados dentro dos intervalos $[d_{r_{i-3},r_i}, \bar{d}_{r_{i-3},r_i}]$, a fim de continuar a pesquisa na árvore de busca (note que o espaço de busca deixa de ser uma árvore binária).

Além do aumento exponencial no espaço de busca do problema, resultados computacionais apresentados em [Cassoli, Bordiaux et al. \(2015\)](#) e [Gonçalves, Mucherino et al. \(2017\)](#) revelaram a principal dificuldade do *iBP*: mesmo com amostras refinadas, não há garantia de que uma solução será encontrada.

Essencialmente, existem duas razões para essa dificuldade:

- As reordenações apresentadas em [Gonçalves, Mucherino et al. \(2017\)](#) e [Lavor, Liberti e Mucherino \(2013\)](#) contêm pares $\{r_{i-3}, r_i\}$ cujas distâncias intervalares não estão associadas a dados de RMN. Isto é,

$$[d_{r_{i-3},r_i}, \bar{d}_{r_{i-3},r_i}] = [d_{r_{i-3},r_i}^{\min}, d_{r_{i-3},r_i}^{\max}], \quad (4.1)$$

implicando que os valores das amostras deverão ser selecionados considerando um círculo, em vez de dois arcos;

- O processo de amostragem “transforma” o *iBP* em uma heurística.

Exercício 39. *Por que o *iBP* não garante mais que uma solução do problema será encontrada, mesmo com um custo exponencial?*

Resultados recentes [Alves, Lavor et al. \(2018\)](#), [Alves e Lavor \(2017\)](#) e [Lavor e Alves \(2019\)](#), usando álgebra de Clifford, propõem uma alternativa para evitar

o processo de amostragem, permitindo que o *iBP* explore o espaço de busca sem “perder” soluções. No entanto, para aplicar esses resultados, uma nova reordenação teve que ser definida para evitar, principalmente, o que acontece em (4.1).

4.2 A Reordenação *hc* (hand-crafted)

Dado um grafo $G = (V, E, d)$ de um DMDGP associado à cadeia principal de uma molécula de proteína ($\{N^k, C_\alpha^k, C^k\}$, $k = 1, \dots, p$), incluindo átomos de oxigênio O^k , ligados a C^k , e átomos de hidrogênio H_α^k e H^k , ligados a C_α^k e C^k , respectivamente (ver Fig. 4.2, para $p = 3$), temos a reordenação *hc* (do inglês *hand-crafted*) proposta em [Lavor, Liberti, Donald et al. \(2019\)](#):

$$\begin{aligned} hc = \{ & N^1, H^1, H^{1'}, C_\alpha^1, N^1, H_\alpha^1, C^1, C_\alpha^1, \dots, \\ & H^i, C_\alpha^i, O^{i-1}, N^i, H^i, C_\alpha^i, N^i, H_\alpha^i, C^i, C_\alpha^i, \dots, \\ & H^p, C_\alpha^p, O^{p-1}, N^p, H^p, C_\alpha^p, N^p, H_\alpha^p, C^p, C_\alpha^p, O^p, C^p, O^{p'} \}, \end{aligned} \quad (4.2)$$

onde $i = 2, \dots, p-1$, $H^{1'}$ é o segundo hidrogênio ligado a N^1 , e $O^{p'}$ é o segundo oxigênio ligado a C^p (a Fig. 4.2 ilustra essa ordenação para $p = 3$).

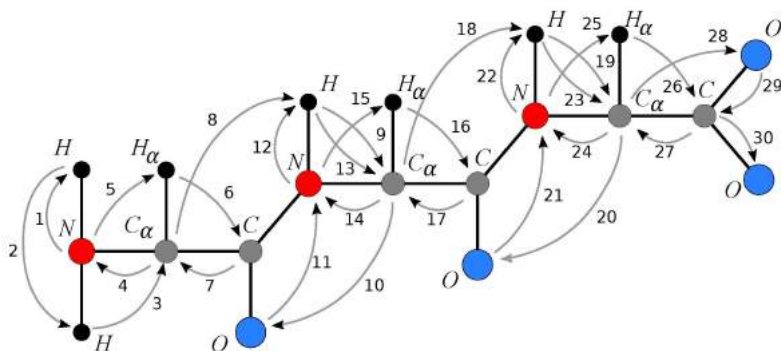


Figura 4.2: Reordenação *hc*.

Usando a reordenação *hc* e algumas propriedades da geometria de proteínas,

demonstra-se que as distâncias entre os pares de átomos de hidrogênio

$$\{H^{1'}, H_{\alpha}^1\}, \dots, \{H_{\alpha}^{i-1}, H^i\}, \{H^i, H_{\alpha}^i\}, \{H_{\alpha}^i, H^{i+1}\}, \dots, \{H^p, H_{\alpha}^p\}, \quad (4.3)$$

para $i = 2, \dots, p - 1$, são condições *suficientes* para representar o espaço de soluções do DMDGP como uma *árvore de busca* [Lavor, Liberti, Donald et al. \(2019\)](#). Além disso, usando o conceito de *quiralidade* [Donald \(2011\)](#), que define a orientação do tetraedro formado por $\{N^1, H^1, H^{1'}, C_{\alpha}^1\}$ e $\{C_{\alpha}^i, N^i, H_{\alpha}^i, C^i\}$, implicando em apenas uma posição possível para C_{α}^1 e C^i , também é provado, em [Lavor, Liberti, Donald et al. \(2019\)](#), que as ramificações na árvore de busca ocorrem apenas nos seguintes átomos de hidrogênio:

$$\{H_{\alpha}^1, \dots, H^i, H_{\alpha}^i, \dots, H^p, H_{\alpha}^p\}. \quad (4.4)$$

Como principais consequências desse resultado, destacamos:

- Se as distâncias relacionadas aos pares (4.4) forem valores precisos, o espaço de busca do DMDGP associado será finito, representado como uma árvore binária;
- Se as distâncias relacionadas aos pares (4.4) forem valores precisos e houver pelo menos uma distância adicional (dos dados de RMN) para cada hidrogênio da lista (4.4), haverá apenas uma solução do DMDGP associado que pode ser encontrada em tempo linear.

