

Volume 71, 2014

Editores

Fernando Rodrigo Rafaeli (Editor Chefe)

Universidade Federal de Uberlândia - UFU
Uberlândia, MG, Brasil

Vanessa Avansini Botta Pirani (Editor Adjunto)

Universidade Estadual Paulista - UNESP
Presidente Prudente, SP, Brasil

Alexandre Loureiro Madureira

Laboratório Nacional de Computação Científica - LNCC
Petrópolis, RJ, Brasil

Edson Luiz Cataldo Ferreira

Universidade Federal Fluminense - UFF
Niterói, RJ, Brasil

Jorge Manuel Vieira Capela

Universidade Estadual Paulista - UNESP
Araraquara, SP, Brasil

Sandra Augusta Santos

Universidade Estadual de Campinas - UNICAMP
Campinas, SP, Brasil

A Sociedade Brasileira de Matemática Aplicada e Computacional - SBMAC publica, desde as primeiras edições do evento, monografias dos cursos que são ministrados nos CNMAC.

Para a comemoração dos 25 anos da SBMAC, que ocorreu durante o XXVI CNMAC em 2003, foi criada a série **Notas em Matemática Aplicada** para publicar as monografias dos minicursos ministrados nos CNMAC, o que permaneceu até o XXXIII CNMAC em 2010.

A partir de 2011, a série passa a publicar, também, livros nas áreas de interesse da SBMAC. Os autores que submeterem textos à série Notas em Matemática Aplicada devem estar cientes de que poderão ser convidados a ministrarem minicursos nos eventos patrocinados pela SBMAC, em especial nos CNMAC, sobre assunto a que se refere o texto.

O livro deve ser preparado em **Latex (compatível com o Miktex versão 2.9)**, as **figuras em eps** e deve ter entre **80 e 150 páginas**. O texto deve ser redigido de forma clara, acompanhado de uma excelente revisão bibliográfica e de **exercícios de verificação de aprendizagem** ao final de cada capítulo.

Veja todos os títulos publicados nesta série na página
http://www.sbmac.org.br/p_notas.php

UM CONVITE À GEOMETRIA DE DISTÂNCIAS

Carlile Lavor
clavor@ime.unicamp.br

Departamento de Matemática Aplicada
Instituto de Matemática, Estatística e Computação Científica
Universidade Estadual de Campinas

Leo Liberti
leoliberti@gmail.com
École Polytechnique e IBM TJ Watson Research Center



Sociedade Brasileira de Matemática Aplicada e Computacional

São Carlos - SP, Brasil
2014

Coordenação Editorial: Maria do Socorro Nogueira Rangel

Coordenação Editorial da Série: Fernando Rodrigo Rafaeli

Editora: SBMAC

Capa: Matheus Botossi Trindade

Patrocínio: SBMAC

Copyright ©2014 by Carlile Lavor e Leo Liberti. Direitos reservados, 2014 pela SBMAC. A publicação nesta série não impede o autor de publicar parte ou a totalidade da obra por outra editora, em qualquer meio, desde que faça citação à edição original.

Catálogo elaborado pela Biblioteca do IBILCE/UNESP
Bibliotecária: Maria Luiza Fernandes Jardim Froner

Lavor, Carlile

Um Convite à Geometria de Distâncias - São Carlos, SP :
SBMAC, 2014, 57 p., 21.5 cm - (Notas em Matemática
Aplicada; v. 71)

e-ISBN 978-85-8215-057-3

1. Geometria de Distâncias 2. Geometria Molecular 3. Branch and Prune
I. Lavor, Carlile II. Liberti, Leo. IV. Título. V. Série

CDD - 51

Agradecimentos

Agradecemos a oportunidade dada pela Sociedade Brasileira de Matemática Aplicada e Computacional (SBMAC) e 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, agradecemos ao Prof. Rubens Sampaio, pelo incentivo, ao Prof. José Mario Martinez, ao Prof. Tibérius Bonates, ao Prof. Douglas Gonçalves e ao Prof. Nolmar Melo, pelas sugestões de melhoria do texto, e aos futuros colegas Germano Abud, Jorge Alencar e Felipe Fidalgo, pelas críticas e auxílio na preparação do texto.

Conteúdo

Prefácio	xi
1 Introdução	1
1.1 Um Pouco de História	1
1.2 Aplicações	2
1.3 Notação e Conceitos Básicos	3
2 O PGD	7
2.1 Definição do PGD	7
2.2 Encarando o PGD	8
2.3 Número de Soluções do PGD	8
2.4 Complexidade do PGD	9
2.5 Instâncias do PGD	11
3 Do Contínuo para o Discreto	15
3.1 Otimização Contínua e o PGD	15
3.2 Finitude do PGD	16
3.3 Ordens para o PGD	19
4 O PDGD₃	23
4.1 Definição do PDGD ₃	23
4.2 Encarando o PDGD ₃	23
4.3 Complexidade do PDGD ₃	27
4.4 O Algoritmo BP para o PDGD ₃	27
5 O PDGDM	31
5.1 As Hipóteses do PDGD ₃	31
5.2 Definição do PDGDM	32
5.3 Complexidade do PDGDM	35
5.4 Simetrias do PDGDM	36
6 GD e Geometria Molecular	41
6.1 O PDGDM e Estruturas de Proteínas	41
6.2 Ordens nas Proteínas	42
6.3 O BP com Custo Polinomial	45
7 Conclusão	49

Prefácio

Geometria de Distâncias é um tema da Matemática Aplicada que investiga as relações existentes entre três situações:

- Distâncias entre objetos relacionados a um determinado problema,
- Distâncias entre pontos (representando tais objetos) em um dado espaço geométrico,
- Localização desses pontos, possivelmente em um espaço geométrico distinto.

Procuramos apresentar o texto de maneira não tradicional, tentando motivar as definições e proposições da Geometria de Distâncias por meio de exemplos. Isso não quer dizer, de maneira alguma, que dispensamos a precisão e a abstração da Matemática. Apenas deixamos de lado o excesso de formalismo, principalmente por se tratar de um convite.

Os exercícios devem ser considerados parte essencial do texto. Todos eles foram criados cuidadosamente para ajudar na construção e aquisição dos conceitos fundamentais da Geometria de Distâncias (não são exercícios “desafios”, exceto o último deles). Selecionamos várias referências bibliográficas, onde aí, sim, um esforço maior e dedicação serão necessários.

Como pré-requisitos, apenas conhecimentos básicos de Geometria Analítica serão suficientes. O mais importante: manter a mente aberta, curiosa e com disposição para aprender.

Outra característica do texto é que ele não é uma história contada novamente, de maneira diferente. Trata-se de uma “estrela”, principalmente porque são tópicos desenvolvidos pelos autores e vários colaboradores, com destaque para os professores Nelson Maculan (da Universidade Federal do Rio de Janeiro) e Antonio Mucherino (da *Université de Rennes 1*). Por isso mesmo, serão inevitáveis os erros cometidos. Estendemos o convite para que nos enviem todos os comentários, sugestões e erros encontrados, a serem incorporados em futuras edições. Agradecemos, antecipadamente.

Para se ter uma ideia geral do texto, faremos uma breve descrição de cada capítulo.

No Capítulo 1, falamos um pouco da história da Geometria de Distâncias, suas aplicações e uma breve revisão dos principais conceitos utilizados. O Capítulo 2 apresenta o principal problema da Geometria de Distâncias, chamado de PGD, onde se baseia todo o texto. Sem uma boa compreensão desse capítulo, toda a leitura seguinte ficará prejudicada. Na verdade, o texto é bem sequencial e não há capítulos independentes. Ou seja, a ordem aqui é fundamental, conceito utilizado também no final do Capítulo 3, que apresenta a maneira clássica de abordar o PGD, usando métodos contínuos, e faz a passagem para o mundo discreto do

PGD. Os Capítulos 4 e 5 tratam de duas classes de problemas que aparecem justamente por conta da estrutura combinatória do PGD, fortemente relacionada com a ordem mencionada acima. No Capítulo 6, destacamos uma das aplicações da Geometria de Distâncias relacionada à Geometria Molecular. Algumas conclusões são apresentadas no Capítulo 7.

Durham e New York, 4 de julho de 2014.

Carlile Lavor
Leo Liberti

Capítulo 1

Introdução

1.1 Um Pouco de História

Geometria de Distâncias (GD) é uma área de pesquisa consolidada, tendo a Matemática e a Computação como áreas fundamentais em seu alicerce. O conceito de distância [14] é essencial para a experiência humana e a GD o coloca como objeto principal de estudo em uma dada estrutura geométrica.

Atualmente, **o problema fundamental da GD é determinar um conjunto de pontos em um dado espaço geométrico, cujas distâncias, entre alguns deles, são conhecidas.**

Considera-se que a GD surgiu em 1928, quando Menger [53] caracterizou vários conceitos geométricos usando a ideia de distância. Entretanto, apenas com os resultados de Blumenthal [8], em 1953, o tema se tornou, de fato, uma nova área do conhecimento, conhecida a partir de então por Geometria de Distâncias.

Naquela época, a principal questão da GD era encontrar condições necessárias e suficientes para decidir se uma dada matriz é uma matriz de distâncias D , 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 D . Note que, nesse caso, todas as distâncias são consideradas conhecidas.

A primeira menção explícita ao problema fundamental da GD (definido acima), onde não são conhecidas todas as distâncias, foi dada por Yemini [70], em 1978. Essa mudança faz toda a diferença, tanto do ponto de vista prático quanto teórico. Em várias aplicações, não se conhecem todas as distâncias, e, neste caso, o problema perde a garantia de poder ser resolvido facilmente [17].

Outro momento importante, principalmente relacionado à aplicação da GD ao cálculo de estruturas moleculares, é o lançamento do livro de Crippen e Havel [13], em 1988, considerados os pioneiros na aplicação da GD ao cálculo de estrutura de proteínas.

Em 2013, publicado pela Springer, aparece o primeiro livro editado, integralmente dedicado à GD [58], reunindo diferentes aplicações e pesquisadores em GD. No mesmo ano, em junho de 2013, é realizado o primeiro workshop internacional sobre o tema [4], contando com palestrantes de renomadas instituições internacionais (Princeton University, IBM TJ Watson Research Center, University of Cambridge, École Polytechnique, Institut Pasteur, École Normale Supérieure, SUTD-MIT International DesignCenter). O evento também contou com o apoio de várias sociedades científicas brasileiras e internacionais (Sociedade Brasileira de Matemática,

Sociedade Brasileira de Matemática Aplicada e Computacional, Sociedade Brasileira de Computação, Sociedade Brasileira de Pesquisa Operacional, La Asociación Latino-Iberoamericana de Investigación Operativa, The International Federation of Operational Research Societies, Associação Brasileira de Estatística e Sociedade Astronômica Brasileira) e das principais agências de fomento à pesquisa do Brasil (CNPq, CAPES e FAPESP), confirmando a integração que a GD permite entre várias áreas do conhecimento.

Acreditamos que esse seja o primeiro texto didático sobre GD. Diferentemente da abordagem clássica, usando métodos contínuos [48], iremos apresentar o tema baseados em sua estrutura combinatória, pois além de facilitar a compreensão dos principais conceitos envolvidos, é também uma maneira nova de considerar o problema. Curiosamente, essa nova visão se originou em um contexto de computação quântica [61], quando propusemos aplicar o algoritmo de Grover [27] em um problema de GD relacionado ao cálculo de estruturas moleculares [36]. Esse trabalho foi publicado nos anais de um congresso realizado também em Natal, desenvolvido durante um encontro científico dos autores, no *Politecnico di Milano*, em junho de 2005.

1.2 Aplicações

Além da teoria matemática associada à GD, o interesse por esse tópico de pesquisa explica-se pela riqueza e variedade de suas aplicações. Podemos citar, como exemplo, aplicações em astronomia, bioquímica, estatística, nanotecnologia, robótica e telecomunicações. Em [46], juntamente com os colegas Nelson Maculan e Antonio Mucherino, apresentamos um *survey* sobre GD, destacando a teoria e aplicações do tema.

Em astronomia, o problema está relacionado à determinação da posição de estrelas, utilizando informação de distância entre elas [51].

Em bioquímica, o problema aparece na determinação de estruturas tridimensionais de moléculas de proteínas, utilizando dados de ressonância magnética nuclear [16]. Foi através desse problema, que tivemos o primeiro contato com GD. Boa parte da pesquisa que desenvolvemos nessa área foi motivada pelas dificuldades encontradas ao lidar com esse problema.

Em estatística, o problema está relacionado à visualização de dados [18] e redução dimensional [42]. Nesse caso, todas as distâncias são conhecidas entre os pontos, que estão em “alta dimensão” (por exemplo, no \mathbb{R}^n), e o problema é como representá-los em \mathbb{R}^2 ou \mathbb{R}^3 , justamente para se ter uma ideia visual dos dados. Essa aplicação também está ligada a um tema atual de pesquisa, denominado *Big Data* [52].

Em nanotecnologia, o problema possui semelhanças com o problema em bioquímica, mas só que em escala “nano” [33].

Em robótica, o problema aparece no posicionamento dos braços de um robô para realizar uma tarefa específica [59]. Existe uma estreita relação entre esse problema e os cálculos relacionados à geometria molecular [19].

Em telecomunicações, o problema está relacionado à localização de redes de sensores sem fio, onde alguns pontos são previamente fixados, por exemplo, os roteadores [46, 70].

Pela própria natureza teórica e imensa variedade de aplicações, a GD torna-se um tópico exemplar em Matemática Aplicada, onde vários conceitos fundamentais da Matemática (métrica, dimensão, simetria, incerteza) e da Computação (algoritmos,

solubilidade de problemas, complexidade de algoritmos) podem ser apresentados de uma maneira integrada, facilitando a construção de uma visão ampla e profunda da Matemática, sem preconceitos e barreiras que insistem em eleger “uma matemática superior” a outras.

1.3 Notação e Conceitos Básicos

Como mencionamos no Prefácio, apenas conhecimentos elementares de Geometria Analítica serão suficientes para acompanhar todo o texto. Obviamente, iremos assumir também que o leitor tem alguma familiaridade com a linguagem matemática envolvendo conceitos básicos de lógica, teoria de conjuntos e funções, necessária em um curso de Geometria Analítica. A seguir, listamos os principais conceitos que serão utilizados, com as devidas notações.

- **Conjuntos:**

- $x \in A$ significa que x é um elemento do conjunto A .
- $A \subset B$ significa que o conjunto A está contido no conjunto B , ou seja, todo elemento de A é um elemento de B .
- $A \cap B = \{x : x \in A \text{ e } x \in B\}$ é o conjunto formado pelos elementos que estão em A e B .
- $A \cup B = \{x : x \in A \text{ ou } x \in B\}$ é o conjunto formado pelos elementos que estão em A ou B , incluindo os elementos que estão em $A \cap B$.
- $A - B = \{x : x \in A \text{ e } x \notin B\}$ é o conjunto A , excluindo os elementos de B que estão em A .
- O conjunto A é infinito enumerável se existe uma bijeção $f : \mathbb{N} \rightarrow A$, isto é, uma correspondência um-a-um entre os elementos de A e os números naturais $\mathbb{N} = \{1, 2, 3, \dots\}$.
- O conjunto A é finito se existe uma bijeção $f : \{1, 2, 3, \dots, n\} \rightarrow A$, onde $n \in \mathbb{N}$. A cardinalidade de A , denotada por $|A|$, é a quantidade de elementos de A .
- O conjunto A é infinito não enumerável se existe uma injeção $f : \mathbb{R} \rightarrow A$, onde \mathbb{R} é o conjunto dos números reais.

• **Vetores e Matrizes:**

- Um vetor $x \in \mathbb{R}^n$ será representado por uma matriz coluna. Por exemplo, $x \in \mathbb{R}^2$ será escrito como $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, onde $x_1, x_2 \in \mathbb{R}$.
- $\mathbb{R}^{m \times n}$ é o conjunto das matrizes com entradas reais, m linhas e n colunas.
- Dada $M \in \mathbb{R}^{m \times n}$, a matriz transposta $M^T \in \mathbb{R}^{n \times m}$ é a matriz obtida de M , trocando as linhas pelas colunas.
- Dada $M \in \mathbb{R}^{n \times n}$, existe a inversa de M , denotada por M^{-1} , se $MM^{-1} = M^{-1}M = I$, onde I é a matriz identidade. Por exemplo, se $I \in \mathbb{R}^{2 \times 2}$, então $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Se existe M^{-1} , dizemos que M é inversível (lembre que o produto de matrizes não é comutativo, isto é, existem matrizes A, B tais que $AB \neq BA$).
- Dada $M \in \mathbb{R}^{m \times n}$, com $m < n$, dizemos que M tem posto completo se existe uma submatriz $M' \in \mathbb{R}^{m \times m}$, ou seja, uma matriz formada com m linhas e m colunas de M , onde M' é inversível.
- Dados $x, y \in \mathbb{R}^n$, o produto interno entre x e y , denotado por $x \cdot y$, é definido por $x \cdot y = x_1y_1 + \dots + x_ny_n$.
- Dado $x \in \mathbb{R}^n$, a norma euclidiana de x , denotada por $\|x\|$, é definida por $\|x\| = \sqrt{x \cdot x} = \sqrt{x_1^2 + \dots + x_n^2}$.

