

UNICAMP

UNIVERSIDADE ESTADUAL DE
CAMPINAS

Instituto de Matemática, Estatística e
Computação Científica

RODRIGO LABIAK

Geometria de Distâncias e o Problema do Loop Fechado

Campinas

2021

Rodrigo Labiak

Geometria de Distâncias e o Problema do Loop Fechado

Tese apresentada ao Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Doutor em Matemática.

Orientador: Plamen Emilov Kochloukov

Coorientador: Carlile Campos Lavor

Este exemplar corresponde à versão final da Tese defendida pelo aluno Rodrigo Labiak e orientada pelo Prof. Dr. Plamen Emilov Kochloukov.

Campinas

2021

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Ana Regina Machado - CRB 8/5467

L113g Labiak, Rodrigo, 1993-
Geometria de distâncias e o problema do loop fechado / Rodrigo Labiak. –
Campinas, SP : [s.n.], 2021.

Orientador: Plamen Emilov Kochloukov.
Coorientador: Carlile Campos Lavor.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de
Matemática, Estatística e Computação Científica.

1. Geometria de distâncias. 2. Loop fechado. 3. Algoritmos branch-and-prune. I. Kochloukov, Plamen Emilov, 1958-. II. Lavor, Carlile Campos, 1968-. III. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: Distance geometry and the loop closure problem

Palavras-chave em inglês:

Distance geometry

Loop closure

Branch-and-prune algorithms

Área de concentração: Matemática

Titulação: Doutor em Matemática

Banca examinadora:

Carlile Campos Lavor [Coorientador]

Marcos Eduardo Ribeiro do Valle Mesquita

Sueli Irene Rodrigues Costa

Douglas Soares Gonçalves

Michael Ferreira de Souza

Data de defesa: 16-03-2021

Programa de Pós-Graduação: Matemática

Identificação e informações acadêmicas do(a) aluno(a)

- ORCID do autor: <https://orcid.org/0000-0002-8811-383X>

- Currículo Lattes do autor: <http://lattes.cnpq.br/1264136564060763>

**Tese de Doutorado defendida em 16 de março de 2021 e aprovada
pela banca examinadora composta pelos Profs. Drs.**

Prof(a). Dr(a). CARLILE CAMPOS LAVOR

Prof(a). Dr(a). MARCOS EDUARDO RIBEIRO DO VALLE MESQUITA

Prof(a). Dr(a). SUELI IRENE RODRIGUES COSTA

Prof(a). Dr(a). DOUGLAS SOARES GONÇALVES

Prof(a). Dr(a). MICHAEL FERREIRA DE SOUZA

A Ata da Defesa, assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria de Pós-Graduação do Instituto de Matemática, Estatística e Computação Científica.

Este trabalho é dedicado os meus pais.

Agradecimentos

Agradeço inicialmente a Deus, por ter sido meu refúgio seguro nos momentos de angústia e incertezas.

Aos professores Carlile Campos Lavor, Fernando Eduardo Torres Orihuela e Plamen Emilov Kochloukov por terem me orientado neste trabalho e por todas as oportunidades que me proporcionaram durante esses quatro anos.

À toda minha família pelo apoio, que mesmo com a distância se fez presente em todos os momentos. Em especial à minha mãe, meu pai, meu irmão, minhas avós e meus avôs.

À minha namorada, Daniele, que me apoiou e deu forças para não desistir no caminho.

Aos amigos que fiz ao longo do mestrado e doutorado.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq, cujo apoio financeiro tornou possível a realização deste trabalho. Nº Processo: 155543/2019-9.

Resumo

O principal problema de Geometria de Distâncias (GD) é encontrar as posições de um conjunto de pontos, considerando conhecidas somente algumas distâncias entre eles. Utilizando resultados de GD, como simetrias e a discretização do espaço de busca, encontraremos soluções para uma classe de problemas denominada Problema do Loop Fechado (PLF). Esses problemas possuem diversas aplicações, destacando-se a Geometria Molecular. Basicamente, os problemas consistem em determinar a posição dos átomos em segmentos de molécula, de forma que os segmentos estejam geometricamente consistentes com toda a estrutura da molécula. Apresentamos uma nova abordagem para resolver o PLF a qual envolve a construção de uma ordem para os átomos do PLF que satisfaça as hipótese do Problema Intervalar de Geometria de Distâncias em Dimensão 3 (*iPDGD*₃), permitindo a utilização do Algoritmo Interval Branch and Prune (iBP) para encontrar as soluções do PLF. Por fim, apresentamos resultados numéricos de testes preliminares que apontam um bom desempenho do iBP em instâncias do PLF.

Palavras-chave: Geometria de Distâncias. Loop Fechado. Branch and Prune.

Abstract

The main problem of Distance Geometry (DG) is to find the positions of a set of points, considering only some known distances between them. Using results of DG, such as symmetries and the search space discretization, we will find solutions to a class of problems called Loop-Closure Problem. This problem has many applications, an important one being Molecular Geometry. Basically, the problems consist in determining the position of the atoms in molecule segments, so that the segments are geometrically consistent with the entire structure of the molecule. We present a new approach to solve the PLF which involves the construction of an order for the atoms of the PLF that satisfies the hypothesis of the Interval Distance Geometry Problem in Dimension 3 ($iPDGD_3$), allowing the use of the Interval Branch and Prune Algorithm (iBP) to find PLF solutions. Finally, we present numerical results of preliminary tests that indicate a good performance of the iBP in instances of the PLF.

Keywords: Distance Geometry. Loop Closure. Branch and Prune.

Lista de ilustrações

Figura 1 – Interseção de Esferas.	16
Figura 2 – Interseção de duas esferas e um casca esférica.	21
Figura 3 – Ilustração da utilização de distância intervalar para realizar podas na árvore de busca.	24
Figura 4 – Ilustração do problema canônico do loop fechado onde os três corpos rígidos são dados por: $(C_{\alpha 3}, C_3, N_1, C_{\alpha 1}), (C_{\alpha 1}, C_1, N_2, C_{\alpha 2})$ e $(C_{\alpha 2}, C_2, N_3, C_{\alpha 3})$	26
Figura 5 – Ilustração dos vetores z_i, r_i^T, r_i^σ	28
Figura 6 – Ilustração dos ângulos α_i, η_i, ξ_i	29
Figura 7 – Relação entre os ângulos τ_i e δ_i	30
Figura 8 – Relação do ângulo de restrição θ com \hat{r}_i^T e \hat{r}_{i-1}^σ	30
Figura 9 – Ilustração do Loop Fechado Generalizado.	43
Figura 10 – Ilustração dos ângulos ϕ e ψ em um resíduo.	43
Figura 11 – Ilustração da avaliação dos loops encontrados pelo algoritmo CCD.	45
Figura 12 – Ilustração das conformações de loops válidos e inválidos encontrados pelo algoritmo CCD.	47
Figura 13 – Ilustração do problema do loop fechado.	49
Figura 14 – Distâncias máxima e mínima de $C_{\alpha 2}$ até N_1	51
Figura 15 – Ilustração do lado direito da árvore de busca do Algoritmo iBP para uma amostra de distância intervalar.	52
Figura 16 – Ilustração da generalização do Problema do Loop Fechado.	54
Figura 17 – Ilustração da estimativa de distância para a Generalização do Problema Loop Fechado.	55
Figura 18 – Ilustração do problema para trechos de moléculas formados por 4 resíduos.	57
Figura 19 – Ilustração dos ângulos ψ e ϕ que são rotacionados no algoritmo CCD em loops formados por quatro resíduos para obter todas as conformações desses loops.	58
Figura 20 – Ilustração do cálculo da distância intervalar.	58
Figura 21 – Lado esquerdo da árvore de busca do iBP para o Problema do Loop Fechado de quatro resíduos.	60
Figura 22 – Exemplo de Loop Fechado formado por oito resíduos de uma molécula.	60
Figura 23 – Ilustração do cálculo da distância intervalar.	62
Figura 24 – Lado esquerdo da árvore de busca do iBP para o Problema do Loop Fechado de oito resíduos.	62
Figura 25 – Exemplo de Loop Fechado formado por doze resíduos de uma molécula.	63

Figura 26 – Ilustração do cálculo da distância intervalar. 64

Lista de abreviaturas e siglas

UNICAMP	Universidade Estadual de Campinas
IMECC	Instituto de Matemática, Estatística e Computação Científica
BP	Branch and Prune
GD	Geometria de Distâncias
PGD	Problema de Geometria de Distâncias
PDGD	Problema Discretizavel de Geometria de Distâncias
iBP	Intervalar Branch and Prune
CCS	Cyclic Coordinate Descent
iPDGD	Problema Discretizavel de Geometria de Distâncias intervalar
PLF	Problema do Loop Fechado
RMSD	Root Mean Square Deviation
LMDE	Local Distance Matrix Error
CSJD	Algoritmo proposto por Coutias

Sumário

	Introdução	14
1	GEOMETRIA DE DISTÂNCIAS	15
1.1	Interseções de n Esferas no \mathbb{R}^n	16
1.2	Algoritmo BP	16
1.3	O Algoritmo iBP	20
2	O PROBLEMA DO LOOP FECHADO	25
2.1	Algoritmo CSJD	26
2.1.1	Derivação das equações para loop tripeptídeo fechado	27
2.1.2	Algoritmo	41
2.2	Generalização	42
2.3	Método Cyclic Coordinate Descent	42
3	O PROBLEMA DO LOOP FECHADO EM GEOMETRIA DE DISTÂNCIAS	48
3.1	Instâncias para o Problema	48
3.2	O Problema do Loop Fechado e Problema Discretizável de Geometria de Distâncias	49
3.3	Distâncias Intervalares no Problema do Loop Fechado	50
3.4	Algoritmo iBP e o Problema do Loop Fechado	51
3.5	Generalizações	53
4	COMPARAÇÕES ENTRE OS MÉTODOS: CSJD, CCS E IBP	56
4.1	Loops formados por quatro resíduos de molécula	57
4.2	Loops formados por oito resíduos de moléculas	59
4.3	Loops formados por 12 resíduos de moléculas	61
4.4	Comparações Gerais	64
5	EXPERIMENTOS COMPUTACIONAIS	66
5.1	RMSD	66
5.2	LDME	66
5.3	Comparando o CSJD e o iBP	67
6	CONSIDERAÇÕES FINAIS	72

REFERÊNCIAS 73

Introdução

O principal problema da Geometria de Distância (GD) é encontrar as posições de conjuntos de pontos, considerando conhecidas somente algumas distâncias entre os pontos (CRIPPEN; HAVEL, 1988). Neste trabalho, utilizando resultados de GD e ilustrações para facilitar a visualização e compreensão, encontraremos soluções para uma classe de problemas denominada Problemas do Loop Fechado (PLF) (COUTSIAS et al., 2004).

O trabalho apresenta uma versão do algoritmo Branch and Prune (BP) (LIBERTI et al., 2014a), para o PLF, onde descrevemos as condições necessárias para sua utilização e uma análise detalhada sobre o espaço de busca do algoritmo. Além disso, o algoritmo generalizado, onde algumas distâncias podem não ter valores exatos, denominado interval BP (iBP) (GOLÇALVES et al., 2017) é estudado detalhadamente devido sua importância na solução do nosso problema.

O principal objetivo deste trabalho é utilizar a GD e o iBP para encontrar as soluções do PLF (COUTSIAS et al., 2004). Resumidamente, o problema consiste em encontrar estruturas de um segmento em uma molécula que sejam geometricamente consistentes com toda a estrutura da molécula. A aplicação que merece destaque deste problema é em Geometria Molecular (TRAMONTANO; LEPLAE; MOREA, 2001), onde é necessário modelar segmentos de inserções ou exclusões quando o restante da estrutura da molécula é bem conhecido.

Apresentamos um estudo detalhado de dois métodos existentes na literatura para solução do Problema do Loop fechado: o Cyclic Coordinate Descent (CCD) (CANUTESCU; DUNBRACK, 2003) e um método denominado de CSJD (COUTSIAS et al., 2004).

O principal desafio encontrado para a utilização da GD e do algoritmo BP no Problema do Loop Fechado foi o de encontrar a ordem necessária para satisfazer a definição do Problema Discretizável de Geometria de Distâncias ($PDGD_3$) (LIBERTI et al., 2014a). Entretanto, com a definição generalizada, denominada Problema Discretizável de Geometria de Distâncias intervalar ($iPDGD_3$), apresentada em (GOLÇALVES et al., 2017), foi possível definir uma ordem para utilização do algoritmo iBP.

Por fim, analisamos e comparamos os três métodos, apontando as condições necessárias para utilização dos mesmos e as vantagens da abordagem proposta.

1 Geometria de Distâncias

Neste capítulo, vamos falar sobre Geometria de Distâncias (GD), apresentaremos as principais definições e resultados sobre o GD que serão utilizados no decorrer do trabalho. Além disso, vamos apresentar e analisar os Algoritmos Branch and Prune (BP) e Interval Branch and Prune (iBP) que podem ser encontrados em (LIBERTI et al., 2014b) e (GOLÇALVES et al., 2017), respectivamente.

O principal problema em Geometria de Distâncias é encontrar as posições de um determinado conjunto de pontos, considerando conhecidas apenas informações sobre as distâncias entre alguns desses pontos (CRIPPEN; HAVEL, 1988). Segue abaixo a definição formal do problema (LIBERTI et al., 2014a).

Definição 1. *Problema de Geometria de Distâncias (PGD). Dado um inteiro positivo K e um grafo simples ponderado e não direcionado $G = (V, E, d)$, onde $d : E \rightarrow \mathbb{R}_+$, encontre uma função $x : V \rightarrow \mathbb{R}^K$ tal que*

$$\forall \{u, v\} \in E \quad \|x_u - x_v\| = d_{uv}, \quad (1.1)$$

onde $x_u = x(u)$, $d_{uv} = d(\{u, v\})$ e $\|\cdot\|$ denota a norma Euclidiana.

Uma função $x : V \rightarrow \mathbb{R}^K$ é chamada de realização de V . Dizemos que x é uma realização válida ou factível se satisfaz a Equação (1.1).

Neste trabalho, estamos interessados em utilizar uma versão "discreta" do PGD (LIBERTI et al., 2014a), onde o espaço de busca é finito. Essa nova versão do problema envolve uma ordenação nos vértices de G .

Definição 2. (LIBERTI et al., 2014a) *Problema Discretizável de Geometria de Distâncias (PDGD₃): Considere um grafo $G = (V, E, d)$ de um PGD, com $K = 3$, e uma ordem total 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 no mínimo três antecessores a_i, b_i, c_i com $\{\{a_i, v_i\}, \{b_i, v_i\}, \{c_i, v_i\}\} \subseteq E$ satisfazendo

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

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

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

Apresentaremos a seguir um algoritmo, chamado Branch and Prune (BP), que resolve o $PDGD_3$ explorando o espaço de busca, representado por uma árvore binária. A operação mais importante do BP é a interseção de esferas.

1.1 Interseções de n Esferas no \mathbb{R}^n

A interseção de n esferas no \mathbb{R}^n pode resultar em quatro diferentes tipos de soluções (COOPE, 2000), sendo elas:

- um ponto,
- dois pontos,
- conjunto vazio,
- conjunto não enumerável de pontos.

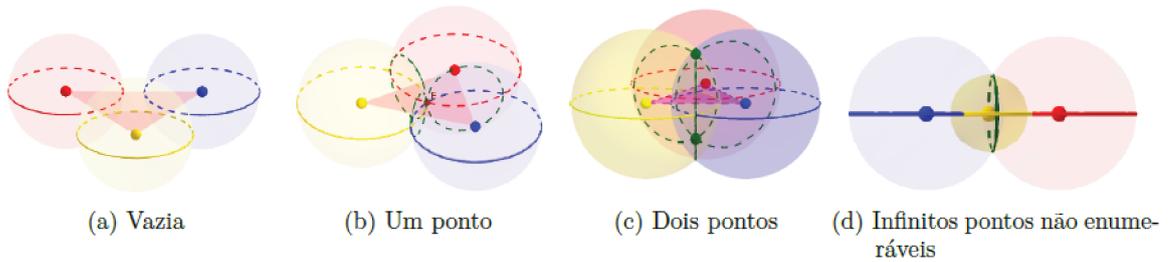


Figura 1 – Interseção de Esferas.

Veja a ilustração para $n = 3$ na Figura 1. O artigo de (COOPE, 2000) apresenta estratégias algébricas e numéricas para encontrar os pontos de interseções das esferas, sob a condição que a matriz originada dos centros das esferas possua posto completo. Para o caso de posto incompleto a análise é feita no artigo de (MAIOLI; GOLÇALVES; LAVOR, 2017).

1.2 Algoritmo BP

Seja $G = (V, E, d)$ um grafo associado a um PGD. Sejam $n = |V|$, $m = |E|$ e $N(v) = \{u \in V | \{u, v\} \in E\}$ para todo $v \in V$, o conjunto de vértices adjacentes à v . Para uma ordem total $<$ sobre V , seja $\gamma(v) = \{u \in V | u < v\}$ o conjunto dos antecessores de v e seja $\rho(v) = |\gamma(v)| + 1$ a posição de v nesta ordem. Além disso, considere $U_v = \{u_1, u_2, u_3 \in N(v) \cap \gamma(v) | \rho(u_1) > \rho(u_2) > \rho(u_3)\}$. Dado $V' \subset V$, denotamos por $G[V']$ o subgrafo de G induzido por V' .

Considere uma instância do $PDGD_3$. Seja $v \in V$ com nível $\rho(v) > 3$, $G^v = G[\gamma(v) \cup \{v\}]$ e x uma realização válida de $G[\gamma(v)]$. A seguir, vamos estudar o número de extensões válidas, uma extensão válida é dada por um ponto x_v que satisfaz todas as distâncias conhecidas para o vértice v , para a realização de G^v .

Lema 1. *Se $|N(v) \cap \gamma(v)| = 3$, então existem no máximo duas extensões distintas de x que são válidas para G^v . Além disso, se existe uma extensão válida, então, com probabilidade 1, existem exatamente duas extensões válidas.*

Demonstração. Sabemos que $|N(v) \cap \gamma(v)| = 3$. Assim, $U_v = N(v) \cap \gamma(v) = \{u_1, u_2, u_3\}$ e v pertence à interseção de exatamente 3 esferas em \mathbb{R}^3 , onde cada esfera tem centro u_j , raio $d_{u_j v}$, $u_j \in U_v$. A posição $z \in \mathbb{R}^3$ de v deve então satisfazer:

$$\begin{aligned} \forall u_j \in U_v, \|z - x_{u_j}\| = d_{u_j v} &\Rightarrow \|z - x_{u_j}\|^2 = d_{u_j v}^2 \\ &\Rightarrow \langle z - x_{u_j}, z - x_{u_j} \rangle = d_{u_j v}^2 \\ &\Rightarrow \langle z, z \rangle - 2\langle z, x_{u_j} \rangle + \langle x_{u_j}, x_{u_j} \rangle = d_{u_j v}^2 \\ &\Rightarrow \|z\|^2 - 2x_{u_j} \cdot z + \|x_{u_j}\|^2 = d_{u_j v}^2. \end{aligned} \quad (1.2)$$

Considerando $u_1 \in U_v$ e subtraindo a linha indexada por u_1 da equação (1.2) das outras linhas de (1.2), obtemos o sistema:

$$\begin{cases} 2(x_{u_2} - x_{u_1}) \cdot z = (\|x_{u_2}\|^2 - d_{u_2 v}^2) - (\|x_{u_1}\|^2 - d_{u_1 v}^2) \\ 2(x_{u_3} - x_{u_1}) \cdot z = (\|x_{u_3}\|^2 - d_{u_3 v}^2) - (\|x_{u_1}\|^2 - d_{u_1 v}^2) \\ \|z\|^2 - 2x_{u_1} \cdot z + \|x_{u_1}\|^2 = d_{u_1 v}^2. \end{cases} \quad (1.3)$$

Note que o sistema (1.3) é composto por 2 equações lineares e uma única equação quadrática no 3-vetor z . Escrevemos as equações lineares como um sistema $Az = b$, onde

$$A = \begin{bmatrix} 2(x_{u_2} - x_{u_1}) \\ 2(x_{u_3} - x_{u_1}) \end{bmatrix}, \quad b = \begin{bmatrix} (\|x_{u_2}\|^2 - d_{u_2 v}^2) - (\|x_{u_1}\|^2 - d_{u_1 v}^2) \\ (\|x_{u_3}\|^2 - d_{u_3 v}^2) - (\|x_{u_1}\|^2 - d_{u_1 v}^2) \end{bmatrix}.$$

A matriz A tem dimensão 2×3 e $b \in \mathbb{R}^2$. Suponha que a matriz A não tem posto completo. Então, pela condição de dependência linear,

$$\xi_{u_2}(x_{u_2} - x_{u_1}) + \xi_{u_3}(x_{u_3} - x_{u_1}) = 0,$$

para alguns coeficientes $\xi_{u_2}, \xi_{u_3} \neq 0$, implicando que x_{u_3} pertence ao espaço gerado por $\{x_{u_2}, x_{u_1}\}$, ou seja, $x_{u_3}, x_{u_2}, x_{u_1}$ são colineares, o que contradiz a hipótese da desigualdade triangular estrita da Definição 2. Portanto, A tem posto completo. Assim, sem perda de generalidade, assumamos que a matriz quadrada B formada pelas primeiras duas colunas de A seja invertível. Defina z_B como o vetor formado das primeiras duas componentes de z . Então, pela parte linear do sistema (1.3), temos:

$$Bz_B = b - Nz_3,$$

onde $N = \begin{bmatrix} 2(x_{u_2,3} - x_{u_1,3}) \\ 2(x_{u_3,3} - x_{u_1,3}) \end{bmatrix}$, resultando em $z_B = B^{-1}(b - Nz_3)$ como uma função $z_B(z_3)$ de z_3 . Substituindo z_B no sistema (1.3) como $z_B(z_k)$, obtemos:

$$\begin{aligned} d_{u_1,v}^2 &= \| B^{-1}(b - Nz_3) \|^2 + z_3^2 - 2x_{u_1,B}(B^{-1}(b - Nz_3)) - 2x_{u_1,3}z_3 + \| x_{u_1,B} \|^2 \\ &\quad + x_{u_1,3}^2 \\ &= \| B^{-1}b - B^{-1}Nz_3 \|^2 + z_3^2 - 2x_{u_1,B}B^{-1}b + 2x_{u_1,B}B^{-1}Nz_3 - 2x_{u_1,3}z_3 \\ &\quad + \| x_{u_1,B} \|^2 + x_{u_1,3}^2. \end{aligned}$$

Tomando $\bar{b} = B^{-1}b$ e $\bar{N} = B^{-1}N$, temos:

$$\begin{aligned} d_{u_1,v}^2 &= \| \bar{b} - \bar{N}z_3 \|^2 + z_3^2 - 2x_{u_1,B}\bar{b} + 2x_{u_1,B}\bar{N}z_3 - 2x_{u_1,3}z_3 + \| x_{u_1,B} \|^2 + x_{u_1,3}^2 \\ &= \| \bar{b} \|^2 - 2\langle \bar{b}, \bar{N}z_3 \rangle + \| \bar{N}z_3 \|^2 + z_3^2 - 2x_{u_1,B}\bar{b} + 2x_{u_1,B}\bar{N}z_3 - 2x_{u_1,3}z_3 \\ &\quad + \| x_{u_1,B} \|^2 + x_{u_1,3}^2 \\ &= (\| \bar{N} \|^2 + 1)z_3^2 + \langle \bar{b}, \bar{b} \rangle + \langle x_{u_1,B}, x_{u_1,B} \rangle - 2\langle x_{u_1,B}, \bar{b} \rangle \\ &\quad - 2((\bar{b} - x_{u_1,B})\bar{N} + x_{u_1,3})z_3 + x_{u_1,3}^2. \end{aligned}$$

Obtemos, então, a seguinte equação quadrática na variável z_3 :

$$\begin{aligned} (\| \bar{N} \|^2 + 1)z_3^2 + \| x_{u_1,B} - \bar{b} \|^2 - 2((\bar{b} - x_{u_1,B})\bar{N} + x_{u_1,3})z_3 + x_{u_1,3}^2 \\ - d_{u_1,v}^2 = 0. \end{aligned} \quad (1.4)$$

Se o discriminante da Equação (1.4) é negativo, então não existe extensão válida de \bar{x} para v e o resultado segue. Se o discriminante é não negativo, temos soluções z'_3, z''_3 , gerando os pontos $z' = (z_B(z'_3), z'_3)$, $z'' = (z_B(z''_3), z''_3) \in \mathbb{R}^3$, que são distintos com probabilidade 1, pois o discriminante 0 tem probabilidade 0 (LIBERTI et al., 2014b). As realizações estendidas, distintas com probabilidade 1, são dadas por (x, z') e (x, z'') . \square

Geometricamente, o Lema 1 é interpretado como a interseção de 3 esferas no \mathbb{R}^3 , o que fornece um conjunto solução com no máximo dois pontos.

Lema 2. *Se $|N(v) \cap \gamma(v)| > 3$, então, com probabilidade 1, existe no máximo uma extensão de x .*

Demonstração. Seja $S \subseteq N(v) \cap \gamma(v)$, satisfazendo $|S| = 4$ e $S \ni U_v$. Assim, existe pelo menos um ponto x_v tal que (x, x_v) é uma realização de $G[S \cup \{v\}]$, em relação ao sistema

$$\forall u \in S, \quad x_{v,1}^2 - 2x_{u,1}x_{v,1} + x_{u,1}^2 + x_{v,2}^2 - 2x_{u,2}x_{v,2} + x_{u,2}^2 + x_{v,3}^2 - 2x_{u,3}x_{v,3} + x_{u,3}^2 = d_{uv}^2 \quad (1.5)$$

ou o sistema não tem solução. No último caso, o resultado segue. Assuma agora que existe um ponto x_v satisfazendo o sistema (1.5). Já que os pontos x_u são conhecidos, $\forall u \in S$, obtemos que o sistema (1.5) é um sistema quadrático com 3 variáveis e 4 equações. Assim,

como na demonstração do Lema 1, obtemos um sistema linear equivalente ao sistema (1.5). Além disso, com probabilidade 1, os pontos $\{x_u | u \in S\}$ são não coplanares, implicando que o sistema associado tenha exatamente uma solução. \square

A Interpretação geométrica do Lema 2 é vista como interseção de 4 esferas no \mathbb{R}^3 , o que gera um conjunto solução com no máximo um ponto.