Exercício 40. *Embora as distâncias precisas e adicionais sejam hipóteses muito fortes, explique como essa informação conecta a cardinalidade do conjunto de soluções com a complexidade computacional do problema.*

Como os átomos H^i, H_{α}^i estão no mesmo aminoácido, a distância associada $d(H^i, H_{\alpha}^i)$ é detectada pela RMN. Embora os átomos H_{α}^{i-1}, H^i estejam em aminoácidos consecutivos, há apenas um ângulo de torção (definido por $\{N^{i-1}, C_{\alpha}^{i-1}, C^{i-1}, H^i\}$ relacionado com a posição de H^i , pois o ângulo de torção definido por $\{C_{\alpha}^{i-1}, C^{i-1}, N^{i-1}, H^i\}$ é, em geral, fixado em π radianos [Donald \(2011\)](#)). Na pior das hipóteses, supondo que a distância $d(H_{\alpha}^{i-1}, H^i)$ não esteja disponível, podemos usar informações “implícitas” associadas ao fato de que a distância não foi detectada [Agra et al. \(2017\)](#) ou algumas estimativas dadas em [Wüthrich \(1986\)](#).

A reordenação *hc* foi “desenhada à mão” por conta de um resultado teórico que diz que encontrar ordenações DMDGP é um problema NP-difícil [Cassoli, Günlük et al. \(2015\)](#). Entretanto, um resultado um tanto surpreendente é que o problema

se torna polinomial quando permitimos repetições “cuidadas” de vértices [Lavor, Souza et al. \(s.d.\)](#).

Para mais detalhes sobre a reordenação *hc*, ver [Lavor, Liberti, Donald et al. \(2019\)](#).

5

Geometria Molecular: do 3D ao 5D

5.1 Esferas e o DMDGP

Também como consequência da definição do DMDGP, temos uma interpretação geométrica interessante para o algoritmo BP. Dada uma ordenação v_1, \dots, v_n nos vértices do grafo $G = (V, E, d)$ associado, consideremos o sistema quadrático relacionado a $v_i, i = 4, \dots, n$,

$$\|x_i - x_{i-1}\|^2 = d_{i-1,i}^2,$$

$$\|x_i - x_{i-2}\|^2 = d_{i-2,i}^2,$$

$$\|x_i - x_{i-3}\|^2 = d_{i-3,i}^2,$$

onde $x_i \in \mathbb{R}^3$ é a posição procurada do vértice v_i , dados $x_{i-1}, x_{i-2}, x_{i-3} \in \mathbb{R}^3$ e as distâncias correspondentes $d_{i-1,i}, d_{i-2,i}, d_{i-3,i}$.

Geometricamente, resolver o sistema acima é obter a interseção de três esferas (ver Figura 5.1), $S(x_{i-1}, d_{i-1,i}), S(x_{i-2}, d_{i-2,i}), S(x_{i-3}, d_{i-3,i})$, centradas em $x_{i-1}, x_{i-2}, x_{i-3}$ e com raios $d_{i-1,i}, d_{i-2,i}, d_{i-3,i}$, respectivamente, dada por

$$D = S(x_{i-1}, d_{i-1,i}) \cap S(x_{i-2}, d_{i-2,i}) \cap S(x_{i-3}, d_{i-3,i}).$$

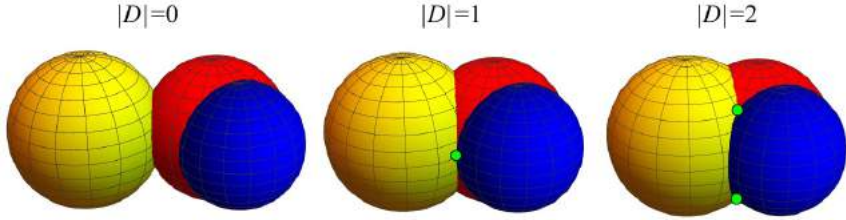


Figura 5.1: Interseção de três esferas.

Supondo que os pontos $x_{i-1}, x_{i-2}, x_{i-3}$ não sejam colineares, temos apenas três possibilidades:

1. $|D| = 0$;
2. $|D| = 1$;
3. $|D| = 2$.

Se não houver erros nas informações das distâncias [Lavor, Liberti e Mucherino \(2013\)](#), o primeiro caso não acontece. Teremos o segundo caso, quando os pontos $x_i, x_{i-1}, x_{i-2}, x_{i-3}$ forem coplanares, e o terceiro é o caso geral. Ou seja, sem erro algum nas distâncias dadas, temos

$$D = \{x_i^0, x_i^1\},$$

onde $x_i^0, x_i^1 \in \mathbb{R}^3$ representam o *branching* em cada iteração do BP (podendo acontecer $x_i^0 = x_i^1$, como mencionado acima).

Recordemos que podemos particionar as arestas de G como

$$E = E_d \cup E_p,$$

onde E_d são as arestas de discretização e E_p são as arestas de poda.

Quando $E_p \neq \emptyset$, temos esferas adicionais que devem ser acrescentadas à interseção anterior. Isto é, devemos fazer

$$P = D \cap \left(\bigcap_{j < i-3} S(x_j, d_{j,i}) \right).$$

Essa nova interseção, $\bigcap_{j < i-3} S(x_j, d_{j,i})$, é justamente o teste de viabilidade (o *pruning*), gerando apenas duas possibilidades para P :

- $|P| = 0$,
- $|P| = 1$.

No primeiro caso, houve alguma inviabilidade (obrigando um retorno na árvore do BP), e no segundo, temos o ponto que satisfaz todas as distâncias dadas aos pontos anteriores já posicionados.

Nos dois subproblemas fundamentais envolvidos na resolução de um DMDGP (o *branching* e o *pruning* do BP), a operação matemática envolvida é a *interseção de esferas* [Maioli, Lavor e Gonçalves \(2017\)](#).

Exercício 41. *Mesmo quando temos distâncias intervalares $d_{i-3,i}$, mostre que o problema continua sendo interseção de esferas, com a diferença que, nesse caso, temos a interseção entre duas esferas e uma casca esférica.*

A próxima seção apresenta um novo “modelo” para o espaço 3D, em que esferas são *objetos primitivos*.

5.2 O Modelo Conforme

No capítulo 2, usamos matrizes em $\mathbb{R}^{4 \times 4}$ para calcular as coordenadas cartesianas de uma dada molécula, em função das coordenadas internas. No \mathbb{R}^4 , além de rotações, podemos representar translações do \mathbb{R}^3 por meio de matrizes. Isto é, a translação $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$, definida por

$$T(x) = x + v,$$

onde $v \in \mathbb{R}^3$, pode ser representada no \mathbb{R}^4 por

$$\begin{bmatrix} 1 & 0 & 0 & v_1 \\ 0 & 1 & 0 & v_2 \\ 0 & 0 & 1 & v_3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ 1 \end{bmatrix} = \begin{bmatrix} x_1 + v_1 \\ x_2 + v_2 \\ x_3 + v_3 \\ 1 \end{bmatrix}.$$

A representação matricial permite que a composição de translações e rotações seja facilmente calculada pelo produto das matrizes associadas.

