



UNIVERSIDADE ESTADUAL DE
CAMPINAS

Instituto de Matemática, Estatística e
Computação Científica

WAGNER ALAN APARECIDO DA ROCHA

Ordenação em Vértices de Grafos de Proteínas

Campinas

2022

Wagner Alan Aparecido da Rocha

Ordenação em Vértices de Grafos de Proteínas

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 Aplicada.

Orientador: Carlile Campos Lavor

Coorientador: Ícaro Putinhon Caruso

Este trabalho corresponde à versão final da Tese defendida pelo aluno Wagner Alan Aparecido da Rocha e orientada pelo Prof. Dr. Carlile Campos Lavor.

Campinas

2022

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

R582o Rocha, Wagner Alan Aparecido da, 1993-
Ordenação em vértices de grafos de proteínas / Wagner Alan Aparecido da Rocha. – Campinas, SP : [s.n.], 2022.

Orientador: Carlile Campos Lavor.

Coorientador: Ícaro Putinhon Caruso.

Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Geometria de distâncias. 2. Ressonância magnética nuclear. 3. Proteínas - Estrutura. I. Lavor, Carlile Campos, 1968-. II. Caruso, Ícaro Putinhon. III. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: A new vertex order for protein graphs

Palavras-chave em inglês:

Distance geometry

Nuclear magnetic resonance

Proteins - Structure

Área de concentração: Matemática Aplicada

Titulação: Doutor em Matemática Aplicada

Banca examinadora:

Carlile Campos Lavor [Orientador]

Douglas Soares Gonçalves

Kelly Cristina Poldi

Luiz Leduino de Salles Neto

Rafael Santos de Oliveira Alves

Data de defesa: 07-07-2022

Programa de Pós-Graduação: Matemática Aplicada

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

- ORCID do autor: <https://orcid.org/0000-0002-3894-4002>

- Currículo Lattes do autor: <http://lattes.cnpq.br/5171808268756071>

**Tese de Doutorado defendida em 07 de julho de 2022 e aprovada
pela banca examinadora composta pelos Profs. Drs.**

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

Prof(a). Dr(a). KELLY CRISTINA POLDI

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

Prof(a). Dr(a). LUIZ LEDUINO DE SALLES NETO

Prof(a). Dr(a). RAFAEL SANTOS DE OLIVEIRA ALVES

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.

Dedico este trabalho a minha família.



Agradecimentos

Em primeiro lugar gostaria de agradecer o meu orientador Professor Carlile Lavor, sou muito grato por ter aceitado me orientar, ter me apresentado um problema de Matemática Aplicada tão instigante e interessante que virou tema de minha tese, ser tão paciente em todas as nossas discussões, ter me aconselhado e apoiado em muitas decisões que tomei ao longo do meu Doutorado. Carrego comigo a certeza de que fiz um amigo pra vida toda. Também gostaria de externar meus mais sinceros agradecimentos ao meu coorientador, o Professor Ícaro Putinhon Caruso, por ter aceitado a tarefa de participar deste trabalho enfrentando a barreira inerente de qualquer projeto interdisciplinar. A sua ajuda mudou completamente o meu entendimento sobre a parte biológica do problema que tratei nesta tese.

Ao Professor Leo Liberti eu agradeço por não ter medido esforços para que eu conseguisse realizar um período sanduíche na École Polytechnique de Paris, minha passagem por esta instituição mudou completamente o nível deste trabalho e me proporcionou uma experiência de vida única e inesquecível, de modo que, eu posso afirmar sem nenhum constrangimento que essa experiência mudou completamente a minha vida!

Aos Professores Michael Ferreira de Souza e Antonio Mucherino eu agradeço por todo o tempo dedicado e todas as discussões sobre o Problema de Geometria de Distâncias. Todo o tempo que passei na presença de vocês eu cresci profissionalmente na área que escolhi como carreira de vida!

A Professora Kelly Cristina Poldi e aos Professores Douglas Soares Gonçalves, Luiz Leduino de Salles Neto e Rafael Santos de Oliveira Alves eu apresento meus sinceros agradecimentos por terem lido cuidadosamente este texto, pelas discussões realizadas no dia da minha defesa, pelas críticas apresentadas e pelas sugestões de correções para que eu pudesse melhorar a minha tese.

A todas as Professoras e todos os Professores que fizeram parte da minha caminhada ao longo dos onze anos que passei pela UNICAMP, o meu desenvolvimento profissional se deu graças a interação que pude realizar com todos vocês! O meu mais genuíno

agradecimento a todas e todos por *ensignarem* muito mais do que conhecimento matemático e científico na minha vida!

A todas e todos funcionários da secretaria de pós-graduação do IMECC eu gostaria de agradecer por todo trabalho prestado, olhando para trás eu vejo o quanto ajuda de vocês foi imprescindível para que a minha caminhada pela pós-graduação no IMECC fosse possível.

A todos os heróis anônimos que trabalham pela UNICAMP, sejam nos restaurantes universitários, na bibliotecas, na limpeza dos espaços físicos, entre tantos outros. Meu mais profundo agradecimento a vocês que tornam possível a vida na universidade, que “*fazem da coragem a espinha dorsal, que são a real coluna social*” da UNICAMP.

Para a minha mãe Irma e meu pai João eu não encontrei uma palavra no nosso vocabulário para expressar o tamanho da minha gratidão. Muito obrigado pelo apoio irrestrito que me deram durante todo o período que estive na UNICAMP. A maneira que vocês me ensinaram a enfrentar a vida, sempre com humildade, me fez superar qualquer limite imposto pela sociedade e chegar até aqui!

Ao meu irmão Paulo eu agradeço por ter sido a minha primeira referência sobre a importância dos estudos nas nossas vidas. Muito obrigado por tudo que me ensinou e ainda me ensina! A minha cunhada Luiza eu sou grato por ser um grande exemplo de coragem e resistência. Ao Matheus, muito obrigado por ter feito a vida do tio mais leve e divertida desde a sua chegada nesse mundo.

A Incrível Cientista Letícia, minha mais que amiga, eu agradeço por todo o companheirismo ao longo de todo o meu doutorado. Dificilmente conseguiria me lembrar de todas as situações em que sua ajuda foi imprescindível durante esse período. No entanto, gostaria de lembrar que sua incansável ajuda e confiança foram vitais na minha aprendizagem da língua inglesa, o que tornou possível grande parte dos feitos desse trabalho!

A todos os amigos e amigas que fizeram parte da minha vida nestes onze anos de UNICAMP meu muito obrigado! A companhia de vocês me fez ser quem eu sou hoje e, conseqüentemente, isso foi crucial para a conclusão deste trabalho. Para não correr o risco de esquecer de nenhum de vocês, não vou citar ninguém nominalmente, mas se você me proporcionou algum momento de felicidade, seja em um rolê sem sentido, um churrasco, uma festa, se estudamos juntos, se frequentei sua salinha ou seu laboratório, se você frequentou minha salinha, se foi meu aluno ou minha aluna, se foi meu professor ou professora, se discutimos juntos, se participamos de algum rolê cultural ou até mesmo se ficamos bêbados juntos; saiba que você me ajudou a construir este trabalho!

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001. Este trabalho também contou com o apoio do Conselho Nacional de Desenvolvimento Científico

e Tecnológico (CNPq) - Número de Processo 156812/2019-3. Este trabalho só pode ser realizado graças ao financiamento que estas duas agências de fomento me propiciaram. O Programa de Excelência Acadêmica (Proex) da CAPES financiou 16 meses deste trabalho; o CNPq financiou este projeto por 26 meses; e o Programa Institucional de Internacionalização (PrInt) da CAPES financiou 6 meses deste trabalho. Gostaria de externar meus sinceros agradecimentos a estas agências de fomento que são imprescindíveis para o desenvolvimento da ciência brasileira.

Para encerrar os agradecimentos, este trabalho é a realização de um sonho, um sonho que fica compartilhado com todo aquele que se dispuser a ler esta tese. Deste modo, a melhor maneira que eu encontro para terminar esta seção é citando o Poeta *Sérgio Vaz*:

“Enquanto eles capitalizam a realidade, eu socializo meus sonhos”.

*“Poucas palavras, tio, vou ser breve
Se a história é nossa deixa que nós escreve!”*

Renan Inquérito.

Resumo

Apresentando-se como uma nova face da geometria, desde sua criação até os dias atuais, a teoria de Geometria de Distâncias mostrou-se muito atrativa na modelagem de vários problemas inversos. Nesta tese, propomo-nos trabalhar com o problema inverso da conformação tridimensional de proteínas considerando um conjunto de restrições de distâncias entre seus átomos: um problema fundamental da área de geometria computacional para a determinação de estruturas biomoleculares. Este é um problema NP-difícil, conhecido na literatura como Problema de Geometria de Distâncias Moleculares (PGDM). Para que a modelagem deste problema seja condizente com a realidade, consideramos que as distâncias que definem o PGDM são provenientes de experimentos de Ressonância Magnética Nuclear (RMN). Modelando este problema com um grafo simples, ponderado e não-direcionado, descrevemos uma ordenação nos vértices deste grafo que viabiliza a discretização do espaço de busca de soluções. Esta abordagem combinatória permite a utilização do algoritmo *interval Branch & Prune* (*iBP*), que já é bem estabelecido da literatura. No entanto, a observação de relações geométricas definidas em alguns átomos relacionados a cadeia principal das proteínas nos possibilitou a proposta de um novo algoritmo, baseado na estratégia *iBP*, para resolver o PGDM descrito por dados experimentais de RMN. Como método de análise da eficácia do algoritmo proposto, resultados de testes computacionais em instâncias geradas a partir de estruturas do *Protein Data Bank* (PDB) nos mostram vantagens deste algoritmo sobre o *iBP*.

Palavras-chave: Geometria de Distâncias; Estrutura Molecular 3D; Problema Discretizável de Geometria de Distâncias; Ressonância Magnética Nuclear.

Abstract

Presenting itself as a new face of geometry, from its creation to the present day, the theory of Distance Geometry proved to be very attractive in modeling several inverse problems. This thesis proposes to work with the inverse problem of protein three-dimensional structure conformation considering a set of distance constraints on the protein atoms: a fundamental problem in the computational geometry area for determining biomolecular structures. This inverse problem is an NP-hard problem known in the literature as Molecular Distance Geometry Problem (MDGP). For the modeling of this problem to be consistent with reality, we consider that the distances that define the MDGP come from Nuclear Magnetic Resonance (NMR) experiments. Modeling this problem with a simple, weighted, and undirected graph, we describe an ordering in the vertices of this graph that allows the discretization of the solution search space. This combinatorial approach enables the application of the interval Branch & Prune (*iBP*) algorithm, which is already well established in the literature. However, the observation of geometric relationships in some atoms related to the backbone of proteins allowed us to propose a new algorithm, based on the *iBP* strategy, to solve the MDGP described by experimental NMR data. As a method of analyzing the effectiveness of the proposed algorithm, the results of computational tests on instances generated from Protein Data Bank (PDB) structures show us the advantages of this algorithm over *iBP*.

Keywords: Distance Geometry; 3D Molecular Structure; Discretizable Distance Geometry Problem; Nuclear Magnetic Resonance.

Lista de ilustrações

Figura 1 – Lista de aminoácidos. Imagem de COJOCARI (2010), modificada. . . .	27
Figura 2 – Estrutura geral de um aminoácido.	28
Figura 3 – Processo de junção de dois aminoácidos na formação de um dipeptídeo.	28
Figura 4 – Gráfico de Ramachandran. Imagem de RAMACHANDRAN; RAMAKRISHNAN; SASISEKHARAN (1963).	29
Figura 5 – Dois pontos de vista de uma α -hélice. Modelo 1 da estrutura 1dep do PDB.	30
Figura 6 – Arranjos estruturais folhas- β . Modelo 1 da estrutura 5o6f do PDB.	30
Figura 7 – Molécula de Mioglobina. Modelo 1 da estrutura 2spl do PDB.	31
Figura 8 – Molécula relacionada à proteína <i>spike</i> do <i>SARS-CoV-2</i> (covid-19). Modelo 1 da estrutura 7lm8 do PDB.	31
Figura 9 – Exemplos de espectros de RMN unidimensional (esquerda) e bidimensional (direita). Imagem de NELSON; COX (2012), modificada.	33
Figura 10 – Casos possíveis para a intersecção de três esferas no espaço.	41
Figura 11 – Exemplo de árvore binária com 8 nós.	45
Figura 12 – Espaço de busca de um PDGD com sete vértices.	46
Figura 13 – Exemplo iterativo do funcionamento do BP. <i>Link</i> contido no <i>QR code</i> : https://bit.ly/3t1Fx1Z	48
Figura 14 – Disposição dos pontos x_{i_3} , x_{i_2} , x_{i_1} e x_i no espaço, ressaltando os dois planos definidos por estes quatro pontos. A distância entre alguns destes pontos é destacada pela linha contínua azul, e, θ , ϑ e τ evidenciam todos os ângulos necessários para definir a localização geométrica dos quatro pontos.	50
Figura 15 – Circunferência tangente aos vetores normais dos planos $\{x_{i_3}, x_{i_2}, x_{i_1}\}$ e $\{x_{i_2}, x_{i_1}, x_i\}$ com centro c na reta formada pelos pontos x_{i_2} e x_{i_1} . O ponto p é dado pela intersecção desta circunferência com o plano $\{x_{i_3}, x_{i_2}, x_{i_1}\}$	51

Figura 16 – Interseção das esferas $S_2^i = S(x_{i_2}, d_{i,i_2})$ e $S_1^i = S(x_{i_1}, d_{i,i_1})$, representadas, respectivamente, pelas cores azul e ciano. O plano π que contém esta intersecção é apresentado em verde.	52
Figura 17 – Representação geométrica dos vetores v , r e s . Aqui os vetores ortogonais \hat{r} e s foram encontrados pela aplicação do processo de ortogonalização de Gram-Schmidt nos vetores \hat{r} e \hat{v}	54
Figura 18 – Triângulo reto contido no plano $\{x_{i_2}, x_{i_1}, x_i\}$ definido pelos pontos x_{i_2} , c e x_i	54
Figura 19 – Ilustração da rotação do ponto x através do eixo definido pela reta e que contém a origem O do espaço.	57
Figura 20 – Ângulo triédrico com vértice no ponto S , como descrito no Teorema 3.1.	58
Figura 21 – Ângulo triédrico formado pelos pontos x_{i_3} , x_{i_2} , x_{i_1} e x_i com vértice em x_{i_2}	58
Figura 22 – Descrição geométrica de vetores que evidenciam a projeção ortogonal do ponto x_w no plano $\pi = \{x_z, x_u, x_v\}$	75
Figura 23 – Representação do i -ésimo resíduo de uma proteína, destacando todos os átomos envolvidos e os ângulos de torção ω_i , ϕ_i e ψ_i	79
Figura 24 – Representação do plano peptídico no entorno do i -ésimo resíduo de uma proteína.	79
Figura 25 – Quiralidade dos átomos $\{N, C_\alpha, C, H_\alpha\}$	81
Figura 26 – (a) Uma posição espacial para os átomos C^{i-1} , N^i , H_N^i , C_α^i , H_α^i e C^i se $\phi_i = 30^\circ$. (b) Planos $\{C^{i-1}, N^i, C_\alpha^i\}$, $\{N^i, C_\alpha^i, C^i\}$ e $\{N^i, C_\alpha^i, H_\alpha^i\}$ identificados na estrutura da figura (a). (c) Estrutura da figura (b) rotacionada de modo a identificar um ponto de vista de interesse. (d) Projeção ortogonal da estrutura da figura (c) no plano, identificando os ângulos de torção $\phi_i = \{C^{i-1}, N^i, C_\alpha^i, C^i\}$, $\tau_{C^{i-1}, H_N^i} = \{C^{i-1}, N^i, C_\alpha^i, H_N^i\}$, $\tau_{C^i, H_\alpha^i} = \{C^i, N^i, C_\alpha^i, H_\alpha^i\}$ e $\tau_{H_N^i, H_\alpha^i} = \{H_N^i, N^i, C_\alpha^i, H_\alpha^i\}$ como ângulos planares.	83
Figura 27 – Relações que descrevem a dinâmica entre os ângulos de torção ϕ_i , τ_{C^{i-1}, H_N^i} , τ_{C^i, H_α^i} e $\tau_{H_N^i, H_\alpha^i}$. O caso $0^\circ \leq \phi_i \leq 180^\circ$ é ilustrado pela sequência de imagens (a)→(c)→(d)→(e)→(b), e o caso $-180^\circ < \phi_i \leq 0^\circ$ é ilustrado pela sequência de imagens (a)→(f)→(g)→(h)→(b).	85
Figura 28 – Identificação dos ângulos de torção ϕ_i , τ_{C^{i-1}, H_N^i} , τ_{C^i, H_α^i} e $\tau_{H_N^i, H_\alpha^i}$ para as situações (I) e (II).	86

Figura 29 – (a) Uma posição espacial para os átomos C^{i-1} , N^i , C_α^i , $H_{\delta_2}^i$, $H_{\delta_3}^i$, C_α^i , H_α^i e C^i se $\phi_i = 20^\circ$. (b) Planos $\{C^{i-1}, N^i, C_\alpha^i\}$, $\{N^i, C_\alpha^i, C^i\}$ e $\{N^i, C_\alpha^i, H_{\delta_3}^i\}$ identificados na estrutura da figura (a). (c) Estrutura da figura (b) rotacionada de modo a identificar um ponto de vista de interesse. (d) Projeção ortogonal da estrutura da figura (c) no plano, identificando os ângulos de torção $\phi_i = \{C^{i-1}, N^i, C_\alpha^i, C^i\}$, $\tau_{C^{i-1}, H_{\delta_3}^i} = \{C^{i-1}, N^i, C_\alpha^i, H_{\delta_3}^i\}$, $\tau_{C^i, H_\alpha^i} = \{C^i, N^i, C_\alpha^i, H_\alpha^i\}$ e $\tau_{H_{\delta_3}^i, H_\alpha^i} = \{H_{\delta_3}^i, N^i, C_\alpha^i, H_\alpha^i\}$ como ângulos planares.	88
Figura 30 – Identificação dos ângulos de torção ϕ_i , τ_{C^{i-1}, H_k^j} , τ_{C^i, H_α^i} e $\tau_{H_k^j, H_\alpha^i}$ para as situações (i) e (ii).	90
Figura 31 – Posição dos átomos C^{i-1} , C_α^i , H_α^i e C^i no plano da projeção ortogonal se $\phi_i = 0^\circ$	93
Figura 32 – Disposição no plano da projeção ortogonal dos átomos H_k^j , C^{i-1} , C_α^i , H_α^i e C^i , considerando $\phi_i = 0^\circ$, $ \phi_i = 180^\circ$, $\phi_i = {}^1\phi_i^{*+}$ e $\phi_i = {}^1\phi_i^{*-}$, para o Caso 1.	95
Figura 33 – Dinâmica do ângulo ϕ_i com relação aos ângulos τ_{C^{i-1}, H_k^j} , τ_{C^i, H_α^i} e $ \tau_{H_k^j, H_\alpha^i} $ para o Caso 1.	96
Figura 34 – Disposição no plano da projeção ortogonal dos átomos H_k^j , C^{i-1} , C_α^i , H_α^i e C^i , considerando $\phi_i = 0^\circ$, $ \phi_i = 180^\circ$, $\phi_i = {}^2\phi_i^{*+}$ e $\phi_i = {}^2\phi_i^{*-}$, para o Caso 2.	98
Figura 35 – Dinâmica do ângulo ϕ_i com relação aos ângulos τ_{C^{i-1}, H_k^j} , τ_{C^i, H_α^i} e $ \tau_{H_k^j, H_\alpha^i} $ para o Caso 2.	100
Figura 36 – Disposição no plano da projeção ortogonal dos átomos H_k^j , C^{i-1} , C_α^i , H_α^i e C^i , considerando $\phi_i = 0^\circ$, $ \phi_i = 180^\circ$, $\phi_i = {}^3\phi_i^{*+}$ e $\phi_i = {}^3\phi_i^{*-}$, para o Caso 3.	101
Figura 37 – Dinâmica do ângulo ϕ_i com relação aos ângulos τ_{C^{i-1}, H_k^j} , τ_{C^i, H_α^i} e $ \tau_{H_k^j, H_\alpha^i} $ para o Caso 3.	102
Figura 38 – Disposição no plano da projeção ortogonal dos átomos H_k^j , C^{i-1} , C_α^i , H_α^i e C^i , considerando $\phi_i = 0^\circ$, $ \phi_i = 180^\circ$, $\phi_i = {}^4\phi_i^{*+}$ e $\phi_i = {}^4\phi_i^{*-}$, para o Caso 4.	104
Figura 39 – Dinâmica do ângulo ϕ_i com relação aos ângulos τ_{C^{i-1}, H_k^j} , τ_{C^i, H_α^i} e $ \tau_{H_k^j, H_\alpha^i} $ para o Caso 4.	105
Figura 40 – (a) Uma posição espacial para os átomos N^i , C_α^i , C^i , H_α^i , N^{i+1} e H_N^{i+1} onde $\psi_i = 40^\circ$. (b) Planos $\{N^i, C_\alpha^i, C^i\}$, $\{C_\alpha^i, C^i, H_\alpha^i\}$, $\{C_\alpha^i, C^i, N^{i+1}\}$ e $\{C_\alpha^i, C^i, H_N^{i+1}\}$ identificados na estrutura da figura (a). (c) Estrutura da figura (b) rotacionada identificando um ponto de vista de interesse. (d) Projeção ortogonal da estrutura da figura (c) no plano; os ângulos de torção $\psi_i = \{N^i, C_\alpha^i, C^i, N^{i+1}\}$, $\tau_{N^i, H_\alpha^i} = \{N^i, C_\alpha^i, C^i, H_\alpha^i\}$, $\tau_{N^{i+1}, H_N^{i+1}} = \{N^{i+1}, C_\alpha^i, C^i, H_N^{i+1}\}$ e $\tau_{H_\alpha^i, H_N^{i+1}} = \{H_\alpha^i, C_\alpha^i, C^i, H_N^{i+1}\}$ são identificados como ângulos planares.	108

Figura 41 – (a) Uma posição espacial para os átomos N^i , C_α^i , C^i , H_α^i , N^{i+1} e $H_{\delta_3}^{i+1}$ onde $\psi_i = 40^\circ$. (b) Planos $\{N^i, C_\alpha^i, C^i\}$, $\{C_\alpha^i, C^i, H_\alpha^i\}$, $\{C_\alpha^i, C^i, N^{i+1}\}$ e $\{C_\alpha^i, C^i, H_{\delta_3}^{i+1}\}$ identificados na estrutura da figura (a). (c) Estrutura da figura (b) rotacionada identificando um ponto de vista de interesse. (d) Projeção ortogonal da estrutura da figura (c) no plano; os ângulos de torção $\psi_i = \{N^i, C_\alpha^i, C^i, N^{i+1}\}$, $\tau_{N^i, H_\alpha^i} = \{N^i, C_\alpha^i, C^i, H_\alpha^i\}$, $\tau_{N^{i+1}, H_{\delta_3}^{i+1}} = \{N^{i+1}, C_\alpha^i, C^i, H_{\delta_3}^{i+1}\}$ e $\tau_{H_\alpha^i, H_{\delta_3}^{i+1}} = \{H_\alpha^i, C_\alpha^i, C^i, H_{\delta_3}^{i+1}\}$ são identificados como ângulos planares.	109
Figura 42 – Posição dos átomos N^i , C^i , H_N^{i+1} e N^{i+1} no plano da projeção ortogonal se $\psi_i = 0^\circ$ para as situações (i) e (ii).	111
Figura 43 – Conjunto de átomos de interesse para uma proteína composta por p resíduos de aminoácidos. Aqui $H_N^i = H_N^i$ ou $H_N^i = H_{\delta_3}^i$, a depender se o resíduo de aminoácido i é a prolina ou não.	114
Figura 44 – Nova ordem <i>hand-crafted</i> aplicada a uma proteína composta por quatro resíduos de aminoácidos.	115
Figura 45 – Metade da região da árvore do espaço de busca de soluções referente ao i -ésimo resíduo de aminoácido da ordem <i>hc</i> . Os arcos de circunferência simbolizam a existência de uma infinidade de posições possíveis para a imersão dos átomos que estão sobre eles. As setas representam distâncias “extras”, i.e., não usadas no processo de discretização, da ordem, cuja diferença no traço trazem a seguinte informação: a linha tracejada representa a existência de uma distância precisa; as linhas contínuas descrevem a existência de uma distância intervalar; e as linhas ponto-traço a possibilidade de existência de uma distância intervalar.	117
Figura 46 – Árvore da Figura 45 considerando informações de quiralidade local e da planaridade imposta pela ligação peptídica.	118
Figura 47 – Árvore da Figura 46 evidenciando a mudança das restrições de distância dos átomos H_N^i e H_α^i para os átomos N^i e C^i , respectivamente.	122
Figura 48 – Disposição no plano da projeção ortogonal dos átomos H_k^j , N^i , C^i , N^{i+1} e H_N^{i+1} , com $\psi_i = 0^\circ$, $ \psi_i = 180^\circ$, $\psi_i = {}^1\psi_i^{*+}$ e $\psi_i = {}^1\psi_i^{*-}$, para o Caso 1.	141
Figura 49 – Dinâmica do ângulo ψ_i com relação aos ângulos τ_{N^i, H_k^j} , $\tau_{N^{i+1}, H_N^{i+1}}$ e $ \tau_{H_k^j, H_N^{i+1}} $ para o Caso 1.	142
Figura 50 – Disposição no plano da projeção ortogonal dos átomos H_k^j , N^i , C^i , N^{i+1} e H_N^{i+1} , com $\psi_i = 0^\circ$, $ \psi_i = 180^\circ$, $\psi_i = {}^2\psi_i^{*+}$ e $\psi_i = {}^2\psi_i^{*-}$, para o Caso 2.	143
Figura 51 – Dinâmica do ângulo ψ_i com relação aos ângulos τ_{N^i, H_k^j} , $\tau_{N^{i+1}, H_N^{i+1}}$ e $ \tau_{H_k^j, H_N^{i+1}} $ para o Caso 2.	144
Figura 52 – Disposição no plano da projeção ortogonal dos átomos H_k^j , N^i , C^i , N^{i+1} e H_N^{i+1} , com $\psi_i = 0^\circ$, $ \psi_i = 180^\circ$, $\psi_i = {}^3\psi_i^{*+}$ e $\psi_i = {}^3\psi_i^{*-}$, para o Caso 3.	145

Figura 53 – Dinâmica do ângulo ψ_i com relação aos ângulos τ_{N^i, H_k^j} , $\tau_{N^{i+1}, H_{\aleph}^{i+1}}$ e $\left \tau_{H_k^j, H_{\aleph}^{i+1}} \right $ para o Caso 3.	147
Figura 54 – Disposição no plano da projeção ortogonal dos átomos H_k^j , N^i , C^i , N^{i+1} e H_{\aleph}^{i+1} , com $\psi_i = 0^\circ$, $ \psi_i = 180^\circ$, $\psi_i = {}^4\psi_i^{*+}$ e $\psi_i = {}^4\psi_i^{*-}$, para o Caso 4.	148
Figura 55 – Dinâmica do ângulo ψ_i com relação aos ângulos τ_{N^i, H_k^j} , $\tau_{N^{i+1}, H_{\aleph}^{i+1}}$ e $\left \tau_{H_k^j, H_{\aleph}^{i+1}} \right $ para o Caso 4.	149
Figura 56 – Disposição no plano da projeção ortogonal dos átomos H_k^j , N^i , C^i , N^{i+1} e H_{\aleph}^{i+1} , com $\psi_i = 0^\circ$, $ \psi_i = 180^\circ$, $\psi_i = {}^1\psi_i^{*+}$ e $\psi_i = {}^1\psi_i^{*-}$, para o Caso 1.	151
Figura 57 – Dinâmica do ângulo ψ_i com relação aos ângulos τ_{N^i, H_k^j} , $\tau_{N^{i+1}, H_{\aleph}^{i+1}}$ e $\left \tau_{H_k^j, H_{\aleph}^{i+1}} \right $ para o Caso 1.	153
Figura 58 – Disposição no plano da projeção ortogonal dos átomos H_k^j , N^i , C^i , N^{i+1} e H_{\aleph}^{i+1} , com $\psi_i = 0^\circ$, $ \psi_i = 180^\circ$, $\psi_i = {}^2\psi_i^{*+}$ e $\psi_i = {}^2\psi_i^{*-}$, para o Caso 2.	154
Figura 59 – Dinâmica do ângulo ψ_i com relação aos ângulos τ_{N^i, H_k^j} , $\tau_{N^{i+1}, H_{\aleph}^{i+1}}$ e $\left \tau_{H_k^j, H_{\aleph}^{i+1}} \right $ para o Caso 2.	155
Figura 60 – Disposição no plano da projeção ortogonal dos átomos H_k^j , N^i , C^i , N^{i+1} e H_{\aleph}^{i+1} , com $\psi_i = 0^\circ$, $ \psi_i = 180^\circ$, $\psi_i = {}^3\psi_i^{*+}$ e $\psi_i = {}^3\psi_i^{*-}$, para o Caso 3.	156
Figura 61 – Dinâmica do ângulo ψ_i com relação aos ângulos τ_{N^i, H_k^j} , $\tau_{N^{i+1}, H_{\aleph}^{i+1}}$ e $\left \tau_{H_k^j, H_{\aleph}^{i+1}} \right $ para o Caso 3.	158
Figura 62 – Disposição no plano da projeção ortogonal dos átomos H_k^j , N^i , C^i , N^{i+1} e H_{\aleph}^{i+1} , com $\psi_i = 0^\circ$, $ \psi_i = 180^\circ$, $\psi_i = {}^4\psi_i^{*+}$ e $\psi_i = {}^4\psi_i^{*-}$, para o Caso 4.	159
Figura 63 – Dinâmica do ângulo ψ_i com relação aos ângulos τ_{N^i, H_k^j} , $\tau_{N^{i+1}, H_{\aleph}^{i+1}}$ e $\left \tau_{H_k^j, H_{\aleph}^{i+1}} \right $ para o Caso 4.	160



Lista de tabelas

- Tabela 1 – Restrições de distâncias entre átomos de uma proteína tomadas com base no conjunto de hipóteses apresentado na Seção 4.1. Aqui, $k = 2, 3, \dots, p$. 116
- Tabela 2 – Resultados numéricos obtidos em instâncias artificiais geradas a partir de estruturas do PDB. No processamento dos algoritmos, consideramos que cada intervalo foi amostrado com três pontos ($D = 3$). Os valores de referência foram tomados na estrutura do PDB. 129
- Tabela 3 – Resultados numéricos obtidos em instâncias artificiais geradas a partir de estruturas do PDB. No processamento dos algoritmos, consideramos que cada intervalo foi amostrado com três pontos ($D = 5$). Os valores de referência foram tomados na estrutura do PDB. 130



Lista de algoritmos

Algoritmo 1 – Estrutura principal do algoritmo BP.	48
Algoritmo 2 – Estrutura principal do algoritmo i BP em \mathbb{R}^3	77
Algoritmo 3 – Estrutura principal do algoritmo i TBP.	123

Sumário

1	INTRODUÇÃO	21
1.1	Contextualização	21
1.2	Objetivos	23
1.3	Organização	24
2	PROTEÍNAS: O QUE SÃO E COMO DETERMINAR SUA ESTRUTURA	25
2.1	Conceitos Básicos Sobre Proteínas	27
2.2	Métodos Experimentais para Caracterização da Estrutura 3D de Proteínas	31
2.2.1	Ressonância Magnética Nuclear	32
3	PROBLEMA DISCRETIZÁVEL DE GEOMETRIA DE DISTÂNCIAS	35
3.1	Problema de Geometria de Distâncias	35
3.2	Problema Discretizável de Geometria de Distâncias em \mathbb{R}^3	37
3.2.1	Características de uma Solução do PDGD	40
3.2.2	O Algoritmo <i>Branch & Prune</i>	44
3.2.3	Uma Maneira Intuitiva de Calcular a Imersão de um Vértice no \mathbb{R}^3	49
3.3	Problema Discretizável de Geometria de Distâncias Intervalares em \mathbb{R}^3	62
4	ESTRUTURA TRIDIMENSIONAL DE PROTEÍNAS E RESTRIÇÕES DE DISTÂNCIAS	78
4.1	Restrições de Distâncias	78
4.2	Relação dos Ângulos de Torção ϕ_i com Distâncias Entre Hidrogênios Próximos	81
4.2.1	Primeiro Caso de Estudo	82
4.2.2	Segundo Caso de Estudo	86
4.2.3	Terceiro Caso de Estudo	87
4.2.4	Quarto Caso de Estudo	89
4.2.5	Relação Geral para Determinar o Ângulo de Torção ϕ_i	91

4.3	Relação dos Ângulos de Torção ψ_i com Distâncias Entre Hidrogênios Próximos	107
5	CARACTERIZAÇÃO DA ESTRUTURA 3D DO <i>BACKBONE</i> DE PROTEÍNAS	113
5.1	Uma Nova Ordem PDGD _{<i>i</i>} Para Grafos de Proteínas	113
5.2	Um Novo Algoritmo do Tipo <i>i</i> BP Aplicado a Proteínas	119
5.3	Experimentos Computacionais	122
6	CONCLUSÕES	132
	REFERÊNCIAS	134
APÊNDICE A	DESCRIÇÕES GEOMÉTRICAS UTILIZADAS NA SE- ÇÃO 4.3	140
A.1	Situação (<i>i</i>): $\tau_{N^{i+1}, H_N^{i+1}} \geq 0^\circ$	140
A.2	Situação (<i>ii</i>): $\tau_{N^{i+1}, H_N^{i+1}} < 0^\circ$	151

CAPÍTULO

1 Introdução

1.1 Contextualização

Na revista *Nature* de 1953, o bioquímico norte-americano James Watson e o biofísico inglês Francis Crick propuseram o que mais tarde ficou conhecido como o primeiro modelo 3D correto da estrutura tridimensional do DNA: a famosa dupla hélice ([WATSON; CRICK, 1953](#)). O modelo de dupla hélice, apresentado neste trabalho, foi baseado em uma única imagem de difração de raios-X, chamada *Photo 51*¹ ([NATURE, 2013](#)), e nas informações do bioquímico austríaco-americano Erwin Chargaff, fornecidas a eles por meio de um contato privado nos anos anteriores, sugerindo que o DNA tinha bases emparelhadas. Nesta mesma edição da *Nature*, foram apresentados cinco trabalhos que forneceram evidências experimentais que corroboraram com o modelo proposto por Watson e Crick ([PAPERS, 2003](#)). Um deles foi um trabalho de Franklin e Gosling, que apresentou seus dados de difração de raios-X e uma análise que apoiou o modelo de Watson e Crick. Esta edição também apresentou um artigo sobre a estrutura do DNA de Wilkins e colaboradores ([WILKINS; STOKES; WILSON, 1953](#)). Com esse modelo, comprovado experimentalmente em 1958, inaugurou-se uma revolução na compreensão do processo de hereditariedade. Watson, Crick e Wilkins receberam conjuntamente o Prêmio Nobel de Fisiologia e Medicina em 1962 ([NOBELPRIZE, 1962](#)), infelizmente após a morte de Franklin, em 1958. Ainda hoje, a quem esta descoberta deve ser atribuída é objeto de debate ([MADDOX, 2003](#)).

A nota histórica acima evidencia um dos primeiros trabalhos em que se constatou a íntima relação entre a estrutura tridimensional e a função de uma molécula. Não obstante, esta relação é particularmente verdadeira no caso das proteínas: macromoléculas fundamentais em qualquer sistema vivo.

¹ Feita em maio de 1952 pela química Rosalind Franklin e seu aluno de doutorado Raymond Gosling. Watson e Crick tiveram acesso a esta imagem, sem o conhecimento de Franklin, através do físico britânico Maurice Wilkins, colaborador de Franklin.

Por observações experimentais, é possível estabelecer um conjunto de restrições de distância entre alguns átomos das proteínas. Assim, nos problemas relacionados ao cálculo de estruturas 3D das proteínas, estabelece-se o seguinte problema inverso: como obter a posição espacial de todos os átomos de uma molécula dadas apenas distâncias entre alguns destes átomos? Este problema é NP-difícil (SAXE, 1980) e conhecido como Problema de Geometria de Distâncias Moleculares (PGDM) (CRIPPEN; HAVEL, 1988).

Nossa proposta é aplicar a teoria de Geometria de Distâncias (CRIPPEN; HAVEL, 1988) no cálculo de estruturas tridimensionais de proteínas cujas informações geométricas destas proteínas são provenientes de experimentos de Ressonância Magnética Nuclear (RMN) (WÜTHRICH, 1986). Diferentemente dos métodos tradicionais que são baseados em otimização contínua (LIBERTI et al., 2014), nosso intuito é explorar a natureza combinatória do PGDM utilizando as informações experimentais.

Como a posição no espaço que a molécula de proteína está localizada é irrelevante, fixa-se previamente os três primeiros átomos desta molécula e, com isso, descarta-se estruturas obtidas por translações e rotações. Definindo-se uma ordem apropriada entre os átomos da proteína, a distância entre quarto átomo aos três primeiros também pode ser conhecida (LAVOR et al., 2012a), com isso, têm-se duas possibilidades para a posição do quarto átomo. Essa é a chave para a discretização do problema (LAVOR et al., 2012b). Para tornar evidentes os aspectos combinatórios envolvidos neste problema, ele pode ser modelado por um grafo simples, ponderado e não-direcionado, onde cada vértice é associado a um átomo da proteína e quando a distância entre dois átomos é conhecida, estabelece-se uma aresta entre os vértices que representam estes átomos com peso dado pelo valor da distância. Assim, resolver este problema é obter uma imersão do grafo no espaço 3D, em que as distâncias euclidianas calculadas entre pares de átomos sejam iguais aos pesos das arestas correspondentes (LAVOR et al., 2012a).

O PGDM garante distâncias entre átomos próximos na molécula, assim, é possível definir uma ordem entre estes átomos de modo que a distância de cada átomo a , pelo menos, três átomos anteriores sejam conhecidas. A classe de problemas do PGDM com esta ordem é chamada de Problema Discretizável de Geometria de Distâncias (PDGD) (MUCHERINO; LAVOR; LIBERTI, 2012). Continuamos com um problema NP-difícil, mas com estas informações, demonstra-se que o espaço de busca tem uma estrutura de árvore binária, uma vez que os três primeiros átomos são fixados, a partir do quarto átomo têm-se duas possibilidades para a posição, no espaço 3D, de cada um destes átomos (MUCHERINO; LAVOR; LIBERTI, 2012). Entretanto, para cada átomo em que se conhece a distância deste átomo para outros quatro, pode-se testar as duas posições possíveis encontradas para este átomo, reduzindo o espaço de busca e direcionando a descida na árvore binária, o que é muito importante, uma vez que uma solução do problema pode ser definida por um caminho da raiz da árvore até um nó folha, onde todas as distâncias

conhecidas sejam satisfeitas (LAVOR et al., 2012a). O algoritmo *Branch & Prune* (BP) proposto em LIBERTI; LAVOR; MACULAN (2008), explora, de maneira organizada, o espaço de busca e encontra todas as soluções do PDGD.

1.2 Objetivos

Como nosso intuito pe considerar dados de RMN, para representarmos uma situação mais próxima da realidade devemos considerar as incertezas dos dados experimentais. Assim, naturalmente, vamos supor que as distâncias fornecidas pela RMN são intervalos de números reais, ou seja, distâncias intervalares, e não valores precisos (LAVOR; LIBERTI; MUCHERINO, 2013). A primeira tentativa de resolver o problema considerando distâncias intervalares foi baseada na amostragem de valores dentro dos intervalos relativos a estas distâncias em uma extensão do algoritmo BP, chamado de *interval Branch & Prune* (*iBP*) (LAVOR; LIBERTI; MUCHERINO, 2013). Essa ideia tornou o algoritmo BP uma heurística, já que não é mais possível garantir que sempre será encontrada uma solução, e isso é simplesmente justificado pelo fato de que o *iBP* pode não encontrar uma solução caso a distância correta não esteja entre os valores selecionados no processo de amostragem. Além disso, amostras refinadas aumentam exponencialmente o espaço de busca, a árvore associada deixa de ser binária (LAVOR; LIBERTI; MUCHERINO, 2013). No entanto, o trabalho LAVOR et al. (2021b) mostra que existe uma melhoria significativa nesta abordagem quando se considera todas as distâncias que se têm a disposição para a realização dos vértices envolvidos no problema. Com este processo consegue-se restringir o intervalo de possibilidades para a imersão de cada átomo, considerando as distâncias deste átomo para outros átomos prévios na ordem definida. Na prática, isto significa que utilizando mais informações conseguimos restringir o erro natural da medida de modo que a distância intervalar tenha um comprimento menor, o que auxilia na discretização destas distâncias intervalares.

O trabalho de LAVOR et al. (2011) inicia uma nova era na caracterização de uma proteína utilizando Geometria de Distâncias, uma vez que descreve uma ordem PDGD nos átomos iniciais da cadeia principal, da proteína, e nos átomos de hidrogênio ligados a esta cadeia. Entretanto, a proposta deste trabalho leva em conta que se conhecem distâncias entre átomos de hidrogênio que nem sempre são detectados pelo experimento de RMN. Dois anos depois, uma outra ordem foi proposta em LAVOR; LIBERTI; MUCHERINO (2013), desta vez considerando toda a cadeia principal e os átomos de hidrogênio ligados a ela. Estes dois trabalhos foram os embriões do trabalho LAVOR et al. (2019), que usamos como base para a construção desta tese.

Nesta tese, modelando um conjunto de átomos de uma proteína por um grafo ponderado e não-direcionado, descrevemos uma nova ordem PDGD que leva em conta aspectos teóricos, experimentais e geométricos sobre as proteínas. Para isso, consideramos uma ordenação do conjunto de vértices e a combinamos com duas relações geométricas entre átomos das proteínas que descrevemos neste trabalho. Assim, obtivemos uma ordem que considera o maior número de informações disponíveis na imersão de cada átomo. Isto possibilitou a proposta de um novo algoritmo, baseado na estratégia do *iBP*, para caracterizar a posição espacial dos átomos da cadeia principal de proteínas.

1.3 Organização

Este trabalho está organizado da seguinte maneira: no próximo capítulo, tratamos de conceitos básicos sobre proteínas e o experimento de Ressonância Magnética Nuclear; no [Capítulo 3](#), apresentamos o Problema Discretizável de Geometria de Distâncias em \mathbb{R}^3 e algumas definições e resultados que utilizamos para descrever matematicamente o problema proposto; no [Capítulo 4](#), estabelecemos as hipóteses da modelagem do nosso problema e apresentamos duas relações geométricas que observamos no conjunto de átomos que consideramos na nossa modelagem; no [Capítulo 5](#), propomos uma nova ordenação de vértices do grafo que modela o nosso problema de interesse, um novo algoritmo baseado na estratégia do *iBP* que trabalha com esta nova ordem e alguns resultados computacionais comparando a nossa proposta de algoritmo com o *iBP*; no último capítulo finalizamos com a conclusão deste trabalho. Ao final, um apêndice traz algumas discussões complementares mais detalhadas de alguns pontos descritos na tese.

Informações para o leitor:

- Todos os resultados novos que propusemos neste trabalho foram considerados como **proposições**. Lemas são resultados auxiliares que utilizamos na demonstração das nossas proposições. Qualquer resultado da literatura foi considerado como teorema.
- Como o tema deste trabalho é de natureza interdisciplinar, ele, também, foi pensado com o intuito de poder ser uma referência para pessoas que não sejam da área da matemática.

CAPÍTULO

2 Proteínas: O Que São e Como Determinar Sua Estrutura

O início da história das proteínas acontece no século XVIII com a descoberta de que substâncias do meio vivo, como o sangue, o leite, a clara de ovo (albúmen), entre outros, coagulam em temperaturas elevadas ou em um meio ácido. Substâncias com estas propriedades foram chamadas de albuminoides (semelhante ao albúmen). O próximo ponto nesta cronologia foi a descoberta, no início do século XIX, de que os principais elementos que constituem as células vivas eram substâncias albuminoides. O termo “proteína” foi apresentado ao mundo científico pelo químico orgânico holandês Gerardus Johannes Mulder em 1838 (MULDER, 1838a; MULDER, 1838b). No entanto, este termo foi cunhado pelo químico e investigador científico sueco Jöns Jacob Berzelius, considerado “o pai da química sueca”. Este termo tem origem na palavra grega *πρωτεῖος* (*proteios*) que em português significa algo como “primeiro, primitivo, de primeira importância”. Berzelius sugeriu este termo a Mulder, pois pensava que as substâncias albuminoides eram a base de todos os seres vivos. Toda esta história de origem pode ser consultada em VICKERY (1950).

O interesse científico sobre as proteínas não parou de crescer. Já na virada do século XX, cientistas descobriram que sua degradação liberava aminoácidos, isto é, compostos orgânicos que contêm um grupo amina ($-NH_3^+$) e um grupo carboxila ($-CO_2^-$). Por volta do ano 1900, 12 aminoácidos distintos já tinham sido identificados pela degradação de proteínas. Em 1902, diante destas evidências, o médico alemão e professor de química fisiológica, Franz Hofmeister, sugeriu que a estrutura das proteínas seriam dadas por uma cadeia de aminoácidos. Em 1906 o número de aminoácidos identificados subiu para 15; em 1935 para 18; e em 1940, completa-se a lista dos 20 aminoácidos essenciais AMABIS; MARTHO (1997).

Em geral, as proteínas naturais assumem conformações estruturais próprias, que estão relacionadas com suas funcionalidades biológicas. No início dos anos 1960, o químico estadunidense Christian B. Anfinsen e seus colaboradores descobriram que existe uma relação entre a estrutura 3D da proteína e sua sequência de aminoácidos (ANFINSSEN et

al., 1961). Em 1972, Anfinsen, o químico Stanford Moore e o bioquímico William H. Stein, ambos estadunidenses, compartilharam o prêmio Nobel de química (NOBELPRIZE, 1972) pelas contribuições que fizeram acerca da molécula de ribonuclease. Como é descrito na última referência, estes cientistas foram agraciados com tal prêmio pelo seguinte motivo:

- Anfinsen “*por seu trabalho sobre ribonuclease, especialmente no que diz respeito à conexão entre a sequência de aminoácidos e a conformação biologicamente ativa*”;
- Moore e Stein “*por sua contribuição para o entendimento da conexão entre a estrutura química e a atividade catalítica do centro ativo da molécula de ribonuclease*”.

Ainda na linha da correspondência da geometria de uma proteína com sua função, os genes de um organismo vivo são, indiretamente, os responsáveis pelas características físicas desse organismo, mas as proteínas correspondentes é que determinam, de fato, essas características (DONALD; ISTRAIL; PEVZNER, 2011).