Lema 3. *Utilizando a notação do Lema 1, se \bar{x} é uma realização para $G[U_v]$, então z'' é uma reflexão de z' , com respeito ao plano dado pelos 3 pontos de \bar{x} .*

Demonstração. Note que as duas soluções encontradas no Lema 1, geometricamente, pertencem à interseção de 3 esferas em \mathbb{R}^3 . Além disso, sabe-se que qualquer esfera em \mathbb{R}^3 é simétrica com respeito a qualquer plano passando por seu centro. Portanto a interseção de 3 esferas em \mathbb{R}^3 é simétrica com respeito ao plano contendo todos os centros das 3 esferas. Assim, os pontos z, z' são reflexões entre si. \square

A seguir, apresentamos um pseudo-código do algoritmo BP para instâncias do tipo $PDGD_3$ (MUCHERINO; LAVOR; LIBERTI, 2012), onde k é representa o vértice de posição k na ordem da Definição 2, n é quantidade de vértices da ordem, d são as distâncias conhecidas entre os átomos e $x_k^{(i)}$ representa a i -ésima possível posição que o vértice k pode assumir.

Algoritmo 1 O algoritmo BP

```

1:  $BP(k, n, d)$ 
2: para  $(i=1,2)$  fazer
3:   calcular as  $i$ -ésimas posições para o vértice  $k$ :  $x_k^{(i)}$ ;
4:   verificar a factibilidade das posições  $x_k^{(i)}$ ;
5:   se (a posição  $x_k^{(i)}$  é factível) então
6:     se  $k = n$  então
7:       uma solução é encontrada
8:     senão
9:        $BP(k + 1, n, d)$ 
10:    fim se
11:  senão
12:    o ramo correspondente é eliminado
13:  fim se
14: fim para

```

Em resumo, dado uma ordem nos vértices de V que satisfaz a Definição 2, o Algoritmo BP age recursivamente na ordem explorando a árvore de posições possíveis para os vértices. A cada vértice da ordem, cujo posicionamento é determinado, é definido

um novo ramo para árvore de busca, esta fase do algoritmo é conhecida como fase de busca. Para vértices que possuem informações de distâncias adicionais conhecidas são realizadas as etapas de verificações da factibilidade das posições encontradas para os vértices através da checagem das informações adicionais, esta fase é chamada de fase de poda.

1.3 O Algoritmo iBP

Nesta seção, vamos apresentar uma generalização do Algoritmo BP, chamada de Algoritmo *iBP*, o qual é mais robusto, no sentido de possibilitar a aplicação deste algoritmo em diversos problemas envolvendo estrutura de proteínas.

Seja $G = (V, E, d)$ um grafo de um PDGD associado a uma proteína. Dado um vértice $v \in V$, denotamos por $v - h$ o vértice de nível $\rho(v) - h$ para algum $h < \rho(v)$. As distâncias entre átomos conectados por ligações covalentes são conhecidas com precisão aceitável (CRIPPEN; HAVEL, 1988). Assim, consideramos as distâncias $d_{u_1,v}, d_{u_2,v}$ exatas para qualquer vértice v tal que $\rho(v) > 3$. As distâncias $d_{u_3,v}$, bem como outras distâncias relativas a átomos próximos detectados por experimentos, podem não possuir valores exatos.

Sempre que as distâncias entre um átomo i até seus adjacentes são exatas, geometricamente estamos trabalhando para encontrar a interseção entre três esferas, onde consideramos que os vértices adjacentes $\{a, b, c\}$ são não colineares, resultando em no máximo dois pontos. Entretanto, quando uma das distâncias é um valor intervalar, geometricamente buscamos encontrar a interseção entre duas esferas e uma casca esférica, resultando em dois arcos sobre a circunferência determinada pela interseção das duas esferas. Veja a ilustração na Figura 2, retirada de (GOLÇALVES et al., 2017). Quando encontramos o caso em que a interseção é dada por dois arcos, para manter o conjunto discreto de possibilidades de posições para o átomo i , devemos selecionar um número finito D de posições dos arcos. Na figura 2, a, b, c representam os pontos fixos, são conhecidas as distâncias exatas dos vértices correspondentes aos pontos a, b até o vértice i que está sendo realizado, mas a distância entre o vértice correspondente ao ponto c até o vértice i é dada por um intervalo.

Definição 3. (GOLÇALVES et al., 2017) *Discretização Intervalar do Problema de Geometria de Distâncias em Dimensão 3 (iPDGD₃):* Dado um grafo simples, ponderado e não direcionado $G = (V, E, d)$, onde $E' \subset E$ é um subconjunto de arestas para as quais suas distâncias são dadas por valores exatos, dizemos que G representa uma instância do *iPDGD₃* se existe uma ordem total nos vértices de V satisfazendo:

$G_C = (V_C, E_C) = G[\{1, 2, 3\}]$, é uma clique, tal que $E_C \subset E'$;

$\forall i \in \{4, \dots, |V|\}$ existem $\{a, b, c\}$ tais que

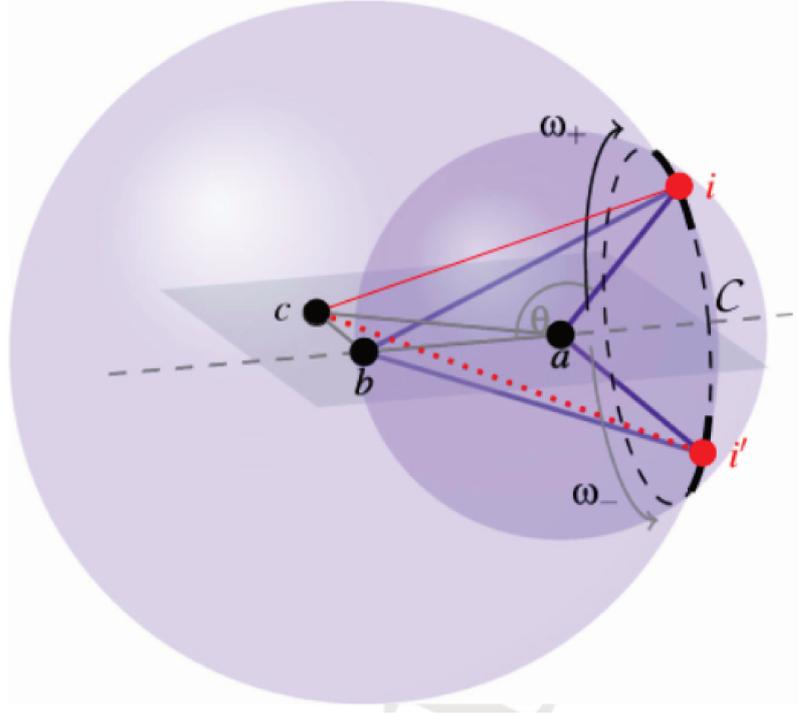


Figura 2 – Interseção de duas esferas e um casca esférica.

1. $a < i$, $b < i$, $c < i$;
2. $\{\{b, i\}, \{c, i\}\} \subset E'$, $\{a, i\} \subset E$;
3. $\Delta_S(a, b, c) > 0$,

onde $\Delta_S(a, b, c)$ é a área do triângulo formado por a, b, c .

Note que esta ordem utiliza as distâncias intervalares para realizar a discretização e também as podas na árvore de busca. A seguir, vamos apresentar alguns resultados conhecidos sobre o Algoritmo iBP (GOLÇALVES et al., 2017):

Lema 4. *Seja q o número de amostras retiradas da distância intervalar. Se $|N(v) \cap \gamma(v)| = 3$, então existem no máximo $2q$ extensões distintas de x que são válidas para G^v . Se existe uma extensão válida, então, com probabilidade 1, existem exatamente $2q$ extensões válidas distintas.*

Demonstração. Já que $|N(v) \cap \gamma(v)| = 3$, então $U_v = N(v) \cap \gamma(v) = \{u_1, u_2, u_3\}$. Tomando u_3 , tal que, d_{vu_3} é dado por um intervalo, temos que a equação (1.2) é dada por

$$\forall u \in U_v - \{u_3\}, (\|z - x_{u_i}\| = d_{u_i v}) \text{ e } (\|z - x_{u_3}\| \in d_{u_3, v})$$

e a terceira linha da Equação (1.3) é dada por;

$$\|z\|^2 - 2x_{u_3} \cdot z + \|x_{u_3}\|^2 \in d_{u_3, v}^2,$$

onde $d_{u_3,v}^2$ denota o conjunto $\{(d_{u_3,v}^j)^2 | j \leq q\}$ das q distâncias $d_{u_3,v}$. Temos então, q diferentes equações quadráticas

$$\|z\|^2 - 2x_{u_3} \cdot z + \|x_{u_3}\|^2 = (d_{u_3,v}^j)^2, \quad (1.6)$$

onde $j \leq q$. Para cada uma dessas equações, segue pelo Lema 1 que existem dois diferentes valores para z com probabilidade 1, como queríamos demonstrar. \square

Lema 5. *Seja \bar{x} uma realização válida de $G[U_v]$ e sejam $\{z^{1j}, z^{2j} | j \leq q\} \subseteq \mathbb{R}^3$ as $2q$ posições do vértice v dadas pelo Lema 4, onde z^{1j}, z^{2j} são obtidas da equação (1.6), indexadas por j . Então, z^{2j} é uma reflexão de z^{1j} com respeito ao plano dado pelos 3 pontos de \bar{x} .*

Demonstração. Como qualquer esfera em \mathbb{R}^3 é simétrica com respeito a qualquer plano passando pelo seu centro, a interseção de 3 esferas é simétrica ao plano através dos seus 3 centros. Então, para cada valor d_{uv}^j , temos que z^{2j} é simétrico a z^{1j} para cada $j \leq q$. \square

A seguir, apresentamos um Pseudo-Código do iBP, retirado de (GOLÇALVES et al., 2017), onde i representa o i -ésimo vértice da ordem da Definição 3, n é o número de vértices da ordem, d representa o conjunto de distância exatas conhecidas e D o conjunto de distâncias intervalares conhecidas.

Algoritmo 2 O algoritmo iBP

```

1:  $iBP(i, n, d, D)$ 
2: se ( $i > n$ ) então
3:   uma solução é encontrada
4:   print conformação respectiva
5: senão
6:   calcular coordenadas
7:   se ( $d_{ci}$  é um intervalo) então
8:     calcule os dois arcos e adicione-os a lista  $L$ 
9:   senão
10:    calcule os dois vetores posição e adicione-os a lista  $L$ 
11:   fim se
12:   para  $j = 1, \dots, |L|$  fazer
13:     se ( $L(j)$  é um arco) então
14:       tomar  $D$  amostras do arco; faça  $N = D$ ;
15:     senão
16:        $N = 1$ 
17:     fim se
18:     verifique a factibilidade das posições calculadas
19:     para  $k = 1, \dots, N$  fazer
20:       se ( $x_i^{j,k}$  é factível) então
21:          $iBP(i + 1, n, d, D)$ ;
22:       fim se
23:     fim para
24:   fim para
25: fim se

```

Em resumo, dada uma ordem nos vértices de V que satisfaz a Definição 3, o Algoritmo iBP age recursivamente na ordem explorando a árvore de posições possíveis para os átomos. A cada vez que calculamos uma posição para um novo átomo, também definimos um novo ramo para a árvore de busca. Chamamos esta fase no Algoritmo iBP de fase de busca. Após um átomo ter sua posição calculada, é verificado a factibilidade desta posição através da verificação das informações adicionais obtidas das propriedades da molécula. Chamamos esta fase no Algoritmo iBP de fase de poda.

Na Figura 3, ilustramos a utilização de distâncias intervalares para a realização de podas no espaço de busca. Nesta ilustração, observamos que a distância intervalar d_{uv} permite que somente 2 dos valores encontrados para distância d_{uv} sejam válidos, ao invés dos 4 inicialmente encontrados durante a “exploração” da árvore de busca, sendo eles r_1^2, r_2^2 . Consequentemente, o número de possíveis realizações do vértice v diminui de 16 para 8.

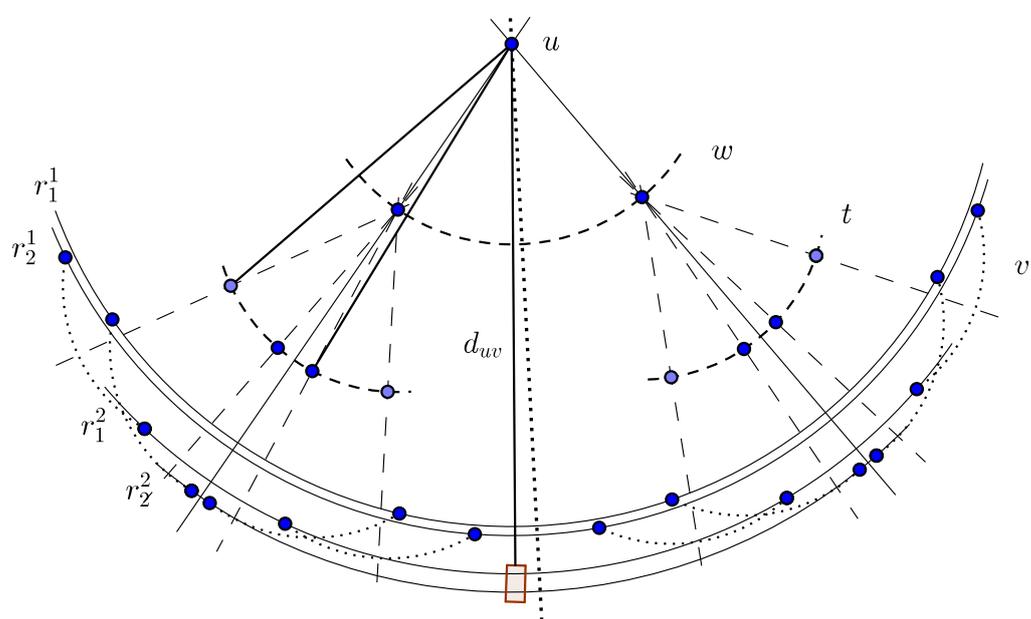


Figura 3 – Ilustração da utilização de distância intervalar para realizar podas na árvore de busca.

2 O Problema do Loop Fechado

Neste capítulo, vamos apresentar o problema do Loop Fechado e descrever os principais resultados da literatura sobre este problema. O principal trabalho utilizado para o estudo do problema foi (COUTSIAS et al., 2004). Resumidamente, o problema consiste em encontrar estruturas de um segmento em uma molécula que sejam geometricamente consistentes com toda a estrutura da molécula. A aplicação que merece destaque deste problema é em modelagem por homologia (TRAMONTANO; LEPLAE; MOREA, 2001), onde é necessário modelar segmentos de inserções ou exclusões quando o restante da estrutura da molécula é bem conhecido.

Em robótica, ligações que permitem uma estrutura girar em torno de outra em um ângulo fixo são chamadas Pares de Rotação ou R-pares. Um sistema de estruturas com seis ligações rotáveis é uma “6R”-ligação (HARTENBERG; DANAVIT, 1964).

O problema considerado neste trabalho é um caso especial do problema das seis ligações rotáveis e possui uma descrição simples: Consideram-se todos os movimentos de uma cadeia molecular que envolve mudança em somente seis torções da cadeia principal da molécula e busca-se determinar a posição de todos os átomos pertencentes a esta cadeia utilizando informações sobre distâncias e ângulos conhecidos entre alguns desses átomos.

A cadeia é arranjada de forma a possuir três pares de rotações consecutivas, implicando que os segmentos entre os pares formam uma cadeia maior com três ou quatro corpos rígidos, respectivamente denominados de caso fechado e caso aberto.

A Figura 4, retirada de (COUTSIAS et al., 2004), é uma ilustração para o Problema do Loop fechado descrito acima. O caso aberto é exemplificado quando a cadeia é formada pelos seguintes quatro corpos rígidos: $(N_1, C_{\alpha 1})$, $(C_{\alpha 1}, C_1, N_2, C_{\alpha 2})$, $(C_{\alpha 2}, C_2, N_3, C_{\alpha 3})$ e $(C_{\alpha 3}, C_3)$. O caso fechado é obtido neste exemplo quando consideramos os seguintes três corpos rígidos: $(C_{\alpha 3}, C_3, N_1, C_{\alpha 1})$, $(C_{\alpha 1}, C_1, N_2, C_{\alpha 2})$ e $(C_{\alpha 2}, C_2, N_3, C_{\alpha 3})$. Para o caso fechado, cada um dos três corpos rígidos possui duas ligações que o conecta com os outros dois corpos rígidos. Definindo uma linha que conecta as duas ligações sobre cada um dos corpos rígidos, esta linha será chamada de eixo de rotação para seu respectivo corpo rígido.

Os movimentos permitidos nessa estrutura são somente os realizados individualmente, por dois dos três corpos rígidos, girando em torno de seus eixos de rotação $(C_{\alpha 1} - C_{\alpha 2}$ e $C_{\alpha 2} - C_{\alpha 3})$ e da rotação deste dois corpos rígidos simultaneamente em torno do eixo de rotação do terceiro $(C_{\alpha 1} - C_{\alpha 3})$. Desta forma, todos os movimentos permitidos podem ser descritos através de rotações de ângulos τ_i , $i = 1, 2, 3$, sobre os eixos de rotação. Se considerarmos os ângulos θ_i , $i = 1, 2, 3$, entre as ligações dos corpos rígidos com valores

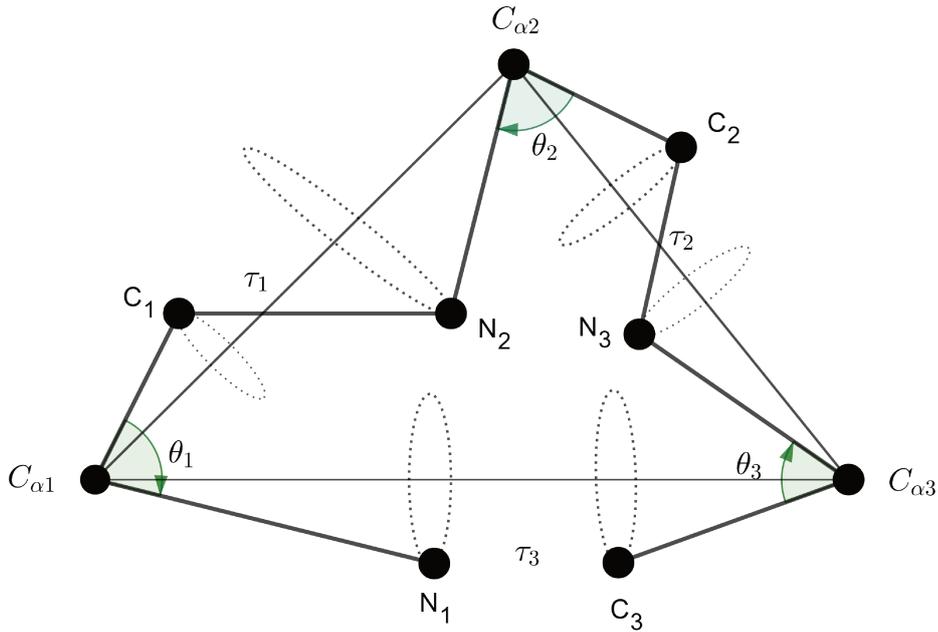


Figura 4 – Ilustração do problema canônico do loop fechado onde os três corpos rígidos são dados por: $(C_{\alpha 3}, C_3, N_1, C_{\alpha 1})$, $(C_{\alpha 1}, C_1, N_2, C_{\alpha 2})$ e $(C_{\alpha 2}, C_2, N_3, C_{\alpha 3})$.

fixos, as possíveis posições assumidas pela estrutura através dos movimentos satisfazendo estas restrições formam um conjunto discreto (COUTSIAS et al., 2004). Segue abaixo a definição formal do problema.

Definição 4. *Problema do Loop Fechado: Encontre todas as possíveis estruturas de uma cadeia molecular obtida através de mudanças nos valores dos ângulos τ_i , $i = 1, 2, 3$, considerando as seguintes restrições:*

- As posições dos átomos $C_{\alpha 1}, C_{\alpha 2}, C_{\alpha 3}$ são conhecidas.
- As distâncias entre os átomos de um mesmo corpo rígido são conhecidas e fixas.
- Os ângulos θ_i , $i = 1, 2, 3$ são conhecidos e fixos.

Para facilitar a interpretação do problema pelo leitor vamos exibir a seguir na Tabela 1 a Matriz de Adjacência do Problema Canônico do Loop Fechado.

2.1 Algoritmo CSJD

Na literatura, encontramos diversos trabalhos voltados a resolver o problema de determinar as conformações de loops fechados com seis torções, tais como (GO; SCHE- RAGA, 1970) e (LEE; LIANG, 1988). Entretanto, a grande maioria deles aborda o problema

Tabela 1 – Matriz de Adjacência do PLF

	$C_{\alpha 1}$	C_1	N_2	$C_{\alpha 2}$	C_2	N_3	$C_{\alpha 3}$	C_3	N_1
$C_{\alpha 1}$	1	1	1	1	0	0	1	1	1
C_1	1	1	1	1	0	0	0	0	1
N_2	1	1	1	1	1	0	0	0	0
$C_{\alpha 2}$	1	1	1	1	1	1	1	0	0
C_2	0	0	1	1	1	1	1	0	0
N_3	0	0	0	1	1	1	1	1	0
$C_{\alpha 3}$	1	0	0	1	1	1	1	1	1
C_3	1	0	0	0	0	1	1	1	1
N_1	1	1	0	0	0	0	1	1	1

através da redução para encontrar raízes de polinômios. A seguir, vamos descrever em detalhes o trabalho de Countsias et al. sobre encontrar conformações de loop fechado de seis torções e o algoritmo chamado CSJD. (COUTSIAS et al., 2004).

O artigo considera uma versão simplificada do problema do loop fechado: Dada uma cadeia molecular, onde as distâncias entre os vértices e ângulos entre eles são fixos, encontre todas as conformações obtidas através de movimentos que envolvem mudanças feitas em no máximo seis torções da cadeia principal da molécula. Para facilitar a apresentação, abordamos o caso do loop tripeptídeo fechado.

2.1.1 Derivação das equações para loop tripeptídeo fechado

Nesta seção, vamos descrever a derivação do problema do Loop Fechado de seis torções, utilizando uma representação simplificada e exemplificada pela Figura 6. De acordo com (COUTSIAS et al., 2004), o problema é completamente caracterizado considerando conhecidos os ângulos α_i , um dos eixos de rotação, os ângulos ξ_i, η_i formados por ligações rotacionáveis em cada $C_{\alpha i}$ e os ângulos diedrais δ_i . Entretanto, considerar os vértices $C_{\alpha i}$ fixos é equivalente a considerar um dos três corpos rígidos fixos. Desta forma, assim como na Figura 4, os átomos $N_1, C_{\alpha 1}, C_{\alpha 3}$ e C_3 são fixos no espaço e todas as possíveis posições dos demais átomos são determinadas a partir de algumas informações de distâncias e ângulos entre os átomos.

Um resultado bem conhecido na literatura para o caso do Loop Fechado com seis ângulos de torção livres diz que o número de restrições é igual ao número de graus de liberdade (GO; SCHERAGA, 1970). No exemplo da Figura 4, existem 3 restrições dadas pelos ângulos $\theta_i, i = 1, 2, 3$, e três variáveis dadas pelos ângulos de rotação $\tau_i, i = 1, 2, 3$, dos C_i e N_{i+1} em torno do seu eixo de rotação $C_{\alpha i} - C_{\alpha i+1}$, onde $i + 1 = 4$ é equivalente a $i = 1$. As rotações pelos ângulos τ_i preservam todas as distâncias e ângulos entre os átomos, exceto os três ângulos θ_i .

Os movimentos dados pelas torções iniciais do problemas são representados

pelos movimentos dos três ângulos τ_i . Fixando os ângulos θ_i , restringimos os ângulos τ_i a um conjunto discreto de valores encontrados como raízes de uma equação polinomial nas correspondentes variáveis u_i , com u_i dado pela transformação (ver a Figura 5, reedição de imagem disponível em (COUTSIAS et al., 2004)).

$$u_i = \operatorname{tg} \frac{\tau_i}{2} \rightarrow \operatorname{sen} \tau_i = \frac{2u_i}{1+u_i^2}, \quad \cos \tau_i = \frac{1-u_i^2}{1+u_i^2}, \quad i = 1, 2, 3.$$

Detalhamos a seguir os métodos propostos por (COUTSIAS et al., 2004) e (CANUTESCU; DUNBRACK, 2003). Assim o leitor pode saltar para o Capítulo 3.