Exercício 42. *Verifique que as matrizes (2.1) codificam, ao mesmo tempo, uma translação e duas rotações.*

Ao “passarmos” para o \mathbb{R}^4 , estamos interessados em fazer operações com os “representantes” dos pontos do \mathbb{R}^3 , tirando proveito do “espaço aumentado”.

Para representar pontos do espaço 3D no \mathbb{R}^4 , precisamos incluir mais um vetor na base canônica $\{e_1, e_2, e_3\}$, que chamaremos de e_4 , dado por

$$e_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Ou seja, o representante $X \in \mathbb{R}^4$ de $x = (x_1, x_2, x_3) \in \mathbb{R}^3$ será dado por

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix},$$

onde, abusando um pouco da notação, escrevemos simplesmente

$$X = x + e_4.$$

Essa representação do espaço 3D é chamada de *Modelo Homogêneo*, muito utilizada em computação gráfica [Fernandes, Lavor e Oliveira Neto \(2017\)](#). Uma das características desse modelo é que, na verdade, o ponto X pode ser representado por

$$X = x + te_4,$$

onde t é um escalar não nulo. Para mais detalhes, sugerimos [Kanatani \(2015\)](#).

Além das operações que definem o espaço vetorial real \mathbb{R}^3 (adição de vetores e multiplicação por escalar), já vimos a necessidade de introduzir uma outra operação associada a conceitos geométricos fundamentais, como *comprimento*, *distância* e *ângulo*, dada pelo *produto interno* $u \cdot v$, definido por

$$u \cdot v = u_1v_1 + u_2v_2 + u_3v_3, \quad (5.1)$$

onde $u = (u_1, u_2, u_3) \in \mathbb{R}^3$ e $v = (v_1, v_2, v_3) \in \mathbb{R}^3$, que caracteriza o *Modelo Euclidiano* do espaço 3D.

O produto interno (5.1) satisfaz as seguintes propriedades, $\forall u, v, w \in \mathbb{R}^3$ e $\forall \alpha \in \mathbb{R}$, também usadas para apresentar de maneira abstrata o conceito de produto interno [Lima \(2016\)](#):

1. Comutatividade:

$$u \cdot v = v \cdot u;$$

2. Distributividade:

$$(u + v) \cdot w = (u \cdot w) + (v \cdot w),$$

$$u \cdot (v + w) = (u \cdot v) + (u \cdot w);$$

3. Homogeneidade:

$$(\alpha u) \cdot v = \alpha(u \cdot v),$$

$$u \cdot (\alpha v) = \alpha(u \cdot v);$$

4. Positividade:

$$u \neq 0 \Rightarrow u \cdot u > 0.$$

Com o conceito de distância, podemos definir operações que preservam distâncias, as isometrias [Lima \(2016\)](#). De forma precisa, uma *isometria* no \mathbb{R}^3 é uma função $f: \mathbb{R}^3 \rightarrow \mathbb{R}^3$ tal que, $\forall u, v \in \mathbb{R}^3$,

$$\|f(u) - f(v)\| = \|u - v\|.$$

Pode-se demonstrar [Lima \(2016\)](#) que f tem a forma

$$f(u) = Au + b, \tag{5.2}$$

onde $A \in \mathbb{R}^{3 \times 3}$, $b \in \mathbb{R}^3$ e A tem inversa dada por $A^{-1} = A^T$. Matrizes com essa propriedade são chamadas *ortogonais* [Lima \(2016\)](#), que podem também ser definidas, de maneira equivalente, por

$$\|Au - Av\| = \|u - v\|$$

ou

$$(Au) \cdot (Av) = u \cdot v.$$

De (5.2), vemos que uma isometria pode ser representada por uma matriz ortogonal, a menos de uma translação. Como o Modelo Homogêneo permite “incorporar” a translação em uma matriz, seria possível transladar ortogonalmente no

Espaço Homogêneo? Em termos matemáticos: podemos representar isometrias do \mathbb{R}^3 por meio de operações ortogonais no \mathbb{R}^4 ? Em símbolos, teríamos

$$X \cdot Y = k \|x - y\|^2, \quad (5.3)$$

onde $x, y \in \mathbb{R}^3$, k é uma constante real não nula e X, Y são os representantes de x, y no \mathbb{R}^4 , respectivamente.

De (5.3), temos que

$$x = y \Rightarrow X \cdot X = 0, \quad (5.4)$$

implicando na perda da positividade do produto interno em \mathbb{R}^4 .

Exercício 43. *Prove que não é possível satisfazer a propriedade (5.3) em \mathbb{R}^4 , mesmo abrindo mão da positividade do produto interno.*

Vamos “subir” mais uma dimensão, considerando o representante de $x \in \mathbb{R}^3$ por

$$X = x + x_4 e_4 + x_5 e_5.$$

Assumindo que e_5 é ortogonal a $\{e_1, e_2, e_3, e_4\}$ e mantendo a ortogonalidade de e_4 em relação a $\{e_1, e_2, e_3\}$, obtemos:

$$\begin{aligned} X \cdot X = 0 &\Rightarrow \\ (x + x_4 e_4 + x_5 e_5) \cdot (x + x_4 e_4 + x_5 e_5) = 0 &\Rightarrow \\ x \cdot x + x_4^2 (e_4 \cdot e_4) + x_5^2 (e_5 \cdot e_5) = 0 &\Rightarrow \\ x_4^2 (e_4 \cdot e_4) + x_5^2 (e_5 \cdot e_5) = -\|x\|^2. \end{aligned}$$

A última igualdade obriga, pelo menos, um dos produtos ($e_4 \cdot e_4$ ou $e_5 \cdot e_5$) ter valor negativo. Vamos assumir que

$$e_4 \cdot e_4 = 1 \quad (5.5)$$

e

$$e_5 \cdot e_5 = -1. \quad (5.6)$$

A expressão

$$X \cdot X = 0$$

exige que os representantes dos pontos em \mathbb{R}^3 , agora em \mathbb{R}^5 , sejam *vetores nulos* ($X \cdot X = 0$ e $X \neq 0$), fato possível com a perda da positividade do produto interno.

Para que os elementos adicionais que formam a base do \mathbb{R}^5 também sejam vetores nulos, podemos fazer a seguinte “mudança de base”:

$$e_0 = \frac{e_4 + e_5}{2}$$

e

$$e_\infty = e_5 - e_4.$$

Exercício 44. Verifique que, de fato, e_0 e e_∞ são vetores nulos, i.e.,

$$e_0 \cdot e_0 = 0$$

e

$$e_\infty \cdot e_\infty = 0.$$

Com a nova base $\{e_0, e_1, e_2, e_3, e_\infty\}$, devemos calcular $X \cdot Y$, considerando

$$X = x_0 e_0 + x + x_\infty e_\infty \quad (5.7)$$

e

$$Y = y_0 e_0 + y + y_\infty e_\infty, \quad (5.8)$$

onde $x = x_1 e_1 + x_2 e_2 + x_3 e_3$ e $y = y_1 e_1 + y_2 e_2 + y_3 e_3$ são os pontos do \mathbb{R}^3 que queremos representar no \mathbb{R}^5 .