- **Grafos** (esse tema, na verdade, é uma área na fronteira entre a Matemática e a Computação, com “vida própria”, muito rica teoricamente e com uma infinidade de aplicações [28]. Apresentaremos aqui, apenas algumas definições básicas):

- Dado um conjunto finito V e um outro conjunto E formado por pares (não ordenados) de elementos de V , temos um grafo $G = (V, E)$, onde V é o conjunto de vértices do grafo e E é o conjunto das arestas do grafo. Podemos representar um grafo no plano, usando pontos para os vértices e segmentos de retas (ou curvas) para as arestas, como ilustrado abaixo.
- Um grafo $G = (V, E)$ é dito simples se $\{x, y\} \in E \Rightarrow x \neq y$ (ver Figura 1.1).



(a) Grafo simples

(b) Grafo não simples

Figura 1.1: Exemplos de grafos simples e não simples.

- Um grafo $G = (V, E)$ é dito conectado se 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 $\{a, b\} \in E$, onde $a \in A$ e $b \in B$ (ver Figura 1.2).
- Um grafo $G = (V, E)$ é dito completo se E contém todos os pares possíveis, ou seja, $a, b \in V \Rightarrow \{a, b\} \in E$ (ver Figura 1.3).
- Uma clique em um grafo $G = (V, E)$ é um outro grafo $G' = (V', E')$, onde $V' \subset V$, $E' \subset E$ e G' é completo (ver Figura 1.4).

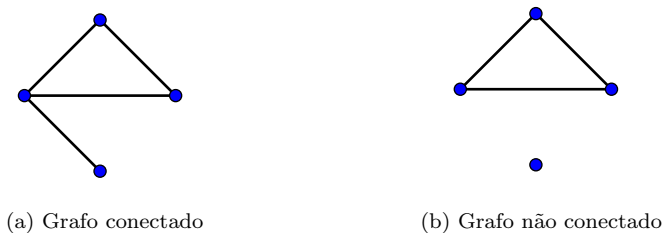


Figura 1.2: Exemplos de grafos conectado e não conectado.

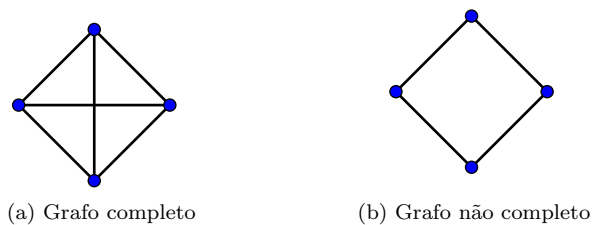


Figura 1.3: Exemplos de grafos completo e não completo.

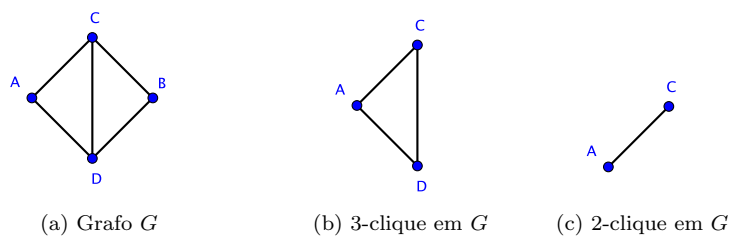


Figura 1.4: Exemplos de cliques em um grafo G .

- Quando associamos a cada aresta do grafo um número real, ou seja, definimos uma função $d : E \rightarrow \mathbb{R}$, temos um grafo $G = (V, E, d)$ com pesos nas arestas (ver Figura 1.5).

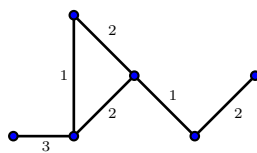


Figura 1.5: Grafo com pesos nas arestas.

Capítulo 2

O Problema de Geometria de Distâncias (PGD)

2.1 Definição do PGD

Como já comentamos, o problema fundamental da GD, que denotaremos por PGD, é determinar um conjunto de pontos, em um dado espaço geométrico, cujas distâncias entre alguns deles são conhecidas. Dependendo da aplicação, esses pontos podem representar estrelas, pontos de articulação de um robô, átomos ou pessoas. Aqui, já dá para sentir o poder da Matemática, pois podemos representar todos esses objetos usando um único conceito matemático: grafo. Na verdade, cada um desses objetos será representado por um vértice de um grafo e, se a distância entre eles é conhecida, teremos uma aresta entre os vértices correspondentes.

Formalmente, temos a seguinte **Definição do PGD**:

Definição 2.1. *Dado um inteiro $K > 0$ e um grafo simples $G = (V, E, d)$, conectado e com pesos nas arestas $d : E \rightarrow (0, \infty)$, encontre uma função $x : V \rightarrow \mathbb{R}^K$ tal que*

$$\forall \{u, v\} \in E, \|x(u) - x(v)\| = d(u, v). \quad (2.1.1)$$

Resolver o problema é, então, associar cada vértice de G a um único ponto em \mathbb{R}^K , satisfazendo as equações (2.1.1). Ou seja, ao posicionar os vértices $u, v \in V$, tais que $\{u, v\} \in E$, em \mathbb{R}^K , temos que “acertar” a distância calculada $\|x(u) - x(v)\|$ com o valor dado $d(u, v)$.

A função x é chamada de *realização de G* . Uma realização de um grafo é uma “representação” de seus vértices em algum espaço euclidiano \mathbb{R}^K (note que a dimensão do espaço faz parte da entrada do problema, assim como o grafo G).

Iremos considerar, na maioria dos casos, $K = 2$ ou $K = 3$, mas todos os resultados podem ser estendidos para \mathbb{R}^K [46]. Uma realização que satisfaz todas as equações (2.1.1) é dita uma *realização válida*.

Para simplificar a notação, usaremos x_u, x_v , no lugar de $x(u), x(v)$, e d_{uv} , no lugar de $d(u, v)$. Neste texto, utilizaremos a norma euclidiana (já definida na Seção 1.3), mas dependendo da aplicação, outras normas podem ser empregadas [46].

Com a Definição 2.1, além da representação abstrata de qualquer tipo de objeto (por meio dos vértices de um grafo), podemos também utilizar diferentes métricas e, até mesmo, alterar o espaço de realização do grafo. Em astronomia, por exemplo, pode ser mais adequado considerar espaços não euclidianos [51].

Exercício 2.1. *Pode haver mais de uma solução de um PGD? O conjunto solução pode ser vazio?*

Exercício 2.2. *Quando “desenhamos” um grafo no papel, estamos resolvendo um PGD?*

2.2 Encarando o PGD

Considerando a norma euclidiana na definição do PGD e fixando $K = 2$, obtemos o seguinte sistema de equações:

$$\sqrt{(x_{u1} - x_{v1})^2 + (x_{u2} - x_{v2})^2} = d_{uv}, \quad \forall \{u, v\} \in E, \quad (2.2.2)$$

onde $x_u^T = (x_{u1}, x_{u2})$ e $x_v^T = (x_{v1}, x_{v2})$. Temos, então, um sistema com $2|V|$ variáveis e $|E|$ equações.

Podemos reescrever o sistema (2.2.2) da seguinte forma:

$$(x_{u1} - x_{v1})^2 + (x_{u2} - x_{v2})^2 = d_{uv}^2, \quad \forall \{u, v\} \in E, \quad (2.2.3)$$

obtendo um sistema quadrático.

Encarar o PGD “de frente”, ou seja, tentar resolver o sistema de equações (2.1.1), ou o sistema quadrático associado, parece não ser uma boa ideia, já que existem evidências de que não seja possível obter uma fórmula fechada para a solução do problema, envolvendo somas, diferenças, produtos, frações, potências e radicais [5].

Uma outra possibilidade seria tentar resolver o problema numericamente, que também pode apresentar dificuldades [46].

Exercício 2.3. *Considere um PGD com $K = 2$, $V = \{u, v, r\}$, $E = \{\{u, v\}, \{v, r\}\}$ e $d_{uv} = d_{vr} = 1$. Resolva graficamente o PGD.*

Exercício 2.4. *Considerando o exercício anterior, como ficaria a solução do problema se acrescentarmos $\{u, r\}$ em E , com $d_{ur} = 1$?*

Antes de apresentarmos a maneira clássica de resolver o PGD, consideremos dois aspectos importantes do problema: cardinalidade do conjunto solução e complexidade computacional.

2.3 Número de Soluções do PGD

Qual a importância do conhecimento da quantidade de soluções do PGD? É possível ter essa informação, antes de resolver o problema?

Além da importância teórica, a cardinalidade do conjunto de soluções de uma instância do PGD pode ajudar na solução do problema e fornecer informações relacionadas à aplicação que originou o PGD. Isso ficará mais claro no Capítulo 5.

Pelos dois exercícios da seção anterior, vimos que a quantidade de soluções de um PGD pode ser infinita. Mas isso sempre acontece?

Usando a desigualdade triangular, sabemos que, dados 3 pontos x_u, x_v, x_r em \mathbb{R}^2 , por exemplo, a desigualdade abaixo deve ser satisfeita:

$$d_{uv} + d_{vr} \geq d_{ur},$$

onde d_{uv}, d_{vr}, d_{ur} são as distâncias entre os pontos dados.

Isso significa que, ao acrescentarmos a aresta $\{u, r\}$ em E (em nosso último exercício), o problema poderia não ter solução, caso a desigualdade triangular não fosse satisfeita.

Pode-se verificar facilmente que, com uma dada solução de um PGD, podemos obter uma quantidade infinita (não enumerável) de realizações válidas distintas através de rotações e translações. Eliminando, portanto, as soluções obtidas via rotações e translações, sabemos, até o momento, que o conjunto solução de um PGD pode ser de três tipos:

- Vazio,
- Finito,
- Infinito não enumerável.

Falta considerar apenas mais um caso: conjunto solução com uma quantidade infinita, mas enumerável. É possível? A resposta é não. Obviamente, como não temos como apresentar um exemplo, temos que provar a afirmação. Infelizmente, não conhecemos uma maneira simples de justificar essa impossibilidade. Mas o resultado pode ser obtido usando geometria algébrica [6].

Exercício 2.5. *Considere um PGD com $K = 2$, $V = \{u, v, r, s\}$, $E = \{\{u, v\}, \{u, r\}, \{v, r\}, \{v, s\}\}$ e $d_{uv} = d_{ur} = d_{vr} = d_{vs} = 1$. A menos de rotações e translações, quantas soluções existem?*

Exercício 2.6. *Considerando o problema anterior, quantas soluções existirão, ao acrescentarmos $\{u, s\}$ em E , com $d_{us} = 2$?*

No exercício anterior, deu para perceber o “salto” do contínuo para o discreto, sem passar pelo caso “intermediário” (infinito enumerável)?

2.4 Complexidade do PGD

Nesta seção, estamos interessados em saber qual é dificuldade computacional de se resolver o PGD. Vamos considerar, inicialmente, que o grafo associado é completo.

Considere um PGD com $K = 1$, $V = \{u, v, r\}$, $E = \{\{u, v\}, \{u, r\}, \{v, r\}\}$, $d_{uv} = d_{vr} = 1$ e $d_{ur} = 2$. Fixando $x_u = 0$ e $x_v = 1$, temos

$$\begin{aligned} \|x_r - x_u\| &= 2 \\ \|x_r - x_v\| &= 1, \end{aligned}$$

ou ainda, elevando ao quadrado ambos os termos das igualdades,

$$\begin{aligned} x_r^2 - 2x_r x_u + x_u^2 &= 4 \\ x_r^2 - 2x_r x_v + x_v^2 &= 1. \end{aligned}$$

Subtraindo uma equação da outra, obtemos:

$$-2x_r x_u + 2x_r x_v + x_u^2 - x_v^2 = 3 \Rightarrow 2x_r(x_v - x_u) = x_v^2 - x_u^2 + 3.$$

Usando os valores fixados para x_u e x_v , temos:

$$2x_r = 4 \Rightarrow x_r = 2.$$

É claro, poderíamos ter resolvido o problema simplesmente fazendo um desenho, mas o interessante desse procedimento é que ele pode ser generalizado para o \mathbb{R}^K . No \mathbb{R}^2 , vejamos o que acontece.

Considere, então, um PGD com $K = 2$, $V = \{u, v, r, s\}$ e $E = \{\{u, v\}, \{u, r\}, \{u, s\}, \{v, r\}, \{v, s\}, \{r, s\}\}$. Fixando u, v, r , ou seja, determinando $x_u, x_v, x_r \in \mathbb{R}^2$ de tal modo que $\|x_u - x_v\| = d_{uv}$, $\|x_u - x_r\| = d_{ur}$ e $\|x_v - x_r\| = d_{vr}$, podemos montar um sistema quadrático para posicionar s , da seguinte forma:

$$\begin{aligned}\|x_s - x_u\| &= d_{us} \\ \|x_s - x_v\| &= d_{vs} \\ \|x_s - x_r\| &= d_{rs}.\end{aligned}$$

Elevando ao quadrado ambos os termos das igualdades,

$$\begin{aligned}\|x_s\|^2 - 2(x_s \cdot x_u) + \|x_u\|^2 &= d_{us}^2 \\ \|x_s\|^2 - 2(x_s \cdot x_v) + \|x_v\|^2 &= d_{vs}^2 \\ \|x_s\|^2 - 2(x_s \cdot x_r) + \|x_r\|^2 &= d_{rs}^2,\end{aligned}$$

e subtraindo a primeira equação das outras duas, obtemos:

$$\begin{aligned}2(x_s \cdot x_v) - 2(x_s \cdot x_u) &= \|x_v\|^2 - \|x_u\|^2 + d_{us}^2 - d_{vs}^2 \\ 2(x_s \cdot x_r) - 2(x_s \cdot x_u) &= \|x_r\|^2 - \|x_u\|^2 + d_{us}^2 - d_{rs}^2.\end{aligned}$$

Para ficar claro quem são as variáveis do problema, podemos reescrever as equações acima de uma forma ainda melhor:

$$\begin{aligned}2(x_v - x_u) \cdot x_s &= \|x_v\|^2 - \|x_u\|^2 + d_{us}^2 - d_{vs}^2 \\ 2(x_r - x_u) \cdot x_s &= \|x_r\|^2 - \|x_u\|^2 + d_{us}^2 - d_{rs}^2.\end{aligned}$$

Ou seja, temos um sistema linear

$$Ax = b,$$

onde

$$\begin{aligned}A &= 2 \begin{bmatrix} x_{v1} - x_{u1} & x_{v2} - x_{u2} \\ x_{r1} - x_{u1} & x_{r2} - x_{u2} \end{bmatrix}, \\ b &= \begin{bmatrix} \|x_v\|^2 - \|x_u\|^2 + d_{us}^2 - d_{vs}^2 \\ \|x_r\|^2 - \|x_u\|^2 + d_{us}^2 - d_{rs}^2 \end{bmatrix}\end{aligned}$$

e

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

Se a matriz A for inversível, temos uma única solução x^* , dada por

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

Caso não exista a inversa de A , o que fazemos? Podemos, por exemplo, fixar outros três vértices diferentes de $\{u, v, r\}$ para tentar obter um sistema com matriz inversível.

Conclusão: se o grafo G de um PGD for completo (supondo existência de solução), podemos obter a realização de G resolvendo vários sistemas lineares. Claro, para que isso seja possível, as matrizes associadas teriam que ser inversíveis.

Exercício 2.7. *No exemplo anterior, como poderíamos fixar u, v, r ?*

Exercício 2.8. *No exemplo anterior, que condições sobre os pontos x_u, x_v, x_r podemos impor para garantir que o sistema linear associado tenha solução?*

Exercício 2.9. *No exemplo anterior, se o sistema linear tem solução, pode-se garantir o mesmo para o sistema quadrático que o originou?*

Exercício 2.10. *Ainda considerando o exemplo anterior, se nenhuma das outras possibilidades para a escolha dos três primeiros vértices ($\{u, v, s\}, \{u, r, s\}, \{v, r, s\}$) produzir uma matriz inversível, significa que o problema não tem solução?*

Admitindo que todas as matrizes associadas a um PGD com grafo completo possam ser invertidas, o problema tem uma única solução que pode ser obtida com um custo computacional proporcional a $|V|$. Diz-se, então, que o problema pode ser resolvido em tempo linear [17].