Considerando os seguintes vetores unitários:

$$\hat{z}_i = \frac{C_{\alpha i} C_{\alpha i+1}}{|C_{\alpha i} C_{\alpha i+1}|}, \quad \hat{r}_i^\tau = \frac{C_{\alpha i} C_i}{|C_{\alpha i} C_i|}, \quad \hat{r}_i^\sigma = \frac{C_{\alpha i+1} N_{i+1}}{|C_{\alpha i+1} N_{i+1}|}, \quad (2.1)$$

definimos os ângulos fixos:

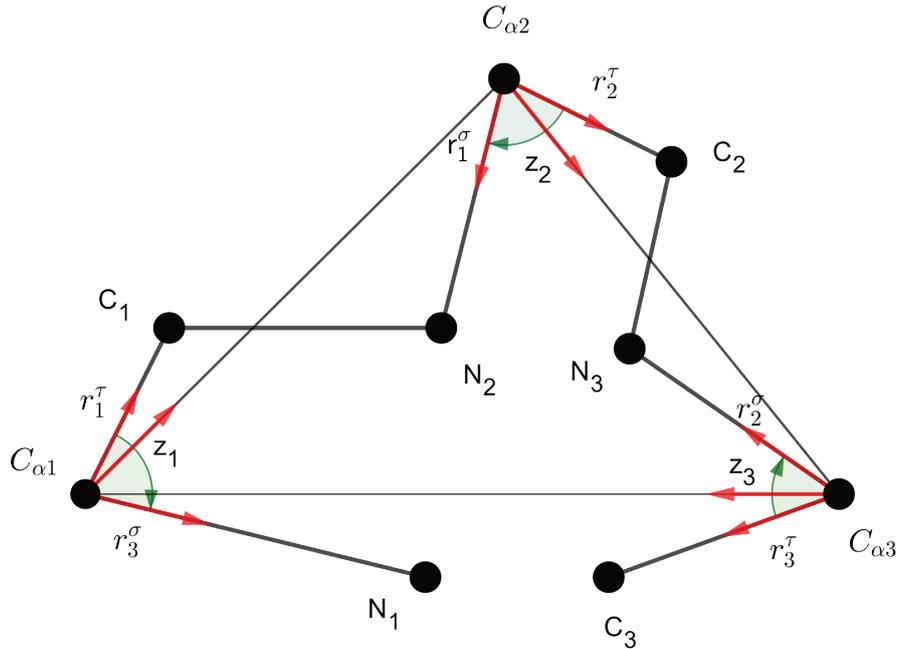


Figura 5 – Ilustração dos vetores $z_i, r_i^\tau, r_i^\sigma$.

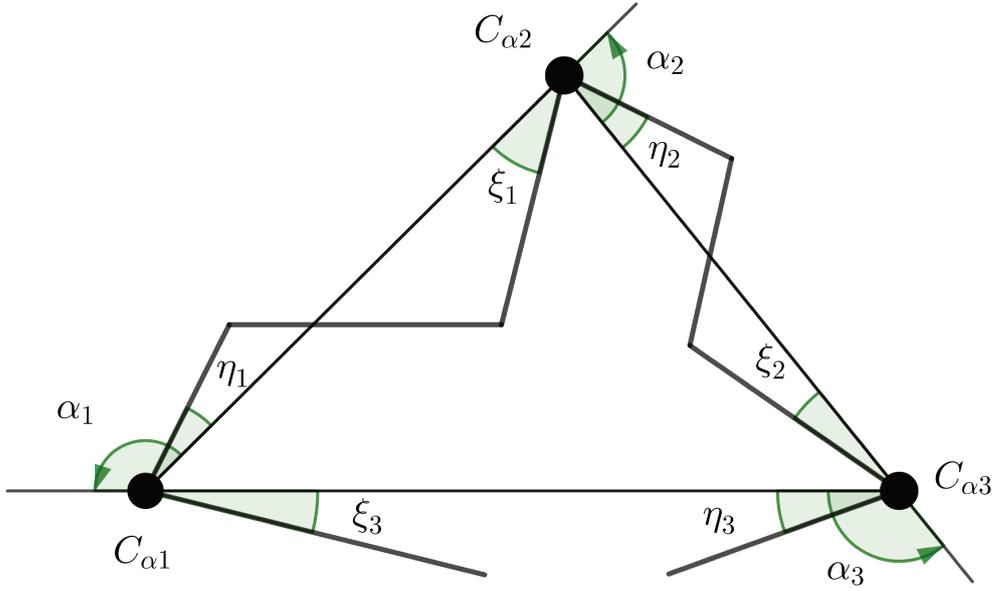
$$\alpha_i = \cos^{-1}(\hat{z}_i \cdot \hat{z}_{i-1}), \quad (2.2)$$

$$\eta_i = \cos^{-1}(\hat{z}_i \cdot \hat{r}_i^\tau), \quad (2.3)$$

$$\xi_i = \cos^{-1}(-\hat{z}_i \cdot \hat{r}_i^\sigma), \quad (2.4)$$

onde α_i, η_i e ξ_i são todos tomados no intervalo $[0, 2\pi]$ (ver Figura 6, retirada de (COUTSIAS et al., 2004)).

Agora, definimos um sistema de coordenadas local, respeitando a regra da mão direita, por três vetores unitários $(\hat{x}_i, \hat{y}_i, \hat{z}_i)$ para a rotação de τ_i , onde o eixo de referência


 Figura 6 – Ilustração dos ângulos α_i, η_i, ξ_i .

\hat{y} é definido por

$$\hat{y} = \frac{\hat{z}_3 \times \hat{z}_1}{|\hat{z}_3 \times \hat{z}_1|}. \quad (2.5)$$

Dessa forma, ele é perpendicular a todos os \hat{z}_i definidos na equação (2.1), com $\hat{x}_i = \hat{y} \times \hat{z}_i$. Os ângulos τ_i podem ser escritos em termos dos ângulos de rotação de \hat{r}_i^T sobre \hat{z}_i , neste mesmo sistema local de coordenadas, definindo σ_i como o ângulo de rotação de \hat{r}_i^σ sobre \hat{z}_i .

Os ângulos τ_i e σ_i estão “relacionados”, pois \hat{r}_i^T e \hat{r}_i^σ são rotacionados como um corpo rígido. A relação entre τ_i e σ_i é dada por

$$\sigma_i = \tau_i + \delta_i, \quad (2.6)$$

onde δ_i é o ângulo de torção definido pelos três vetores $(C_i C_{\alpha i}, C_{\alpha i} C_{\alpha i+1}, C_{\alpha i+1} N_{i+1})$. Ilustramos a relação da equação (2.6) na Figura 7, imagem retirada de (COUTSIAS et al., 2004).

Como podemos ver, na Figura 8, construída a partir de Figura encontrada em (COUTSIAS et al., 2004), os vetores unitários \hat{r}_i^T e \hat{r}_i^σ estão expressos em termos dos vetores unitários e ângulos acima definidos como

$$\hat{r}_i^T = \cos \eta_i \hat{z}_i + \sin \eta_i (\cos \tau_i \hat{x}_i + \sin \tau_i \hat{y}), \quad (2.7)$$

$$\hat{r}_{i-1}^\sigma = -\cos \xi_i \hat{z}_{i-1} + \sin \xi_i (\cos \sigma_{i-1} \hat{x}_{i-1} + \sin \sigma_{i-1} \hat{y}), \quad (2.8)$$

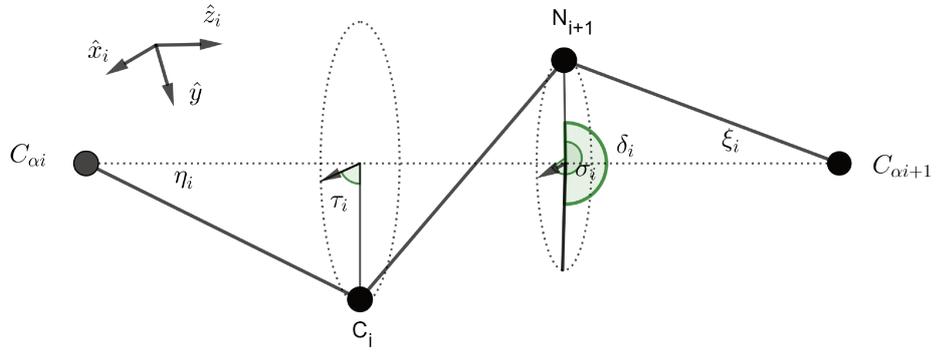


Figura 7 – Relação entre os ângulos τ_i e δ_i .

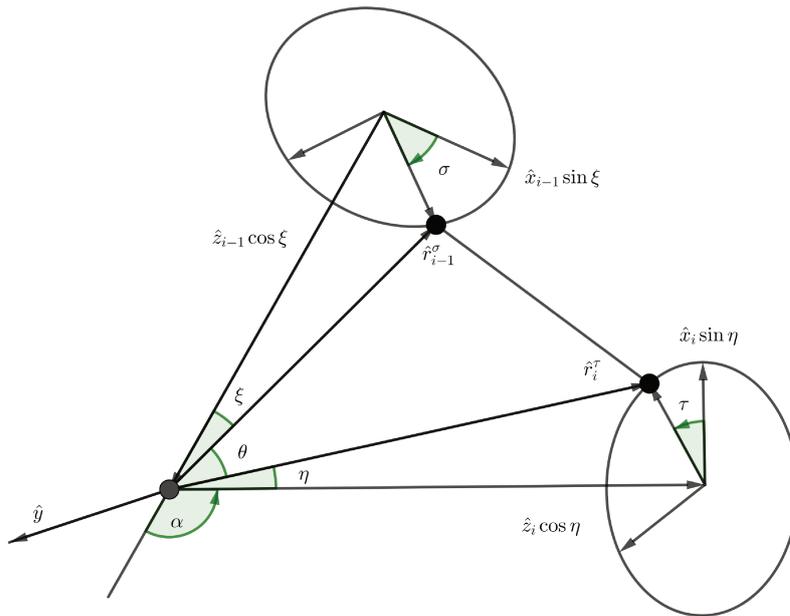


Figura 8 – Relação do ângulo de restrição θ com \hat{r}_i^τ e \hat{r}_{i-1}^σ .

e o ângulo de restrição θ_i pode ser descrito em termos de \hat{r}_i^τ e \hat{r}_{i-1}^σ por

$$\hat{r}_i^\tau \cdot \hat{r}_{i-1}^\sigma = \cos \theta_i. \quad (2.9)$$

Substituindo as equações (2.7) e (2.8) na equação (2.9), obtemos as seguintes

equações:

$$\begin{aligned}
\cos \theta_i &= \hat{r}_i^\tau \cdot \hat{r}_{i-1}^\sigma \\
&= (\cos \eta_i \hat{z}_i + \text{sen} \eta_i (\cos \tau_i \hat{x}_i + \text{sen} \tau_i \hat{y})) \cdot (-\cos \xi_i \hat{z}_{i-1} \\
&\quad + \text{sen} \xi_{i-1} (\cos \sigma_{i-1} \hat{x}_{i-1} + \text{sen} \sigma_{i-1} \hat{y})) \\
&= -\cos \eta_i \cos \xi_{i-1} \hat{z}_i \cdot \hat{z}_{i-1} + \cos \eta_i \text{sen} \xi_{i-1} \cos \sigma_{i-1} \hat{z}_i \cdot \hat{x}_{i-1} \\
&\quad + \cos \eta_i \text{sen} \xi_{i-1} \text{sen} \sigma_{i-1} \hat{z}_i \cdot \hat{y} + \text{sen} \eta_i \cos \tau_i \text{sen} \xi_{i-1} \cos \sigma_{i-1} \hat{x}_i \cdot \hat{x}_{i-1} \\
&\quad + \text{sen} \eta_i \text{sen} \tau_i \text{sen} \xi_{i-1} \cos \sigma_{i-1} \hat{z}_i \cdot \hat{y} + \hat{x}_{i-1} + \text{sen} \eta_i \cos \tau_i \text{sen} \xi_{i-1} \text{sen} \sigma_{i-1} \hat{x}_i \cdot \hat{y} \\
&\quad + \text{sen} \eta_i \text{sen} \tau_i \text{sen} \xi_{i-1} \text{sen} \sigma_{i-1} \hat{y} \cdot \hat{y} - \text{sen} \eta_i \cos \tau_i \cos \xi_{i-1} \hat{x}_i \cdot \hat{z}_{i-1} \\
&\quad - \text{sen} \eta_i \text{sen} \tau_i \cos \xi_{i-1} \hat{y} \cdot \hat{z}_{i-1} \\
&= -\cos \eta_i \cos \xi_i \cos \alpha_i + \cos \eta_i \text{sen} \xi_{i-1} \cos \sigma_{i-1} \text{sen} \alpha_i \\
&\quad + \text{sen} \eta_i \cos \tau_i \text{sen} \xi_{i-1} \cos \sigma_{i-1} \cos \alpha_i + \text{sen} \eta_i \text{sen} \tau_i \text{sen} \xi_{i-1} \text{sen} \sigma_{i-1} \\
&\quad + \text{sen} \eta_i \cos \tau_i \cos \xi_{i-1} \text{sen} \alpha_i \\
&= -\cos \eta_i \cos \xi_i \cos \alpha_i + \text{sen} \alpha_i (\cos \eta_i \text{sen} \xi_{i-1} \cos \sigma_{i-1} + \text{sen} \eta_i \cos \tau_i \cos \xi_{i-1}) \\
&\quad + \text{sen} \xi_{i-1} \text{sen} \eta_i (\cos \eta_i \cos \sigma_{i-1} \cos \alpha_i + \text{sen} \tau_i \text{sen} \sigma_{i-1}), \tag{2.10}
\end{aligned}$$

com $i = 1, 2, 3$, onde os produtos internos são dados por:

$$\begin{aligned}
\hat{z}_i \cdot \hat{z}_{i-1} &= \cos \alpha_i, \quad \hat{z}_i \cdot \hat{y} = 0, \quad \hat{z}_i \cdot \hat{x}_{i-1} = \text{sen} \alpha_i, \\
\hat{x}_i \cdot \hat{x}_{i-1} &= \cos \alpha_i, \quad \hat{x}_i \cdot \hat{y} = 0, \quad \hat{z}_{i-1} \cdot \hat{x}_i = -\text{sen} \alpha_i.
\end{aligned}$$

Para a eliminação do termo σ_{i-1} da equação (2.10), escrevemos as equações (2.10) como séries de Fourier dupla, isto é,

$$0 = a_i + b_i \cos \sigma_{i-1} + c_i \cos \tau_i + d_i \cos \sigma_{i-1} \cos \tau_i + e_i \text{sen} \sigma_{i-1} \text{sen} \tau_i, \tag{2.11}$$

onde os coeficientes são:

$$\begin{aligned}
a_i &= -\cos \theta_i - \cos \eta_i \cos \xi_{i-1} \cos \alpha_i, \\
b_i &= \text{sen} \alpha_i \text{sen} \xi_{i-1} \cos \eta_i, \\
c_i &= \text{sen} \alpha_i \cos \xi_{i-1} \text{sen} \eta_i, \\
d_i &= \cos \alpha_i \text{sen} \xi_{i-1} \text{sen} \eta_i, \\
e_i &= \text{sen} \xi_{i-1} \text{sen} \eta_i.
\end{aligned}$$

Definindo

$$w_i := \text{tg} \frac{\sigma_i}{2}, \quad u_i := \text{tg} \frac{\tau_i}{2}, \tag{2.12}$$

e usando as fórmulas de arco metade, obtemos:

$$\cos \tau = \frac{1 - u^2}{1 + u^2}, \quad \text{sen} \tau = \frac{2u}{1 + u^2}, \quad u = \text{tg} \frac{\tau}{2}. \tag{2.13}$$

Isso implica que podemos reescrever as equações (2.11) em um sistema de três equações biquadráticas em w_i, u_i , onde $i = 1, 2, 3$,

$$0 = a_i + b_i \frac{1 - w_{i-1}^2}{1 + w_{i-1}^2} + c_i \frac{1 - u_i^2}{1 + u_i^2} + d_i \frac{1 - w_{i-1}^2}{1 + w_{i-1}^2} \cdot \frac{1 - u_i^2}{1 + u_i^2} + e_i \frac{2w_{i-1}}{1 + w_{i-1}^2} \cdot \frac{2u_i}{1 + u_i^2}, \quad (2.14)$$

ou, equivalentemente,

$$\begin{aligned} 0 = & a_i(1 + w_{i-1}^2)(1 + u_i^2) + b_i(1 - w_{i-1}^2)(1 + u_i^2) \\ & + c_i(1 + w_{i-1}^2)(1 - u_i^2) + d_i(1 - w_{i-1}^2)(1 - u_i^2) + 4e_i w_{i-1}^2 u_i. \end{aligned} \quad (2.15)$$

Expandindo e reagrupando, obtemos

$$A_i w_{i-1}^2 + B_i w_{i-1}^2 + C_i w_{i-1} u_i + D_i u_i^2 + E_i = 0, \quad (2.16)$$

onde

$$\begin{aligned} A_i &= a_i - b_i - c_i + d_i = -\cos \theta_i - \cos(\alpha_i - \xi_{i-1} - \eta_i), \\ B_i &= a_i - b_i + c_i - d_i = -\cos \theta_i - \cos(\alpha_i + \xi_{i-1} - \eta_i), \\ C_i &= 4e_i = 4\text{sen} \xi_{i-1} \text{sen} \eta_i, \\ D_i &= a_i + b_i - c_i - d_i = -\cos \theta_i - \cos(\alpha_i - \xi_{i-1} + \eta_i), \\ E_i &= a_i + b_i + c_i + d_i = -\cos \theta_i - \cos(\alpha_i + \xi_{i-1} + \eta_i). \end{aligned}$$

Utilizando a equação (2.6) para eliminar w_{i-1} da equação (2.16), obtemos

$$w_i = \text{tg}\left(\frac{\tau_i}{2} + \frac{\delta_i}{2}\right) = \frac{\text{tg}\left(\frac{\tau_i}{2}\right) + \text{tg}\left(\frac{\delta_i}{2}\right)}{1 - \text{tg}\left(\frac{\delta_i}{2}\right)\text{tg}\left(\frac{\tau_i}{2}\right)} = \frac{u_i + \Delta_i}{1 - \Delta_i u_i}, \quad (2.17)$$

onde introduzimos $\Delta = \text{tg}\left(\frac{\delta}{2}\right)$.

A equação (2.16) é, então, reescrita como

$$A_i \left(\frac{u_{i-1} + \Delta_{i-1}}{1 - \Delta_{i-1} u_{i-1}} \right)^2 u_i^2 + B_i \left(\frac{u_{i-1} + \Delta_{i-1}}{1 - \Delta_{i-1} u_{i-1}} \right)^2 + C_i \frac{u_{i-1} + \Delta_{i-1}}{1 - \Delta_{i-1} u_{i-1}} u_i + D_i u_i^2 + E_i = 0.$$

Multiplicando a equação acima por $(1 - \Delta_{i-1} u_{i-1})^2$ e reagrupando, obtemos:

$$\begin{aligned} 0 &= A_i (u_{i-1} + \Delta_{i-1})^2 u_i^2 + B_i (u_{i-1} + \Delta_{i-1})^2 + C_i (u_{i-1} + \Delta_{i-1}) (1 - \Delta_{i-1} u_{i-1}) u_i \\ &+ D_i u_i^2 (1 - \Delta_{i-1} u_{i-1})^2 + E_i (1 - \Delta_{i-1} u_{i-1})^2 \\ &= A_i (u_{i-1}^2 + 2u_{i-1} \Delta_{i-1} + \Delta_{i-1}^2) u_i^2 + B_i (u_{i-1}^2 + 2u_{i-1} \Delta_{i-1} + \Delta_{i-1}^2) \\ &+ C_i (u_{i-1} - \Delta_{i-1} u_{i-1}^2 + \Delta_{i-1} - \Delta_{i-1}^2 u_{i-1}) u_i + D_i u_i^2 (1 - 2\Delta_{i-1} u_{i-1} + \Delta_{i-1} u_{i-1}^2) \\ &+ E_i (1 - 2\Delta_{i-1} u_{i-1} + \Delta_{i-1} u_{i-1}^2). \end{aligned}$$

Como

$$\Delta = \frac{\text{sen} \delta}{1 + \cos \delta} \text{ e } \Delta^2 = \frac{1 - \cos \delta}{1 + \cos \delta}, \quad (2.18)$$

a equação acima pode ser reescrita como

$$\begin{aligned}
0 &= A_i(u_{i-1}^2 + 2u_{i-1}\frac{\text{sen}\delta_{i-1}}{1 + \cos\delta_{i-1}} + \frac{1 - \cos\delta_{i-1}}{1 + \cos\delta_{i-1}})u_i^2 + B_i(u_{i-1}^2 + 2u_{i-1}\frac{\text{sen}\delta_{i-1}}{1 + \cos\delta_{i-1}} \\
&\quad + \frac{1 - \cos\delta_{i-1}}{1 + \cos\delta_{i-1}}) + C_i(u_{i-1} + u_{i-1}^2\frac{\text{sen}\delta_{i-1}}{1 + \cos\delta_{i-1}} + \frac{\text{sen}\delta_{i-1}}{1 + \cos\delta_{i-1}} - \frac{1 - \cos\delta_{i-1}}{1 + \cos\delta_{i-1}}u_{i-1})u_i \\
&\quad + D_i(1 - 2u_{i-1}\frac{\text{sen}\delta_{i-1}}{1 + \cos\delta_{i-1}} + u_{i-1}^2\frac{1 - \cos\delta_{i-1}}{1 + \cos\delta_{i-1}})u_i^2 + E_i(1 - 2u_{i-1}\frac{\text{sen}\delta_{i-1}}{1 + \cos\delta_{i-1}} \\
&\quad + u_{i-1}^2\frac{1 - \cos\delta_{i-1}}{1 + \cos\delta_{i-1}}) \\
&= A_i(u_{i-1}^2\frac{1 + \cos\delta_{i-1}}{2} + u_{i-1}\text{sen}\delta_{i-1} + \frac{1 - \cos\delta_{i-1}}{2})u_i^2 + B_i(u_{i-1}^2\frac{1 + \cos\delta_{i-1}}{2} \\
&\quad + u_{i-1}\text{sen}\delta_{i-1} + \frac{1 - \cos\delta_{i-1}}{2}) + C_i(u_{i-1}\cos\delta_{i-1} - u_{i-1}^2\frac{\text{sen}\delta_{i-1}}{2} + \frac{\text{sen}\delta_{i-1}}{2})u_i \\
&\quad + D_i(\frac{1 + \cos\delta_{i-1}}{2} - u_{i-1}\text{sen}\delta_{i-1} + u_{i-1}^2\frac{1 - \cos\delta_{i-1}}{2}) + E_i(\frac{1 + \cos\delta_{i-1}}{2} - u_{i-1}\text{sen}\delta_{i-1} \\
&\quad + u_{i-1}^2\frac{1 - \cos\delta_{i-1}}{2}).
\end{aligned}$$

Agora, vamos reagrupar os termos em função da potência das variáveis u_i e u_{i-1} . Iniciamos com a potência zero para as duas variáveis:

$$\begin{aligned}
B_i(\frac{1 - \cos\delta_{i-1}}{2}) + E_i(\frac{1 + \cos\delta_{i-1}}{2}) &= \frac{B_i}{2} - \frac{\cos\delta_{i-1}}{2}B_i + \frac{E_i}{2} + \frac{\cos\delta_{i-1}}{2}E_i \\
&= \frac{(-\cos\theta_i - \cos(\alpha_i + \xi_{i-1} - \eta_i))}{2} + \frac{(-\cos\theta_i - \cos(\alpha_i + \xi_{i-1} + \eta_i))}{2} \\
&\quad - \cos\delta_{i-1}\frac{(-\cos\theta_i - \cos(\alpha_i + \xi_{i-1} - \eta_i))}{2} \\
&\quad + \cos\delta_{i-1}\frac{(-\cos\theta_i - \cos(\alpha_i + \xi_{i-1} + \eta_i))}{2} \\
&= -\cos\theta_i - \frac{\cos(\alpha_i + \xi_{i-1} - \eta_i)}{2} - \frac{\cos(\alpha_i + \xi_{i-1} + \eta_i)}{2} \\
&\quad + \frac{\cos\delta_{i-1}\cos(\alpha_i + \xi_{i-1} - \eta_i)}{2} - \frac{\cos\delta_{i-1}\cos(\alpha_i + \xi_{i-1} + \eta_i)}{2} \\
&= -\cos\theta_i - \frac{\cos(\alpha_i + \xi_{i-1})\cos\eta_i}{2} - \frac{\text{sen}(\alpha_i + \xi_{i-1})\text{sen}\eta_i}{2} \\
&\quad + \frac{\text{sen}(\alpha_i + \xi_{i-1})\text{sen}\eta_i}{2} - \frac{\cos(\alpha_i + \xi_{i-1})\cos\eta_i}{2} \\
&\quad + \frac{\cos\delta_{i-1}(\cos(\alpha_i + \xi_{i-1})\cos\eta_i + \text{sen}(\alpha_i + \xi_{i-1})\text{sen}\eta_i)}{2} \\
&\quad - \frac{\cos\delta_{i-1}(\cos(\alpha_i + \xi_{i-1})\cos\eta_i - \text{sen}(\alpha_i + \xi_{i-1})\text{sen}\eta_i)}{2} \\
&= -\cos\theta_i - \cos(\alpha_i + \xi_{i-1})\cos\eta_i + \cos\delta_{i-1}\text{sen}(\alpha_i + \xi_{i-1})\text{sen}\eta_i.
\end{aligned}$$

Considerando os termos com somente u_i :

$$C_i\frac{\text{sen}\delta_{i-1}}{2} = \frac{4\text{sen}\delta_{i-1}\text{sen}\xi_{i-1}\text{sen}\eta_i}{2} = 2\text{sen}\delta_{i-1}\text{sen}\xi_{i-1}\text{sen}\eta_i.$$

Considerando os termos com somente u_{i-1} :

$$\begin{aligned}
B_i \text{sen} \delta_{i-1} - E_i \text{sen} \delta_{i-1} &= (B_i - E_i) \text{sen} \delta_{i-1} \\
&= (-\cos \theta_i - \cos(\alpha_i + \xi_{i-1} - \eta_i) + \cos \theta_i + \cos(\alpha_i + \xi_{i-1} + \eta_i)) \text{sen} \delta_{i-1} \\
&= (-\cos(\alpha_i + \delta_{i-1}) \cos \eta_i - \text{sen}(\alpha_i + \xi_{i-1}) \text{sen} \eta_i \\
&+ \cos(\alpha_i + \xi_{i-1}) \cos \eta_i - \text{sen}(\alpha_i + \xi_{i-1}) \text{sen} \eta_i) \text{sen} \delta_{i-1} \\
&= -2 \text{sen}(\alpha_i + \xi_{i-1}) \text{sen} \eta_i \text{sen} \delta_{i-1}.
\end{aligned}$$