Exercício 45. Mostre que

$$e_0 \cdot e_\infty = -1.$$

Usando (5.7) e (5.8), temos:

$$\begin{aligned} X \cdot Y &= (x_0 e_0 + x + x_\infty e_\infty) \cdot (y_0 e_0 + y + y_\infty e_\infty) \\ &= x \cdot y - (x_0 y_\infty + x_\infty y_0). \end{aligned}$$

Para $X = Y$,

$$X \cdot X = \|x\|^2 - 2x_0 x_\infty.$$

Ou seja,

$$X \cdot X = 0 \Rightarrow \|x\|^2 - 2x_0 x_\infty = 0.$$

Fazendo $x_0 = 1$:

$$x_\infty = \frac{1}{2} \|x\|^2 \Rightarrow X = e_0 + x + \frac{1}{2} \|x\|^2 e_\infty.$$

Portanto, um ponto $x \in \mathbb{R}^3$, representado como vetor nulo no \mathbb{R}^5 , é dado por

$$X = e_0 + x + \frac{1}{2} \|x\|^2 e_\infty.$$

Exercício 46. *Considerando*

$$X = e_0 + x + \frac{1}{2}\|x\|^2 e_\infty$$

e

$$Y = e_0 + y + \frac{1}{2}\|y\|^2 e_\infty,$$

prove que

$$X \cdot Y = -\frac{1}{2}\|x - y\|^2.$$

De forma resumida, o *Modelo Conforme* é o \mathbb{R}^5 munido desse novo produto interno, em que operações que preservam o produto interno entre pontos representados por

$$X = e_0 + x + \frac{1}{2}\|x\|^2 e_\infty, \quad (5.9)$$

preservam distâncias entre os pontos associados no \mathbb{R}^3 .

Pode-se dar uma interpretação para os vetores nulos e_0 e e_∞ [Kanatani \(2015\)](#):

- e_0 representa a origem do \mathbb{R}^3 ;
- e_∞ representa “o infinito” do \mathbb{R}^3 .

Finalmente, seja S um objeto do Espaço Conforme, dado por

$$S = C - \frac{r^2}{2} e_\infty,$$

onde

$$C = e_0 + c + \frac{1}{2}\|c\|^2 e_\infty,$$

$c \in \mathbb{R}^3$ e $r \in \mathbb{R}$.

Considerando os pontos X que pertencem a S , descritos como

$$X \cdot S = 0,$$

onde

$$X = e_0 + x + \frac{1}{2}\|x\|^2 e_\infty, \quad x \in \mathbb{R}^3,$$

e usando as propriedades do produto interno do Modelo Conforme,

$$X \cdot e_\infty = e_0 \cdot e_\infty = -1,$$

obtemos:

$$\begin{aligned} X \cdot S &= X \cdot \left(C - \frac{r^2}{2} e_\infty \right) \\ &= (X \cdot C) - \frac{r^2}{2} (X \cdot e_\infty) \\ &= -\frac{1}{2} \|x - c\|^2 + \frac{r^2}{2}. \end{aligned}$$

Ou seja,

$$X \cdot S = 0 \Leftrightarrow \|x - c\|^2 = r^2,$$

implicando que, no Modelo Conforme,

$$C - \frac{r^2}{2} e_\infty$$

representa uma esfera em \mathbb{R}^3 , onde C é o representante em \mathbb{R}^5 do centro da esfera e r seu raio.

Além dessa forma compacta de representação de esferas, que se estende naturalmente para o \mathbb{R}^n , o Modelo Conforme ganha um poder ainda maior quando se define uma álgebra de Clifford (ou *álgebra geométrica*) sobre ele, permitindo interseções de esferas, mesmo considerando distâncias intervalares. Para mais detalhes, ver [Lavor e Alves \(2019\)](#).

6

Conclusão

O cálculo da estrutura 3D de uma molécula de proteína, utilizando dados experimentais de RMN, é uma das aplicações mais importantes da Geometria de Distâncias (GD), cujo problema fundamental é encontrar um conjunto de pontos em um dado espaço geométrico, conhecendo distâncias entre alguns desses pontos.

A abordagem combinatória que apresentamos tem importantes consequências:

- Modelagem do problema como imersão de grafos no espaço 3D;
- Representação do espaço de busca como uma árvore binária;
- Desenvolvimento do algoritmo Branch & Prune (BP);
- Descoberta de simetrias que caracterizam todas as soluções matemáticas do problema, obtidas em tempo linear, a partir de uma solução dada por qualquer método.

Todos os resultados acima estão relacionados à definição de uma classe de problemas da GD, o *Discretizable Molecular Distance Geometry Problem* (DMDGP), que considera as distâncias provenientes da RMN como valores exatos, a fim de garantir a rigidez do grafo associado e iniciar o estudo do problema.

A primeira tentativa de levar em conta as distâncias intervalares (as incertezas da RMN) foi baseada na amostragem de valores dentro dos intervalos numéricos correspondentes. Essa ideia tornou o algoritmo BP uma heurística, já que não é mais possível garantir que uma solução sempre será encontrada. Além disso, amostras refinadas aumentam exponencialmente o espaço de busca (a árvore associada deixa de ser binária) e, mesmo assim, a solução pode ser “perdida”, caso a distância correta esteja entre os valores selecionados no processo de amostragem.

O passo fundamental para enfrentar essas dificuldades foi o “desenho” de novas ordenações para os vértices do DMDGP, considerando as informações obtidas pelos experimentos de RMN, bem como a própria geometria das proteínas.

Dada uma ordenação v_1, \dots, v_n nos vértices do DMDGP, o algoritmo BP calcula duas possibilidades para o vértice v_i (em função das posições $x_{i-3}, x_{i-2}, x_{i-1} \in \mathbb{R}^3$, previamente calculadas para os vértices $v_{i-3}, v_{i-2}, v_{i-1}$) e verifica se há distâncias adicionais $d_{j,i}$, para $j < i-3$, dadas pelos experimentos de RMN. Quando isso acontece, temos a interseção de quatro ou mais esferas (centradas em $x_j, x_{i-3}, x_{i-2}, x_{i-1}$ e com raios $d_{j,i}, d_{i-3,i}, d_{i-2,i}, d_{i-1,i}$), resultando em dois casos possíveis:

- um único ponto é obtido, indicando que a busca deve continuar, considerando o vértice v_{i+1} ;
- a interseção é vazia, obrigando “retornos” na árvore do BP.