A busca pela compreensão das proteínas para a maquinaria biológica mobiliza cientistas de diversas áreas do conhecimento, uma vez que estas são responsáveis por inúmeras funções no meio fisiológico, como por exemplo, a construção celular, o transporte e a recepção de sinais e a regulação de entrada e de saída de íons (NELSON; COX, 2012). Por outro lado, algumas doenças como o mal de Alzheimer, a doença de Huntington e o diabetes tipo II, estão diretamente relacionadas com o enovelamento defeituoso das proteínas (DOBSON, 2003). Outro aspecto importante, de interesse global da indústria farmacêutica, é que o desenvolvimento de novos medicamentos pode ser baseado nas estruturas tridimensionais das proteínas (ANDRICOPULO; SALUM; ABRAHAM, 2009).

Criado em 1971 sob liderança do cientista Walter Hamilton e sediado no Laboratório Nacional de Brookhaven, Nova Iorque, EUA, o *RCSB Protein Data Bank* (PDB) (BERMAN et al., 2000) é um repositório para estruturas tridimensionais de proteínas. Na sua criação, ele possuía a estrutura de 7 proteínas distintas. Cinquenta anos depois, esse banco de dados possui mais de 172.894 estruturas distintas, que foram caracterizadas por diferentes métodos experimentais, sendo os experimentos de difração de raios-X e RMN os dois mais utilizados para este fim. O PDB é uma importantíssima fonte de dados e está relacionado com o avanço do conhecimento sobre as proteínas, pois é acessado por pesquisadores de todo o globo terrestre gratuitamente. Para a realização deste trabalho, utilizamos ativamente este banco de dados.

2.1 Conceitos Básicos Sobre Proteínas

Após os comentários históricos feitos no início deste capítulo, de maneira simplificada, vamos dizer que proteínas são um encadeamento linear de aminoácidos conectados uns aos outros por ligações peptídicas, isto é, ligações entre o átomo de carbono do grupo carboxílico de um aminoácido com o átomo de nitrogênio do grupo amina de outro aminoácido, em que a repetição de aminoácidos é permitida na construção desta cadeia (CREIGHTON, 2010). Deste modo, as proteínas também podem ser chamadas de polipeptídeos. Como vimos anteriormente, existem 20 aminoácidos padrão distintos que podem compor uma proteína. Quanto a estrutura dos aminoácidos, a Figura 1 apresenta a estrutura molecular de cada um destes aminoácidos.

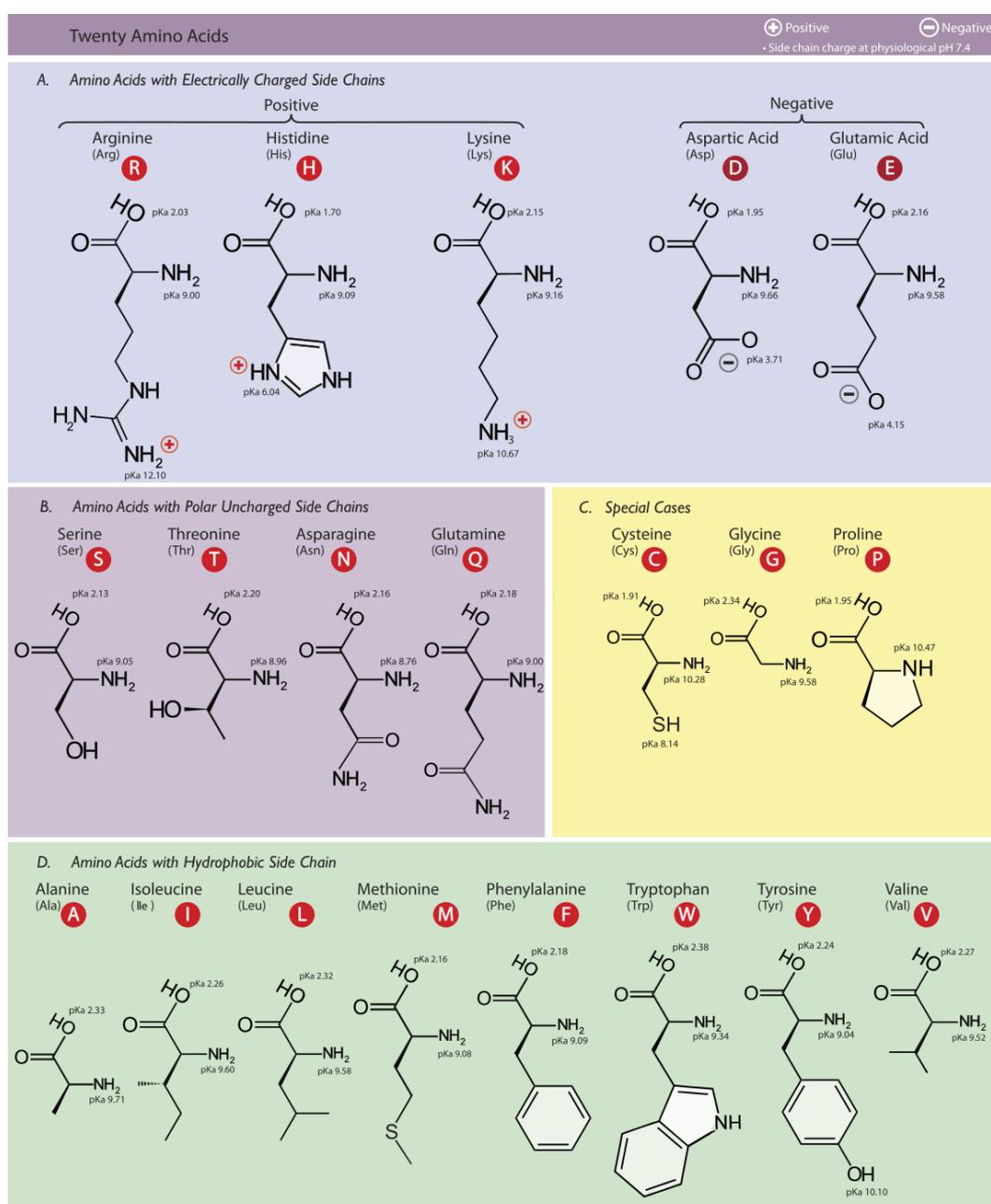


Figura 1 – Lista de aminoácidos. Imagem de COJOCARI (2010), modificada.

De maneira geral, cada aminoácido possui um carbono central em sua cadeia, denominado carbono α (C_α). Este carbono realiza quatro ligações covalentes, uma com um grupo amina ($-NH_2$), outra com um grupo carboxila ($-COOH$), outra com um átomo de hidrogênio (H_α), e outra com uma cadeia lateral, ou grupo R (cadeia lateral), oriunda do aminoácido. A Figura 2 ilustra esta estrutura geral dos aminoácidos. A cadeia lateral diferencia todos os aminoácidos encontrados na natureza quanto à estrutura, tamanho, carga elétrica e hidrofobicidade. Vale ressaltar que o aminoácido prolina é o único que possui um grupo amina distinto dos outros 19 aminoácidos. Para a prolina, o átomo de nitrogênio de seu grupo amina, o átomo C_α e sua cadeia lateral formam uma estrutura rígida em anel. Desta maneira, seu grupo amina possui um átomo de hidrogênio a menos que estes grupos dos outros aminoácidos (CREIGHTON, 2010).

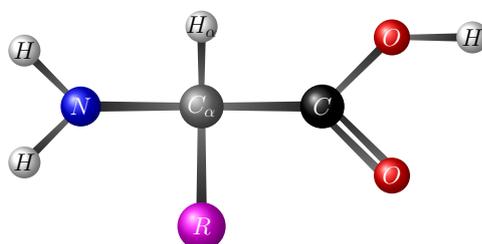


Figura 2 – Estrutura geral de um aminoácido.

Para a formação das proteínas, aminoácidos adjacentes são conectados por ligações peptídicas onde cada ocorrência desta ligação resulta na liberação de uma molécula de H_2O como subproduto. A Figura 3 evidencia este processo para um dipeptídeo, isto é, uma molécula de proteína formada por dois resíduos de aminoácidos.

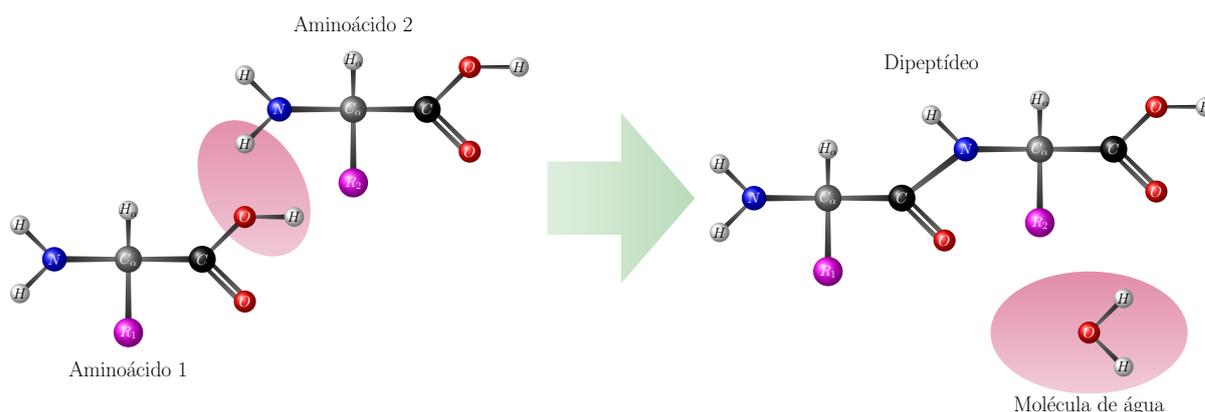


Figura 3 – Processo de junção de dois aminoácidos na formação de um dipeptídeo.

A cadeia principal de uma proteína é descrita pela sequência de triplas de átomos $\{N^i, C_\alpha^i, C^i\}$, onde i denota o i -ésimo resíduo de aminoácido. Apesar da Figura 3 sugerir uma conformação linear desta cadeia, forças interatômicas fazem com que a estrutura da proteína se encurve e se dobre, resultando em uma configuração 3D característica para cada proteína. O problema de se descrever a estrutura de uma proteína dado sua sequência

de aminoácidos é denominado *Protein Folding Problem*, que em português é descrito como: Problema de Enovelamento de Proteínas (NEUMAIER, 1997).

Os pioneiros no estudo da estrutura 3D das proteínas foram os bioquímicos estadunidenses Linus Pauling e Robert Corey. O estudo que eles fizeram sobre as ligações peptídicas os permitiram concluir que os átomos de C_α de resíduos de aminoácidos adjacentes são separados por três ligações covalentes, a saber $C_\alpha^i - C^i - N^{i+1} - C_\alpha^{i+1}$, de modo que a ligação peptídica impõe uma rigidez neste segmento da proteína, observando-se que estes quatro átomos são praticamente planares. Eles também observaram que os outros átomos da cadeia principal das proteínas não possuem nenhuma estrutura rígida, o que permite a definição de ângulos que descrevem torções nesta cadeia no espaço (NELSON; COX, 2012).

Estes ângulos de torção definidos na cadeia principal das proteínas são chamados de ângulos ω , ϕ e ψ , e cada um deles é definido em 4 átomos específicos desta cadeia. A geometria destes ângulos está descrita no Capítulo 4, por enquanto vamos salientar que: ω é o ângulo rígido caracterizado nas ligações peptídicas e os ângulos ϕ e ψ são os flexíveis. No entanto, foi observado experimentalmente que estes ângulos, apesar de flexíveis, não assumem todos os valores possíveis. Isto acontece em razão de restrições impostas pelas nuvens eletrônicas dos átomos de hidrogênio e oxigênio ao redor das ligações peptídicas (SCHLICK, 2013). O mapa que descreve os valores possíveis para os ângulos ϕ e ψ foi descrito primeiramente em RAMACHANDRAN; RAMAKRISHNAN; SASISEKHARAN (1963), e é chamado de *gráfico de Ramachandran* em homenagem ao primeiro cientista a descrever tal mapa, o físico indiano G. N. Ramachandran. A Figura 4 apresenta a imagem deste mapa no trabalho que o originou.

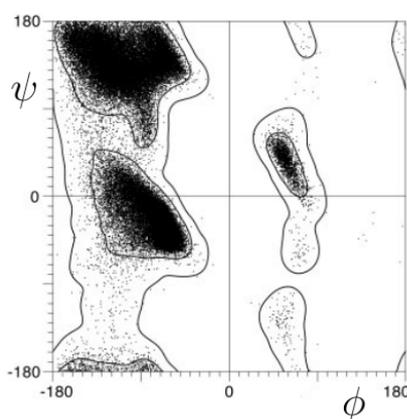


Figura 4 – Gráfico de Ramachandran. Imagem de RAMACHANDRAN; RAMAKRISHNAN; SASISEKHARAN (1963).

Como descrito em NELSON; COX (2012), a estrutura 3D das proteínas é classificada com as seguintes subdivisões estruturais:

- **Estrutura Primária:** o principal elemento desta subdivisão é a descrição da

sequência de aminoácidos que compõem uma proteína. Portanto, esta subdivisão retrata o enovelamento das proteínas no espaço 3D.

- **Estrutura Secundária:** esta estrutura refere-se a arranjos estáveis de uma sequência de resíduos de aminoácidos que dão origem a padrões estruturais característicos. Os arranjos mais relevantes são as α -hélices e as folhas- β . Devido ao caráter particular destes arranjos, podemos observar que:
 - **α -hélices:** são os tipos de conformações mais comuns. Sua estrutura segue o formato de uma curva helicoidal, que é definida pelos átomos de uma sequência de resíduos de aminoácidos. Esta espiral gira em torno de um eixo longitudinal imaginário, onde a cadeia lateral de cada resíduo permanece em sua parte exterior. A [Figura 5](#) ilustra esta estrutura.

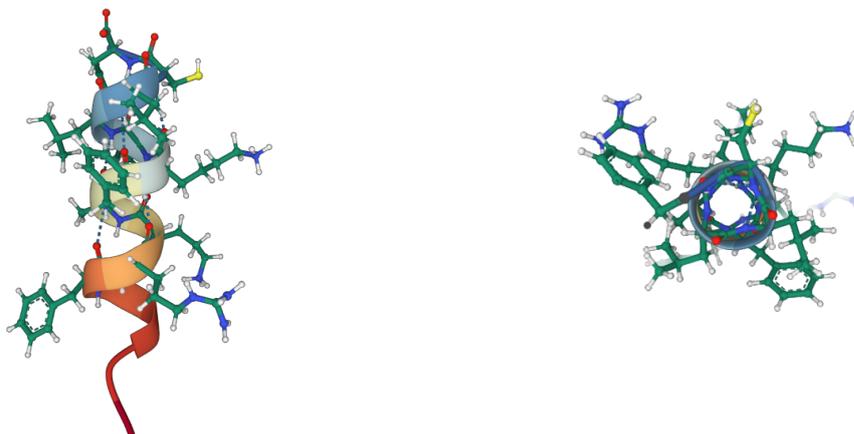


Figura 5 – Dois pontos de vista de uma α -hélice. Modelo 1 da estrutura 1dep do PDB.

- **folhas- β :** este arranjo estrutural é definido por um zigue-zague na cadeia polipeptídica. Neste caso, as cadeias laterais se alternam nos aminoácidos adjacentes, apontando em direções opostas. A [Figura 6](#) ilustra esta estrutura.



Figura 6 – Arranjos estruturais folhas- β . Modelo 1 da estrutura 5o6f do PDB.

- **Estrutura Terciária:** esta subdivisão reflete a organização tridimensional que engloba todos os átomos da proteína. A [Figura 7](#) ilustra esta estrutura;

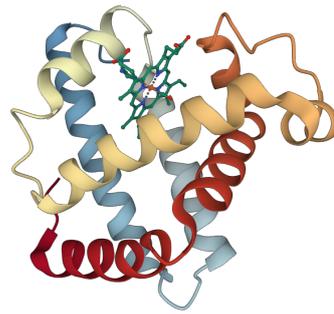


Figura 7 – Molécula de Mioglobina. Modelo 1 da estrutura 2spl do PDB.

- **Estrutura Quaternária:** esta subdivisão estrutural das proteínas refere-se às subestruturas independentes (cadeias) de proteína, isto é, que são cadeias polipeptídicas independentes mas fazem parte da mesma molécula de proteína em um específico estado oligomérico (monômero, dímero, trímero, ...). A [Figura 8](#) ilustra uma molécula de proteína composta por 5 cadeias.

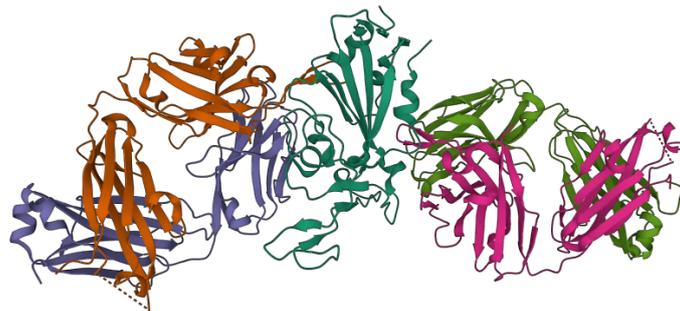


Figura 8 – Molécula relacionada à proteína *spike* do *SARS-CoV-2* (covid-19). Modelo 1 da estrutura 7lm8 do PDB.

Os conceitos sobre proteínas que apresentamos nesta seção não são aprofundados o suficiente para um completo entendimento sobre o assunto. No entanto, enfatizamos que ele é satisfatório para o pleno entendimento das discussões que realizamos nos próximos capítulos desta tese.

2.2 Métodos Experimentais para Caracterização da Estrutura 3D de Proteínas

Os três métodos experimentais mais utilizados para a determinação da estrutura 3D de proteínas em alta resolução são: a Cristalografia de Raio-X (CR-X), a Ressonância Magnética Nuclear (RMN) e a crio-Microscopia Eletrônica (cME) (BERMAN et al., 2000). Dentre os três métodos citados, a CR-X e a cME são, respectivamente, os métodos mais e

menos utilizados. Ressaltamos que estes três métodos não competem entre si, pois, são técnicas que se complementam. Em conjunto, elas fornecem uma imagem detalhada da estrutura e da dinâmica de macromoléculas em resolução atômica, o que pode ser utilizado para uma melhor compreensão dos processos biológicos no nível molecular (BRÜNGER, 1997).

Até o início dos anos 1980, a informação estrutural de proteínas em resolução atômica só podia ser determinada pela técnica de difração de raio-X (DRENTH; MESTERS, 2007). Proposta pelo químico e biofísico suíço Kurt Wüthrich, a técnica de RMN passou a ser considerada para determinação de estruturas tridimensionais de macromoléculas biológicas em solução (WÜTHRICH, 1989). O feito de Wüthrich resultou no Nobel de Química, conjuntamente com o químico estadunidense John Fenn e o químico japonês Koichi Tanaka em 2002 (NOBELPRIZE, 2002).

Com o advento da RMN, passou a ser possível caracterizar estruturas proteicas, com elevada precisão, em uma condição mais próxima da situação natural dos organismos vivos: estas medidas são realizadas à temperatura ambiente e as amostras não necessitam passar pelo processo de cristalização (WÜTHRICH, 1986).

Como este trabalho supõe que os dados experimentais são decorrentes do experimento de RMN, não vamos discorrer sobre os outros experimentos citados. No entanto, sugerimos as referências (DRENTH; MESTERS, 2007; FRANK, 2018) para as outras técnicas.

2.2.1 Ressonância Magnética Nuclear

Na RMN, campos magnéticos de grande magnitude ($> 14 T$) e pulsos de radiofrequência ($60 - 1000 MHz$) são aplicados para o estudo do ambiente magnético dos núcleos atômicos. O ambiente local dos núcleos está relacionado com a frequência de absorção da ressonância. Como resultado deste experimento se tem um espectro que apresenta informações sobre os deslocamentos químicos e interações locais entre os átomos que apresentam núcleos ressonantes. Assim, as frequências de absorção dos diferentes átomos podem ser diferenciadas, mas para a obtenção de imagens sem muitas ambiguidades é necessário que exista uma distinção clara entre os diferentes sinais. Desta maneira, o tamanho das moléculas que podem ser analisadas via RMN é limitado (KEELER, 2011).

De uma maneira mais formal, a base deste experimento reside no fato de que é possível alterar o momento angular de *spins* de átomos que possuem *spins* nucleares semi-inteiros ou inteiros não nulos com o fornecimento de energia em forma de frequência. Se dois átomos estão espacialmente próximos, pelo acoplamento *spin-spin* entre os núcleos destes átomos, seus *spins* nucleares interagem (interação dipolo-dipolo) e modifica-se a frequência necessária para mudar um *spin* de estado (KEELER, 2011).

As afirmações descritas no parágrafo anterior envolvem muitos elementos da estrutura atômica dos átomos, que por si só já necessitariam de uma explicação que fogem do escopo deste trabalho. Então, aquele parágrafo foi considerado somente para iluminar alguns pontos teóricos importantes que envolvem este experimento, mas para qualquer aprofundamento no assunto sugerimos uma leitura da referência citada.

O elemento químico que possui momento angular de *spin* nuclear semi-inteiro de maior abundância na terra é o átomo de hidrogênio, seu *spin* nuclear é $1/2$. No entanto, para a caracterização de proteínas pela RMN, pode-se considerar os isótopos carbono-13 e nitrogênio-15, também com *spin* nuclear $1/2$, que representam apenas uma pequena fração dos isótopos estáveis destes elementos, aproximadamente 1,1% para o carbono-13 e menos de 0,05% para o nitrogênio-15 (GÜNTHER, 2013). Assim, a utilização da RMN é muito mais eficaz na medição de moléculas com um grande número de átomos de hidrogênio. Como vimos na seção anterior, este é o caso das proteínas.

A transferência de magnetização através do acoplamento *spin-spin* entre os átomos de hidrogênio via relaxamento cruzado é denominada *Nuclear Overhauser Effect* (NOE). Com esse processo, os picos no espectro resultante do experimento são ligeiramente alterados, o que torna possível a estimativa de distâncias entre pares de átomos de hidrogênio espacialmente próximos, com distância inferior a $5 \sim 6 \text{ \AA}$ ($1 \text{ \AA} = 10^{-10} \text{ m}$). Os valores de distância obtidos contêm erros sistemáticos e aleatórios, devido aos inúmeros caminhos de transferência de magnetização e à dinâmica molecular interna (VÖGELI et al., 2016).

O espectro resultante do NOE é um espectro unidimensional que pode ter uma análise complicada. No entanto, o experimento *Nuclear Overhauser Effect Spectroscopy* (NOESY), que é a consideração de um espectro bidimensional para o NOE, apresenta um resultado de análise mais factível. A Figura 9 ilustra o espectro destes dois experimentos.

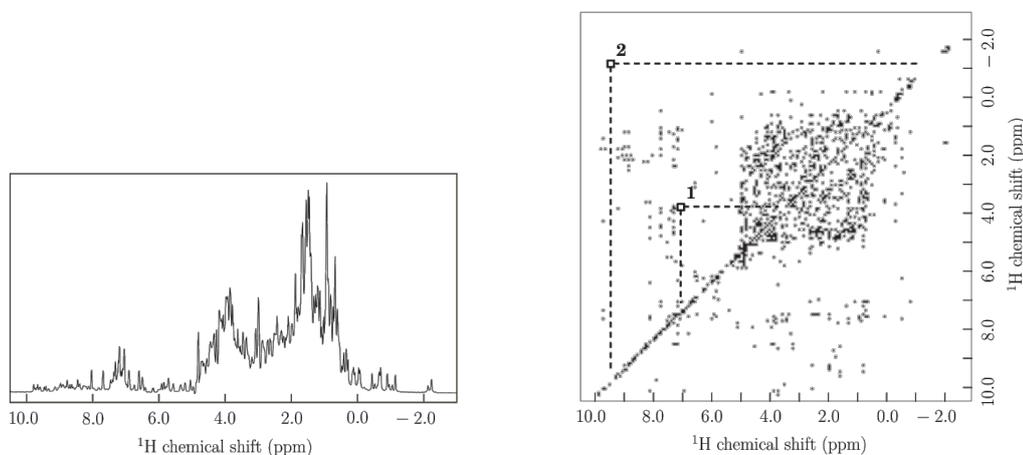


Figura 9 – Exemplos de espectros de RMN unidimensional (esquerda) e bidimensional (direita). Imagem de NELSON; COX (2012), modificada.

As informações resultantes dos experimentos de RMN não são tão detalhadas quanto as provenientes da cristalografia de raio-X. No entanto, a solução que contém a molécula medida pela RMN não é cristalizada, assim ela incorpora os efeitos causados pela variação térmica da solução.

Pela discussão apresentada nesta seção, no contexto desta tese, vamos considerar que o experimento NOESY de RMN é capaz de nos fornecer uma estimativa da distância euclidiana entre alguns átomos de hidrogênio em uma proteína.

CAPÍTULO

3 Problema Discretizável de Geometria de Distâncias

Em 1928, o matemático austríaco-americano Karl Menger caracterizou um grande número de conceitos geométricos em termos de distâncias (como por exemplo, congruência e conjuntos de conectividade) no trabalho (MENGER, 1928). Os resultados encontrados por Menger e o trabalho do matemático americano Leonard M. Blumenthal no livro BLUMENTHAL (1953) originaram o corpo do que conhecemos hoje por Geometria de Distâncias (GD).

Na primeira seção deste capítulo definimos o Problema de Geometria de Distâncias, discutimos um pouco sobre o problema e apresentamos alguns resultados que são importantes no decorrer da tese. Já na segunda seção, definimos o Problema Discretizável de Geometria de Distâncias e trabalhamos com a parte geométrica fundamental utilizada na construção de um algoritmo para encontrar soluções para o problema. Na terceira seção, descrevemos o Problema Discretizável de Geometria de Distâncias Intervalares fundamentando todos os conceitos necessários para o desenvolvimento da aplicação de nosso interesse.

3.1 Problema de Geometria de Distâncias

Definimos o Problema de Geometria de Distâncias (PGD) da seguinte maneira:

Definição 3.1. Sejam $G = (V, E, d)$ um grafo simples, ponderado e não-direcionado, onde V é seu conjunto de vértices, E seu conjunto de arestas e $d : E \rightarrow \mathbb{R}_+$, e K um número inteiro positivo. O PGD consiste em encontrar uma função

$$\begin{aligned} x & : V \rightarrow \mathbb{R}^K \\ v & \mapsto x(v) \end{aligned} ,$$

chamada *realização*, tal que

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

sendo $\|\cdot\|$ uma norma.

Uma realização é dita *válida* se satisfaz o conjunto de equações de (3.1). O conjunto de todas as equações descritas em (3.1) é denominado uma *instância* do PGD. Uma vez que temos uma realização válida, digamos x , uma solução do PGD é dada por:

$$X = \{x(v) \mid v \in V\}.$$

Uma abordagem clássica para o PGD é uma formulação utilizando otimização contínua, e isto se dá de maneira intuitiva considerando uma função $f : \mathbb{R}^{K^n} \rightarrow \mathbb{R}_+$, com $n = |V|$, tal que

$$f(x(v_1), x(v_2), \dots, x(v_n)) = \sum_{\{v_i, v_j\} \in E} \left(\|x(v_i) - x(v_j)\| - d(\{v_i, v_j\}) \right)^2. \quad (3.2)$$

Agora, note que uma condição necessária e suficiente para x ser uma realização válida é ter $f(x(v_1), \dots, x(v_n)) = 0$. Como o menor valor que f pode assumir é zero, basta tomarmos o seguinte problema:

$$\min_{(x(v_1), \dots, x(v_n)) \in \mathbb{R}^{K^n}} f(x(v_1), \dots, x(v_n)). \quad (3.3)$$

Assim, queremos encontrar um mínimo global do problema descrito em (3.3). Não obstante, esta tarefa é bem complexa, uma vez que f não é convexa e possui um número de mínimos locais que aumenta exponencialmente com o número de vértices do grafo associado ao problema, o que pode ser consultado em (LIBERTI et al., 2014). Dificuldade que ainda é intensificada pois diferenciar um mínimo local de um mínimo global é uma tarefa sensível, uma vez que muitos métodos de otimização contínua dispõem apenas de informações locais, isto é, do valor da função e suas derivadas no ponto de análise. O trabalho LAVOR; LIBERTI; MACULAN (2006) evidencia que resolver este problema é difícil do ponto de vista da complexidade computacional, assim como do ponto de vista prático.

Apesar de termos argumentado um pouco sobre a dificuldade de encontrar uma solução do PGD utilizando uma abordagem de otimização contínua, encontrar uma solução para este problema é difícil independentemente da abordagem utilizada. E para quantificar o quão difícil este problema é, o trabalho SAXE (1980) apresenta um resultado que nos diz que o PGD é NP-completo para $K = 1$ e NP-difícil para $K \geq 2$. Assim, de maneira simplificada, encontrar uma solução para o PGD pode exigir um custo computacional elevado, proporcional a $2^{|V|}$.

Uma outra possibilidade que poderia ser levantada, seria tentar resolver o sistema de equações não lineares definido em (3.1) de maneira analítica. No entanto, [BAJAJ \(1988\)](#) nos traz uma conjectura de que não é possível obter uma fórmula fechada para resolver este sistema analiticamente se a norma em questão é a euclidiana.

O campo de Geometria de Distâncias é vasto, existem inúmeras aplicações e inúmeros métodos para encontrar soluções para o PGD. Aqui, ficamos restritos a discutir uma modelagem de estruturas moleculares, entretanto, recomendamos a referência [LIBERTI et al. \(2014\)](#) caso o leitor queira estudar outras aplicações ou aprender um pouco mais sobre diferentes metodologias utilizadas em Geometria de Distâncias.

3.2 Problema Discretizável de Geometria de Distâncias em \mathbb{R}^3

A ideia de explorar uma ordem para vértices em um grafo para investigar sua rigidez apareceu primeiramente em [HENNEBERG \(1886\)](#). De fato, ordens de vértices são importantes para a solução de muitos problemas modelados por grafos, o que pode ser consultado em [BODLAENDER et al. \(2012\)](#).

Uma ordem trilateral em um grafo consiste na ordenação de seus vértices de modo que os quatro primeiros vértices definam uma clique, isto é, existe uma aresta entre quaisquer dois destes vértices, e do quinto vértice até o último, cada vértice é adjacente a pelo menos quatro vértices anteriores. Como descrito em [EREN et al. \(2004\)](#), a construção desta ordem em um grafo do PGD em \mathbb{R}^3 implica que existe um algoritmo, com tempo linear, para encontrar a única solução deste problema. No entanto, não conseguimos construir uma ordem como esta no grafo do que modela a nossa aplicação de interesse.

Na linha de estabelecer uma ordem nos vértices do grafo que nos garanta, pelo menos, a finitude do número de soluções do PGD, o *Discretizable Vertex Ordering Problem* trabalha em uma ordenação dos vértices do grafo de modo que cada vértice desta enumeração é definido levando-se em conta o maior número possível de arestas aos antecessores. [LAVOR et al. \(2012\)](#) discute detalhadamente este problema e mostra que ele pode ser resolvido em tempo polinomial. Já a referência [LAVOR et al. \(2017\)](#) nos mostra que a existência dessa ordem é apenas uma condição suficiente para a finitude do PGD.

Para prosseguirmos com esta ideia de definir uma ordenação nos vértices do grafo do PGD, necessitamos dos seguintes conceitos:

Definição 3.2. Sejam $G = (V, E, d)$ um grafo simples, ponderado e não-direcionado e $<$ uma ordem sobre os elementos de V . Então, dado um vértice $v \in V$, temos que:

1. O conjunto dos vértices predecessores do vértice v é dado por:

$$\gamma(v) = \{u \in V \mid u < v\};$$

2. O conjunto dos vértices adjacentes ao vértice v é dado por:

$$N(v) = \{u \in V \mid \{u, v\} \in E\};$$

3. O conjunto dos predecessores adjacentes de v é dado por:

$$U_v = N(v) \cap \gamma(v).$$

Obs.: diremos que $k_v = |U_v|$.

Se considerarmos que os predecessores adjacentes de todos os vértices de V respeitam certa condição, que é apresentada na próxima definição, teremos que o número k_v é um valor crítico para o PGD. Como na aplicação de interesse temos que $K = 3$, esta criticidade acontece em torno do valor de $k_v = 3$. Dizemos que este valor é crítico uma vez que $k_v < 3$, $k_v = 3$ e $k_v > 3$, para qualquer vértice $v \in V$, implicam em três situações completamente distintas no PGD. No primeiro caso, $k_v < 3$, o número de soluções do PGD provavelmente é infinita e não enumerável. Já no segundo caso, $k_v = 3$, existem $2^{|V|-3}$ soluções para o PGD. Por fim, no terceiro caso, $k_v > 3$, o PGD pode ser resolvido unicamente, com exceção de uma reflexão simples, em tempo linear. Estas afirmações podem ser consultadas em [LAVOR et al. \(2012\)](#).

Na nossa aplicação de interesse, como já dissemos, não garantimos uma ordem trilateral para ordenação dos vértices do grafo do PGD, no entanto temos garantias que para todos os vértices deste grafo, $k_v \geq 3$. Portanto, para prosseguir nesta direção consideremos a seguinte definição:

Definição 3.3. Sejam $G = (V, E, d)$ um grafo simples, ponderado e não-direcionado associado a uma instância do PGD, com $K = 3$, $\|\cdot\|$ a norma euclidiana e uma ordem v_1, \dots, v_n nos vértices de V . O Problema Discretizável de Geometria de Distâncias (PDGD) consiste em determinar todas as instâncias do PGD que satisfazem as seguintes suposições:

$S1$: Existe uma realização válida para os vértices v_1, v_2 e v_3 ;

$S2$: Para todo vértice v_i , com $i = 4, \dots, n$, existe $\{v_{i3}, v_{i2}, v_{i1}\} \subset U_{v_i}$, onde:

$$d(\{v_{i3}, v_{i1}\}) < d(\{v_{i3}, v_{i2}\}) + d(\{v_{i2}, v_{i1}\}).$$

Obs.: Neste trabalho, por conta do nosso interesse da aplicação de GD em geometria molecular, definimos o PDGD fixando $K = 3$. No entanto, pode-se definir o PDGD com K sendo um número natural qualquer maior que zero. Esta definição geral pode ser encontrada na referência [LIBERTI et al. \(2014\)](#).

Primeiramente vamos observar que o PDGD é NP-difícil, como mostra a referência [MUCHERINO; LAVOR; LIBERTI \(2012\)](#). Agora, vamos observar quais são as implicações das suposições da Definição 3.3. Para tanto, notemos que a suposição $S1$ nos garante que existe uma realização válida para os vértices v_1 , v_2 e v_3 , isto significa que os pontos $x(v_1)$, $x(v_2)$ e $x(v_3)$ possuem posições fixas em \mathbb{R}^3 . Com isso, temos a informação de distância entre quaisquer dois desses pontos, assim, os vértices v_1 , v_2 , v_3 são uma clique. Já a suposição $S2$ nos garante que os pontos $x(v_1)$, $x(v_2)$ e $x(v_3)$ não são colineares (justificamos essa afirmação no próximo parágrafo). Estes fatos nos garantem que o conjunto solução X do PDGD contém somente soluções incongruentes, isto é, o conjunto solução do PDGD não contém soluções obtidas por rotações e translações de outras soluções, com exceção de uma reflexão simples. O que pode ser observado na referência [MUCHERINO; LAVOR; LIBERTI \(2012\)](#).

Notemos que a suposição $S2$ nos garante que para qualquer vértice v_i , com $i = 4, \dots, n$, tem-se $k_{v_i} \geq 3$, uma vez que o número de vértices antecessores a v_i tem de ser pelo menos 3. Isto significa que pode existir um conjunto de índices $\mathcal{J} \subset \{5, 6, \dots, n\}$, tal que se $j \in \mathcal{J}$ então $k_{v_j} > 3$. Vamos considerar que $\mathcal{J} \subsetneq \{5, 6, \dots, n\}$, uma vez que a igualdade entre os conjuntos resultaria que $k_{v_i} > 3$, para todo $i = 5, \dots, n$, e neste caso o PDGD seria resolvido em tempo linear, como descrito anteriormente. Notemos, também, que a suposição $S2$ garante-nos que após os vértices do conjunto de predecessores adjacentes de v_i serem imersos no \mathbb{R}^3 , pelo menos três deles não pertencem a uma mesma reta. Isto é facilmente constatado, uma vez que a igualdade na desigualdade triangular acontece se, e somente se, os 3 pontos do \mathbb{R}^3 descritos na desigualdade são colineares. Como a suposição $S2$ descreve uma desigualdade triangular estrita para as distâncias de pelo menos três vértices de U_{v_i} , a condição de colinearidade para estes pontos se torna impossível. Esta suposição também garante que o conjunto solução X do PDGD é finito, como descrito em [MUCHERINO; LAVOR; LIBERTI \(2012\)](#).

Das três observações apresentadas acima, concluímos que uma modelagem descrita pelo PDGD tem um número finito de soluções distintas.

Na definição do PDGD, se assumirmos na suposição $S2$ que $v_{i_3} = v_{i-1}$, $v_{i_2} = v_{i-2}$ e $v_{i_1} = v_{i-3}$, teremos o Problema Discretizável de Geometria de Distâncias Moleculares (PDGDM), que é uma classe mais restrita de problemas do PDGD.

O PDGDM tem uma rica estrutura combinatória, com propriedades não presentes na versão mais geral do problema ([LIBERTI et al., 2014](#)). Esta abordagem combinatória, fundamentada em toda a teoria de Geometria de Distâncias, possibilitou a descoberta de simetrias que caracterizam todas as soluções matemáticas do problema ([MUCHERINO; LAVOR; LIBERTI, 2012](#)). A principal consequência desse resultado é que torna-se suficiente encontrar uma única solução, usando qualquer método, pois todas as soluções restantes podem ser facilmente encontradas em tempo linear ([MUCHERINO; LAVOR; LIBERTI,](#)

2012). No trabalho LAVOR et al. (2021a) apresentamos um algoritmo com complexidade ótima para encontrar o conjunto que caracteriza as simetrias do PDGDM. Apesar desta grande vantagem do PDGDM sobre o PDGD, para desenvolver os resultados deste trabalho, não conseguimos definir uma ordem PDGDM nos vértices do grafo associado ao PGD. Portanto, as discussões abordadas neste trabalho não vão abordar nenhuma particularidade exclusiva do PDGDM.

Deste ponto em diante no texto a norma $\|\cdot\|$ sempre é a euclidiana, $n = |V|$ e, para simplificar a notação, utilizamos somente o índice do vértice nas quantidades que dependem deste vértice, ou seja, $x(v_i) = x_{v_i} = x_i$ e $d(\{v_i, v_j\}) = d_{v_i, v_j} = d_{i, j}$.

3.2.1 Características de uma Solução do PDGD

Nesta seção, vamos observar quais seriam as condições e como seria um possível método para encontrar uma solução do PDGD, para isso, vamos seguir as suposições dadas na definição do PDGD e constataremos quais seriam as consequências.

Da suposição $S2$, temos que os vértices $\{v_1, v_2, v_3\}$ são predecessores adjacentes de v_4 , logo $k_4 = 3$. Então $S2$ nos garante que $\{\{v_1, v_4\}, \{v_2, v_4\}, \{v_3, v_4\}\} \in E$. Isto significa que a distância do vértice v_4 para os seus predecessores adjacentes é conhecida. Assim, temos o sistema na variável x_4 :

$$\begin{cases} \|x_4 - x_1\| = d_{4,1}, \\ \|x_4 - x_2\| = d_{4,2}, \\ \|x_4 - x_3\| = d_{4,3}. \end{cases} \quad (3.4)$$

Como x_4 é um ponto do \mathbb{R}^3 e é desconhecido, podemos observar cada equação do Sistema (3.4), após elevar ambos os lados da igualdade ao quadrado, como sendo uma esfera no espaço, centrada no ponto x_j de raio $d_{4,j}$, com $j = 1, 2, 3$. Portanto, encontrar o valor de x_4 , isto é, resolver tal sistema de equações, seria o equivalente algébrico do problema geométrico de encontrar a intersecção das três esferas mencionadas anteriormente. Como o apelo geométrico, na maioria das vezes, nos é mais intuitivo vamos seguir por este caminho. Para tanto, vamos analisar todas as possibilidades da intersecção de três esferas no espaço. Fazemos esta análise com o auxílio da Figura 10.

Assim, com este apelo geométrico, sabemos que existem quatro casos para a intersecção de três esferas no espaço:

- Caso 1: Intersecção vazia. Isto ocorre quando o centro das três esferas não estão suficientemente próximos uns dos outros;
- Caso 2: Intersecção igual a um ponto. Isto ocorre quando o ponto de intersecção é coplanar ao plano definido pelo centro das 3 esferas;

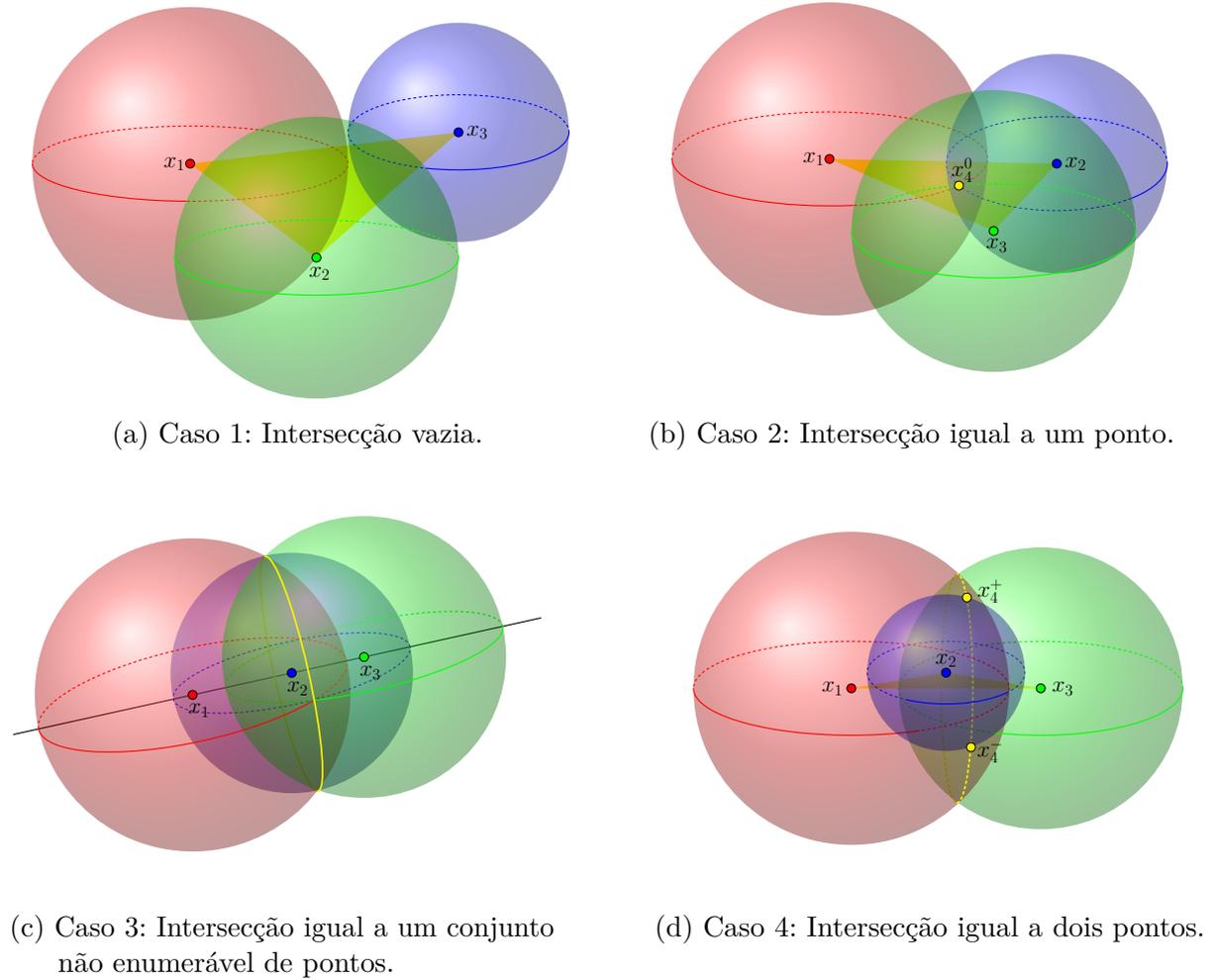


Figura 10 – Casos possíveis para a intersecção de três esferas no espaço.

Caso 3: Intersecção igual um conjunto não enumerável de pontos. Isto ocorre se o centro das três esferas são colineares;

Caso 4: Intersecção igual a dois pontos. Este é o caso geral da intersecção de 3 esferas, e acontece quando os centros das 3 esferas estão suficiente próximos, não são colineares e o ponto de intersecção não pertence ao plano formado pelo centro das 3 esferas. É interessante ressaltar que os dois pontos deste caso são simétricos com relação ao plano definido pelo centro das três esferas.

Portanto, os casos descritos anteriormente, representam todas as possibilidades de solução do Sistema (3.4). No entanto, sabemos que S^2 nos garante que x_1, x_2 e x_3 definem um plano em \mathbb{R}^3 , uma vez que não são colineares, conforme discutido anteriormente. Logo, como o Sistema (3.4) faz parte do PDGD, o Caso 3 é impossível. Como já citamos anteriormente, o PDGD possui solução, com isso, também, podemos desconsiderar o Caso 1. Assim, a solução do Sistema (3.4), pode ser um ponto coplanar aos pontos x_1, x_2 e x_3 , digamos x_4^0 , ou dois pontos não coplanares com estes pontos e simétricos com relação a este plano, digamos x_4^+ ou x_4^- . Na Subseção 3.2.3, mostramos algebricamente que x_4^+ e x_4^-

são simétricos com relação ao plano definido por x_1 , x_2 e x_3 . Assim, diremos que o ponto x_4^0 acontece em um caso degenerado do Caso 4, isto é, quando $x_4^+ = x_4^-$. Desta maneira, vamos considerar que o Caso 2 está contido dentro do Caso 4. Assim, para fins de cálculo vamos admitir que o Caso 4 é a opção de solução do Sistema (3.4). Logo, consideramos que há duas posições possíveis para a realização do vértice v_4 . Desta maneira, fixando uma destas posições, a princípio de maneira arbitrária, estamos aptos a seguir para a imersão do próximo vértice da ordem de V , o vértice v_5 .

Um processo similar ao descrito anteriormente pode ser aplicado para a imersão de todo vértice v_i , com $i = 4, \dots, n$, onde $k_i = 3$, a única diferença é quais seriam os três vértices predecessores adjacentes do vértice v_i , o que implicaria somente em renomear as variáveis do sistema de Equações (3.4). E com isso, teríamos duas opções para a realização de v_i e, arbitrariamente, poderíamos fixar uma delas, de maneira completamente análoga a já descrita.