Considerando os termos $u_{i-1}u_i$:

$$C_i \left(\frac{1 + \cos \delta_{i-1}}{2} - \frac{11 \cos \delta_{i-1}}{2} \right) = C_i \cos \delta_{i-1} = 4 \text{sen} \xi_{i-1} \text{sen} \eta_i \text{sen} \delta_{i-1}.$$

Considerando os termos u_{i-1}^2 :

$$\begin{aligned}
B_i \frac{1 + \cos \delta_{i-1}}{2} + E_i \frac{1 - \cos \delta_{i-1}}{2} &= \frac{B_i}{2} + \frac{E_i}{2} + \frac{B_i \cos \delta_{i-1}}{2} - \frac{E_i \cos \delta_{i-1}}{2} \\
&= \frac{\cos \theta_i}{2} - \frac{\cos(\alpha_i + \xi_{i-1} - \eta_i)}{2} - \frac{\cos \theta_i}{2} - \frac{\cos(\alpha_i + \xi_{i-1} + \eta_i)}{2} \\
&- \frac{\cos \theta_i \cos \delta_{i-1}}{2} - \frac{\cos(\alpha_i + \xi_{i-1} - \eta_i) \cos \delta_{i-1}}{2} \\
&+ \frac{\cos \theta_i \cos \delta_{i-1}}{2} + \frac{\cos(\alpha_i + \xi_{i-1} + \eta_i) \cos \delta_{i-1}}{2} \\
&= -\cos \theta_i - \frac{\cos(\alpha_i + \xi_{i-1}) \cos \eta_i}{2} - \frac{\text{sen}(\alpha_i + \xi_{i-1}) \text{sen} \eta_i}{2} \\
&- \frac{\cos(\alpha_i + \xi_{i-1}) \cos \eta_i}{2} + \frac{\text{sen}(\alpha_i + \xi_{i-1}) \text{sen} \eta_i}{2} \\
&- \frac{\cos(\alpha_i + \xi_{i-1}) \cos \eta_i \cos \delta_{i-1}}{2} \\
&- \frac{\text{sen}(\alpha_i + \xi_{i-1}) \text{sen} \eta_i \cos \delta_{i-1}}{2} \\
&+ \frac{\cos(\alpha_i + \xi_{i-1}) \cos \eta_i \cos \delta_{i-1}}{2} \\
&- \frac{\text{sen}(\alpha_i + \xi_{i-1}) \text{sen} \eta_i \cos \delta_{i-1}}{2} \\
&= -\cos \theta_i - \cos(\alpha_i + \xi_{i-1}) \cos \eta_i - \text{sen}(\alpha_i + \xi_{i-1}) \text{sen} \eta_i \cos \delta_{i-1}.
\end{aligned}$$

Considerando os termos u_i^2 :

$$\begin{aligned}
A_i \frac{1 - \cos \delta_{i-1}}{2} + D_i \frac{1 + \cos \delta_{i-1}}{2} &= \frac{A_i}{2} + \frac{D_i}{2} - \frac{A_i \cos \delta_{i-1}}{2} + \frac{D_i \cos \delta_{i-1}}{2} \\
&= -\frac{\cos \theta_i}{2} - \frac{\cos(\alpha_i - \xi_{i-1} - \eta_i)}{2} - \frac{\cos \theta_i}{2} - \frac{\cos(\alpha_i - \xi_{i-1} + \eta_i)}{2} \\
&+ \frac{\cos \theta_i \cos \delta_{i-1}}{2} - \frac{\cos \theta_i \cos \delta_{i-1}}{2} + \frac{\cos(\alpha_i - \xi_{i-1} - \eta_i) \cos \delta_{i-1}}{2} \\
&- \frac{\cos(\alpha_i - \xi_{i-1} + \eta_i) \cos \delta_{i-1}}{2} \\
&= -\cos \theta_i - \frac{\cos(\alpha_i - \xi_{i-1}) \cos \eta_i}{2} - \frac{\text{sen}(\alpha_i - \xi_{i-1}) \text{sen} \eta_i}{2} \\
&- \frac{\cos(\alpha_i - \xi_{i-1}) \cos \eta_i}{2} + \frac{\text{sen}(\alpha_i - \xi_{i-1}) \text{sen} \eta_i}{2} \\
&+ \frac{\cos(\alpha_i - \xi_{i-1}) \cos \eta_i \cos \delta_{i-1}}{2} + \frac{\text{sen}(\alpha_i - \xi_{i-1}) \text{sen} \eta_i \cos \delta_{i-1}}{2} \\
&- \frac{\cos(\alpha_i - \xi_{i-1}) \cos \eta_i \cos \delta_{i-1}}{2} + \frac{\text{sen}(\alpha_i - \xi_{i-1}) \text{sen} \eta_i \cos \delta_{i-1}}{2} \\
&= -\cos \theta_i - \cos(\alpha_i - \xi_{i-1}) \cos \eta_i + \text{sen}(\alpha_i - \xi_{i-1}) \text{sen} \eta_i \cos \delta_{i-1}.
\end{aligned}$$

Considerando os termos $u_{i-1}^2 u_i$:

$$-C_i \frac{\text{sen} \delta_{i-1}}{2} = -\frac{4 \text{sen} \xi_{i-1} \text{sen} \eta_i \text{sen} \delta_{i-1}}{2} = -2 \text{sen} \xi_{i-1} \text{sen} \eta_i \text{sen} \delta_{i-1}$$

Considerando os termos $u_{i-1} u_i^2$:

$$\begin{aligned}
A_i \text{sen} \delta_{i-1} - D_i \text{sen} \delta_{i-1} &= (A_i - D_i) \text{sen} \delta_{i-1} \\
&= (-\cos \theta_i - \cos(\alpha_i - \xi_{i-1} - \eta_i) + \cos \theta_i + \cos(\alpha_i - \xi_{i-1} + \eta_i)) \text{sen} \delta_{i-1} \\
&= \text{sen} \delta_{i-1} (-\cos(\alpha_i - \xi_{i-1}) \cos \eta_i - \text{sen}(\alpha_i - \xi_{i-1}) \text{sen} \eta_i + \cos(\alpha_i - \xi_{i-1}) \cos \eta_i \\
&- \text{sen}(\alpha_i - \xi_{i-1}) \text{sen} \eta_i) \\
&= -2 \text{sen} \delta_{i-1} \text{sen}(\alpha_i - \xi_{i-1}) \text{sen} \eta_i.
\end{aligned}$$

Considerando os termos $u_{i-1}^2 u_i^2$:

$$\begin{aligned}
A_i \frac{1 + \cos \delta_{i-1}}{2} + D_i \frac{1 - \cos \delta_{i-1}}{2} &= \frac{A_i}{2} + \frac{D_i}{2} + \frac{A_i \cos \delta_{i-1}}{2} - \frac{D_i \cos \delta_{i-1}}{2} \\
&= \frac{-\cos \theta_i}{2} - \frac{\cos \theta_i \cos \delta_{i-1}}{2} - \frac{\cos \theta_i}{2} + \frac{\cos \theta_i \cos \delta_{i-1}}{2} \\
&\quad - \frac{\cos(\alpha_i - \xi_{i-1} - \eta_i)}{2} - \frac{\cos(\alpha_i - \xi_{i-1} + \eta_i)}{2} - \frac{\cos(\alpha_i - \xi_{i-1} - \eta_i) \cos \delta_{i-1}}{2} \\
&\quad + \frac{\cos(\alpha_i - \xi_{i-1} + \eta_i) \cos \delta_{i-1}}{2} \\
&= -\cos \theta_i - \frac{\cos(\alpha_i - \xi_{i-1}) \cos \eta_i}{2} - \frac{\text{sen}(\alpha_i - \xi_{i-1}) \text{sen} \eta_i}{2} \\
&\quad - \frac{\cos(\alpha_i - \xi_{i-1}) \cos \eta_i}{2} + \frac{\text{sen}(\alpha_i - \xi_{i-1}) \text{sen} \eta_i}{2} \\
&\quad - \frac{\cos(\alpha_i - \xi_{i-1}) \cos \eta_i \cos \delta_{i-1}}{2} - \frac{\text{sen}(\alpha_i - \xi_{i-1}) \text{sen} \eta_i \cos \delta_{i-1}}{2} \\
&\quad + \frac{\cos(\alpha_i - \xi_{i-1}) \cos \eta_i \cos \delta_{i-1}}{2} - \frac{\text{sen}(\alpha_i - \xi_{i-1}) \text{sen} \eta_i \cos \delta_{i-1}}{2} \\
&= -\cos \theta_i - \cos(\alpha_i - \xi_{i-1}) \cos \eta_i - \text{sen}(\alpha_i - \xi_{i-1}) \text{sen} \eta_i \cos \delta_{i-1}.
\end{aligned}$$

Agora, podemos reescrever as equações (2.11) como equações biquadráticas em $u_i = \text{tg} \frac{\tau_i}{2}$:

$$P_1(u_3, u_1) := \sum_{j,k=0}^2 p_{jk}^{(1)} u_3^j u_1^k = 0, \quad (2.19)$$

$$P_2(u_1, u_2) := \sum_{j,k=0}^2 p_{jk}^{(2)} u_1^j u_2^k = 0, \quad (2.20)$$

$$P_3(u_2, u_3) := \sum_{j,k=0}^2 p_{jk}^{(3)} u_2^j u_3^k = 0, \quad (2.21)$$

onde os coeficientes $p_{jk}^{(1)}, p_{jk}^{(2)}, p_{jk}^{(3)}$ são definidos em termos dos ângulos fixos $\alpha_i, \theta_i, \xi_{i-1}$ e δ_{i-1} , da seguinte forma:

$$\begin{aligned}
p_{22}^{(i)} &= -\cos \theta_i - \cos \xi_{i-1} \cos(\alpha_i - \eta_i) - \cos \delta_{i-1} \text{sen} \xi_{i-1} \text{sen}(\alpha_i - \eta_i) \\
p_{21}^{(i)} &= -2 \text{sen} \xi_{i-1} \text{sen} \eta_i \text{sen} \delta_{i-1} \\
p_{20}^{(i)} &= -\cos \theta_i - \cos(\alpha_i + \xi_{i-1}) \cos \eta_i - \text{sen}(\alpha_i + \xi_{i-1}) \text{sen} \eta_i \cos \delta_{i-1} \\
p_{12}^{(i)} &= -2 \text{sen} \delta_{i-1} \text{sen}(\alpha_i - \xi_{i-1}) \text{sen} \eta_i \\
p_{11}^{(i)} &= 4 \cos \delta_{i-1} \text{sen} \xi_{i-1} \text{sen} \eta_i \\
p_{10}^{(i)} &= -2 \text{sen}(\alpha_i + \xi_{i-1}) \text{sen} \eta_i \text{sen} \delta_{i-1} \\
p_{02}^{(i)} &= -\cos \theta_i - \cos(\alpha_i - \xi_{i-1}) \cos \eta_i + \text{sen}(\alpha_i - \xi_{i-1}) \text{sen} \eta_i \cos \delta_{i-1} \\
p_{01}^{(i)} &= 2 \text{sen} \delta_{i-1} \text{sen} \xi_{i-1} \text{sen} \eta_i \\
p_{00}^{(i)} &= -\cos \theta_i - \cos \xi_{i-1} \cos(\alpha_i + \eta_i) + \cos \delta_{i-1} \text{sen} \eta_i \text{sen}(\alpha_i + \xi_{i-1}).
\end{aligned}$$

Finalmente, as equações (2.19), (2.20), (2.21) são reescritas como

$$\begin{aligned} P_1(u_3, u_1) &= \sum_{j=0}^2 \left(\sum_{k=0}^2 p_{jk}^{(1)} u_3^j \right) u_1^k = \sum_{j=0}^2 L_j u_1^j, \\ P_2(u_1, u_2) &= \sum_{k=0}^2 \left(\sum_{j=0}^2 p_{jk}^{(2)} u_2^k \right) u_1^j = \sum_{k=0}^2 M_k u_1^k, \\ P_3(u_2, u_3) &= \sum_{j=0}^2 \left(\sum_{k=0}^2 p_{jk}^{(3)} u_3^k \right) u_2^j = \sum_{j=0}^2 N_j u_2^j, \end{aligned} \quad (2.22)$$

onde

$$L_j := L_j(u_3) := \sum_{k=0}^2 p_{jk}^{(1)} u_3^k, \quad M_k := M_k(u_2) := \sum_{j=0}^2 p_{jk}^{(2)} u_2^j, \quad \text{e} \quad N_j := N_j(u_3) := \sum_{k=0}^2 p_{jk}^{(3)} u_3^k.$$

Para encontrar uma solução comum entre as três equações acima, vamos precisar utilizar o Método da Resultante para Sistema de Polinômios (COUTSIAS et al., 2004). A resultante de um sistema de equações polinomiais é uma condição necessária e suficiente para existência de uma solução comum entre as equações do sistema. Resumidamente, para dois polinômios, digamos $F_n(u)$ e $F_m(u)$ de graus n e m , respectivamente, terem uma solução u em comum é necessário a existência de um fator em comum. Ou seja, é necessário que existam polinômios $g(u)$ e $h(u)$ de graus menores que $n - 1$ e $m - 1$, respectivamente, tais que

$$gF_m + hF_n = 0.$$

Dessa forma, obtemos um sistema de $m + n$ equações lineares homogêneas para determinar os coeficientes de g e h . A resultante é o determinante da matriz associada ao sistema. Vamos analisar este método para o caso de polinômios de segundo grau com somente uma variável. Sejam

$$\begin{aligned} f_1(u) &= a_2 u^2 + a_1 u + a_0, \\ f_2(u) &= b_2 u^2 + b_1 u + b_0. \end{aligned}$$

Supondo que estes polinômios possuem uma raiz em comum, digamos u^* , então os polinômios devem ser da forma

$$\begin{aligned} f_1(u) &= a_2(u - u^*)(u - u_1) = 0, \\ f_2(u) &= b_2(u - u^*)(u - u_2) = 0. \end{aligned}$$

Além disso, existem dois polinômios de grau 1, $g(u) = b_2(u - u_2)$ e $h(u) = -a_2(u - u_1)$, tais que

$$g(u)f_1(u) + h(u)f_2(u) = 0. \quad (2.23)$$

Como as raízes são desconhecidas, escrevemos

$$g(u) = g_1u + g_0, \quad h(u) = h_1(u) + h_0,$$

implicando que a equação (2.23) é reescrita como

$$(g_1u + g_0)(a_2u + a_1u + a_0) + (h_1u + h_0)(b_2u + b_1u + b_0) = 0.$$

Reagrupando em função das potências de u , obtemos

$$\begin{aligned} (g_1a_2 + h_1b_2)u^3 &+ (g_1a_1 + g_0a_2 + h_1b_1 + h_0b_2)u^2 \\ &+ (g_0a_1 + g_1a_0 + h_0b_1 + h_1b_0)u + (g_0a_0 + h_0b_0) = 0, \end{aligned}$$

gerando a forma matricial

$$\begin{pmatrix} g_1 & g_0 & h_1 & h_0 \end{pmatrix} \begin{pmatrix} a_2 & a_1 & a_0 & 0 \\ 0 & a_2 & a_1 & a_0 \\ b_2 & b_1 & b_0 & 0 \\ 0 & b_2 & b_1 & b_0 \end{pmatrix} \begin{pmatrix} u^3 \\ u^2 \\ u \\ 1 \end{pmatrix} = 0. \quad (2.24)$$

Para que os vetores da direita e da esquerda sejam não nulos simultaneamente, precisamos que a matriz dos coeficientes não tenha posto completo. Ou seja, precisamos que o determinante da matriz de coeficientes seja nulo, condição necessária e suficiente para que existam vetores não nulos.

Dessa forma, a resultante dos polinômios (2.19) e (2.20), cujo “anulamento” garante uma raiz em comum em u_1 , é dada pelo determinante

$$R_8(u_2, u_3) = \begin{vmatrix} L_2 & L_1 & L_0 & 0 \\ 0 & L_2 & L_1 & L_0 \\ M_2 & M_1 & M_0 & 0 \\ 0 & M_2 & M_1 & M_0 \end{vmatrix} = \begin{vmatrix} L_2 & L_0 \\ M_2 & M_0 \end{vmatrix}^2 - \begin{vmatrix} L_2 & L_1 \\ M_2 & M_1 \end{vmatrix} \begin{vmatrix} L_1 & L_0 \\ M_1 & M_0 \end{vmatrix}. \quad (2.25)$$

Como todos os elementos não nulos são produtos de duas equações quadráticas em u_2 e duas equações quadráticas em u_3 , a resultante é uma biquadrática nessas variáveis e possui a forma

$$R_8(u_2, u_3) = \sum_{j,k=0}^4 q_{jk} u_2^j u_3^k. \quad (2.26)$$

Observe que existem 25 termos q_{jk} , dados pelos produtos dos termos $a_{jk} := p_{jk}^{(1)}$ e $b_{jk} := p_{jk}^{(2)}$.

Para obter R_8 como uma equação biquadrática na variável u_2 , utilizamos as funções biquadráticas Q_j em u_3 :

$$R_8 = \sum_{j=0}^4 \left(\sum_{k=0}^4 q_{jk} u_3^k \right) u_2^j = \sum_{j=0}^4 Q_j u_2^j.$$

Agora, queremos eliminar a variável u_2 . Utilizando o método da resultante novamente, obtemos um polinômio de grau 16 na variável u_3 , dado por

$$R_{16} = \det(S),$$

onde a matriz S é

$$S(u_3) := \sum_{k=0}^4 S_k u_3^k = \begin{pmatrix} N_2 & N_1 & N_0 & 0 & 0 & 0 \\ 0 & N_2 & N_1 & N_0 & 0 & 0 \\ 0 & 0 & N_2 & N_1 & N_0 & 0 \\ 0 & 0 & 0 & N_2 & N_1 & N_0 \\ Q_4 & Q_3 & Q_2 & Q_1 & Q_0 & 0 \\ 0 & Q_4 & Q_3 & Q_2 & Q_1 & Q_0 \end{pmatrix}. \quad (2.27)$$

Dessa forma,

$$S_k := \begin{pmatrix} c_{2k} & c_{1k} & c_{0k} & 0 & 0 & 0 \\ 0 & c_{2k} & c_{1k} & c_{0k} & 0 & 0 \\ 0 & 0 & c_{2k} & c_{1k} & c_{0k} & 0 \\ 0 & 0 & 0 & c_{2k} & c_{1k} & c_{0k} \\ q_{4k} & q_{3k} & q_{2k} & q_{1k} & q_{0k} & 0 \\ 0 & q_{4k} & q_{3k} & q_{2k} & q_{1k} & q_{0k} \end{pmatrix},$$

onde $c_{ij} := p_{jk}^{(3)}$, com $c_{i3} = c_{i4} = 0$, $i = 0, 1, 2$.

O cálculo dos coeficientes do polinômio para uma aproximação da solução exige cálculos adicionais descritos abaixo. Prosseguimos com a expansão de Laplace da Equação (2.27) pelos menores, determinante de matrizes quadradas menores obtidas pela “exclusão” de uma ou mais linhas ou colunas, complementares de ordem 3. Primeiro, vamos rearranjar as linhas da matriz S através de operações fundamentais:

$$\begin{aligned} \det S &= \begin{vmatrix} N_2 & N_1 & N_0 & 0 & 0 & 0 \\ 0 & N_2 & N_1 & N_0 & 0 & 0 \\ 0 & 0 & N_2 & N_1 & N_0 & 0 \\ 0 & 0 & 0 & N_2 & N_1 & N_0 \\ Q_4 & Q_3 & Q_2 & Q_1 & Q_0 & 0 \\ 0 & Q_4 & Q_3 & Q_2 & Q_1 & Q_0 \end{vmatrix} \\ &= - \begin{vmatrix} N_2 & N_1 & N_0 & 0 & 0 & 0 \\ Q_4 & Q_3 & Q_2 & Q_1 & Q_0 & 0 \\ 0 & 0 & N_2 & N_1 & N_0 & 0 \\ 0 & N_2 & N_1 & N_0 & 0 & 0 \\ 0 & Q_4 & Q_3 & Q_2 & Q_1 & Q_0 \\ 0 & 0 & 0 & N_2 & N_1 & N_0 \end{vmatrix} \\ &= - \sum_{1 \leq i_1 < i_2 < i_3 \leq 6} (-1)^{i_1+i_2+i_3} \det S(1, 2, 3; i_1, i_2, i_3) \times \det S(4, 5, 6; i_4, i_5, i_6), \end{aligned}$$

onde $S(1, 2, 3; i_1, i_2, i_3)$ é a submatriz 3×3 de S formada pelos elementos nas linhas 1, 2, 3 e colunas $i_1 < i_2 < i_3$. Também, $i_4 < i_5 < i_6$ e i_4, i_5, i_6 diferentes de i_1, i_2, i_3 . Agora, introduzimos a submatriz 3×5

$$P := \begin{pmatrix} N_2 & N_1 & N_0 & 0 & 0 & 0 \\ Q_4 & Q_3 & Q_2 & Q_1 & Q_0 & 0 \\ 0 & 0 & N_2 & N_1 & N_0 & 0 \end{pmatrix}$$

e os minors 3×3 , denominados $T(i, j, k)$, são formados pelas colunas i, j, k de P . Então, a expansão de Laplace de S em termos dos minors sobre as linhas 1, 2 e 3 e seu complemento, as linhas 4, 5 e 6, pode ser escrita simplesmente da forma:

$$\begin{aligned} \det S &= T(1, 2, 3)T(3, 4, 5) - T(1, 2, 4)T(2, 4, 5) + T(1, 2, 5)T(2, 3, 5) \\ &\quad + T(1, 3, 4)T(1, 4, 5) - T(1, 3, 5)T(1, 3, 5) + T(1, 4, 5)T(1, 2, 5) \end{aligned}$$

O cálculo da resultante é feito calculando os nove valores de $T(i, j, k)$ acima. Note que esses termos são somas de produtos da forma $N_\alpha N_\beta Q_\gamma$. Ou seja, eles são polinômios na variável u_3 e possuem grau 8. Trabalhando com esses produtos, obtemos as seguintes igualdades:

$$T(1, 2, 3) = \begin{vmatrix} N_2 & N_1 & N_0 \\ Q_4 & Q_3 & Q_2 \\ 0 & 0 & N_2 \end{vmatrix} = N_2 \begin{vmatrix} N_2 & N_1 \\ Q_4 & Q_3 \end{vmatrix},$$

$$T(1, 2, 4) = \begin{vmatrix} N_2 & N_1 & N_0 \\ Q_4 & Q_3 & Q_2 \\ 0 & 0 & N_1 \end{vmatrix} = N_1 \begin{vmatrix} N_2 & N_1 \\ Q_4 & Q_3 \end{vmatrix},$$

$$T(1, 2, 5) = \begin{vmatrix} N_2 & N_1 & N_0 \\ Q_4 & Q_3 & Q_2 \\ 0 & 0 & N_0 \end{vmatrix} = N_0 \begin{vmatrix} N_2 & N_1 \\ Q_4 & Q_3 \end{vmatrix},$$

$$T(1, 3, 4) = \begin{vmatrix} N_2 & N_1 & N_0 \\ Q_4 & Q_2 & Q_1 \\ 0 & N_2 & N_1 \end{vmatrix} = -N_2 \begin{vmatrix} Q_2 & Q_1 \\ N_2 & N_1 \end{vmatrix} - Q_4 N_0 N_1,$$

$$T(1, 3, 5) = \begin{vmatrix} N_2 & N_1 & 0 \\ Q_4 & Q_2 & Q_0 \\ 0 & N_2 & N_0 \end{vmatrix} = -N_2 \begin{vmatrix} N_2 & N_0 \\ Q_2 & Q_0 \end{vmatrix} - Q_4 N_0^2,$$

$$T(1, 4, 5) = \begin{vmatrix} N_2 & 0 & 0 \\ Q_4 & Q_1 & Q_0 \\ 0 & N_1 & N_0 \end{vmatrix} = -N_2 \begin{vmatrix} N_1 & N_0 \\ Q_1 & Q_0 \end{vmatrix},$$

$$T(2, 3, 5) = \begin{vmatrix} N_1 & 0 & 0 \\ Q_3 & Q_2 & Q_0 \\ 0 & N_2 & N_0 \end{vmatrix} = -N_1 \begin{vmatrix} N_2 & N_0 \\ Q_2 & Q_0 \end{vmatrix} - N_0^2 Q_3,$$

$$T(2, 4, 5) = \begin{vmatrix} N_1 & 0 & 0 \\ Q_3 & Q_1 & Q_0 \\ 0 & N_1 & N_0 \end{vmatrix} = -N_1 \begin{vmatrix} N_1 & N_0 \\ Q_1 & Q_0 \end{vmatrix},$$

$$T(3, 4, 5) = \begin{vmatrix} N_0 & 0 & 0 \\ Q_2 & Q_1 & Q_0 \\ N_2 & N_1 & N_0 \end{vmatrix} = -N_0 \begin{vmatrix} N_1 & N_0 \\ Q_1 & Q_0 \end{vmatrix}.$$

Dessa forma, o cálculo da expressão do polinômio de grau 16 foi reduzido ao cálculo de somente quatro determinantes 2×2 distintos.

2.1.2 Algoritmo

Nesta seção, vamos apresentar a ideia de algoritmo utilizado no trabalho de (COUTSIAS et al., 2004). Como as equações polinomiais foram determinadas na seção anterior, vamos agora apresentar resumidamente como são encontrados os valores dos ângulos e distâncias utilizados na derivação acima e também qual o método utilizado para encontrar as raízes deste polinômio de grau 16.