Portanto, avançando ou retornando na árvore de busca do BP, a principal operação realizada é a interseção de esferas. Teremos sempre três esferas (pelas hipóteses do DMDGP) e, em alguns momentos, quatro esferas ou mais (quando distâncias adicionais são geradas pelos experimentos de RMN).

Classicamente, a interseção de esferas pode ser calculada resolvendo um sistema quadrático, mas por conta da geometria das proteínas, esse procedimento pode ser substituído por multiplicação de matrizes, usando coordenadas homogêneas. Com distâncias exatas, a segunda opção tem se mostrado mais eficiente. Entretanto, como algumas distâncias $d_{i-3,i}$, além das distâncias $d_{j,i}$ ($j < i-3$), são obtidas pelos experimentos de RMN, seus valores devem ser representados por distâncias intervalares, no lugar de valores precisos. Técnicas de amostragem sobre tais intervalos (incluindo aritmética intervalar e suas generalizações) foram propostas, mas os resultados não são animadores.

Usando coordenadas homogêneas, o espaço 3D é representado com 4 dimensões, “ganhando” a linearidade das translações. Uma das maneiras de justificar o aumento de mais uma dimensão (e chegar no Espaço Conforme) é que, com

5 dimensões, translações no espaço 3D podem ser representadas ortogonalmente, abrindo mão da positividade do produto interno usual. De maneira simplificada, podemos interpretar o Modelo Conforme como o espaço 5D com produto interno que não respeita mais a positividade.

Uma propriedade fundamental do Modelo Conforme (de grande importância para o nosso problema) é que a esfera torna-se um objeto tão primitivo quanto pontos, retas e planos. Mesmo considerando distâncias imprecisas, por conta dos dados de RMN, o que teremos, no caso geral, é a interseção entre esferas (com raios precisos dados pelas ligações e ângulos “covalentes”) e cascas esféricas (com raios definidos pelas distâncias intervalares da RMN).

O Modelo Conforme inclui a esfera na lista de objetos primitivos, mas para a manipulação desses objetos é necessário uma linguagem mais poderosa do que a Álgebra Linear: a Álgebra Geométrica. Nesse novo ambiente, operadores e objetos são representados como elementos de mesma natureza, os *multivetores*, abrindo espaço para o desenvolvimento de uma nova linguagem com poder, tanto simbólico quanto computacional, superior ao da Álgebra Linear. Em Computação Gráfica e Robótica, sem falar na Física, esse processo já está em pleno desenvolvimento, e o que estamos propondo é que a área de Geometria Molecular ganhe novo impulso por conta das novas possibilidades teóricas e computacionais oferecidas pela Álgebra Geométrica.

Uma outra vantagem que vislumbramos é a possibilidade de integrar, em uma única linguagem, as características combinatórias do problema (mantendo a estrutura de árvore de busca e simetrias associadas) com os aspectos “contínuos” dos dados experimentais.

Concluindo: considerar o cálculo da estrutura 3D de proteínas como um problema em Geometria de Distâncias, descrito na linguagem da Álgebra Geométrica Conforme, parece ser o caminho possível para considerar distâncias intervalares provenientes da RMN, sem a necessidade do processo de amostragem, e ainda manter as propriedades da abordagem combinatória.

Bibliografia

- A. Agra, R. Figueiredo, C. Lavor, N. Maculan, A. Pereira e C. Requejo (2017). “Feasibility check for the distance geometry problem: an application to molecular conformations”. *Int. Trans. Oper. Res.* 24.5, pp. 1023–1040. MR: [3663137](#). Zbl: [1371.90086](#) (ver p. [42](#)).
- J. Alencar, C. Lavor e L. Liberti (2019). “Realizing Euclidean distance matrices by sphere intersection”. *Discrete Appl. Math.* 256, pp. 5–10. MR: [3926356](#). Zbl: [07027409](#) (ver p. [5](#)).
- R. Alves, C. Lavor, C. Souza e M. Souza (2018). “Clifford algebra and discretizable distance geometry”. *Math. Methods Appl. Sci.* 41.11, pp. 4063–4073. MR: [3824543](#). Zbl: [1397.92523](#) (ver p. [40](#)).
- R. Alves e C. Lavor (2017). “Geometric algebra to model uncertainties in the discretizable molecular distance geometry problem”. *Adv. Appl. Clifford Algebr.* 27.1, pp. 439–452. MR: [3619377](#). Zbl: [1373.92148](#) (ver p. [40](#)).
- B. Anderson, P. Belhumeur, T. Eren, D. Goldenberg, S. Morse, W. Whiteley e R. Yang (2009). “Graphical properties of easily localizable sensor networks”. *Wireless Networks* 15, pp. 177–191 (ver pp. [5](#), [19](#)).
- C. Bajaj (1988). “The algebraic degree of geometric optimization problems”. *Discrete Comput. Geom.* 3.2, pp. 177–191. MR: [920702](#). Zbl: [0647.90087](#) (ver p. [2](#)).
- R. Benedetti e J.-J. Risler (1990). *Real algebraic and semi-algebraic sets*. Actualités Mathématiques. Hermann, Paris, p. 340. MR: [1070358](#). Zbl: [0694.14006](#) (ver p. [17](#)).