Agora, para a imersão de um vértice v_i , com $i = 5, \dots, n$, onde $k_i > 3$. Observamos, primeiramente, que já é conhecida a posição no \mathbb{R}^3 dos vértices $v_1, v_2, v_3, \dots, v_{i-1}$ e que, de maneira arbitrária, escolhemos a posição de todos os vértices $u < v_i$ em $k_u = 3$. Seja $U_i = \{v_{i_1}, v_{i_2}, v_{i_3}, \dots, v_{i_{k_i}}\}$ o conjunto dos vértices predecessores adjacentes de v_i , logo, o equivalente ao Sistema (3.4), para este caso, é:

$$\begin{cases} \|x_i - x_{i_{k_i}}\| = d_{i,i_{k_i}}, \\ \vdots \\ \|x_i - x_{i_3}\| = d_{i,i_3}, \\ \|x_i - x_{i_2}\| = d_{i,i_2}, \\ \|x_i - x_{i_1}\| = d_{i,i_1}. \end{cases} \quad (3.5)$$

Novamente, observamos que cada equação do sistema acima, após elevar ambos os lados da igualdade ao quadrado, representa uma esfera no espaço, centrada no ponto x_i de raio d_{i,i_ℓ} , com $\ell = 1, \dots, k_i$. Portanto, encontrar uma solução do Sistema (3.5), isto é, obter o valor de x_i , seria o equivalente algébrico do problema geométrico de encontrar a intersecção das k_i esferas mencionadas anteriormente. Logo, o aspecto geométrico do problema pode nos ajudar a encontrar o valor de x_i . Considerando que $S_1^i = S(x_{i_1}, d_{i,i_1})$, $S_2^i = S(x_{i_2}, d_{i,i_2})$, \dots , $S_{k_i}^i = S(x_{i_{k_i}}, d_{i,i_{k_i}})$ sejam as k_i esferas que queremos interseccionar, então a solução do Sistema (3.5) é dada por:

$$X_i = S_1^i \cap S_2^i \cap S_3^i \cap \dots \cap S_{k_i}^i. \quad (3.6)$$

Agora, a fim de utilizar o conhecimento adquirido na discussão sobre a intersecção de três esferas, vamos reescrever a igualdade de (3.6) da seguinte maneira:

$$\begin{aligned}
X_i &= S_1^i \cap S_2^i \cap S_3^i \cap \cdots \cap S_{k_i}^i \\
&= (S_1^i \cap S_2^i \cap S_3^i) \cap (S_1^i \cap S_2^i \cap S_4^i) \cap \cdots \cap (S_1^i \cap S_2^i \cap S_{k_i}^i) \\
&= \bigcap_{\ell=3}^{k_i} (S_1^i \cap S_2^i \cap S_\ell^i).
\end{aligned} \tag{3.7}$$

Como discutimos anteriormente, $S_1^i \cap S_2^i \cap S_\ell^i = \{(1,2,\ell)x_i^+\} \cup \{(1,2,\ell)x_i^-\}$, onde $(1,2,\ell)x_i^\pm$ são as duas possibilidades para a intersecção das esferas S_1^i , S_2^i e S_ℓ^i que são simétricas com relação ao plano definido pelos pontos x_{i_1} , x_{i_2} e x_{i_ℓ} . Portanto, podemos reescrever a Equação (3.7) da seguinte maneira:

$$\begin{aligned}
X_i &= \bigcap_{\ell=3}^{k_i} (S_1^i \cap S_2^i \cap S_\ell^i) \\
&= \bigcap_{\ell=3}^{k_i} \left(\{(1,2,\ell)x_i^+\} \cup \{(1,2,\ell)x_i^-\} \right).
\end{aligned} \tag{3.8}$$

Dessa maneira, nos é intuitivo que as possibilidades para X_i são:

1. $X_i = \emptyset$;
2. $|X_i| = 1$, com $X_i = \{x_i^-\}$ ou $X_i = \{x_i^+\}$. Este caso acontece se o centro das k_i esferas não forem coplanares;
3. $X_i = \{x_i^-, x_i^+\}$ onde podemos ter $x_i^- = x_i^+$. Este caso acontece se o centro das k_i esferas forem coplanares.

Os detalhes sobre as afirmações acima podem ser consultadas em (MAIOLI; LAVOR; GONÇALVES, 2017).

Na terceira afirmação, como a condição necessária para este caso acontecer é a coplanaridade de todas as k_i esferas, a menos que seja explicitado o contrário, consideramos este caso como improvável. Portanto, para um vértice v_i com $k_i > 3$, podemos ter uma ou nenhuma possibilidade para sua posição espacial. Visto que já argumentamos que o PDGD possui solução, vamos ressaltar que quando vamos imergir o vértice v_i no \mathbb{R}^3 , já posicionamos $i - 1$ vértices no espaço e que para alguns deles, os que possuem exatamente três predecessores adjacentes, uma escolha arbitrária para sua posição foi feita. Assim, para entender como o Sistema (3.5) pode não ter solução no PDGD, primeiramente, vamos definir os seguintes conjuntos:

Definição 3.4. Sejam $G = (V, E, d)$ um grafo simples, ponderado e não-direcionado e $<$ uma ordem sobre os elementos de V . Então:

$$\begin{aligned} \mathcal{B} &\equiv \{v \in V \mid k_v = 3\}, \quad \mathcal{B}_i \equiv \{u \in \mathcal{B} \mid u < v_i\}, \quad \text{com } p = |\mathcal{B}| \text{ e } p_i = |\mathcal{B}_i|, \\ \mathcal{P} &\equiv \{v \in V \mid k_v > 3\}, \quad \mathcal{P}_i \equiv \{u \in \mathcal{P} \mid u < v_i\}, \quad \text{com } q = |\mathcal{P}| \text{ e } q_i = |\mathcal{P}_i|. \end{aligned} \quad (3.9)$$

Logo, notemos que antes de imergir o vértice v_i , em p_i vértices fizemos uma escolha arbitrária, entre duas opções possíveis, de qual seria o ponto considerado para a posição espacial destes vértices, e em q_i vértices determinamos a sua imersão sem nenhuma escolha. Com isso, notamos que existe um carácter combinatorial nas escolhas feitas para a imersão de cada um dos p_i vértices de \mathcal{B}_i . Este carácter combinatorial nos garante que a solução do PDGD existe mesmo que não seja factível a solução do Sistema (3.5) para o vértice v_i , uma vez que, esta factibilidade está condicionada as escolhas feitas para a posição espacial dos vértices de \mathcal{B}_i . Então, uma escolha adequada para a realização dos vértices de \mathcal{B}_i garantirá que sempre existe uma imersão factível para qualquer vértice do PDGD, e, conseqüentemente, uma escolha errada, para a realização destes vértices, resultará em uma impossibilidade de realizar um vértice v_i . Desta forma, caso a impossibilidade de realização de um vértice $v_i \in \mathcal{P}$ aconteça, temos de retroceder na ordem de vértices e escolher a outra opção de realização para algum vértice de \mathcal{B}_i . Isto implica em realizar, novamente, todos os vértices posteriores ao vértice cuja posição espacial foi trocada.

Portanto, com o que acabamos de discutir, temos uma maneira sistemática de realizar qualquer vértice da ordem definida no grafo do PDGD. Isto nos mostrou um modo, ainda que não sistematizado, de encontrarmos uma solução deste problema. Na próxima seção, vamos sistematizar um algoritmo capaz de encontrar uma solução para o PDGD.

3.2.2 O Algoritmo *Branch & Prune*

Nesta seção, vamos sistematizar o que foi discutido na seção anterior na forma do algoritmo *Branch & Prune* (optamos por usar o nome deste algoritmo em inglês visto que na vasta literatura que existe sobre GD, este é o nome utilizado). *Branch & Prune* significa algo como “ramifica e poda” em português, o motivo deste nome fica claro no decorrer desta seção.

A sistematização dos conceitos apresentados na seção anterior, se dá de maneira natural associando uma árvore binária, pois no máximo temos duas opções de posição no espaço para cada vértice de V , na organização do espaço combinatorial intrínseco onde buscamos as soluções do PDGD. Para tanto, observamos que o conceito de árvore binária é amplamente utilizado em teoria de grafos, definições voltadas para este contexto podem ser encontradas em [CORMEN et al. \(2009\)](#). No entanto, neste trabalho não utilizamos o conceito completo de árvore binária, mas sim a ideia deste conceito.

Dizemos que uma árvore binária é um conjunto finito de nós (pontos), que possui uma maneira específica de conectar estes nós. Toda árvore possui um nó gerador, que é denominado raiz. Quanto as conexões entre os nós, a raiz pode ter no máximo duas conexões, no entanto, os outros nós podem ter até de três conexões. A profundidade de um nó é a quantidade de nós que existem no caminho que conecta este nó e a raiz. O conjunto de nós com a mesma profundidade é chamado de nível da árvore. Os nós de uma árvore binária podem ter grau zero, um ou dois, dependendo do número de conexões que ele possui com nós do próximo nível. Nós de grau zero são denominados folhas e conexões de nós de um nível com nós do próximo nível de ramo.

A Figura 11 apresenta um exemplo de árvore binária com 8 nós, $\{n_1, n_2, \dots, n_8\}$, onde o nó 1 é a raiz. Os nós 2 e 3 estão no nível 1, os nós 4, 5 e 6 no nível 2, e os nós 7 e 8 no nível 3. Os nós 4, 6, 7 e 8 são folhas, já o grau do nó 3 é um, e dos nós 1, 2 e 5 é dois.

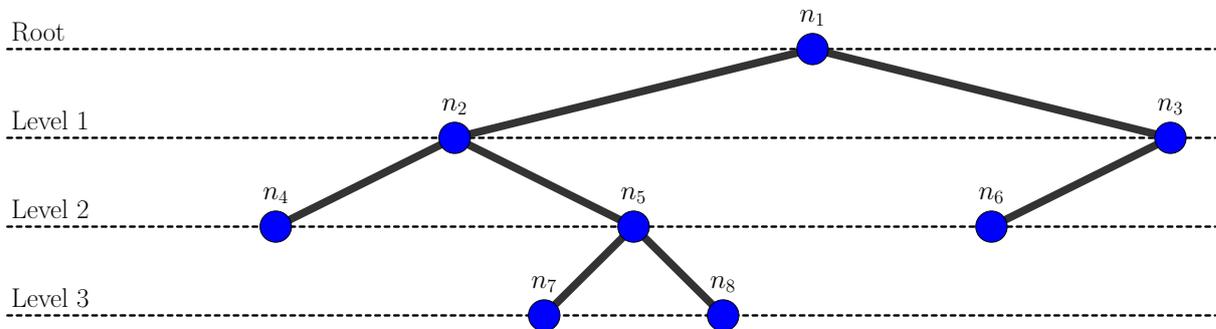


Figura 11 – Exemplo de árvore binária com 8 nós.

Definimos como espaço de busca, o espaço das possibilidades de realizações de todos os vértices do grafo do PDGD, e vamos utilizar uma árvore binária para modelar este espaço de busca. Para tanto, de maneira natural vamos identificar os nós da árvore como realizações dos vértices, e posicionamos a realização do terceiro vértice como raiz da árvore. Seguindo com esta ideia, posicionamos cada uma das duas possíveis realizações de um vértice no mesmo nível, e diferenciamos os dois ramos que partem de um mesmo vértice realizado com os sinais de “-”, para o ramo mais a esquerda, e “+”, para o ramo mais a direita. Como o espaço de busca representa todas as possíveis soluções para o PDGD, a árvore binária que utilizamos para modelar este espaço tem, com exceção dos vértices do nível $n - 1$, dois ramos para todos os nós. Portanto, a partir da raiz da árvore, temos representada todas as possíveis posições para as realizações dos vértices v_i , para $i = 4, \dots, n$. Isto significa, duas possibilidades para x_4 , quatro para x_5 , oito para $x_6, \dots, 2^{i-3}$ para $x_i, \dots, e 2^{n-3}$ para x_n .

A Figura 12 apresenta o espaço de busca, modelado por uma árvore binária, para um PDGD onde o número de vértices de V é sete. Cada ramo da árvore está associado a um sinal de “-” ou “+” conforme descrito anteriormente, no entanto, somente evidenciamos os sinais dos dois primeiros ramos.

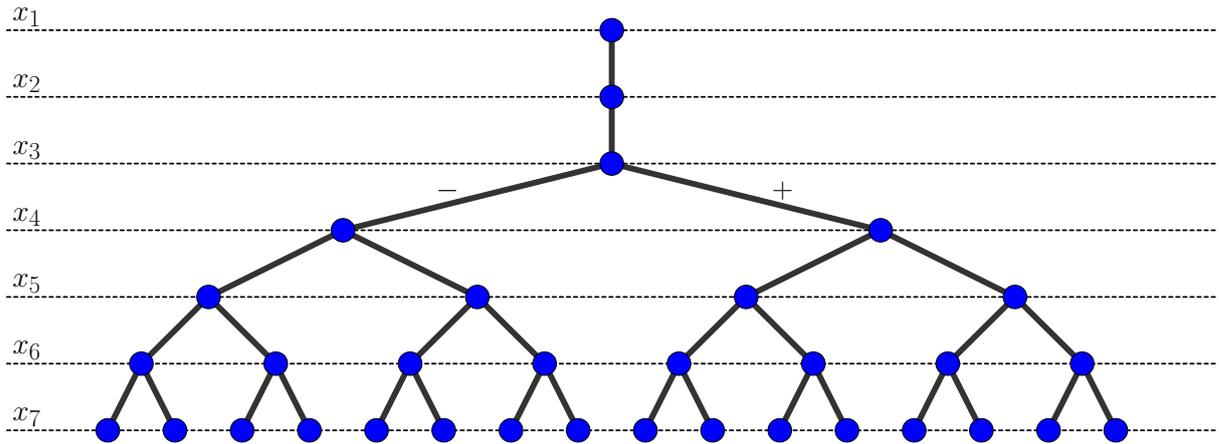


Figura 12 – Espaço de busca de um PDGD com sete vértices.

Portanto, de maneira intuitiva, encontrar uma solução do PDGD significa descer até o nó do último nível da árvore. Note que, se chegamos a uma folha da árvore, então conseguimos evidenciar uma sequência de sinais “+” ou “-”, um para cada nível da árvore, que nos informou qual o caminho que tomamos para chegar até este nó. Observe também que a recíproca deste caso é verdadeira, isto é, se temos uma sequência de sinais “+” ou “-”, um para cada nível da árvore, que nos informa qual o caminho que devemos tomar, então chegamos até o último nó da árvore.

Assim, pela maneira que o espaço de busca foi modelado, fica claro que se exploramos toda a árvore podemos encontrar, ao menos um, caminho para chegarmos até o último nó da árvore. No entanto, observamos que o número de possíveis caminhos que existe é de 2^{n-3} , o que não nos parece muito razoável computacionalmente. Todavia, olhando para o PDGD, a realização de um vértice $v_i \in \mathcal{P}$ possui uma ou nenhuma opção de realização, como já foi discutido anteriormente. Isto significa que quando formos realizar estes vértices temos a oportunidade de descartar um, ou os dois dos ramos que seriam criados a partir do nó x_{i-1} . Denominamos a passagem sobre ramos na árvore como ramificação e o descarte de ramos como poda. Da definição dos conjuntos \mathcal{B} e \mathcal{P} , em (3.9), observa-se facilmente que $\mathcal{B} \cup \mathcal{P} = \{v_4, v_5, \dots, v_n\}$ e que $\mathcal{B} \cap \mathcal{P} = \emptyset$. Assim, diremos que todos os vértices, com exceção dos três primeiros, do grafo do PDGD estão relacionados a um processo de ramificação ou a um de poda.

Notemos que a poda em um ramo, digamos o ramo “-” do nó x_{i-1} , implica que todos os nós da árvore que estão localizados nos níveis mais altos que o nível deste nó podem ser descartados, pois a realização de quaisquer vértices representados por estes nós não é possível dado que x_i^- não é uma realização factível. Com isso, temos que o processo de poda reduz o espaço de busca.

Antes de seguir com a sistematização algorítmica de encontrar uma solução para o PDGD, vamos observar como se dá o funcionamento de algoritmos da classe *Backtracking*, que geralmente são utilizadas para encontrar soluções de problemas computacionais que

envolvem a satisfação de restrições. A lógica destes algoritmos é construir um candidato a solução e abandoná-lo (*backtrack*) se for determinado que este candidato não poderá ser estabelecido como uma solução válida. Caso um *backtrack* ocorra, o algoritmo inicia do último candidato à solução que ainda poderia ser factível e realiza uma escolha diferente da que levou à infactibilidade da solução no passo anterior. Como descrito na referência [HALL; COLLIS \(1995\)](#) o termo *Backtracking* foi cunhado pelo matemático estadunidense Derrick H. Lehmer na década de 1950. No entanto, uma das primeiras descrições deste algoritmo apareceu no século XIX em [LUCAS \(1893\)](#). Atualmente, este algoritmo é utilizado em uma ampla gama de problemas combinatoriais como descrito em [ROSSI; BEEK; WALSH \(2006\)](#).

Isto posto, tendo em mente toda a discussão que apresentamos até este ponto nesta seção, fica muito intuitivo pensar em um algoritmo com uma lógica de *Backtracking* para encontrar as soluções do PDGD. De fato, utilizando este conceito os autores de ([LIBERTI; LAVOR; MACULAN, 2008](#)) propuseram o algoritmo *Branch & Prune*, que a partir deste ponto do texto denominaremos como BP, que utiliza esta lógica algorítmica na construção de uma solução para o PDGD. Este algoritmo utiliza todos os vértices de poda como restrições, e com isso ele tem indicadores que envolvem a satisfação ou não de uma restrição. O nome *Branch & Prune*, como já dissemos, significa algo como “Ramifica e Poda”, assim, a escolha do nome fica evidente se nos atermos ao fato de que os vértices do grafo do PDGD ou estão associados a um processo de ramificação ou a um processo de poda. Assim, permitindo ramificações ou realizando podas, o BP constrói uma sequência de sinais “+” ou “-”, um sinal para cada ramo da árvore, e vai descendo na árvore que descreve o espaço de busca. O algoritmo termina quando ele consegue realizar o último vértice de V , e com isso encontrar uma solução para o PDGD.

Para a sistematização do BP, ainda temos de tratar alguns pontos que podem gerar algum tipo de ambiguidade nos passos deste algoritmo. Para isto, observemos o seguinte:

1. Como a realização dos três primeiros vértices é dada pela definição do PDGD, a clique inicial pode ser fixa, a princípio em qualquer posição do \mathbb{R}^3 que respeite a distância estabelecida entre estes vértices. No entanto, salientamos que pontos estratégicos podem ser escolhidos, de modo a facilitar o processo de cálculo computacional que é realizado pelo algoritmo.
2. Para garantirmos que todos os ramos da árvore do espaço de busca sejam visitados, vamos adotar, sem nenhum motivo especial, que a ordem da escolha do ramo a ser considerado na tentativa de descer na árvore será sempre o ramo “-” como primeira opção, e caso seja constatada uma poda neste ramo, consideramos o ramo “+”. Todavia, destacamos esta ordem não é a única possível.

Assim, dadas estas considerações, o pseudo-código do [Algoritmo 1](#) sistematiza a estrutura geral do BP aplicado ao PDGD.

Algoritmo 1: Estrutura principal do algoritmo BP.

Entrada: $G = (V, E, d)$, $X = \{x_1, x_2, x_3\}$ e $i = 4$
Saída: $X = \{x_v \in \mathbb{R}^3 \mid v \in V\}$

```

1 Função BP( $G, X, i$ )
2   para todo  $s \in \{-1, +1\}$  faça
3     compute  $x_i^s$ ;
4     se ( $x_i^s$  é factível) então
5        $X \leftarrow X \cup \{x_i^s\}$ ;
6       se ( $i = n$ ) então
7         retorne  $X$ ;
8       senão
9         BP( $G, X, i + 1$ );
10      fim
11    fim
12  fim
13  retroceda para o primeiro  $v_j < v_i$ , tal que  $v_j \in \mathcal{B}$  e troque o ramo escolhido;
14   $X \leftarrow X \setminus \{x_j, x_{j+1}, \dots, x_{i-1}\}$ ;
15  BP( $G, X, j$ );
16 fim

```

A imagem iterativa contida no *QR code* da [Figura 13](#), evidencia o funcionamento geral do BP quanto a descida na árvore do espaço de busca, e com isso a obtenção de uma solução, em um exemplo onde o grafo do PDGD possui sete vértices em que $k_4 = 3$, $k_5 = 3$, $k_6 = 4$ e $k_7 = 5$.



Figura 13 – Exemplo iterativo do funcionamento do BP. *Link* contido no *QR code*: <https://bit.ly/3t1Fx1Z>

Quanto ao BP, ele pode ser exponencial no pior caso, o que é consistente com o fato, que vimos anteriormente neste capítulo, do PDGD ser NP-difícil. No entanto, a referência [LAVOR et al. \(2012a\)](#) evidencia que do ponto de vista de eficiência e confiabilidade, o BP possui uma performance muito satisfatória. Entretanto, a referência [MUCHERINO; LAVOR; LIBERTI \(2012\)](#) mostra que em alguns casos, os quais não envolvem a nossa aplicação de interesse, se a instância do PDGD não possuir um conjunto de distâncias

bem dimensionado, o cálculo com ponto flutuante utilizado na realização dos vértices pode gerar erros numéricos grandes que, em alguns casos, impedem que o BP encontre soluções.

Quanto a utilizar o BP para encontrar todas as soluções do PDGD, note que a princípio isto só ocorrerá se explorarmos todo o espaço de busca. Como já dissemos, cada poda reduz o espaço de busca, no entanto, dependendo do número de podas, este processo ainda pode ser muito custoso computacionalmente.

Na próxima seção, trabalhamos em uma maneira de realizar um vértice em \mathbb{R}^3 . Salientamos que existem inúmeros processos distintos que podem cumprir esta tarefa, e a nossa escolha se baseia no fato de ser uma maneira intuitiva de se fazê-lo.

3.2.3 Uma Maneira Intuitiva de Calcular a Imersão de um Vértice no \mathbb{R}^3

Nesta seção, apresentamos um modo de calcular a imersão dos vértices no PDGD. Para exemplificar e expor outras maneiras de realização de vértices do PDGD, uma maneira sistemática de realização utilizando intersecção de esferas pode ser encontrada na referência [PETITJEAN \(2013\)](#). O trabalho [\(LAVOR et al., 2017\)](#) apresenta um modo de se realizar tais vértices dispondo de Álgebra Linear. Uma metodologia de realização utilizando Geometria Projetiva pode ser encontrada na referência [GONÇALVES; MUCHERINO \(2014\)](#), neste mesmo trabalho é mostrado que esta maneira apresenta uma eficiência computacional muito grande. Já a referência [ALVES et al. \(2018\)](#) realiza tais vértices utilizando Álgebra Geométrica Conforme.

Para a maneira de realização que apresentamos neste trabalho, optamos por fazê-la de uma maneira geométrica e que julgamos bem intuitiva. Ela foi construída utilizando como base a ideia de realização apresentada na referência [\(GONÇALVES et al., 2017\)](#). Deste ponto em diante neste trabalho, utilizamos muitos conceitos básicos de geometria analítica sobre vetores. Todos os resultados utilizados serão explicitados, mas não os demonstramos. Caso o leitor deseje lembrá-los, recomendamos a referência [\(WEXLER, 1962\)](#).

Pelo que já discutimos, sabemos que só existem vértices que estão associados a processos de ramificação ou de poda, e o que ocasiona esta diferença é o número de predecessores adjacentes que estes vértices possuem. Assim, se pensarmos no caráter geométrico da intersecção de esferas, notaremos que basta utilizarmos um processo que realize qualquer vértice utilizando somente três de seus predecessores adjacentes, visto que, isto implicará na existência de duas possibilidades para sua posição no \mathbb{R}^3 . Observamos que considerar estes dois pontos é a realidade dos vértices relacionados ao processo de ramificação, e quanto aos vértices de poda, podemos utilizar as restrições de distância associadas aos outros predecessores adjacentes deste vértice simplesmente verificando se os pontos encontrados satisfazem estas restrições.

Seja $G = (V, E, d)$ o grafo associado ao PDGD, com a ordem $\{v_1, v_2, \dots, v_n\}$ sobre os vértices de V . Desta maneira, queremos calcular a imersão no \mathbb{R}^3 do vértice v_i , com $i = 4, 5, \dots, n$. Assim, vamos considerar que v_{i_1}, v_{i_2} e v_{i_3} sejam vértices predecessores adjacentes de v_i . Portanto, queremos realizar o vértice v_i sabendo que já é conhecida a posição no \mathbb{R}^3 dos seus predecessores adjacentes. Vamos considerar que as duas posições para v_i são x_i^+ e x_i^- .

Supondo que os pontos $x_{i_3}, x_{i_2}, x_{i_1}$ e x_i não sejam coplanares, eles definem dois planos em \mathbb{R}^3 . A Figura 14 representa esta situação.

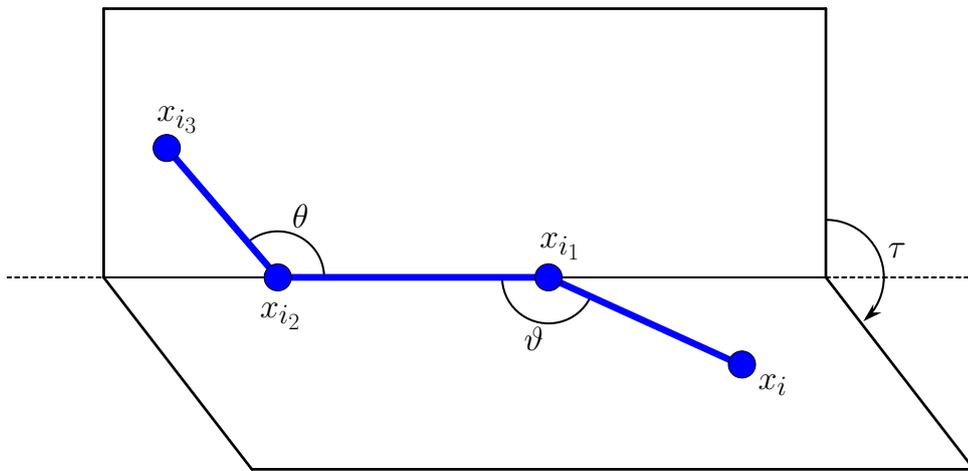


Figura 14 – Disposição dos pontos $x_{i_3}, x_{i_2}, x_{i_1}$ e x_i no espaço, ressaltando os dois planos definidos por estes quatro pontos. A distância entre alguns destes pontos é destacada pela linha contínua azul, e, θ , ϑ e τ evidenciam todos os ângulos necessários para definir a localização geométrica dos quatro pontos.

Como x_{i_3}, x_{i_2} e x_{i_1} já estão fixados, conhecemos a distância entre todos estes pontos e, portanto, pela lei dos cossenos, sabemos o valor do ângulo θ . As distâncias que envolvem o ponto x_i são dadas pelo sistema de equações a seguir, que é uma parte da instância do problema:

$$\begin{cases} \|x_i - x_{i_3}\| = d_{i,i_3}, \\ \|x_i - x_{i_2}\| = d_{i,i_2}, \\ \|x_i - x_{i_1}\| = d_{i,i_1}. \end{cases} \quad (3.10)$$

Agora, visto que também temos as distâncias entre os pontos x_{i_2}, x_{i_1} e x_i , de maneira análoga a anterior, conseguimos obter o valor do ângulo ϑ pela lei dos cossenos.

Para obter o ângulo τ , denominado de ângulo diedral ou ângulo de torção, necessitamos de um processo um pouco mais elaborado que envolve todas as distâncias entre os quatro pontos. A maneira de encontrarmos este valor é apresentada mais à frente nesta seção. Por hora, vamos considerar que o ângulo τ seja livre, ou seja, vamos desconsiderar que a distância entre os pontos x_{i_3} e x_i seja fixa. A família das circunferências tangentes

aos vetores normais dos planos $\{x_{i_3}, x_{i_2}, x_{i_1}\}$ e $\{x_{i_2}, x_{i_1}, x_i\}$, com centro c na reta formada pelos pontos x_{i_2} e x_{i_1} são descritas por este ângulo de torção τ . Se tomamos a circunferência que passa pelo ponto x_i , teremos uma relação entre o ponto x_i e o ângulo de torção $\tau \in [-180^\circ, 180^\circ]$. A Figura 15 ilustra geometricamente esta descrição.

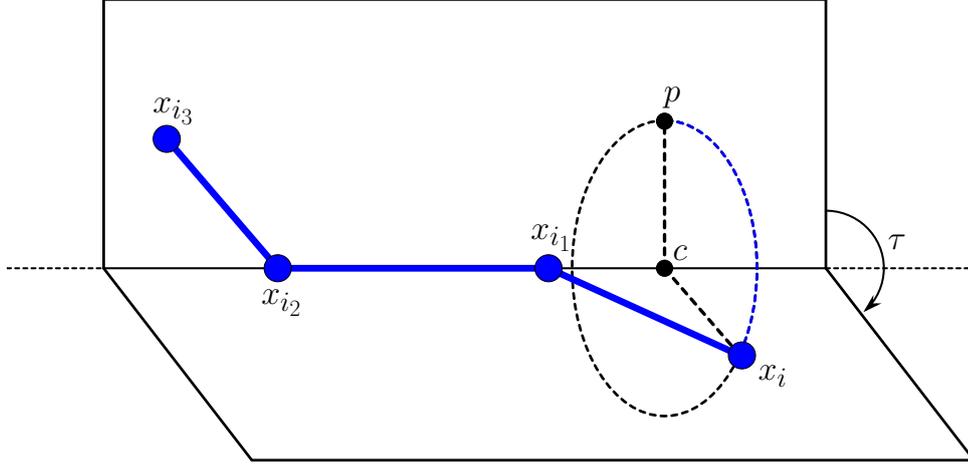


Figura 15 – Circunferência tangente aos vetores normais dos planos $\{x_{i_3}, x_{i_2}, x_{i_1}\}$ e $\{x_{i_2}, x_{i_1}, x_i\}$ com centro c na reta formada pelos pontos x_{i_2} e x_{i_1} . O ponto p é dado pela intersecção desta circunferência com o plano $\{x_{i_3}, x_{i_2}, x_{i_1}\}$.

Conseqüentemente, notemos que o ponto c é dado pela intersecção da reta formada por x_{i_2} e x_{i_1} com o plano π definido pelos pontos c , x_i e p . Podemos descrever, geometricamente, o plano π pela equação:

$$\hat{r} \cdot (x - x_i) = 0, \quad (3.11)$$

onde $\hat{r} = r/\|r\|$, com $r = x_{i_1} - x_{i_2}$, e x é um ponto genérico do \mathbb{R}^3 . Sobre a reta que passa por x_{i_2} e x_{i_1} , podemos descrevê-la pela equação:

$$u(t) = x_{i_2} + \hat{r}t, \text{ com } t \in \mathbb{R}.$$

Como c é um ponto que pertence a reta u , então existe $\kappa \in \mathbb{R}$ tal que:

$$c = u(\kappa) = x_{i_2} + \kappa\hat{r}. \quad (3.12)$$

E como c também pertence ao plano π , temos que:

$$\hat{r} \cdot (c - x_i) = 0. \quad (3.13)$$

Se combinamos as Equações (3.12) e (3.13), temos o seguinte valor para κ :

$$\begin{aligned} \hat{r} \cdot (c - x_i) &= 0 && \Leftrightarrow \\ \hat{r} \cdot [(x_{i_2} + \hat{r}\kappa) - x_i] &= 0 && \Leftrightarrow \\ \kappa &= \hat{r} \cdot x_i - \hat{r} \cdot x_{i_2}. \end{aligned} \quad (3.14)$$

Notemos que se elevamos os dois lados das equações definidas no Sistema (3.10) ao quadrado, temos três esferas, $S_1^i = S(x_{i_1}, d_{i,i_1})$, $S_2^i = S(x_{i_2}, d_{i,i_2})$ e $S_3^i = S(x_{i_3}, d_{i,i_3})$, cuja intersecção resulta nos dois valores possíveis para x_i . Desta maneira, como o centro das esferas S_2^i e S_1^i pertencem a reta u , temos que a circunferência resultante da intersecção destas duas esferas também possui o centro na reta u . A Figura 16 ilustra a intersecção das esferas S_2^i e S_1^i , evidencia o plano π e a circunferência de intersecção.

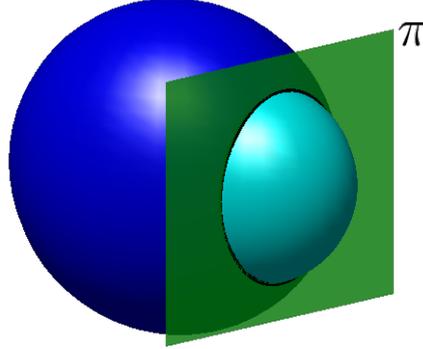


Figura 16 – Intersecção das esferas $S_2^i = S(x_{i_2}, d_{i,i_2})$ e $S_1^i = S(x_{i_1}, d_{i,i_1})$, representadas, respectivamente, pelas cores azul e ciano. O plano π que contém esta intersecção é apresentado em verde.

Da equação das esferas S_2^i e S_1^i , vamos observar o seguinte:

$$\begin{cases} S_2^i = S(x_{i_2}, d_{i,i_2}) \\ S_1^i = S(x_{i_1}, d_{i,i_1}) \end{cases} \Leftrightarrow \begin{cases} \|x_i - x_{i_2}\|^2 = d_{i,i_2}^2 \\ \|x_i - x_{i_1}\|^2 = d_{i,i_1}^2 \end{cases} \Leftrightarrow \begin{cases} \|x_i\|^2 - 2x_i \cdot x_{i_2} + \|x_{i_2}\|^2 = d_{i,i_2}^2 \\ \|x_i\|^2 - 2x_i \cdot x_{i_1} + \|x_{i_1}\|^2 = d_{i,i_1}^2 \end{cases}$$

Subtraindo a primeira da segunda equação, temos:

$$\begin{aligned} -2(x_{i_1} - x_{i_2}) \cdot x_i + \|x_{i_1}\|^2 - \|x_{i_2}\|^2 &= d_{i,i_1}^2 - d_{i,i_2}^2 \\ (x_{i_1} - x_{i_2}) \cdot x_i &= \frac{d_{i,i_2}^2 - \|x_{i_2}\|^2 - d_{i,i_1}^2 + \|x_{i_1}\|^2}{2} \\ \hat{r} \cdot x_i &= \frac{1}{2\|r\|} (d_{i,i_2}^2 - \|x_{i_2}\|^2 - d_{i,i_1}^2 + \|x_{i_1}\|^2). \end{aligned}$$

Assim, utilizando a quantidade obtida na relação anterior na Equação (3.14), temos o seguinte valor para κ :

$$\begin{aligned} \kappa &= \hat{r} \cdot x_i - \hat{r} \cdot x_{i_2} \\ &= \left[\frac{1}{2\|r\|} (d_{i,i_2}^2 - \|x_{i_2}\|^2 - d_{i,i_1}^2 + \|x_{i_1}\|^2) \right] - \hat{r} \cdot x_{i_2} \\ &= \frac{1}{2\|r\|} (d_{i,i_2}^2 - \|x_{i_2}\|^2 - d_{i,i_1}^2 + \|x_{i_1}\|^2 - 2r \cdot x_{i_2}) \\ &= \frac{1}{2\|x_{i_1} - x_{i_2}\|} (d_{i,i_2}^2 - \|x_{i_2}\|^2 - d_{i,i_1}^2 + \|x_{i_1}\|^2 - 2(x_{i_1} - x_{i_2}) \cdot x_{i_2}) \end{aligned}$$

logo,

$$\kappa = \frac{d_{i,i_2}^2 + d_{i_1,i_2}^2 - d_{i,i_1}^2}{2d_{i_1,i_2}}. \quad (3.15)$$

Portanto, dado o valor de κ , pela Equação (3.12) fica definido o ponto c com relação aos pontos x_{i_2} , x_{i_1} e com as distâncias d_{i_1,i_2} , d_{i,i_2} e d_{i,i_1} .

Desta maneira, a circunferência da Figura 15 pode ser parametrizada se rotacionarmos o ponto p com relação ao eixo definido pela reta que passa por x_{i_2} e x_{i_1} . Como sabemos, a parametrização desta circunferência não é única, no entanto, o ponto p é o candidato natural para realizar esta parametrização, pois ele pertence ao plano $\{x_{i_3}, x_{i_2}, x_{i_1}\}$ que é definido *a priori*. Assim, se o ângulo τ é fixado, então o plano $\{x_{i_2}, x_{i_1}, x_i\}$ fica definido e, com isso, conseguimos o valor de x_i . A parametrização desta circunferência é a relação entre o ponto x_i e o ângulo τ que salientamos anteriormente.

Agora, notemos que a reta que passa pelos pontos c e p é dada por:

$$u^\perp(t) = c + \hat{s}t, \text{ com } t \in \mathbb{R},$$

onde o vetor \hat{s} é tal que $\hat{s} \perp \hat{r}$ e $\|\hat{s}\| = 1$.

Como a reta u^\perp foi construída passando por p , se ρ é o raio da circunferência da Figura 15, então temos que:

$$u^\perp(\rho) = p = c + \rho\hat{s}. \quad (3.16)$$

Para encontrarmos a direção \hat{s} , vamos considerar o vetor $v = x_{i_3} - x_{i_2}$. Notemos que os vetores v , \hat{r} e \hat{s} estão contidos no plano $\{x_{i_3}, x_{i_2}, x_{i_1}\}$ e que os vetores \hat{r} e v são linearmente independentes, uma vez que os pontos x_{i_3} , x_{i_2} e x_{i_1} não são colineares. Portanto, conseguiremos dois vetores ortogonais no plano definido por \hat{r} e v se aplicarmos o processo de ortogonalização de Gram-Schmidt (STRANG, 2016) no conjunto $\{\hat{r}, \hat{v}\}$, aqui $\hat{v} = v/\|v\|$. Assim, seguindo com esta estratégia, fixamos \hat{r} como o primeiro vetor do processo e, com isso, temos que um vetor ortogonal a \hat{r} é:

$$s = \hat{v} - (\hat{v} \cdot \hat{r})\hat{r}.$$

A Figura 17 ilustra os vetores v , r e s . Note que a direção tomada para os vetores r e v resultaram no vetor s com a orientação apresentada nesta figura.

Logo, o vetor unitário \hat{s} é obtido por meio da divisão de s por sua norma, que é calculada facilmente por:

$$\begin{aligned} \|s\|^2 &= \|\hat{v} - (\hat{v} \cdot \hat{r})\hat{r}\|^2 \\ &= 1 - 2(\hat{v} \cdot \hat{r})(\hat{v} \cdot \hat{r}) + (\hat{v} \cdot \hat{r})^2 \\ &= 1 - (\hat{v} \cdot \hat{r})^2 \end{aligned}$$

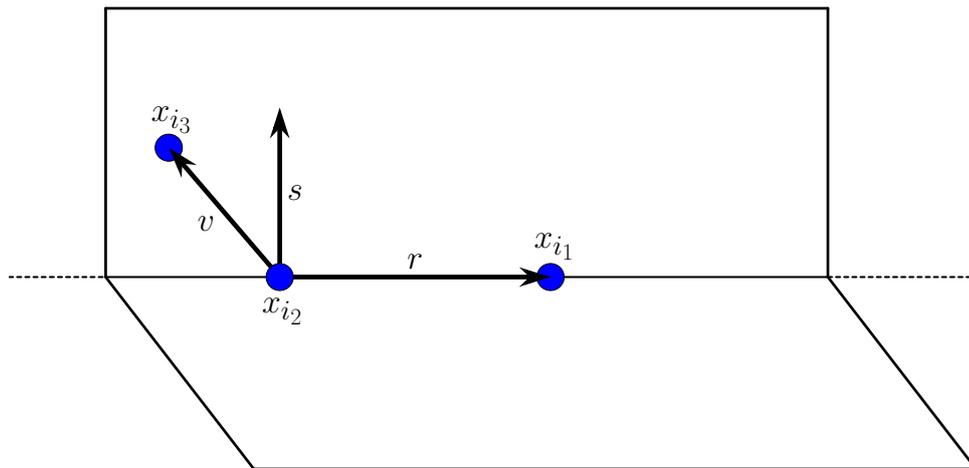


Figura 17 – Representação geométrica dos vetores v , r e s . Aqui os vetores ortogonais \hat{r} e s foram encontrados pela aplicação do processo de ortogonalização de Gram-Schmidt nos vetores \hat{r} e \hat{v} .

mas, $\hat{v} \cdot \hat{r} = \cos \theta$, portanto:

$$\begin{aligned} \|s\| &= \sqrt{1 - \cos^2 \theta} \\ &= \sin \theta \end{aligned}$$

Com isso, temos que:

$$\hat{s} = \frac{\hat{v} - (\hat{v} \cdot \hat{r})\hat{r}}{\sin \theta} = \frac{\hat{v} - \hat{r} \cos \theta}{\sin \theta}. \quad (3.17)$$

Vale ressaltar novamente que $\sin \theta \neq 0$, pois os pontos x_{i_3} , x_{i_2} e x_{i_1} não são colineares.

Agora, para obtermos o valor de ρ , basta olharmos para o triângulo definido pelos pontos x_{i_2} , c e x_i definidos no plano $\{x_{i_2}, x_{i_1}, x_i\}$. A Figura 18 ilustra este triângulo.

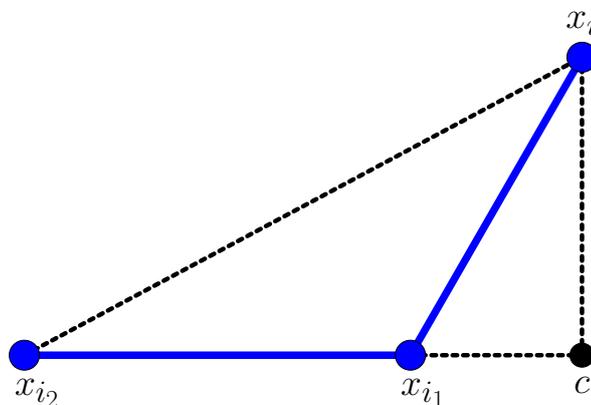


Figura 18 – Triângulo reto contido no plano $\{x_{i_2}, x_{i_1}, x_i\}$ definido pelos pontos x_{i_2} , c e x_i .

Dada a maneira com que o ponto c foi construído, notemos que o ângulo $\widehat{x_{i_2}cx_i}$ é reto. Como $\rho = \|c - x_i\|$, utilizando o teorema de Pitágoras no triângulo $\Delta x_{i_2}cx_i$, temos que:

$$\begin{aligned} \|x_{i_2} - x_i\|^2 &= \|c - x_{i_2}\|^2 + \|c - x_i\|^2 \Leftrightarrow \\ d_{i,i_2}^2 &= \|(x_{i_2} + \kappa\hat{r}) - x_{i_2}\|^2 + \rho^2 \Leftrightarrow \\ d_{i,i_2}^2 &= \|\kappa\hat{r}\|^2 + \rho^2 \end{aligned}$$

logo,

$$\rho = \sqrt{d_{i,i_2}^2 - \kappa^2}. \quad (3.18)$$

Como a Equação (3.17) apresenta \hat{s} em função do seno e do cosseno de θ , para encontrar estes valores, vamos observar que a lei dos cossenos aplicada ao triângulo $x_{i_3}x_{i_2}x_{i_1}$ resulta em:

$$d_{i_1,i_3}^2 = d_{i_2,i_3}^2 + d_{i_1,i_2}^2 - 2d_{i_2,i_3}d_{i_1,i_2} \cos \theta. \quad (3.19)$$

Desta maneira, com base na Equação (3.15), definiremos a quantidade $\bar{\kappa}$ da seguinte maneira:

$$\bar{\kappa} \equiv \frac{d_{i_2,i_3}^2 + d_{i_1,i_2}^2 - d_{i_1,i_3}^2}{2d_{i_1,i_2}}. \quad (3.20)$$

Comparando a quantidade definida em (3.20) com a Equação (3.19), temos que:

$$\bar{\kappa} = d_{i_2,i_3} \cos \theta. \quad (3.21)$$

Agora, definindo $\bar{\rho}$ como:

$$\bar{\rho} \equiv \sqrt{d_{i_2,i_3}^2 - \bar{\kappa}^2}, \quad (3.22)$$

e utilizando a Equação (3.21), conseguimos a seguinte relação para $\bar{\rho}$ e $\sin \theta$:

$$\begin{aligned} \bar{\rho} &= \sqrt{d_{i_2,i_3}^2 - \bar{\kappa}^2} \\ &= \sqrt{d_{i_2,i_3}^2 - (d_{i_2,i_3} \cos \theta)^2} \\ &= \sqrt{d_{i_2,i_3}^2 (1 - \cos^2 \theta)} \\ &= d_{i_2,i_3} \sin \theta. \end{aligned} \quad (3.23)$$

Logo, podemos reescrever o valor de \hat{s} , dado pela Equação (3.17), utilizando as quantidades $\bar{\kappa}$ e $\bar{\rho}$, apresentadas respectivamente pelas Equações (3.21) e (3.23), da seguinte maneira:

$$\begin{aligned} \hat{s} &= \frac{\hat{v} - \hat{r} \cos \theta}{\sin \theta} = \frac{v/d_{i_2,i_3} - \hat{r}\bar{\kappa}/d_{i_2,i_3}}{\bar{\rho}/d_{i_2,i_3}} \\ &= \frac{v - \bar{\kappa}\hat{r}}{\bar{\rho}}. \end{aligned} \quad (3.24)$$

Portanto, utilizando os valores de c , \hat{s} , κ , ρ , $\bar{\kappa}$ e $\bar{\rho}$, descritos, respectivamente, pelas Equações (3.12), (3.24), (3.15), (3.18), (3.20) e (3.22), na Equação (3.16), temos que o ponto p é dado por:

$$p = x_{i_2} + \kappa \hat{r} + \frac{\rho}{\bar{\rho}}(v - \bar{\kappa} \hat{r}). \quad (3.25)$$

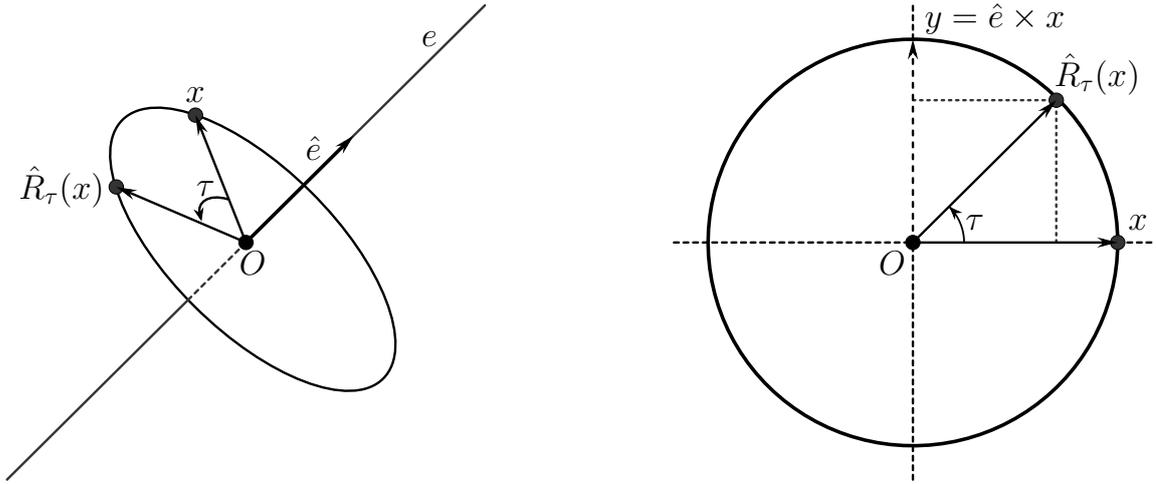
Com isso, o ponto p fica descrito em função dos pontos x_{i_3} , x_{i_2} e x_{i_1} , e das distâncias d_{i_1,i_2} , d_{i_1,i_3} , d_{i_2,i_3} , d_{i,i_2} e d_{i,i_1} .