Bom, mas já sabemos que esse caso não retrata a grande maioria das aplicações, onde faltam distâncias. Aplicando o procedimento anterior, obteríamos apenas uma realização parcial do grafo. Antes de continuar nossa aventura com o PGD, enunciamos um resultado teórico sobre a **Complexidade Computacional do PGD** [63]:

Teorema 2.1. *O PGD é um problema NP-completo para $K = 1$ e NP-difícil para $K \geq 2$.*

Isso significa, de uma maneira simplificada, que resolver um PGD pode exigir um elevado custo computacional, proporcional a $2^{|V|}$.

Exercício 2.11. *Para resolvermos uma instância do PGD em tempo linear, o grafo associado precisa ser necessariamente completo?*

Exercício 2.12. *Podemos ter uma única solução, mesmo o grafo não sendo completo?*

2.5 Instâncias do PGD

Para testar qualquer algoritmo para o PGD, até mesmo para analisar a relação entre a estrutura do grafo e o conjunto solução, é importante ter à disposição um conjunto de instâncias. Nesta seção, apresentamos uma maneira de gerar instâncias do PGD para $K = 2$ e $K = 3$.

Como esse procedimento foi desenvolvido pensando no PGD associado a problemas de cálculo de estrutura molecular, iniciamos construindo uma estrutura 3D de uma molécula artificial. Para simplificar (o que não significa, de maneira alguma, diminuir a complexidade computacional do problema), consideremos uma sequência de átomos ligados covalentemente, sem nenhuma ramificação. Ou seja, cada átomo está ligado a apenas dois outros, exceto o primeiro e o último da sequência, que denotaremos por $1, \dots, n$.

Para definir uma estrutura espacial de nossa molécula artificial, podemos usar o sistema de coordenadas cartesianas, $x_1, \dots, x_n \in \mathbb{R}^3$, definindo a posição de cada átomo, ou o sistema de *coordenadas internas* [66], dado 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). Cada ângulo de torção $\omega_{i-3,i}$ é, na verdade, o

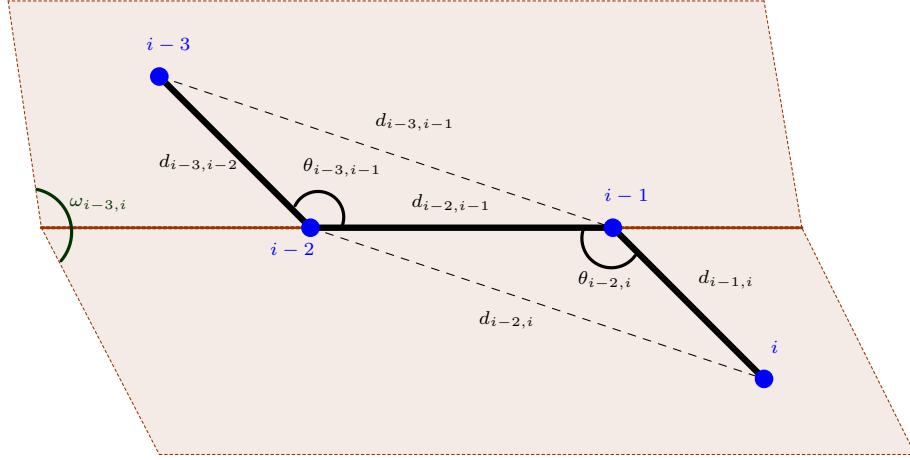


Figura 2.1: Ângulo de torção $\omega_{i-3,i}$.

ângulo entre os vetores normais dos planos definidos pelos átomos $i-3, i-2, i-1$ e $i-2, i-1, i$, respectivamente (ver Figura 2.1).

Uma maneira ainda mais fácil de gerar instâncias, também sem reduzir a complexidade do problema, é fixar os comprimentos das ligações covalentes (por exemplo, $d_{i-1,i} = 1,526$) e os valores dos ângulos planos (por exemplo, $\theta_{i-2,i} = 1,91$ radianos). Dessa maneira, a menos de rotações e translações, a estrutura de nossa molécula artificial ficará determinada pelos valores dos ângulos de torção $\omega_{1,4}, \dots, \omega_{n-3,n}$, cada um podendo variar no intervalo $[0, 2\pi]$.

Para gerar, então, uma instância do PGD com $K = 3$, podemos escolher aleatoriamente valores para $\omega_{i-3,i} \in [0, 2\pi]$ e, por exemplo, selecionar pares de pontos i, j cujas distâncias euclidianas $d_{i,j}$ sejam menores do que um determinado valor. Para simular instâncias do PGD associadas ao cálculo de estrutura molecular, podemos selecionar pares de pontos i, j tais que $d_{i,j} \leq 5$ [34].

Mas como fazer a seleção dos pares de pontos i, j sem saber o valor das distâncias $d_{i,j}$? Para isso, precisamos obter as coordenadas cartesianas a partir das coordenadas internas, que podem ser obtidas usando as fórmulas seguintes:

$$\begin{bmatrix} x_{i1} \\ x_{i2} \\ x_{i3} \\ 1 \end{bmatrix} = B_1 B_2 \cdots B_i \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \quad \forall i = 1, \dots, n,$$

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

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

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

Exercício 2.13. *Como nossa molécula artificial pode ser descrita usando coordenadas cartesianas ou coordenadas internas, qual é a diferença em usar um ou outro sistema para gerar instâncias do PGD?*

Usando as matrizes acima e fixando os comprimentos das ligações covalentes $d_{1,2}, d_{2,3}$ e o valor do ângulo plano $\theta_{1,3}$, os três primeiros átomos da sequência são dados por:

$$\begin{aligned} x_1 &= \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \\ x_2 &= \begin{bmatrix} -d_{1,2} \\ 0 \\ 0 \end{bmatrix}, \\ x_3 &= \begin{bmatrix} -d_{1,2} + d_{2,3} \cos \theta_{1,3} \\ d_{2,3} \sin \theta_{1,3} \\ 0 \end{bmatrix}. \end{aligned}$$

Para gerar instâncias com $K = 2$, fazemos $\omega_{i-3,i} = 0, \forall i = 4, \dots, n$, e escolhemos aleatoriamente valores para $d_{i-1,i}$ e $\theta_{i-2,i}$.

Capítulo 3

Do Contínuo para o Discreto

3.1 Otimização Contínua e o PGD

Como vimos no Capítulo 2, o PGD pode ser descrito por um sistema quadrático, mas a maneira clássica de abordar o problema é representá-lo como um problema de otimização contínua [46].

Para entendermos essa abordagem, consideremos um PGD com $K = 2$, $V = \{u, v, s\}$, $E = \{\{u, v\}, \{v, s\}\}$, onde o sistema quadrático associado é

$$\begin{aligned}(x_{u1} - x_{v1})^2 + (x_{u2} - x_{v2})^2 &= d_{uv}^2 \\ (x_{v1} - x_{s1})^2 + (x_{v2} - x_{s2})^2 &= d_{vs}^2,\end{aligned}$$

que pode ser reescrito como

$$\begin{aligned}(x_{u1} - x_{v1})^2 + (x_{u2} - x_{v2})^2 - d_{uv}^2 &= 0 \\ (x_{v1} - x_{s1})^2 + (x_{v2} - x_{s2})^2 - d_{vs}^2 &= 0.\end{aligned}$$

Considerando a função $f: \mathbb{R}^6 \rightarrow \mathbb{R}$, definida por

$$\begin{aligned}f(x_{u1}, x_{u2}, x_{v1}, x_{v2}, x_{s1}, x_{s2}) &= \left((x_{u1} - x_{v1})^2 + (x_{u2} - x_{v2})^2 - d_{uv}^2 \right)^2 \\ &\quad + \left((x_{v1} - x_{s1})^2 + (x_{v2} - x_{s2})^2 - d_{vs}^2 \right)^2,\end{aligned}$$

não é difícil perceber que a solução $x^* \in \mathbb{R}^6$ do PGD associado pode ser encontrada resolvendo o seguinte problema:

$$\min_{x \in \mathbb{R}^6} f(x).$$

Isto é, queremos encontrar o ponto x^* em \mathbb{R}^6 que retorna o menor valor de f .

Exercício 3.1. *No problema acima, é possível dizer qual é o menor valor de f ? O resultado vale para qualquer PGD?*

Exercício 3.2. *Considerando ainda o mesmo problema, qual é a diferença entre resolver o sistema quadrático*

$$\begin{aligned}(x_{u1} - x_{v1})^2 + (x_{u2} - x_{v2})^2 - d_{uv}^2 &= 0 \\ (x_{v1} - x_{s1})^2 + (x_{v2} - x_{s2})^2 - d_{vs}^2 &= 0\end{aligned}$$

e a equação

$$f(x) = 0, x \in \mathbb{R}^6,$$

onde f é dada por

$$f(x_{u1}, x_{u2}, x_{v1}, x_{v2}, x_{s1}, x_{s2}) = \left((x_{u1} - x_{v1})^2 + (x_{u2} - x_{v2})^2 - d_{uv}^2 \right)^2 + \left((x_{v1} - x_{s1})^2 + (x_{v2} - x_{s2})^2 - d_{vs}^2 \right)^2$$

Ou seja, podemos encarar o PGD como um problema de minimização. Apesar de ser a abordagem clássica para o PGD, a grande dificuldade agora é que a função a ser minimizada tem vários mínimos locais e o que se deseja é encontrar o mínimo global, ou seja, o menor dos mínimos locais (Figura 3.1). Para complicar ainda mais a situação, a quantidade de mínimos locais cresce exponencialmente com o tamanho do problema, dado pela quantidade de vértices do grafo associado [46]. Em otimização, diferenciar um mínimo local de um global é um problema complicado [21, 50].

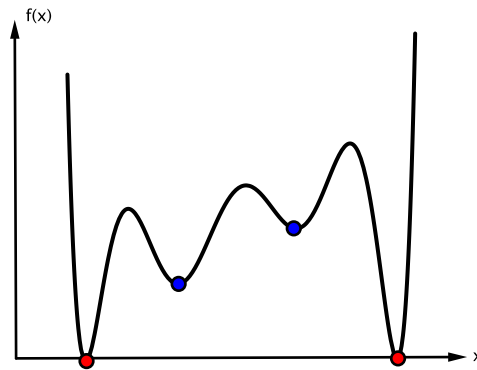


Figura 3.1: Mínimos locais e globais.

Exercício 3.3. *É possível existir mais de um mínimo global para o problema de otimização associado ao PGD?*

Exercício 3.4. *A quantidade de mínimos globais também pode crescer exponencialmente com o tamanho do problema?*

3.2 Finitude do PGD

Supondo que o conjunto solução de um PGD seja não vazio, já sabemos que ele é não enumerável ou finito. No caso finito, além de podermos aplicar os métodos clássicos, podemos explorar a estrutura do grafo que define o problema para pensar em novos métodos de solução [39].

A primeira questão é: como identificar se o conjunto solução é finito ou não enumerável? Nesta seção, iremos analisar que condições podem garantir a finitude do conjunto solução de um PGD.

Consideremos o mesmo problema da Seção 2.4 ($K = 2$), alterando apenas o valor de K para 3. Isto é, temos um PGD com $K = 3$, $V = \{u, v, r, s\}$ e $E =$

$\{\{u, v\}, \{u, r\}, \{u, s\}, \{v, r\}, \{v, s\}, \{r, s\}\}$. Fixando também u, v, r (podemos fazer isso usando as matrizes da Seção 2.5), obtemos o mesmo sistema quadrático:

$$\begin{aligned} \|x_s - x_u\|^2 &= d_{us}^2 \\ \|x_s - x_v\|^2 &= d_{vs}^2 \\ \|x_s - x_r\|^2 &= d_{rs}^2. \end{aligned}$$

Fazendo as contas e subtraindo a primeira equação das outras duas, obtemos, novamente:

$$\begin{aligned} 2(x_v - x_u) \cdot x_s &= \|x_v\|^2 - \|x_u\|^2 + d_{us}^2 - d_{vs}^2 \\ 2(x_r - x_u) \cdot x_s &= \|x_r\|^2 - \|x_u\|^2 + d_{us}^2 - d_{rs}^2. \end{aligned}$$

Até aqui, nenhuma diferença pode ser notada. Entretanto, obtendo explicitamente o sistema linear associado, temos:

$$\begin{bmatrix} x_{v1} - x_{u1} & x_{v2} - x_{u2} & x_{v3} - x_{u3} \\ x_{r1} - x_{u1} & x_{r2} - x_{u2} & x_{r3} - x_{u3} \end{bmatrix} \begin{bmatrix} x_{s1} \\ x_{s2} \\ x_{s3} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} \|x_v\|^2 - \|x_u\|^2 + d_{us}^2 - d_{vs}^2 \\ \|x_r\|^2 - \|x_u\|^2 + d_{us}^2 - d_{rs}^2 \end{bmatrix}.$$

Não temos mais uma matriz 2×2 , mas uma matriz 2×3 . E agora, temos solução única?

Podemos reescrever o sistema acima como

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

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

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

implicando que não temos mais uma única solução! Para cada valor de $x_{s3} \in \mathbb{R}$, obtemos valores para x_{s1} e x_{s2} . Finalmente, para obtermos a solução do nosso PGD, devemos retornar ao sistema quadrático associado, escolher uma das equações (por exemplo, $\|x_s - x_u\|^2 = d_{us}^2$) e resolvê-la usando a solução do sistema linear acima.

Geometricamente, temos a interseção entre uma reta, dada por

$$\begin{bmatrix} x_{s1} \\ x_{s2} \end{bmatrix} = A - B [x_{s3}],$$

onde

$$\begin{aligned} A &= \frac{1}{2} \begin{bmatrix} x_{v1} - x_{u1} & x_{v2} - x_{u2} \\ x_{r1} - x_{u1} & x_{r2} - x_{u2} \end{bmatrix}^{-1} \begin{bmatrix} \|x_v\|^2 - \|x_u\|^2 + d_{us}^2 - d_{vs}^2 \\ \|x_r\|^2 - \|x_u\|^2 + d_{us}^2 - d_{rs}^2 \end{bmatrix}, \\ B &= \begin{bmatrix} x_{v1} - x_{u1} & x_{v2} - x_{u2} \\ x_{r1} - x_{u1} & x_{r2} - x_{u2} \end{bmatrix}^{-1} \begin{bmatrix} x_{v3} - x_{u3} \\ x_{r3} - x_{u3} \end{bmatrix}, \end{aligned}$$

e uma esfera, dada por

$$\|x_s - x_u\| = d_{us},$$

resultando em 3 possibilidades (Figura 3.2):

- 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).

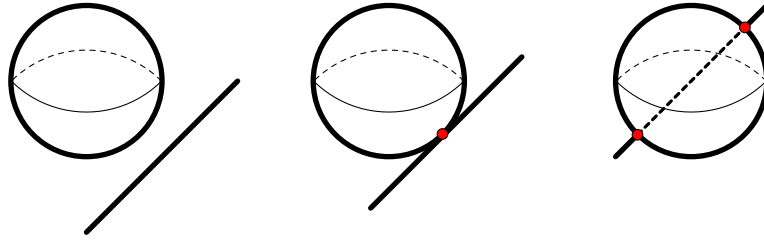


Figura 3.2: Interseções entre uma reta e uma esfera.

Exercício 3.5. *Algebricamente, qual é o significado das interpretações geométricas acima?*

Exercício 3.6. *Mostre a equivalência entre o sistema quadrático original, dado por*

$$\begin{aligned} \|x_s - x_u\|^2 &= d_{us}^2 \\ \|x_s - x_v\|^2 &= d_{vs}^2 \\ \|x_s - x_r\|^2 &= d_{rs}^2, \end{aligned}$$

e o novo sistema, dado por

$$\left\| \begin{bmatrix} \mathbf{x}_{s1} \\ \mathbf{x}_{s2} \\ \mathbf{x}_{s3} \end{bmatrix} - \begin{bmatrix} x_{u1} \\ x_{u2} \\ x_{u3} \end{bmatrix} \right\|^2 = d_{us}^2,$$

$$\begin{bmatrix} \mathbf{x}_{s1} \\ \mathbf{x}_{s2} \end{bmatrix} = \begin{bmatrix} x_{v1} - x_{u1} & x_{v2} - x_{u2} \\ x_{r1} - x_{u1} & x_{r2} - x_{u2} \end{bmatrix}^{-1} \times \left[\frac{1}{2} \begin{bmatrix} \|x_v\|^2 - \|x_u\|^2 + d_{us}^2 - d_{vs}^2 \\ \|x_r\|^2 - \|x_u\|^2 + d_{us}^2 - d_{rs}^2 \end{bmatrix} - \begin{bmatrix} x_{v3} - x_{u3} \\ x_{r3} - x_{u3} \end{bmatrix} \right] [\mathbf{x}_{s3}],$$

onde as variáveis são $\mathbf{x}_{s1}, \mathbf{x}_{s2}, \mathbf{x}_{s3} \in \mathbb{R}$.