Primeiramente, os coeficientes do polinômio são determinados a partir dos ângulos α_i, η_i, ξ_i e δ_i . Inicialmente, são determinados os ângulos η_3, ξ_1, δ_3 das coordenadas de $N_1, C_{\alpha 1}, C_{\alpha 3}$ e C_3 . Em seguida, todos os outros ângulos são calculados, a partir das distâncias e ângulos conhecidos. Dessa forma, utilizando a derivação apresentada anteriormente, todos os coeficientes do polinômio de grau 16 são determinados.

Em seguida, utiliza-se o Método de Sturm (HENRICI, 1974, Seção 6) para determinar o número de zeros reais em um intervalo arbitrário. Os intervalos são bisectados e refinados até todas as soluções estarem em intervalos distintos. Por fim, utilizando o método da secante, as soluções são refinadas.

Encontrando os zeros do polinômio calculamos, $u_3 = \text{tg} \frac{\tau_3}{2}$, depois é determinado $u_2 = \text{tg} \frac{\tau_2}{2}$, através da Equação (2.26), e, finalmente, $u_1 = \text{tg} \frac{\tau_1}{2}$, utilizando uma das Equações de (2.22). Como calculamos anteriormente δ_i , dos quatro átomos posicionados ou das distâncias conhecidas, utilizamos a Equação (2.6) e obtemos os valores de σ_i .

Após serem calculados todos os valores dos nossos ângulos, definimos o frame inicial. Isto é, posicionamos no espaço os átomos do corpo rígido $(C_{\alpha 1}, N_1, C_3, C_{\alpha 3})$, com o vetor unidade \hat{z}_3 sendo definido a partir das posições dos átomos $C_{\alpha 1}$ e $C_{\alpha 3}$. Em seguida, considere $C_{\alpha 3} \cdot \hat{z}_1$ com a única restrição de que o ângulo entre \hat{z}_1 e \hat{z}_3 é dado por α_1 . Conhecendo \hat{z}_1 e \hat{z}_3 , calculamos \hat{z}_2 através da expressão $\hat{z}_2 = -\hat{z}_1 - \hat{z}_3$. Prosseguimos utilizando a Equação (2.5) para determinar \hat{y} e \hat{x}_i , a partir dos valores de \hat{z}_i .

Por fim, as posições de todos os átomos são determinadas a partir de τ_i e σ_i . Primeiro, o frame referencial é definido. O vetor unitário \hat{z}_3 é determinado das coordenadas de $C_{\alpha 1}$ e $C_{\alpha 3}$. \hat{z}_1 é um conjunto arbitrário, exceto que o ângulo entre \hat{z}_1 e \hat{z}_3 é α_1 . Em seguida, $\hat{z}_2 = -\hat{z}_1 - \hat{z}_3$ é calculado. Posteriormente, \hat{y} e \hat{x}_i são calculados a partir de \hat{z}_i . Em seguida, \hat{r}_i^T e \hat{r}_i^σ ($i = 1, 2$) são obtidos, no frame referencial, de τ_i e σ_i usando as Equações (2.7). Todas as posições dos átomos são determinadas a partir desses vetores no frame referencial. Os vetores unitários que definem $\tau_3^{(0)}$ são determinados das coordenadas fixas de $N_1, C_{\alpha 1}, C_{\alpha 3}$ e C_3 . Todos os átomos são então rotacionados sobre \hat{z}_3 por um ângulo de $(\tau_3^{(0)} - \tau_3)$ para retornar ao frame original.

2.2 Generalização

O Problema do Loop Fechado é definido basicamente através de três rotações sobre os eixos $C_{\alpha i} - C_{\alpha i+1}$ ($i = 1, 2, 3$) e três restrições sobre estas rotações que asseguram que os ângulos de ligação entre as duas ligações rotacionáveis $N_i - C_{\alpha i}$ e $C_{\alpha i} - C_i$ sejam fixadas. Observe a ilustração na Figura 4. Note que a cadeia de átomos entre os átomos $C_{\alpha i}$ possui rotação rígida, ou seja, não possuem movimentos independentes. Desta forma, o problema é caracterizado completamente quando são dados os ângulos α_i entre os eixos de rotação, os ângulos ξ_i, η_i formados pelas ligações rotacionáveis de cada $C_{\alpha i}$, as arestas do triângulo $C_{\alpha i}, C_{\alpha i+1}, C_{\alpha i+2}$ e os ângulos diedrais δ_i . Como na derivação das equações do loop fechado não é utilizada qualquer informação sobre a estrutura da molécula entre os átomos $C_{\alpha i}$, o algoritmo descrito acima pode ser utilizado sem nenhuma modificação para problemas com mais estruturas rígidas entre os átomos C_α (veja a Figura 9). Isto é, os ângulos α_i, ξ_i, η_i e δ_i que definem o problema são obtidos da mesma forma como ilustrado nas Figuras 6 e 7.

2.3 Método Cyclic Coordinate Descent

Outro método encontrado na literatura para a solução do Problema do Loop Fechado em proteínas é chamado de CCD (Cyclic Coordinate Descent) (CANUTESCU; DUNBRACK, 2003). O CCD procura os loops através de um procedimento iterativo que modifica sequencialmente os ângulos diedrais (ϕ e ψ) (veja a Figura 10, retirada de (CANUTESCU; DUNBRACK, 2003)) de cada resíduo do *backbone* da proteína.

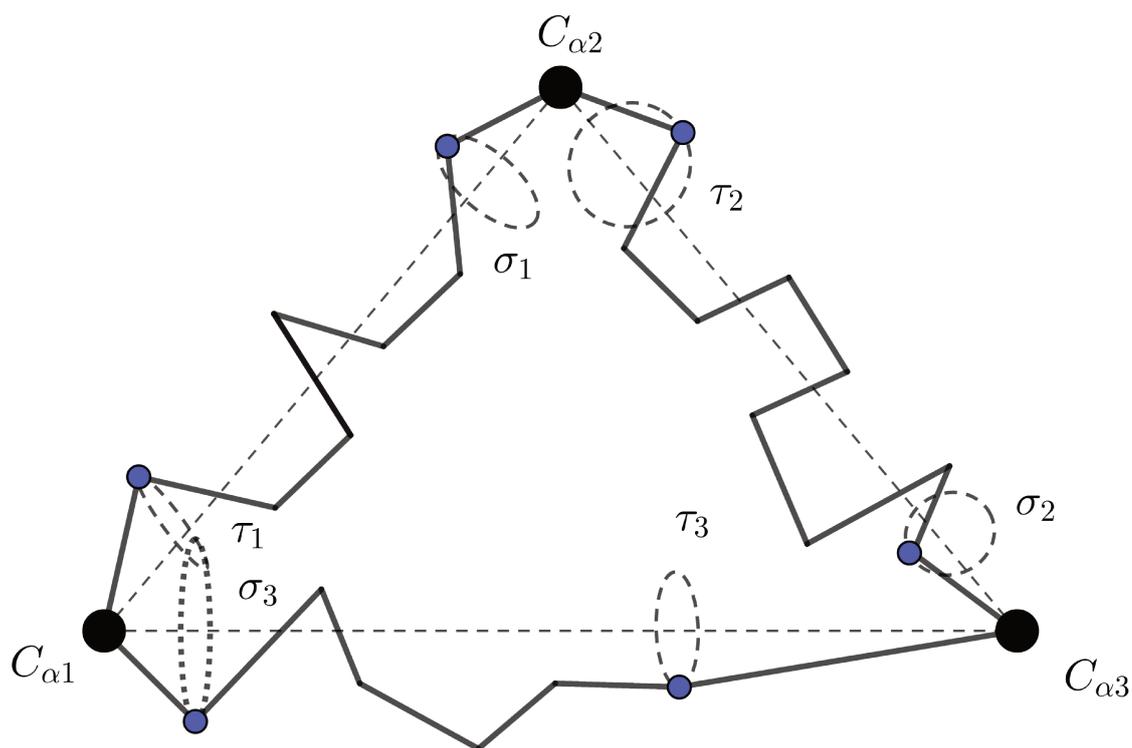


Figura 9 – Ilustração do Loop Fechado Generalizado.

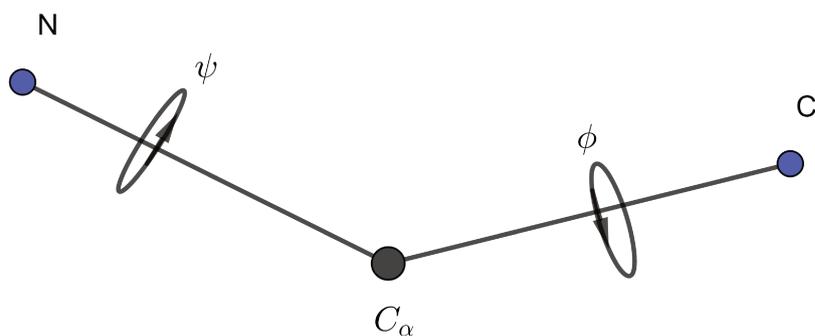


Figura 10 – Ilustração dos ângulos ϕ e ψ em um resíduo.

Definimos a N -âncora e a C -âncora para serem os N e C terminais residuais que suportam o loop e permanecem fixos durante todo o cálculo. Os resíduos do loop são enumerados de 0 à n , onde o 0-ésimo resíduo é o N -âncora e o n -ésimo resíduo é o C -âncora.

O método exige uma configuração inicial do loop para todos os resíduos, que pode ser obtida de qualquer fonte (obtida do banco de dados de proteínas ou a partir de valores aleatórios para os ângulos diedrais, comprimentos de ligações e ângulos de ligação no backbone). De toda forma, sempre teremos valores iniciais para todos ângulos diedrais, comprimentos de ligação e ângulos de ligação para todos os resíduos do loop.

Considere a configuração inicial do loop com os resíduos do 0 á n . O resíduo 0 do loop coincide com a N -âncora, enquanto que a posição do resíduo n não coincide com a posição fixada da C -âncora. O procedimento iterativo consiste em ajustar os ângulos diedrais (ϕ e ψ) em cada resíduo de modo que os átomos do resíduo n , (N , C_α e C) no backbone obtido, estejam sobrepostos aos átomos fixos da C -âncora. O progresso no processo de procura de loop é avaliado pelas distância entre os átomos do backbone em movimento no resíduo n do loop e sua posição desejada na C -âncora.

Considere a Figura 11, construída a partir de imagem disponível em (CANU-TESCU; DUNBRACK, 2003). Sejam F_1, F_2 e F_3 os vetores que representam as posições do C -âncora, isto é, as posições desejadas dos átomos do resíduo n . As posições dos átomos do resíduo n durante o processo são representadas por M_{01}, M_{02}, M_{03} e M_1, M_2, M_3 antes e depois, respectivamente, de uma mudança em um ângulo diedral de qualquer resíduo no loop. O eixo de rotação, contendo O_1, O_2, O_3 , é dado pelas direções das ligações correspondentes ao ângulo diedral que é modificado, onde O_1, O_2, O_3 são os pontos de referência no eixo de rotação para os átomos do resíduo n .

Para verificarmos a sobreposição entre as posições atual e desejada dos átomos, a soma dos quadrados das distâncias, S , deve ser minimizada:

$$S = \|\vec{F}_1\vec{M}_1\|^2 + \|\vec{F}_2\vec{M}_2\|^2 + \|\vec{F}_3\vec{M}_3\|^2, \quad (2.28)$$

onde

$$\vec{F}_i\vec{M}_i = \vec{O}_i\vec{M}_i - \vec{O}_i\vec{F}_i, \quad (2.29)$$

para $i = 1, 2, 3$. Definindo $\vec{O}_1\vec{M}_{01} = \vec{r}_1$, $\vec{O}_1\vec{F}_1 = \vec{f}_1$ e um sistema de coordenadas ortogonal local $\hat{r}_1, \hat{s}_1, \hat{\theta}$, onde \hat{r}_1 é o vetor unitário de \vec{r}_1 e $\hat{\theta}$ é o vetor unitário ao longo do eixo de rotação, podemos escrever

$$\vec{O}_1\vec{M}_1 = r_1 \cos \theta \hat{r}_1 + r_1 \sin \theta \hat{s}_1. \quad (2.30)$$

Substituindo a Equação (2.30) na Equação (2.29), obtemos:

$$\vec{F}_1\vec{M}_1 = r_1 \cos \theta \hat{r}_1 + r_1 \sin \theta \hat{s}_1 - \vec{f}_1 \equiv \vec{d}_1. \quad (2.31)$$

Utilizando um cálculo análogo, obtemos equações similares para o segundo e terceiro átomo do resíduo n . Seguimos com o cálculo do quadro das distâncias entre os

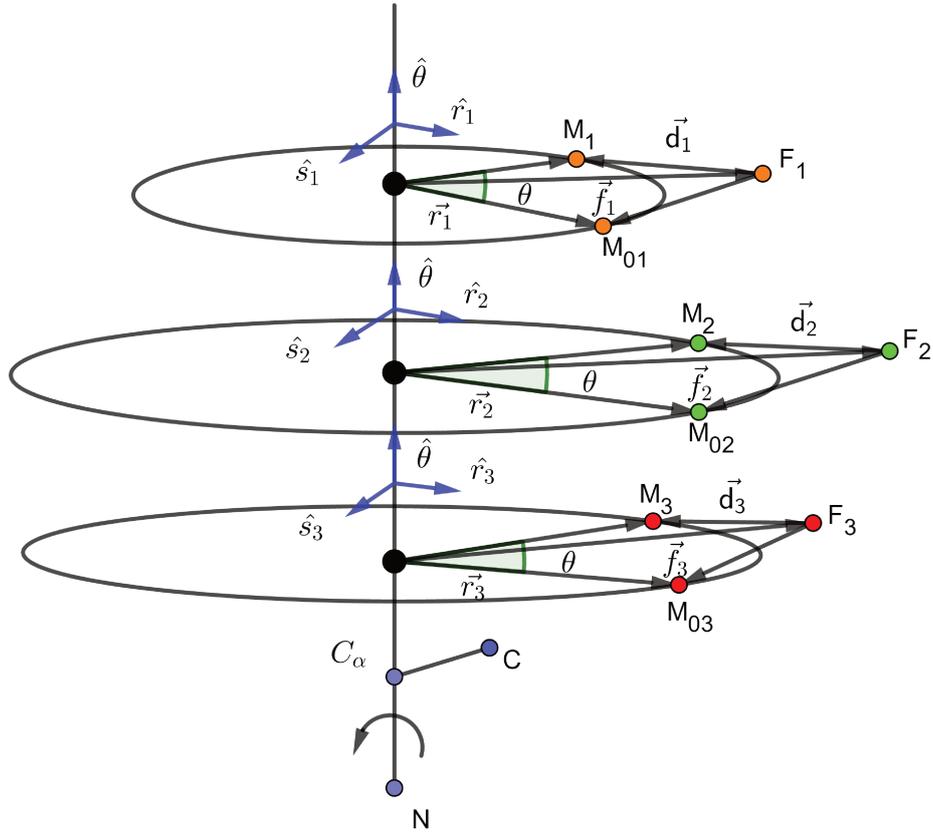


Figura 11 – Ilustração da avaliação dos loops encontrados pelo algoritmo CCD.

átomos;

$$\begin{aligned}
 \|\vec{d}_1\|^2 &= r_1^2 + f_1^2 - 2r_1 \cos \theta (\vec{f}_1 \cdot \hat{r}_1) - 2r_1 \sin \theta (\vec{f}_1 \cdot \hat{s}_1), \\
 \|\vec{d}_2\|^2 &= r_2^2 + f_2^2 - 2r_2 \cos \theta (\vec{f}_2 \cdot \hat{r}_2) - 2r_2 \sin \theta (\vec{f}_2 \cdot \hat{s}_2), \\
 \|\vec{d}_3\|^2 &= r_3^2 + f_3^2 - 2r_3 \cos \theta (\vec{f}_3 \cdot \hat{r}_3) - 2r_3 \sin \theta (\vec{f}_3 \cdot \hat{s}_3).
 \end{aligned} \tag{2.32}$$

A Equação (2.28) pode ser reescrita como

$$S = \|\vec{d}_1\|^2 + \|\vec{d}_2\|^2 + \|\vec{d}_3\|^2, \tag{2.33}$$

implicando em

$$\frac{dS}{d\theta} = \frac{d(\|\vec{d}_1\|^2)}{d\theta} + \frac{d(\|\vec{d}_2\|^2)}{d\theta} + \frac{d(\|\vec{d}_3\|^2)}{d\theta}, \tag{2.34}$$

onde

$$\frac{d(\|\vec{d}_i\|^2)}{d\theta} = 2r_i \sin \theta (\vec{f}_i \cdot \hat{r}_i) - 2r_i \cos \theta (\vec{f}_i \cdot \hat{s}_i), \tag{2.35}$$

para $i = 1, 2, 3$. Fixando $\frac{dS}{d\theta} = 0$, para obter os pontos críticos, obtemos o valor de $\tan \alpha$, onde α é o ângulo de rotação que produz os valores extremos para a soma dos quadrados das distâncias:

$$\operatorname{tg}\alpha = \frac{(\vec{f}_1 \cdot \hat{s}_1)r_1 + (\vec{f}_2 \cdot \hat{s}_2)r_2 + (\vec{f}_3 \cdot \hat{s}_3)r_3}{(\vec{f}_1 \cdot \hat{r}_1)r_1 + (\vec{f}_2 \cdot \hat{r}_2)r_2 + (\vec{f}_3 \cdot \hat{r}_3)r_3}. \quad (2.36)$$

Aplicando a função inversa da função tangente, obtemos dois valores para o ângulo α , distintos um do outro por exatamente π radianos. Escolhemos o valor de α que gera um valor positivo para a segunda derivada de S em relação θ , a qual é obtida das Equações (2.34) e (2.35). Entretanto, na prática, o valor de α é obtido por um caminho diferente, o qual é descrito a seguir.

Das equações (2.32) e (2.33), escrevemos S

$$S = a - b \cos \theta - c \operatorname{sen} \theta, \quad (2.37)$$

onde

$$\begin{aligned} a &= r_1^2 + r_2^2 + r_3^2 + f_1^2 + f_2^2 + f_3^2, \\ b &= 2(r_1(\vec{f}_1 \cdot \hat{r}_1) + r_2(\vec{f}_2 \cdot \hat{r}_2) + r_3(\vec{f}_3 \cdot \hat{r}_3)), \\ c &= 2(r_1(\vec{f}_1 \cdot \hat{s}_1) + r_2(\vec{f}_2 \cdot \hat{s}_2) + r_3(\vec{f}_3 \cdot \hat{s}_3)). \end{aligned} \quad (2.38)$$

Multiplicando os dois termos da Equação (2.37) por $\frac{\sqrt{b^2 + c^2}}{\sqrt{b^2 + c^2}}$ e definindo

$$\cos \alpha = \frac{b}{\sqrt{b^2 + c^2}}, \quad \operatorname{sen} \alpha = \frac{c}{\sqrt{b^2 + c^2}}, \quad (2.39)$$

obtemos

$$\begin{aligned} S &= a - \sqrt{b^2 + c^2} \cos \alpha \cos \theta - \sqrt{b^2 + c^2} \operatorname{sen} \alpha \operatorname{sen} \theta \\ &= a - \sqrt{b^2 + c^2} (\cos \alpha \cos \theta - \operatorname{sen} \alpha \operatorname{sen} \theta) \end{aligned} \quad (2.40)$$

$$= a - \sqrt{b^2 + c^2} \cos(\theta - \alpha). \quad (2.41)$$

Dessa forma, para $\theta = \alpha$, S assume um valor mínimo.

Na Figura 12, ilustramos as conformações encontradas pelo Algoritmo CCD para os loops. Em preto, ilustramos os átomos do primeiro e último resíduos cujo posições são consideradas conhecidas para utilização do Algoritmo CCD. Em azul, ilustramos as conformações do loop que são obtidas e consideradas soluções factíveis pelo Algoritmo CCD, o critério utilizado para que uma conformação seja factível é que as posições encontradas para os átomos do último resíduo coincidam com as posição já conhecidas para estes átomos. Em vermelho, ilustramos uma conformação considerada não factível, pois as posições encontradas para os átomos do último resíduo não coincidem com as posições já conhecidas para estes átomos.

3 O Problema do Loop Fechado em Geometria de Distâncias

Neste capítulo, vamos apresentar o método proposto no trabalho para encontrar as soluções do PLF. Na primeira seção, vamos descrever uma forma de criar instâncias para o Problema do Loop Fechado (PLF). Na segunda seção, vamos justificar o motivo de não existir uma ordem satisfazendo a Definição 2. Na terceira seção encontra-se o resultado principal do trabalho, nesta seção é descrita a estratégia para encontrar uma estimativa para uma distância intervalar à qual possibilita construir uma ordem para os átomos do PLF que satisfazem a Definição 3. Além disso, a ordem satisfazendo a Definição 3 é construída nesta seção. A seguir, na quarta seção, apresentamos e demonstramos uma releitura de resultados conhecidos para GD para o PLF. Por fim, na quinta seção, nós apresentamos uma generalização do PLF, já conhecida na literatura (COUTSIAS et al., 2004), juntamente com a estratégia para estimar uma distância intervalar para a generalização do PLF. Além disso, nós construímos uma ordem, satisfazendo a Definição 3, para a generalização do PLF.

3.1 Instâncias para o Problema

Nesta seção, vamos explicar detalhadamente o processo de geração de instâncias que representem o PLF. Estamos interessados em solucionar o problema para o caso fechado. Isto é, quando temos três corpos rígidos conectados entre si, como ilustrado pela Figura 4.

Iniciamos construindo uma molécula artificial. Para os átomos em um mesmo corpo rígido, as distâncias entre eles são todas conhecidas. Além disso, como os corpos rígidos são conectados entre si, um ao outro por um átomo em comum, consideramos que as distâncias entre as extremidades de diferentes corpos rígidos são conhecidas. Sem perda de generalidade, vamos supor que as distâncias entre os átomos conectados por ligações covalentes e os ângulos entre três átomos consecutivos no mesmo corpo rígido são fixos.

Já que a localização dos primeiros três átomos é considerada conhecida no problema, escolhemos três pontos não colineares no \mathbb{R}^3 . Pela Definição 4, conhecemos duas distâncias dos três átomos já fixados para qualquer um dos outros átomos ainda não determinados. Escolhendo um desses para a distância desconhecida entre este novo átomo e o já fixado considerando uma distância aleatória, obtemos duas possibilidades para o quarto átomo e escolhemos uma delas. Para os seguintes átomos, teremos, pelo menos, três distâncias a átomos já fixados, implicando em duas possibilidades para o novo átomo, onde escolhemos uma dessas possibilidades para ser o átomo da nossa molécula artificial.

Dessa forma, para criar a instância do problema precisamos somente calcular as distâncias que são conhecidas entre os átomos da molécula e ignorar as demais.

3.2 O Problema do Loop Fechado e Problema Discretizável de Geometria de Distâncias

Nesta seção, apresentaremos uma justificativa para a inexistência de uma ordem nos átomos do PLF que satisfaça as hipóteses da Definição 2.

A primeira hipótese da Definição 2 exige que a posição dos três primeiros átomos da ordem seja conhecida, esta hipótese é satisfeita pelo átomos $C_{\alpha 1}, C_{\alpha 2}, C_{\alpha 3}$, cujas posições são conhecidas de acordo com a Definição 4.

A segunda hipótese da Definição 2 exige que, para todos os átomos $x_i, i = 4, 5, \dots, n$, existam três antecessores $\{a, b, c\}$, cujas distâncias exatas até os átomos x_i sejam conhecidas. Entretanto, para determinar o quarto átomo na ordem necessitamos que este átomo possua distâncias exatas conhecidas até os três átomos com posições conhecidas, $C_{\alpha 1}, C_{\alpha 2}, C_{\alpha 3}$. Condição que não é satisfeita por nenhum dos demais átomos da nossa molécula. Por exemplo, o átomo N_1 possui distâncias exatas conhecidas até os átomos $C_{\alpha 1}$ e $C_{\alpha 3}$, mas não possui distância conhecida até o átomo $C_{\alpha 2}$. Veja a Figura 13.

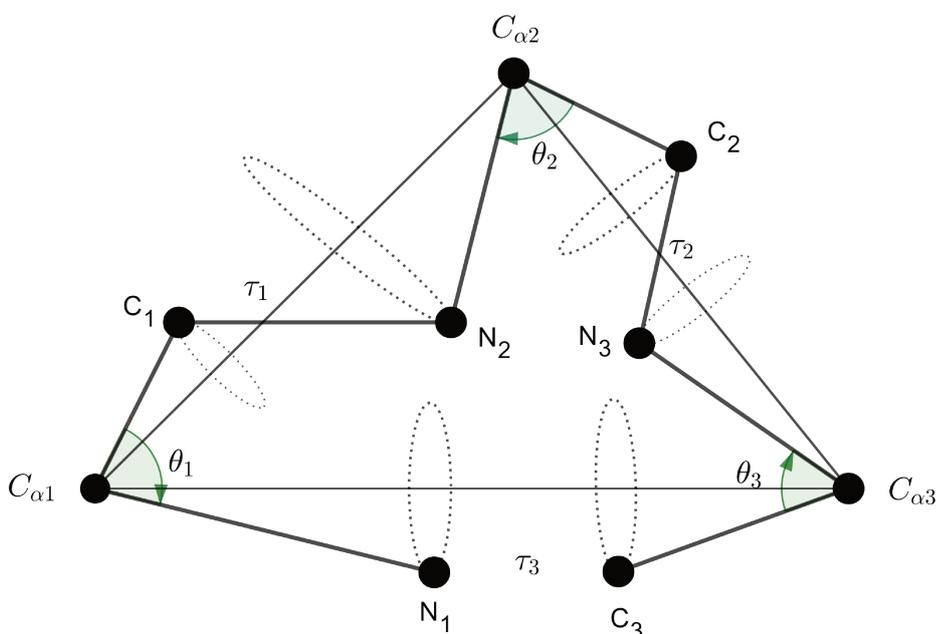


Figura 13 – Ilustração do problema do loop fechado.