Agora, dado que temos o ponto p , vamos descrever como podemos realizar a parametrização da circunferência da Figura 15 rotacionando este ponto em torno da reta definida pelos pontos x_{i_2} e x_{i_1} . Para isso, primeiramente, vamos observar o que devemos fazer para descrever a rotação de um ponto no espaço. Esta tarefa é mais simples se o centro da circunferência, que descreve esta rotação, é a origem. Como a traslação que deve ser feita para reposicionar este ponto no espaço não envolve nenhuma dificuldade matemática, além de uma operação de soma, vamos entender como podemos rotacionar um ponto no \mathbb{R}^3 em torno da origem e, em seguida, ver como podemos trasladar esta circunferência para uma posição de interesse. Para garantirmos que os diferentes eixos de rotação construam uma mesma rotação, desconsiderando a operação de traslação, vamos observar o seguinte:

- O eixo de rotação que contém a origem é descrito por uma reta paralela a reta que define o eixo de rotação original;
- A orientação dos dois eixos de rotação considerados é a mesma.
- A reta definida pelo centro da rotação e o ponto a ser rotacionado, nos dois casos considerados, são paralelas e possuem a mesma orientação;
- A distância do ponto rotacionado para o eixo de rotação é a mesma nos dois casos;

Para prosseguir com a discussão, vamos considerar que r , p , c e ρ sejam, respectivamente, a reta que descreve o eixo de rotação, o ponto a ser rotacionado, o centro de rotação, na reta r , do ponto p , e o raio da circunferência definida pela rotação. Assim, se e e x são o novo par, eixo de rotação e ponto a ser rotacionado, tal que a origem O pertence a reta e . Tem-se que, $x = \rho \hat{x} = p - c$. Então, as quatro considerações anteriores são satisfeitas para os pares $\{r, p\}$ e $\{e, x\}$. Seja $\hat{R}_x(\tau)$ a rotação de um ângulo τ do ponto x em torno do eixo e . A Figura 19 (a) ilustra esta rotação geometricamente. A parametrização da circunferência apresentada na Figura 19 (b) pode ser dada por:

$$\begin{aligned} \hat{R}_x(\tau) &= \rho \hat{x} \cos \tau + \rho \hat{y} \sin \tau \\ &= \rho \hat{x} \cos \tau + \rho(\hat{e} \times \hat{x}) \sin \tau \\ &= (\rho \hat{x}) \cos \tau + \hat{e} \times (\rho \hat{x}) \sin \tau \end{aligned}$$



(a) Representação geométrica da rotação do ponto x em torno do eixo de rotação definido pela reta e .

(b) Vista axial da circunferência definida por esta rotação. O ponto O é a origem do \mathbb{R}^3 .

Figura 19 – Ilustração da rotação do ponto x através do eixo definido pela reta e que contém a origem O do espaço.

com $\tau \in [-180^\circ, 180^\circ]$. Como $\hat{e} = \hat{r}$, a rotação do ponto p em torno do eixo r em um ângulo τ , que definimos como $R_\tau(p)$, é obtida simplesmente transladando $\hat{R}_\tau(x)$ de uma quantidade c . Portanto, $R_\tau(p)$ é dada por:

$$\begin{aligned} R_p(\tau) &= c + \hat{R}_x(\tau) \\ &= c + (\rho \hat{x}) \cos \tau + \hat{e} \times (\rho \hat{x}) \sin \tau \\ &= c + (p - c) \cos \tau + \hat{r} \times (p - c) \sin \tau. \end{aligned} \quad (3.26)$$

Isto posto, podemos escrever a relação entre o ponto x_i e o ângulo τ , em função dos valores de c e p apresentados nas Equações (3.12) e (3.25), respectivamente, da seguinte maneira:

$$x_i(\tau) = x_{i_2} + \kappa \hat{r} + \frac{\rho}{\rho} \left((v - \bar{\kappa} \hat{r}) \cos \tau + \hat{r} \times v \sin \tau \right). \quad (3.27)$$

Desta maneira, a parametrização da circunferência da Figura 15 é dada tomando $\tau \in (-180^\circ, 180^\circ]$ na Equação anterior.

Assim, dada a Equação (3.27), para encontrar as duas posições no espaço para o vértice v_i , basta nós estabelecermos a relação do ângulo τ com a distância dos pontos x_{i_3} e x_i . Uma vez estabelecida esta relação, utilizamos esta equação para encontrar as posições para o ponto x_i . Como já salientamos, o processo para descrever o ângulo de torção τ é um pouco mais elaborado que utilizar somente a lei dos cossenos. No entanto, ele pode ser encontrado utilizando o que a referência POGORELOV (1987) descreve como “lei dos cossenos para um ângulo triédrico”. Esta lei é descrita pelo teorema apresentado a seguir:

Teorema 3.1. (Lei dos cossenos para um ângulo triédrico): Sejam α , β e γ os ângulos formados pelas 3 semi-retas de um triedro com vértice S , e τ o ângulo diedral oposto ao ângulo γ (a Figura 20 ilustra o ângulo triédrico aqui descrito). Então:

$$\cos \gamma = \cos \alpha \cos \beta + \sin \alpha \sin \beta \cos \tau. \tag{3.28}$$

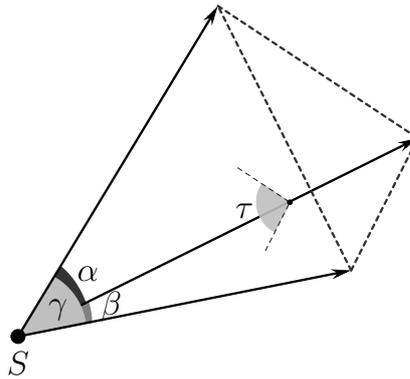


Figura 20 – Ângulo triédrico com vértice no ponto S , como descrito no Teorema 3.1.

Demonstração. A demonstração deste teorema não apresenta nenhuma dificuldade que devemos salientar e ela pode ser encontrada na pág 278 do livro [POGORELOV \(1987\)](#) □

Assim, considerando as posições dos pontos x_{i_3} , x_{i_2} , x_{i_1} e x_i da [Figura 14](#), podemos interpretar um ângulo triédrico com vértice no ponto x_{i_2} , e com isso tirar as conclusões apresentadas no Teorema 3.1 para uma relação dos ângulos definidos entre estes quatro pontos. A [Figura 21](#) evidencia o ângulo triédrico mencionado considerando os pontos x_{i_3} , x_{i_2} , x_{i_1} e x_i .

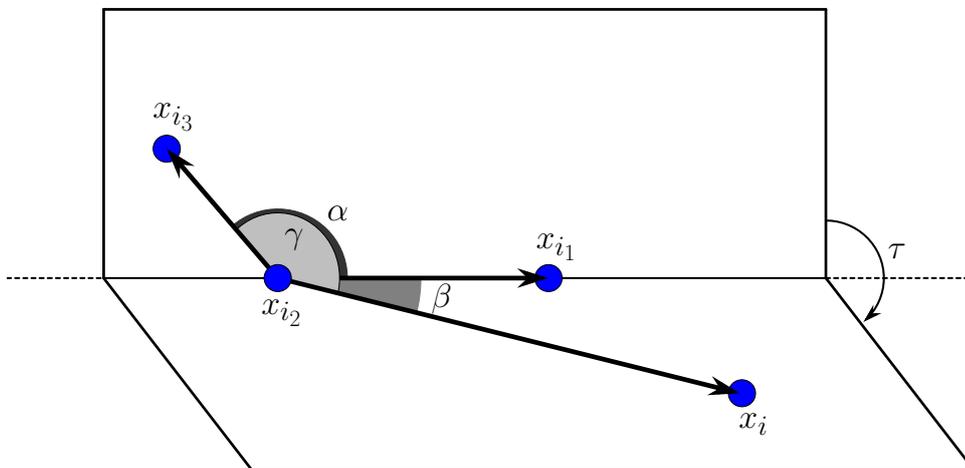


Figura 21 – Ângulo triédrico formado pelos pontos x_{i_3} , x_{i_2} , x_{i_1} e x_i com vértice em x_{i_2} .

Desta maneira, para termos o cosseno dos ângulos α , β e γ em função das distâncias entre os pontos x_{i_3} , x_{i_2} , x_{i_1} e x_i , podemos utilizar a lei dos cossenos, que, para cada um destes ângulos, nos diz que:

$$d_{i_1,i_3}^2 = d_{i_1,i_2}^2 + d_{i_2,i_3}^2 - 2d_{i_1,i_2}d_{i_2,i_3} \cos \alpha \quad (3.29)$$

$$d_{i,i_1}^2 = d_{i,i_2}^2 + d_{i_1,i_2}^2 - 2d_{i,i_2}d_{i_1,i_2} \cos \beta \quad (3.30)$$

$$d_{i,i_3}^2 = d_{i,i_2}^2 + d_{i_2,i_3}^2 - 2d_{i,i_2}d_{i_2,i_3} \cos \gamma \quad (3.31)$$

Agora, observemos que a Equação (3.15) pode ser reescrita como:

$$d_{i,i_1}^2 = d_{i_1,i_2}^2 + d_{i,i_2}^2 - 2d_{i_1,i_2} \cdot \kappa, \quad (3.32)$$

desta forma, se compararmos as Equações (3.32) e (3.29), então podemos fazer a seguinte associação pra o valor de κ e o ângulo β :

$$\kappa = d_{i,i_2} \cos \beta. \quad (3.33)$$

Utilizando a relação (3.33) para descrever κ na Equação (3.18), temos que ρ é dado, função do ângulo β , da seguinte maneira:

$$\begin{aligned} \rho &= \sqrt{d_{i,i_2}^2 - \kappa^2} \\ &= \sqrt{d_{i,i_2}^2 - (d_{i,i_2} \cos \beta)^2} \\ &= d_{i,i_2} \sin \beta. \end{aligned} \quad (3.34)$$

Agora, notando que o ângulo θ da Figura 14 é definido da mesma maneira que o ângulo α da Figura 21, das Equações (3.21) e (3.23), temos que:

$$\begin{aligned} \bar{\kappa} &= d_{i_2,i_3} \cos \alpha \\ \bar{\rho} &= d_{i_2,i_3} \sin \alpha \end{aligned} \quad (3.35)$$

Assim, dadas as relações de κ e ρ com o ângulo β e de $\bar{\kappa}$ e $\bar{\rho}$ com o ângulo α , notemos que:

$$\begin{aligned} d_{i_2,i_3}^2 &= d_{i_2,i_3}^2 \cdot 1 = d_{i_2,i_3}^2 (\cos^2 \alpha + \sin^2 \alpha) = d_{i_2,i_3}^2 \cos^2 \alpha + d_{i_2,i_3}^2 \sin^2 \alpha = \bar{\kappa}^2 + \bar{\rho}^2 \\ d_{i,i_2}^2 &= d_{i,i_2}^2 \cdot 1 = d_{i,i_2}^2 (\cos^2 \beta + \sin^2 \beta) = d_{i,i_2}^2 \cos^2 \beta + d_{i,i_2}^2 \sin^2 \beta = \kappa^2 + \rho^2 \end{aligned}$$

Logo, utilizando estes valores na Equação (3.31), temos o seguinte valor para o cosseno do ângulo γ :

$$\cos \gamma = \frac{(\bar{\kappa}^2 + \bar{\rho}^2) + (\kappa^2 + \rho^2) - d_{i,i_3}^2}{2d_{i,i_2}d_{i_2,i_3}}. \quad (3.36)$$

Portanto, dadas as relações de cossenos e senos dos ângulos α , β e γ , apresentadas nas Equações (3.33), (3.34), (3.35) e (3.36), aplicando-as no resultado do Teorema 3.1, descrito pela Equação (3.28), temos que:

$$\begin{aligned} \cos \gamma &= \cos \alpha \cos \beta + \sin \alpha \sin \beta \cos \tau && \Leftrightarrow \\ \frac{\kappa^2 + \bar{\kappa}^2 + \rho^2 + \bar{\rho}^2 - d_{i,i_3}^2}{2d_{i_2,i_3}d_{i,i_2}} &= \left(\frac{\bar{\kappa}}{d_{i_2,i_3}} \right) \left(\frac{\kappa}{d_{i,i_2}} \right) + \left(\frac{\bar{\rho}}{d_{i_2,i_3}} \right) \left(\frac{\rho}{d_{i,i_2}} \right) \cos \tau && \Leftrightarrow \\ \cos \tau &= \frac{(\kappa - \bar{\kappa})^2 + \rho^2 + \bar{\rho}^2 - d_{i,i_3}^2}{2\rho\bar{\rho}} \end{aligned}$$

Desta maneira, o cosseno do ângulo de torção τ é dado por:

$$\cos \tau = \frac{\nu - d_{i,i_3}^2}{2\mu}, \quad (3.37)$$

onde,

$$\nu = (\kappa - \bar{\kappa})^2 + \rho^2 + \bar{\rho}^2 \quad \text{e} \quad \mu = \rho\bar{\rho}.$$

Agora que estabelecemos a relação entre o cosseno do ângulo de torção e a distância entre os pontos x_{i_3} e x_i , basta aplicar este valor de ângulo na Equação (3.27), que descreve a posição do ponto x_i em função deste ângulo de torção, para encontrar a posição do ponto x_i . No entanto, como já foi discutido anteriormente, esta equação deve produzir duas possibilidades para o valor de x_i , x_i^+ e x_i^- . Assim, observando que o valor de $x_i(\tau)$ descrito por esta equação necessita dos valores de seno e cosseno do ângulo τ , atentemo-nos ao fato de que dado o valor do cosseno de τ , descrito pela Equação (3.37), o valor do seno de τ fica dado por:

$$\sin \tau = \pm \sqrt{1 - \cos^2 \tau}.$$

Logo, estes dois valores para o seno de τ evidenciam as duas possibilidades para a posição de x_i . Com a finalidade de justificar que os dois valores do seno de τ produzem as duas posições, simétricas com relação ao plano $\{x_{i_3}, x_{i_2}, x_{i_1}\}$, para o ponto x_i , pela paridade das funções seno e cosseno podemos escrever:

$$\begin{aligned} x_i(\pm|\tau|) &= x_{i_2} + \kappa\hat{r} + \frac{\rho}{\rho} \left((v - \bar{\kappa}\hat{r}) \cos(\pm|\tau|) + \hat{r} \times v \sin(\pm|\tau|) \right) \\ &= x_{i_2} + \kappa\hat{r} + \frac{\rho}{\rho} \left((v - \bar{\kappa}\hat{r}) \cos \tau \pm \hat{r} \times v \sin |\tau| \right) \end{aligned} \quad (3.38)$$

Assim, os dois valores do seno de τ podem ser associados com a orientação tomada pelo ângulo de torção τ com relação ao plano $\{x_{i_3}, x_{i_2}, x_{i_1}\}$. O que tem todo sentido geométrico, visto que, a Equação (3.27) descreve a posição do ponto x_i rotacionando um ponto pertencente ao plano $\{x_{i_3}, x_{i_2}, x_{i_1}\}$. Isto justifica a simetria do ponto x_i com relação

a este plano. Portanto, as duas posições para a realização do vértice v_i são dadas pela Equação (3.27), da seguinte maneira:

$$x_i^\pm = x_i(\pm|\tau|) = x_{i_2} + \kappa\hat{r} + \frac{\rho}{\rho} \left((v - \bar{\kappa}\hat{r}) \cos \tau \pm \hat{r} \times v \sin |\tau| \right) \quad (3.39)$$

com τ dado pela Equação (3.37). Desta maneira, a relação acima descreve as duas possibilidades para a imersão do vértice v_i no \mathbb{R}^3 em função dos pontos x_{i_3} , x_{i_2} e x_{i_1} , e das distâncias d_{i_1,i_2} , d_{i_1,i_3} , d_{i_2,i_3} , d_{i,i_2} , d_{i,i_1} e d_{i,i_3} .

Como salientamos no início desta subsecção, a descrição que fizemos utiliza somente três predecessores adjacentes do vértice v_i . No caso em que $v_i \in \mathcal{P}$, isto é, possui mais de três predecessores adjacentes, para verificar as restrições de distância dos $k_i - 3$ predecessores adjacentes de v_i que não utilizamos no cálculo dos pontos x_i^\pm , podemos calcular as quantidades

$$\epsilon_\pm = \max_{\ell \in \{4, \dots, k_i\}} \left\{ \|x_i^\pm - x_{i_\ell}\| - d_{i,i_\ell} \right\}$$

e, desta maneira, o ponto x_i^\pm é factível se $\epsilon_\pm = 0$. Vale ressaltar que a imersão no \mathbb{R}^3 de todos os vértices de U_i é conhecida antes de calcularmos as posições x_i^\pm . Deste modo, as quantidades acima podem ser calculadas de maneira trivial.

A maneira que descrevemos a imersão do vértice v_i leva em conta os predecessores adjacentes v_{i_3} , v_{i_2} e v_{i_1} . No entanto, para descrever este cálculo de um modo geral, isto é, com base em quaisquer três predecessores adjacentes de v_i , digamos $\{v_{i_j}, v_{i_k}, v_{i_\ell}\} \subset U_i$, com $j, k, \ell \in \{1, 2, 3, \dots, k_i\}$, podemos rearranjar os índices da relação (3.39) e escrever as duas posições no \mathbb{R}^3 do vértice v_i , que são simétricas com relação ao plano $\{x_{i_j}, x_{i_k}, x_{i_\ell}\}$, da seguinte maneira:

$${}^{(j,k,\ell)}x_i^\pm(\tau_{i_\ell, i_k, i_j, i}) = \mathbf{a}(i_k, i_j, i) + \mathbf{b}(i_\ell, i_k, i_j, i) \cos \tau_{i_\ell, i_k, i_j, i} \pm \mathbf{c}(i_\ell, i_k, i_j, i) \sin |\tau_{i_\ell, i_k, i_j, i}|, \quad (3.40)$$

com,

$$\begin{aligned} \mathbf{a}(u, v, w) &= x_u + \kappa(u, v, w)\hat{r}(u, v), \\ \mathbf{b}(z, u, v, w) &= \frac{\rho^2(u, v, w)}{\mu(z, u, v, w)} \left(r(u, z) - \kappa(u, v, z)\hat{r}(u, v) \right), \end{aligned} \quad (3.41)$$

$$\mathbf{c}(z, u, v, w) = \hat{r}(u, v) \times \mathbf{b}(z, u, v, w),$$

$$\cos \tau_{z,u,v,w} = \frac{\nu(z, u, v, w) - d_{z,w}^2}{2\mu(z, u, v, w)}, \quad (3.42)$$

$$\sin |\tau_{z,u,v,w}| = \sqrt{1 - \cos^2 \tau_{z,u,v,w}}$$

onde,

$$\begin{aligned}
r(u, v) &= x_v - x_u, & \hat{r}(u, v) &= \frac{r(u, v)}{d_{u,v}} \\
\kappa(u, v, w) &= \frac{d_{u,w}^2 + d_{u,v}^2 - d_{v,w}^2}{2d_{u,v}}, & \rho^2(u, v, w) &= d_{u,w}^2 - \kappa^2(u, v, w), \\
\nu(z, u, v, w) &= \left(\kappa(u, v, w) - \kappa(u, v, z) \right)^2 + \rho^2(u, v, w) + \rho^2(u, v, z), \\
\mu(z, u, v, w) &= \sqrt{\rho^2(u, v, w)\rho^2(u, v, z)}.
\end{aligned} \tag{3.43}$$

Portanto, a Equação (3.40) descreve um processo geral para a imersão no \mathbb{R}^3 de qualquer vértice do grafo G associado ao PDGD. Na próxima seção, tratamos o Problema Discretizável de Geometria de Distâncias Intervalares, que é o aporte teórico que utilizamos na aplicação de interesse.

3.3 Problema Discretizável de Geometria de Distâncias Intervalares em \mathbb{R}^3

Como descrevemos no Capítulo 2, a RMN é um experimento cujas medidas possuem erros associados a ela. Assim, para representarmos uma situação mais próxima da realidade, devemos considerar as incertezas dos dados experimentais. Desta maneira, para levar em conta este fato na modelagem do problema, é necessário considerar que as distâncias fornecidas pela RMN são distâncias intervalares, ou seja, intervalos de números reais. Os conceitos do PGD e do PDGD que apresentamos nas duas seções anteriores não são suficientes para considerar esta nova suposição. No entanto, a ideia da generalização do PGD e do PDGD é similar as definições já apresentadas.

Nesta seção, tratamos da extensão intervalar do PGD, do PDGD e do BP. Para uma descrição mais completa destes novos conceitos recomendamos as referências [LIBERTI et al. \(2014\)](#), [GONÇALVES et al. \(2017\)](#) e [LAVOR et al. \(2021b\)](#).

Definimos o Problema de Geometria de Distâncias Intervalares (PGDi) em \mathbb{R}^3 da seguinte maneira:

Definição 3.5. Seja $G = (V, E, \mathbf{d})$ um grafo simples, ponderado e não-direcionado, com $\mathbf{d} : E \rightarrow \mathcal{I}(\mathbb{R}_+)$, onde $\mathcal{I}(\mathbb{R}_+)$ é um conjunto de intervalos em \mathbb{R}_+ . O PGDi \mathbb{R}^3 consiste em encontrar uma função $x : V \rightarrow \mathbb{R}^3$, tal que

$$\forall \{u, v\} \in E, \|x(u) - x(v)\| \in \mathbf{d}(\{u, v\}) \equiv [\underline{d}_{u,v}, \bar{d}_{u,v}]. \tag{3.44}$$

A existência de soluções deste problema é discutida em [LIBERTI et al. \(2014\)](#). Neste trabalho, vamos considerar que este problema possui solução.

Da definição do PDG*i*, notamos que:

1. Como $\mathbf{d}(\{u, v\}) = \{d \in \mathbb{R}_+ \mid \underline{d}_{u,v} \leq d \leq \bar{d}_{u,v}\}$, o sistema de equações de (3.44) pode ser reescrito como:

$$\forall \{u, v\} \in E, \underline{d}_{u,v} \leq \|x_u - x_v\| \leq \bar{d}_{u,v} \quad (3.45)$$

2. Se $\underline{d}_{u,v} = \bar{d}_{u,v} = d_{u,v}$, $\forall \{u, v\} \in E$, o sistema de equações de (3.44) é reescrito como: $\forall \{u, v\} \in E, \|x_u - x_v\| = d_{u,v}$. Então recuperamos a definição do PGD.

A segunda afirmação anterior nos mostra porque podemos considerar o PGD*i* como uma generalização do PGD. Já a primeira afirmação nos mostra uma maneira alternativa de escrever as restrições de distâncias do PGD*i*. Consideramos que esta nova representação é mais intuitiva, desta maneira, vamos utilizar a representação dada no sistema (3.45) para descrever a instância do PGD*i*.

Como salientamos no [Capítulo 2](#), conseguimos considerar que a distância entre alguns pares de átomos é precisa, enquanto que para outros pares temos que a distância é um intervalo. Para modelar este fato, vamos definir os seguintes conjuntos:

Definição 3.6. Sejam $G = (V, E, \mathbf{d})$ um grafo simples, ponderado e não-direcionado com $\mathbf{d} : E \rightarrow \mathcal{I}(\mathbb{R}_+)$. Então:

$$\begin{aligned} E_0 &\equiv \{\{u, v\} \in E \mid \underline{d}_{u,v} = \bar{d}_{u,v}\}, \\ E_I &\equiv \{\{u, v\} \in E \mid \underline{d}_{u,v} < \bar{d}_{u,v}\}. \end{aligned} \quad (3.46)$$

Da Definição 3.6, temos, trivialmente, que $E = E_0 \cup E_I$ e $E_0 \cap E_I = \emptyset$. De maneira intuitiva, observamos que o conjunto de distâncias precisas e intervalares do PGD*i* são representadas, respectivamente, por E_0 e E_I .

Como a definição do conjunto de vértices predecessores adjacentes de um vértice do PGD que apresentamos na Definição 3.2 leva em conta somente a existência de arestas no grafo, e a única diferença entre o PGD*i* e o PGD é descrita na natureza do peso das arestas de E , isto é, se é preciso ou intervalar, então, a mesma definição de vértices predecessores adjacentes pode ser aplicada no grafo $G = (V, E, \mathbf{d})$.

Levando em conta as definições do conjunto de predecessores adjacentes de um vértice, dos conjuntos de arestas precisas e intervalares de um grafo $G = (V, E, \mathbf{d})$. O Problema Discretizável de Geometria de Distâncias Intervalares (PDGD*i*) em \mathbb{R}^3 é definido da seguinte maneira:

Considerando que $S_1^i = S(x_{i_1}, d_{i,i_1})$ e $S_2^i = S(x_{i_2}, d_{i,i_2})$ sejam as duas esferas e que $\mathcal{S}_i^3 = S(x_{i_3}, [d_{i,i_3}, \bar{d}_{i,i_3}])$, ..., $\mathcal{S}_{k_i}^i = S(x_{i_{k_i}}, [d_{i,i_{k_i}}, \bar{d}_{i,i_{k_i}}])$ sejam as $k_i - 2$ cascas esféricas que queremos interseccionar, então a solução do Sistema (3.47) é dada por:

$$\begin{aligned} X_i &= S_1^i \cap S_2^i \cap \mathcal{S}_3^i \cap \cdots \cap \mathcal{S}_{k_i}^i \\ &= (S_1^i \cap S_2^i \cap \mathcal{S}_3^i) \cap (S_1^i \cap S_2^i \cap \mathcal{S}_4^i) \cap \cdots \cap (S_1^i \cap S_2^i \cap \mathcal{S}_{k_i}^i) \\ &= \bigcap_{\ell=3}^{k_i} (S_1^i \cap S_2^i \cap \mathcal{S}_\ell^i) \end{aligned} \quad (3.48)$$

Primeiramente, vamos entender qual é o resultado da intersecção de duas esferas com uma casca esférica, digamos $S_1^i \cap S_2^i \cap \mathcal{S}_3^i$, e, então estabelecer como é dado o conjunto x_i , em (3.48). Nos parece intuitivo que a intersecção $S_1^i \cap S_2^i \cap \mathcal{S}_3^i$ seja dada por dois arcos de circunferências simétricos com relação ao plano $\{x_{i_3}, x_{i_2}, x_{i_1}\}$. De fato, este é o caso, mas para justificar este resultado, vamos nos atentar aos seguintes lemas e proposições:

Lema 3.1. A função $\tau^+ : [\sqrt{\nu - 2\mu}, \sqrt{\nu + 2\mu}] \longrightarrow [0^\circ, 180^\circ]$, tal que

$$\tau^+(d) = \arccos\left(\frac{\nu - d^2}{2\mu}\right) \quad (3.49)$$

com $0 < 2\mu \leq \nu$ é contínua e bijetora.

Demonstração. Para mostrar que τ^+ é uma função bijetora, vamos observar que:

- A função $\alpha : [\sqrt{\nu - 2\mu}, \sqrt{\nu + 2\mu}] \rightarrow [-1, 1]$, tal que $\alpha(d) = -\frac{1}{2\mu}d^2 + \frac{\nu}{2\mu}$ é uma função polinomial de segundo grau contínua e bijetora;
- A função $g : [-1, 1] \rightarrow [0^\circ, 180^\circ]$, tal que $g(t) = \arccos t$ é contínua e bijetora.

Escrevendo $\tau^+(d) = g(\alpha(d)) = g \circ \alpha(d)$, observamos que τ^+ é composição das duas funções definidas acima. Logo τ^+ é a composição de duas funções contínuas e bijetoras. Portanto, também é uma função contínua e bijetora. \square

Obs.: Os resultados que utilizamos no Lema 3.1 sobre funções contínuas e bijetoras podem ser consultadas em qualquer bom livro de análise. Recomendamos como referência os livros (DUGUNDJI, 1966) ou (APOSTOL, 1974).

Lema 3.2. A função $\tau^- : [\sqrt{\nu - 2\mu}, \sqrt{\nu + 2\mu}] \longrightarrow [-180^\circ, 0^\circ]$, tal que

$$\tau^-(d) = -\arccos\left(\frac{\nu - d^2}{2\mu}\right) \quad (3.50)$$

com, $0 < 2\mu \leq \nu$ é contínua e bijetora.

Demonstração. A demonstração deste resultado é análoga a do Lema 3.1. \square

Lema 3.3. Sejam $\mathbf{d} = [\underline{d}, \bar{d}] \subset [\sqrt{\nu - 2\mu}, \sqrt{\nu + 2\mu}]$ e $\tau^\pm(d)$ as funções definidas nos lemas anteriores. Então:

$$\begin{aligned}\tau^+(\mathbf{d}) &= \{t \in [0^\circ, 180^\circ] : \underline{\tau} \leq t \leq \bar{\tau}\} && \equiv \mathcal{T}^+, \\ \tau^-(\mathbf{d}) &= \{t \in [-180^\circ, 0^\circ] : -\bar{\tau} \leq t \leq -\underline{\tau}\} && \equiv \mathcal{T}^-, \end{aligned}$$

onde,

$$\underline{\tau} = \arccos\left(\frac{\nu - \underline{d}^2}{2\mu}\right) \quad \text{e} \quad \bar{\tau} = \arccos\left(\frac{\nu - \bar{d}^2}{2\mu}\right).$$

Demonstração. Este resultado segue direto do fato de que τ^\pm são contínuas e bijetoras. \square

Queremos entender como podemos utilizar a expressão de (3.42) para descrever o comportamento de $\tau_{z,u,v,w} \in (-180^\circ, 180^\circ]$ em função da distância $d_{w,z}$. No entanto, dos lemas anteriores observamos duas expressões que definem esta dependência em intervalos angulares complementares, mais diferentes do que necessitamos. Para uma representação completa, que combina os resultados anteriores, nos atentemos ao próximo lema:

Lema 3.4. Sejam $\mathbf{d} = [\underline{d}, \bar{d}] \subset [\sqrt{\nu - 2\mu}, \sqrt{\nu + 2\mu}]$ e a relação

$$\begin{aligned}\tau &: [\sqrt{\nu - 2\mu}, \sqrt{\nu + 2\mu}] \longrightarrow [-180^\circ, 180^\circ] \\ d &\longmapsto \tau(d) = \tau^\pm(d)\end{aligned}$$

onde $\tau^\pm(d)$ são as funções definidas nos lemas anteriores. Então:

$$\tau(\mathbf{d}) = \mathcal{T} = \mathcal{T}^- \cup \mathcal{T}^+.$$

Demonstração. Da definição da relação τ e pelo resultado do Lema 3.3, tem-se:

$$\begin{aligned}\mathcal{T} &= \tau(\mathbf{d}) \\ &= \{t \in (-180^\circ, 180^\circ] : \tau(\underline{d}) \leq t \leq \tau(\bar{d})\} \\ &= \{t \in (-180^\circ, 0^\circ) : \tau(\underline{d}) \leq t \leq \tau(\bar{d})\} \cup \{t \in [0^\circ, 180^\circ] : \tau(\underline{d}) \leq t \leq \tau(\bar{d})\} \\ &= \{t \in (-180^\circ, 0^\circ) : \tau^-(\bar{d}) \leq t \leq \tau^-(\underline{d})\} \cup \{t \in [0^\circ, 180^\circ] : \tau^+(\underline{d}) \leq t \leq \tau^+(\bar{d})\} \\ &= \{t \in (-180^\circ, 0^\circ) : -\bar{\tau} \leq t \leq -\underline{\tau}\} \cup \{t \in [0^\circ, 180^\circ] : \underline{\tau} \leq t \leq \bar{\tau}\} \\ &= \mathcal{T}^- \cup \mathcal{T}^+\end{aligned}$$

\square

Assim, a proposição a seguir evidencia a relação entre o ângulo $\tau_{z,u,v,w}$ e a distância $d_{w,z}$ de (3.43).

Proposição 3.1. Sejam $d_{v,z}$, $d_{v,u}$, $d_{u,z}$, $d_{w,u}$, $d_{w,v}$ distâncias precisas e $\mathbf{d}_{w,z} = [\underline{d}_{w,z}, \bar{d}_{w,z}]$ uma distância intervalar. Tomando $\tau_{z,u,v,w} := \tau$, onde $\nu = \nu(z, u, v, w)$ e $\mu = \mu(z, u, v, w)$, conforme dado em (3.43). Então

$$\tau_{z,u,v,w}(\mathbf{d}_{w,z}) = \mathcal{T}_{z,u,v,w} = \mathcal{T}_{z,u,v,w}^- \cup \mathcal{T}_{z,u,v,w}^+ = [-\bar{\tau}_{z,u,v,w}, -\underline{\tau}_{z,u,v,w}] \cup [\underline{\tau}_{z,u,v,w}, \bar{\tau}_{z,u,v,w}],$$

onde,

$$\underline{\tau}_{z,u,v,w} = \arccos\left(\frac{\nu(z, u, v, w) - \underline{d}_{w,z}^2}{2\mu(z, u, v, w)}\right) \quad \text{e} \quad \bar{\tau}_{z,u,v,w} = \arccos\left(\frac{\nu(z, u, v, w) - \bar{d}_{w,z}^2}{2\mu(z, u, v, w)}\right).$$

Demonstração. Como $0 < 2\mu(z, u, v, w) \leq \nu(z, u, v, w) \forall z, u, v, w \in V$, este resultado segue direto do Lema 3.4. \square

Agora, para observar como descrevemos o ponto x_i em função do intervalo angular da proposição anterior, vamos seguir com os seguintes lemas.

Lema 3.5. A parametrização do círculo no espaço $\mathcal{C} : (-180^\circ, 180^\circ] \rightarrow \mathbb{R}^3$, descrita por $\mathcal{C}(\theta) = \mathbf{a} + \mathbf{b} \cos \theta + \mathbf{c} \sin \theta$, com $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^3$ e $\mathbf{b} \perp \mathbf{c}$ é contínua e injetora.

Demonstração. Como cada uma das coordenadas da função \mathcal{C} é escrita como composição de funções contínuas (soma, multiplicação, cosseno e seno), então \mathcal{C} é uma função contínua.

Para provar que \mathcal{C} é injetora, sejam $\theta_1, \theta_2 \in (-180^\circ, 180^\circ]$, então:

$$\begin{aligned} \mathcal{C}(\theta_1) &= \mathcal{C}(\theta_2) && \Leftrightarrow \\ \mathbf{a} + \mathbf{b} \cos \theta_1 + \mathbf{c} \sin \theta_1 &= \mathbf{a} + \mathbf{b} \cos \theta_2 + \mathbf{c} \sin \theta_2 && \Leftrightarrow \\ (\cos \theta_1 - \cos \theta_2) \mathbf{b} &= -(\sin \theta_1 - \sin \theta_2) \mathbf{c} \end{aligned}$$

Como $\mathbf{b} \perp \mathbf{c}$, então $\cos \theta_1 - \cos \theta_2 = \sin \theta_1 - \sin \theta_2 = 0$. Portanto, $\cos \theta_1 = \cos \theta_2$ e $\sin \theta_1 = \sin \theta_2$. Destas duas igualdades, temos que $\theta_1 = \theta_2 + 360^\circ \cdot k$, com $k \in \mathbb{Z}$. Como $\theta_1, \theta_2 \in (-180^\circ, 180^\circ]$, temos que $k = 0$ é o único valor de k possível. Logo, temos que $\mathcal{C}(\theta_1) = \mathcal{C}(\theta_2) \Rightarrow \theta_1 = \theta_2$. Assim, concluímos que a função \mathcal{C} é injetora. \square

Lema 3.6. Sejam \mathcal{C} a parametrização do lema anterior e $\mathcal{T} = [\underline{\tau}, \bar{\tau}] \subset (-180^\circ, 180^\circ]$. Então $\mathcal{A} \equiv \mathcal{C}(\mathcal{T})$ define um arco de circunferência.

Demonstração. Como \mathcal{C} é injetora, ela associa cada valor de ângulo a um único ponto no espaço. Como \mathcal{C} descreve uma circunferência, temos que $\mathcal{C}(\mathcal{T})$ representa um arco desta circunferência. \square

Teorema 3.2. Seja $f : A \rightarrow B$ uma função. Se $A_\ell \subset A \forall \ell = 1, \dots, N$, então:

1. $f \left(\bigcup_{\ell=1}^N A_\ell \right) = \bigcup_{\ell=1}^N f(A_\ell)$;
2. $f \left(\bigcap_{\ell=1}^N A_\ell \right) \subseteq \bigcap_{\ell=1}^N f(A_\ell)$, com igualdade se f é injetora.

Demonstração. A demonstração deste resultado não apresenta nenhuma dificuldade que devemos salientar e ela pode ser consultada em [DUGUNDJI \(1966\)](#) ou [APOSTOL \(1974\)](#). \square

Lema 3.7. Sejam $d_{v,z}$, $d_{v,u}$, $d_{u,z}$, $d_{w,u}$, $d_{w,v}$ distâncias precisas. Então a função

$${}^{(z,u,v)}x_w : (-180^\circ, 180^\circ] \longrightarrow \mathbb{R}^3,$$

tal que,

$${}^{(z,u,v)}x_w(\tau) = \mathbf{a}(u, v, w) + \mathbf{b}(z, u, v, w) \cos \tau + \mathbf{c}(z, u, v, w) \sin \tau,$$

com, $\mathbf{a}(u, v, w)$, $\mathbf{b}(z, u, v, w)$ e $\mathbf{c}(z, u, v, w)$ dados em (3.41), é contínua e injetora.

Demonstração. Como $\mathbf{b}(z, u, v, w) \perp \mathbf{c}(z, u, v, w) \forall z, u, v, w \in V$, este resultado segue direto do Lema 3.5. \square

Assim, com os resultados apresentados, já temos o ferramental matemático necessário para demonstrar a proposição seguinte:

Proposição 3.2. Sejam $S_v^w = S(x_v, d_{w,v})$ e $S_u^w = S(x_u, d_{w,u})$ duas esferas e a casca esférica $\mathcal{S}_w^z = S(x_z, \mathbf{d}_{w,z})$, com S_v^w , S_u^w , $\mathcal{S}_w^z \subset \mathbb{R}^3$. Então:

$$S_v^w \cap S_u^w \cap \mathcal{S}_w^z = {}^{(z,u,v)}x_w(\tau_{z,u,v,w}(\mathbf{d}_{w,z})) = {}^{(z,u,v)}\mathcal{A}_w^- \cup {}^{(z,u,v)}\mathcal{A}_w^+$$

onde ${}^{(z,u,v)}x_w$ é a função definida no Lema 3.7, $\tau_{z,u,v,w}$ é definido na Proposição 3.1 e ${}^{(z,u,v)}\mathcal{A}_w^\pm$ são dois arcos da mesma circunferência, simétricos com relação ao plano $\{x_z, x_u, x_v\}$.

Demonstração. Como já salientamos anteriormente, a intersecção $S_v^w \cap S_u^w \cap \mathcal{S}_w^z$ é um problema geométrico cujo equivalente algébrico é encontrar a solução do sistema não linear na variável x_w :

$$\begin{cases} d_{w,z}^2 \leq \|x_w - x_z\|^2 \leq d_{w,z}^2, \\ \|x_w - x_u\|^2 = d_{w,u}^2, \\ \|x_w - x_v\|^2 = d_{w,v}^2. \end{cases} \quad (3.51)$$

Deste sistema, temos que as distâncias precisas $d_{v,z}$, $d_{v,u}$, $d_{u,z}$, $d_{w,u}$, $d_{w,v}$ e a distância intervalar $\mathbf{d}_{w,z} = [\underline{d}_{w,z}, \bar{d}_{w,z}]$ são conhecidas (as distâncias precisas $d_{v,z}$, $d_{v,u}$, $d_{u,z}$ não estão explícitas no Sistema (3.51), mas como os pontos x_z , x_u e x_v são conhecidos, essas distâncias podem ser calculadas trivialmente). Assim, dado o conjunto de distâncias precisas, a distância intervalar e os pontos x_z , x_u e x_v , temos todos os parâmetros para descrever ${}^{(z,u,v)}x_w(\tau_{z,u,v,w}(\mathbf{d}_{w,z}))$ conforme a Proposição 3.1 e o Lema 3.7.

Podemos verificar sem muitos problemas que ${}^{(z,u,v)}x_w(\tau_{z,u,v,w}(\mathbf{d}_{w,z}))$ é uma solução do Sistema (3.51), no entanto, devido a toda discussão que fizemos na Subseção 3.2.3, vamos omitir estas contas e assumir este resultado.

Agora, da Proposição 3.1, temos que

$$\tau_{z,u,v,w}(\mathbf{d}_{w,z}) = \mathcal{T}_{z,u,v,w} = \mathcal{T}_{z,u,v,w}^- \cup \mathcal{T}_{z,u,v,w}^+$$

onde $\mathcal{T}_{z,u,v,w}^-$ e $\mathcal{T}_{z,u,v,w}^+$ são dois intervalos angulares. Utilizando este resultado na solução de (3.51), temos que:

$${}^{(z,u,v)}x_w(\tau_{z,u,v,w}(\mathbf{d}_{w,z})) = {}^{(z,u,v)}x_w(\mathcal{T}_{z,u,v,w}) = {}^{(z,u,v)}x_w(\mathcal{T}_{z,u,v,w}^- \cup \mathcal{T}_{z,u,v,w}^+).$$

Do item 1 do Teorema 3.2, temos que:

$${}^{(z,u,v)}x_w(\mathcal{T}_{z,u,v,w}^- \cup \mathcal{T}_{z,u,v,w}^+) = {}^{(z,u,v)}x_w(\mathcal{T}_{z,u,v,w}^-) \cup {}^{(z,u,v)}x_w(\mathcal{T}_{z,u,v,w}^+) = {}^{(z,u,v)}\mathcal{A}_w^- \cup {}^{(z,u,v)}\mathcal{A}_w^+,$$

onde ${}^{(z,u,v)}\mathcal{A}_w^\pm$ representam dois arcos de circunferência (resultado que segue direto do Lema 3.6).

Para finalizar, como ${}^{(z,u,v)}x_w$ descreve uma circunferência contida em um plano ortogonal ao plano $\{x_z, x_u, x_v\}$, ${}^{(z,u,v)}x_w(0^\circ) \in \{x_z, x_u, x_v\}$ e os intervalos angulares $\mathcal{T}_{z,u,v,w}^\pm$ são simétricos com relação a 0° , temos que os arcos ${}^{(z,u,v)}\mathcal{A}_w^\pm$ são simétricos com relação ao plano $\{x_z, x_u, x_v\}$. \square

Portanto, esta proposição justifica nossa intuição inicial sobre o resultado da intersecção de duas esferas e uma casca esférica. Desta maneira, aplicando o resultado desta proposição na expressão para X_i em (3.48), temos:

$$X_i = \bigcap_{\ell=3}^{k_i} (S_1^i \cap S_2^i \cap S_\ell^i) = \bigcap_{\ell=3}^{k_i} ({}^{(i_1, i_2, i_\ell)}\mathcal{A}_i^+ \cup {}^{(i_1, i_2, i_\ell)}\mathcal{A}_i^-). \quad (3.52)$$

Assim, concluímos que a solução do Sistema (3.47) é dado pela intersecção de $k_i - 2$ arcos de circunferência. Para entender melhor como podemos proceder para realizar a intersecção destes arcos de circunferência, vamos nos atentar, novamente, a sequência de resultados a seguir.

Lema 3.8. Seja a operação binária $\oplus : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$, com

$$\oplus(a, b) = a \oplus b = a + b + \gamma_{[a,b]}, \text{ onde } \gamma_{[a,b]} = \begin{cases} -2\pi & \text{se } a + b > \pi \\ 0 & \text{se } |a + b| \leq \pi \\ 2\pi & \text{se } a + b < -\pi \end{cases} \quad (3.53)$$

O subconjunto $(-\pi, \pi] \subset \mathbb{R}$ é fechado com relação a operação \oplus .

Demonstração. Queremos provar que $\forall (a, b) \in (-\pi, \pi] \times (-\pi, \pi]$, temos que $a \oplus b \in (-\pi, \pi]$. Para isso, notemos que, se $a, b \in (-\pi, \pi]$, então

$$\begin{cases} -\pi < a \leq \pi \\ -\pi < b \leq \pi \end{cases} \Rightarrow -2\pi < a + b \leq 2\pi \Leftrightarrow \begin{cases} -2\pi < a + b < -\pi \\ -\pi \leq a + b \leq \pi \\ \pi < a + b < 2\pi \end{cases}$$

Assim, vamos observar que:

$$(i) \ a + b \in (-2\pi, -\pi) \Rightarrow a \oplus b = \begin{cases} a + b + 2\pi < (-2\pi) + 2\pi = 0, \\ a + b + 2\pi > (-\pi) + 2\pi = \pi. \end{cases}$$

Logo, $0 < a \oplus b < \pi$;

$$(ii) \ a + b \in [-\pi, \pi] \Rightarrow a \oplus b = \begin{cases} a + b \leq \pi, \\ a + b \geq -\pi. \end{cases}$$

Logo, $-\pi \leq a \oplus b \leq \pi$;

$$(iii) \ a + b \in (\pi, 2\pi) \Rightarrow a \oplus b = \begin{cases} a + b - 2\pi < (2\pi) - 2\pi = 0, \\ a + b - 2\pi > (\pi) - 2\pi = -\pi. \end{cases}$$

Logo, $-\pi < a \oplus b < 0$.

Dos três itens anteriores, temos que $\forall (a, b) \in (-\pi, \pi] \times (-\pi, \pi] \Rightarrow a \oplus b \in (-\pi, \pi]$. \square

Proposição 3.3. Sejam d_{v,z_0} , d_{u,z_0} , d_{v,z_1} , d_{u,z_1} , $d_{v,u}$, $d_{w,u}$, $d_{w,v}$ distâncias precisas e ${}^{(z,u,v)}x_w$ a função definida no Lema 3.7. então

$${}^{(z_1,u,v)}x_w(\tau) = {}^{(z_0,u,v)}x_w(\varphi_{z_0,u,v,z_1} \oplus \tau),$$

onde \oplus é a operação definida no Lema 3.8 e φ_{z_0,u,v,z_1} é o ângulo de torção entre os planos $\{x_{z_0}, x_u, x_v\}$ e $\{x_{z_1}, x_u, x_v\}$.