- S. J. L. Billinge, P. M. Duxbury, D. S. Gonçalves, C. Lavor e A. Mucherino (2016). “Assigned and unassigned distance geometry: applications to biological molecules and nanostructures”. *4OR* 14.4, pp. 337–376. MR: 3568084. Zbl: 1364. 51005 (ver p. 5).
- (2018). “Recent results on assigned and unassigned distance geometry with applications to protein molecules and nanostructures”. *Ann. Oper. Res.* 271.1, pp. 161–203. MR: 3871082. Zbl: 07032593 (ver p. 1).
- L. M. Blumenthal (1953). *Theory and applications of distance geometry*. Oxford, at the Clarendon Press, pp. xi+347. MR: 0054981. Zbl: 0050 . 38502 (ver pp. 3, 18).
- H. L. Bodlaender, F. V. Fomin, A. M. C. A. Koster, D. Kratsch e D. M. Thilikos (2012). “A note on exact algorithms for vertex ordering problems on graphs”. *Theory Comput. Syst.* 50.3, pp. 420–432. MR: 2885638. Zbl: 1253 . 68164 (ver p. 19).
- A. C. Brown (1864). “On the theory of isomeric compounds”. *Transactions of the Royal Society of Edinburgh* 23, pp. 707–719 (ver p. 16).
- A. Cassioli, B. Bordiaux, G. Bouvier, A. Mucherino, R. Alves, L. Liberti, M. Nilges, C. Lavor e T. E. Malliavin (2015). “An algorithm to enumerate all possible protein conformations verifying a set of distance constraints”. *BMC Bioinformatics* 16, pp. 16–23 (ver pp. 2, 40).
- A. Cassioli, O. Günlük, C. Lavor e L. Liberti (2015). “Discretization vertex orders in distance geometry”. *Discrete Appl. Math.* 197, pp. 27–41. MR: 3398958. Zbl: 1321. 05029 (ver pp. 34, 42).
- R. Connelly (2005). “Generic global rigidity”. *Discrete Comput. Geom.* 33.4, pp. 549–563. MR: 2132290. Zbl: 1072 . 52016 (ver p. 19).
- T. M. Costa, H. Bouwmeester, W. A. Lodwick e C. Lavor (2017). “Calculating the possible conformations arising from uncertainty in the molecular distance geometry problem using constraint interval analysis”. *Inform. Sci.* 415/416, pp. 41–52. MR: 3681200 (ver p. 2).
- V. Costa, A. Mucherino, C. Lavor, A. Cassioli, L. M. Carvalho e N. Maculan (2014). “Discretization orders for protein side chains”. *J. Global Optim.* 60.2, pp. 333–349. MR: 3257687. Zbl: 1312. 90069 (ver p. 11).
- G. M. Crippen e T. F. Havel (1988). *Distance geometry and molecular conformation*. Vol. 15. Chemometrics Series. Research Studies Press, Ltd., Chichester; John Wiley & Sons, Inc., New York, pp. x+541. MR: 975025. Zbl: 1066 . 51500 (ver p. 5).
- C. D’Ambrosio, K. Vu, C. Lavor, L. Liberti e N. Maculan (2017). “New error measures and methods for realizing protein graphs from distance data”. *Dis-*

- crete Comput. Geom.* 57.2, pp. 371–418. MR: 3602859. Zbl: 1358.05085 (ver p. 2).
- M. M. Deza e E. Deza (2009). *Encyclopedia of distances*. Com um CD-ROM para Windows, Macintosh e UNIX. Springer-Verlag, Berlin, pp. xiv+590. MR: 2538177. Zbl: 1167.51001 (ver p. 15).
- B. R. Donald (2011). *Algorithms in Structural Molecular Biology*. Boston: MIT Press (ver pp. 1, 42).
- L. Fernandes, C. Lavor e M. Oliveira Neto (2017). *Álgebra Geométrica e Aplicações*. Vol. 85. Notas em Matemática Aplicada (ver pp. 13, 47).
- F. Fidalgo, D. S. Gonçalves, C. Lavor, L. Liberti e A. Mucherino (2018). “A symmetry-based splitting strategy for discretizable distance geometry problems”. *J. Global Optim.* 71.4, pp. 717–733. MR: 3827287. Zbl: 1405.90134 (ver p. 26).
- C. A. Floudas e C. E. Gounaris (2009). “A review of recent advances in global optimization”. *J. Global Optim.* 45.1, pp. 3–38. MR: 2533740. Zbl: 1180.90245 (ver p. 2).
- M. Fréchet (1935). “Sur La Definition Axiomatique D’Une Classe D’Espaces Vectoriels Distancias Applicables Vectoriellement Sur L’Espace de Hilbert”. *Ann. of Math.* 36.3. Ver uma nota adicional no artigo Schoenberg (1935), pp. 705–718. MR: 1503246.
- K. Gibson e H. A. Scheraga (1997). “Energy minimization of rigid-geometry polypeptides with exactly closed disulfide loops”. *Journal of Computational Chemistry* 18, pp. 403–415 (ver p. 11).
- H. Gluck (1975). “Almost all simply connected closed surfaces are rigid”, 225–239. *Lecture Notes in Math.*, Vol. 438. MR: 0400239 (ver p. 18).
- D. S. Gonçalves e A. Mucherino (2014). “Discretization orders and efficient computation of Cartesian coordinates for distance geometry”. *Optim. Lett.* 8.7, pp. 2111–2125. MR: 3263243. Zbl: 1302.90177 (ver p. 20).
- D. S. Gonçalves, A. Mucherino, C. Lavor e L. Liberti (2017). “Recent advances on the interval distance geometry problem”. *J. Global Optim.* 69.3, pp. 525–545. MR: 3714551. Zbl: 1382.90084 (ver pp. 2, 40).
- J. Graver, B. Servatius e H. Servatius (1993). *Combinatorial rigidity*. Vol. 2. Graduate Studies in Mathematics. American Mathematical Society, Providence, RI, pp. x+172. MR: 1251062. Zbl: 0788.05001 (ver p. 17).
- P. Güntert (1998). “Structure calculation of biological macromolecules from NMR data”. *Quarterly Reviews of Biophysics* 31, pp. 145–237 (ver p. 1).

- F. Harary (1969). *Graph theory*. Addison-Wesley Publishing Co., Reading, Mass.-Menlo Park, Calif.-London, pp. ix+274. MR: 0256911. Zbl: 0182.57702 (ver p. 15).
- T. F. Havel e K. Wüthrich (1984). “A distance geometry program for determining the structures of small proteins and other macromolecules from nuclear magnetic resonance measurements of ^1H - ^1H proximities in solution”. *Bulletin of Mathematical Biology* 46, pp. 673–698 (ver p. 11).
- B. Hendrickson (1992). “Conditions for unique graph realizations”. *SIAM J. Comput.* 21.1, pp. 65–84. MR: 1148818. Zbl: 0756.05047 (ver pp. 18, 19).
- L. Henneberg (1886). *Statik der starren Systeme*. Darmstadt: Verlag von Arnold Bergsträsser, p. 374 (ver p. 19).
- B. Jackson e T. Jordán (2005). “Connected rigidity matroids and unique realizations of graphs”. *J. Combin. Theory Ser. B* 94.1, pp. 1–29. MR: 2130278. Zbl: 1076.05021 (ver p. 17).
- (2008). “On the rigidity of molecular graphs”. *Combinatorica* 28.6, pp. 645–658. MR: 2488744. Zbl: 1199.52036 (ver p. 19).
- K. Kanatani (2015). *Understanding geometric algebra. Hamilton, Grassmann, and Clifford for computer vision and graphics*. CRC Press, Boca Raton, FL, pp. xv+192. MR: 3643031. Zbl: 1319.68004 (ver pp. 47, 51).
- G. Laman (1970). “On graphs and rigidity of plane skeletal structures”. *J. Engrg. Math.* 4, pp. 331–340. MR: 0269535. Zbl: 0213.51903 (ver p. 19).
- C. Lavor e L. Liberti (2014). *Um Convite à Geometria de Distâncias*. Vol. 71. Notas em Matemática Aplicada (ver p. viii).
- C. Lavor, A. Mucherino, L. Liberti e N. Maculan (2011). “On the computation of protein backbones by using artificial backbones of hydrogens”. *J. Global Optim.* 50.2, pp. 329–344. MR: 2799569. Zbl: 1219.90209 (ver p. 37).
- C. Lavor, M. Souza, L. Mariano e L. Liberti (s.d.). “On the polynomiality of finding K DMDGP re-orders”. aceito para publicação (ver p. 43).
- C. Lavor (2006). “On generating instances for the molecular distance geometry problem”. Em: *Global optimization*. Vol. 84. Nonconvex Optim. Appl. Springer, New York, pp. 405–414. MR: 2206961. Zbl: 1100.90034 (ver pp. 12, 14).
- C. Lavor e R. Alves (2019). “Oriented conformal geometric algebra and the molecular distance geometry problem”. *Adv. Appl. Clifford Algebr.* 29.1, Art. 9, 9. MR: 3882392. Zbl: 07027677 (ver pp. 40, 52).
- C. Lavor, R. Alves, W. Figueiredo, A. Petraglia e N. Maculan (2015). “Clifford algebra and the discretizable molecular distance geometry problem”. *Adv. Appl.*