3.3 Distâncias Intervalares no Problema do Loop Fechado

Sejam $C_{\alpha_1}, C_{\alpha_2}, C_{\alpha_3}$ os átomos inicialmente fixados no espaço. Gostaríamos que o próximo átomo tenha as três distâncias exatas conhecidas para os três átomos já fixados. Infelizmente, este não é o caso, pois todos os átomos ainda não realizados possuem exatamente 2 distâncias exatas conhecidas para seus antecessores. Este problema inviabiliza a construção da nossa ordem $PDGD_3$ e não permite a aplicação imediata do BP para solucionar o problema. Entretanto, podemos estimar uma distância intervalar de um desses átomos até um dos átomos fixados inicialmente.

Consideramos o átomo N_1 e os átomos $(C_{\alpha_3}, C_{\alpha_2}, C_{\alpha_1})$ fixos no espaço. Pelas informações do problema, além das distâncias entre os átomos já fixados, também são conhecidas as seguintes distâncias entre átomos fixos e N_1 :

$$\begin{cases} \| N_1 - C_{\alpha_1} \| = d_{C_{\alpha_1}N_1}, \\ \| N_1 - C_{\alpha_3} \| = d_{C_{\alpha_3}N_1}. \end{cases}$$

Nosso objetivo agora é estimar a distância do átomo fixo C_{α_2} até o átomo N_1 . Para isso, consideramos a reta r passando pelos átomos C_{α_1} e C_{α_3} e calcule a distância do átomo C_{α_2} até a reta r . Em seguida, determinamos os pontos B e A sobre a reta r satisfazendo as relações $d_{C_{\alpha_2}r} = d_{C_{\alpha_2}B}$ e $d_{N_1r} = d_{N_1A}$, e defina

$$\gamma = \cos^{-1}(\vec{v} \cdot \vec{\omega}), \quad (3.1)$$

$$\text{onde } \vec{\omega} = \frac{\overrightarrow{BN_1}}{\| \overrightarrow{BN_1} \|}, \vec{v} = \frac{\overrightarrow{BA}}{\| \overrightarrow{BA} \|}.$$

Dessa forma, obtemos a seguinte estimativa para a distância $d_{C_{\alpha_2}N_1}$:

$$d_{C_{\alpha_2}N_1} = [m, n], \quad (3.2)$$

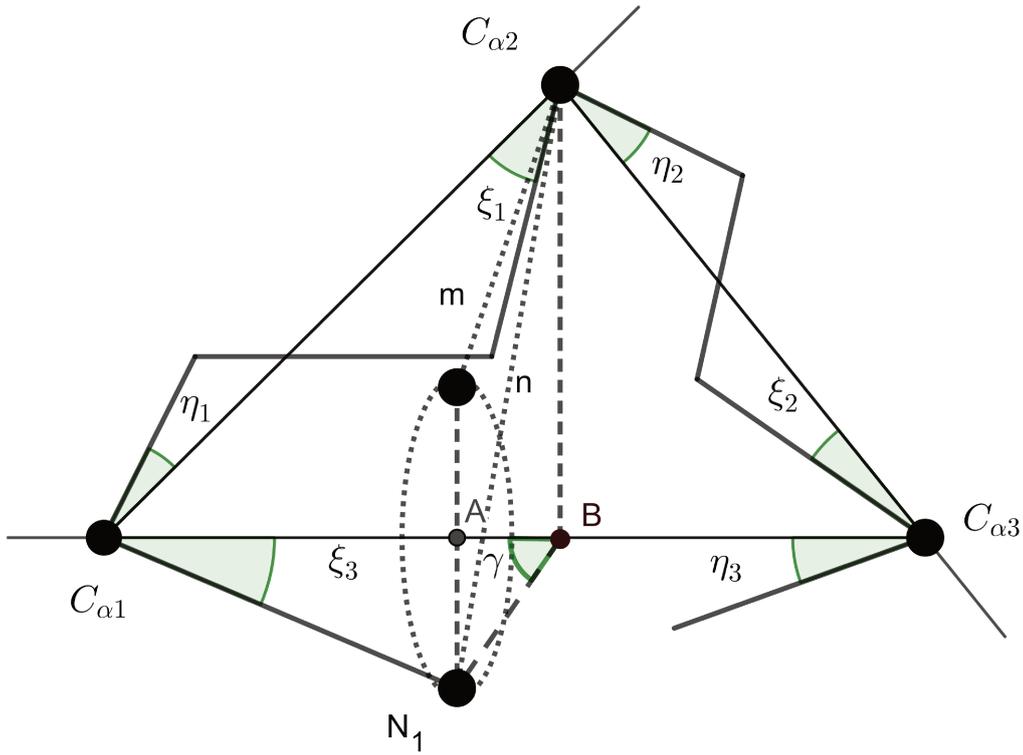
$$\text{onde } m = \sqrt{d_{C_{\alpha_2}B}^2 + d_{N_1B}^2 - 2d_{C_{\alpha_2}B}d_{N_1B} \cos\left(\frac{\pi}{2} - \gamma\right)}$$

$$\text{e } n = \sqrt{d_{C_{\alpha_2}B}^2 + d_{N_1B}^2 - 2d_{C_{\alpha_2}B}d_{N_1B} \cos\left(\frac{\pi}{2} + \gamma\right)}. \text{ Veja a ilustração na Figura 14.}$$

Assim, podemos construir a ordem que satisfaz as hipóteses da Definição 3, dada pela seguinte sequência:

$$C_{\alpha_2}, C_{\alpha_1}, C_{\alpha_3}, N_1, C_3, N_3, C_2, N_2, C_1. \quad (3.3)$$

A primeira hipótese da Definição 3, exige que as posições dos três primeiros átomos sejam conhecidas. Esta hipótese é cumprida pelos átomos $C_{\alpha_1}, C_{\alpha_2}$ e C_{α_3} que tem posições conhecidas garantidas pela Definição 4. A segunda hipótese da Definição 3 exige que para todo átomo $x_i, i = 4, 5, \dots, n$, sejam conhecidos três antecessores $\{a, b, c\}$, cujas


 Figura 14 – Distâncias máxima e mínima de $C_{\alpha 2}$ até N_1 .

distâncias até o átomo x_i devem ser duas conhecidas por valores exatos e a terceira pode ser conhecida por um intervalo real ou um valor exato. Para o quarto átomo da ordem, N_1 , são conhecidos três antecessores $C_{\alpha 1}$, $C_{\alpha 2}$ e $C_{\alpha 3}$, onde $d_{N_1 C_{\alpha 1}}$, $d_{N_1 C_{\alpha 3}}$ são conhecidas e exatas e $d_{N_1 C_{\alpha 2}}$ é uma distância intervalar conhecida, satisfazendo a segunda hipótese. Para todos os outros átomos da ordem são conhecidos, pelo menos, três antecessores $\{a, b, c\}$ para quais as distâncias $d_{x_i a}$, $d_{x_i b}$, $d_{x_i c}$ são conhecidas e exatas, satisfazendo a segunda hipótese da Definição 3.

3.4 Algoritmo iBP e o Problema do Loop Fechado

Nesta seção, vamos apresentar observações sobre o número de soluções para o PLF que são decorrentes da existência da ordem (3.3).

Observação 1. *Seja G o grafo que satisfaz as hipóteses do Problema do Loop Fechado. Considere a ordem nos vértices dada em (3.3). Para cada valor de distância intervalar temos, no máximo, 64 possibilidades de conformações para a sequência de átomos não necessariamente factíveis.*

Demonstração. Como estamos considerando uma amostra para a distância intervalar, temos que, para cada átomo a partir do quarto vértice na ordem, existem 3 antecessores

adjacentes. Pelo Lema 1, temos no máximo duas extensões válidas distintas dos três antecessores considerados. Dessa forma, como nossa ordem no grafo possui 9 átomos, temos que o número máximo de conformações é dado por $2^{9-3} = 64$. \square

Note que, neste resultado, estamos desconsiderando informações adicionais que podem resultar em podas em nossa árvore de busca. Observe a ilustração da Observação 1 na Figura 15.

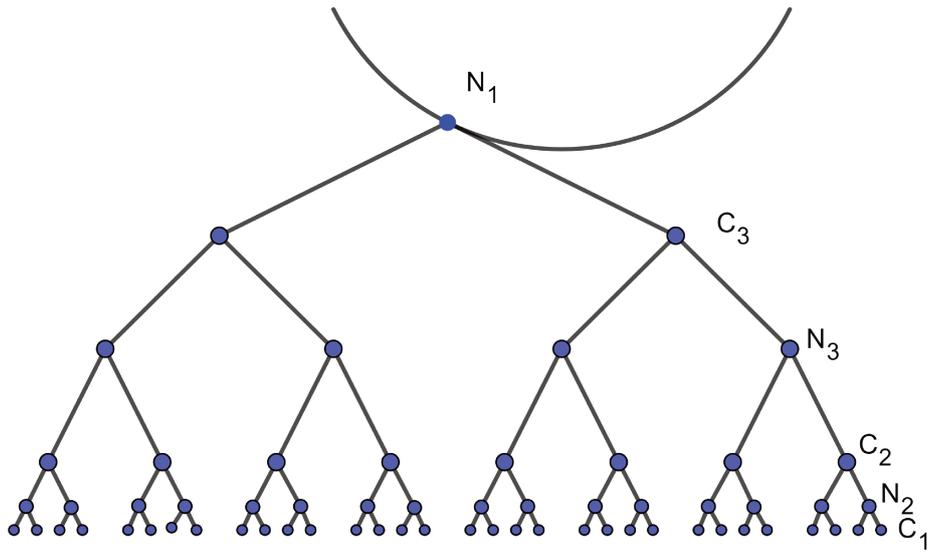


Figura 15 – Ilustração do lado direito da árvore de busca do Algoritmo iBP para uma amostra de distância intervalar.

Observação 2. Se x é uma conformação válida do PLF, então x' , a reflexão parcial de x através dos 3 primeiros pontos, é também uma conformação válida para o PLF.

Demonstração. Seja x uma conformação válida do PLF. Para que os pontos de x satisfaçam todas as distâncias na instância, vamos verificar que os pontos de x' , os pontos que são reflexões dos pontos de x através do plano π formado por $C_{\alpha 1}, C_{\alpha 2}, C_{\alpha 3}$, também satisfazem todas as distâncias na instância. Vamos dividir a demonstração em dois casos.

1º caso: Seja um ponto $x_i \in x$ tal que $d_{C_{\alpha i} x_i}$ é conhecido. Devemos mostrar que $x'_i \in x'$ satisfaz $d_{C_{\alpha i} x_i} = d_{C_{\alpha i} x'_i}$.

De fato, seja a o ponto de interseção da reta passando por x_i e x'_i e o plano contendo os três pontos $C_{\alpha 1}, C_{\alpha 2}, C_{\alpha 3}$. Como x'_i é a reflexão de x_i , temos que $d_{x_i a} = d_{x'_i a}$. Além disso, note que o segmento $aC_{\alpha i}$ pertence ao triângulo $\Delta aC_{\alpha i}x_i$ e ao triângulo $\Delta aC_{\alpha i}x'_i$. Ainda, pela propriedade de reflexão, temos que os ângulos $\angle x_i a C_{\alpha i} = \angle x'_i a C_{\alpha i} = 90^\circ$. Então, pelo caso LAL (lado, ângulo, lado) de congruência de triângulos, temos que $d_{C_{\alpha i} x_i} = d_{C_{\alpha i} x'_i}$.

2º Caso: Sejam dois pontos $x_i, x_j \in x$ tal que $d_{x_i x_j}$ é conhecido. Devemos mostrar que $x'_i, x'_j \in x'$ satisfaz $d_{x_i x_j} = d_{x'_i x'_j}$.

Sejam a, b os pontos de interseção do plano contendo os três pontos $C_{\alpha_1}, C_{\alpha_2}, C_{\alpha_3}$ com as retas passando por x_i, x'_i e x_j, x'_j , respectivamente. Inicialmente, devemos mostrar que os pontos $x_i, x'_i, x_j, x'_j, a, b$ pertencem a um mesmo plano.

Considere as retas r e r' definidas pelos pontos x_i, x_j e x'_i, x'_j , respectivamente. Então, temos que r e r' são reflexões entre si através do plano π . Se temos que r intercepta o plano π em um ponto c , então r' intercepta o plano π no ponto c . Dessa forma, r e r' definem um novo plano, digamos π' . Se r não intercepta o plano π , então r' não intercepta o plano π , pois x_i e x'_i são reflexões um do outro pelo plano π . Analogamente, para x_j e x'_j , temos que as retas r e r' são paralelas e distintas, que também definem um plano. Além disso, como a e b podem ser escritos como combinação linear de pontos nesse plano, em ambos os casos, segue-se que $x_i, x'_i, x_j, x'_j, a, b$ pertencem a um mesmo plano. Agora, considere os triângulos $\Delta x_i a b$ e $\Delta x'_i a b$. Como $d_{x_i a} = d_{x'_i a}$ e o segmento ab pertence aos dois triângulos, também pela propriedade de reflexão, segue pelo caso LAL de congruência de triângulos que os triângulos $\Delta x_i a b$ e $\Delta x'_i a b$, são congruentes. Então, obtemos que $\angle x_i b a = \angle x'_i b a$, implicando em $\angle x_i b x_j = \angle x'_i b x'_j$ e $d_{x_i b} = d_{x'_i b}$. Além disso, sabemos que $d_{x_j b} = d_{x'_j b}$. Novamente, pelo caso LAL de congruência de triângulos, obtemos que $\Delta x_i x_j b$ é congruente a $\Delta x'_i x'_j b$. Portanto, $d_{x_i x_j} = d_{x'_i x'_j}$. \square

3.5 Generalizações

Nesta seção, vamos considerar o Problema do Loop Fechado com os corpos rígidos possuindo mais átomos e em diferentes quantidades. Formalmente, definimos o problema generalizado da seguinte forma:

Definição 5. *Considere os corpos rígidos $(C_{\alpha_1}, C_{11}, \dots, C_{1n}, C_{\alpha_2})$, $(C_{\alpha_2}, C_{21}, \dots, C_{2m}, C_{\alpha_3})$ e $(C_{\alpha_3}, C_{31}, \dots, C_{3s}, C_{\alpha_1})$, onde os átomos $C_{\alpha_1}, C_{\alpha_2}, C_{\alpha_3}$ são fixos e os movimentos permitidos na estrutura são realizados individualmente por cada um dos corpos rígidos. Também, os ângulos entre três átomos consecutivos de dois diferentes corpos rígidos são conhecidos. Além disso, os movimentos de cada corpo rígido são restritos a girar em torno de seus eixos de rotação dados por $(C_{\alpha_i} - C_{\alpha_{i+1}})$, onde $\alpha_4 = \alpha_1$.*

Observe uma ilustração da generalização com $n = 7, m = 8, s = 6$ na Figura 16.

Assim, como no caso inicial, para resolver o problema devemos definir uma ordem nos vértices que satisfaça a Definição 3.

Sejam $C_{\alpha_2}, C_{\alpha_1}, C_{\alpha_3}$ os três átomos fixados dados pela Definição 5. Devemos estimar uma das distâncias desconhecidas do átomo fixo C_{α_2} até um dos átomos pertencente

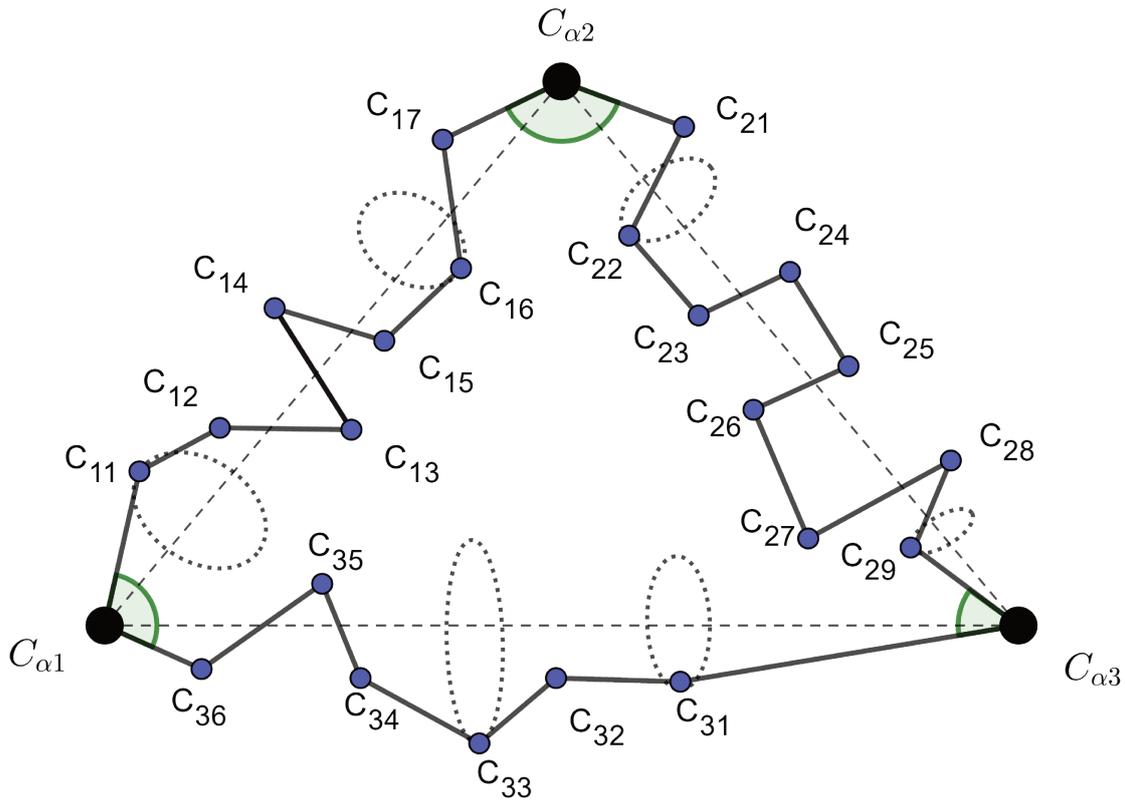


Figura 16 – Ilustração da generalização do Problema do Loop Fechado.

ao corpo rígido $(C_{\alpha 3}, C_{31}, \dots, C_{3s}, C_{\alpha 1})$. Sem perda de generalidade, vamos estimar a distância de $C_{\alpha 2}$ até C_{3s} (o cálculo até qualquer outro átomo é análogo).

Considere a reta r passando pelos átomos $C_{\alpha 1}$ e $C_{\alpha 3}$ e calcule a distância do átomo $C_{\alpha 2}$ até a reta r . Agora, determine os pontos B e A sobre a reta r satisfazendo as equações $d_{C_{\alpha 2}r} = d_{C_{\alpha 2}B}$ e $d_{N_{3s}r} = d_{N_{3s}A}$. Definindo

$$\gamma = \cos^{-1}(\vec{v} \cdot \vec{w}), \quad (3.4)$$

onde $\vec{w} = \frac{\overrightarrow{BN_{3s}}}{\|\overrightarrow{BN_{3s}}\|}$, $\vec{v} = \frac{\overrightarrow{BA}}{\|\overrightarrow{BA}\|}$, obtemos a seguinte estimativa para a distância $d_{C_{\alpha 2}N_{3s}}$:

$$d_{C_{\alpha 2}N_{3s}} = [p, q] \quad (3.5)$$

onde

$$p = \sqrt{d_{C_{\alpha 2}B}^2 + d_{N_{3s}B}^2 - 2d_{C_{\alpha 2}B}d_{N_{3s}B} \cos\left(\frac{\pi}{2} - \gamma\right)}$$

$$q = \sqrt{d_{C_{\alpha 2}B}^2 + d_{N_{3s}B}^2 - 2d_{C_{\alpha 2}B}d_{N_{3s}B} \cos\left(\frac{\pi}{2} + \gamma\right)}. \text{ Veja a ilustração para o caso em que } n = 7, m = 9, s = 6 \text{ na Figura 17.}$$

A ordem nos vértices que satisfaz todas as hipóteses da Definição 3 é dada por

$$\begin{aligned} & C_{\alpha 2}, C_{\alpha 1}, C_{\alpha 3}, C_{3s}, C_{31}, C_{2m}, C_{21}, C_{1n}, C_{11}, C_{12}, C_{13}, \dots, C_{1n-1}, C_{22}, \\ & C_{23}, \dots, C_{2m-1}, C_{32}, C_{33}, \dots, C_{3s-1}. \end{aligned} \quad (3.6)$$

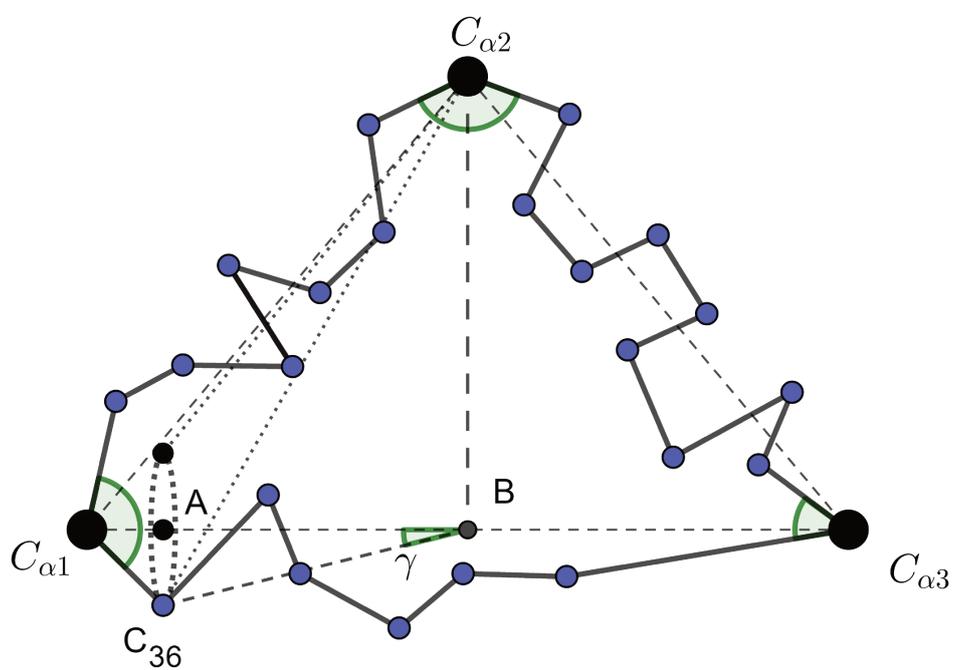


Figura 17 – Ilustração da estimativa de distância para a Generalização do Problema Loop Fechado.

4 Comparações entre os métodos: CSJD, CCS e iBP

Neste capítulo, vamos comparar os métodos apresentados nos capítulos anteriores para alguns casos de loops que ocorrem em moléculas e podem ser modelados como Loops Fechados. Iniciamos descrevendo a ideia básica do método apresentado por (COUTSIAS *et al.*, 2004) chamado de CSJD, seguimos com uma rápida descrição do método apresentado em (CANUTESCU; DUNBRACK, 2003) denominado CCD e do método que estamos utilizando neste trabalho, chamado iBP. O CSJD é um método baseado em derivar equações polinomiais para identificar os possíveis valores dos ângulos τ_i e δ_i , ilustrados na Figura 7, para posicionar os átomos a partir destas informações. O CCD é um método baseado em modificar uma configuração inicial do loop, considerada conhecida, através do ajuste dos ângulos diedrais em cada resíduo. Enquanto o iBP trabalha a partir do posicionamento ordenado dos átomos. Ou seja, o método do iBP exige uma ordem nos átomos da molécula para obter sua conformação.

No CSJD, após a construção do polinômio, é utilizado o Método de Sturm (HENRICI, 1974, Seção 6) para determinar o número de soluções do polinômio em um determinado intervalo. Em seguida, os intervalos são bisectados e refinados até todas as soluções estarem separadas. Então, cada intervalo é refinado até que a solução esteja contida em um intervalo pequeno o suficiente de modo que qualquer ponto do intervalo satisfaça um erro pré-determinado, prossegue-se utilizando essas soluções para se determinar os possíveis ângulos σ_i e τ_i e, com essas informações, constrói-se a estrutura. Notemos que, neste método, se uma raiz real para o polinômio é encontrada, então as estruturas construídas a partir dessa raiz não são submetidas a nenhuma checagem para avaliar se as posições encontradas para os átomos nas estruturas satisfazem as distâncias conhecidas entre eles.

No CCD, sempre é considerado um “modelo” inicial para o loop procurado, sendo ele obtido de algum banco de dados de proteínas ou construído artificialmente a partir dos dados das distâncias e ângulos entre os átomos que compõem o loop. Posteriormente, os novos loops são encontrados a partir de alterações nos ângulos diedrais (ϕ e ψ). Antes de realizar uma mudança em determinado ângulo diedral, é realizado o procedimento descrito na Seção 2.3 para determinar o melhor movimento para aquele determinado ângulo. O procedimento é realizado iniciando com o ângulo diedral ϕ do primeiro resíduo da cadeia, seguido pelo ângulo diedral ψ do segundo resíduo e assim sucessivamente. As condições de parada são dadas pelos resíduo inicial e final do trecho da proteína que forma o loop, estejam sobrepostos com a N -âncora e C -âncora ou atinjam um limite de 5000 passos no

processo.

No iBP, após estimar a distância intervalar, tomamos valores neste intervalo que obtivemos e calculamos as possíveis soluções para ele. No caso canônico (o loop possui 9 átomos), encontramos, para cada valor, 64 possibilidades, sendo as últimas 32 dadas pela reflexão das primeiras 32 através do plano formado pelos três átomos inicialmente conhecidos. Além disso, observando a ordem, notamos que, após no máximo nove átomos, todos os seguintes possuem no mínimo 4 antecessores. Na ordem (3.6), para cada átomo a seguir, temos uma etapa de verificação da factibilidade do loop que estamos construindo. Dessa forma, o número de loops factíveis é amplamente reduzido. Isso ocorre devido a cada amostra retirada da distância intervalar gerar, com probabilidade 1, duas ou nenhuma soluções factíveis entre as 64 possíveis.