Demonstração. Como vamos considerar duas variações da função ${}^{(z,u,v)}x_w$ e seus valores constantes envolvem muitas variáveis em comum, para simplificar a notação, vamos considerar:

$$\varphi_{z_0,u,v,z_1} = \varphi_{0,1}, \ \mathbf{a}(u, v, w) = \mathbf{a}, \ \mathbf{b}(z_\ell, u, v, w) = \mathbf{b}_\ell, \ \mathbf{c}(z_\ell, u, v, w) = \mathbf{c}_\ell, \text{ com } \ell = 0, 1.$$

Inicialmente, vamos observar que, se $\theta_1, \theta_2 \in (-180^\circ, 180^\circ]$, então:

$$\begin{aligned}\cos(\theta_1 \oplus \theta_2) &= \cos(\theta_1 + \theta_2 + \gamma_{[a,b]}) = \cos(\theta_1 + \theta_2) \cos(\gamma_{[a,b]}) - \sin(\theta_1 + \theta_2) \sin(\gamma_{[a,b]}), \\ \sin(\theta_1 \oplus \theta_2) &= \sin(\theta_1 + \theta_2 + \gamma_{[a,b]}) = \sin(\theta_1 + \theta_2) \cos(\gamma_{[a,b]}) + \sin(\gamma_{[a,b]}) \cos(\theta_1 + \theta_2).\end{aligned}$$

Como a quantidade $\gamma_{[\theta_1, \theta_2]}$ é um múltiplo de 360° , temos que $\cos(\gamma_{[a,b]}) = 1$ e $\sin(\gamma_{[a,b]}) = 0$. Logo, das expressões acima, concluímos que $\cos(\theta_1 \oplus \theta_2) = \cos(\theta_1 + \theta_2)$ e $\sin(\theta_1 \oplus \theta_2) = \sin(\theta_1 + \theta_2)$. Desta maneira:

$$\begin{aligned}{}^{(z_0, u, v)}x_w(\varphi_{0,1} \oplus \tau) &= \mathbf{a} + \mathbf{b}_0 \cos(\tau \oplus \varphi_{0,1}) + \mathbf{c}_0 \sin(\tau \oplus \varphi_{0,1}) \\ &= \mathbf{a} + \mathbf{b}_0 \cos(\tau + \varphi_{0,1}) + \mathbf{c}_0 \sin(\tau + \varphi_{0,1}) \\ &= \mathbf{a} + (\mathbf{b}_0 \cos \varphi_{0,1} + \mathbf{c}_0 \sin \varphi_{0,1}) \cos \tau + \\ &\quad (\mathbf{c}_0 \cos \varphi_{0,1} - \mathbf{b}_0 \sin \varphi_{0,1}) \sin \tau\end{aligned}$$

Como ${}^{(z_1, u, v)}x_w(\tau) = \mathbf{a} + \mathbf{b}_1 \cos \tau + \mathbf{c}_1 \sin \tau$, esta proposição ficará provada se verificarmos que:

$$\begin{cases} \mathbf{b}_1 &= \mathbf{b}_0 \cos \varphi_{0,1} + \mathbf{c}_0 \sin \varphi_{0,1}, \\ \mathbf{c}_1 &= \mathbf{c}_0 \cos \varphi_{0,1} - \mathbf{b}_0 \sin \varphi_{0,1}. \end{cases} \quad (3.54)$$

Pelas expressões de $\mathbf{b}(z, u, v, w)$ e $\mathbf{c}(z, u, v, w)$ apresentadas em (3.41), observamos que:

$$1. \mathbf{b}_\ell \cdot \mathbf{b}_\ell = \mathbf{c}_\ell \cdot \mathbf{c}_\ell = \rho^2; \quad 2. \mathbf{c}_\ell = \hat{r} \times \mathbf{b}_\ell; \quad 3. \hat{r} \perp \mathbf{b}_\ell;$$

com $\ell = 0, 1$.

Como os planos $\pi_{z_0} = \{x_{z_0}, x_u, x_v\}$ e $\pi_{z_1} = \{x_{z_1}, x_u, x_v\}$ podem ser parametrizados, respectivamente, por $\pi_{z_0}(t, s) = x_u + \hat{r}t + \mathbf{b}_0s$ e $\pi_{z_1}(t, s) = x_u + \hat{r}t + \mathbf{b}_1s$, com $t, s \in \mathbb{R}$, temos que, $\varphi_{0,1}$ é o vetor formado pelos vetores \mathbf{c}_0 e \mathbf{c}_1 . Para descrever o ângulo $\varphi_{0,1}$ em função dos vetores \mathbf{b}_ℓ e \mathbf{c}_ℓ , com $\ell = 0, 1$, vamos nos atentar para as propriedades dos produtos vetorial e interno para vetores em \mathbb{R}^3 , que podem ser consultadas em [THORNTON; MARION \(2004\)](#):

$$\vec{v}_1 \times \vec{v}_2 = -\vec{v}_2 \times \vec{v}_1 \quad (\text{P1})$$

$$\vec{v}_1 \cdot (\vec{v}_2 \times \vec{v}_3) = \vec{v}_3 \cdot (\vec{v}_1 \times \vec{v}_2) = \vec{v}_3 \cdot (\vec{v}_1 \times \vec{v}_2) \quad (\text{P2})$$

$$(\vec{v}_1 \times \vec{v}_2) \times (\vec{v}_3 \times \vec{v}_4) = [(\vec{v}_1 \times \vec{v}_2) \cdot \vec{v}_4] \vec{v}_3 - [(\vec{v}_1 \times \vec{v}_2) \cdot \vec{v}_3] \vec{v}_4 \quad (\text{P3})$$

$$(\vec{v}_1 \times \vec{v}_2) \cdot (\vec{v}_3 \times \vec{v}_4) = (\vec{v}_1 \cdot \vec{v}_3)(\vec{v}_2 \cdot \vec{v}_4) - (\vec{v}_1 \cdot \vec{v}_4)(\vec{v}_2 \cdot \vec{v}_3) \quad (\text{P4})$$

$$\vec{v}_1 \cdot \vec{v}_1 = \vec{v}_1 \cdot \vec{v}_2 \Leftrightarrow \vec{v}_1 = \vec{v}_2 \quad (\text{P5})$$

Desta maneira, temos que:

$$\begin{aligned}
 \cos \varphi_{0,1} &= \frac{\mathbf{c}_0 \cdot \mathbf{c}_1}{\|\mathbf{c}_0\| \|\mathbf{c}_1\|} = \frac{(\hat{r} \times \mathbf{b}_0) \cdot (\hat{r} \times \mathbf{b}_1)}{\rho^2} \\
 &\stackrel{(P4)}{=} \frac{1}{\rho^2} (\hat{r} \cdot \hat{r}) (\mathbf{b}_0 \cdot \mathbf{b}_1) - \underbrace{(\hat{r} \cdot \mathbf{b}_1)}_{=0} \underbrace{(\mathbf{b}_0 \cdot \hat{r})}_{=0} \\
 &= \frac{1}{\rho^2} \mathbf{b}_0 \cdot \mathbf{b}_1
 \end{aligned} \tag{3.55}$$

$$\begin{aligned}
 \sin \varphi_{0,1} &= \frac{\|\mathbf{c}_0 \times \mathbf{c}_1\|}{\|\mathbf{c}_0\| \|\mathbf{c}_1\|} = \frac{\|(\hat{r} \times \mathbf{b}_0) \times (\hat{r} \times \mathbf{b}_1)\|}{\rho^2} \\
 &\stackrel{(P3)}{=} \frac{1}{\rho^2} \|((\hat{r} \times \mathbf{b}_0) \cdot \mathbf{b}_1) \hat{r} - ((\hat{r} \times \mathbf{b}_0) \cdot \hat{r}) \mathbf{b}_1\| \\
 &\stackrel{(P2)}{=} \frac{1}{\rho^2} \|((\hat{r} \times \mathbf{b}_0) \cdot \mathbf{b}_1) \hat{r} - \underbrace{((\hat{r} \times \hat{r}) \cdot \mathbf{b}_0)}_{=0} \mathbf{b}_1\| \\
 &= \frac{1}{\rho^2} (\hat{r} \times \mathbf{b}_0) \cdot \mathbf{b}_1 \stackrel{(P2)}{=} \frac{1}{\rho^2} (\mathbf{b}_1 \times \hat{r}) \cdot \mathbf{b}_0 \stackrel{(P1)}{=} -\frac{1}{\rho^2} (\hat{r} \times \mathbf{b}_1) \cdot \mathbf{b}_0
 \end{aligned} \tag{3.56}$$

Logo, $\mathbf{c}_0 \cdot \mathbf{c}_1 = \mathbf{b}_0 \cdot \mathbf{b}_1 = \rho^2 \cos \varphi_{0,1}$ e $\mathbf{c}_0 \cdot \mathbf{b}_1 = -\mathbf{c}_1 \cdot \mathbf{b}_0 = \rho^2 \sin \varphi_{0,1}$.

Pelas expressões dadas anteriormente, vamos observar que:

$$\begin{aligned}
 \mathbf{b}_1 \cdot (\mathbf{b}_0 \cos \varphi_{0,1} + \mathbf{c}_0 \sin \varphi_{0,1}) &= \mathbf{b}_1 \cdot \mathbf{b}_0 \cos \varphi_{0,1} + \mathbf{b}_1 \cdot \mathbf{c}_0 \sin \varphi_{0,1} \\
 &= \rho^2 \cos \varphi_{0,1} + \rho^2 \sin \varphi_{0,1} = \rho^2 \\
 &= \mathbf{b}_1 \cdot \mathbf{b}_1 \\
 \mathbf{c}_1 \cdot (\mathbf{c}_0 \cos \varphi_{0,1} - \mathbf{b}_0 \sin \varphi_{0,1}) &= \mathbf{c}_1 \cdot \mathbf{c}_0 \cos \varphi_{0,1} - \mathbf{c}_1 \cdot \mathbf{b}_0 \sin \varphi_{0,1} \\
 &= \rho^2 \cos \varphi_{0,1} + \rho^2 \sin \varphi_{0,1} = \rho^2 \\
 &= \mathbf{c}_1 \cdot \mathbf{c}_1
 \end{aligned}$$

Portanto, aplicando (P5) nas igualdades obtidas acima, temos que as equações de (3.54) são verdadeiras. \square

Desta proposição, podemos escrever, de maneira direta, o seguinte corolário:

Corolário 3.1. Sejam ${}^{(z,u,v)}x_w$, \oplus e φ_{z_0,u,v,z_1} , respectivamente, a função, a operação binária e o ângulo de torção utilizados na proposição anterior. Se $\mathcal{T} = [\underline{\tau}, \bar{\tau}] \subset (-180^\circ, 180^\circ]$, então

$${}^{(z_1,u,v)}x_w(\mathcal{T}) = {}^{(z_0,u,v)}x_w(\varphi_{z_0,u,v,z_1} \oplus \mathcal{T}),$$

onde $\varphi \oplus \mathcal{T} \equiv \{\theta \in (-180^\circ, 180^\circ] \mid \min\{\varphi \oplus \underline{\tau}, \varphi \oplus \bar{\tau}\} \leq \theta \leq \max\{\varphi \oplus \underline{\tau}, \varphi \oplus \bar{\tau}\}\}$.

Desta maneira, levando em conta todos os resultados que descrevemos nesta seção, podemos escrever a seguinte proposição, que trata da solução do sistema não linear (3.47):

Proposição 3.4. A solução do sistema não linear (3.47), cujo problema geométrico equivalente é interseccionar as esferas $S_1^i = S(x_{i_1}, d_{i_1, i_1})$, $S_2^i = S(x_{i_2}, d_{i_2, i_2})$ e as cascas esféricas $S_3^i = S(x_{i_3}, \mathbf{d}_{i_3, i_3})$, \dots , $S_{k_i}^i = S(x_{i_{k_i}}, \mathbf{d}_{i_3, i_{k_i}})$, é dada por

$${}^{(i_3, i_2, i_1)}x_i(\mathcal{T}_i),$$

com:

$$\mathcal{T}_i = \bigcap_{\ell=3}^{k_i} \varphi_{i_3, i_2, i_1, i_\ell} \oplus \mathcal{T}_{i_\ell, i_2, i_1, i} = \bigcap_{\ell=3}^{k_i} \left(\varphi_{i_3, i_2, i_1, i_\ell} \oplus \mathcal{T}_{i_\ell, i_2, i_1, i}^- \cup \varphi_{i_3, i_2, i_1, i_\ell} \oplus \mathcal{T}_{i_\ell, i_2, i_1, i}^+ \right), \quad (3.57)$$

onde,

- ${}^{(u, v, w)}x_w$ é a função definida no Lema 3.7;
- $\mathcal{T}_{z, u, v, w} = \mathcal{T}_{z, u, v, w}^- \cup \mathcal{T}_{z, u, v, w}^+ = \tau_{z, u, v, w}(\mathbf{d}_{w, z})$, sendo $\tau_{z, u, v, w}$ a relação descrita no Lema 3.4;
- $\varphi_{z, u, v, w}$ é o ângulo de torção entre os planos $\{x_u, x_v, x_w\}$ e $\{x_u, x_v, x_z\}$;
- \oplus é a operação binária definida no Lema 3.8.

Demonstração. Para construir a demonstração deste resultado, vamos denotar:

- A Proposição 3.2 por (R1);
- O Corolário 3.1 por (R3);
- O item 1 do Teorema 3.2 por (R2);
- O item 2 do Teorema 3.2 por (R4).

Já vimos que a solução de (3.47) é dada por $X_i = \bigcap_{\ell=3}^{k_i} (S_1^i \cap S_2^i \cap S_\ell^i)$. Assim:

$$\begin{aligned} X_i &= \bigcap_{\ell=3}^{k_i} (S_1^i \cap S_2^i \cap S_\ell^i) \stackrel{(R1)}{=} \bigcap_{\ell=3}^{k_i} \left({}^{(i_\ell, i_2, i_1)}\mathcal{A}_i^+ \cup {}^{(i_\ell, i_2, i_1)}\mathcal{A}_i^- \right) \\ &= \bigcap_{\ell=3}^{k_i} \left({}^{(i_\ell, i_2, i_1)}x_i(\mathcal{T}_{i_\ell, i_2, i_1, i}^-) \cup {}^{(i_\ell, i_2, i_1)}x_i(\mathcal{T}_{i_\ell, i_2, i_1, i}^-) \right) \\ &\stackrel{(R2)}{=} \bigcap_{\ell=3}^{k_i} \left({}^{(i_\ell, i_2, i_1)}x_i(\mathcal{T}_{i_\ell, i_2, i_1, i}^- \cup \mathcal{T}_{i_\ell, i_2, i_1, i}^+) \right) = \bigcap_{\ell=3}^{k_i} \left({}^{(i_\ell, i_2, i_1)}x_i(\mathcal{T}_{i_\ell, i_2, i_1, i}) \right) \\ &\stackrel{(R3)}{=} \bigcap_{\ell=3}^{k_i} \left({}^{(i_3, i_2, i_1)}x_i(\varphi_{i_3, i_2, i_1, i_\ell} \oplus \mathcal{T}_{i_\ell, i_2, i_1, i}) \right) \\ &\stackrel{(R4)}{=} {}^{(i_3, i_2, i_1)}x_i \left(\bigcap_{\ell=3}^{k_i} \varphi_{i_3, i_2, i_1, i_\ell} \oplus \mathcal{T}_{i_\ell, i_2, i_1, i} \right) \end{aligned}$$

Logo, $X_i = (i_3, i_2, i_1) x_i (\mathcal{T}_i)$, onde:

$$\mathcal{T}_i = \bigcap_{\ell=3}^{k_i} \varphi_{i_3, i_2, i_1, i_\ell} \oplus \mathcal{T}_{i_\ell, i_2, i_1, i} = \bigcap_{\ell=3}^{k_i} \left(\varphi_{i_3, i_2, i_1, i_\ell} \oplus \mathcal{T}_{i_\ell, i_2, i_1, i}^- \cup \varphi_{i_3, i_2, i_1, i_\ell} \oplus \mathcal{T}_{i_\ell, i_2, i_1, i}^+ \right).$$

□

Agora, antes de observar as implicações da solução de (3.47), que é descrita pela proposição acima, vamos ressaltar os seguintes pontos sobre o cálculo de \mathcal{T}_i :

1. Para obter os ângulos de torção $\underline{\tau}_{i_\ell, i_2, i_1, i}$, $\bar{\tau}_{i_\ell, i_2, i_1, i}$ e $\varphi_{i_3, i_2, i_1, i_\ell}$, com $\ell = 3, \dots, k_i$, é necessário que conheçamos as distâncias precisas d_{i_1, i_ℓ} , d_{i_2, i_ℓ} e d_{i_1, i_2} . No entanto, como $v_{i_\ell}, v_{i_2}, v_{i_1} \in U_i$, temos que as posições x_{i_ℓ} , x_{i_2} e x_{i_1} são conhecidas. Logo, estes valores de distâncias podem ser calculados de maneira trivial.
2. Podemos interpretar o ângulo de torção $\varphi_{i_3, i_2, i_1, i_\ell}$ como uma mudança do referencial $\{x_{i_\ell}, x_{i_2}, x_{i_1}\}$ para o referencial $\{x_{i_3}, x_{i_2}, x_{i_1}\}$. Deste modo, a orientação deste ângulo tem um papel fundamental nesta mudança. Podemos utilizar a expressão do cosseno do ângulo de torção apresentada em (3.42) para calcular o valor absoluto de $\varphi_{i_3, i_2, i_1, i_\ell} \forall \ell = 3, \dots, k_i$. Na próxima proposição, apresentamos uma maneira de obter a orientação destes ângulos.

Proposição 3.5. Dados $x_z, x_u, x_v, x_w \in \mathbb{R}^3$ não colineares. O sinal do ângulo de torção $\varphi_{z,u,v,w} = \{x_z, x_u, x_v, x_w\}$ é dado por:

$$\text{sign}(\varphi_{z,u,v,w}) = \text{sign}([(x_v - x_u) \times (x_z - x_u)] \cdot (x_w - x_u)).$$

Demonstração. Para provar este resultado, vamos observar que a orientação do ângulo $\varphi_{z,u,v,w}$ é dada com base no plano $\pi = \{x_z, x_u, x_v\}$ e no ponto x_w . Como o plano π divide o espaço em dois semiespaços, o sinal de $\varphi_{z,u,v,w}$ é dado conforme o semiespaço que o ponto x_w está localizado. Para realizar os cálculos, vamos considerar que $\vec{v} = x_z - x_u$, $\vec{r} = x_v - x_u$ e x_w^0 é a projeção ortogonal do ponto x_w no plano π , que, matematicamente, é descrita por:

$$x_w^0 = \text{proj}_\pi(x_w) = x_u + \hat{s} \cdot (x_w - x_u) \hat{s} + \hat{r} \cdot (x_w - x_u) \hat{r},$$

onde $\vec{s} = \hat{v} - (\hat{v} \cdot \hat{r}) \hat{r}$ e \hat{s} é o vetor de π que é ortonormal a \hat{r} .

Obs.: A notação \hat{x} representa o vetor unitário na direção de \vec{x} .

Desta maneira, $\vec{w} = x_w - x_w^0$ é um vetor normal ao plano π . No entanto, $\hat{r} \times \hat{s}$ é o vetor normal unitário que dá a orientação do plano π . A Figura 22 evidencia a situação entre os pontos x_z, x_u, x_v, x_w e x_w^0 geometricamente. Assim, para verificar em qual semiespaço o ponto x_w pertence, podemos calcular $(\hat{r} \times \hat{s}) \cdot \hat{w}$. Geometricamente, esta quantidade nos

diz se os vetores $\hat{r} \times \hat{s}$ e \hat{w} estão na mesma direção ou em direções opostas. Como $\hat{r} \perp \hat{s}$, temos que $\hat{r} \times \hat{s}$ é um vetor unitário. Logo:

$$(\hat{r} \times \hat{s}) \cdot \hat{w} = \begin{cases} 1 & \text{se } \hat{r} \times \hat{s} \text{ e } \hat{w} \text{ estão na mesma direção,} \\ -1 & \text{se } \hat{r} \times \hat{s} \text{ e } \hat{w} \text{ estão em direções opostas.} \end{cases}$$

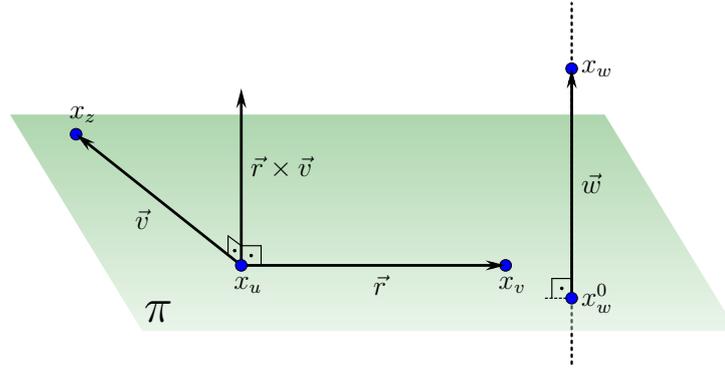


Figura 22 – Descrição geométrica de vetores que evidenciam a projeção ortogonal do ponto x_w no plano $\pi = \{x_z, x_u, x_v\}$.

Assim, $(\hat{r} \times \hat{s}) \cdot \hat{w}$ representa o sinal do ângulo de torção $\varphi_{z,u,v,w}$. Para observar o resultado desta proposição, notemos que:

$$\begin{aligned} \text{sign}(\varphi_{z,u,v,w}) &= (\hat{r} \times \hat{s}) \cdot \hat{w} = \hat{r} \times \hat{s} \cdot \frac{x_w - x_u + \hat{s} \cdot (x_w - x_u)\hat{s} + \hat{r} \cdot (x_w - x_u)\hat{r}}{\|\vec{w}\|} \\ &= \frac{1}{\|\vec{r}\|\|\vec{s}\|\|\vec{w}\|} (\vec{r} \times \vec{s}) \cdot (x_w - x_u) = \frac{1}{\|\vec{r}\|\|\vec{s}\|\|\vec{w}\|} (\vec{r} \times \vec{v}) \cdot (x_w - x_u) \\ &= \frac{1}{\|\vec{r}\|\|\vec{s}\|\|\vec{w}\|} [(x_v - x_u) \times (x_z - x_u)] \cdot (x_w - x_u) \\ &= \text{sign}([(x_v - x_u) \times (x_z - x_u)] \cdot (x_w - x_u)) \end{aligned}$$

□

A sequência de resultados que apresentamos nesta seção nos levaram a uma maneira de obter a solução X_i do sistema não linear (3.47). Pela maneira que esta solução foi construída, notamos que esta solução representa arcos de uma circunferência no \mathbb{R}^3 , e que estes arcos são dados com base em um intervalo angular \mathcal{T}_i que parametriza esta circunferência. Pela expressão de \mathcal{T}_i apresentada na (3.57), notamos facilmente que se as distâncias intervalares são, na verdade, distâncias precisas, \mathcal{T}_i pode ser um conjunto vazio, dado por um ou por dois pontos. Com isso, recuperamos a situação discutida na Subseção 3.2.1, assim, $|X_i| \in \{0, 1, 2\}$.

De uma maneira mais geral, podemos dizer o seguinte sobre a solução X_i de (3.47):

1. $X_i = \emptyset$ se $\mathcal{T}_i = \emptyset$;

2. X_i é um conjunto discreto de pontos se \mathcal{T}_i é um conjunto discreto;
3. X_i é um conjunto não enumerável de pontos se \mathcal{T}_i é um conjunto não enumerável.

Com base nas distâncias intervalares \mathbf{d}_{i,i_ℓ} , com $\ell = 3, \dots, k_i$, pela expressão de \mathcal{T}_i (3.57) e os três casos mencionados acima, podemos dizer que:

- X_i é vazio ou um conjunto discreto de pontos se existe $3 \leq l \leq k_i$ tal que $\mathbf{d}_{i,i_l} = d_{i,i_l}$;
- X_i é dado pela união de dois arcos de circunferência simétricos com relação ao plano $\{x_{i_3}, x_{i_2}, x_{i_1}\}$ se $v_i \in \mathcal{B}$. Caso $\mathbf{d}_{i,i_3} = d_{i,i_3}$, então cada arco é dado por um único ponto.

Portanto, observada a estrutura do conjunto X_i , salientamos que o algoritmo BP não pode ser utilizado para resolver o PDGD i , pois o BP não é descrito para lidar com a situação em que X_i é um conjunto não enumerável de pontos. No entanto, o algoritmo *interval Branch & Prune* (*iBP*), considerado como uma extensão do BP, é proposto em LAVOR; LIBERTI; MUCHERINO (2013) para lidar com o impasse enfrentado pelo BP. Em linhas gerais, o *iBP* possui a mesma estrutura do BP, mas realiza uma amostragem no conjunto X_i se o mesmo é dado por um conjunto não enumerável de pontos. É intuitivo que este processo de amostragem pode selecionar um conjunto de pontos que resulte somente em ramos infactíveis na árvore de busca. De fato, os trabalhos GONÇALVES et al. (2017) e D'AMBROSIO et al. (2017) mostram que uma amostra grande aumenta exponencialmente o espaço de busca e uma amostra pequena pode resultar na não obtenção de uma solução do PDGD i .

O trabalho LAVOR et al. (2021b) mostra que uma estratégia de redução de arcos aprimora o processo de amostragem do *iBP* aplicado ao PDGDM com distâncias intervalares em \mathbb{R}^3 . A maneira que construímos o conjunto X_i leva em conta esta estratégia de redução de arcos, que foi incorporada no cálculo do intervalo angular \mathcal{T}_i em (3.57). Desta maneira, propomos o pseudo-código do Algoritmo 2 como a estrutura geral do *iBP* aplicado ao PDGD i em \mathbb{R}^3 .

Depois do processo de amostragem, o *iBP* utiliza o mesmo mecanismo que o BP usa para explorar o espaço de busca. No entanto, o número de caminhos possíveis do espaço de busca do *iBP* é muito maior que o do BP, como dissemos anteriormente, ele aumenta exponencialmente com o número de pontos tomados em cada amostragem. Assim, a discussão que realizamos sobre o BP na Subseção 3.2.2 pode ser estendida, tomando as precauções necessárias, para o *iBP*.

Nesta seção, apresentamos aspecto teórico do PDGD i e um algoritmo para encontrar uma solução deste problema. No próximo capítulo, observamos aspectos geométricos das proteínas que utilizamos para modelar o problema de interesse.

Algoritmo 2: Estrutura principal do algoritmo i BP em \mathbb{R}^3 .

Entrada: $G = (V, E, \mathbf{d})$, $X = \{x_1, x_2, x_3\}$, $i = 4$, $b_v = 1$ e $N_v \geq 1 \forall v \in V$
Saída: $X = \{x_v \in \mathbb{R}^3 \mid v \in V\}$

1 **Função** i BP(G, X, i, b_i, N_i)
2 **se** ($b_i = 1$) **então**
3 **compute** \mathcal{T}_i como foi definido em (3.57);
4 **se** ($\mathcal{T}_i \neq \emptyset$) **então**
5 **considere** $T_i = \{\tau_1, \dots, \tau_{N_i}\}$ uma amostra do intervalo \mathcal{T}_i com N_i
6 pontos;
7 **compute** as quantidades $\mathbf{a}_i(i_2, i_1, i)$, $\mathbf{b}_i(i_3, i_2, i_1, i)$, $\mathbf{c}_i(i_3, i_2, i_1, i)$
8 definidas em (3.41);
9 **senão**
10 $T_i \leftarrow \emptyset$;
11 **fim**
12 **se** ($T_i \neq \emptyset$) **então**
13 $X \leftarrow X \cup \{\mathbf{a}_i(i_2, i_1, i) + \mathbf{b}_i(i_3, i_2, i_1, i) \cos \tau_{b_i} + \mathbf{c}_i(i_3, i_2, i_1, i) \sin \tau_{b_i}\}$;
14 **se** ($i = |V|$) **então**
15 **retorne** X ;
16 **senão**
17 i BP($G, X, i + 1, b_{i+1}, N_{i+1}$);
18 **fim**
19 **senão**
20 **retroceda para** para o primeiro $v_j < v_i$ tal que $b_j < N_j$;
21 $b_j \leftarrow b_j + 1$;
22 $X \leftarrow X \setminus \{x_j, x_{j+1}, \dots, x_{i-1}\}$;
23 $b_{j+1}, b_{j+2}, \dots, b_{i-1} \leftarrow 1$;
24 i BP(G, X, j, b_j, N_j);
25 **fim**

CAPÍTULO

4 Estrutura Tridimensional de Proteínas e Restrições de Distâncias

Neste trabalho, estamos interessados na determinação da posição dos átomos da cadeia principal das proteínas, cujo nome em inglês é *backbone*. Dado que esta nomenclatura é bastante difundida na literatura (NELSON; COX, 2012), optamos por utilizar esta denominação em língua estrangeira para referenciar esta sequência de átomos nas proteínas. Neste capítulo, apresentamos as hipóteses que consideramos para construir um conjunto de distâncias entre alguns átomos das proteínas. Estas hipóteses são baseadas em conhecimento experimental sobre a geometria das proteínas e no experimento de RMN. Dado este conjunto de distâncias, apresentamos duas relações geométricas que evidenciam como ângulos de torção do *backbone* das proteínas podem ser descritos com base em distâncias entre átomos de hidrogênio próximos.

4.1 Restrições de Distâncias

Como vimos no Capítulo 2, proteínas são cadeias de resíduos de aminoácidos. Assim, o *backbone* da proteína é descrito por uma sequência da terna de átomos $\{N^i, C_\alpha^i, C^i\}$, com $i = 1, \dots, p$, onde p é o número de aminoácidos desta proteína. A descrição desta proteína fica completa acrescentando átomos H_N^i ligados a átomos N^i , átomos O^i ligados a átomos C^i , cadeias laterais R^i e átomos H_α^i ligadas a átomos C_α^i . A Figura 23 evidencia as ligações presentes no resíduo i e nos resíduos imediatamente adjacentes. Esta figura também evidencia os ângulos de torção (ou diedrais) ω_i , ϕ_i e ψ_i , definidos no *backbone* da proteína pelas 4-uplas de átomos:

$$\omega_i = \{C_\alpha^{i-1}, C^{i-1}, N^i, C_\alpha^i\}, \phi_i = \{C^{i-1}, N^i, C_\alpha^i, C^i\}, \psi_i = \{N^i, C_\alpha^i, C^i, N^{i+1}\}. \quad (4.1)$$

Como descrito em NELSON; COX (2012), uma aproximação teórica nos diz que os ângulos de torção ω_i admitem somente dois valores distintos, $\omega = 0^\circ$, caso *cis*, ou $\omega = 180^\circ$, caso *trans*. Portanto, independentemente dos valores assumidos por ω_i , os quatro átomos

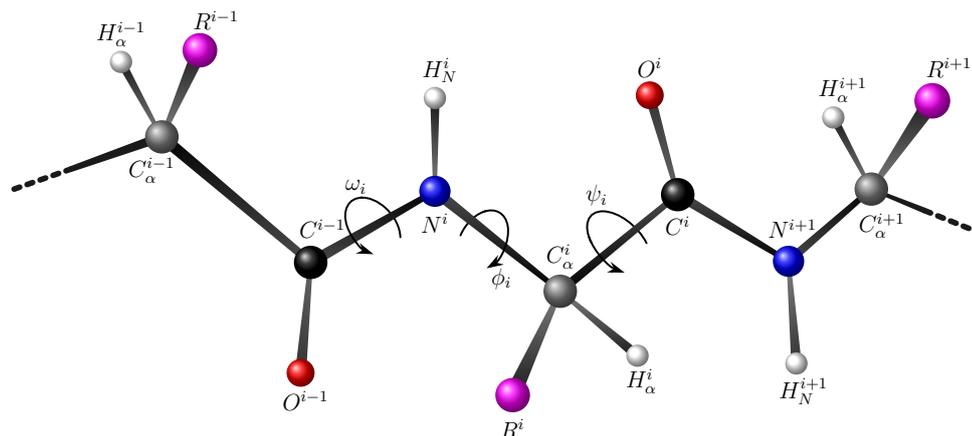


Figura 23 – Representação do i -ésimo resíduo de uma proteína, destacando todos os átomos envolvidos e os ângulos de torção ω_i , ϕ_i e ψ_i .

que descrevem este ângulo de torção fazem parte de um mesmo plano. Esta planaridade se dá pela ligação peptídica, que é rígida e planar. Os átomos H_N^i e O^i também fazem parte deste plano (NELSON; COX, 2012). A Figura 24 evidencia a disposição destes átomos para os resíduos de aminoácidos $i - 1$, i e $i + 1$.

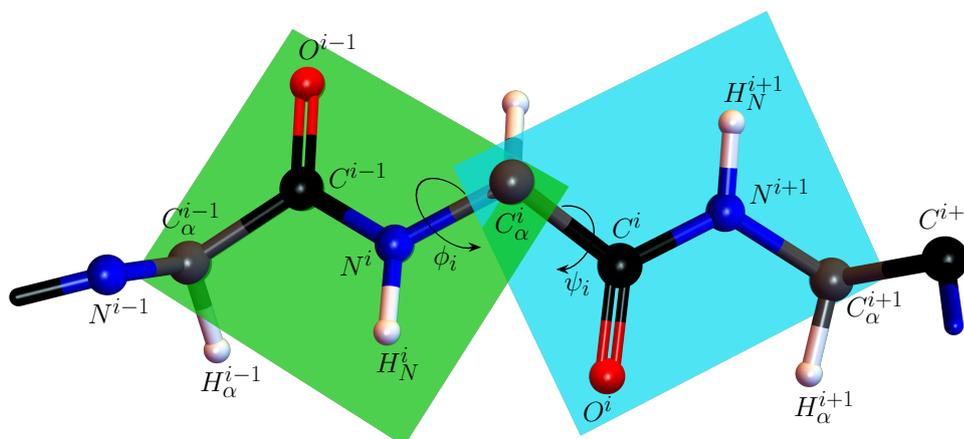


Figura 24 – Representação do plano peptídico no entorno do i -ésimo resíduo de uma proteína.

Como as cadeias laterais não respeitam nenhum padrão, uma vez que variam de proteína para proteína, limitar-nos-emos em trabalhar na determinação da posição dos átomos do *backbone*, pois uma vez que estes átomos são fixados, uma estratégia como a apresentada em COSTA et al. (2014) pode ser utilizada para fixar os átomos das cadeias laterais. Vale observar que o experimento NOESY de RMN (WÜTHRICH, 1986) fornece somente uma estimativa para as distância entre os átomos de hidrogênio próximos, assim, não se têm informações experimentais suficientes para encontrar somente as posições espaciais dos átomos do *backbone* utilizando uma abordagem discreta. Isto se dá, uma vez que não conseguimos definir uma ordem PDGD i que apresente algumas distâncias “extras” entre átomos de resíduos não adjacentes.

Como o nosso intuito é construir o *backbone* de proteínas com base em um conjunto de restrições de distâncias, primeiramente, vamos observar entre quais átomos são conhecidos os valores de distância pela teoria e quais propriedades conformacionais podem resultar em restrições de distâncias. Para isso, vamos enumerar afirmações que tem relação direta com informações destas restrições:

1. A sequência de resíduos de aminoácidos que compõe qualquer proteína é conhecida *a priori* (DONALD; ISTRAIL; PEVZNER, 2011).
2. A distância entre átomos de uma molécula de proteína separados por uma ligação covalente é conhecida (ENGH; HUBER, 1991). Esta é uma informação experimental.
3. Todos os ângulos planos definidos na estrutura proteica são fixos e conhecidos, cujo módulo não é 0° , nem 180° (ENGH; HUBER, 1991). Esta também é uma informação experimental.
4. A distância entre todos os átomos da molécula de proteína que fazem parte um mesmo grupo peptídico são conhecidas (NELSON; COX, 2012). Estas informações de distâncias se dão pela planaridade deste grupo.
5. O experimento NOESY de RMN fornece uma estimativa para distância entre pares de átomos de hidrogênio conhecidos e próximos ($< 5 \sim 6 \text{ \AA}$) (WÜTHRICH, 1986).
6. Uma estimativa para os ângulos de torção ϕ_i e ψ_i , para todo $i = 1, \dots, p$, é conhecida pela predição da rede neural TALOS-N (SHEN; BAX, 2013). Esta predição é dada utilizando informações experimentais de deslocamentos químicos provenientes do experimento de RMN.
7. A propriedade de quiralidade local, observada pela teoria de química das proteínas, nos garante a orientação de qualquer tetraedro formado por 4 átomos separados por no máximo duas ligações covalentes (DONALD; ISTRAIL; PEVZNER, 2011). A Figura 25 ilustra essa propriedade para o tetraedro formado pelos átomos $\{N, C_\alpha, C, H_\alpha\}$.

Com base nas sete afirmações anteriores, vamos assumir que as proteínas consideradas neste trabalho satisfazem o seguinte conjunto de hipóteses:

H1 : Todos os seus átomos são conhecidos.

H2 : A distância exata entre todos os seus átomos separados por uma ligação covalente é conhecida.

H3 : A distância exata entre todos os seus átomos separados por duas ligações covalentes é conhecida.

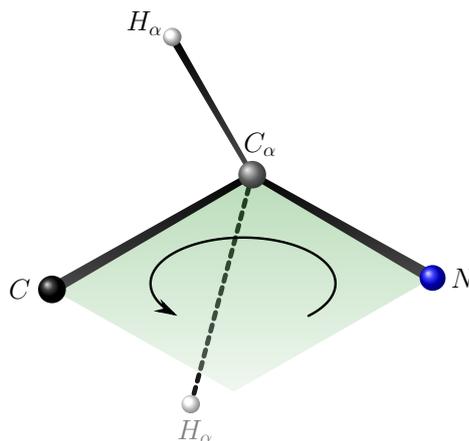


Figura 25 – Quiralidade dos átomos $\{N, C_\alpha, C, H_\alpha\}$.

- H4** : A Informação de Quiralidade Local (IQL) nos garante que a orientação de todo tetraedro definido por quatro átomos que são separados por no máximo duas ligações covalentes é conhecida.
- H5** : A distância entre todos os átomos que pertencem a um mesmo plano peptídico é conhecida.
- H6** : As distâncias fornecidas pelos experimentos de RMN estão associadas a pares de átomos conhecidos.
- H7** : O experimento de RMN fornece uma distância intervalar para cada par de átomos de hidrogênio que estão distantes entre si a no máximo 5 Å.
- H8** : Um intervalo angular para todos os ângulos de torção ϕ_i e uma distância intervalar entre os átomos C^{i-1} e C^i são conhecidos.
- H9** : Um intervalo angular para todos os ângulos de torção ψ_i e uma distância intervalar entre os átomos N^i e N^{i+1} são conhecidos.

Nas Seções 4.2 e 4.3 utilizamos as informações sobre as proteínas presentes neste conjunto de hipóteses para investigar como os ângulos de torção ϕ_i e ψ_i podem ser descritos com relação a restrições de distâncias no entorno do *backbone*.

4.2 Relação dos Ângulos de Torção ϕ_i com Distâncias Entre Hidrogênios Próximos

Nesta seção, através de observações geométricas, vamos construir uma relação entre o ângulo de torção $\phi_i \equiv \{C^{i-1}, N^i, C_\alpha^i, C^i\}$ e outros ângulos de torção definidos no entorno dos átomos que compõe este ângulo, em especial átomos de hidrogênio. A construção

desta relação é baseada no estudo de quatro casos, bem delimitados, descritos nas quatro subseções seguintes, de modo que a relação apresentada na última subseção desta seção descreve as observações feitas em cada uma das 4 subseções anteriores. Salientamos ainda que, no estudo dos quatro casos iniciais fizemos uso de um apelo geométrico, que consideramos intuitivo, e na subseção final utilizamos todo ferramental matemático formal para justificar os argumentos apresentados.

4.2.1 Primeiro Caso de Estudo

Inicialmente, vamos supor que o resíduo de aminoácido R_i , que gera a parte do *backbone* da proteína que compõe o ângulo ϕ_i , não seja a Prolina. Desta maneira, podemos afirmar que existe um átomo H_N^i ligado ao átomo N^i e um átomo H_α^i ligado ao átomo C_α^i . Estas afirmações ficam justificadas pela observação dos 20 aminoácidos da [Figura 1](#). No caso em que R_i é a Glicina, que possui $H_{\alpha_2}^i$ e $H_{\alpha_3}^i$ ligados ao átomo C_α^i , identificamos $H_\alpha^i = H_{\alpha_2}^i$, pois os átomos N^i , C_α^i , C^i e $H_{\alpha_2}^i$ possuem a mesma quiralidade dos átomos N^i , C_α^i , C^i e H_α^i se R_i não é a Glicina ([ALIPANAHI et al., 2013](#)). Observando as considerações teóricas feitas na Subseção [3.2.3](#), se fixamos os átomos C^{i-1} , N^i e C_α^i e o ângulo de torção ϕ_i é conhecido, então ao menos uma posição espacial dos átomos H_N^i , H_α^i e C^i pode ser determinada, e, com isso, na nomenclatura do PGD todos estes átomos são imersos no \mathbb{R}^3 . Para exemplificar esta afirmação, considerando as restrições contidas nas hipóteses apresentadas na Seção [4.1](#) para os átomos C^{i-1} , N^i , C_α^i , H_N^i , H_α^i e C^i , cujos valores numéricos foram retirados de ([ENGH; HUBER, 1991](#)), e supondo que $\phi_i = +30^\circ$, a [Figura 26](#) (a) apresenta uma posição espacial para os átomos mencionados, cuja estrutura foi calculada fixando os três primeiros átomos e os outros três foram realizados, um a um, pela relação de [\(3.40\)](#) considerando as cliques $\{C^{i-1}, N^i, C_\alpha^i, C^i\}$, $\{C^{i-1}, N^i, C_\alpha^i, H_N^i\}$ e $\{N^i, C_\alpha^i, C^i, H_\alpha^i\}$.

Identificando os planos $\{C^{i-1}, N^i, C_\alpha^i\}$, $\{N^i, C_\alpha^i, C^i\}$ e $\{N^i, C_\alpha^i, H_\alpha^i\}$ na estrutura da [Figura 26](#) (a), resultamos na [Figura 26](#) (b). Agora, rotacionando esta estrutura de forma a identificar o ponto de vista onde o vetor formado pelos pontos N^i e C_α^i seja perpendicular a folha deste texto, obtemos a [Figura 26](#) (c). Projetando a estrutura da [Figura 26](#) (c) ortogonalmente no plano formado pela folha do texto, conseguimos identificar os ângulos de torção $\phi_i \equiv \{C^{i-1}, N^i, C_\alpha^i, C^i\}$, $\tau_{C^{i-1}, H_N^i} \equiv \{C^{i-1}, N^i, C_\alpha^i, H_N^i\}$, $\tau_{C^i, H_\alpha^i} \equiv \{C^i, N^i, C_\alpha^i, H_\alpha^i\}$ e $\tau_{H_N^i, H_\alpha^i} \equiv \{H_N^i, N^i, C_\alpha^i, H_\alpha^i\}$ como ângulos planos, e se consideramos a orientação do ângulo de torção ϕ_i como referência, estes ângulos estão no intervalo $(-180^\circ, 180^\circ]$, [Figura 26](#) (d).

Se conhecemos a posição espacial dos átomos C^{i-1} , N^i , H_N^i , C_α^i , H_α^i e C^i os ângulos de torção ϕ_i , τ_{C^{i-1}, H_N^i} , τ_{C^i, H_α^i} e $\tau_{H_N^i, H_\alpha^i}$ são obtidos de maneira trivial pela relação descrita na Equação [\(3.42\)](#), que relaciona um ângulo de torção com as distâncias entre os quatro pontos que o define. Notemos que, em certo sentido, a recíproca da afirmação anterior é

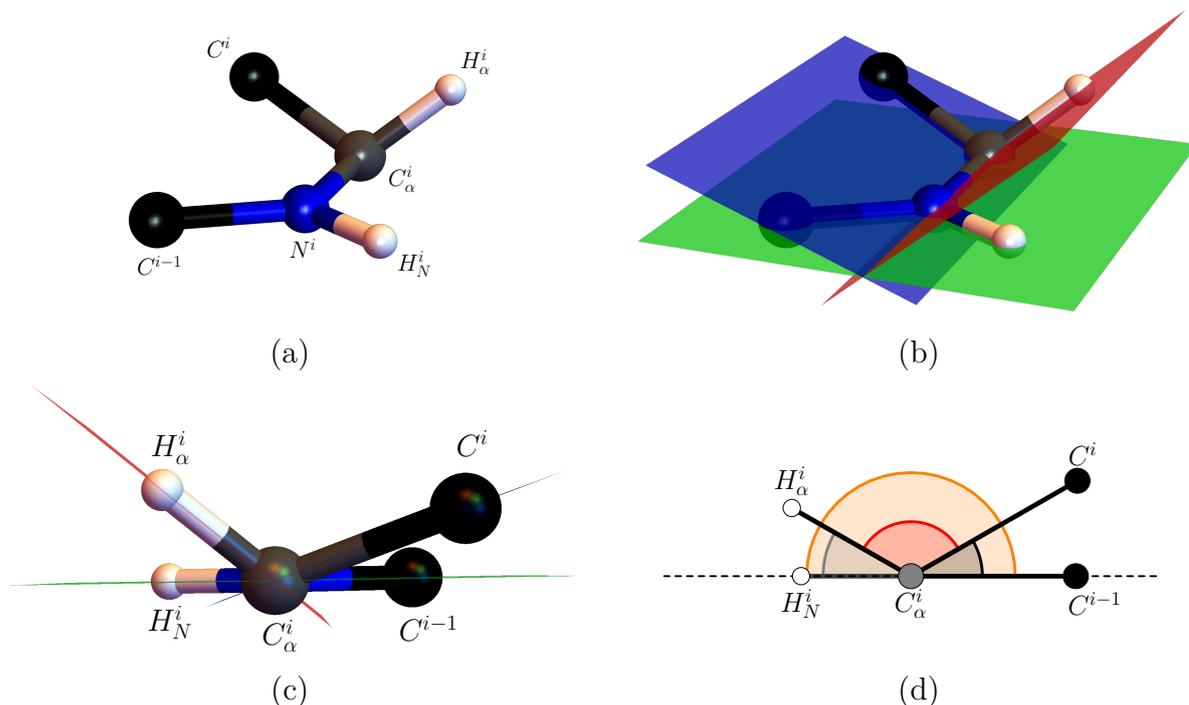


Figura 26 – (a) Uma posição espacial para os átomos C^{i-1} , N^i , H_N^i , C_α^i , H_α^i e C^i se $\phi_i = 30^\circ$. (b) Planos $\{C^{i-1}, N^i, C_\alpha^i\}$, $\{N^i, C_\alpha^i, C^i\}$ e $\{N^i, C_\alpha^i, H_\alpha^i\}$ identificados na estrutura da figura (a). (c) Estrutura da figura (b) rotacionada de modo a identificar um ponto de vista de interesse. (d) Projeção ortogonal da estrutura da figura (c) no plano, identificando os ângulos de torção $\phi_i = \{C^{i-1}, N^i, C_\alpha^i, C^i\}$, $\tau_{C^{i-1}, H_N^i} = \{C^{i-1}, N^i, C_\alpha^i, H_N^i\}$, $\tau_{C^i, H_\alpha^i} = \{C^i, N^i, C_\alpha^i, H_\alpha^i\}$ e $\tau_{H_N^i, H_\alpha^i} = \{H_N^i, N^i, C_\alpha^i, H_\alpha^i\}$ como ângulos planares.

verdadeira, isto é, se os ângulos de torção ϕ_i , τ_{C^{i-1}, H_N^i} , τ_{C^i, H_α^i} e $\tau_{H_N^i, H_\alpha^i}$ são dados, então ao menos uma posição espacial dos átomos que os compõe podem ser obtida. Para isto, basta fixar três átomos, estabelecer três cliques de quatro átomos e utilizar a relação de (3.40) para realizar os outros três átomos.