- Clifford Algebr.* 25.4, pp. 925–942. MR: 3413612. Zbl: 1327 . 15052 (ver p. 32).
- C. Lavor, J. Lee, A. Lee-St. John, L. Liberti, A. Mucherino e M. Sviridenko (2012). “Discretization orders for distance geometry problems”. *Optim. Lett.* 6.4, pp. 783–796. MR: 2902263. Zbl: 1258 . 90096 (ver p. 22).
- C. Lavor, L. Liberti, B. R. Donald, B. Worley, B. Bardiaux, T. E. Malliavin e M. Nilges (2019). “Minimal NMR distance information for rigidity of protein graphs”. *Discrete Appl. Math.* 256, pp. 91–104. MR: 3926362. Zbl: 1405 . 05178 (ver pp. 41–43).
- C. Lavor, L. Liberti, W. A. Lodwick e T. Mendonça da Costa (2017). *An introduction to distance geometry applied to molecular geometry*. SpringerBriefs in Computer Science. Springer, Cham, pp. ix+54. MR: 3676181. Zbl: 1403 . 92001 (ver p. 5).
- C. Lavor, L. Liberti, N. Maculan e A. Mucherino (2012a). “Recent advances on the discretizable molecular distance geometry problem”. *European J. Oper. Res.* 219.3, pp. 698–706. MR: 2898950. Zbl: 1253 . 05132 (ver p. 17).
- (2012b). “The discretizable molecular distance geometry problem”. *Comput. Optim. Appl.* 52.1, pp. 115–146. MR: 2925767. Zbl: 1259 . 90153 (ver pp. 20, 22, 34).
- C. Lavor, L. Liberti e A. Mucherino (2013). “The interval Branch-and-Prune algorithm for the discretizable molecular distance geometry problem with inexact distances”. *J. Global Optim.* 56.3, pp. 855–871. MR: 3070134. Zbl: 1272 . 90074 (ver pp. 29, 37, 40, 45).
- C. Lavor, N. Maculan, M. Souza e R. Alves (2017). *Álgebra e geometria no cálculo de estrutura molecular*. Publicações Matemáticas do IMPA. 31º Colóquio Brasileiro de Matemática. Instituto Nacional de Matemática Pura e Aplicada (IMPA), Rio de Janeiro, pp. v+110. MR: 3753697. Zbl: 06921573 (ver p. viii).
- L. Liberti e C. Lavor (2016). “Six mathematical gems from the history of distance geometry”. *Int. Trans. Oper. Res.* 23.5, pp. 897–920. MR: 3508435. Zbl: 1362 . 51002 (ver p. 3).
- (2017). *Euclidean distance geometry. An introduction*. Springer Undergraduate Texts in Mathematics and Technology. Springer, Cham, pp. xiii+133. MR: 3702008. Zbl: 06738024 (ver p. 5).
- L. Liberti, C. Lavor, J. Alencar e G. Abud (2013). “Counting the number of solutions of K DMDGP instances”. Em: *Geometric science of information*. Vol. 8085. Lecture Notes in Comput. Sci. Springer, Heidelberg, pp. 224–230. MR: 3126049. Zbl: 1405 . 05083 (ver p. 36).

- L. Liberti, C. Lavor e N. Maculan (2008). “A branch-and-prune algorithm for the molecular distance geometry problem”. *Int. Trans. Oper. Res.* 15.1, pp. 1–17. MR: 2387010. Zbl: 1136.92037 (ver pp. 20, 22, 34).
- L. Liberti, C. Lavor, N. Maculan e A. Mucherino (2014). “Euclidean distance geometry and applications”. *SIAM Rev.* 56.1, pp. 3–69. MR: 3246296. Zbl: 1292.51010 (ver pp. 2, 15, 19).
- L. Liberti, C. Lavor, A. Mucherino e N. Maculan (2011). “Molecular distance geometry methods: from continuous to discrete”. *Int. Trans. Oper. Res.* 18.1, pp. 33–51. MR: 2779655. Zbl: 1219.90177 (ver p. 5).
- L. Liberti, B. Masson, J. Lee, C. Lavor e A. Mucherino (2014). “On the number of realizations of certain Henneberg graphs arising in protein conformation”. *Discrete Appl. Math.* 165, pp. 213–232. MR: 3162289. Zbl: 1288.05121 (ver pp. 17, 34, 35).
- L. Liberti, G. Swirszcz e C. Lavor (2016). “Distance geometry on the sphere”. Em: *Discrete and computational geometry and graphs*. Vol. 9943. Lecture Notes in Comput. Sci. Springer, Cham, pp. 204–215. MR: 3595584. Zbl: 06679461 (ver p. 3).
- E. L. Lima (2016). *Algebra linear*. Coleção Matemática Universitária. IMPA, Rio de Janeiro (ver pp. 47, 48).
- D. S. Maioli, C. Lavor e D. S. Gonçalves (2017). “A note on computing the intersection of spheres in \mathbb{R}^n ”. *ANZIAM J.* 59.2, pp. 271–279. MR: 3731729. Zbl: 1386.51007 (ver p. 46).
- K. Menger (1928). “Untersuchungen über allgemeine Metrik”. *Math. Ann.* 100.1, pp. 75–163. MR: 1512479 (ver p. 3).
- N. J. M. Moreira, L. T. Duarte, C. Lavor e C. Torezzan (2018). “A novel low-rank matrix completion approach to estimate missing entries in Euclidean distance matrix”. *Comput. Appl. Math.* 37.4, pp. 4989–4999. MR: 3848577. Zbl: 1400.15027 (ver p. 5).
- A. Mucherino, C. Lavor e L. Liberti (2012a). “Exploiting symmetry properties of the discretizable molecular distance geometry problem”. *J. Bioinform. Comput. Biol.* 10.3, 1242009(1–15). MR: 2996476. Zbl: 1258.90100 (ver p. 36).
- (2012b). “The discretizable distance geometry problem”. *Optim. Lett.* 6.8, pp. 1671–1686. MR: 2996476. Zbl: 1258.90100 (ver pp. 20–22).
- A. Mucherino, C. Lavor, L. Liberti e N. Maculan, ed. (2013). *Distance geometry. Theory, methods, and applications*. Springer, New York, pp. xvi+420. MR: 3059590 (ver p. 5).
- C. Mueller, B. Martin e A. Lumsdaine (2007). “A comparison of vertex ordering algorithms for large graph visualization”. Em: *IEEE Proc. of the 6th Interna-*