Exercício 3.7. *Mostre também que a escolha da equação quadrática para formar o novo sistema é indiferente.*

Se a matriz $\begin{bmatrix} x_{v1} - x_{u1} & x_{v2} - x_{u2} \\ x_{r1} - x_{u1} & x_{r2} - x_{u2} \end{bmatrix}$ não for inversível, ou ainda, se não pudermos selecionar uma matriz inversível 2×2 da matriz original

$$\begin{bmatrix} x_{v1} - x_{u1} & x_{v2} - x_{u2} & x_{v3} - x_{u3} \\ x_{r1} - x_{u1} & x_{r2} - x_{u2} & x_{r3} - x_{u3} \end{bmatrix},$$

significa que não poderemos mais garantir que o conjunto solução seja finito?

A discussão acima nos sugere dois aspectos importantes sobre a finitude de um PGD:

1. Para cada vértice $s \in V$ a ser realizado em \mathbb{R}^3 , devem existir arestas $\{u, s\}, \{v, s\}, \{r, s\} \in E$ tais que os vértices $u, v, r \in V$ já tenham sido realizados, para que se possa gerar um sistema quadrático com $x_s \in \mathbb{R}^3$ como única variável, dado por

$$\begin{aligned} \|x_s - x_u\|^2 &= d_{us}^2 \\ \|x_s - x_v\|^2 &= d_{vs}^2 \\ \|x_s - x_r\|^2 &= d_{rs}^2. \end{aligned}$$

2. Para que esse sistema tenha, no máximo, duas soluções, a matriz do sistema linear obtido subtraindo uma equação das outras duas deve ter posto completo.

Exercício 3.8. *Que interpretação geométrica pode ser dada ao exigirmos que a matriz $\begin{bmatrix} x_{v1} - x_{u1} & x_{v2} - x_{u2} & x_{v3} - x_{u3} \\ x_{r1} - x_{u1} & x_{r2} - x_{u2} & x_{r3} - x_{u3} \end{bmatrix}$ tenha posto completo?*

Exercício 3.9. *Em \mathbb{R}^2 , como ficariam as condições acima?*

Exercício 3.10. *Na prática, que condição é mais importante?*

Exercício 3.11. *Tendo mais de 3 arestas na condição 1 acima, como fazer a escolha?*

3.3 Ordens para o PGD

A parte fundamental das duas condições acima é a seguinte: “Para cada vértice $s \in V$ a ser realizado em \mathbb{R}^3 , devem existir arestas $\{u, s\}, \{v, s\}, \{r, s\} \in E$ tais que os vértices $u, v, r \in V$ já tenham sido realizados”.

Destaquemos as duas informações realmente importantes, relacionadas ao vértice $s \in V$ que deve ser posicionado em \mathbb{R}^3 :

1. Existem $u, v, r \in V$ tais que $\{\{u, s\}, \{v, s\}, \{r, s\}\} \subset E$,
2. $x_u, x_v, x_r \in \mathbb{R}^3$ fazem parte de uma realização “parcial” válida.

A ideia que conecta essas duas informações está relacionada ao conceito de **ordem nos vértices** do grafo $G = (V, E, d)$ do PGD. Ou seja, se existir uma ordem em V satisfazendo as condições 1 e 2 acima, podemos garantir (a menos de rotações e translações e supondo que os pontos associados aos vértices u, v, r não sejam colineares) que o conjunto solução do problema é finito.

Recordemos que resolver um PGD é obter uma realização válida $x: V \rightarrow \mathbb{R}^3$ do grafo associado, o que implica, por sua vez, definir um ponto $x_s \in \mathbb{R}^3$ para cada $s \in V$, satisfazendo todas as equações do sistema

$$\forall \{u, v\} \in E, \|x_u - x_v\| = d_{uv}.$$

Uma solução do problema pode ser representada, então, como um elemento do $\mathbb{R}^{3|V|}$.

Com essa estrutura de ordem em V , devemos ainda considerar todo o espaço de busca ($\mathbb{R}^{3|V|}$) para tentar encontrar uma solução? Antes de considerar essa questão, de extrema importância, analisemos com mais cuidado a condição 1 acima, supondo que a condição 2 é sempre satisfeita.

Uma pergunta imediata é a seguinte: como podemos descobrir se tal ordem existe? Inicialmente, vamos definir precisamente nosso novo problema, **Ordenação no PGD**:

Definição 3.1. *Dado o grafo $G = (V, E, d)$ de um PGD com $K = 3$, existe uma ordem $<$ em V tal que os 3 primeiros vértices formam uma clique e, $\forall s \in V$, a partir do quarto elemento, existem $u, v, r \in V$, com $u < s, v < s, r < s$ e $\{u, s\}, \{v, s\}, \{r, s\} \in E$?*

Supondo que a clique inicial seja dada (o que, de fato, acontece na maioria das aplicações), o problema pode ser resolvido em tempo polinomial, usando a ideia de posicionar os vértices seguintes considerando o maior número possível de arestas aos antecessores [35]. Isto é, para a clique dada, pode-se dizer, em tempo polinomial, se existe ou não tal ordem.

Exercício 3.12. *A ordem depende da clique inicial?*

Exercício 3.13. *Descreva o método de ordenação mencionado acima.*

Exercício 3.14. *Qual é o custo computacional desse método?*

Exercício 3.15. *Demonstre que se o método pára sem enumerar todos os vértices, tal ordem não existe.*

Consideremos um PGD com $V = \{a, b, c, u, v, r\}$ e $E = \{\{a, b\}, \{a, c\}, \{a, u\}, \{b, c\}, \{b, v\}, \{c, r\}, \{u, v\}, \{u, r\}, \{v, r\}\}$. Para $K = 2$, queremos encontrar uma ordem tal que os dois primeiros vértices formem uma clique e, a partir do terceiro, existam dois vértices anteriores ligados a ele. Considerando todas as cliques possíveis, vejamos o que acontece:

- Iniciando com a clique $\{a, b\}$, temos os seguintes vértices como candidatos à terceira posição: c, u, v, r . O próximo vértice seria c , pois $\{\{a, c\}, \{b, c\}\} \subset E$, mas depois dele, não haveria mais candidatos.
- Iniciando com a clique $\{a, c\}$, temos os seguintes vértices como candidatos à terceira posição: b, u, v, r . O próximo vértice seria b , pois $\{\{a, b\}, \{c, b\}\} \subset E$, mas depois dele, não haveria mais candidatos.
- Iniciando com a clique $\{a, u\}$, não haveria nenhum candidato para ocupar a terceira posição.
- Iniciando com a clique $\{b, c\}$, temos os seguintes vértices como candidatos à terceira posição: a, u, v, r . O próximo vértice seria a , pois $\{\{b, a\}, \{c, a\}\} \subset E$, mas depois dele, não haveria mais candidatos.

- Iniciando com a clique $\{b, v\}$, não haveria nenhum candidato para ocupar a terceira posição.
- Iniciando com a clique $\{c, r\}$, não haveria nenhum candidato para ocupar a terceira posição.
- Iniciando com a clique $\{u, v\}$, temos os seguintes vértices como candidatos à terceira posição: a, b, c, r . O próximo vértice seria r , pois $\{\{u, r\}, \{v, r\}\} \subset E$, mas depois dele, não haveria mais candidatos.
- Iniciando com a clique $\{u, r\}$, temos os seguintes vértices como candidatos à terceira posição: a, b, c, v . O próximo vértice seria v , pois $\{\{u, v\}, \{r, v\}\} \subset E$, mas depois dele, não haveria mais candidatos.
- Iniciando com a clique $\{v, r\}$, temos os seguintes vértices como candidatos à terceira posição: a, b, c, u . O próximo vértice seria u , pois $\{\{v, u\}, \{r, u\}\} \subset E$, mas depois dele, não haveria mais candidatos.

Portanto, não existe tal ordem! Podemos concluir, então, que o conjunto solução é não enumerável? Não, pois se sabe que o grafo associado não tem uma quantidade não enumerável de realizações [31] (Figura 3.3).

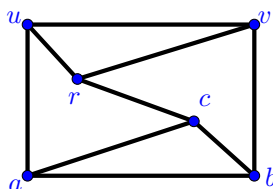


Figura 3.3: Grafo sem ordenação no PGD.

Ou seja, a existência da ordem que procuramos é uma condição apenas suficiente para a finitude do PGD. Essa questão está relacionada a uma outra área de pesquisa chamada *Rigidez de Grafos* [26].

Capítulo 4

O Problema Discretizável de Geometria de Distâncias (PDGD₃)

4.1 Definição do PDGD₃

Usando a informação de ordem nos vértices do grafo de um PGD, temos um novo problema, que chamaremos de **Problema Discretizável de Geometria de Distâncias**. Apesar da definição poder ser estendida para um inteiro positivo qualquer [46], iremos considerar apenas o caso $K = 3$, que chamaremos de **PDGD₃** [55]:

Definição 4.1. *Considere o grafo $G = (V, E, d)$ de um PGD com $K = 3$ e uma ordem em V , denotada por v_1, \dots, v_n , tal que*

- existe uma realização válida para v_1, v_2, v_3 ;
- para todo v_i , $i = 4, \dots, n$, existem, pelo menos, 3 vértices anteriores a_i, b_i, c_i com $\{\{a_i, v_i\}, \{b_i, v_i\}, \{c_i, v_i\}\} \subset E$ tais que

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

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

$$\forall \{v_i, v_j\} \in E, \quad \|x_{v_i} - x_{v_j}\| = d_{v_i v_j}.$$

O PDGD₃, na verdade, é um caso particular do PGD, onde $K = 3$ e uma ordem é dada como parte da entrada do problema.

4.2 Encarando o PDGD₃

Graças à estrutura de ordem que temos nos vértices do grafo associado ao PDGD₃, podemos atacar o problema aproveitando essa informação que, de fato, aparece nas aplicações [41]. Essencialmente, a ideia é realizar vértice por vértice, seguindo a ordem dada.

Consideremos um PGD com $K = 3$, dado pelo grafo $G = (V, E)$:

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

Podemos fazer a ordenação, seguindo o procedimento descrito na Seção 3.3, resultando no seguinte:

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

Com essa ordem, trata-se de um PDGD₃ (supondo pesos nas arestas satisfazendo as respectivas desigualdades triangulares), pois $\{r, q, t\}$ é uma clique (com uma realização válida) 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 $r = v_1, q = v_2, t = v_3, s = v_4, v = v_5, u = v_6, p = v_7$, o grafo do nosso PDGD₃ é dado por

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

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

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

Pelas hipóteses do PDGD₃, v_1, v_2, v_3 já estão 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 com $x_{v_4} \in \mathbb{R}^3$ como única variável, obtemos:

$$\begin{aligned} \|x_{v_4} - x_{v_1}\|^2 &= d_{v_1 v_4}^2 \\ \|x_{v_4} - x_{v_2}\|^2 &= d_{v_2 v_4}^2 \\ \|x_{v_4} - x_{v_3}\|^2 &= d_{v_3 v_4}^2. \end{aligned}$$

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

$$Ax_{v_4} = b,$$

onde $A \in \mathbb{R}^{2 \times 3}$ e $b \in \mathbb{R}^2$. Como A tem posto completo (hipótese do PDGD₃), obtemos x_{v_4} em função de um parâmetro $\lambda \in \mathbb{R}$, denotado por $x_{v_4}(\lambda)$. Substituindo $x_{v_4}(\lambda)$ em uma das equações do sistema quadrático acima, por exemplo, na equação $\|x_{v_4} - x_{v_1}\|^2 = d_{v_1 v_4}^2$, obtemos a seguinte equação do segundo grau em λ :

$$\|x_{v_4}(\lambda)\|^2 - 2x_{v_4}(\lambda) \cdot x_{v_1} + \|x_{v_1}\|^2 - d_{v_1 v_4}^2 = 0.$$

Com os valores de λ , obtemos os possíveis resultados para x_{v_4} .

Exercício 4.1. *Geometricamente, o que significa obter duas raízes iguais?*

Exercício 4.2. *Poderíamos obter raízes complexas?*

Para cada posição de v_4 ($x_{v_4}^0$ e $x_{v_4}^1$), fazendo o mesmo procedimento acima, podemos também obter duas posições para v_5 . Ou seja, considerando

$$\begin{aligned} a_5 &= v_2 \\ b_5 &= v_3 \\ c_5 &= v_4 \end{aligned}$$

e escolhendo $x_{v_4}^0$, temos o seguinte sistema quadrático:

$$\begin{aligned} \|x_{v_5} - x_{v_2}\|^2 &= d_{v_2 v_5}^2 \\ \|x_{v_5} - x_{v_3}\|^2 &= d_{v_3 v_5}^2 \\ \|x_{v_5} - x_{v_4}^0\|^2 &= d_{v_4 v_5}^2. \end{aligned}$$

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

$$\|x_{v_5}^0 - x_{v_1}\| = d_{v_1 v_5} \text{ ou } \|x_{v_5}^1 - x_{v_1}\| = d_{v_1 v_5}?$$

Supondo que os pontos $\{x_{v_1}, x_{v_2}, x_{v_3}, x_{v_4}^0\}$ não sejam coplanares, apenas uma das equações deve ser satisfeita, digamos a primeira. Ou seja, descartamos $x_{v_5}^1$ e consideremos $x_{v_5}^0$.

Exercício 4.3. *Qual seria a diferença, caso tivéssemos selecionado $x_{v_4}^1$?*

Exercício 4.4. *Qual o motivo da hipótese sobre planaridade dos pontos $\{x_{v_1}, x_{v_2}, x_{v_3}, x_{v_4}^0\}$?*

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

$$\begin{aligned} \|x_{v_6} - x_{v_2}\|^2 &= d_{v_2 v_6}^2 \\ \|x_{v_6} - x_{v_3}\|^2 &= d_{v_3 v_6}^2 \\ \|x_{v_6} - x_{v_5}^0\|^2 &= d_{v_5 v_6}^2, \end{aligned}$$

e obtemos as duas soluções possíveis: $x_{v_6}^0$ e $x_{v_6}^1$. 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, $x_{v_6}^0$), mas ao mesmo tempo, lembrar que há outra possibilidade.

Vamos ao vértice seguinte: v_7 . Considerando

$$\begin{aligned} a_7 &= v_4 \\ b_7 &= v_5 \\ c_7 &= v_6, \end{aligned}$$

temos um novo sistema quadrático,

$$\begin{aligned} \|x_{v_7} - x_{v_4}\|^2 &= d_{v_4 v_7}^2 \\ \|x_{v_7} - x_{v_5}\|^2 &= d_{v_5 v_7}^2 \\ \|x_{v_7} - x_{v_6}^0\|^2 &= d_{v_6 v_7}^2, \end{aligned}$$

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

$$\|x_{v_7}^0 - x_{v_1}\| = d_{v_1 v_7} \text{ ou } \|x_{v_7}^1 - x_{v_1}\| = d_{v_1 v_7}$$

e

$$\|x_{v_7}^0 - x_{v_2}\| = d_{v_2 v_7} \text{ ou } \|x_{v_7}^1 - x_{v_2}\| = d_{v_2 v_7}.$$

Agora, temos duas possibilidades:

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

O primeiro caso pode ocorrer, caso a seleção que fizemos para v_6 tenha sido “errada”. Então, temos que “retornar”, escolher $x_{v_6}^1$ 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 $x_{v_7}^0$ e $x_{v_7}^1$.

Mas vamos supor que tenha ocorrido o primeiro caso. Então, retornamos, escolhemos $x_{v_6}^1$ e resolvemos um novo sistema quadrático, obtendo $y_{v_7}^0$ e $y_{v_7}^1$. Como temos $\{v_1, v_7\} \in E$ (supondo que os pontos $\{x_{v_1}, x_{v_2}, x_{v_4}, x_{v_6}^1\}$ não sejam coplanares), apenas uma das posições é válida, digamos $y_{v_7}^1$. Portanto, finalmente, encontramos uma solução do problema: $\{x_{v_1}, x_{v_2}, x_{v_3}, x_{v_4}^0, x_{v_5}^0, x_{v_6}^1, y_{v_7}^1\}$.

Exercício 4.5. *No momento da escolha entre $x_{v_6}^0$ e $x_{v_6}^1$, é possível antecipar que escolhendo $x_{v_6}^0$ chegaríamos num ponto inviável?*

Exercício 4.6. *Usando o procedimento acima, o que garante que encontraremos, de fato, uma solução?*

Exercício 4.7. *Na prática, que critério pode ser adotado para decidir que aresta deve ser utilizada, no caso de existirem várias arestas adicionais e supondo que todas elas sejam teoricamente “válidas”.*

4.3 Complexidade do PDGD₃

O custo computacional do procedimento acima pode ser elevado (exponencial), como o exemplo sugere. É possível criarmos um algoritmo polinomial para o PDGD₃?