Pelas hipóteses teóricas apresentadas na Seção 4.1, as distâncias entre os átomos que formam os ângulos de torção τ_{C^{i-1}, H_N^i} e τ_{C^i, H_α^i} são conhecidas *a priori*, assim o valor absoluto destes ângulos pode ser calculado pela relação (3.42). Agora, observamos que o valor absoluto do ângulo de torção $\tau_{H_N^i, H_\alpha^i}$ também pode ser calculado pela mesma relação se a distância entre H_N^i e H_α^i for conhecida. Como o experimento *NOESY* da RMN é capaz de fornecer a distância entre os átomos H_N^i e H_α^i ($\|H_N^i - H_\alpha^i\| < 5 \text{ \AA}$ (GÜNTERT, 1998)) então, o valor absoluto do ângulo de torção $\tau_{H_N^i, H_\alpha^i}$ também pode ser obtido *a priori*.

Agora, para estudar a relação entre os ângulos de torção da estrutura da Figura 26 (d), observando a sua geometria, concluímos que:

$$\left| \tau_{C^{i-1}, H_N^i} \right| = \left| \tau_{H_N^i, H_\alpha^i} \right| + \left| \tau_{C^i, H_\alpha^i} \right| + \phi_i,$$

assim, no caso deste exemplo, o ângulo de torção ϕ_i é dado por:

$$\phi_i = \left(\left| \tau_{C^{i-1}, H_N^i} \right| - \left| \tau_{C^i, H_\alpha^i} \right| \right) - \left| \tau_{H_N^i, H_\alpha^i} \right|, \quad (4.2)$$

e pode ser calculado imediatamente após o estabelecimento dos valores absolutos dos ângulos de torção τ_{C^{i-1}, H_N^i} , $\tau_{H_N^i, H_\alpha^i}$ e τ_{C^i, H_α^i} .

Pela hipótese **H4**, apresentada na Seção 4.1, o tetraedro formado pelos átomos N^i , C_α^i , H_α^i e C^i da estrutura da Figura 26 (a) é rígido e de quiralidade conhecida. Assim, se observamos a Figura 26 (d) e considerarmos que o ângulo de torção ϕ_i pode variar no intervalo $(-180^\circ, 180^\circ]$, percebemos que: se variamos continuamente o ângulo ϕ_i de 0° a 180° , então duas situações distintas ocorrem:

- Em um primeiro momento, o valor de $\left| \tau_{H_N^i, H_\alpha^i} \right|$ diminui de $180^\circ - \left| \tau_{C^i, H_\alpha^i} \right|$ até 0° .
- Em um segundo momento, o valor de $\left| \tau_{H_N^i, H_\alpha^i} \right|$ cresce de 0° até $\left| \tau_{C^i, H_\alpha^i} \right|$.

Para ilustrar esse comentário e apresentar uma relação que descreve o ângulo de torção ϕ_i nas condições destacadas anteriormente, primeiramente vamos definir a seguinte quantidade:

$$\phi_{i,i,N} \equiv \left| \tau_{C^{i-1}, H_N^i} \right| - \left| \tau_{C^i, H_\alpha^i} \right|. \quad (4.3)$$

Agora, utilizando esta quantidade, a sequência de imagens (a)→(c)→(d)→(e)→(b) da Figura 27 ilustram o comentário feito no parágrafo anterior e destacam uma relação para descrever o ângulo ϕ_i em cada situação. Já a sequência de imagens (a)→(f)→(g)→(h)→(b), da mesma figura ilustram o caso quando o ângulo ϕ_i é negativo, que de certa forma é contrário ao caso anterior, uma vez que, em um primeiro momento o valor absoluto do ângulo de torção $\tau_{H_N^i, H_\alpha^i}$ aumenta até o valor de 180° , e depois, em um segundo momento, decresce e atinge um valor limite quando $\phi_i = 180^\circ$.

Sobre as observações da Figura 27, notamos que as relações entre os ângulos de torção podem ser escritas com base no valor absoluto do ângulo de torção $\tau_{H_N^i, H_\alpha^i}$ da seguinte forma:

1. Se $0^\circ \leq \left| \tau_{H_N^i, H_\alpha^i} \right| \leq \phi_{i,i,N}$, então das relações apresentadas nas Figura 27 (a), (c), (d) e (e), tem-se:

$$\begin{cases} \text{(a), (c) e (d)} & \Rightarrow \phi_i = \phi_{i,i,N} - \left| \tau_{H_N^i, H_\alpha^i} \right|, \\ \text{(d) e (e)} & \Rightarrow \phi_i = \phi_{i,i,N} + \left| \tau_{H_N^i, H_\alpha^i} \right|. \end{cases} \quad (4.4)$$

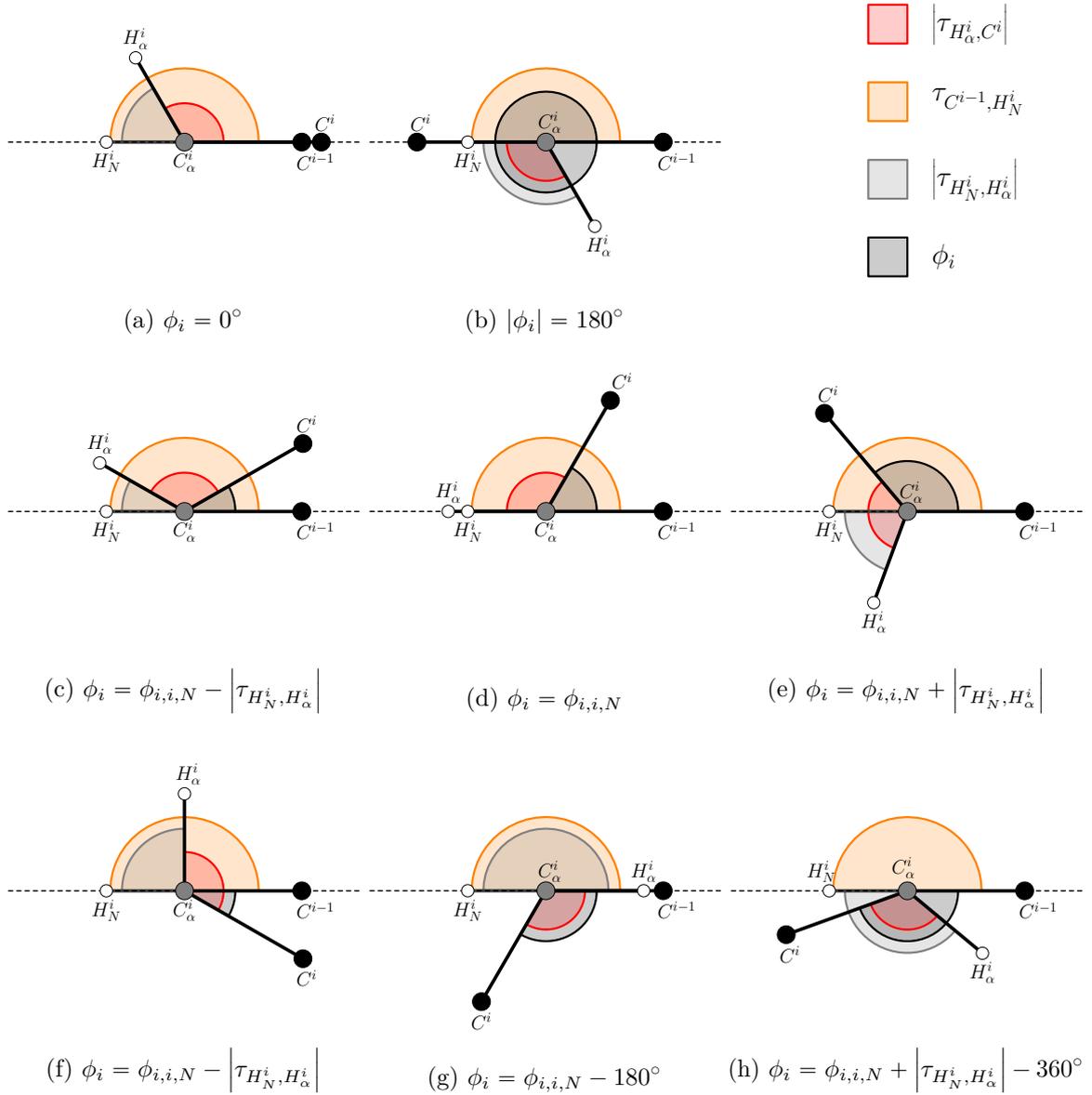


Figura 27 – Relações que descrevem a dinâmica entre os ângulos de torção ϕ_i , τ_{C^{i-1}, H_N^i} , τ_{C^i, H_α^i} e $\tau_{H_N^i, H_\alpha^i}$. O caso $0^\circ \leq \phi_i \leq 180^\circ$ é ilustrado pela sequência de imagens (a)→(c)→(d)→(e)→(b), e o caso $-180^\circ < \phi_i \leq 0^\circ$ é ilustrado pela sequência de imagens (a)→(f)→(g)→(h)→(b).

2. Se $\phi_{i,i,N} \leq |\tau_{H_N^i, H_\alpha^i}| \leq 180^\circ - \phi_{i,i,N}$, então das relações apresentadas nas Figura 27 (a), (b), (e) e (f), tem-se:

$$\begin{cases} \text{(e) e (b)} & \Rightarrow \phi_i = \phi_{i,i,N} + |\tau_{H_N^i, H_\alpha^i}|, \\ \text{(a) e (f)} & \Rightarrow \phi_i = \phi_{i,i,N} - |\tau_{H_N^i, H_\alpha^i}|. \end{cases} \quad (4.5)$$

3. Se $180^\circ - \phi_{i,i,N} < \left| \tau_{H_N^i, H_\alpha^i} \right| \leq 180^\circ$, então das relações apresentadas nas [Figura 27](#) (b), (f), (g) e (h), tem-se:

$$\begin{cases} \text{(f) e (g)} & \Rightarrow \phi_i = \phi_{i,i,N} - \left| \tau_{H_N^i, H_\alpha^i} \right|, \\ \text{(g), (h) e (b)} & \Rightarrow \phi_i = \phi_{i,i,N} + \left| \tau_{H_N^i, H_\alpha^i} \right| - 360^\circ. \end{cases} \quad (4.6)$$

Desta maneira, para este exemplo, os três itens acima nos mostram uma maneira de relacionar os ângulos de torção ϕ_i , τ_{C^{i-1}, H_N^i} , τ_{C^i, H_α^i} e $\tau_{H_N^i, H_\alpha^i}$. O fato de observarmos duas possibilidades para a determinação do ângulo ϕ_i é natural, pois, as relações que descrevem o ângulo ϕ_i são baseadas no valor absoluto de $\tau_{H_N^i, H_\alpha^i}$.

4.2.2 Segundo Caso de Estudo

Na construção do exemplo da subseção anterior, assumimos que $\left| \tau_{C^{i-1}, H_N^i} \right| = 180^\circ$ o que, na realidade, significa que os átomos C^{i-1} , N^i , H_N^i , C_α^i são coplanares. No entanto, por conta de movimentos termodinâmicos dos átomos em uma molécula, esta coplanaridade dificilmente é observada na prática, mas observa-se que $\left| \tau_{C^{i-1}, H_N^i} \right| \approx 180^\circ$ ([MATTHEWS, 2016](#)). Desta maneira, para construir uma relação entre ângulos de torção, como a apresentada em cada uma das situações da [Figura 27](#) para este caso “mais próximo” da realidade, vamos observar o caso em que $\left| \tau_{C^{i-1}, H_N^i} \right| = 180^\circ - \varepsilon$, com ε pequeno ($\leq 8^\circ$) ([MATTHEWS, 2016](#)). Para isso, vamos considerar uma situação hipotética como a do exemplo da [Figura 26](#), mas, nesse caso, considerando duas situações: (I) $\tau_{C^{i-1}, H_N^i} = 180^\circ - \varepsilon$ e (II) $\tau_{C^{i-1}, H_N^i} = -180^\circ + \varepsilon$. Para a análise destas situações, consideremos as duas imagens da [Figura 28](#), análogas a da [Figura 26](#) (d).

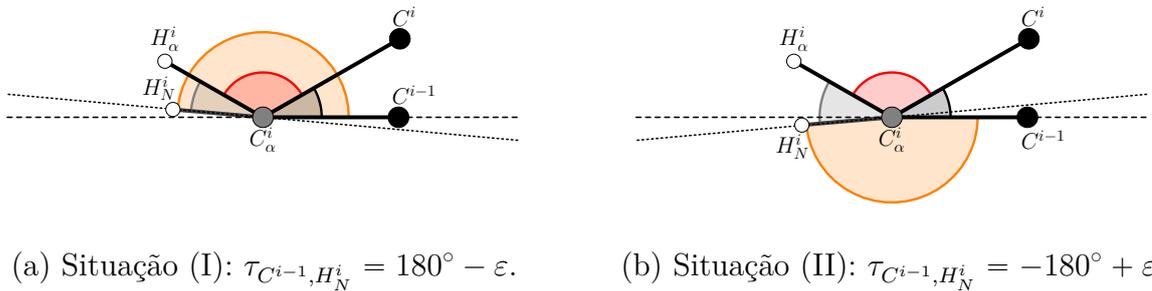


Figura 28 – Identificação dos ângulos de torção ϕ_i , τ_{C^{i-1}, H_N^i} , τ_{C^i, H_α^i} e $\tau_{H_N^i, H_\alpha^i}$ para as situações (I) e (II).

Generalizando a relação (4.3) para:

$$\phi_{i,i,N} \equiv \tau_{C^{i-1}, H_N^i} - \left| \tau_{C^i, H_\alpha^i} \right|, \quad (4.7)$$

e observando a geometria das [Figura 28](#) (a) e (b), concluímos que:

$$\begin{cases} (I) & : \tau_{C^{i-1}, H_N^i} = \left| \tau_{H_N^i, H_\alpha^i} \right| + \left| \tau_{C^i, H_\alpha^i} \right| + \phi_i, \\ (II) & : -\tau_{C^{i-1}, H_N^i} + \left| \tau_{H_N^i, H_\alpha^i} \right| + \left| \tau_{C^i, H_\alpha^i} \right| + \phi_i = 360^\circ, \end{cases}$$

ou, equivalentemente,

$$(I) : \phi_i = \phi_{i,i,N} - \left| \tau_{H_N^i, H_\alpha^i} \right|, \quad (4.8)$$

$$(II) : \phi_i = \phi_{i,i,N} - \left| \tau_{H_N^i, H_\alpha^i} \right| + 360^\circ. \quad (4.9)$$

Assim, percebemos que a expressão que relaciona os ângulos de torção ϕ_i , τ_{C^{i-1}, H_N^i} , τ_{C^i, H_α^i} e $\tau_{H_N^i, H_\alpha^i}$ muda com relação à troca de orientação do ângulo de torção τ_{C^{i-1}, H_N^i} .

Observando atentamente as relações (4.2) e (4.8), notamos que elas são iguais, assim, encontramos um primeiro indício de que uma relação geral para descrever o ângulo de torção ϕ_i para a situação (I) será a mesma que a apresentada na Subseção 4.2.1. No entanto, neste ponto do texto só nos interessa observar que uma relação geral para descrever o ângulo ϕ_i deve satisfazer pelo menos as situações (I) e (II).

4.2.3 Terceiro Caso de Estudo

Para dar mais um passo na construção de uma relação geral entre ângulos de torção que nos apresente o valor do ângulo ϕ_i , vamos observar um exemplo onde o resíduo R_i é a Prolina. Para tanto, vamos considerar uma estrutura com os átomos C^{i-1} , N^i , C_δ^i , $H_{\delta_2}^i$, $H_{\delta_3}^i$, C_α^i , H_α^i e C^i se $\phi_i = 20^\circ$. A [Figura 29](#) (a) apresenta esta estrutura, que foi construída de maneira análoga a estrutura da [Figura 26](#) (a), fazendo somente as adequações necessárias para realização dos novos átomos. Neste exemplo, assumimos um caso teórico: os átomos C^{i-1} , N^i , C_δ^i e C_α^i são coplanares. Novamente, o nosso objeto de estudo é a figura planar da [Figura 29](#) (d).

Observando atentamente a [Figura 29](#) (d), nota-se que:

$$\tau_{C^{i-1}, H_{\delta_3}^i} = \left| \tau_{H_{\delta_3}^i, H_\alpha^i} \right| + \left| \tau_{C^i, H_\alpha^i} \right| + \phi_i. \quad (4.10)$$

Generalizando a relação (4.7) para:

$$\phi_{i,i,k} \equiv \tau_{C^{i-1}, H_k^i} - \left| \tau_{C^i, H_\alpha^i} \right|, \text{ onde } k = N, \delta_3, \quad (4.11)$$

podemos escrever a Equação (4.10), como:

$$\phi_i = \phi_{i,i,k} - \left| \tau_{H_k^i, H_\alpha^i} \right|, \text{ com } k = \delta_3. \quad (4.12)$$

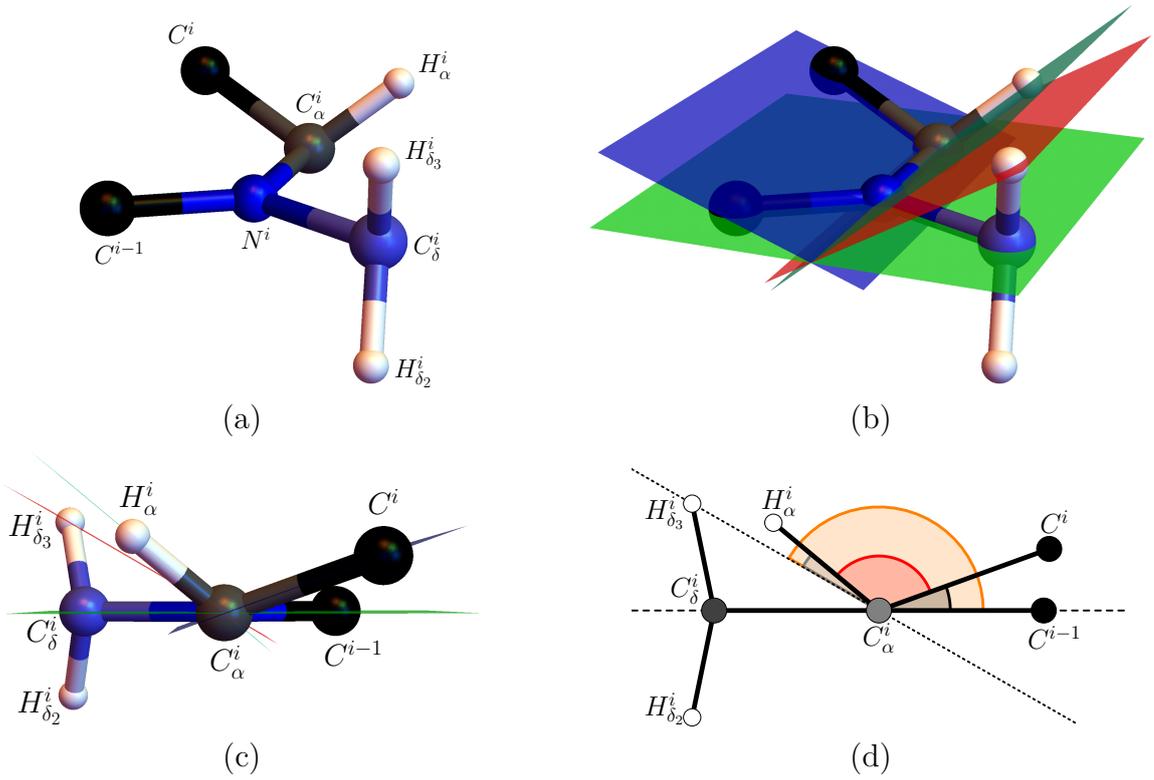


Figura 29 – (a) Uma posição espacial para os átomos C^{i-1} , N^i , C_α^i , H_α^i , $H_{\delta_2}^i$, $H_{\delta_3}^i$, C_δ^i e C^i se $\phi_i = 20^\circ$. (b) Planos $\{C^{i-1}, N^i, C_\alpha^i\}$, $\{N^i, C_\alpha^i, C^i\}$ e $\{N^i, C_\alpha^i, H_{\delta_3}^i\}$ identificados na estrutura da figura (a). (c) Estrutura da figura (b) rotacionada de modo a identificar um ponto de vista de interesse. (d) Projecção ortogonal da estrutura da figura (c) no plano, identificando os ângulos de torção $\phi_i = \{C^{i-1}, N^i, C_\alpha^i, C^i\}$, $\tau_{C^{i-1}, H_{\delta_3}^i} = \{C^{i-1}, N^i, C_\alpha^i, H_{\delta_3}^i\}$, $\tau_{C^i, H_\alpha^i} = \{C^i, N^i, C_\alpha^i, H_\alpha^i\}$ e $\tau_{H_{\delta_3}^i, H_\alpha^i} = \{H_{\delta_3}^i, N^i, C_\alpha^i, H_\alpha^i\}$ como ângulos planares.

Agora, observando as imagens da [Figura 29](#) (d) e [Figura 28](#) (a) notamos que em ambos os casos tem-se que $\tau_{C^{i-1}, H_k^i} > 0$, com $k = N, \delta_3$. Observamos, também, que se tomamos $k = N, \delta_3$, então a relação (4.12) descreve o ângulo de torção ϕ_i do exemplo da situação (I) da Subseção 4.2.2, descrito pela Equação (4.8), e do exemplo da [Figura 29](#) (d).

Como vimos na Subseção 4.2.1, uma relação que observa a dinâmica entre os ângulos de torção ϕ_i e $\tau_{H_N^i, H_\alpha^i}$ leva em conta que a variação do ângulo ϕ_i resulta na diminuição ou no aumento do módulo do ângulo $\tau_{H_N^i, H_\alpha^i}$. Considerando os exemplos das [Figura 28](#) e [Figura 29](#) (d), a hipótese **H5** nos garante que o tetraedro formado pelos átomos N^i , C_α^i , H_α^i e C^i é rígido, assim se supormos que ϕ_i é variável, notamos que o ângulo τ_{C^{i-1}, H_k^i} , com $k = N, \delta_3$, tem influência direta na magnitude da mudança do valor absoluto do ângulo $\tau_{H_k^i, H_\alpha^i}$. Esta observação pode ser constatada com mais clareza se notarmos que a linha pontilhada das [Figura 28](#) e [Figura 29](#) (d) representa uma reta que divide o plano de modo que, no caso dinâmico, o comportamento do valor absoluto do ângulo $\tau_{H_k^i, H_\alpha^i}$, com $k = N, \delta_3$, muda quando o átomo H_α^i cruza essa linha: para de decrescer e começa a crescer ou vice-versa.

Desta maneira, consideramos que $\left| \tau_{H_k^i, H_\alpha^i} \right| = 0^\circ$ ou $\left| \tau_{H_k^i, H_\alpha^i} \right| = 180^\circ$ estão associados a valores críticos para o ângulo ϕ_i .

Novamente, neste ponto do texto não apresentamos uma relação que evidencia a dinâmica dos ângulos de torção ϕ_i e $\tau_{H_{\delta_3}^i, H_\alpha^i}$, como fizemos na Subseção 4.2.1. No entanto, percebemos uma semelhança deste caso de estudo com um dos casos estudados na Subseção 4.2.2, o exemplo apresentado na Figura 28 (a). Semelhança esta que foi descrita pela relação (4.12), com $k = N$, δ_3 . No próximo caso de estudo, vemos um outro exemplo que nos permite intuir que uma relação que mostra a dinâmica entre os ângulos ϕ_i e $\tau_{H_k^i, H_\alpha^i}$ é baseada no valor do ângulo τ_{C^{i-1}, H_k^i} , e na descrição geral do ângulo ϕ_i ficará claro que a semelhança destes dois casos reside no fato de que $\left| \tau_{C^i, H_\alpha^i} \right| \leq \tau_{C^{i-1}, H_k^i} \leq 180^\circ$.

Na construção inicial deste exemplo, assumimos que os átomos C^{i-1} , N^i , C_δ^i e C_α^i são coplanares. No entanto, observando a Figura 29 (d), notamos que se considerarmos o caso em que o ângulo de torção $\left| \tau_{C^{i-1}, C_\delta^i} \right| = 180^\circ - \varepsilon$, então esta variação de ε no ângulo $\tau_{C^{i-1}, C_\delta^i}$ é transferida diretamente para o valor do ângulo de torção $\tau_{C^{i-1}, H_{\delta_3}^i}$, e dadas as proporções do ângulo ε , ainda temos que $\left| \tau_{C^i, H_\alpha^i} \right| \leq \tau_{C^{i-1}, H_{\delta_3}^i} \pm \varepsilon \leq 180^\circ$. Deste modo, uma mesma relação pode descrever a dinâmica do ângulo ϕ_i para o caso teórico ($\varepsilon = 0^\circ$) e para um caso “mais próximo” da realidade ($\varepsilon \neq 0^\circ$).

4.2.4 Quarto Caso de Estudo

Neste caso, vamos considerar que o ângulo de torção ϕ_i seja proveniente de um resíduo de aminoácido R_i qualquer. Assim, afirmamos que existe um átomo H_α^i ligado ao átomo C_α^i . Para o caso em que R_i é a Glicina, levamos em conta as mesmas considerações apresentadas no início da Subseção 4.2.1. Como já foi visto, se o ângulo de torção ϕ_i é conhecido, então ao menos uma posição espacial dos átomos C^{i-1} , N^i , C_α^i , H_α^i e C^i é conhecida. Além do mais, neste caso de estudo vamos considerar que a distância entre os átomos de hidrogênio H_k^j e H_α^i , com $j < i$, seja conhecida por advento da experiência de RMN. *A priori*, não temos nenhuma garantia de que a distância entre o átomo H_k^j e qualquer átomo do *backbone* proveniente do resíduo R_i seja conhecida. No entanto, como descrevemos na Subseção 3.2.2, o algoritmo *BEP* resolve o PDGD realizando um vértice de cada vez em uma ordem pré estabelecida, assim, por este algoritmo, se vamos realizar o vértice v_i , então qualquer vértice v_j com $j < i$ já está realizado, ou seja, conhecemos a posição espacial de suas realizações x_1, x_2, \dots, x_{i-1} . Desta maneira, a distância entre todos os vértices v_1, v_2, \dots, v_{i-1} pode ser calculada de maneira trivial. Por conta desta particularidade do algoritmo *BEP*, dependendo da ordem definida nos vértices do grafo, a distância do átomo de hidrogênio H_k^j aos átomos C^{i-1} , N^i e C_α^i pode ser calculada na iminência de realizar do átomo C^i , e com isso, antes de estabelecermos o valor do ângulo de torção ϕ_i .

Desta maneira, vamos considerar os dois exemplos apresentados na [Figura 30](#), que foram construídos considerando $\phi = 30^\circ$ e apresentam duas situações distintas para a posição do átomo de hidrogênio H_k^j .

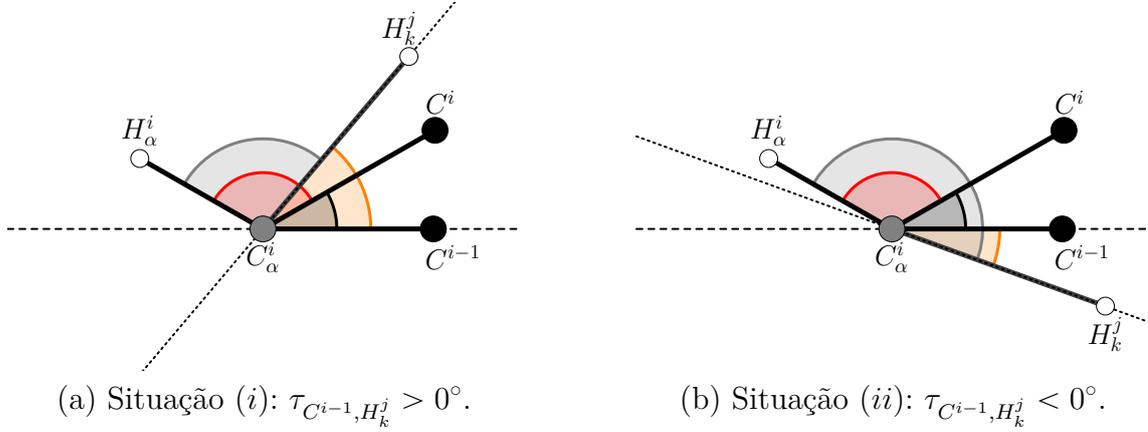


Figura 30 – Identificação dos ângulos de torção ϕ_i , τ_{C^{i-1}, H_k^j} , τ_{C^i, H_α^i} e $\tau_{H_k^j, H_\alpha^i}$ para as situações (i) e (ii).

Generalizando a relação (4.11) para:

$$\phi_{i,j,k} \equiv \tau_{C^{i-1}, H_k^j} - \left| \tau_{C^i, H_\alpha^i} \right|, \quad (4.13)$$

com $j \leq i$ e k representando um átomo de hidrogênio qualquer. Da geometria das imagens apresentadas na [Figura 30](#), notamos que o ângulo ϕ_i , nestas duas situações, é dado por:

$$(i) : \phi_i = \phi_{i,j,k} + \left| \tau_{H_k^j, H_\alpha^i} \right|, \quad (4.14)$$

$$(ii) : \phi_i = \phi_{i,j,k} + \left| \tau_{H_k^j, H_\alpha^i} \right|. \quad (4.15)$$

Observando que a quantidade (4.13) também é uma generalização para (4.7), se comparamos as relações que encontramos para as situações (i) e (ii) com o observado para as situações (I) e (II) do exemplo da Subseção 4.2.2, isto é, comparar as relações (4.8) e (4.14) e as relações (4.9) e (4.15), nota-se que: embora nas situações (I) e (i) tem-se que $\tau_{C^{i-1}, H_k^j} > 0$, o ângulo ϕ_i é descrito de maneira diferente para os dois casos. O mesmo acontece para as situações (II) e (ii), no entanto, neste caso, ambas as situações apresentam $\tau_{C^{i-1}, H_k^j} < 0$.

Desta maneira, constatamos que somente o sinal de τ_{C^{i-1}, H_k^j} não é suficiente para distinguir as relações que descrevem o ângulo de torção ϕ_i e, portanto, estas relações devem ser dependentes do valor de τ_{C^{i-1}, H_k^j} . Na próxima subseção apresentaremos uma relação geral para descrever o comportamento dinâmico entre os ângulos de torção ϕ_i , τ_{C^{i-1}, H_k^j} , τ_{C^i, H_α^i} e $\tau_{H_k^j, H_\alpha^i}$ e a similaridade ou não de todos os casos estudados até aqui são completamente explicadas.

4.2.5 Relação Geral para Determinar o Ângulo de Torção ϕ_i

Nesta subseção construímos uma relação geral entre os ângulos de torção ϕ_i , τ_{C^{i-1}, H_k^j} , τ_{C^i, H_α^i} e $\tau_{H_k^j, H_\alpha^i}$. Para dar um passo a mais nesta direção, vamos observar que as relações que descrevemos em todos os exemplos anteriores, entre estes ângulos, podem ser interpretadas como composições de rotações no plano (neste trabalho não detalhamos o conceito de rotação em \mathbb{R}^3 , mas sugerimos as referências (HOFFMAN; KUNZE, 1971; HESTENES, 1999) como consulta). Para entender melhor esta afirmação, primeiramente vamos lembrar que os ângulos de torção são provenientes de tetraedros que possuem uma orientação, conforme descrito na Subseção 3.2.3. Desta maneira, salientamos que a projeção ortogonal da estrutura proteica no plano, formada por estes ângulos, foi tomada de forma que um ângulo de torção, positivo ou negativo, representa, no caso planar, uma rotação no sentido anti-horário ou horário, respectivamente. Como estamos interessados em uma relação entre ângulos, vamos considerar sem perda de generalidade que vamos rotacionar pontos de uma circunferência de raio unitário centrada no ponto $C_\alpha^i = (0, 0)$. Assim, o ponto C^i é representado por uma rotação de ângulo $\phi_i = \{C^{i-1}, N^i, C_\alpha^i, C^i\}$ do ponto C^{i-1} . Um processo similar é considerado para os outros átomos e ângulos. Logo, considerando que $R(\theta)$ seja uma matriz de rotação 2×2 de ângulo θ e centro em C_α^i , tem-se que:

$$C^i = R(\phi_i) C^{i-1}, \quad C^i = R(-\tau_{C^i, H_\alpha^i}) H_\alpha^i, \quad H_\alpha^i = R(\tau_{H_k^j, H_\alpha^i}) H_k^j, \quad H_k^j = R(\tau_{C^{i-1}, H_k^j}) C^{i-1}.$$

Portanto, das rotações apresentadas anteriormente, concluí-se que:

$$\begin{aligned} R(\phi_i) C^{i-1} &= R(-\tau_{C^i, H_\alpha^i}) H_\alpha^i && \Leftrightarrow \\ R(\phi_i) C^{i-1} &= R(-\tau_{C^i, H_\alpha^i}) R(\tau_{H_k^j, H_\alpha^i}) H_k^j && \Leftrightarrow \\ R(\phi_i) C^{i-1} &= R(-\tau_{C^i, H_\alpha^i}) R(\tau_{H_k^j, H_\alpha^i}) R(\tau_{C^{i-1}, H_k^j}) C^{i-1} && \Leftrightarrow \\ R(\phi_i) C^{i-1} &= R(\tau_{C^{i-1}, H_k^j} - \tau_{C^i, H_\alpha^i} + \tau_{H_k^j, H_\alpha^i}) C^{i-1} \end{aligned} \quad (4.16)$$

Logo,

$$\left[R(\phi_i) - R(\tau_{C^{i-1}, H_k^j} - \tau_{C^i, H_\alpha^i} + \tau_{H_k^j, H_\alpha^i}) \right] C^{i-1} = \vec{\mathbf{0}},$$

onde $\vec{\mathbf{0}}$ é a matriz 2×2 nula. Como o ponto $C^{i-1} \neq (0, 0)$ e os ângulos de torção ϕ_i e $\tau_{H_k^j, H_\alpha^i}$ são arbitrários, então:

$$R(\phi_i) = R(\tau_{C^{i-1}, H_k^j} - \tau_{C^i, H_\alpha^i} + \tau_{H_k^j, H_\alpha^i}). \quad (4.17)$$

Uma vez que, $R(\theta) = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$, da Equação (4.17), tem-se que:

$$\begin{cases} \cos(\phi_i) &= \cos(\tau_{C^{i-1}, H_k^j} - \tau_{C^i, H_\alpha^i} + \tau_{H_k^j, H_\alpha^i}), \\ \sin(\phi_i) &= \sin(\tau_{C^{i-1}, H_k^j} - \tau_{C^i, H_\alpha^i} + \tau_{H_k^j, H_\alpha^i}). \end{cases} \quad (4.18)$$

Como geometricamente não conseguimos distinguir os casos em que um ângulo de torção é igual a -180° ou 180° , deste ponto em diante, vamos considerar que essas duas condições estão representadas pelo ângulo positivo. Logo, das duas igualdades de (4.18), se $\phi_i \in (-180^\circ, 180^\circ]$, concluímos que:

$$\phi_i = \begin{cases} \tau_{C^{i-1}, H_k^j} - \tau_{C^i, H_\alpha^i} + \tau_{H_k^j, H_\alpha^i} & \text{se } |\tau_{C^{i-1}, H_k^j} - \tau_{C^i, H_\alpha^i} + \tau_{H_k^j, H_\alpha^i}| < 180^\circ, \\ \tau_{C^{i-1}, H_k^j} - \tau_{C^i, H_\alpha^i} + \tau_{H_k^j, H_\alpha^i} + 360^\circ & \text{se } \tau_{C^{i-1}, H_k^j} - \tau_{C^i, H_\alpha^i} + \tau_{H_k^j, H_\alpha^i} < -180^\circ, \\ \tau_{C^{i-1}, H_k^j} - \tau_{C^i, H_\alpha^i} + \tau_{H_k^j, H_\alpha^i} - 360^\circ & \text{se } \tau_{C^{i-1}, H_k^j} - \tau_{C^i, H_\alpha^i} + \tau_{H_k^j, H_\alpha^i} > 180^\circ, \\ 180^\circ & \text{se } |\tau_{C^{i-1}, H_k^j} - \tau_{C^i, H_\alpha^i} + \tau_{H_k^j, H_\alpha^i}| = 180^\circ. \end{cases} \quad (4.19)$$

Desta maneira, observamos que as quatro relações apresentadas em (4.19) são possibilidades para descrever a dinâmica dos ângulos de torção. Agora, vamos nos atentar aos seguintes fatos:

- A hipótese **H4** (IQL) nos garante que a orientação do tetraedro formado pelos átomos N^i , C_α^i , C^i e H_α^i é conhecida, e **ALIPANAHI et al. (2013)** nos diz que a única orientação possível para este caso é a positiva. Portanto, $|\tau_{C^i, H_\alpha^i}| = \tau_{C^i, H_\alpha^i}$;
- O ângulo de torção $\tau_{C^{i-1}, H_k^j} \in (-180^\circ, 180^\circ]$ é um valor fixo dado, uma vez que:
 1. Se $j < i$, então a posição dos átomos C^{i-1} , N^i , C_α^i e H_k^j no \mathbb{R}^3 é conhecida e a orientação do tetraedro formado por estes átomos pode ser aferida de maneira trivial.
 2. Se $j = i$, então pela hipótese **H5** (plano peptídico) tem-se que o ângulo τ_{C^{i-1}, H_k^i} é conhecido com valor fixo.
- Como o ângulo de torção $\tau_{H_k^j, H_\alpha^i} \in (-180^\circ, 180^\circ]$, então $\tau_{H_k^j, H_\alpha^i} = \pm |\tau_{H_k^j, H_\alpha^i}|$ e o sinal deste ângulo é definido pela orientação do tetraedro formado pelos átomos H_k^j , N^i , C_α^i e H_α^i , que deve ser identificada de alguma maneira. No entanto, pelo conjunto de hipóteses que estamos considerando, as restrições de distâncias que envolvem estes átomos aplicadas a relação (3.42) nos garantem, somente, o valor absoluto de $\tau_{H_k^j, H_\alpha^i}$.

Dadas estas três observações, utilizando a quantidade descrita em (4.13), podemos reescrever as relações de (4.19) como:

$$\phi_i = \begin{cases} \phi_{i,j,k} \pm |\tau_{H_k^j, H_\alpha^i}| & \text{se } |\phi_{i,j,k} \pm |\tau_{H_k^j, H_\alpha^i}|| < 180^\circ, \\ \phi_{i,j,k} \pm |\tau_{H_k^j, H_\alpha^i}| + 360^\circ & \text{se } \phi_{i,j,k} \pm |\tau_{H_k^j, H_\alpha^i}| < -180^\circ, \\ \phi_{i,j,k} \pm |\tau_{H_k^j, H_\alpha^i}| - 360^\circ & \text{se } \phi_{i,j,k} \pm |\tau_{H_k^j, H_\alpha^i}| > 180^\circ, \\ 180^\circ & \text{se } |\phi_{i,j,k} \pm |\tau_{H_k^j, H_\alpha^i}|| = 180^\circ. \end{cases} \quad (4.20)$$

E assim, levando em conta as informações de distâncias que temos *a priori*, existem, a princípio, 7 possibilidades para a relação entre os ângulos de torção. Notamos ainda que algumas destas relações nós já observamos na descrição dos exemplos que tratamos ao longo desta seção.

Agora, para identificar a orientação do tetraedro definido pelo ângulo de torção $\tau_{H_k^j, H_\alpha^i}$, isto é, identificar o seu sinal, vamos voltar a interpretação de rotacionar pontos em uma circunferência unitária com centro na origem. Com um raciocínio análogo ao que foi apresentado no processo descrito em (4.16), podemos escrever a posição do átomo H_α^i com relação ao ângulo de torção ϕ_i da seguinte maneira:

$$H_\alpha^i = R(\tau_{C^i, H_\alpha^i} + \phi_i) C^{i-1}. \quad (4.21)$$

Como o ângulo τ_{C^{i-1}, H_k^j} é fixo, o átomo H_k^j possui posição fixa nesta circunferência. Então, a orientação do ângulo $\tau_{H_k^j, H_\alpha^i}$ será dada conforme a posição do átomo H_α^i , que por sua vez é dependente do ângulo ϕ_i . Uma vez que, $\phi_i \in (-180^\circ, 180^\circ] = (-180^\circ, 0^\circ] \cup [0^\circ, 180^\circ]$, primeiramente, vamos considerar disposição dos átomos C^{i-1} , C_α^i , H_α^i e C^i se $\phi_i = 0^\circ$. Esta configuração é destacada na Figura 31. Desta disposição dos átomos, nota-se que o plano pode ser dividido nas 4 regiões disjuntas duas a duas:

- $\mathcal{R}_1 = \{(\rho, \theta) \in [0, +\infty) \times (-180^\circ, 180^\circ] \mid 0^\circ \leq \theta < \tau_{C^i, H_\alpha^i}\}$;
- $\mathcal{R}_2 = \{(\rho, \theta) \in [0, +\infty) \times (-180^\circ, 180^\circ] \mid \tau_{C^i, H_\alpha^i} \leq \theta \leq 180^\circ\}$;
- $\mathcal{R}_3 = \{(\rho, \theta) \in [0, +\infty) \times (-180^\circ, 180^\circ] \mid -180^\circ + \tau_{C^i, H_\alpha^i} < \theta < 0^\circ\}$;
- $\mathcal{R}_4 = \{(\rho, \theta) \in [0, +\infty) \times (-180^\circ, 180^\circ] \mid -180^\circ < \theta \leq -180^\circ + \tau_{C^i, H_\alpha^i}\}$.

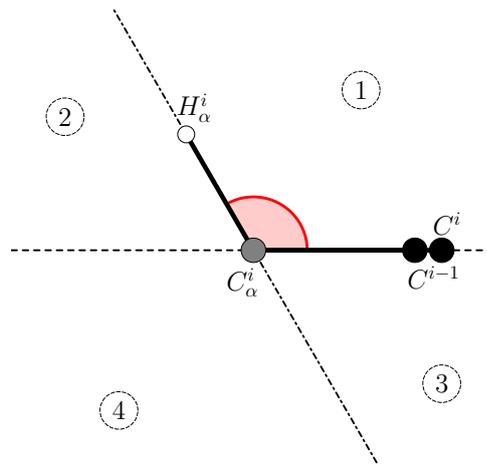


Figura 31 – Posição dos átomos C^{i-1} , C_α^i , H_α^i e C^i no plano da projeção ortogonal se $\phi_i = 0^\circ$.

Assim, para identificar a orientação do ângulo $\tau_{H_k^j, H_\alpha^i}$, vamos investigar os 4 casos definidos se $H_k^j \in \mathcal{R}_m$, com $m = 1, 2, 3, 4$. No entanto, como o átomo H_k^j fica definido pelo ângulo de torção τ_{C^{i-1}, H_k^j} , então, podemos reescrever estes 4 casos como:

- Caso 1: $0^\circ \leq \tau_{C^{i-1}, H_k^j} < \tau_{C^i, H_\alpha^i}$;
- Caso 2: $\tau_{C^i, H_\alpha^i} \leq \tau_{C^{i-1}, H_k^j} \leq 180^\circ$;
- Caso 3: $-180^\circ + \tau_{C^i, H_\alpha^i} < \tau_{C^{i-1}, H_k^j} < 0^\circ$;
- Caso 4: $-180^\circ < \tau_{C^{i-1}, H_k^j} \leq -180^\circ + \tau_{C^i, H_\alpha^i}$.

Na análise destes casos, vamos identificar como é descrito o ângulo ϕ_i em intervalos para o valor de $\left| \tau_{H_k^j, H_\alpha^i} \right|$. Para definir estes intervalos, vamos considerar os valores para o ângulo ϕ_i tais que $\left| \tau_{H_k^j, H_\alpha^i} \right| = 0^\circ$ ou $\left| \tau_{H_k^j, H_\alpha^i} \right| = 180^\circ$, denotando os mesmos por ${}^m\phi_i^{*+}$, o ângulo não negativo, ou ${}^m\phi_i^{*-}$, o ângulo não positivo, com m representando o caso considerado. Denominamos estes valores como críticos, pois veremos que estes pontos são responsáveis pela inversão da orientação do ângulo de torção $\tau_{H_k^j, H_\alpha^i}$. Também vamos utilizar a quantidade definida em (4.13) na descrição destes casos. Assim, as relações que escrevem o ângulo ϕ_i para os quatro casos são apresentadas a seguir:

Caso 1: $0^\circ \leq \tau_{C^{i-1}, H_k^j} < \tau_{C^i, H_\alpha^i}$

Para este caso, a [Figura 32](#) evidencia a disposição dos átomos H_k^j , C^{i-1} , C^i , H_α^i e C^i para as seguintes situações: $\phi_i = 0^\circ$, $|\phi_i| = 180^\circ$, $\phi_i = {}^1\phi_i^{*+}$ e $\phi_i = {}^1\phi_i^{*-}$. Portanto, observa-se que os valores ${}^1\phi_i^{*\pm}$ podem ser descritos em função da quantidade $\phi_{i,j,k}$, que é descrita por (4.13).