- tional Asia-Pacific Symposium on Visualization, Sydney, NSW, Australia, 5-7 Feb. 2007*, pp. 141–148 (ver p. 19).
- L. S. Neto, C. Lavor e W. A. Lodwick (s.d.). “A constrained interval approach to the generalized distance geometry problem”. aceito para publicação (ver p. 2).
- J. Nielsen e B. Roth (1999). “On the kinematic analysis of robotic mechanisms”. *International Journal of Robotics Research* 18, pp. 1147–1160 (ver p. 5).
- P. Nucci, L. T. Nogueira e C. Lavor (2013). “Solving the discretizable molecular distance geometry problem by multiple realization trees”. Em: *Distance geometry*. Springer, New York, pp. 161–176. MR: 3051953. Zbl: 1366.92096 (ver p. 26).
- S. Sallaume, S. Martins, L. Ochi, W. Gramacho, C. Lavor e L. Liberti (2013). “A discrete search algorithm for finding the structure of protein backbones and side chains”. *International Journal of Bioinformatics Research and Applications* 9, pp. 261–270 (ver p. 11).
- C. P. Santiago, C. Lavor, S. A. Monteiro e A. Kroner-Martins (2018). “A new algorithm for the small-field astrometric point-pattern matching problem”. *J. Global Optim.* 72.1, pp. 55–70. MR: 3846926. Zbl: 06996782 (ver p. 5).
- R. dos Santos Carvalho, C. Lavor e F. Protti (2008). “Extending the geometric build-up algorithm for the molecular distance geometry problem”. *Inform. Process. Lett.* 108.4, pp. 234–237. MR: 2457932. Zbl: 1191.68772 (ver p. 20).
- J. B. Saxe (1979). “Embeddability of weighted graphs in k-space is strongly NP-hard”. Em: *Proc. of the 17th Allerton Conference in Communications, Control e Computing*, pp. 480–489 (ver p. 7).
- I. J. Schoenberg (1935). “Remarks to Maurice Fréchet’s article “Sur la définition axiomatique d’une classe d’espace distanciés vectoriellement applicable sur l’espace de Hilbert””. *Ann. of Math. (2)* 36.3. Uma nota ao artigo **frechet**, pp. 724–732. MR: 1503248 (ver pp. 3, 4, 58).
- M. J. Sippl e H. A. Scheraga (1986). “Cayley-Menger coordinates”. *Proc. Nat. Acad. Sci. U.S.A.* 83.8, pp. 2283–2287. MR: 839626. Zbl: 0587.51013 (ver p. 4).
- M. Sitharam e Y. Zhou (2004). “A tractable approximate combinatorial 3D rigidity characterization”. Em: *Proc. of the Fifth Workshop on Automated Deduction in Geometry* (ver p. 19).
- M. Souza, C. Lavor, A. Muritiba e N. Maculan (2013). “Solving the molecular distance geometry problem with inaccurate distance data”. *BMC Bioinformatics* 14, S71–S76 (ver p. 2).

- M. Souza, A. E. Xavier, C. Lavor e N. Maculan (2011). “Hyperbolic smoothing and penalty techniques applied to molecular structure determination”. *Oper. Res. Lett.* 39.6, pp. 461–465. MR: 2854456. Zbl: 1235.90121 (ver p. 2).
- J. Sylvester (1877). “Chemistry and algebra”. *Nature* 17, pp. 284–284 (ver p. 16).
- T.-S. Tay e W. Whiteley (1985). “Generating isostatic frameworks”. *Structural Topology* 11. Texto em duas línguas Francês e Inglês, pp. 21–69. MR: 804977. Zbl: 0574.51025 (ver p. 18).
- H. Thompson (1967). “Calculation of cartesian coordinates and their derivatives from internal molecular coordinates”. *Journal of Chemical Physics* 47, pp. 3407–3410 (ver p. 11).
- B. Worley, F. Delhommel, F. Cordier, T. E. Malliavin, B. Bardiaux, N. Wolff, M. Nilges, C. Lavor e L. Liberti (2018). “Tuning interval branch-and-prune for protein structure determination”. *J. Global Optim.* 72.1, pp. 109–127. MR: 3846929. Zbl: 06996785 (ver p. 2).
- K. Wüthrich (1986). *NMR of proteins and nucleic acids*. New York: Wiley (ver pp. 1, 42).
- Y. Yemini (1978). “The positioning problem — a draft of an intermediate summary”. Em: *Proc. of the Conference on Distributed Sensor Networks*, pp. 137–145 (ver p. 5).

Títulos Publicados — 32º Colóquio Brasileiro de Matemática

Emergence of Chaotic Dynamics from Singularities – *Pablo G. Barrientos, Santiago Ibáñez, Alexandre A. Rodrigues e J. Ángel Rodríguez*

Nonlinear Dispersive Equations on Star Graphs – *Jaime Angulo Pava e Márcio Cavalcante de Melo*

Scale Calculus and M-Polyfolds An Introduction – *Joa Weber*

Real and Complex Gaussian Multiplicative Chaos – *Hubert Lacoin*

Rigidez em Grafos de Proteínas – *Carlile Lavor*

Gauge Theory in Higher Dimensions – *Daniel G. Fadel e Henrique N. Sá Earp*

Elementos da Teoria de Aprendizagem de Máquina Supervisionada – *Vladimir Pestov*

Función Gamma: Propriedades Clásicas e Introducción a su Dinâmica – *Pablo Diaz e Rafael Labarca*

Introdução à Criptografia com Curvas Elípticas – *Sally Andria, Rodrigo Gondim e Rodrigo Salomão*

O Teorema dos Quatro Vértices e sua Recíproca – *Mário Jorge Dias Carneiro e Ronaldo Alves Garcia*

Uma Introdução Matemática a Blockchains – *Augusto Teixeira*



Instituto de
Matemática
Pura e Aplicada

ISBN 978-85-244-0483-2



9 788524 404832