Na Seção 2.4, mencionamos que o PGD é um problema NP-difícil para $K \geq 2$. Já que o PDGD₃ é um subproblema do PGD, com conjunto solução finito, é natural questionar se houve alguma alteração na complexidade computacional do problema. Pode ser demonstrado que o PDGD₃ ainda continua sendo NP-difícil [55].

Recordemos que na ordem associada ao grafo $G = (V, E)$ do PDGD₃, para todo $v_i \in V$, $i = 4, \dots, n$, existem, pelo menos, 3 vértices anteriores a_i, b_i, c_i com $\{\{a_i, v_i\}, \{b_i, v_i\}, \{c_i, v_i\}\} \subset E$. Vimos no exemplo anterior que, dependendo da instância do PDGD₃, podemos ter mais de 3 vértices. Supondo que a_i, b_i, c_i e o vértice adicional gerem pontos não coplanares e que o sistema quadrático associado tenha solução, teremos apenas uma solução, o que diminui o custo computacional do procedimento [11].

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”. Generalizando esse fato, podemos dizer que, se para todo $v_i \in V$, $i = 5, \dots, n$, existirem, pelo menos, 4 vértices anteriores a_i, b_i, c_i, d_i com $\{\{a_i, v_i\}, \{b_i, v_i\}, \{c_i, v_i\}, \{d_i, v_i\}\} \subset E$, supondo que $\{a_i, b_i, c_i, d_i\}$ não sejam coplanares, o PDGD₃ poderia ser resolvido em tempo linear!

Podemos interpretar os “3 vértices anteriores” como um “ponto crítico”: eles garantem a finitude (de forma suficiente, mas não necessária), no “limite” de atingir a polinomialidade, obtida com 4 vértices anteriores. Será coincidência o fato da dimensão do espaço de realização do grafo ser também 3?

Exercício 4.8. *O que quer dizer a afirmação acima: “eles garantem a finitude (de forma suficiente, mas não necessária)”?*

Exercício 4.9. *Se temos um PDGD₃ com conjunto solução não vazio, onde $\forall v_i \in V$, $i = 5, \dots, n$, existem (pelo menos) 4 vértices anteriores a_i, b_i, c_i, d_i com $\{\{a_i, v_i\}, \{b_i, v_i\}, \{c_i, v_i\}, \{d_i, v_i\}\} \subset E$ e $\{a_i, b_i, c_i, d_i\}$ não coplanares, podemos afirmar que a solução é única?*

4.4 O Algoritmo BP para o PDGD₃

A ordem do PDGD₃ garante a finitude do conjunto solução do problema (supondo que as respectivas desigualdades triangulares são satisfeitas de forma estrita) e, além disso, “organiza” o espaço onde devemos fazer a busca por uma solução. Apesar do **espaço de busca** continuar sendo contínuo (\mathbb{R}^{3n} , com n vértices), a ordem indica o caminho que deve ser percorrido para encontrar uma solução do problema, como vimos na seção anterior.

Na verdade, a ordem induz uma estrutura de **árvore binária** no espaço de busca (Figura 4.1), onde a raiz representa o ponto x_{v_1} e os dois nós seguintes representam os pontos x_{v_2} e x_{v_3} , inicialmente fixados. A partir do quarto nível da árvore, temos as representações de todas as possíveis posições para os vértices v_i , $i = 4, \dots, n$. Ou seja, 2 possibilidades para v_4 , 4 para v_5 , 8 para v_6 , 16 para $v_7, \dots, 2^{i-3}$ para $v_i, \dots, 2^{n-3}$ para v_n .

Como pode ser percebido pelo exemplo anterior, decidimos fazer uma “busca pela esquerda”, ou seja, quando não temos arestas adicionais, optamos por escolher a solução $x_{v_i}^0$ do sistema quadrático associado. Outras opções podem ser definidas

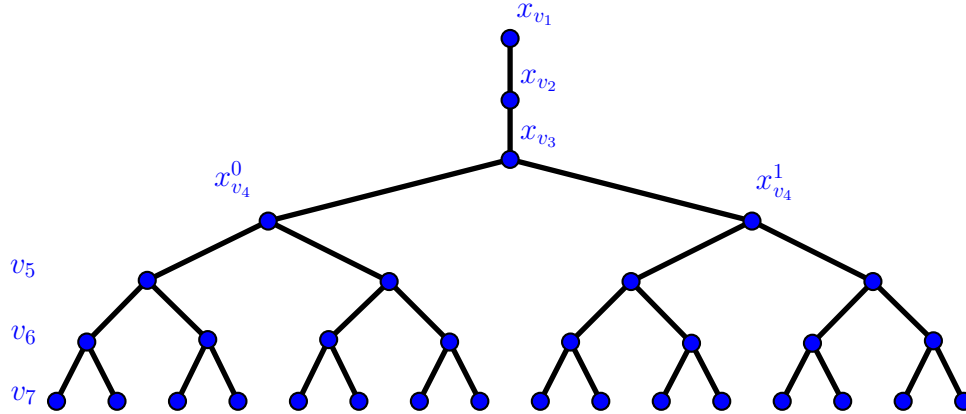


Figura 4.1: Árvore binária.

[20, 60], 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 PDGD₃ 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 do grafo do PDGD₃ 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” o caminho que deve ser percorrido, ao reduzir o espaço de busca.

Para encontrar uma solução, devemos tentar descer pela árvore, da raiz até o último nível, passando por todos os testes de viabilidade. Chegando no último nível, a solução é dada pelo caminho percorrido entre o primeiro e o último nível da árvore, considerando apenas os nós viáveis.

Vimos que, em alguns casos, temos que retroceder na árvore para refazer o caminho. Esses retrocessos é que podem aumentar (exponencialmente) o custo computacional da busca. Podemos descer pela árvore, 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 $\{\{a_i, v_i\}, \{b_i, v_i\}, \{c_i, v_i\}, \{d_i, v_i\}\} \subset E$.

Nesses dois casos, uma solução é encontrada rapidamente, com um custo computacional proporcional a n .

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

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

pois não há nenhum teste de viabilidade a ser feito.

No segundo caso, basta resolver os sistemas quadráticos

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

dando solução única, também com nenhum teste de viabilidade a ser feito, já que o único teste a ser realizado foi incluído no sistema (estamos supondo que o sistema tem solução).

A “natureza” do conjunto E_d é a mesma em qualquer PDGD₃, enquanto que o conjunto E_p depende de cada instância do problema. As arestas de discretização E_d definem os sistemas quadráticos e as arestas de poda E_p definem os testes de viabilidade.

Olhando com mais cuidado, podemos perceber que existe uma “geometria das arestas”, também como consequência da ordem do PDGD₃.

Consideremos um sistema quadrático genérico associado a v_i , $i = 4, \dots, n$:

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

Geometricamente, resolver esse sistema é obter a interseção de 3 esferas (Figura 4.2), $S(x_{a_i}, d_{a_i v_i})$, $S(x_{b_i}, d_{b_i v_i})$, $S(x_{c_i}, d_{c_i v_i})$, centradas em $x_{a_i}, x_{b_i}, x_{c_i} \in \mathbb{R}^3$ e com raios $d_{a_i v_i}, d_{b_i v_i}, d_{c_i v_i} \in (0, \infty)$, dada por

$$D = S(x_{a_i}, d_{a_i v_i}) \cap S(x_{b_i}, d_{b_i v_i}) \cap S(x_{c_i}, d_{c_i v_i}).$$

Pela desigualdade triangular estrita do PDGD₃, os pontos a_i, b_i, c_i não são colineares e, portanto, temos apenas 3 possibilidades:

- $|D| = 0$,
- $|D| = 1$,
- $|D| = 2$.

Supondo não haver erro nas informações das distâncias [40], o primeiro caso não acontece. Teremos o segundo caso, quando os pontos $x_{a_i}, x_{b_i}, x_{c_i}, x_{v_i}$ forem coplanares, e o terceiro é o caso geral. Ou seja,

$$D = \{x_{v_i}^0, x_{v_i}^1\},$$

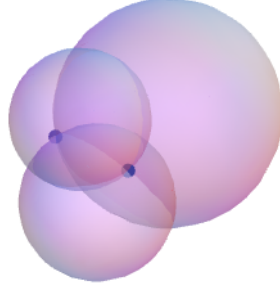


Figura 4.2: Interseção de três esferas.

onde pode acontecer que $x_{v_i}^0 = x_{v_i}^1$.

Quando $E_p \neq \emptyset$, temos esferas adicionais (Figura 4.3) que devem ser acrescentadas à interseção anterior. Ou seja, devemos fazer

$$P = D \cap \left(\bigcap_{i=1}^{|E_p|} S(x_{p_i}, d_{p_i v_i}) \right),$$

onde $p_i < v_i$, $\{p_i, v_i\} \in E$ e $S(x_{p_i}, d_{p_i v_i})$ é a esfera com centro no ponto x_{p_i} e raio $d_{p_i v_i}$. Essa nova interseção, $\bigcap_{i=1}^{|E_p|} S(x_{p_i}, d_{p_i v_i})$, é justamente o teste de viabilidade.

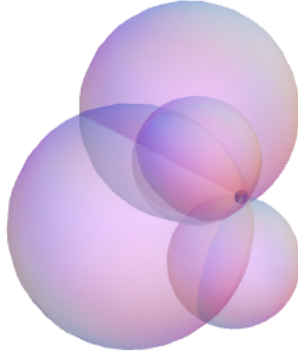


Figura 4.3: Interseção de quatro esferas.

Com as esferas adicionais, temos apenas duas possibilidades:

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

No primeiro caso, houve alguma inviabilidade, e no segundo, temos o ponto que satisfaz todas as distâncias dadas aos pontos anteriores já posicionados.

Essa interpretação geométrica nos faz pensar em outras alternativas de como encarar os dois subproblemas fundamentais envolvidos na resolução do PDGD₃ [2, 3, 23, 24, 58].

O algoritmo para encontrar uma solução do PDGD₃ é chamado de **Branch and Prune** [45], ou simplesmente **BP**.

Exercício 4.10. *Descreva os passos mais importantes do BP.*

Exercício 4.11. *Explique melhor a frase “quando os pontos $x_{a_i}, x_{b_i}, x_{c_i}, x_{v_i}$ forem coplanares”, que aparece no caso $|D| = 1$.*

Exercício 4.12. *Quando $E_p = \emptyset$, é possível saber quantas soluções tem o problema?*

Exercício 4.13. *O BP pode encontrar todas as soluções de um PDGD₃?*

Capítulo 5

O Problema Discretizável de Geometria de Distâncias Moleculares (PDGDM)

5.1 As Hipóteses do PDGD₃

Já sabemos que, para garantir a finitude do conjunto solução do PGD, podemos introduzir uma estrutura de ordem nos vértices do grafo associado. Se tal ordem existe, vimos também que não é difícil encontrá-la. Vejamos agora, com mais cuidado, a estrutura dessa ordem.

Supondo que o problema de encontrar a clique inicial $\{v_1, v_2, v_3\}$ esteja resolvido, o PDGD₃ assume que, para todo v_i , $i = 4, \dots, n$, existem (pelo menos) 3 vértices anteriores a_i, b_i, c_i , com $\{\{a_i, v_i\}, \{b_i, v_i\}, \{c_i, v_i\}\} \subset E$, tais que

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

Admitindo que a informação sobre as distâncias não contém nenhum tipo de erro [7], as desigualdades triangulares são satisfeitas, mas não necessariamente de maneira estrita. Mas a questão não é o fato de exigirmos $<$, no lugar de \leq (já sabemos que a desigualdade estrita é necessária para evitar uma quantidade não enumerável de soluções). O problema é que, dependendo da instância do PDGD₃, pode faltar alguma distância envolvendo os vértices a_i, b_i, c_i fazendo com que o sistema quadrático

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

não tenha solução em \mathbb{R}^3 .

Uma saída seria, então, exigir que, para todo v_i , $i = 4, \dots, n$, as três distâncias entre os vértices a_i, b_i, c_i sejam conhecidas, o que aconteceu no problema da Seção 4.2, mas não foi exigido na definição do PDGD₃ [55]. 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 [40].

Exercício 5.1. *Geometricamente, o que significa o sistema quadrático acima não ter solução.*

Exercício 5.2. *O fato dos vértices a_i, b_i, c_i formarem uma clique garante que a desigualdade triangular seja satisfeita?*

5.2 Definição do PDGDM

Agora, estamos interessados em encontrar uma ordem onde os vértices utilizados na construção de cada sistema quadrático formem uma clique e sejam imediatamente anteriores ao vértice que se deseja realizar. Considerando o mesmo PGD da Seção 4.2, existe uma ordem com tais características? Tente encontrá-la, antes de prosseguir.

Voltemos, então, ao nosso problema: um PGD com $K = 3$, onde o grafo associado $G = (V, E)$ é

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

Considere uma nova ordenação, dada por

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

Para facilitar, denotemos $p = u_1$, $u = u_2$, $v = u_3$, $q = u_4$, $t = u_5$, $r = u_6$ e $s = u_7$ (usamos u_i , no lugar de v_i , para enfatizar que estamos utilizando uma ordem diferente da anterior). Podemos verificar que essa ordem tem as propriedades desejadas, 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 associado, com $x_{u_4} \in \mathbb{R}^3$ como única variável:

$$\begin{aligned} \|x_{u_4} - x_{u_1}\|^2 &= d_{u_1 u_4}^2 \\ \|x_{u_4} - x_{u_2}\|^2 &= d_{u_2 u_4}^2 \\ \|x_{u_4} - x_{u_3}\|^2 &= d_{u_3 u_4}^2. \end{aligned}$$

Escolhemos uma das duas possibilidades para posicionar u_4 (digamos $x_{u_4}^0$) e obtemos o sistema quadrático seguinte para posicionar u_5 :

$$\begin{aligned} \|x_{u_5} - x_{u_2}\|^2 &= d_{u_2 u_5}^2 \\ \|x_{u_5} - x_{u_3}\|^2 &= d_{u_3 u_5}^2 \\ \|x_{u_5} - x_{u_4}^0\|^2 &= d_{u_4 u_5}^2. \end{aligned}$$

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

$$\begin{aligned} \|x_{u_6} - x_{u_3}\|^2 &= d_{u_3 u_6}^2 \\ \|x_{u_6} - x_{u_4}\|^2 &= d_{u_4 u_6}^2 \\ \|x_{u_6} - x_{u_5}^0\|^2 &= d_{u_5 u_6}^2. \end{aligned}$$

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

$$\|x_{u_6}^0 - x_{u_1}\| = d_{u_1 u_6} \text{ ou } \|x_{u_6}^1 - x_{u_1}\| = d_{u_1 u_6}?$$

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

$$\begin{aligned} \|x_{u_6} - x_{u_3}\|^2 &= d_{u_3 u_6}^2 \\ \|x_{u_6} - x_{u_4}\|^2 &= d_{u_4 u_6}^2 \\ \|x_{u_6} - x_{u_5}^1\|^2 &= d_{u_5 u_6}^2. \end{aligned}$$

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

$$\|y_{u_6}^0 - x_{u_1}\| = d_{u_1 u_6} \text{ ou } \|y_{u_6}^1 - x_{u_1}\| = d_{u_1 u_6}?$$

Supondo que os pontos $\{x_{u_1}, x_{u_3}, x_{u_4}^0, x_{u_5}^1\}$ não sejam coplanares, apenas uma das equações deve ser satisfeita, digamos a primeira. Ou seja, descartamos $y_{u_6}^1$ e consideramos $y_{u_6}^0$. O novo sistema quadrático é

$$\begin{aligned} \|x_{u_7} - x_{u_4}\|^2 &= d_{u_4 u_7}^2 \\ \|x_{u_7} - x_{u_5}\|^2 &= d_{u_5 u_7}^2 \\ \|x_{u_7} - y_{u_6}^0\|^2 &= d_{u_6 u_7}^2. \end{aligned}$$

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 ($x_{u_7}^0$ e $x_{u_7}^1$), temos:

$$\|x_{u_7}^0 - x_{u_1}\| = d_{u_1 u_7} \text{ ou } \|x_{u_7}^1 - x_{u_1}\| = d_{u_1 u_7}$$

e

$$\|x_{u_7}^0 - x_{u_3}\| = d_{u_3 u_7} \text{ ou } \|x_{u_7}^1 - x_{u_3}\| = d_{u_3 u_7}.$$

Fazendo o mesmo procedimento realizado com a ordem anterior, vamos supor que $x_{u_7}^1$ seja o ponto selecionado. Portanto, chegamos à solução do problema: $x_{u_1}, x_{u_2}, x_{u_3}, x_{u_4}^0, x_{u_5}^1, y_{u_6}^0, x_{u_7}^1$.