Sobre as quantidades ${}^1\phi_i^{*-} = \phi_{i,j,k}$ e ${}^1\phi_i^{*+} = {}^1\phi_i^{*-} + 180^\circ$, observamos que:

$$\begin{aligned}
 0^\circ &\leq \tau_{C^{i-1}, H_k^j} < \tau_{C^i, H_\alpha^i} &&\Leftrightarrow \\
 -\tau_{C^i, H_\alpha^i} &\leq \phi_{i,j,k} < 0^\circ &&\Leftrightarrow \\
 -\tau_{C^i, H_\alpha^i} &\leq {}^1\phi_i^{*-} < 0^\circ &&\Leftrightarrow \\
 180^\circ - \tau_{C^i, H_\alpha^i} &\leq {}^1\phi_i^{*+} < 180^\circ.
 \end{aligned} \tag{4.22}$$

Agora, utilizando a intuição do desenho geométrico, na [Figura 33](#) sistematizamos as relações que descrevem a dinâmica entre os ângulos de torção ϕ_i , τ_{C^{i-1}, H_k^j} , τ_{C^i, H_α^i} e $\tau_{H_k^j, H_\alpha^i}$. Os casos apresentados se distinguem com relação a orientação e ao intervalo considerado do ângulo $\tau_{H_k^j, H_\alpha^i}$. Estes intervalos são dados com relação as quantidades descritas na [Figura 32](#). Desta maneira, considerando as relações apresentadas na [Figura 33](#), podemos escrever o seguinte lema:

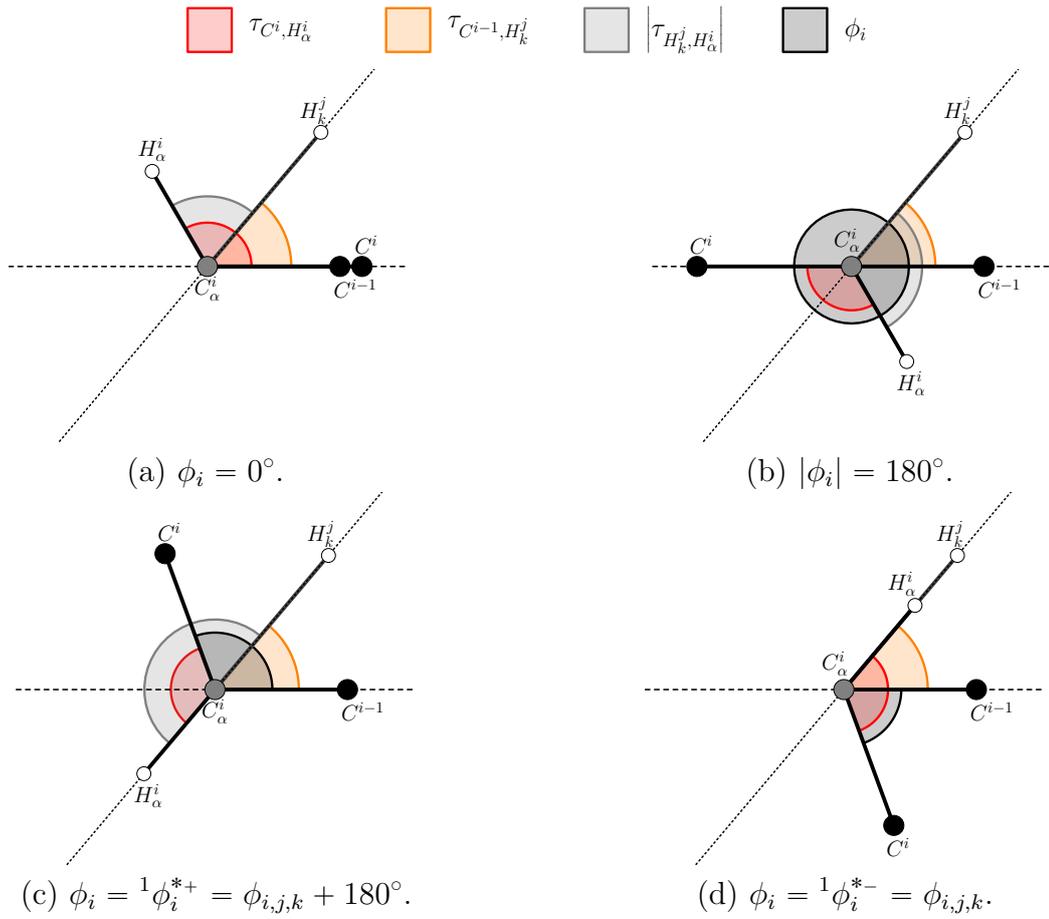


Figura 32 – Disposição no plano da projeção ortogonal dos átomos H_k^j , C^{i-1} , C_α^i , H_α^i e C^i , considerando $\phi_i = 0^\circ$, $|\phi_i| = 180^\circ$, $\phi_i = {}^1\phi_i^{*+}$ e $\phi_i = {}^1\phi_i^{*-}$, para o Caso 1.

Lema 4.1. Sejam ϕ_i , τ_{C^{i-1}, H_k^j} , τ_{C^i, H_α^i} , $\tau_{H_k^j, H_\alpha^i}$ os ângulos apresentados na Figura 33. Considere que os valores de τ_{C^i, H_α^i} e τ_{C^{i-1}, H_k^j} são conhecidos de modo que $0^\circ \leq \tau_{C^{i-1}, H_k^j} < \tau_{C^i, H_\alpha^i}$ e o valor absoluto do ângulo $\tau_{H_k^j, H_\alpha^i}$ é conhecido. Então, o ângulo $\phi_i \in (-180^\circ, 180^\circ]$ pode ser escrito de duas formas, dadas por ϕ_i^+ e ϕ_i^- , em função dos outros três ângulos dados, da seguinte maneira:

$$\phi_i^+ = \phi_{i,j,k} + \left| \tau_{H_k^j, H_\alpha^i} \right|.$$

$$\phi_i^- = \begin{cases} \phi_{i,j,k} - \left| \tau_{H_k^j, H_\alpha^i} \right| & \text{se } \left| \phi_{i,j,k} - \left| \tau_{H_k^j, H_\alpha^i} \right| \right| < 180^\circ, \\ \phi_{i,j,k} - \left| \tau_{H_k^j, H_\alpha^i} \right| + 360^\circ & \text{se } \phi_{i,j,k} - \left| \tau_{H_k^j, H_\alpha^i} \right| \leq -180^\circ. \end{cases}$$

Demonstração. Para demonstrar este lema, vamos observar as 7 possibilidades de se descrever o ângulo de torção ϕ_i descrita pela relação (4.20). Para isso, vamos dividir esta demonstração observando a orientação do ângulo $\tau_{H_k^j, H_\alpha^i}$. Assim:

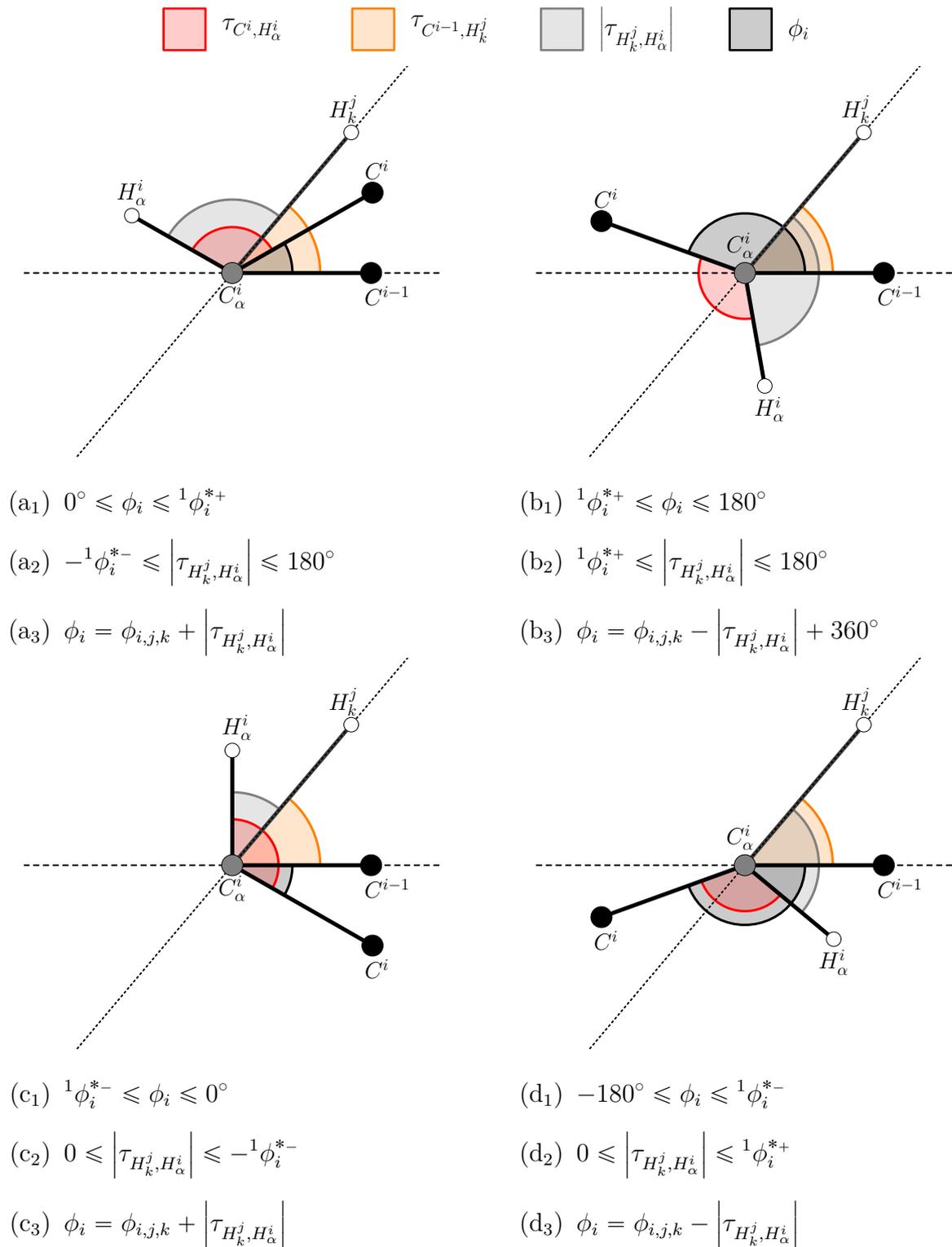


Figura 33 – Dinâmica do ângulo ϕ_i com relação aos ângulos τ_{C^{i-1}, H_k^j} , τ_{C^i, H_α^i} e $|\tau_{H_k^j, H_\alpha^i}|$ para o Caso 1.

- Se $\tau_{H_k^j, H_\alpha^i} > 0$, então:

$$\begin{aligned} 0^\circ &\leq \tau_{H_k^j, H_\alpha^i} \leq 180^\circ && \Leftrightarrow \\ 0^\circ &\leq \left| \tau_{H_k^j, H_\alpha^i} \right| \leq 180^\circ && \Leftrightarrow \\ \phi_{i,j,k} &\leq \phi_{i,j,k} + \left| \tau_{H_k^j, H_\alpha^i} \right| \leq \phi_{i,j,k} + 180^\circ && \Leftrightarrow \\ {}^1\phi_i^{*-} &\leq \phi_{i,j,k} + \left| \tau_{H_k^j, H_\alpha^i} \right| \leq {}^1\phi_i^{*+}. \end{aligned}$$

No entanto, das desigualdades descritas em (4.22) observamos que $-180^\circ < {}^1\phi_i^{*-}$ e que ${}^1\phi_i^{*+} < 180^\circ$. Portanto, $\left| \phi_{i,j,k} + \left| \tau_{H_k^j, H_\alpha^i} \right| \right| < 180^\circ$. Logo, da relação (4.20) concluímos que:

$$\phi_i^+ = \phi_{i,j,k} + \left| \tau_{H_k^j, H_\alpha^i} \right|.$$

- Se $\tau_{H_k^j, H_\alpha^i} < 0$, então:

$$-180^\circ \leq \tau_{H_k^j, H_\alpha^i} \leq 0^\circ \Rightarrow \begin{cases} 1) & -{}^1\phi_i^{*+} < \tau_{H_k^j, H_\alpha^i} \leq 0^\circ \\ 2) & \tau_{H_k^j, H_\alpha^i} = -{}^1\phi_i^{*+} \\ 3) & -180^\circ \leq \tau_{H_k^j, H_\alpha^i} < {}^1\phi_i^{*+} \end{cases}, \text{ desta maneira:}$$

$$1) \begin{cases} -{}^1\phi_i^{*+} < \tau_{H_k^j, H_\alpha^i} \leq 0^\circ & \Leftrightarrow \\ -{}^1\phi_i^{*+} < -\left| \tau_{H_k^j, H_\alpha^i} \right| \leq 0^\circ & \Leftrightarrow \\ \phi_{i,j,k} - {}^1\phi_i^{*+} < \phi_{i,j,k} - \left| \tau_{H_k^j, H_\alpha^i} \right| \leq \phi_{i,j,k} & \Leftrightarrow \\ -180^\circ < \phi_{i,j,k} - \left| \tau_{H_k^j, H_\alpha^i} \right| \leq {}^1\phi_i^{*-} \end{cases}$$

$$2) \begin{cases} \tau_{H_k^j, H_\alpha^i} = -{}^1\phi_i^{*+} & \Leftrightarrow \\ -\left| \tau_{H_k^j, H_\alpha^i} \right| = -{}^1\phi_i^{*+} & \Leftrightarrow \\ \phi_{i,j,k} - \left| \tau_{H_k^j, H_\alpha^i} \right| = -180^\circ \end{cases}$$

$$3) \begin{cases} -180^\circ \leq \tau_{H_k^j, H_\alpha^i} < -{}^1\phi_i^{*+} & \Leftrightarrow \\ -180^\circ \leq -\left| \tau_{H_k^j, H_\alpha^i} \right| < -{}^1\phi_i^{*+} & \Leftrightarrow \\ \phi_{i,j,k} - 180^\circ \leq \phi_{i,j,k} - \left| \tau_{H_k^j, H_\alpha^i} \right| < -180^\circ \end{cases}$$

Logo, das desigualdades descritas em (4.22) observamos que:

1. $\left| \phi_{i,j,k} - \left| \tau_{H_k^j, H_\alpha^i} \right| \right| < 180^\circ$;
2. $\left| \phi_{i,j,k} - \left| \tau_{H_k^j, H_\alpha^i} \right| \right| = 180^\circ$;
3. $\phi_{i,j,k} - \left| \tau_{H_k^j, H_\alpha^i} \right| < -180^\circ$.

Portanto, da relação (4.20) concluímos que:

$$\phi_i^- = \begin{cases} \phi_{i,j,k} - |\tau_{H_k^j, H_\alpha^i}| & \text{se } |\phi_{i,j,k} - \tau_{H_k^j, H_\alpha^i}| < 180^\circ, \\ \phi_{i,j,k} - |\tau_{H_k^j, H_\alpha^i}| + 360^\circ & \text{se } \phi_{i,j,k} - |\tau_{H_k^j, H_\alpha^i}| \leq -180^\circ. \end{cases}$$

□

Caso 2: $\tau_{C^i, H_\alpha^i} \leq \tau_{C^{i-1}, H_k^j} \leq 180^\circ$

Para este caso, a Figura 34 evidencia a disposição dos átomos H_k^j , C^{i-1} , C_α^i , H_α^i e C^i para as seguintes situações: $\phi_i = 0^\circ$, $|\phi_i| = 180^\circ$, $\phi_i = {}^2\phi_i^{*+}$ e $\phi_i = {}^2\phi_i^{*-}$.

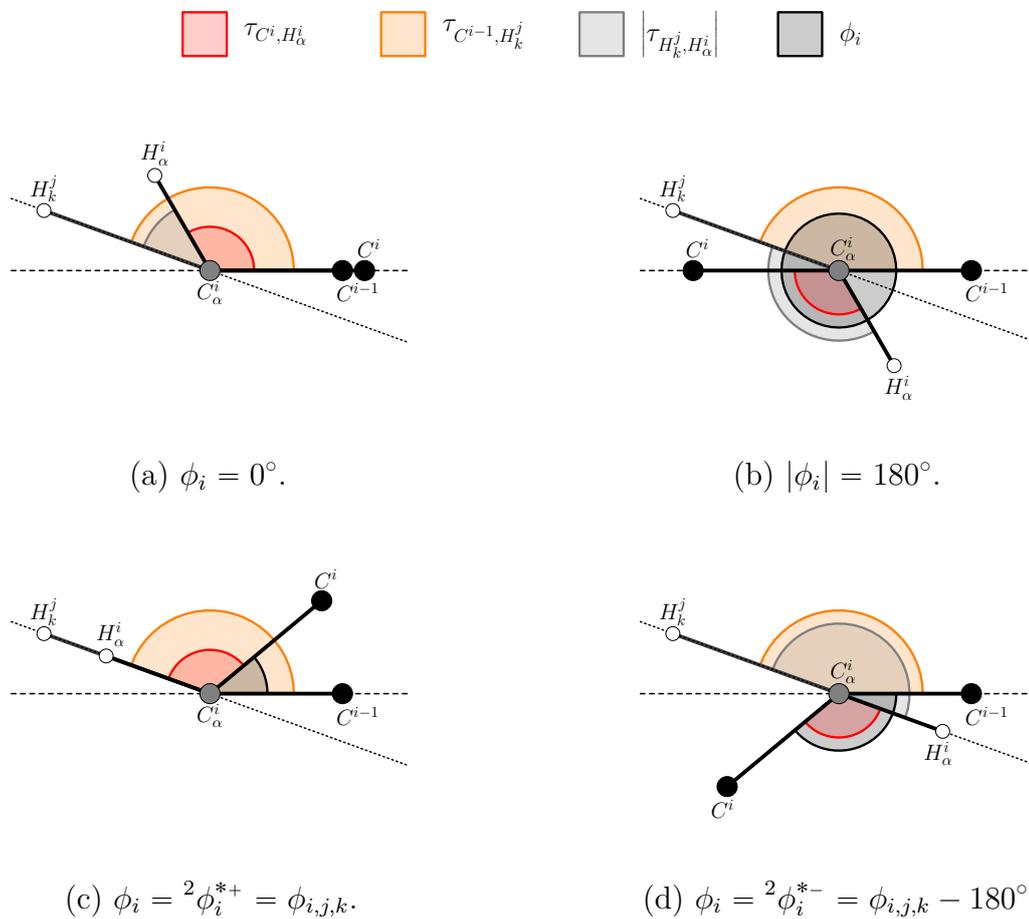


Figura 34 – Disposição no plano da projeção ortogonal dos átomos H_k^j , C^{i-1} , C_α^i , H_α^i e C^i , considerando $\phi_i = 0^\circ$, $|\phi_i| = 180^\circ$, $\phi_i = {}^2\phi_i^{*+}$ e $\phi_i = {}^2\phi_i^{*-}$, para o Caso 2.

Sobre as quantidades ${}^2\phi_i^{*-} = \phi_{i,j,k} - 180^\circ$ e ${}^2\phi_i^{*+} = {}^2\phi_i^{*-} + 180^\circ$, observamos que:

$$\begin{aligned}
 \tau_{C^i, H_\alpha^i} &\leq \tau_{C^{i-1}, H_k^j} \leq 180^\circ && \Leftrightarrow \\
 0^\circ &\leq \phi_{i,j,k} \leq 180^\circ - \tau_{C^i, H_\alpha^i} && \Leftrightarrow \\
 -180^\circ &\leq {}^2\phi_i^{*-} \leq -\tau_{C^i, H_\alpha^i} && \Leftrightarrow \\
 0^\circ &\leq {}^2\phi_i^{*+} \leq 180^\circ - \tau_{C^i, H_\alpha^i}. &&
 \end{aligned} \tag{4.23}$$

Como no caso anterior, pela intuição do desenho geométrico, na [Figura 35](#) sistematizamos as relações que descrevem a dinâmica entre os ângulos de torção ϕ_i , τ_{C^{i-1}, H_k^j} , τ_{C^i, H_α^i} e $\tau_{H_k^j, H_\alpha^i}$.

Assim, considerando as relações apresentadas na [Figura 35](#), podemos escrever o seguinte lema:

Lema 4.2. Sejam ϕ_i , τ_{C^{i-1}, H_k^j} , τ_{C^i, H_α^i} , $\tau_{H_k^j, H_\alpha^i}$ os ângulos apresentados na [Figura 35](#). Considere que os valores de τ_{C^i, H_α^i} e τ_{C^{i-1}, H_k^j} são conhecidos de modo que $0^\circ < \tau_{C^i, H_\alpha^i} \leq \tau_{C^{i-1}, H_k^j} \leq 180^\circ$ e o valor absoluto do ângulo $\tau_{H_k^j, H_\alpha^i}$ é conhecido. Então, o ângulo $\phi_i \in (-180^\circ, 180^\circ]$ pode ser escrito de duas formas, dadas por ϕ_i^+ e ϕ_i^- , em função dos outros três ângulos dados, da seguinte maneira:

$$\begin{aligned}
 \phi_i^+ &= \begin{cases} \phi_{i,j,k} + \left| \tau_{H_k^j, H_\alpha^i} \right| & \text{se } \left| \phi_{i,j,k} + \left| \tau_{H_k^j, H_\alpha^i} \right| \right| \leq 180^\circ, \\ \phi_{i,j,k} + \left| \tau_{H_k^j, H_\alpha^i} \right| - 360^\circ & \text{se } \left| \phi_{i,j,k} + \left| \tau_{H_k^j, H_\alpha^i} \right| \right| > 180^\circ. \end{cases} \\
 \phi_i^- &= \phi_{i,j,k} - \left| \tau_{H_k^j, H_\alpha^i} \right|.
 \end{aligned}$$

Demonstração. A demonstração deste lema pode ser feita de maneira análoga a do [Lema 4.1](#). No entanto, neste caso, consideramos as desigualdades de [\(4.23\)](#) para observar o valor de $\phi_{i,j,k} \pm \left| \tau_{H_k^j, H_\alpha^i} \right|$. \square

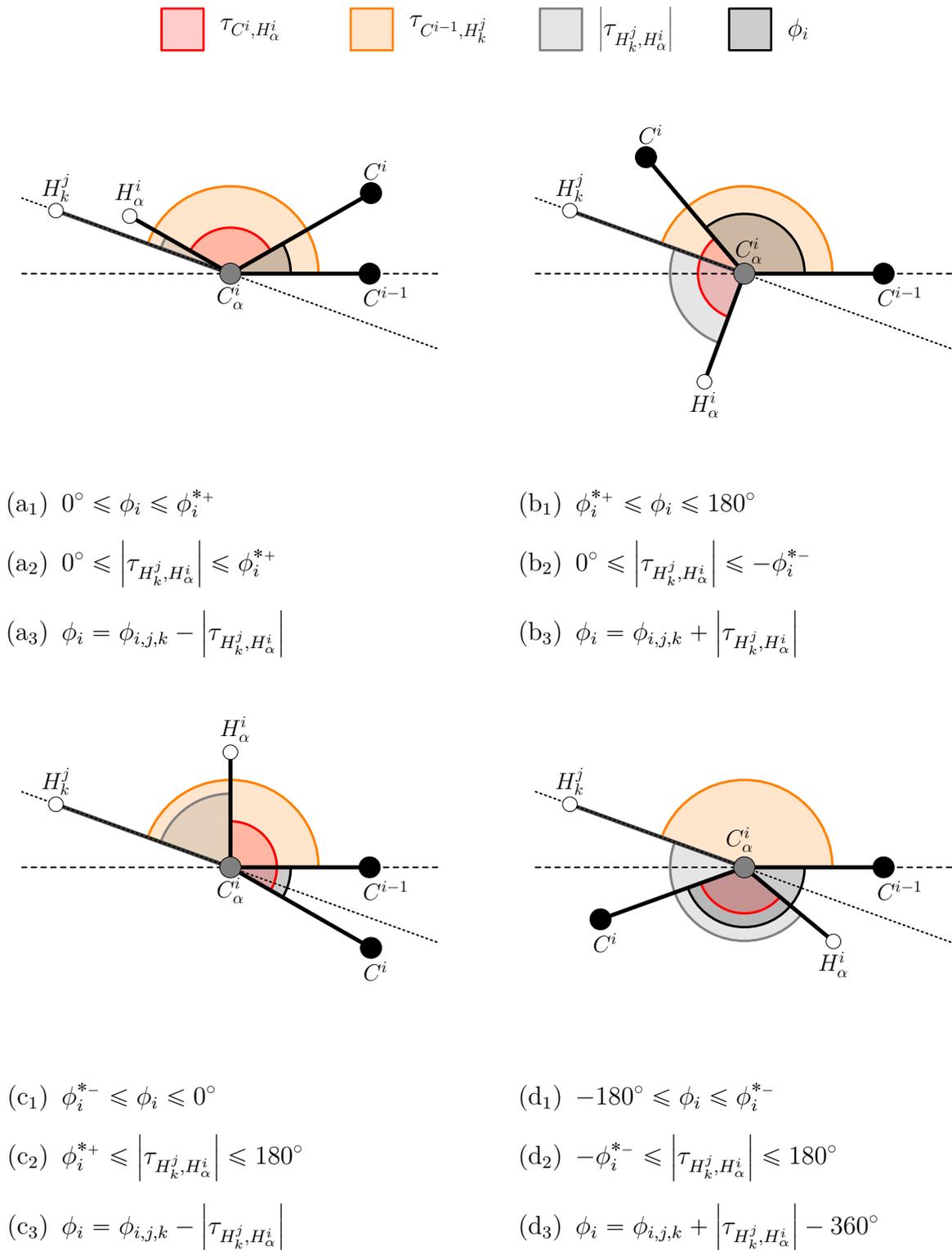


Figura 35 – Dinâmica do ângulo ϕ_i com relação aos ângulos τ_{C^{i-1}, H_k^j} , τ_{C^i, H_α^i} e $|\tau_{H_k^j, H_\alpha^i}|$ para o Caso 2.

Caso 3: $-180^\circ + \tau_{C^i, H_\alpha^i} < \tau_{C^{i-1}, H_k^j} < 0^\circ$

Para este caso, a [Figura 36](#) evidencia a disposição dos átomos H_k^j , C^{i-1} , C_α^i , H_α^i e C^i para as seguintes situações: $\phi_i = 0^\circ$, $|\phi_i| = 180^\circ$, $\phi_i = {}^3\phi_i^{*+}$ e $\phi_i = {}^3\phi_i^{*-}$.

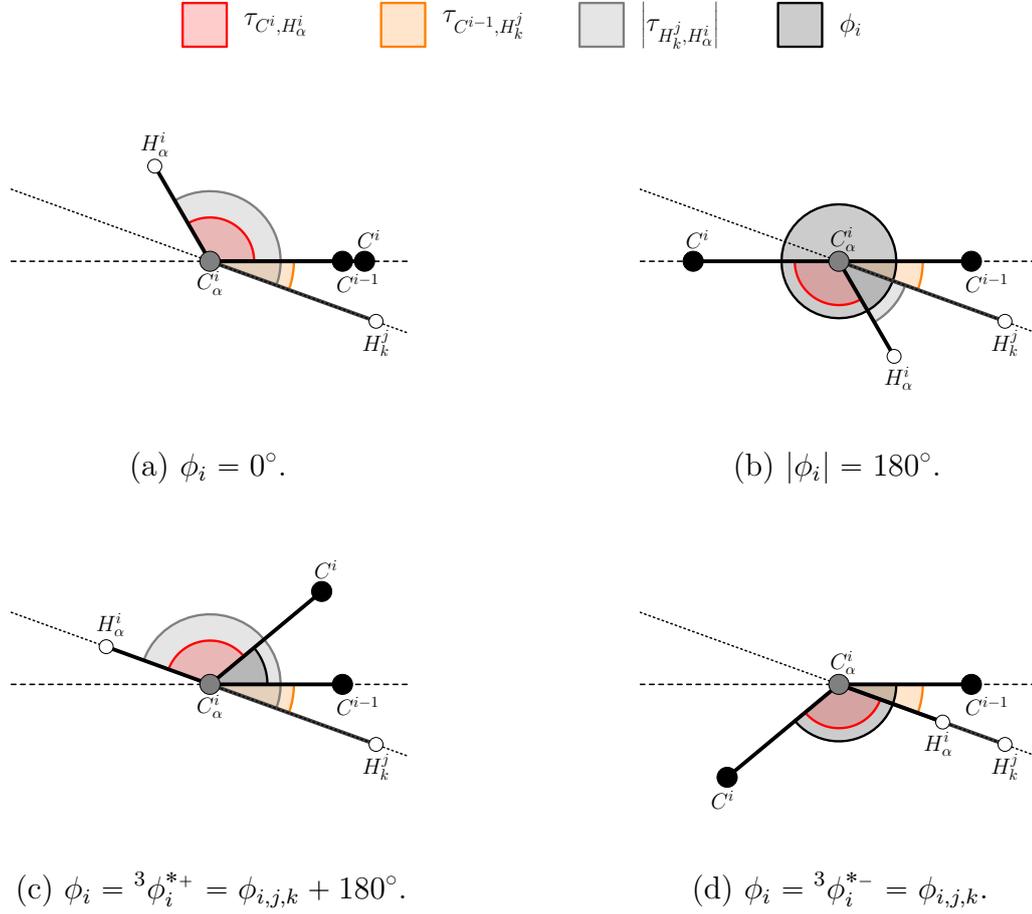


Figura 36 – Disposição no plano da projeção ortogonal dos átomos H_k^j , C^{i-1} , C_α^i , H_α^i e C^i , considerando $\phi_i = 0^\circ$, $|\phi_i| = 180^\circ$, $\phi_i = {}^3\phi_i^{*+}$ e $\phi_i = {}^3\phi_i^{*-}$, para o Caso 3.

Sobre as quantidades ${}^3\phi_i^{*-} = \phi_{i,j,k}$ e ${}^3\phi_i^{*+} = {}^3\phi_i^{*-} + 180^\circ$, observamos que:

$$\begin{aligned}
 -180^\circ + \tau_{C^i, H_\alpha^i} &< \tau_{C^{i-1}, H_k^j} < 0^\circ && \Leftrightarrow \\
 -180^\circ &< \phi_{i,j,k} < -\tau_{C^i, H_\alpha^i} && \Leftrightarrow \\
 -180^\circ &< {}^3\phi_i^{*-} < -\tau_{C^i, H_\alpha^i} && \Leftrightarrow \\
 0^\circ &< {}^3\phi_i^{*+} < 180^\circ - \tau_{C^i, H_\alpha^i}. &&
 \end{aligned} \tag{4.24}$$

Como no caso anterior, pela intuição do desenho geométrico, na [Figura 37](#) sistematizamos as relações que descrevem a dinâmica entre os ângulos de torção ϕ_i , τ_{C^{i-1}, H_k^j} , τ_{C^i, H_α^i} e $\tau_{H_k^j, H_\alpha^i}$.

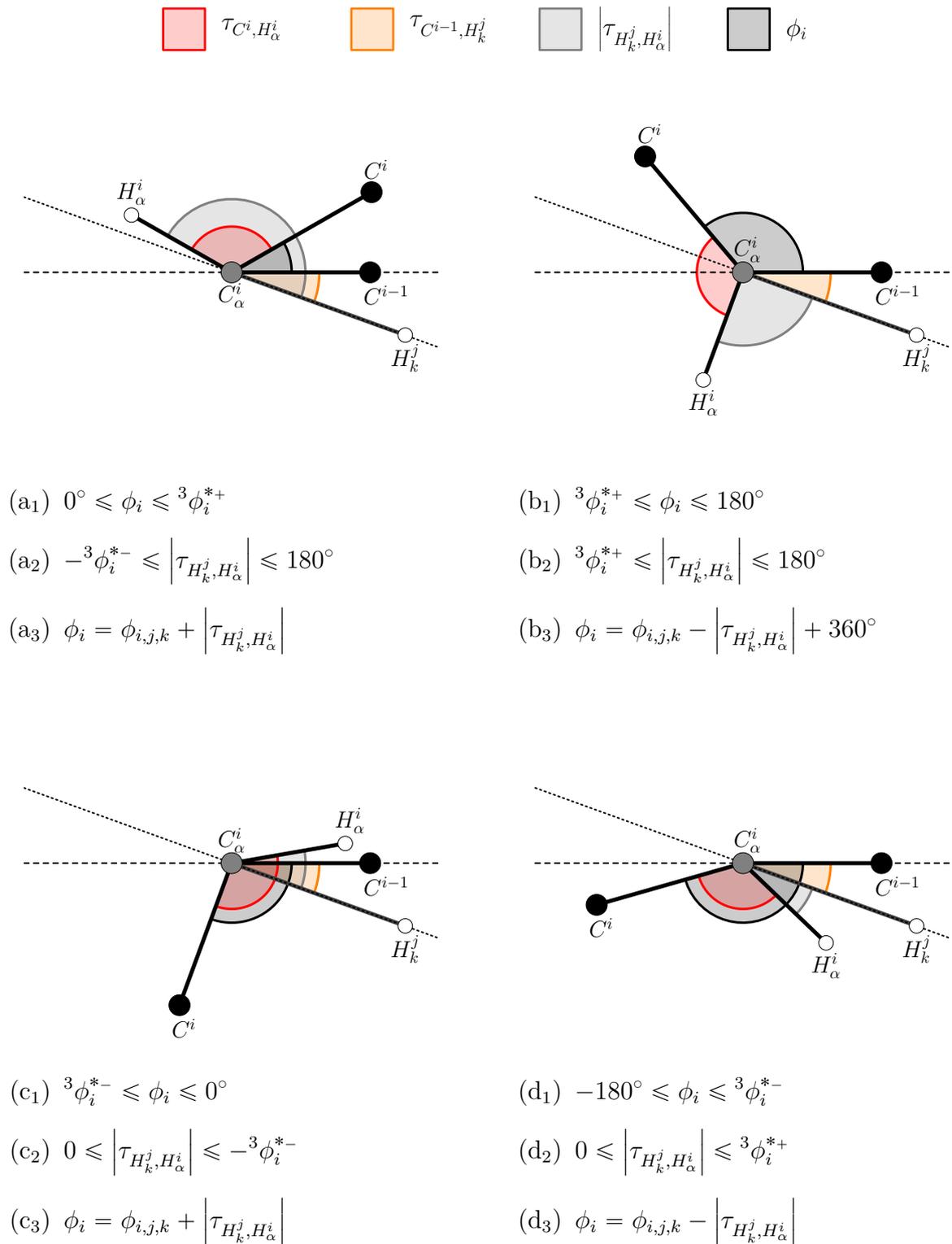


Figura 37 – Dinâmica do ângulo ϕ_i com relação aos ângulos τ_{C^{i-1}, H_k^j} , τ_{C^i, H_α^i} e $|\tau_{H_k^j, H_\alpha^i}|$ para o Caso 3.

Assim, considerando as relações apresentadas na [Figura 37](#), podemos escrever o seguinte lema:

Lema 4.3. Sejam ϕ_i , τ_{C^{i-1}, H_k^j} , τ_{C^i, H_α^i} , $\tau_{H_k^j, H_\alpha^i}$ os ângulos apresentados na [Figura 37](#). Considere que os valores de τ_{C^i, H_α^i} e τ_{C^{i-1}, H_k^j} são conhecidos de modo que $\tau_{C^i, H_\alpha^i} > 0^\circ$ e $-180^\circ + \tau_{C^i, H_\alpha^i} < \tau_{C^{i-1}, H_k^j} < 0^\circ$; e o valor absoluto do ângulo $\tau_{H_k^j, H_\alpha^i}$ é conhecido. Então, o ângulo $\phi_i \in (-180^\circ, 180^\circ]$ pode ser escrito de duas formas, dadas por ϕ_i^+ e ϕ_i^- , em função dos outros três ângulos dados, da seguinte maneira:

$$\begin{aligned} \phi_i^+ &= \phi_{i,j,k} + \left| \tau_{H_k^j, H_\alpha^i} \right| \\ \phi_i^- &= \begin{cases} \phi_{i,j,k} - \left| \tau_{H_k^j, H_\alpha^i} \right| & \text{se } \left| \phi_{i,j,k} - \left| \tau_{H_k^j, H_\alpha^i} \right| \right| < 180^\circ, \\ \phi_{i,j,k} - \left| \tau_{H_k^j, H_\alpha^i} \right| + 360^\circ & \text{se } \left| \phi_{i,j,k} - \left| \tau_{H_k^j, H_\alpha^i} \right| \right| \leq -180^\circ. \end{cases} \end{aligned}$$

Demonstração. A demonstração deste lema pode ser feita de maneira análoga a do [Lema 4.1](#). No entanto, neste caso, consideramos as desigualdades de [\(4.24\)](#) para observar o valor de $\phi_{i,j,k} \pm \left| \tau_{H_k^j, H_\alpha^i} \right|$. \square

Pelo resultado observado nos [Lemas 4.1](#) e [4.3](#), notamos que a relação entre os ângulos de torção analisados é a mesma se $-180^\circ + \tau_{C^i, H_\alpha^i} < \tau_{C^{i-1}, H_k^j} < \tau_{C^i, H_\alpha^i}$. A justificativa para esta observação reside no fato de que os valores ${}^1\phi_i^{*\pm}$ e ${}^3\phi_i^{*\pm}$ são escritos em função da quantidade $\phi_{i,j,k}$ da mesma maneira e que a orientação do ângulo $\tau_{H_k^j, H_\alpha^i}$ é mesma se $\phi_i \in [{}^m\phi_i^{*-}, {}^m\phi_i^{*+}]$ ou $\phi_i \in [-180^\circ, {}^m\phi_i^{*-}] \cup [{}^m\phi_i^{*+}, 180^\circ]$, com $m = 1, 3$.

Caso 4: $-180^\circ < \tau_{C^{i-1}, H_k^j} \leq -180^\circ + \tau_{C^i, H_\alpha^i}$

Para este caso, a [Figura 38](#) evidencia a disposição dos átomos H_k^j , C^{i-1} , C^i , H_α^i e C^i para as seguintes situações: $\phi_i = 0^\circ$, $|\phi_i| = 180^\circ$, $\phi_i = {}^4\phi_i^{*+}$ e $\phi_i = {}^4\phi_i^{*-}$.

Sobre as quantidades ${}^4\phi_i^{*-} = \phi_{i,j,k} + 180^\circ$ e ${}^4\phi_i^{*+} = {}^4\phi_i^{*-} + 180^\circ$, observamos que:

$$\begin{aligned} -180^\circ < \tau_{C^{i-1}, H_k^j} &\leq -180^\circ + \tau_{C^i, H_\alpha^i} && \Leftrightarrow \\ -180^\circ - \tau_{C^i, H_\alpha^i} < \phi_{i,j,k} &\leq -180^\circ && \Leftrightarrow \\ -\tau_{C^i, H_\alpha^i} < {}^4\phi_i^{*-} &\leq 0^\circ && \Leftrightarrow \\ 180^\circ - \tau_{C^i, H_\alpha^i} < {}^4\phi_i^{*+} &\leq 180^\circ. \end{aligned} \tag{4.25}$$

Como no caso anterior, pela intuição do desenho geométrico, na [Figura 39](#) sistematizamos as relações que descrevem a dinâmica entre os ângulos de torção ϕ_i , τ_{C^{i-1}, H_k^j} , τ_{C^i, H_α^i} e $\tau_{H_k^j, H_\alpha^i}$.

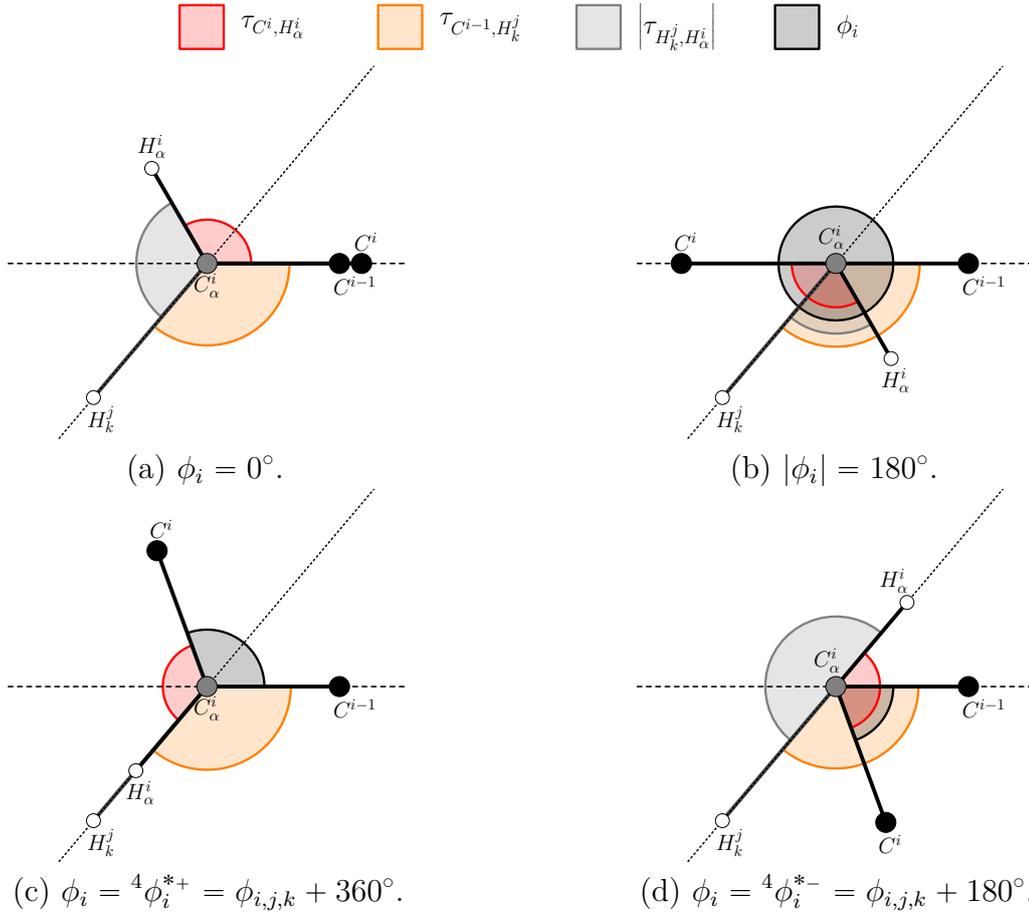


Figura 38 – Disposição no plano da projeção ortogonal dos átomos H_k^j , C^{i-1} , C_α^i , H_α^i e C^i , considerando $\phi_i = 0^\circ$, $|\phi_i| = 180^\circ$, $\phi_i = 4\phi_i^{*+}$ e $\phi_i = 4\phi_i^{*-}$, para o Caso 4.

Assim, considerando as relações apresentadas na Figura 39, podemos escrever o seguinte lema:

Lema 4.4. Sejam ϕ_i , τ_{C^{i-1}, H_k^j} , τ_{C^i, H_α^i} , $\tau_{H_k^j, H_\alpha^i}$ os ângulos apresentados na Figura 39. Considere que os valores de τ_{C^i, H_α^i} e τ_{C^{i-1}, H_k^j} são conhecidos de modo que $\tau_{C^i, H_\alpha^i} > 0^\circ$ e $-180^\circ < \tau_{C^{i-1}, H_k^j} \leq -180^\circ + \tau_{C^i, H_\alpha^i}$; e o valor absoluto do ângulo $\tau_{H_k^j, H_\alpha^i}$ é conhecido. Então, o ângulo $\phi_i \in (-180^\circ, 180^\circ]$ pode ser escrito de duas formas, dadas por ϕ_i^+ e ϕ_i^- , em função dos outros três ângulos dados, da seguinte maneira:

$$\phi_i^+ = \begin{cases} \phi_{i,j,k} + |\tau_{H_k^j, H_\alpha^i}| & \text{se } \left| \phi_{i,j,k} + |\tau_{H_k^j, H_\alpha^i}| \right| < 180^\circ, \\ \phi_{i,j,k} + |\tau_{H_k^j, H_\alpha^i}| + 360^\circ & \text{se } \left| \phi_{i,j,k} + |\tau_{H_k^j, H_\alpha^i}| \right| \leq -180^\circ. \end{cases}$$

$$\phi_i^- = \phi_{i,j,k} - |\tau_{H_k^j, H_\alpha^i}| + 360^\circ.$$

Demonstração. A demonstração deste lema pode ser feita de maneira análoga a do Lema 4.1. No entanto, neste caso, consideramos as desigualdades de (4.25) para observar o valor de $\phi_{i,j,k} \pm |\tau_{H_k^j, H_\alpha^i}|$. □

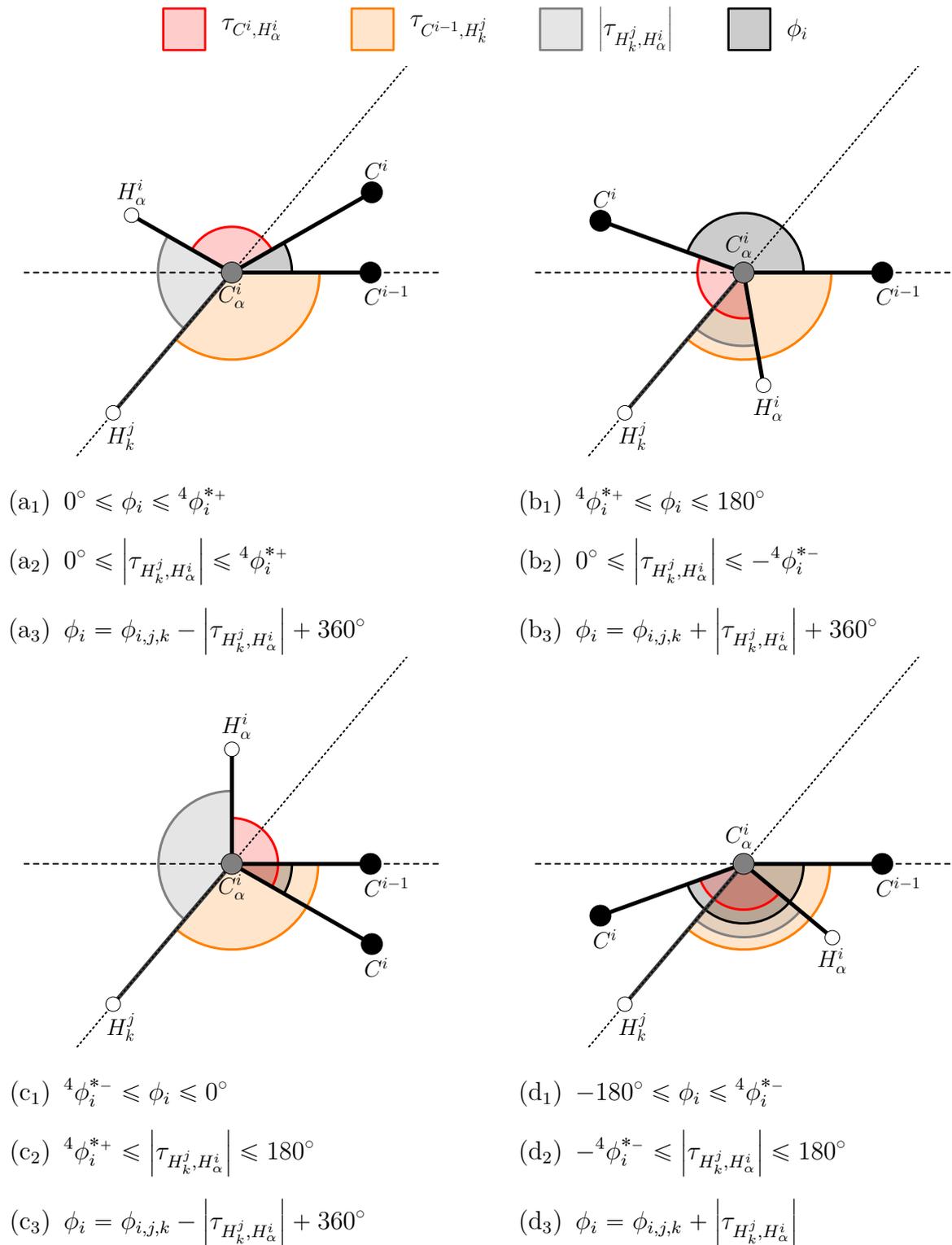


Figura 39 – Dinâmica do ângulo ϕ_i com relação aos ângulos τ_{C^{i-1}, H_k^j} , τ_{C^i, H_α^i} e $|\tau_{H_k^j, H_\alpha^i}|$ para o Caso 4.