4.1 Loops formados por quatro resíduos de molécula

Os loops formados por 4 resíduos contém um total de 12 átomos, devido a cada resíduo ser constituído por três átomos, sendo eles dois átomos de carbono e um átomo de nitrogênio. Esses átomos formam um trecho na cadeia principal de algumas moléculas.

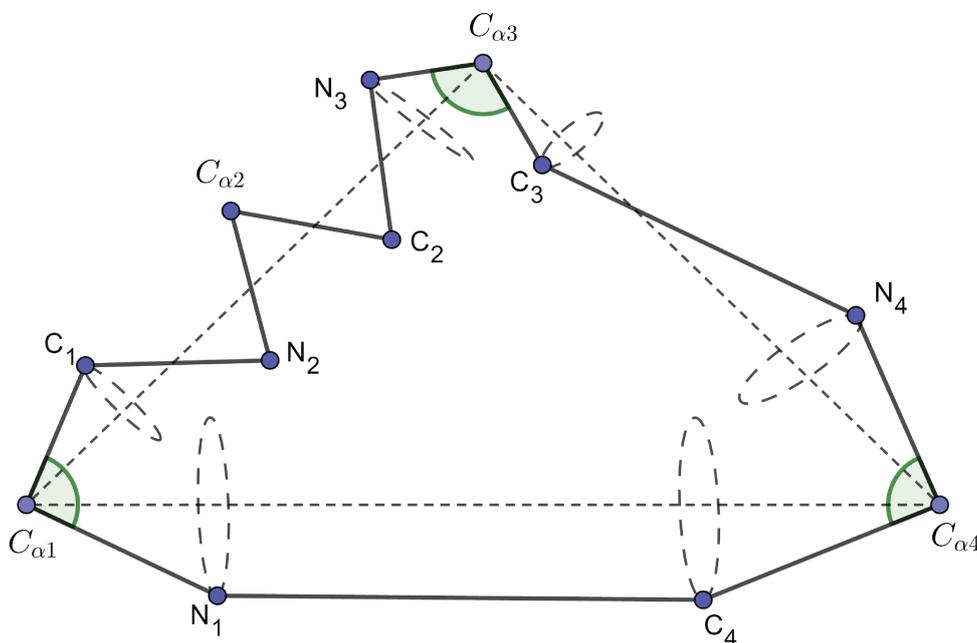


Figura 18 – Ilustração do problema para trechos de moléculas formados por 4 resíduos.

Neste caso, a solução através do CSJD segue a dedução apresentada da Generalização do algoritmo CSJD, (para o posicionamento de todos os átomos é necessário o cálculo de mais alguns ângulos diédricos). Para o CCD, considerando que temos uma conformação inicial do loop, temos que as mudanças ocorrem em exatamente seis ângulos

diedrais (veja a Figura 19). Utilizando o iBP, o cálculo para o posicionamento dos átomos extras, em relação ao caso canônico, é realizado da mesma forma que para os átomos que compõem o problema canônico.

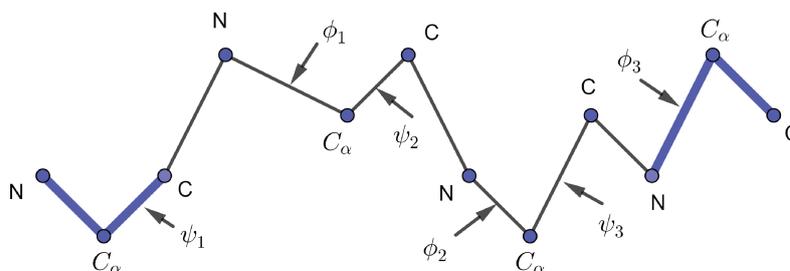


Figura 19 – Ilustração dos ângulos ψ e ϕ que são rotacionados no algoritmo CCD em loops formados por quatro resíduos para obter todas as conformações desses loops.

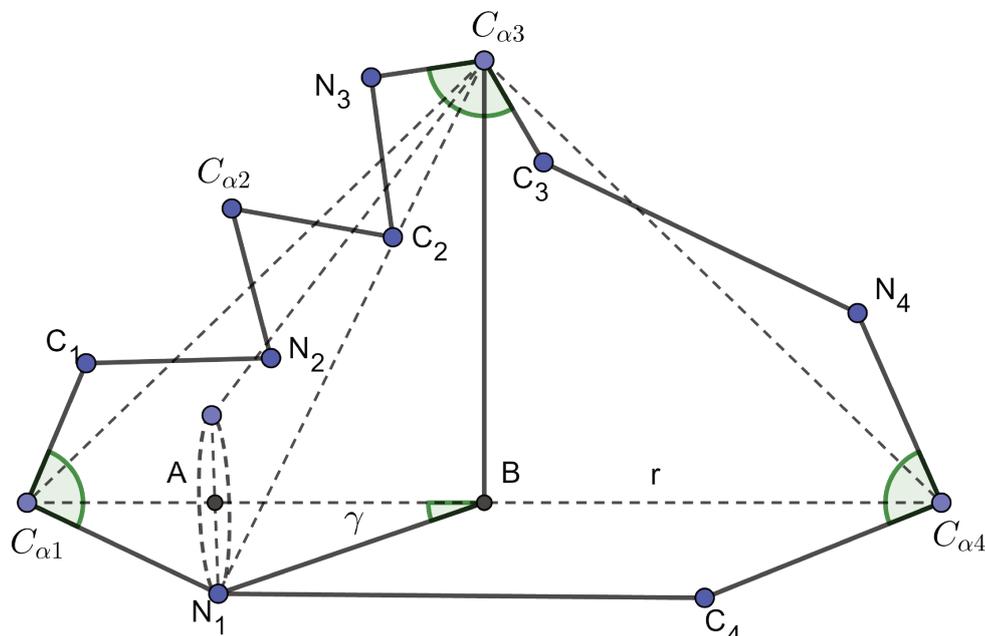


Figura 20 – Ilustração do cálculo da distância intervalar.

O cálculo da distância intervalar e a ordenação dos átomos, considerando o exemplo da ilustração na Figura 20 são dados a seguir. Consideramos o átomo N_1 e os átomos $(C_{\alpha 3}, C_{\alpha 4}, C_{\alpha 1})$ fixos no espaço. Pelas informações do problema, além das distâncias entre os átomos já fixados, também são conhecidas as seguintes distâncias dos átomos fixos até N_1 :

$$\begin{cases} \| N_1 - C_{\alpha 1} \| &= d_{C_{\alpha 1} N_1}, \\ \| N_1 - C_{\alpha 4} \| &= d_{C_{\alpha 3} N_1}. \end{cases}$$

Agora, devemos estimar a distância do átomo fixo $C_{\alpha 3}$ até o átomo N_1 . Para isso, considere a reta r passando pelos átomos $C_{\alpha 1}$ e $C_{\alpha 3}$ e calcule a distância do átomo $C_{\alpha 2}$ até a reta r . Em seguida, determine os pontos B e A sobre a reta r satisfazendo as relações $d_{C_{\alpha 3} r} = d_{C_{\alpha 3} B}$ e $d_{N_1 r} = d_{N_1 A}$. Defina,

$$\gamma = \cos^{-1}(\vec{v} \cdot \vec{\omega}), \quad (4.1)$$

onde $\vec{\omega} = \frac{\overrightarrow{BN_1}}{\| \overrightarrow{BN_1} \|}$, $\vec{v} = \frac{\overrightarrow{BA}}{\| \overrightarrow{BA} \|}$. Dessa forma, obtemos a seguinte estimativa para a distância $d_{C_{\alpha 3} N_1}$:

$$d_{C_{\alpha 3} N_1} = [m, n], \quad (4.2)$$

onde

$$m = \sqrt{d_{C_{\alpha 3} B}^2 + d_{N_1 B}^2 - 2d_{C_{\alpha 3} B}d_{N_1 B} \cos\left(\frac{\pi}{2} - \gamma\right)}$$

e

$$n = \sqrt{d_{C_{\alpha 3} B}^2 + d_{N_1 B}^2 - 2d_{C_{\alpha 3} B}d_{N_1 B} \cos\left(\frac{\pi}{2} + \gamma\right)}. \text{ Veja a ilustração na Figura 20.}$$

A seguir, construímos a ordem que satisfaz as hipóteses da Definição 3, dada pela seguinte sequência nos átomos:

$$C_{\alpha 3}, C_{\alpha 1}, C_{\alpha 4}, N_1, C_4, N_4, C_3, N_3, C_1, N_2, C_{\alpha 2}, C_2. \quad (4.3)$$

A principal diferença entre este caso e o caso canônico para o iBP está no espaço de busca. Neste caso, temos 3 átomos a mais para posicionar, mas para cada um desses átomos extras existem, pelo menos, 4 átomos antecessores adjacentes. Então, temos, no máximo, um ramo na árvore de busca a partir do décimo átomo na ordem. Veja a ilustração na Figura 21.

4.2 Loops formados por oito resíduos de moléculas

Neste caso, nossa sequência conta com oito resíduos de moléculas, ou seja, 24 átomos, sendo conhecidos os três átomos fixos. Para o CSJD, o procedimento é feito identicamente ao descrito na Seção 2.2 do Capítulo 2. No algoritmo CCD, temos movimentos em exatamente 14 ângulos diedrais. Como o procedimento é limitado a um número fixo de passos, então cada ângulo diedral é modificado menos vezes que no caso de loops formados por quatro resíduos. Entretanto, como temos mais ângulos para serem ajustados nesse

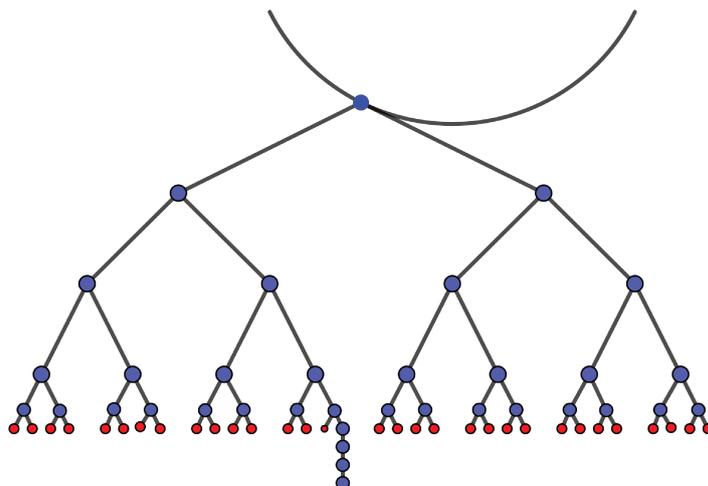


Figura 21 – Lado esquerdo da árvore de busca do iBP para o Problema do Loop Fechado de quatro resíduos.

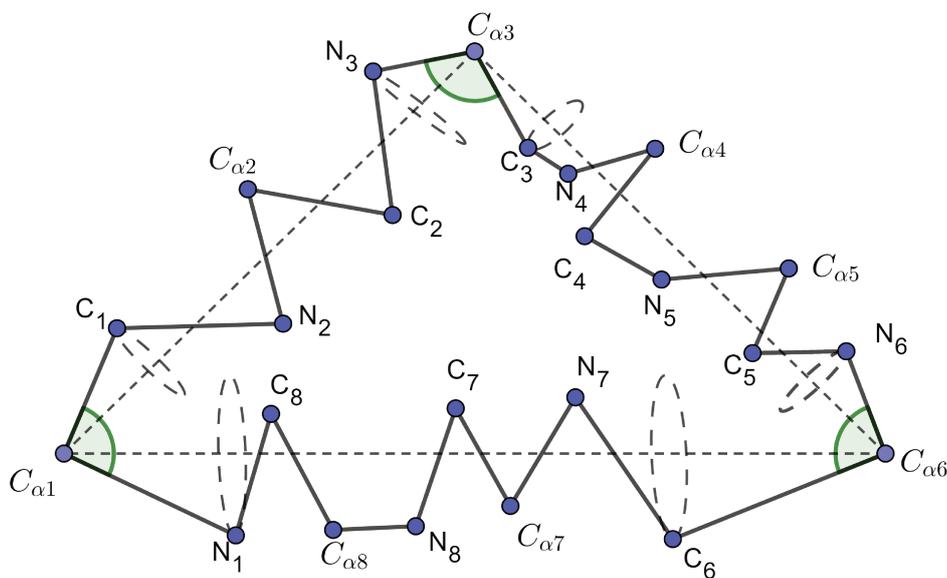


Figura 22 – Exemplo de Loop Fechado formado por oito resíduos de uma molécula.

caso, a chance de obter soluções para o problema é maior, devido a possibilidade de maior número de ajustes em ângulos.

Vamos agora ver como é o procedimento para o iBP. Assim como para outros casos, o primeiro passo é estimar a distância intervalar dos átomos do corpo rígido $C_{\alpha 3}, C_{31}, \dots, C_{3n}, C_{\alpha 1}$, exceto $C_{\alpha 1}$ e $C_{\alpha 3}$ até $C_{\alpha 2}$, pois já são distâncias conhecidas por

hipótese.

Vamos considerar o exemplo ilustrado na Figura 22. Consideramos o átomo N_1 e os átomos $(C_{\alpha 3}, C_{\alpha 6}, C_{\alpha 1})$ fixos no espaço. Pelas informações do problema, além das distâncias entre os átomos já fixados, também são conhecidas as seguintes distâncias dos átomos fixos a N_1 :

$$\begin{cases} \| N_1 - C_{\alpha 1} \| = d_{C_{\alpha 1} N_1}, \\ \| N_1 - C_{\alpha 6} \| = d_{C_{\alpha 6} N_1}. \end{cases}$$

Nosso objetivo é estimar a distância do átomo fixo $C_{\alpha 3}$ até o átomo N_1 . Para isso, considere a reta r passando pelos átomos $C_{\alpha 1}$ e $C_{\alpha 6}$ e calcule a distância do átomo $C_{\alpha 3}$ até a reta r . Em seguida, determine os pontos B e A sobre a reta r satisfazendo as relações $d_{C_{\alpha 3} r} = d_{C_{\alpha 3} B}$ e $d_{N_1 r} = d_{N_1 A}$. Defina,

$$\gamma = \cos^{-1}(\vec{v} \cdot \vec{\omega}), \quad (4.4)$$

onde $\vec{\omega} = \frac{\overrightarrow{BN_1}}{\| \overrightarrow{BN_1} \|}$, $\vec{v} = \frac{\overrightarrow{BA}}{\| \overrightarrow{BA} \|}$. Dessa forma, obtemos a seguinte estimativa para a distância $d_{C_{\alpha 3} N_1}$:

$$d_{C_{\alpha 3} N_1} = [m, n], \quad (4.5)$$

onde

$$m = \sqrt{d_{C_{\alpha 3} B}^2 + d_{N_1 B}^2 - 2d_{C_{\alpha 3} B}d_{N_1 B} \cos\left(\frac{\pi}{2} - \gamma\right)}$$

e

$$n = \sqrt{d_{C_{\alpha 3} B}^2 + d_{N_1 B}^2 - 2d_{C_{\alpha 3} B}d_{N_1 B} \cos\left(\frac{\pi}{2} + \gamma\right)}. \text{ Veja a ilustração na Figura 23.}$$

Portanto, podemos construir a ordem que satisfaz as hipóteses da Definição 3 dada pela seguinte sequência:

$$\begin{aligned} & C_{\alpha 3}, C_{\alpha 1}, C_{\alpha 6}, N_1, C_6, N_6, C_3, N_3, C_1, C_8, C_{\alpha 8}, N_8, C_7, C_{\alpha 7}, \\ & N_7, C_5, C_{\alpha 5}, N_5, C_4, C_{\alpha 4}, N_4, C_2, C_{\alpha 2}, N_2. \end{aligned} \quad (4.6)$$

A árvore de busca do iBP para loops formados por oito resíduos é semelhante à árvore de busca do caso de loops formados por 4 resíduos de uma molécula. Veja a ilustração na Figura 24.

4.3 Loops formados por 12 resíduos de moléculas

Em loops compostos por 12 resíduos, temos 36 átomos que devem ter suas posições determinadas. No algoritmo CSJD, o processo é realizado como descrito na Seção 2.2 do Capítulo 2. Para o algoritmo CCD, temos movimentos em 22 ângulos diedrais,

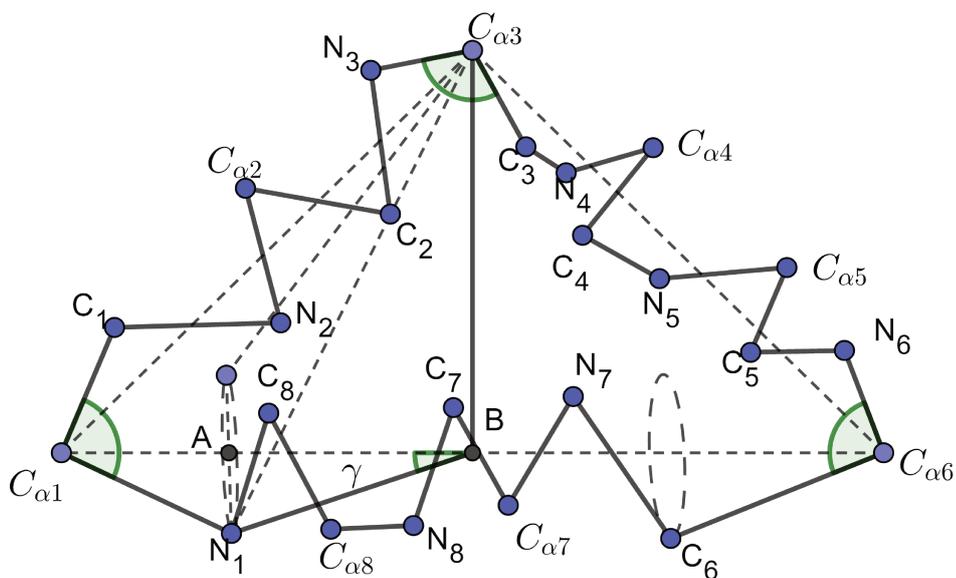


Figura 23 – Ilustração do cálculo da distância intervalar.

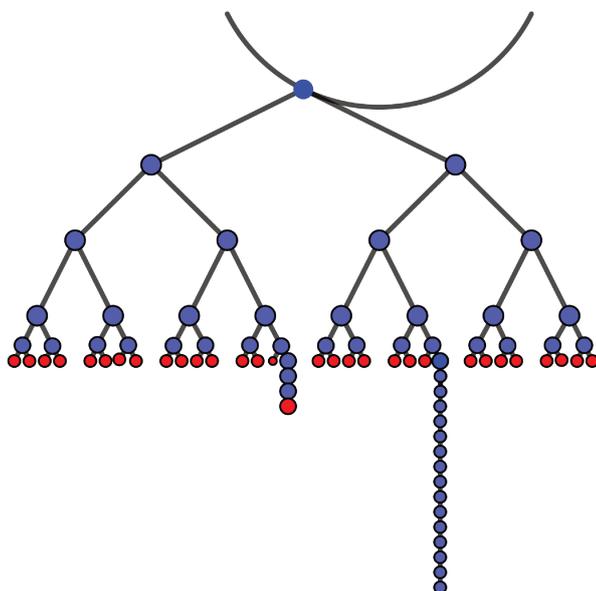


Figura 24 – Lado esquerdo da árvore de busca do iBP para o Problema do Loop Fechado de oito resíduos.

devido ao maior número de possibilidade de movimentos distintos, o que facilita a obtenção de soluções. Seguimos descrevendo o procedimento para o algoritmo iBP. Inicialmente, devemos estimar a distâncias intervalar de um átomo de um corpo rígido até o átomo fixo

que não pertence a este corpo rígido.

Vamos considerar o exemplo ilustrado na Figura 25. Consideramos o átomo N_1 e os átomos $(C_{\alpha 5}, C_{\alpha 9}, C_{\alpha 1})$ fixos no espaço. Pelas informações do problema, além das distâncias entre os átomos já fixados, também são conhecidas as seguintes distâncias dos átomos fixos a N_1 :

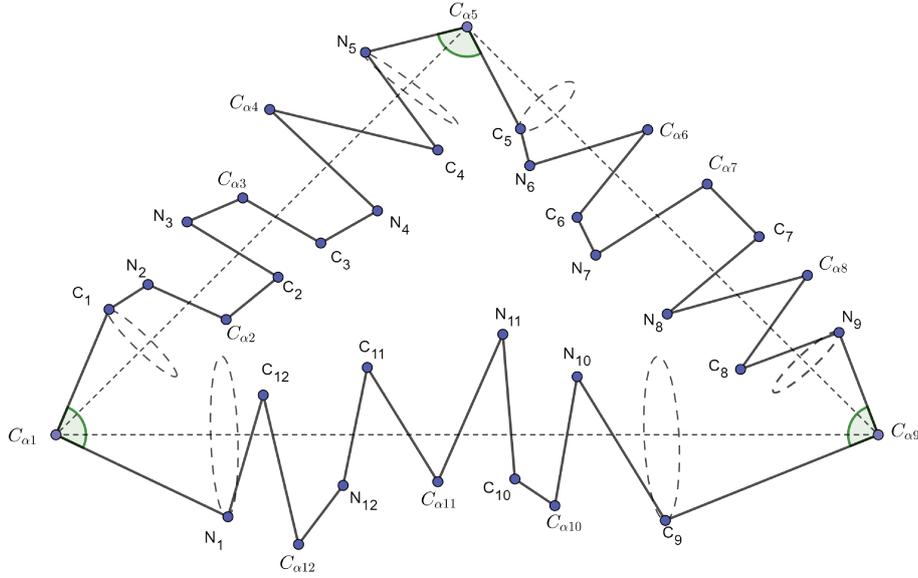


Figura 25 – Exemplo de Loop Fechado formado por doze resíduos de uma molécula.

$$\begin{cases} \| N_1 - C_{\alpha 1} \| = d_{C_{\alpha 1} N_1}, \\ \| N_1 - C_{\alpha 9} \| = d_{C_{\alpha 9} N_1}. \end{cases}$$

Queremos estimar a distância do átomo fixo $C_{\alpha 5}$ até o átomo N_1 . Para isso, considere a reta r passando pelos átomos $C_{\alpha 1}$ e $C_{\alpha 9}$ e calcule a distância do átomo $C_{\alpha 5}$ até a reta r . Em seguida, determine os pontos B e A sobre a reta r satisfazendo as relações $d_{C_{\alpha 5} r} = d_{C_{\alpha 5} B}$ e $d_{N_1 r} = d_{N_1 A}$. Defina,

$$\gamma = \cos^{-1}(\vec{v} \cdot \vec{\omega}), \quad (4.7)$$

onde $\vec{\omega} = \frac{\overrightarrow{BN_1}}{\| \overrightarrow{BN_1} \|}$, $\vec{v} = \frac{\overrightarrow{BA}}{\| \overrightarrow{BA} \|}$. Dessa forma, obtemos a seguinte estimativa para a distância $d_{C_{\alpha 5} N_1}$:

$$d_{C_{\alpha 5} N_1} = [m, n], \quad (4.8)$$

onde

$$m = \sqrt{d_{C_{\alpha 5} B}^2 + d_{N_1 B}^2 - 2d_{C_{\alpha 5} B}d_{N_1 B} \cos\left(\frac{\pi}{2} - \gamma\right)}$$

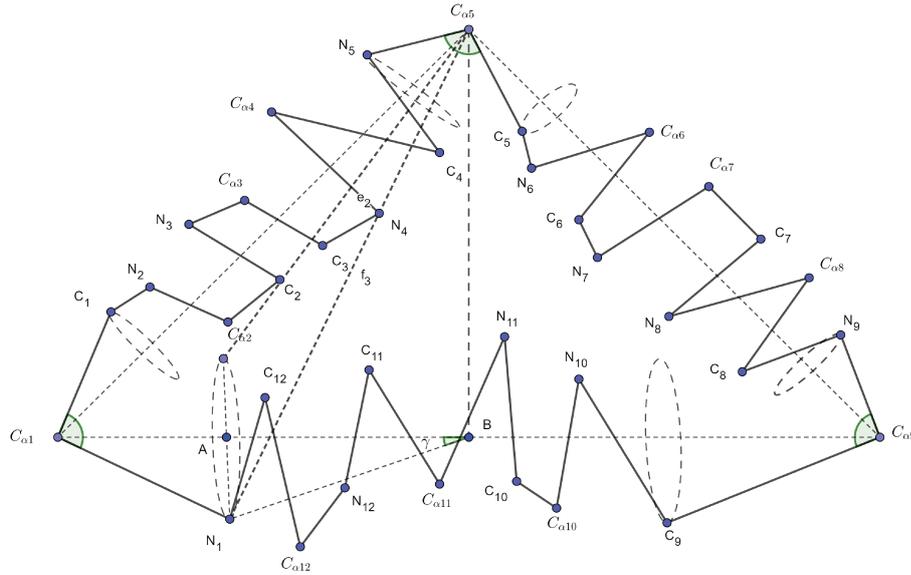


Figura 26 – Ilustração do cálculo da distância intervalar.

e

$$n = \sqrt{d_{C_{\alpha 5}B}^2 + d_{N_1B}^2 - 2d_{C_{\alpha 5}B}d_{N_1B} \cos\left(\frac{\pi}{2} + \gamma\right)}. \text{ Veja a ilustração na Figura 26.}$$

Portanto, podemos construir a ordem que satisfaz as hipóteses da Definição 3, dada pela seguinte sequência:

$$\begin{aligned} &C_{\alpha 5}, C_{\alpha 1}, C_{\alpha 9}, N_1, C_9, N_9, C_5, N_5, C_1, C_{12}, C_{\alpha 12}, N_{12}, C_{11}, C_{\alpha 11}, \\ &N_{11}, C_{10}, C_{\alpha 10}, N_{10}, C_8, C_{\alpha 8}, N_8, C_7, C_{\alpha 7}, N_7, C_6, C_{\alpha 6}, N_6, C_4, \\ &C_{\alpha 4}, N_4, C_3, C_{\alpha 3}, N_3, C_2, C_{\alpha 2}, N_2. \end{aligned} \quad (4.9)$$

A ilustração do espaço de busca para esse caso no iBP é idêntica à ilustração do espaço de busca do caso de loops formados por oito resíduos. Veja a figura 24.