Até o momento, não foi possível perceber qualquer vantagem dessa nova ordem, além do fato de garantir que os sistemas quadráticos tenham solução. Entretanto, com as cliques $\{v_{i-3}, v_{i-2}, v_{i-1}, v_i\}$, para todo $i = 4, \dots, n$, podemos substituir a resolução dos sistemas quadráticos por algo bem mais simples e mais estável numericamente. Antes de prosseguirmos, formalizemos a definição do novo problema:

Problema Discretizável de Geometria de Distâncias Moleculares (PDGDM):

Definição 5.1. *Considere o grafo $G = (V, E, d)$ de um PGD com $K = 3$ e uma ordem em V , denotada por v_1, \dots, v_n , tal que*

- existe uma realização válida para v_1, v_2, v_3 ;
- para todo v_i , $i = 4, \dots, n$, existem, pelo menos, 3 vértices imediatamente anteriores $v_{i-3}, v_{i-2}, v_{i-1}$, onde $\{v_{i-3}, v_{i-2}, v_{i-1}, v_i\}$ é uma clique e

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

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

$$\forall \{v_i, v_j\} \in E, \quad \|x_{v_i} - x_{v_j}\| = d_{v_i v_j}. \quad (5.2.1)$$

5.3 Complexidade do PDGDM

O fato de existirem cliques $\{v_{i-3}, v_{i-2}, v_{i-1}, v_i\}$, $\forall i = 4, \dots, n$, nos dá ainda mais informação sobre a ordem dos vértices do grafo associado ao PGD. Usando as informações de distâncias das cliques $\{v_{i-3}, v_{i-2}, v_{i-1}, v_i\}$, podemos definir “quase” todos os valores abaixo:

- $d_{1,2}, \dots, d_{n-1,n}$: distâncias associadas a vértices consecutivos,
- $\theta_{1,3}, \dots, \theta_{n-2,n}$: ângulos planos associados a três vértices consecutivos,
- $\omega_{1,4}, \dots, \omega_{n-3,n}$: ângulos de torção associados a quatro vértices consecutivos.

Exercício 5.3. *Qual a razão do “quase” acima?*

Lembramos 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.

Os valores $d_{1,2}, \dots, d_{n-1,n}$ são, obviamente, obtidos pela própria definição do PDGDM, e os valores $\theta_{1,3}, \dots, \theta_{n-2,n}$ são obtidos pela Lei dos Cossenos. Entretanto, com as informações do PDGDM, podemos obter apenas os valores dos cossenos dos ângulos de torção, dados por ($\forall i = 4, \dots, n$) [37]:

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

onde

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

Na verdade, com a estrutura de ordem do PDGDM, podemos imaginar que temos uma “molécula artificial” (Figura 5.1), o que explica também o termo extra “moleculares” na nova classe de problemas.

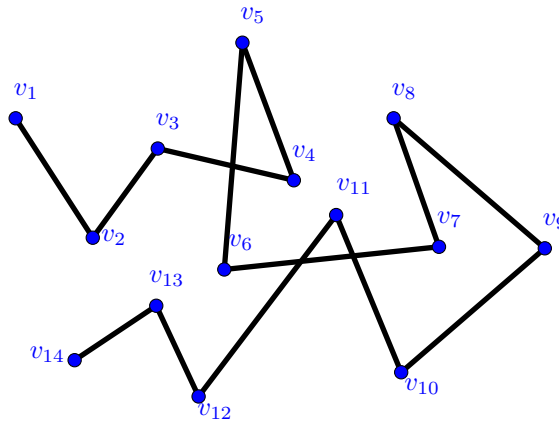


Figura 5.1: Molécula artificial do PDGDM.

Exercício 5.4. *Qual é a única distância que aparece apenas uma vez na fórmula acima para o $\cos(\omega_{i-3,i})$? Por que ela é diferente das demais?*

Lembramos que, para definir uma estrutura 3D de uma molécula, podemos usar o sistema de coordenadas internas, dado justamente pelos valores $d_{1,2}, \dots, d_{n-1,n}$, $\theta_{1,3}, \dots, \theta_{n-2,n}$ e $\omega_{1,4}, \dots, \omega_{n-3,n}$. Mas como dissemos acima, no PDGDM, temos apenas $\cos(\omega_{i-3,i})$, $i = 4, \dots, n$, que geram dois valores possíveis para cada ângulo de torção, já que $\omega_{i-3,i} \in [0, 2\pi]$. Entretanto, isso implica que não precisamos mais resolver nenhum sistema quadrático! E mais, as duas possibilidades para cada ângulo de torção podem ser previamente calculadas usando as informações de distâncias das cliques $\{v_{i-3}, v_{i-2}, v_{i-1}, v_i\}$. Para fazer as podas, com as distâncias adicionais, usamos as matrizes da Seção 2.5 para obter as coordenadas cartesianas das duas possíveis soluções para cada vértice v_i (fixadas as posições de $v_{i-3}, v_{i-2}, v_{i-1}$) e comparar as distâncias calculadas com a distância dada.

Resta uma questão: qual é o custo computacional de se encontrar a ordem do PDGDM? Diferentemente do custo polinomial da ordem do PDGD₃ (com a clique inicial dada *a priori*), encontrar uma ordem para o PDGDM pode ser complicado, pois é um problema NP-difícil [10], assim como o próprio PDGDM [38]. É o “preço” da nova informação!

Escapamos da resolução dos sistemas quadráticos, mas aumentamos exponencialmente o custo de se encontrar a nova ordem. Entretanto, dependendo da aplicação, essa ordem pode ser obtida usando as características do problema. É o que acontece, por exemplo, em problemas de cálculo de estruturas de proteínas [40].

Exercício 5.5. Por que há “troca de sinal” nas fórmulas $d_{i-3,i-2,i-1} = d_{i-3,i-2}^2 + d_{i-2,i-1}^2 - d_{i-3,i-1}^2$ e $d_{i-2,i-1,i} = d_{i-2,i-1}^2 + d_{i-2,i}^2 - d_{i-1,i}^2$?

Exercício 5.6. Tente deduzir a fórmula dada para $\cos(\omega_{i-3,i})$.

5.4 Simetrias do PDGDM

Vimos que a ordem do PDGDM permite visualizar o problema como uma “molécula artificial” e, usando coordenadas internas e as informações de distâncias das cliques $\{v_{i-3}, v_{i-2}, v_{i-1}, v_i\}$, $i = 4, \dots, n$, podemos também previamente calcular os dois valores possíveis para todos os ângulos de torção $\omega_{i-3,i}$.

Existe uma outra propriedade, ainda mais interessante, relacionada à simetria das soluções do problema. Esse é o tema da última seção do texto.

Em nosso problema teste (Seção 4.2), percebemos que as duas posições para v_4 ($x_{v_4}^0$ e $x_{v_4}^1$) podem ser consideradas, já que não há arestas adicionais para poder inviabilizar uma delas. Isso implica que, para qualquer solução encontrada na subárvore esquerda, tendo o nó $x_{v_4}^0$ como raiz, existe uma outra simétrica em relação ao plano definido por $x_{v_1}, x_{v_2}, x_{v_3}$ [38]. Note aqui a relação existente entre finitude do conjunto solução, desigualdade triangular estrita e simetrias.

Uma consequência imediata desse fato é que o conjunto solução tem uma quantidade par de soluções. Entretanto, desde os primeiros resultados computacionais obtidos para o PDGDM [45], a quantidade de soluções era sempre potência de dois. Apenas recentemente, usando Teoria de Grupos, conseguimos apresentar uma demonstração matemática para esse fato [49].

Podemos dar uma ideia da importância desse resultado, considerando o seguinte conjunto:

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

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

Inicialmente, vamos identificar quem são os elementos de S . O primeiro candidato é v_4 , que estará em S se não existir nenhuma aresta $\{u, w\} \in E$ tal que $u + 3 < v_4 \leq w$. Se existir algum $u \in V$ com tal característica, teremos $u < v_4 - 3$, o que é impossível, pois $v_4 - 3 = v_1$, que é o primeiro elemento de V . Ou seja, $v_4 \in S$, para qualquer PDGDM.

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

- Suponha que exista $\{u, v_5\} \in E$, tal que $u < v_5 - 3$. Temos que $v_5 \notin S$ e $\{v_1, v_5\} \in E$, o que implica que apenas uma das possibilidades para v_5 é viável: ou $x_{v_5}^0$ ou $x_{v_5}^1$.
- Supondo que não exista $\{u, v_5\} \in E$, para $u < v_5 - 3$, temos que considerar 2 casos:
 - Se não existe $\{u, w\} \in E$, tal que $u + 3 < v_5 < w$, então $v_5 \in S$.
 - Se existe $\{u, w\} \in E$, tal que $u + 3 < v_5 < w$, então $v_5 \notin S$.

Como o procedimento acima pode ser generalizado, podemos obter o conjunto S usando apenas os dados do PDGDM, ou seja, antes mesmo de aplicar o BP. Mas qual é a importância desse conjunto?

O conjunto S identifica outros **planos de simetria**, próprios de cada PDGDM (como vimos acima, o plano associado aos vértices $\{v_1, v_2, v_3\}$ define um plano de simetria válido para qualquer PDGDM).

Por exemplo, $v_5 \in S$ implica que as duas posições para v_5 são viáveis, $x_{v_5}^0$ e $x_{v_5}^1$. Ao mesmo tempo, como estamos admitindo que o PDGDM tem solução, $x_{v_5}^0$ e $x_{v_5}^1$ fazem parte de soluções do problema, pois não há aresta adicional para tornar tais posições inviáveis.

Considerando nosso problema teste, voltando a usar a notação v_i , temos ($K = 3$):

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

É fácil ver que $S = \{v_4\}$, já que $\{v_1, v_7\} \in E$. Ou seja, existe apenas o plano simétrico (definido por $x_{v_1}, x_{v_2}, x_{v_3}$) que é comum a qualquer PDGDM, implicando que temos apenas 2 soluções para o PDGDM. Como já obtemos uma solução (usando o BP), dada por

$$x_{v_1}, x_{v_2}, x_{v_3}, x_{v_4}^0, x_{v_5}^1, y_{v_6}^0, x_{v_7}^1,$$

temos uma outra, dada por

$$x_{v_1}, x_{v_2}, x_{v_3}, x_{v_4}^1, x_{v_5}^0, y_{v_6}^1, x_{v_7}^0.$$

Note que essa outra solução não precisa ser calculada pelo BP.

Supondo que $\{v_1, v_7\} \notin E$ e $\{v_3, v_7\} \notin E$, como ficaria o conjunto S ? Fazendo as contas, temos:

$$S = \{v_4, v_7\}.$$

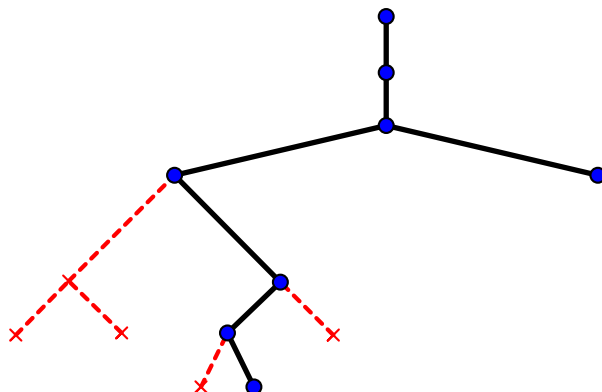


Figura 5.2: Solução obtida pelo **BP**.

Nesse caso, temos um outro plano simétrico, definido por $\{v_4, v_5, v_6\}$.

Vamos facilitar as coisas, representando a primeira solução por uma sequência de zeros e uns e denotando as três primeiras posições por $0, 0, 0$, ou seja,

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

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

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

Considerando agora a simetria no vértice v_4 , obtemos mais duas soluções, a partir das anteriores, dadas por

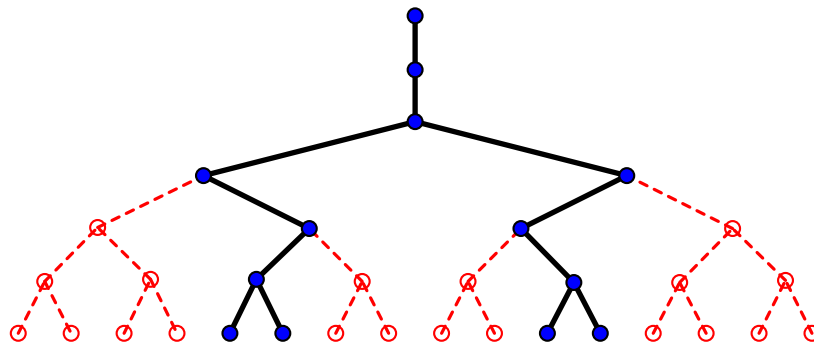


Figura 5.3: Soluções simétricas.

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

e

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

Dois conclusões importantes [44, 56]:

- Podemos saber, *a priori*, apenas usando os dados de qualquer PDGDM, qual é a cardinalidade do conjunto solução do problema, dada por $2^{|S|}$.
- Para encontrar todas as soluções de um PDGDM, basta usar o BP para encontrar uma única solução, pois todas as outras podem ser obtidas pelas simetrias do PDGDM.

Exercício 5.7. *Qual a importância computacional de se saber, a priori, a quantidade de soluções de um PDGDM?*

Exercício 5.8. *Como o custo computacional associado ao uso das propriedades de simetria para obter soluções de um PDGDM é polinomial, como fica a complexidade do PDGDM?*

Capítulo 6

Geometria de Distâncias e Geometria Molecular

6.1 O PDGDM e Estruturas de Proteínas

Atualmente, a aplicação de maior destaque da GD está relacionada à Geometria Molecular, de forma mais específica, ao problema de determinar estruturas tridimensionais de moléculas de proteínas, utilizando dados de ressonância magnética nuclear, que abreviaremos por RMN [68, 69]. Vale mencionar que o Prêmio Nobel de Química, de 2002, foi outorgado ao químico Kurt Wüthrich pelo desenvolvimento de aplicações de RMN à determinação de estruturas de proteínas. Juntamente com as informações da geometria das proteínas, os dados de RMN fornecem distâncias entre átomos próximos e o problema é como utilizar essa informação para determinar a posição de cada átomo da molécula de proteína (Figura 6.1). Ou seja, um problema de GD!



Figura 6.1: Máquina de RMN.

Mas qual é a importância da estrutura 3D de uma molécula? Essa estrutura está intimamente relacionada às suas propriedades físico-químicas. Ou seja, conhecer a estrutura tridimensional ajuda a entender a função da molécula. Um exemplo clássico que ilustra bem esse fato é a descoberta da estrutura 3D do DNA [67].

Em 1953, o físico Maurice Wilkins e a química Rosalind Franklin usaram a difração de raios X (outra técnica para a determinação de estrutura de proteínas [9]) para “fotografar” o DNA. O problema era formular um modelo tridimensional para

a molécula de DNA que se enquadrasse nos resultados da difração de raios X e explicasse algumas propriedades químicas já conhecidas. No mesmo ano, o bioquímico James Watson e o biofísico Francis Crick propuseram um modelo tridimensional – a famosa hélice dupla – que explicava todos os dados disponíveis sobre o DNA. O modelo imediatamente sugeriu o mecanismo de transmissão da informação genética (a característica essencial do modelo é a complementaridade das duas fitas retorcidas de DNA). Watson e Crick perceberam, antes da existência de dados que comprovassem o modelo, que a estrutura proposta poderia ser replicada pela separação das duas fitas e pela síntese de uma fita complementar para cada uma. Em 1958, o biólogo molecular Matthew Meselson e o geneticista Franklin Stahl demonstraram experimentalmente o modelo de replicação do DNA de Watson e Crick. Com o modelo e sua comprovação experimental, inaugurou-se uma revolução na compreensão do processo de hereditariedade. Pela descoberta da estrutura tridimensional da molécula de DNA, James Watson divide o Prêmio Nobel de Medicina com Francis Crick e Maurice Wilkins, em 1962.

Os genes de um organismo vivo (presentes no DNA) são, indiretamente, os responsáveis pelas características físicas desse organismo, mas as proteínas correspondentes é que determinam, de fato, essas características. No interior da célula, o DNA de um gene é transcrito para o RNA mensageiro e essa mensagem é então traduzida para formar a sequência de aminoácidos que dará origem a uma molécula de proteína (Figura 6.2). O processo de transcrição e tradução já é bem compreendido [64]. Entretanto, ainda há muito o que descobrir sobre o mecanismo de formação da molécula de proteína, a partir da sequência de aminoácidos fornecida pelo RNA mensageiro. Este processo é chamado de enovelamento da proteína e o problema associado é conhecido por “Protein Folding Problem” [15].

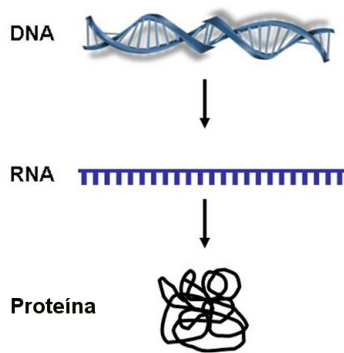


Figura 6.2: DNA e proteína.

Bom, já convencidos de que determinar a estrutura 3D de uma proteína é um problema importante, vejamos sua relação com o PDGDM.

Em 1984 e 1985 [29, 30], Havel e Wüthrich escreveram dois artigos mostrando como a GD pode ser utilizada no cálculo de estrutura de proteínas usando dados de RMN. Mas como dissemos na Introdução, apenas em 1988, com o lançamento do livro “Distance Geometry and Molecular Conformation” [13], Crippen e Havel estabelecem os fundamentos da conexão entre os dois temas de pesquisa. O algoritmo proposto, chamado EMBED, utiliza métodos de álgebra linear e de otimização para tentar resolver o PGD associado.

Nossa proposta é considerar o problema como um PDGDM. Para isso, é necessário definir uma ordem entre os átomos da proteína que induza uma ordem nos

vértices do grafo do PDGDM correspondente, dada por v_1, \dots, v_n . Ou seja, deve existir uma realização válida para v_1, v_2, v_3 e, para todo v_i , $i = 4, \dots, n$, devem existir 3 vértices imediatamente anteriores $v_{i-3}, v_{i-2}, v_{i-1}$, de tal maneira que os vértices $\{v_{i-3}, v_{i-2}, v_{i-1}, v_i\}$ formem uma clique e

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

Esse é o tema da próxima seção.

6.2 Ordens nas Proteínas

Dissemos acima que, juntamente com as informações da geometria das proteínas, os dados de RMN fornecem distâncias entre átomos próximos e o problema é como utilizar essa informação para determinar a posição de cada átomo da molécula de proteína.

Os dados de RMN estão relacionados a átomos próximos, cerca de 5 angstroms (Å), e a informação sobre a geometria das proteínas nos diz que, ao tentarmos determinar a estrutura 3D de uma dada proteína, as distâncias entre átomos ligados covalentemente e os ângulos planos definidos por três átomos ligados consecutivamente são conhecidos, *a priori*. É claro, a proteína não é uma estrutura rígida, mas esses valores podem ser considerados fixos [32, 22].

Isso sugere uma ordem natural entre os átomos da cadeia principal da proteína, formada por uma sequência de 3 átomos que se repetem: N, C $_{\alpha}$, C (Figura 6.3). A cadeia principal é o “esqueleto” da proteína, o que já nos dá uma boa idéia de sua estrutura tridimensional. Nesse texto, iremos nos restringir, essencialmente, à cadeia principal. Em [12, 62], temos propostas para considerar também as cadeias laterais (Figura 6.3), que fazem a distinção entre os 20 tipos de aminoácidos que formam as proteínas [16].

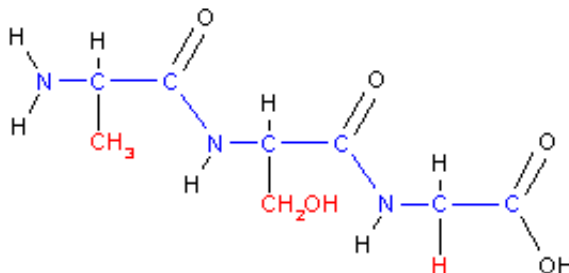


Figura 6.3: Cadeia principal e cadeias laterais de uma proteína.

Como as distâncias entre os átomos i e $i + 3$, na cadeia principal, são menores do que 5Å (em geral), podemos supor que elas são detectadas pelos experimentos de RMN, o que nos daria a ordem procurada. Entretanto, na maior parte dos experimentos, os dados de RMN estão associados a pares de átomos de hidrogênio [68]. Uma saída, seria definir uma ordem envolvendo apenas átomos de hidrogênio.

Mas que ordem considerar, já que a ordem “natural” entre os hidrogênios não satisfaz as condições do PDGDM? Uma possibilidade é incorporar átomos de hidrogênio das cadeias laterais e, além disso, permitir repetições na ordem (Figura 6.4).

Quimicamente, não faz sentido considerar dois átomos na mesma posição, mas podemos fazer isso na ordem dos vértices do grafo associado (a representação de

uma molécula por meio de grafos é antiga [65]). Essa repetição vai garantir que as distâncias $d_{i-3,i}$ sejam conhecidas, podendo até serem nulas. Do ponto de vista computacional, isso tem uma vantagem, pois ao calcularmos novamente a posição de um dado átomo, podemos verificar se os erros numéricos estão sob controle [41].

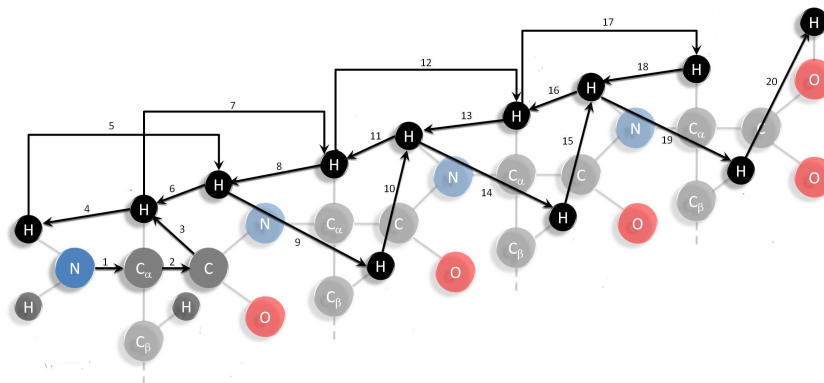


Figura 6.4: Ordem entre átomos de hidrogênio.

Exercício 6.1. Verifique na Figura 6.4 em que pares de átomos a repetição acontece.

Exercício 6.2. O que aconteceria se alguma distância $d_{i-1,i}$ ou $d_{i-2,i}$ fosse nula?

Supondo que, aplicando o algoritmo BP, encontremos a posição dos átomos de hidrogênio ligados à cadeia principal, como determinar a posição dos átomos dessa cadeia, que é o que nos interessa? No Capítulo 4, vimos que a interseção de 4 esferas, sob certas condições, resulta em apenas 1 ponto. Deixamos a resposta para os dois próximos exercícios.

Exercício 6.3. Nas três situações da Figura 6.5, determine os sistemas quadráticos correspondentes às interseções de 4 esferas.

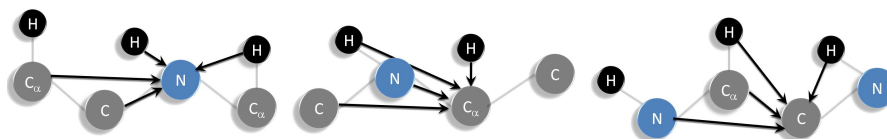


Figura 6.5: Determinando a cadeia principal, usando as posições dos átomos de hidrogênio.

Exercício 6.4. Mostre que, para cada sistema, existe uma única solução.

Até o momento, mencionamos três aspectos importantes do problema que estamos tentando resolver:

1. Distâncias conhecidas apenas entre átomos próximos,
2. Distâncias conhecidas apenas entre átomos de hidrogênio,

3. Necessidade de resolver dois subproblemas: cálculo da posição dos átomos de hidrogênio e dos átomos da cadeia principal.

Na verdade, existe um outro problema, ainda mais complicado, que ainda não temos uma resposta definitiva. Para também deixá-lo destacado, enunciamos a seguir:

- As distâncias provenientes dos dados de RMN, entre átomos de hidrogênio próximos, não são valores precisos.

Essa informação torna o problema muito mais complicado! Alguns resultados preliminares podem ser encontrados em [40, 54]. Lembramos que todos os resultados que apresentamos nesse texto são baseados no fato de que as distâncias envolvidas são números reais, livres de qualquer erro.

Entretanto, sabemos que qualquer medição experimental está sujeita a erros. Ou seja, podemos ter um valor $d_{i,j}$ que não corresponde à distância correta entre os átomos i e j . Numa abordagem inicial, podemos considerar, então, que os dados provenientes dos experimentos de RMN são intervalos de números reais que contêm a distância correta. Mesmo com essa hipótese, que ainda é uma aproximação da realidade, o problema não é trivial.

A boa notícia é que esse novo problema nos deu uma ideia para resolver o problema 3 acima. Podemos inventar uma nova ordem, com duas características principais:

- Consideramos, ao mesmo tempo, átomos de hidrogênio e átomos da cadeia principal;
- Na clique $\{v_{i-3}, v_{i-2}, v_{i-1}, v_i\}$ do grafo do PDGDM associado, para $i = 4, \dots, n$, todas as distâncias $d_{i-1,i}$ e $d_{i-2,i}$ são números reais e apenas as distâncias $d_{i-3,i}$ podem ser “intervalares”.

Exercício 6.5. Baseado na Figura 6.6, verifique que as distâncias $d_{i-1,i}$ e $d_{i-2,i}$ são números reais.

Exercício 6.6. Baseado na Figura 6.6, verifique que as distâncias $d_{i-3,i}$ podem ser intervalos degenerados.

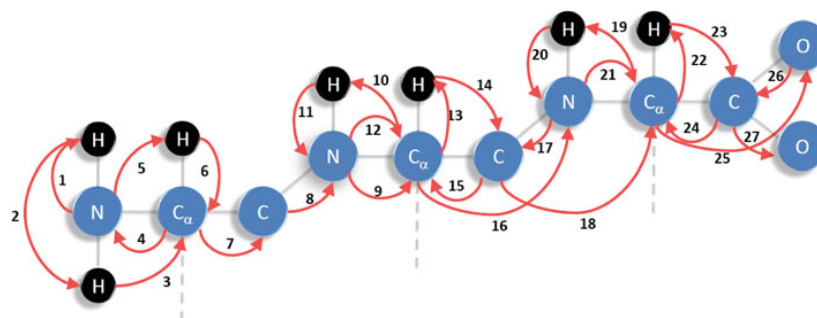


Figura 6.6: Ordem envolvendo átomos de hidrogênio e a cadeia principal.

Exercício 6.7. Como ficaria a árvore de busca do BP, considerando as distâncias $d_{i-3,i}$ intervalares?

Exercício 6.8. Que modificações são necessárias no BP para considerar distâncias intervalares?

6.3 O BP com Custo Polinomial

Já sabemos que, se numa dada instância do PDGDM, para todo $i = 5, \dots, n$, existir uma aresta adicional $\{v_j, v_i\} \in E$ com $j < i - 3$, de tal maneira que o conjunto $\{v_j, v_{i-3}, v_{i-2}, v_{i-1}\}$ gere um conjunto de pontos não-coplanares, haverá uma única realização do grafo associado determinada em tempo linear, ou seja, com um custo computacional proporcional a n .

Em problemas relacionados ao cálculo de estrutura de proteínas, considerando distâncias exatas, verificamos que, em várias instâncias, o desempenho do BP é polinomial [47]. Em parte, isso pode ser explicado pelo fato da cadeia principal de muitas proteínas estar enovelada (Figura 6.2). Quanto mais esticada, menor será a cardinalidade do conjunto das arestas de poda, fazendo com que a árvore do BP se ramifique cada vez mais.

Para se ter uma ideia do “comportamento global” do algoritmo, precisamos pensar na árvore do BP como um todo e não como ela é parcialmente construída em cada passo do algoritmo.

Quando o conjunto das arestas de poda é vazio, $E_p = \emptyset$, a árvore é cheia, representando todo o espaço de busca. Não há dificuldade em encontrar uma solução (em tempo linear), pois basta descer pela árvore, escolhendo qualquer uma das duas possibilidades, em cada passo do BP. Como $E_p = \emptyset$, não há como errar na hora de fazer escolhas. Claro, é impensável encontrar todas as soluções para n grande, já que o conjunto solução tem cardinalidade 2^{n-3} (Figura 6.7).

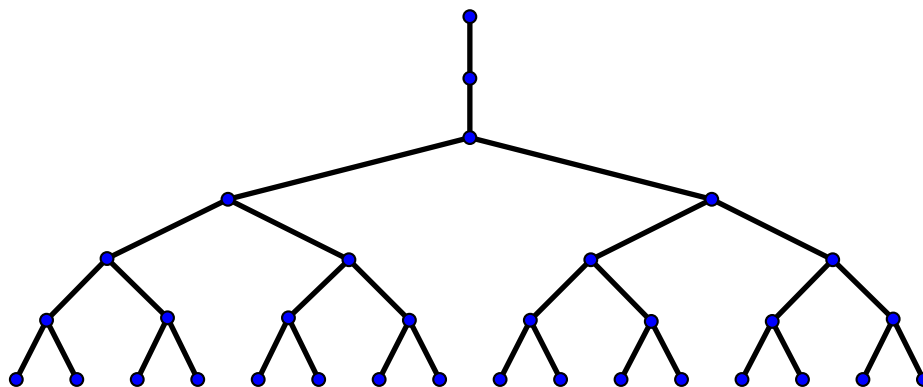


Figura 6.7: Árvore do **BP** com $E_p = \emptyset$.

Supondo a situação descrita no primeiro parágrafo desta seção, ou seja, para todo $i = 5, \dots, n$, existe uma aresta $\{v_j, v_i\} \in E_p$ com $j < i - 3$, temos solução única, encontrada em tempo linear. Nesse caso, sabemos qual é a decisão correta a ser tomada em cada passo do algoritmo (Figura 6.8).

O que aumenta o custo computacional do BP são os retrocessos necessários quando nenhuma das duas posições calculadas, para um dado vértice v , é compatível com as arestas $\{u, v\} \in E_p$, para $u < v - 3$. Isto é, em algum momento, ou melhor, em algum nível anterior da árvore, foi tomada uma decisão equivocada. Isso acontece, precisamente, quando a situação do parágrafo anterior não é atendida. Como essa questão é a causa do retrocesso, merece destaque. Ou seja, a razão matemática para que o BP seja obrigado a fazer retornos na árvore de busca, impedindo que ele “desça livremente”, é a seguinte:

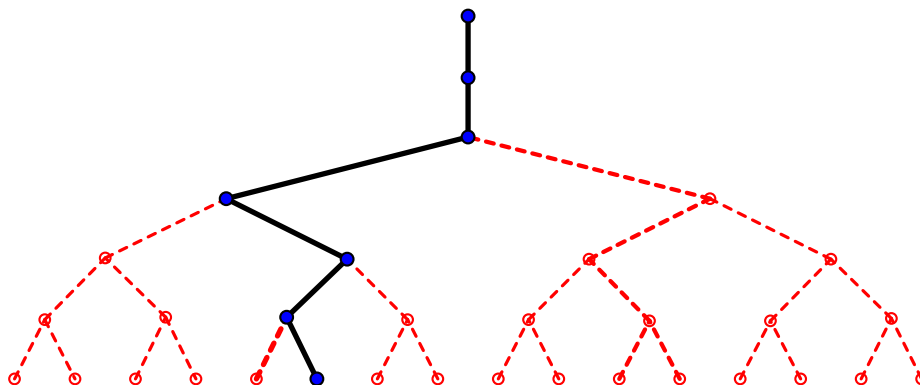


Figura 6.8: Solução única, encontrada em tempo linear.

- Existe (pelo menos) um vértice v_j na ordem dada do PDGDM, $v_1, \dots, v_j, \dots, v_n$, cujos únicos vértices anteriores u , tais que $\{u, v_j\} \in E$, são aqueles utilizados na construção da árvore: $v_{j-3}, v_{j-2}, v_{j-1}$.

Isso significa que, sempre que tal situação ocorre, o nível j da árvore passa a ter o dobro de nós do nível anterior. O problema se agrava ainda mais quando existe um conjunto de vértices consecutivos v_j, \dots, v_{j+k} , com $k \geq 4$, com todos eles na situação destacada acima, expandindo o espaço de busca rapidamente. Imagine, por exemplo, que no nível $j = 50$ da árvore existam $2^{20} = 1.048.576$ posições que satisfazem todas as informações de poda até o momento. Com $k = 5$, a quantidade de soluções possíveis, até então, passaria para $2^{25} = 33.554.432!$

Outra questão importante está relacionada à “quantificação” do erro cometido. Essa é uma questão delicada... Para iniciar sua compreensão, temos que utilizar um resultado sobre **dualidade** do PDGDM [1, 43], assunto não tratado nesse texto:

- Se na ordem v_1, \dots, v_n do PDGDM existe um conjunto de vértices consecutivos v_j, \dots, v_{j+k} , com $k \geq 4$, onde a única aresta $\{u, v\} \in E_p$ é dada por $u = v_j$ e $v = v_{j+k}$, então, na subárvore com raiz em qualquer nó viável do nível j até o nível $j + k$, existe apenas uma única solução (parcial) do problema.

O próximo exercício, o último de nossa pequena aventura pelo mundo da Geometria de Distâncias, é “diferente” de todos os outros.