4.4 Comparações Gerais

Nesta seção, vamos comparar, teoricamente, os resultados obtidos com nossa proposta e os resultados apresentados pelo CCS e CSJD.

Iniciamos analisando o número de átomos fixos necessários em cada método. Para o CSJD, são fixados um dos três corpos rígidos inicialmente e, no final, uma rotação é necessária para encontrar as posições reais dos átomos. Já no algoritmo CCS, é necessário conhecer as posições do primeiro e último resíduos e também uma conformação inicial do trecho da molécula, a qual é construída artificialmente ou retirada do Banco de Dados de Proteínas. Finalmente, utilizando o algoritmo iBP é necessário somente o conhecimento de três átomos iniciais e a ordem construída no Capítulo 3.

Analisemos agora quando uma solução encontrada é considerada válida em cada método. Para o algoritmo CSJD, após ser encontrada as soluções reais do polinômios de grau 16, os átomos são posicionados sem nenhuma outra verificação de factibilidade, assim uma solução é considerada válida se os seus ângulos τ_i corresponderem a uma solução real do polinômio de grau 16. No caso do algoritmo CCS, se uma conformação satisfizer a sobreposição do primeiro e último resíduos, então está é considerada uma solução válida, independente do posicionamento dos átomos dos demais resíduos. No método iBP, podemos analisar o método em duas etapas, como o próprio nome sugere iBP (interval Branch and Prune): a parte de busca de soluções e a parte de poda das soluções que não satisfizerem alguma distância conhecida. A discretização do espaço de busca é dada pelos três antecessores mais próximos do átomos, cuja posição será calculada, garantindo a finitude do espaço de busca (pelo Lema 4 temos no máximo $2q$ possibilidades de realização). Já para os átomos que possuem quatro ou mais antecessores, o método iBP utiliza uma filtragem nas soluções, descartando as que não satisfazem todas as distâncias adicionais. Em particular, no Problema do Loop fechado, a partir do nono átomo da ordem, é sempre realizada a poda, pois todos os átomos a partir do átomo de número nove na ordem possuem pelo menos 4 antecessores. Dessa forma, para o iBP, uma solução é considerada válida se ela satisfaz todas as distâncias conhecidas no problema.

A propagação de erros é um grande problema em diversos algoritmos. O algoritmo iBP dispõe de etapas de verificação da factibilidade do posicionamento dos átomos quando existem distâncias extras. No Problema do Loop Fechado, conhecemos distâncias entre átomos em diferentes posições da ordem. Se houver propagação de erro, a distância entre um átomo no início e no final da sequência será alterada e está conformação será descartada durante a etapa de verificação da solução no iBP. No algoritmo CSJD, uma vez considerada a solução do polinômio, não ocorre a verificação da propagação de erros. No algoritmo CCS, uma solução é considerada factível se os resíduos inicial e final estão sobrepostos com as posições pré-determinadas para eles. Dessa forma, a etapa de verificação dessa condição também soluciona o problema da propagação de erros, mas diversas soluções que não são viáveis são consideradas como válidas, por não ser adotada nenhuma restrição nas conformações do “interior” do loop.

5 Experimentos Computacionais

Os experimentos computacionais para analisar o método sistematizado na forma do algoritmo iBP, apresentado na Seção 3.4, foram realizados no mesmo conjunto de instâncias utilizados nos trabalhos de (CANUTESCU; DUNBRACK, 2003) e (COUTSIAS et al., 2004). Os testes foram realizados em um computador Notebook Intel(R), Core(TM) i7-7500U, 2.90GHz com 8GB RAM, sistema operacional Windows 10, utilizando uma implementação em Wolfram Mathematica.

5.1 RMSD

Neste seção, vamos falar sobre a medida de comparação mais utilizada para verificar a similaridade das coordenadas de duas estruturas atômicas, denominada Root Mean Square Deviation (RMSD) (KUFAREVA; ABAGYAN, 2011). Os valores de RMSD são calculados por

$$RMSD = \sqrt{\frac{1}{n} \sum_1^n d_i^2}, \quad (5.1)$$

onde d_i representa a distância entre os átomos do i -ésimo par.

Quanto mais próximo de zero o valor encontrado pelo teste de RMSD significa que os respectivos pares de átomos das estruturas estão mais próximos ou até mesmo sobrepostos. Dessa forma, como vamos calcular o RMSD das estruturas encontradas pelos métodos iBP e CSJD com as estruturas originais dos loops, resultados próximos de zero correspondem a estruturas mais próximas das estruturas originais dos loops.

5.2 LDME

Nesta seção, vamos apresentar outra medida de qualidade para estruturas de moléculas, chamado de Local Distance Matrix Error (LDME) (SOUZA et al., 2013). O cálculo desta medida é dado por

$$LDME(X) = \sqrt{\frac{1}{|E|} \sum_{\{i,j\} \in E} e_{ij}^2}, \quad (5.2)$$

onde E representa o conjunto de distâncias conhecidas na molécula e

$$e_{ij} = \max[l_{ij} - \|x_i - x_j\|, \|x_i - x_j\| - u_{ij}, 0],$$

onde l_{ij} são os menores valores nas distâncias intervalares conhecidas, u_{ij} são os maiores valores nas distâncias intervalares conhecidas, já para distâncias exatas conhecidas temos que $d_{ij} = l_{ij} = u_{ij}$.

Esta medida de qualidade avalia se as estruturas encontradas satisfazem as informações conhecidas do problema. Desta forma, quanto mais próximo de zero o valor do LDME significa que as estruturas dos loops encontradas, através dos métodos, estão mais próximas de satisfazerem todas as distâncias conhecidas.

5.3 Comparando o CSJD e o iBP

Nesta Seção, serão comparados os resultados numéricos obtidos com a utilização do iBP e os resultados numéricos obtidos com o método CSJD proposto por Coutσίας, os dados numéricos do CSJD foram obtidos de (COUTSIAS et al., 2004). Os loops utilizados para os testes numéricos foram extraídos de (BERMAN et al., 2002). Não vamos utilizar o método CCS nas comparações, pois a comparação entre o método CSJD e o método CCS foi feita em (COUTSIAS et al., 2004) e verificou-se que o método CSJD é superior.

Na Tabela 2, apresentamos os resultados dos testes realizados em loops compostos por 4 resíduos. Um loop composto por 4 resíduos possui 12 átomos, como os três primeiros átomos são fixos, temos para cada amostra retirada da distância intervalar $2^{12-3} = 512$ possibilidades de loops exploradas. Realizamos os testes retirando 1000, 2000 e 10000 amostras da distância intervalar, exceto no loop 1DYS, explorando a factibilidade de $5, 12 \times 10^5, 2 \times 5, 12 \times 10^5$ e $5, 12 \times 10^6$ loops, respectivamente. O número de soluções encontradas em cada instância está ligado a quantidade de amostras retiradas na distância intervalar e ao erro aceitável nas etapas de verificação de factibilidade. O tempo de execução do algoritmo cresce de acordo com a quantidade de amostras retiradas na distância intervalar. O tempo gasto pelo algoritmo em cada loop é de aproximadamente 7×10^{-5} segundos.

Na Tabela 3, mostramos os resultados dos testes realizados em loops formados por 8 resíduos. Nesse caso, temos 24 átomos em cada loop, como os 3 primeiros são fixos, existem 2^{24-3} possibilidades de loops exploradas para cada amostra retirada da distância intervalar. Destacamos que mesmo com o aumento exponencial na quantidade de possibilidade de realização do loops o tempo de execução do algoritmo é semelhante ao verificado em instâncias compostas por somente 4 resíduos, isso ocorre devido a todo átomo, a partir do nono na ordem, possuir, no mínimo, 4 antecessores adjacentes com distâncias exatas conhecidas, implicando no não aparecimento de novas ramificações na árvore de busca. A quantidade de soluções encontradas em cada instância varia entre 2 e 28 dependendo da quantidade de amostras tomadas na distância intervalar e do valor de erro aceito nas etapas de verificação de factibilidade da solução.

Tabela 2 – Loop de 4 resíduos

Loop	Amostras	Distância Intervalar	Precisão	Soluções	CPU (seg)
1DVJ A20-23	1000	0,848166	10^{-4}	12	38,1094
	2000	0,848166	10^{-5}	2	76,7031
	10000	0,848166	10^{-5}	18	387,438
1DYS A47-50	1000	0,273706	10^{-4}	4	36,6563
	2000	0,273706	10^{-4}	8	36,6563
	20000	0,273706	10^{-5}	4	760,891
1EGU A404-407	1000	0,31459	10^{-4}	8	36,0469
	2000	0,31459	10^{-4}	10	72,3906
	10000	0,31459	10^{-5}	2	363,344
1EJO A74-77	1000	1,84316	10^{-3}	4	36,5156
	2000	1,84316	10^{-3}	10	72,875
	10000	1,84316	10^{-4}	2	366,047
1QNR A195-198	1000	0,892762	10^{-4}	2	38,0469
	2000	0,892762	10^{-4}	6	76,2813
	10000	0,892762	10^{-5}	2	381,922
1QOP A44-47	1000	0,828994	10^{-5}	12	36,3125
	2000	0,828994	10^{-5}	28	72,8594
	10000	0,828994	10^{-6}	12	361,141
1TCA A95-98	1000	0,57357	10^{-4}	8	36,1406
	2000	0,57357	10^{-5}	4	72,0156
	10000	0,57357	10^{-6}	4	361,828

Tabela 3 – Loop de 8 resíduos

Loop	Amostras	Distância Intervalar	Precisão	Soluções	CPU (seg)
1CRU A85-92	1000	1,77218	10^{-4}	20	37,3438
	2000	1,77218	10^{-5}	2	71,5
	10000	1,77218	10^{-6}	2	356,001
1CTQ A144-151	1000	1,94185	10^{-3}	18	35,5625
	2000	1,94185	10^{-4}	4	72,9844
	10000	1,94185	10^{-5}	2	364,547
1D8W A334-341	1000	0,752272	10^{-5}	4	38,9844
	2000	0,752272	10^{-6}	2	74,3125
	10000	0,752272	10^{-7}	2	359,187
1DS1 A20-27	1000	0,421043	10^{-4}	10	35,4531
	2000	0,421043	10^{-4}	28	72,0313
	10000	0,421043	10^{-5}	22	364,203
1GK8 A122-129	1000	1,40253	10^{-4}	24	36,5469
	2000	1,40253	10^{-5}	2	70,6094
	10000	1,40253	10^{-6}	2	353,563
1IXH A106-113	1000	0,968921	10^{-4}	8	37,1406
	2000	0,968921	10^{-5}	2	72,4219
	10000	0,968921	10^{-6}	18	362,984
1LAM A420-427	1000	1,6673	10^{-4}	16	36,4688
	2000	1,6673	10^{-5}	4	73,2969
	10000	1,6673	10^{-5}	18	364,969

Tabela 4 – Loop de 12 resíduos

Loop	Amostras	Distância Intervalar	Precisão	Soluções	CPU (seg)
1CTQ A26-37	1000	1,59584	10^{-4}	4	36,3125
	2000	1,59584	10^{-4}	16	72,9219
	10000	1,59584	10^{-5}	10	367,234
1CRU A358-369	1000	2,544209	10^{-4}	2	36,6875
	2000	2,544209	10^{-4}	6	73,625
	10000	2,544209	10^{-5}	2	380,641
1D4O A88-99	1000	0,398333	10^{-4}	26	36
	2000	0,398333	10^{-5}	2	71,9063
	10000	0,398333	10^{-6}	4	365,359
1D8W A46-57	1000	0,915326	10^{-4}	10	36,5625
	2000	0,915326	10^{-5}	4	73,9531
	10000	0,915326	10^{-5}	16	374,578
1DS1 A282- 293	1000	0,985893	10^{-4}	22	36,1094
	2000	0,985893	10^{-5}	4	71,8438
	10000	0,985893	10^{-6}	4	365,054
1DYS A291-302	1000	0,56258	10^{-4}	2	36,1563
	2000	0,56258	10^{-4}	2	73,2818
	10000	0,56258	10^{-5}	2	364,531
1EGU A508-519	1000	1,26185	10^{-4}	24	35,8125
	2000	1,26185	10^{-5}	8	72,8438
	10000	1,26185	10^{-5}	24	358,169

Na Tabela 4, trazemos os resultados dos testes realizados em loops compostos por 12 resíduos. Agora, temos 36 átomos em cada loop, como os 3 primeiros são fixos, temos 2^{36-3} possibilidades de loops para cada amostra retirada na distância intervalar. O tempo de execução do algoritmo é semelhante aos encontradas em instâncias menores e está ligada principalmente ao número de amostras consideradas na distância intervalar e não ao número de loops explorados em cada uma das amostras. A quantidade de soluções encontradas varia entre 2 e 28 dependendo do número de amostras solicitadas inicialmente na distância intervalar e do erro aceito nas etapas de verificação de factibilidade.

Os resultados dos testes nos conjuntos de instâncias, apresentados pelas Tabelas 2, 3 e 4, demonstram que nosso método, sintetizado no Algoritmo iBP, é capaz de resolver problemas de loops fechados maiores, isto é, loop formados por um número maior de resíduos com tempo de execução semelhante a problemas de loops fechados formados por menos resíduos.

A seguir, vamos comparar o número de soluções encontradas pelos métodos iBP e CSJD e também o utilizaremos as medida de RMSD ([KUFAREVA; ABAGYAN, 2011](#)) para medir a qualidade das soluções encontradas por cada método, veja as Tabelas 5, 6 e 7.

Observando as Tabelas 5, 6 e 7, notamos que os números de soluções encontradas

Tabela 5 – Número de soluções em loop de 4 resíduos

Loop	CSJD	RMSD	iBP	RMSD
1DVJ A20-23	4548	0,38	18	0,212532
1DYS A47-50	2234	0,37	4	0,568002
1EGU A74-77	170	0,37	2	0,657145
1EJO A74-77	1564	0,21	2	0,00764602
1QNR A195-198	1064	0,39	2	0,69244
1QOP A44-47	4248	0,61	12	0,0105465
1TCA A95-98	418	0,28	4	0,342475
Média	–	0,372857	–	0,35582664

Tabela 6 – Número de soluções em loop de 8 resíduos

Loop	CSJD	RMSD	iBP	RMSD
1CRU A85-92	2516	0,99	2	$2,77612 \cdot 10^{-6}$
1CTQ A144-151	1754	0,96	2	0,834476
1D8W A334-341	1686	0,37	2	0,000347341
1DS1 A20-27	3506	1,30	4	0,549372
1GK8 A122-129	2362	1,29	2	0,775405
1IXH A106-113	4448	2,36	2	1,07391
1LAM A420-427	2200	0,83	18	0,783793
Média	–	1,014285714	–	0,5739005169

Tabela 7 – Número de soluções em loop de 12 resíduos

Lopp	CSJD	RMSD	iBP	RMSD
1CRU A358- 369	4148	2,00	2	0,501942
1CTQ A26-37	3968	1,86	10	0,726536
1D4O A88-99	1802	1,60	22	0,00414742
1D8W A46-57	3906	2,94	16	0,58969
1DS1 A282- 293	1162	3,10	28	0,0103844
1DYS A291-302	2306	3,04	2	0,710764
1EGU A508-519	2924	2,82	24	0,00684338
Média	–	2,48	–	0,3643296

pelo iBP é muito menor que os números de soluções encontradas pelo CSJD. Além disso, na Tabela 5, apesar dos números menores de soluções fornecidas pelo iBP, temos que os menores valores de RMSD encontradas para as estruturas de cada loop formados por 4 resíduos são, em média, parecidos nos dois métodos. Já, na Tabela 6, observamos que os menores valores de RMSD das soluções dos loops de 8 resíduos obtidas pelo iBP são, em média, razoavelmente melhores que os menores valores de RMSD encontrados nas estruturas obtidas do CSJD. Na Tabela 7, notamos que os menores valores de RMSD encontrados nas soluções fornecidas pelo iBP se mantêm constantes em loops de 12 resíduos quando comparados aos resultados encontrados em loops de 4 e 8 resíduos, mas os menores valores de RMSD encontrados nas soluções fornecidas pelo CSJD em loops de 12 resíduos apresentaram um crescimento significados quando comparados aos resultados obtidos nos

Tabela 8 – RMSD e LDME para soluções iBP de loop de 4 resíduos

Loop	iBP	RMSD	LDME
1DVJ A20-23	18	0,212532	$4,16184 \cdot 10^{-5}$
1DYS A47-50	4	0,568002	$3,0194 \cdot 10^{-3}$
1EGU A74-77	2	0,657145	$1,02276 \cdot 10^{-2}$
1EJ0 A74-77	2	0,00764602	$8,0343 \cdot 10^{-7}$
1QNR A195-198	2	0,69244	$2,62285 \cdot 10^{-3}$
1QOP A44-47	12	0,0105465	$5,22789 \cdot 10^{-7}$
1TCA A95-98	4	0,342475	$6,40289 \cdot 10^{-4}$
Média	–	0,35582664	–

Tabela 9 – RMSD e LDME para soluções iBP de loop de 8 resíduos

Loop	iBP	RMSD	LDME
1CRU A85-92	2	$2,77612 \cdot 10^{-6}$	$6,63657 \cdot 10^{-8}$
1CTQ A144-151	2	0,834476	$2,97047 \cdot 10^{-2}$
1D8W A334-341	2	0,000347341	$1,52366 \cdot 10^{-9}$
1DS1 A20-27	4	0,549372	$4,23032 \cdot 10^{-3}$
1GK8 A122-129	2	0,775405	$4,27788 \cdot 10^{-2}$
1IXH A106-113	2	1,07391	$3,07328 \cdot 10^{-2}$
1LAM A420-427	18	0,783793	$3,67651 \cdot 10^{-2}$
Média	–	0,5739005169	–

Tabela 10 – RMSD e LDME para soluções iBP de loop de 12 resíduos

Lopp	iBP	RMSD	LDME
1CRU A358-369	2	0,501942	$7,1699 \cdot 10^{-4}$
1CTQ A26-37	10	0,726536	$1,94991 \cdot 10^{-3}$
1D4O A88-99	22	0,00414742	$2,21691 \cdot 10^{-7}$
1D8W A46-57	16	0,58969	$2,19708 \cdot 10^{-2}$
1DS1 A282-293	28	0,0103844	$8,12158 \cdot 10^{-7}$
1DYS A291-302	2	0,710764	$1,88575 \cdot 10^{-5}$
1EGU A508-519	24	0,00684338	$1,26411 \cdot 10^{-6}$
Média	–	0,3643296	–

loops formados por 4 e 8 resíduos. Dessa forma, concluímos que as soluções encontradas pelo iBP possuem uma melhor “qualidade” quando avaliadas pela medida RMSD.

A seguir, para destacar a “qualidade” das soluções encontradas pelo iBP, submetemos essas soluções ao teste de medida de qualidade denominado LDME (SOUZA et al., 2013), veja as Tabelas 8, 9 e 10.

Observamos, nas Tabelas 8, 9 e 10, que as soluções encontradas apresentam resultados de LDME em grandezas que variam de 10^{-9} à 10^{-2} , resultados que indica que as soluções encontradas pelo iBP praticamente sobrepõem a solução original conhecida.

6 Considerações Finais

A principal contribuição deste trabalho foi apresentar uma resolução alternativa para o PLF utilizando a GD. Baseado nesta abordagem, foi implementado o Algoritmo denominado iBP que utiliza ferramentas da GD para encontrar soluções do PLF. As principais características deste algoritmo são o tempo de execução depender principalmente do número de amostras retiradas na distância intervalar e a “qualidade” das soluções encontradas.

Primeiramente, determinamos uma ordem para os átomos do PLF que satisfaz a definição do $iPDGD_3$. Observações teóricas sobre a cardinalidade do conjunto de solução e simetrias entre soluções foram apresentadas, demonstradas e ilustradas no Capítulo 3.

Experimentos numéricos, utilizando loops conhecidos na literatura, apresentados em (CANUTESCU; DUNBRACK, 2003) e (COUTSIAS et al., 2004), ilustram o desempenho do iBP. Comparações qualitativas entre o iBP e os métodos CCS e CSJD são apresentadas no Capítulo 4. Em geral, as soluções encontradas pelo iBP passam por testes de factibilidade sempre que existem distâncias extras conhecidas, etapas que não ocorrem nos demais métodos.

Como era esperado, devido a utilização de mais informações nas etapas de testes de factibilidade, o método apresentado se mostrou muito eficiente e as soluções encontradas apresentaram resultados semelhantes ou melhores nos testes de RSMD em relação as soluções encontradas por outros métodos. Além disso, o iBP apresentou resultado satisfatórios no testes de LDME destacando a “qualidade ” de suas soluções.

Referências

- BERMAN, H. M.; BATTISTUZ, T.; BHAT, T. N.; BLUHM, W. F.; BOURNE, P. E.; BURKHARDT, K.; FENG, Z.; GILLILAND, G. L.; IYPE, L.; JAIN, S. et al. The protein data bank. *Acta Crystallographica Section D: Biological Crystallography*, International Union of Crystallography, v. 58, n. 6, p. 899–907, 2002. Citado na página 67.
- CAMARGO, V. S. d.; LAVOR, C. Álgebra geométrica conforme e geometria de distâncias. [sn], 2015. Nenhuma citação no texto.
- CANUTESCU, A. A.; DUNBRACK, J. R. L. Cyclic coordinate descent: A robotics algorithm for protein loop closure. *Protein Science*, Cold Spring Harbor Laboratory, v. 12, p. 963–972, 2003. Citado 7 vezes nas páginas 14, 28, 42, 44, 56, 66 e 72.
- COOPE, I. Reliable computation of the points of intersection of n spheres in r^n . *ANZIAM Journal*, v. 42, p. 461–477, 2000. Citado na página 16.
- COUTSIAS, E. A.; SEOK, C.; JACOBSON, M. P.; DILL, K. A. A kinematic view of loop closure. *Journal of Computational Chemistry*, Wiley Periodicals, Inc, v. 25, n. 4, p. 510–528, 2004. Citado 13 vezes nas páginas 14, 25, 26, 27, 28, 29, 37, 41, 48, 56, 66, 67 e 72.
- CRIPPEN, G.; HAVEL, T. *Distance Geometry and Molecular Conformation*. New York: Wiley, 1988. Citado 3 vezes nas páginas 14, 15 e 20.
- GO, N.; SCHERAGA, H. A. Ring closure and local conformational deformations of chain molecules. *Macromolecules*, ACS Publications, v. 3, n. 2, p. 178–187, 1970. Citado 2 vezes nas páginas 26 e 27.
- GOLÇALVES, D. S.; MUCHERINO, A.; LIBERTI, L.; LAVOR, C. Recent advances on the interval distance geometry. *Journal of Global Optimization*, Springer, v. 69, n. 3, p. 525–545, 2017. Citado 5 vezes nas páginas 14, 15, 20, 21 e 22.
- HARTENBERG, R.; DANAVIT, J. *Kinematic Synthesis of Linkages*. [S.l.]: New York: McGraw-Hill, 1964. Citado na página 25.
- HENRICI, P. *Applied and Computational Complex Analysis*. [S.l.]: Wiley, New York, 1974. Citado 2 vezes nas páginas 41 e 56.
- KUFAREVA, I.; ABAGYAN, R. Methods of protein structure comparison. In: *Homology Modeling*. [S.l.]: Springer, 2011. p. 231–257. Citado 2 vezes nas páginas 66 e 69.
- LEE, H.-Y.; LIANG, C.-G. Displacement analysis of the general spatial 7-link 7r mechanism. *Mechanism and Machine Theory*, Elsevier, v. 23, n. 3, p. 219–226, 1988. Citado na página 26.
- LIBERTI, L.; LAVOR, C.; MACULAN, N.; MUCHERINO, A. Euclidean distance geometry and applications. *SIAM Review*, SIAM, v. 56, n. 1, p. 3–69, 2014. Citado 2 vezes nas páginas 14 e 15.

- LIBERTI, L.; MASSOM, B.; LEE, J.; LAVOR, C.; MUCHERINO, A. On the number of realizations of certain hennegberg graphs arising in protein conformation. *Discrete Applied Mathematics*, Elsevier, v. 165, p. 213–232, 2014. Citado 2 vezes nas páginas 15 e 18.
- MAIOLI, D.; GOLÇALVES, D. S.; LAVOR, C. A note on computing the intersection of spheres in R^n . *The ANZIAM Journal*, v. 59, n. 2, p. 271 – 279, 2017. Citado na página 16.
- MUCHERINO, A.; LAVOR, C.; LIBERTI, L. The discretizable distance geometry problem. *Optimization Letters*, Springer, v. 6, n. 8, p. 1671–1686, 2012. Citado na página 19.
- SOUZA, M.; LAVOR, C.; MURITIBA, A.; MACULAN, N. Solving the molecular distance geometry problem with inaccurate distance data. *BMC bioinformatics*, Springer, v. 14, n. 9, p. 1–6, 2013. Citado 2 vezes nas páginas 66 e 71.
- TRAMONTANO, A.; LEPLAE, R.; MOREA, V. Analysis and assessment of comparative modeling predictions in casp4. *Proteins: Structure, Function, and Bioinformatics*, Wiley Online Library, v. 45, n. S5, p. 22–38, 2001. Citado 2 vezes nas páginas 14 e 25.