Exercício 6.9. *Como esse resultado pode ser utilizado para quantificar o (possível) erro cometido no nível j , notado apenas no nível $j + k$?*

A eficiência do BP, quando aplicado em problemas de cálculo de estrutura de proteínas, está relacionada a dois fatores (supondo distâncias precisas):

- Existem poucos vértices sem informação de poda, onde acontece a duplicação de nós;
- Quando isso acontece, digamos, no vértice v_j , o intervalo onde a situação se mantém, entre v_j e v_{j+k} , é curto. Ou seja, o valor de k é pequeno.

Antes da conclusão propriamente dita, mencionamos que o custo computacional do BP pode ser reduzido, essencialmente, de duas maneiras:

- Usando técnicas de paralelização do algoritmo [25, 57],
- Por meio do conceito de múltiplas árvores [20, 60].

Capítulo 7

Conclusão

Chegamos ao final, mas na verdade, gostaríamos que seja apenas o início de um estudo mais aprofundado em Geometria de Distâncias.

Nosso intuito foi apresentar a Geometria de Distâncias como base para trabalhos de pesquisa ou material de ensino, onde a ideia principal foi mostrar que o mundo da Matemática se torna ainda mais apaixonante quando fazemos a integração entre vários conceitos, motivados por um problema real e desafiador. Ou seja, a mistura é a regra, e não a exceção.

Várias estruturas fundamentais da Matemática, como linguagem, foram consideradas, onde destacamos:

- A necessidade de uma boa definição para tornar uma ideia precisa,
- A compreensão da importância da escolha das hipóteses que devem ser incluídas em um resultado matemático,
- A análise das consequências de um teorema.

Explorando o assunto, vários temas de diferentes ramos da Matemática e da Computação (incluindo Geometria, Álgebra, Combinatória, Estrutura de Dados e Complexidade de Algoritmos) foram apresentados, tais como: dimensão, métrica, simetria, grafo, aproximação numérica, solubilidade de problemas e custo computacional.

Um tópico fundamental, principalmente nas aplicações, que foi apenas levemente considerado no Capítulo 6 está relacionado à **incerteza**. Esse capítulo é diferente dos demais, pois foi dedicado à Geometria Molecular, considerada de grande destaque na Geometria de Distâncias. Nele, procuramos empregar (às vezes, de maneira implícita) os conceitos e resultados apresentados em capítulos anteriores, bem como sensibilizar o leitor de que resolver um problema complexo requer disposição e dedicação para entender as “sutilezas” das relações existentes entre a Matemática e o mundo real. Como mencionamos, é uma das áreas atuais de pesquisa em que estamos trabalhando e necessitando de colaboradores, principalmente no que diz respeito ao tratamento das incertezas envolvidas.

Como pudemos perceber, o tema é bastante rico, envolvendo questões teóricas e computacionais, bem como diversas aplicações. Para prosseguir, além das várias referências bibliográficas citadas, mencionamos também a preparação de um novo livro, em inglês, onde generalizamos os conceitos e resultados aqui apresentados, bem como consideramos outros tópicos mais avançados.

Até o próximo encontro!

Bibliografia

- [1] ABUD, G.; ALENCAR, J. Counting the number of solutions of the discretizable molecular distance geometry problem. In: ANDRIONI, A. et al. (Ed.). *Proceedings of the Workshop on Distance Geometry and Applications*. Universidade Federal do Amazonas, Manaus: [s.n.], 2013. p. 29–32.
- [2] ALVES, R. et al. Adaptive branching in ibp with clifford algebra. In: ANDRIONI, A. et al. (Ed.). *Proceedings of the Workshop on Distance Geometry and Applications*. Universidade Federal do Amazonas, Manaus: [s.n.], 2013. p. 65–69.
- [3] ANDRIONI, A. A clifford algebra approach to the discretizable molecular distance geometry problem. In: ANDRIONI, A. et al. (Ed.). *Proceedings of the Workshop on Distance Geometry and Applications*. Universidade Federal do Amazonas, Manaus: [s.n.], 2013. p. 71–75.
- [4] ANDRIONI, A. et al. *Proceedings of the Workshop on Distance Geometry and Applications*. Manaus: Universidade Federal do Amazonas, 2013.
- [5] BAJAJ, C. The algebraic degree of geometric optimization problems. *Discrete and Computational Geometry*, v. 3, p. 177–191, 1988.
- [6] BENEDETTI, R.; RISLER, J.-J. *Real Algebraic and Semi-algebraic Sets*. Paris: Hermann, 1990.
- [7] BERGER, B.; KLEINBERG, J.; LEIGHTON, T. Reconstructing a three-dimensional model with arbitrary errors. *Journal of the ACM*, v. 46, p. 212–235, 1999.
- [8] BLUMENTHAL, L. *Theory and Applications of Distance Geometry*. Oxford: Oxford University Press, 1953.
- [9] BRÜNGER, A.; NILGES, M. Computational challenges for macromolecular structure determination by x-ray crystallography and solution nmr-spectroscopy. *Quarterly Reviews of Biophysics*, v. 66, p. 49–125, 1993.
- [10] CASSIOLI, A. et al. Discretization vertex orders in distance geometry. *IBM Research Report RC25434*, 2013.
- [11] COOPE, I. Reliable computation of the points of intersection of n spheres in \mathbb{R}^n . *Australian and New Zealand Industrial and Applied Mathematics Journal*, v. 42, p. C461–C477, 2000.
- [12] COSTA, V. et al. Discretization orders for protein side chains. *Journal of Global Optimization*, Springer, a ser publicado.

- [13] CRIPPEN, G. M.; HAVEL, T. F. *Distance geometry and molecular conformation*. [S.l.]: Research Studies Press Taunton, UK, 1988.
- [14] DEZA, M. M.; DEZA, E. *Encyclopedia of distances*. [S.l.]: Springer, 2009.
- [15] DILL, K. A.; MACCALLUM, J. L. The protein-folding problem, 50 years on. *Science*, American Association for the Advancement of Science, v. 338, n. 6110, p. 1042–1046, 2012.
- [16] DONALD, B. R. *Algorithms in structural molecular biology*. [S.l.]: The MIT Press, 2011.
- [17] DONG, Q.; WU, Z. A linear-time algorithm for solving the molecular distance geometry problem with exact inter-atomic distances. *Journal of Global Optimization*, Springer, v. 22, n. 1-4, p. 365–375, 2002.
- [18] DZEMYDA, G.; KURASOVA, O.; ŽILINSKAS, J. *Multidimensional Data Visualization: Methods and Applications*. [S.l.]: Springer, 2013.
- [19] EMIRIS, I. Z.; MOURRAIN, B. Computer algebra methods for studying and computing molecular conformations. *Algorithmica*, Springer, v. 25, n. 2-3, p. 372–402, 1999.
- [20] FIDALGO, F.; RODRIGUEZ, J. Quaternions as a tool for merging multiple realization trees. In: ANDRIONI, A. et al. (Ed.). *Proceedings of the Workshop on Distance Geometry and Applications*. Universidade Federal do Amazonas, Manaus: [s.n.], 2013. p. 119–124.
- [21] FLOUDAS, C. A.; GOUNARIS, C. E. A review of recent advances in global optimization. *Journal of Global Optimization*, Kluwer Academic Publishers, v. 45, n. 1, p. 3–38, 2009.
- [22] GIBSON, K. D.; SCHERAGA, H. A. Energy minimization of rigid-geometry polypeptides with exactly closed disulfide loops. *Journal of Computational Chemistry*, Wiley Online Library, v. 18, n. 3, p. 403–415, 1997.
- [23] GONÇALVES, D. S.; MUCHERINO, A. Discretization orders and efficient computation of cartesian coordinates for distance geometry. *Optimization Letters*, Springer, a ser publicado.
- [24] GRAMACHO, W. et al. A new algorithm to finding discretizable orderings for distance geometry. In: ANDRIONI, A. et al. (Ed.). *Proceedings of the Workshop on Distance Geometry and Applications*. Universidade Federal do Amazonas, Manaus: [s.n.], 2013. p. 149–152.
- [25] GRAMACHO, W. et al. A parallel bp algorithm for the discretizable distance geometry problem. In: *Proceedings of the IEEE Workshop on Parallel Computing and Optimization*. Shangai: [s.n.], 2012. p. 1756–1762.
- [26] GRAVER, J. E.; SERVATIUS, B.; SERVATIUS, H. *Combinatorial rigidity*. [S.l.]: American Mathematical Soc., 1993.
- [27] GROVER, L. K. Quantum mechanics helps in searching for a needle in a haystack. *Physical Review Letters*, APS, v. 79, n. 2, p. 325–328, 1997.
- [28] HARARY, F. *Graph theory. 1994*. [S.l.]: Addison-Wesley, Reading, MA.

- [29] HAVEL, T.; WÜTHRICH, K. A distance geometry program for determining the structures of small proteins and other macromolecules from nuclear magnetic resonance measurements of intramolecular $^1\text{H}-^1\text{H}$ proximities in solution. *Bulletin of Mathematical Biology*, Elsevier, v. 46, n. 4, p. 673–698, 1984.
- [30] HAVEL, T. F.; WÜTHRICH, K. An evaluation of the combined use of nuclear magnetic resonance and distance geometry for the determination of protein conformations in solution. *Journal of Molecular Biology*, Elsevier, v. 182, n. 2, p. 281–294, 1985.
- [31] HUNT, K. Structural kinematics of in-parallel-actuated robot-arms. *Journal of Mechanical Design*, American Society of Mechanical Engineers, v. 105, n. 4, p. 705–712, 1983.
- [32] JACKSON, B.; JORDÁN, T. On the rigidity of molecular graphs. *Combinatorica*, Springer, v. 28, n. 6, p. 645–658, 2008.
- [33] JUHÁS, P. et al. Ab initio determination of solid-state nanostructure. *Nature*, v. 440, n. 7084, p. 655–658, 2006.
- [34] LAVOR, C. On generating instances for the molecular distance geometry problem. In: *Global Optimization: From Theory to Implementation*. [S.l.]: Springer, 2006. p. 405–414.
- [35] LAVOR, C. et al. Discretization orders for distance geometry problems. *Optimization Letters*, Springer, v. 6, n. 4, p. 783–796, 2012.
- [36] LAVOR, C.; LIBERTI, L.; MACULAN, N. Grover’s algorithm applied to the molecular distance geometry problem. In: *Proceedings of the 7th Brazilian Congress of Neural Networks*. Natal: [s.n.], 2005.
- [37] LAVOR, C.; LIBERTI, L.; MACULAN, N. A note on a “branch-and-prune” algorithm for the molecular distance geometry problem. *International Transactions in Operational Research*, Wiley Online Library, v. 18, n. 6, p. 751–752, 2011.
- [38] LAVOR, C. et al. The discretizable molecular distance geometry problem. *Computational Optimization and Applications*, Springer, v. 52, n. 1, p. 115–146, 2012.
- [39] LAVOR, C. et al. Recent advances on the discretizable molecular distance geometry problem. *European Journal of Operational Research*, Elsevier, v. 219, n. 3, p. 698–706, 2012.
- [40] LAVOR, C.; LIBERTI, L.; MUCHERINO, A. The interval branch-and-prune algorithm for the discretizable molecular distance geometry problem with inexact distances. *Journal of Global Optimization*, Springer, v. 56, n. 3, p. 855–871, 2013.
- [41] LAVOR, C. et al. On the computation of protein backbones by using artificial backbones of hydrogens. *Journal of Global Optimization*, Springer, v. 50, n. 2, p. 329–344, 2011.
- [42] LEE, J. A.; VERLEYSEN, M. *Nonlinear dimensionality reduction*. [S.l.]: Springer, 2010.
- [43] LIBERTI, L.; LAVOR, C. On a relationship between graph realizability and distance matrix completion. In: *Optimization Theory, Decision Making, and Operations Research Applications*. [S.l.]: Springer, 2013. p. 39–48.

- [44] LIBERTI, L. et al. Counting the number of solutions of kdmdgp instances. In: *Geometric Science of Information*. [S.l.]: Springer, 2013. p. 224–230.
- [45] LIBERTI, L.; LAVOR, C.; MACULAN, N. A branch-and-prune algorithm for the molecular distance geometry problem. *International Transactions in Operational Research*, Wiley Online Library, v. 15, n. 1, p. 1–17, 2008.
- [46] LIBERTI, L. et al. Euclidean distance geometry and applications. *SIAM Review*, SIAM, v. 56, n. 1, p. 3–69, 2014.
- [47] LIBERTI, L.; LAVOR, C.; MUCHERINO, A. The discretizable molecular distance geometry problem seems easier on proteins. In: *Distance Geometry: Theory, Methods and Applications*. [S.l.]: Springer, 2013. p. 47–60.
- [48] LIBERTI, L. et al. Molecular distance geometry methods: from continuous to discrete. *International Transactions in Operational Research*, Wiley Online Library, v. 18, n. 1, p. 33–51, 2010.
- [49] LIBERTI, L. et al. On the number of realizations of certain henneberg graphs arising in protein conformation. *Discrete Applied Mathematics*, Elsevier, v. 165, p. 213–232, 2014.
- [50] LIMA, R. S.; MARTÍNEZ, J. Solving molecular distance geometry problems using a continuous optimization approach. In: *Distance Geometry: Theory, Methods and Applications*. [S.l.]: Springer, 2013. p. 213–224.
- [51] LINDEGREN, L. et al. The astrometric core solution for the gaia mission. overview of models, algorithms, and software implementation. *Astronomy and Astrophysics*, v. 538, p. 1–47, 2012.
- [52] MAYER-SCHÖNBERGER, V.; CUKIER, K. *Big data: A revolution that will transform how we live, work, and think*. [S.l.]: Houghton Mifflin Harcourt, 2013.
- [53] MENGER, K. Untersuchungen über allgemeine metrik. *Mathematische Annalen*, Springer, v. 100, n. 1, p. 75–163, 1928.
- [54] MUCHERINO, A. On the identification of discretization orders for distance geometry with intervals. In: *Geometric Science of Information*. [S.l.]: Springer, 2013. p. 231–238.
- [55] MUCHERINO, A.; LAVOR, C.; LIBERTI, L. The discretizable distance geometry problem. *Optimization Letters*, Springer, v. 6, n. 8, p. 1671–1686, 2012.
- [56] MUCHERINO, A.; LAVOR, C.; LIBERTI, L. Exploiting symmetry properties of the discretizable molecular distance geometry problem. *Journal of Bioinformatics and Computational Biology*, World Scientific, v. 10, n. 03, 2012.
- [57] MUCHERINO, A. et al. A parallel version of the branch & prune algorithm for the molecular distance geometry problem. In: *Proceedings of the ACS/IEEE International Conference on Computer Systems and Applications*. Hammamet: [s.n.], 2010. p. 1–6.
- [58] MUCHERINO, A. et al. *Distance geometry: theory, methods, and applications*. [S.l.]: Springer, 2013.

- [59] NIELSEN, J.; ROTH, B. On the kinematic analysis of robotic mechanisms. *The International Journal of Robotics Research*, SAGE Publications, v. 18, n. 12, p. 1147–1160, 1999.
- [60] NUCCI, P.; NOGUEIRA, L. T.; LAVOR, C. Solving the discretizable molecular distance geometry problem by multiple realization trees. In: *Distance Geometry*. [S.l.]: Springer, 2013. p. 161–176.
- [61] PORTUGAL, R. et al. Uma introdução à computação quântica. Sociedade Brasileira de Matemática Aplicada e Computacional (SBMAC), 2012.
- [62] SALLAUME, S. et al. A discrete search algorithm for finding the structure of protein backbones and side chains. *International Journal of Bioinformatics Research and Applications*, Inderscience, v. 9, n. 3, p. 261–270, 2013.
- [63] SAXE, J. Embeddability of weighted graphs in k-space is strongly np-hard. In: *Proceedings of the 17th Allerton Conference in Communications, Control and Computing*. [S.l.: s.n.], 1979. p. 480–489.
- [64] SCHLICK, T. *Molecular Modeling and Simulation: An Interdisciplinary Guide*. [S.l.]: Springer, 2002.
- [65] SYLVESTER, J. J. Chemistry and algebra. *Nature*, v. 17, p. 284, 1877.
- [66] THOMPSON, H. B. Calculation of cartesian coordinates and their derivatives from internal molecular coordinates. *The Journal of Chemical Physics*, AIP Publishing, v. 47, n. 9, p. 3407–3410, 1967.
- [67] WATSON, J. D.; CRICK, F. H. Molecular structure of nucleic acids. *Nature*, v. 171, n. 4356, p. 737–738, 1953.
- [68] WÜTHRICH, K. Protein structure determination in solution by nuclear magnetic resonance spectroscopy. *Science*, v. 243, n. 4887, p. 45–50, 1989.
- [69] WÜTHRICH, K. The way to nmr structures of proteins. *Nature Structural & Molecular Biology*, Nature, v. 8, n. 11, p. 923–925, 2001.
- [70] YEMINI, Y. The positioning problem - a draft of an intermediate summary. In: *Proceedings of the Conference on Distributed Sensor Networks*. [S.l.: s.n.], 1978. p. 137–145.