Agora, para sistematizar os 4 casos considerados nesta subseção, a proposição 4.1 apresenta uma relação geral que descreve o ângulo ϕ_i em função dos ângulos de torção τ_{C^{i-1}, H_k^j} , τ_{C^i, H_α^i} e $\tau_{H_k^j, H_\alpha^i}$.

Proposição 4.1. Sejam R_i e R_j resíduos de aminoácido que compõe uma proteína, com $j \leq i$, $\phi_i \equiv \{C^{i-1}, N^i, C_\alpha^i, C^i\}$, $\tau_{C^{i-1}, H_k^j} \equiv \{C^{i-1}, N^i, C_\alpha^i, H_k^j\}$, $\tau_{C^i, H_\alpha^i} \equiv \{C^i, N^i, C_\alpha^i, H_\alpha^i\}$ e $\tau_{H_k^j, H_\alpha^i} \equiv \{H_k^j, N^i, C_\alpha^i, H_\alpha^i\}$ ângulos de torção definidos em átomos destes resíduos. Considere que os valores de τ_{C^i, H_α^i} e τ_{C^{i-1}, H_k^j} são conhecidos de modo que $\tau_{C^i, H_\alpha^i} > 0^\circ$ e o valor absoluto do ângulo $\tau_{H_k^j, H_\alpha^i}$ é conhecido. Então, existem duas possibilidades para se escrever o ângulo $\phi_i \in (-180^\circ, 180^\circ]$, dadas por ϕ_i^+ e ϕ_i^- , em função dos outros três ângulos de torção, que são descritas da seguinte maneira:

$$\phi_i^+ = c_{\phi_i}(\sigma_{i,j,k}^+) + \sigma_{i,j,k}^+ \quad \text{e} \quad \phi_i^- = c_{\phi_i}(\sigma_{i,j,k}^-) + \sigma_{i,j,k}^-, \quad (4.26)$$

onde,

$$\sigma_{i,j,k}^\pm \equiv \phi_{i,j,k} \pm \left| \tau_{H_k^j, H_\alpha^i} \right|, \quad \phi_{i,j,k} \equiv \tau_{C^{i-1}, H_k^j} - \tau_{C^i, H_\alpha^i},$$

$$c_{\phi_i}(\sigma) \equiv \begin{cases} 0 & \text{se } |\sigma| < 180^\circ, \\ 360^\circ \cdot (-1)^{m+1} & \text{se } |\sigma| > 180^\circ, \\ 180^\circ - \sigma & \text{se } |\sigma| = 180^\circ, \end{cases}$$

e

$$m \equiv \begin{cases} 1 & \text{se } -180^\circ + \tau_{C^i, H_\alpha^i} < \tau_{C^{i-1}, H_k^j} < \tau_{C^i, H_\alpha^i}, \\ 2 & \text{se } \tau_{C^i, H_\alpha^i} \leq \tau_{C^{i-1}, H_k^j} \leq 180^\circ, \\ 3 & \text{se } -180 < \tau_{C^{i-1}, H_k^j} \leq -180^\circ + \tau_{C^i, H_\alpha^i}. \end{cases}$$

Demonstração. Este resultado foi construído diretamente dos 4 casos descritos anteriormente nesta subseção. No entanto, neste caso, condensamos a notação utilizada. Desta forma, sua demonstração segue direto dos Lemas 4.1, 4.2, 4.3 e 4.4. \square

O resultado apresentado na proposição acima descreve uma relação entre ângulos de torção, no entanto, valores absolutos dos ângulos τ_{C^{i-1}, H_k^j} , τ_{C^i, H_α^i} , $\tau_{H_k^j, H_\alpha^i}$ podem ser descritos pela relação (3.42) utilizando restrições de distâncias entre os átomos que compõem estes ângulos. A orientação de τ_{C^i, H_α^i} e τ_{C^{i-1}, H_k^j} é tomada via quiralidade local. Já o sinal de τ_{C^{i-1}, H_k^j} , com $j < i$, pode ser calculado pela Proposição 3.5. Assim, podemos dizer que este resultado descreve os ângulos de torção ϕ_i em função de restrições de distâncias de átomos ao entorno destes ângulos.

4.3 Relação dos Ângulos de Torção ψ_i com Distâncias Entre Hidrogênios Próximos

De uma maneira similar a que procedemos na seção anterior, através de observações geométricas, vamos construir uma relação entre o ângulo de torção $\psi_i = \{N^i, C_\alpha^i, C^i, N^{i+1}\}$ e outros ângulos de torção definidos no entorno dos átomos que compõe este ângulo. Na seção anterior estudamos casos particulares mais simples que nos ajudaram a observar uma relação geral, no entanto, nesta seção não vamos apresentar uma discussão detalhada da maneira que fizemos anteriormente. Como o desenvolvimento de rotações descrito em (4.16) foi a chave para descrevermos uma relação geral que evidenciou o valor de ϕ_i , nesta seção, o objetivo inicial é descrever uma relação similar a esta para o ângulo de torção ψ_i .

A dependência direta do átomo de hidrogênio H_α^i ao ângulo de torção ϕ_i , o que é representado em (4.21), é um dos fatores que permitiu a construção apresentada em (4.16). Com sentido de observar uma dependência similar a esta para o ângulo de torção ψ_i , inicialmente, vamos observar que a condição imposta pela ligação peptídica, nos diz que:

- Se R_{i+1} não é a prolina: os átomos C_α^i, C^i, N^{i+1} e H_N^{i+1} são coplanares.
- Se R_{i+1} é a prolina: os átomos C_α^i, C^i, N^{i+1} e C_δ^{i+1} são coplanares.

Assim, a determinação das posições dos átomos C_α^i, C^i e N^{i+1} garante somente uma posição para os átomos H_N^{i+1} ou C_δ^{i+1} , assim estes dois átomos são dependentes dos três átomos anteriores. Como $\psi_i = \{N^i, C_\alpha^i, C^i, N^{i+1}\}$, observamos que os átomos H_N^{i+1} e C_δ^{i+1} são dependentes do ângulo de torção ψ_i . Para o caso em que R_{i+1} é a prolina, ainda observamos que os átomos $C_\alpha^i, C^i, N^{i+1}, C_\alpha^{i+1}$ e C_δ^{i+1} são coplanares, assim vamos considerar como precisas as distâncias do átomo de hidrogênio $H_{\delta_3}^{i+1}$ para os átomos de carbono C_α^i e C^i . Desta maneira, vamos denotar por H_{\aleph}^i , onde $\aleph = N, \delta_3$, o átomo de hidrogênio que tem relação direta com o átomo de nitrogênio N^i .

Para nos situarmos na geometria do entorno do ângulo de torção ψ_i , vamos considerar as Figuras 40 e 41, com estruturas similares as da Figura 26, que representam, respectivamente, os casos em que o resíduo de aminoácido R_{i+1} não é a prolina e é a prolina. Estes casos foram tomados com $\psi_i = 40^\circ$.

Novamente, como estamos interessados em uma relação geral entre ângulos de torção, vamos tomar H_k^j como um átomo de hidrogênio qualquer do j -ésimo resíduo de aminoácido e vamos considerar os ângulos de torção $\psi_i = \{N^i, C_\alpha^i, C^i, N^{i+1}\}$, $\tau_{N^i, H_k^j} \equiv \{N^i, C_\alpha^i, C^i, H_k^j\}$, $\tau_{N^{i+1}, H_\aleph^{i+1}} \equiv \{N^{i+1}, C_\alpha^i, C^i, H_\aleph^{i+1}\}$ e $\tau_{H_k^j, H_\aleph^{i+1}} \equiv \{H_k^j, C_\alpha^i, C^i, H_\aleph^{i+1}\}$. Da mesma maneira que procedemos na Subseção 4.2.5, vamos considerar que a estrutura proteica que define estes ângulos foi projetada ortogonalmente no plano e que estes ângulos

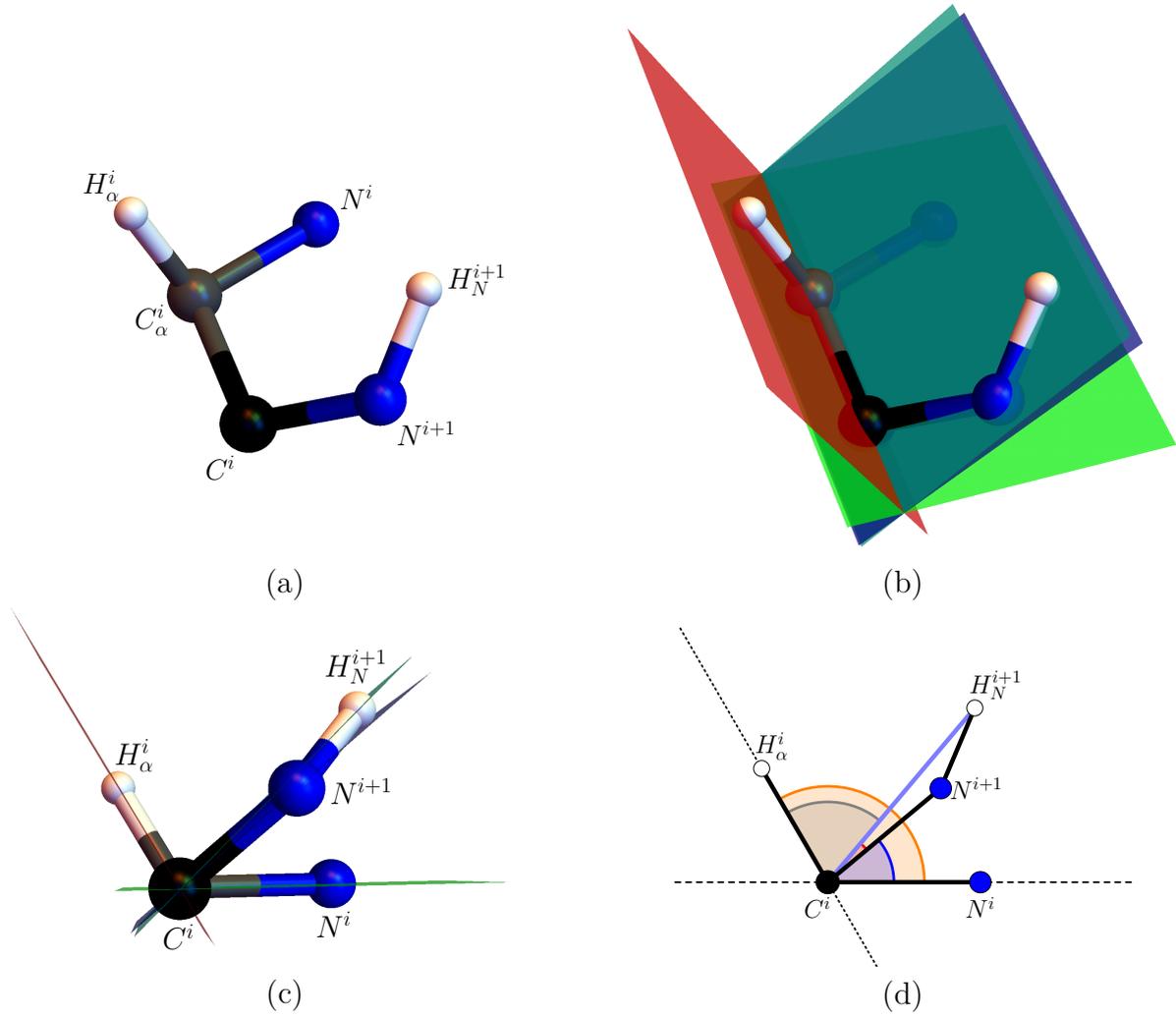


Figura 40 – (a) Uma posição espacial para os átomos N^i , C_α^i , C^i , H_α^i , N^{i+1} e H_N^{i+1} onde $\psi_i = 40^\circ$. (b) Planos $\{N^i, C_\alpha^i, C^i\}$, $\{C_\alpha^i, C^i, H_\alpha^i\}$, $\{C_\alpha^i, C^i, N^{i+1}\}$ e $\{C_\alpha^i, C^i, H_N^{i+1}\}$ identificados na estrutura da figura (a). (c) Estrutura da figura (b) rotacionada identificando um ponto de vista de interesse. (d) Projeção ortogonal da estrutura da figura (c) no plano; os ângulos de torção $\psi_i = \{N^i, C_\alpha^i, C^i, N^{i+1}\}$, $\tau_{N^i, H_\alpha^i} = \{N^i, C_\alpha^i, C^i, H_\alpha^i\}$, $\tau_{N^{i+1}, H_N^{i+1}} = \{N^{i+1}, C_\alpha^i, C^i, H_N^{i+1}\}$ e $\tau_{H_\alpha^i, H_N^{i+1}} = \{H_\alpha^i, C_\alpha^i, C^i, H_N^{i+1}\}$ são identificados como ângulos planares.

são ângulos planares. Assim, sem perda de generalidade, podemos considerar que os átomos N^i , H_α^i , N^{i+1} e H_N^{i+1} são descritos por rotações em uma circunferência de raio unitário centrada no ponto $C^i = (0, 0)$. Se $R(\theta)$ é uma matriz de rotação de ângulo θ e centro em C^i , podemos escrever:

$$\begin{aligned}
 N^{i+1} &= R(\psi_i) N^i, & H_N^{i+1} &= R(\tau_{H_k^j, H_N^{i+1}}) H_k^j, \\
 N^{i+1} &= R(-\tau_{N^{i+1}, H_N^{i+1}}) H_N^{i+1}, & H_k^j &= R(\tau_{N^i, H_k^j}) N^i.
 \end{aligned} \tag{4.27}$$

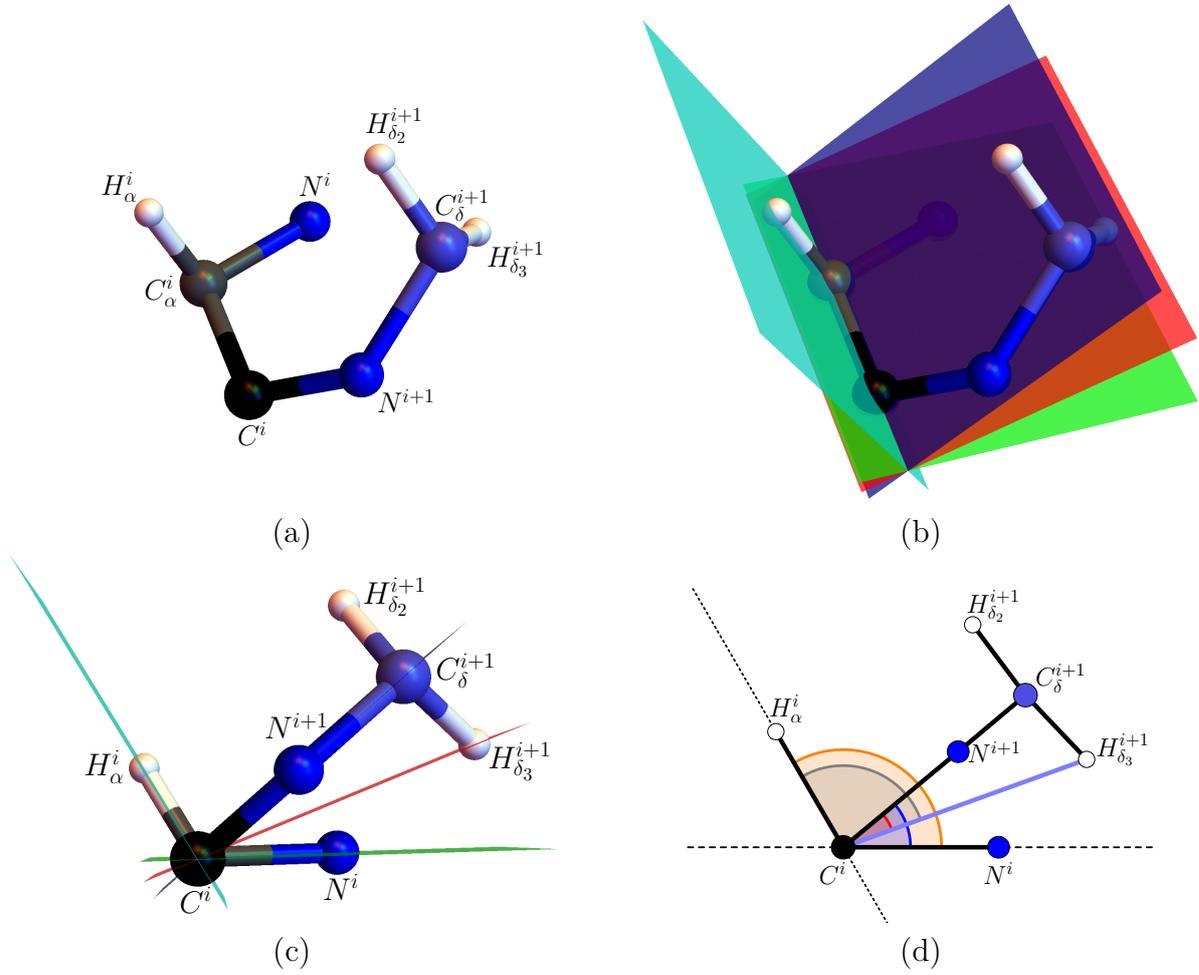


Figura 41 – (a) Uma posição espacial para os átomos N^i , C_α^i , C^i , H_α^i , N^{i+1} e $H_{\delta_3}^{i+1}$ onde $\psi_i = 40^\circ$. (b) Planos $\{N^i, C_\alpha^i, C^i\}$, $\{C_\alpha^i, C^i, H_\alpha^i\}$, $\{C_\alpha^i, C^i, N^{i+1}\}$ e $\{C_\alpha^i, C^i, H_{\delta_3}^{i+1}\}$ identificados na estrutura da figura (a). (c) Estrutura da figura (b) rotacionada identificando um ponto de vista de interesse. (d) Projeção ortogonal da estrutura da figura (c) no plano; os ângulos de torção $\psi_i = \{N^i, C_\alpha^i, C^i, N^{i+1}\}$, $\tau_{N^i, H_\alpha^i} = \{N^i, C_\alpha^i, C^i, H_\alpha^i\}$, $\tau_{N^{i+1}, H_{\delta_3}^{i+1}} = \{N^{i+1}, C_\alpha^i, C^i, H_{\delta_3}^{i+1}\}$ e $\tau_{H_\alpha^i, H_{\delta_3}^{i+1}} = \{H_\alpha^i, C_\alpha^i, C^i, H_{\delta_3}^{i+1}\}$ são identificados como ângulos planares.

Portanto, das rotações apresentadas em (4.27), concluí-se que:

$$\begin{aligned}
 R(\psi_i) N^i &= R\left(-\tau_{N^{i+1}, H_{\delta_3}^{i+1}}\right) H_{\delta_3}^{i+1} && \Leftrightarrow \\
 R(\psi_i) N^i &= R\left(-\tau_{N^{i+1}, H_{\delta_3}^{i+1}}\right) R\left(\tau_{H_k^j, H_{\delta_3}^{i+1}}\right) H_k^j && \Leftrightarrow \\
 R(\psi_i) N^i &= R\left(-\tau_{N^{i+1}, H_{\delta_3}^{i+1}}\right) R\left(\tau_{H_k^j, H_{\delta_3}^{i+1}}\right) R\left(\tau_{N^i, H_k^j}\right) N^i && \Leftrightarrow \\
 R(\psi_i) N^i &= R\left(\tau_{N^i, H_k^j} - \tau_{N^{i+1}, H_{\delta_3}^{i+1}} + \tau_{H_k^j, H_{\delta_3}^{i+1}}\right) N^i. && (4.28)
 \end{aligned}$$

Definindo a quantidade,

$$\psi_{i,j,k} \equiv \tau_{N^i, H_k^j} - \tau_{N^{i+1}, H_{\delta_3}^{i+1}}, \quad (4.29)$$

da última igualdade de (4.28), temos que:

$$\begin{cases} \cos(\psi_i) = \cos(\psi_{i,j,k} + \tau_{H_k^j, H_N^{i+1}}), \\ \sin(\psi_i) = \sin(\psi_{i,j,k} + \tau_{H_k^j, H_N^{i+1}}). \end{cases}$$

Como $\psi_i \in (-180^\circ, 180^\circ]$, analogamente ao que fizemos na seção anterior, das igualdades acima, observamos que existem as seguintes 7 possibilidades para escrevermos o ângulo de torção ψ_i :

$$\psi_i = \begin{cases} \psi_{i,j,k} \pm \left| \tau_{H_k^j, H_N^{i+1}} \right| & \text{se } \left| \psi_{i,j,k} \pm \left| \tau_{H_k^j, H_N^{i+1}} \right| \right| < 180^\circ, \\ \psi_{i,j,k} \pm \left| \tau_{H_k^j, H_N^{i+1}} \right| + 360^\circ & \text{se } \psi_{i,j,k} \pm \left| \tau_{H_k^j, H_N^{i+1}} \right| < -180^\circ, \\ \psi_{i,j,k} \pm \left| \tau_{H_k^j, H_N^{i+1}} \right| - 360^\circ & \text{se } \psi_{i,j,k} \pm \left| \tau_{H_k^j, H_N^{i+1}} \right| > 180^\circ, \\ 180^\circ & \text{se } \left| \psi_{i,j,k} \pm \left| \tau_{H_k^j, H_N^{i+1}} \right| \right| = 180^\circ. \end{cases} \quad (4.30)$$

Para identificar a orientação do ângulo de torção $\tau_{H_k^j, H_N^{i+1}}$, temos que analisar as possibilidades conforme fizemos na Seção 4.2. No entanto, neste caso, temos que observar o fato de que o ângulo $\tau_{N^{i+1}, H_N^{i+1}}$ pode ser positivo ou negativo, dois casos particulares onde isto ocorre estão apresentados nas Figuras 40 e 41, pois o sinal de $\tau_{N^{i+1}, H_N^{i+1}}$ tem efeito direto na magnitude da quantidade $\psi_{i,j,k}$.

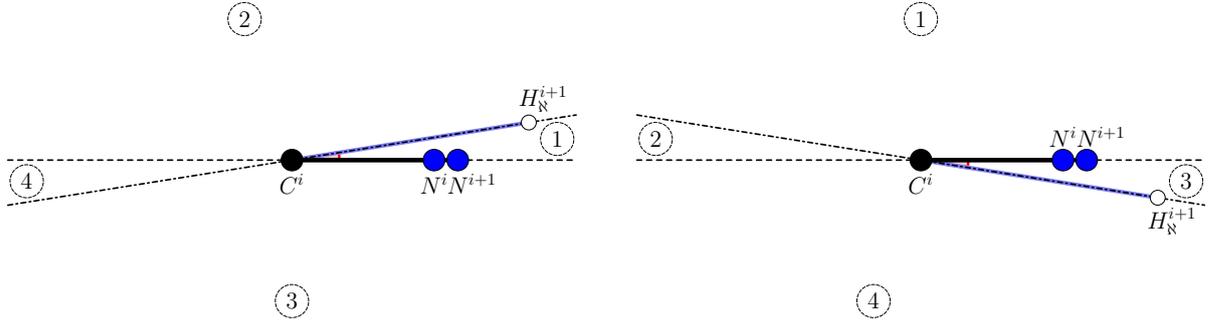
Pela interpretação de rotacionar pontos em uma circunferência unitária com centro na origem, com um raciocínio análogo ao que foi apresentado no processo descrito em (4.28), podemos escrever a posição do átomo H_N^{i+1} com relação ao ângulo ψ_i da seguinte maneira:

$$H_N^{i+1} = R(\psi_i + \tau_{N^{i+1}, H_N^{i+1}}) N^i. \quad (4.31)$$

Como ângulo τ_{N^i, H_k^j} é fixo, o átomo H_k^j possui posição fixa nesta circunferência. Então, a orientação do ângulo $\tau_{N^{i+1}, H_N^{i+1}}$ será dada conforme a posição do átomo H_N^{i+1} , que por sua vez é dependente do ângulo $\psi_i \in (-180^\circ, 180^\circ]$, como mostra a relação (4.31). A Figura 42 evidencia a disposição dos átomos N^i , C^i , H_N^{i+1} e N^{i+1} em $\psi = 0^\circ$ para duas situações: (i) $\tau_{N^{i+1}, H_N^{i+1}} \geq 0^\circ$ e (ii) $\tau_{N^{i+1}, H_N^{i+1}} < 0^\circ$.

Analogamente ao que vimos na Subseção 4.2.5, as quatro regiões do plano explicitadas nas Figura 42 (a) e (b) podem ser descritas com relação ao ângulo de torção τ_{N^i, H_k^j} da seguinte maneira:

- Situação (i): $\tau_{N^{i+1}, H_N^{i+1}} \geq 0^\circ$
 - Caso 1: $0^\circ \leq \tau_{N^i, H_k^j} \leq \tau_{N^{i+1}, H_N^{i+1}}$.
 - Caso 2: $\tau_{N^{i+1}, H_N^{i+1}} < \tau_{N^i, H_k^j} \leq 180^\circ$.



(a) Situação (i): $\tau_{N^{i+1}, H_N^{i+1}} \geq 0^\circ$.

(b) Situação (ii): $\tau_{N^{i+1}, H_N^{i+1}} < 0^\circ$.

Figura 42 – Posição dos átomos N^i , C^i , H_N^{i+1} e N^{i+1} no plano da projeção ortogonal se $\psi_i = 0^\circ$ para as situações (i) e (ii).

- Caso 3: $-180^\circ + \tau_{N^{i+1}, H_N^{i+1}} \leq \tau_{N^i, H_k^j} < 0^\circ$.
- Caso 4: $-180^\circ < \tau_{N^i, H_k^j} < -180^\circ + \tau_{N^{i+1}, H_N^{i+1}}$.
- Situação (ii): $\tau_{N^{i+1}, H_N^{i+1}} < 0^\circ$
 - Caso 1: $0^\circ \leq \tau_{N^i, H_k^j} \leq 180^\circ + \tau_{N^{i+1}, H_N^{i+1}}$.
 - Caso 2: $180^\circ + \tau_{N^{i+1}, H_N^{i+1}} < \tau_{N^i, H_k^j} \leq 180^\circ$.
 - Caso 3: $\tau_{N^{i+1}, H_N^{i+1}} \leq \tau_{N^i, H_k^j} < 0^\circ$.
 - Caso 4: $-180^\circ < \tau_{N^i, H_k^j} < \tau_{N^{i+1}, H_N^{i+1}}$.

De maneira similar a que vimos na Seção 4.2, a descrição do ângulo de torção ψ_i se dá pela análise destes casos. Por essa razão, os detalhes estão apresentados no Apêndice A. Desta maneira, com base nas Proposições A.1 e A.2, temos o seguinte resultado:

Proposição 4.2. Sejam R_j , R_i e R_{i+1} resíduos de aminoácido que compõem uma proteína, com $j \leq i$, $\psi_i \equiv \{N^i, C_\alpha^i, C^i, N^{i+1}\}$, $\tau_{N^i, H_k^j} \equiv \{N^i, C_\alpha^i, C^i, H_k^j\}$, $\tau_{N^{i+1}, H_N^{i+1}} \equiv \{N^{i+1}, C_\alpha^i, C^i, H_N^{i+1}\}$ e $\tau_{H_k^j, H_N^{i+1}} \equiv \{H_k^j, C_\alpha^i, C^i, H_N^{i+1}\}$ ângulos de torção definidos em átomos destes resíduos. Considere que os valores de $\tau_{N^{i+1}, H_N^{i+1}}$ e τ_{N^i, H_k^j} são conhecidos e o valor absoluto do ângulo $\tau_{H_k^j, H_N^{i+1}}$ é conhecido. Então, existem duas possibilidades para se escrever o ângulo $\psi_i \in (-180^\circ, 180^\circ]$, dadas por ψ_i^+ e ψ_i^- , em função dos outros três ângulos de torção, que são descritas da seguinte maneira:

$$\psi_i^+ = c_{\psi_i}(\xi_{i,j,k}^+) + \xi_{i,j,k}^+ \quad \text{e} \quad \psi_i^- = c_{\psi_i}(\xi_{i,j,k}^-) + \xi_{i,j,k}^-, \quad (4.32)$$

onde,

$$\xi_{i,j,k}^{\pm} \equiv \psi_{i,j,k} \pm \left| \tau_{H_k^j, H_{\mathbb{N}}^{i+1}} \right|, \quad \psi_{i,j,k} \equiv \tau_{N^i, H_k^j} - \tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}},$$

$$c_{\psi_i}(\xi) \equiv \begin{cases} 0 & \text{se } |\xi| < 180^\circ, \\ 360^\circ \cdot (-1)^{m+1} \left[\text{sign}^* \left(\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}} \right) \right]^{mod(m,3)} & \text{se } |\xi| > 180^\circ, \\ 180^\circ - \xi & \text{se } |\xi| = 180^\circ, \end{cases}$$

sendo $mod(m, 3)$ o resto da divisão inteira de m por 3, $\text{sign}^*(x) \equiv \begin{cases} x & \text{se } x \geq 0 \\ x & \text{se } x < 0 \end{cases}$, e

$$m \equiv \begin{cases} 1 & \text{se } \lambda_- \leq \tau_{N^i, H_k^j} \leq \lambda_+, \\ 2 & \text{se } \lambda_+ < \tau_{N^i, H_k^j} \leq 180^\circ, \\ 3 & \text{se } -180 < \tau_{N^i, H_k^j} < \lambda_-, \end{cases}$$

com

$$\begin{cases} \lambda_- = \min \left\{ \tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}}, \tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}} - \text{sign}^* \left(\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}} \right) \cdot 180^\circ \right\}, \\ \lambda_+ = \max \left\{ \tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}}, \tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}} - \text{sign}^* \left(\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}} \right) \cdot 180^\circ \right\}. \end{cases}$$

Demonstração. A demonstração deste resultado segue de maneira direta das Proposições A.1 e A.2. \square

Como os valores absolutos dos ângulos τ_{N^i, H_k^j} , $\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}}$, $\tau_{H_k^j, H_{\mathbb{N}}^{i+1}}$ podem ser descritos pela relação (3.42) utilizando restrições de distâncias entre os átomos que compõem estes ângulos. A orientação de $\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}}$ e $\tau_{H_k^j, H_{\mathbb{N}}^{i+1}}$ é tomada via quiralidade local, e o sinal de τ_{N^i, H_k^j} , com $j < i$, pode ser calculado pela Proposição 3.5. Assim, podemos dizer que este resultado descreve os ângulos de torção ψ_i em função de restrições de distâncias de átomos ao entorno destes ângulos.

No próximo capítulo, utilizamos o conjunto de hipóteses teóricas que descrevemos na Seção 4.1 e o combinamos com os resultados das Proposições 4.1 e 4.2 para apresentar uma nova ordem PDGD e um novo algoritmo para caracterizar a posição espacial do *backbone* da proteína.

CAPÍTULO

5 Caracterização da Estrutura 3D do *Backbone* de Proteínas

Neste capítulo, apresentamos a parte central desta tese. Aqui, empregamos as discussões que desenvolvemos nos capítulos anteriores para propor um novo algoritmo que determina a posição espacial dos átomos do *backbone* de proteínas. Para descrevermos este novo algoritmo, que é baseado na estratégia do *iBP* e utiliza algumas particularidades das proteínas, primeiramente vamos definir uma ordem PDGD*i* nos átomos de interesse e, assim, utilizar as relações geométricas apresentadas no capítulo anterior para estabelecermos este novo algoritmo.

5.1 Uma Nova Ordem PDGD*i* Para Grafos de Proteínas

A representação de uma molécula como um conjunto de símbolos atômicos ligados por segmentos foi originalmente descrita em [BROWN \(1864\)](#). Nesta direção, é natural a modelagem de moléculas utilizando um grafo simples, ponderado e não-direcionado $G = (V, E, \mathbf{d})$. Este grafo, como uma abstração matemática, é utilizado para representar os dados desta molécula, isto é, o conjunto de vértices V caracteriza os átomos, o conjunto das arestas E representa a coleção de pares de átomos para os quais conhece-se a distância entre eles, e a função $\mathbf{d} : E \rightarrow \mathcal{I}(\mathbb{R})$ descreve estas distâncias entre os átomos.

Novamente, o objetivo é encontrar uma posição espacial dos átomos do *backbone* de proteínas com base em um conjunto de dados que se assemelha a dados experimentais obtidos em laboratório. No entanto, como salientamos no capítulo anterior, não temos informações experimentais suficientes para considerarmos somente os átomos desta cadeia principal, e assim, devemos considerar também os átomos de hidrogênio nas imediações do *backbone*. A [Figura 43](#) evidencia o conjunto de átomos que vamos levar em conta na nossa modelagem. Na nomenclatura que definimos no capítulo anterior, $H_{\mathbb{N}}^i$ representa o átomo H_N^i se o resíduo de aminoácido R_i não é a prolina e $H_{\delta_3}^i$, caso contrário.

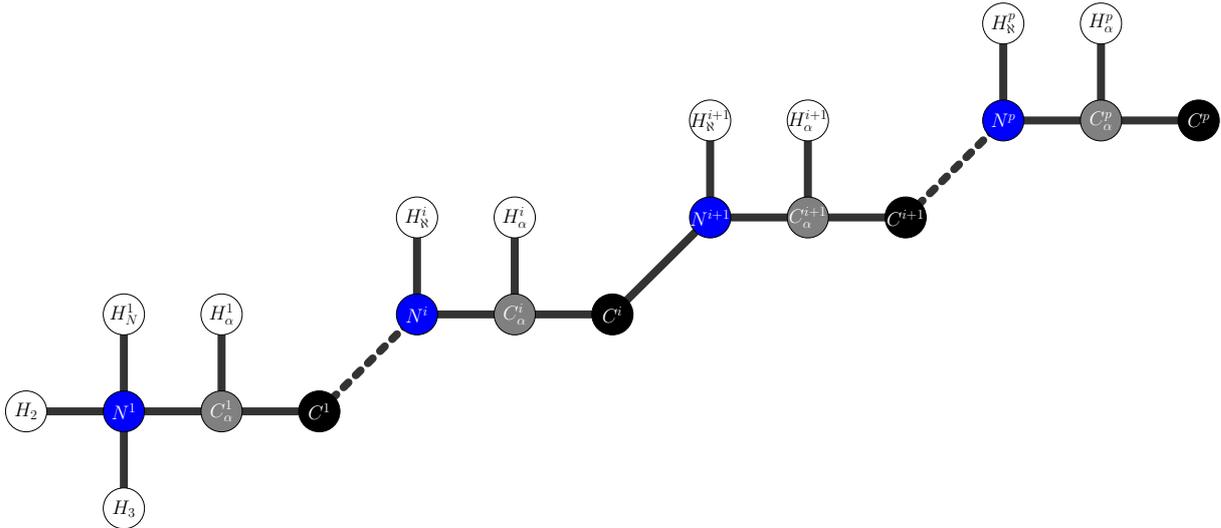


Figura 43 – Conjunto de átomos de interesse para uma proteína composta por p resíduos de aminoácidos. Aqui $H_N^i = H_N^i$ ou $H_N^i = H_{\delta_3}^i$, a depender se o resíduo de aminoácido i é a prolina ou não.

O trabalho [LAVOR et al. \(2011\)](#) foi o pioneiro na modelagem da estrutura das proteínas por meio de uma ordem PDGD. Por mais que o conceito de ordem *hand-crafted* (feita a mão) só tenha surgido em [LAVOR; LIBERTI; MUCHERINO \(2013\)](#), o trabalho de 2011 já propunha uma ordenação nos átomos de proteínas com base em conhecimentos experimentais. Nessa linha, os trabalhos [MUCHERINO \(2015\)](#), [GONÇALVES et al. \(2017\)](#) e [LAVOR et al. \(2019\)](#) também propõem diferentes alternativas para a ordenação dos vértices do grafo que modelam uma proteína. Com base no conhecimento adquirido nestas referências, vamos propor a seguinte ordem *hand-crafted*, que denominamos *hc*, para explorar o conjunto de átomos de nosso interesse:

$$hc = \left\{ H_3, H_2, H_N^1, N^1, C_\alpha^1, H_\alpha^1, C^1, \dots, N^i, H_N^i, C_\alpha^i, C^i, H_\alpha^i, \dots, N^p, H_N^p, C_\alpha^p, C^p, H_\alpha^p \right\}. \quad (5.1)$$

A ordem *hc* difere das ordens descritas nas referências outrora citadas pois sua ordenação consegue descrever os átomos de cada resíduo de aminoácido com base nos ângulos de torção da proteína. Esta afirmação está mais detalhada no decorrer desta seção. A [Figura 44](#) evidencia como esta ordem é tomada nos átomos considerados de uma proteína composta por 4 resíduos de aminoácidos.

Para a ordem *hc*, enunciamos a seguinte proposição:

Proposição 5.1. A ordem *hc* é uma ordem PDGD*i*.

Demonstração. Para provar que a ordem *hc* é de fato uma ordem PDGD*i*, temos que mostrar que ela satisfaz as condições da [Definição 3.7](#). Assim, temos que verificar que: (I) cada vértice v_i desta ordem possui ao menos três predecessores adjacentes que não são colineares em \mathbb{R}^3 e (II) a distância entre v_i e ao menos dois dos seus predecessores

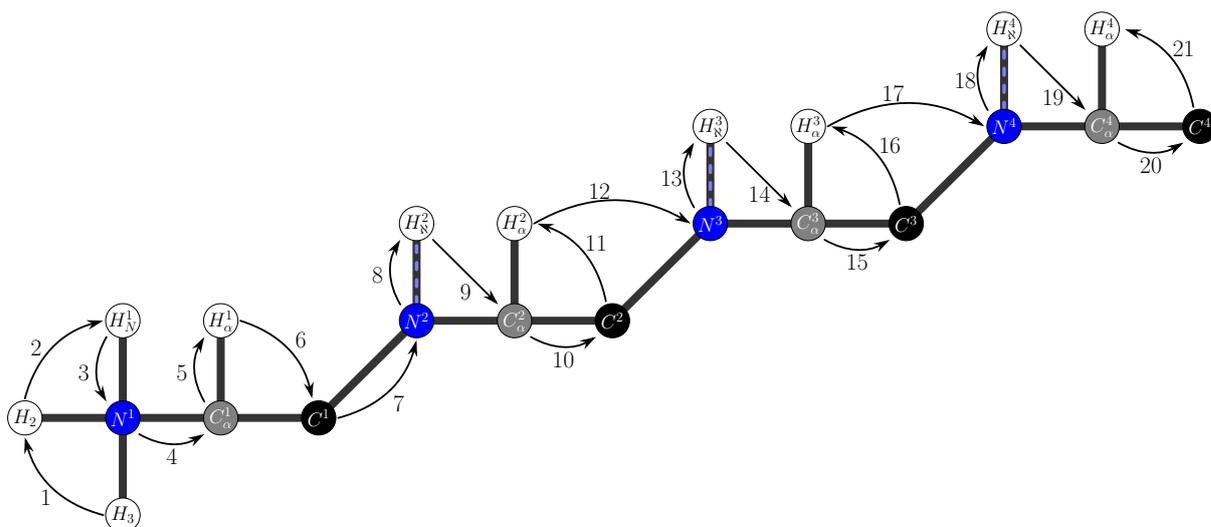


Figura 44 – Nova ordem *hand-crafted* aplicada a uma proteína composta por quatro resíduos de aminoácidos.

adjacentes é precisa. A [Tabela 1](#) apresenta uma possível clique para cada um dos vértices de *hc*. Isto já nos garante que cada vértice desta ordem possui ao menos três vértices predecessores adjacentes. Quanto a posição espacial destes vértices não serem colineares no espaço, na [Seção 4.1](#) salientamos que os ângulos planares definidos nas proteínas não possuem módulo igual a 0° ou 180° . Isto garante que a posição dos vértices predecessores adjacentes a v_i não são colineares. Assim, a condição (I) fica justificada.

Agora, para finalizar esta demonstração, vamos observar que somente as hipóteses [H7](#), [H8](#) e [H9](#) garantem distâncias intervalares. Desta maneira, pela [Tabela 1](#) observamos que estas hipóteses garantem apenas uma distância nas cliques dos átomos H_α^1 , N^k e C^k , com $k = 2, 3, \dots, p$. Portanto, temos a garantia que a distância entre v_i e pelo menos dois de seus predecessores adjacentes é precisa para todo $i = 4, \dots, n$. O que certifica a condição (II) e conclui a demonstração desta proposição.

Tabela 1 – Restrições de distâncias entre átomos de uma proteína tomadas com base no conjunto de hipóteses apresentado na Seção 4.1. Aqui, $k = 2, 3, \dots, p$.

Cliques: $\{v_{i_3}, v_{i_2}, v_{i_1}, v_i\}$				Hipóteses que garantem as distâncias					
v_{i_3}	v_{i_2}	v_{i_1}	v_i	d_{i_3, i_2}	d_{i_3, i_1}	$d_{i_3, i}$	d_{i_2, i_1}	$d_{i_2, i}$	$d_{i_1, i}$
H_3	H_2	H_N^1	N^1	H3	H3	H2	H3	H2	H2
H_2	H_N^1	N^1	C_α^1	H3	H2	H3	H2	H3	H2
H_N^1	N^1	C_α^1	H_α^1	H2	H3	H7	H2	H3	H2
N^1	C_α^1	H_α^1	C^1	H2	H3	H3	H2	H2	H3
N^{k-1}	C_α^{k-1}	C^{k-1}	N^k	H2	H3	H8	H2	H3	H2
C_α^{k-1}	C^{k-1}	N^k	H_N^k	H2	H3	H5	H2	H3	H2
C_α^{k-1}	C^{k-1}	N^k	C_α^k	H2	H3	H5	H2	H3	H2
C^{k-1}	N^k	C_α^k	C^k	H2	H3	H9	H2	H3	H2
N^k	C_α^k	C^k	H_α^k	H2	H3	H3	H2	H2	H3

□

Com base no resultado descrito pela Proposição 5.1, e utilizando todas as informações de restrições de distâncias que o nosso conjunto de hipóteses nos fornece, a Figura 45 descreve metade da região da árvore do espaço de busca de soluções do i -ésimo resíduo de aminoácido da ordem hc .

Como é descrito em GONÇALVES et al. (2017) e LAVOR et al. (2019), informações sobre quiralidade local e a condição planar imposta pela ligação peptídica podem ser utilizadas para reduzir o espaço de busca. Para observar o impacto destas considerações na ordem hc , apresentamos o seguinte resultado.

Proposição 5.2. Considerando o conjunto de hipóteses **H1**, **H2**, \dots , **H9** apresentado na Seção 4.1, o espaço de busca de soluções gerado pela aplicação do i BP na ordem hc apresenta ramificações somente nos átomos H_α^1 , N^i e C^i , com $i = 2, \dots, p$, onde p é o número de resíduos de aminoácidos que compõem a molécula de proteína.

Demonstração. Como todos os vértices da ordem hc podem ser imersos no \mathbb{R}^3 levando em conta as cliques apresentadas na Tabela 1, vamos observar que:

- Os átomos N^1 , C_α^1 , C^1 e H_α^1 podem ser realizados levando em conta, respectivamente, as cliques $\{H_3, H_2, H_N^1, N^1\}$, $\{H_2, H_N^1, N^1, C_\alpha^1\}$, $\{N^1, C_\alpha^1, H_\alpha^1, C^1\}$ e $\{N^i, C_\alpha^i, C^i, H_\alpha^i\}$. Como as distâncias entre todos os átomos destas cliques são

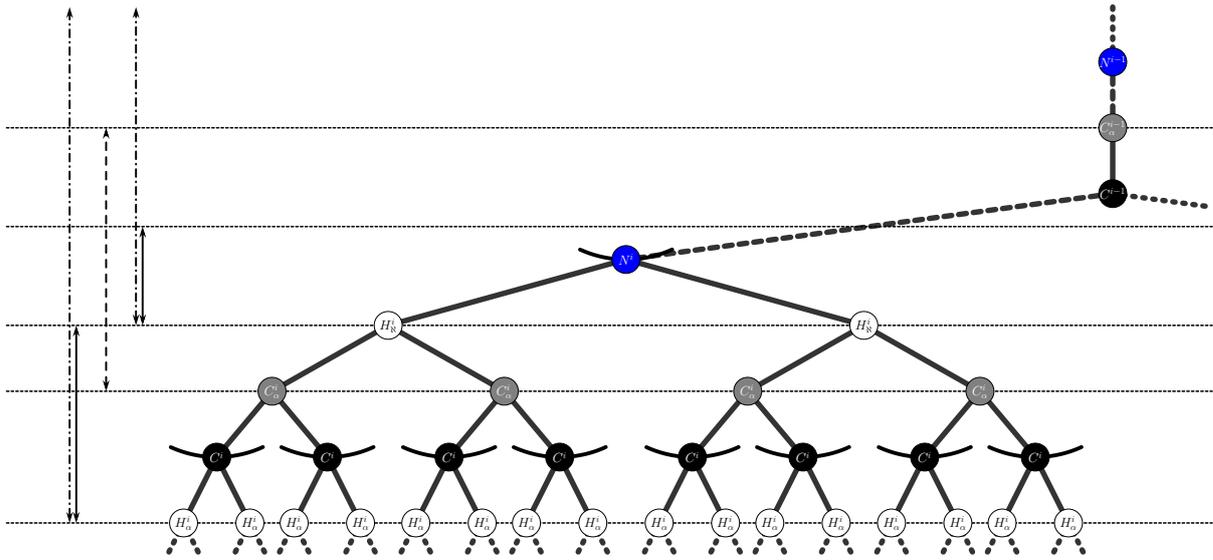


Figura 45 – Metade da região da árvore do espaço de busca de soluções referente ao i -ésimo resíduo de aminoácido da ordem hc . Os arcos de circunferência simbolizam a existência de uma infinidade de posições possíveis para a imersão dos átomos que estão sobre eles. As setas representam distâncias “extras”, i.e., não usadas no processo de discretização, da ordem, cuja diferença no traço trazem a seguinte informação: a linha tracejada representa a existência de uma distância precisa; as linhas contínuas descrevem a existência de uma distância intervalar; e as linhas ponto-traço a possibilidade de existência de uma distância intervalar.

precisas, pela discussão apresentada na Subseção 3.2.1, existem duas possibilidades para a posição de cada átomo. No entanto, a hipótese **H4** garante que existe uma quiralidade local definida nestas 4 cliques de átomos. Assim, podemos fixar uma destas posições com uma informação *a priori*, portanto só existe uma possibilidade factível para a realização de cada um destes átomos.

- Os átomos H_N^i e C_α^i podem ser realizados pelas cliques $\{C_\alpha^{i-1}, C^{i-1}, N^i, H_N^i\}$ e $\{C_\alpha^{i-1}, C^{i-1}, N^i, C_\alpha^i\}$, respectivamente. Como as distâncias entre todos os átomos destas cliques são precisas, e pela hipótese **H5** sabemos que todos eles estão relacionados com o plano peptídico. Conforme apresentado na Subseção 3.2.1, só existe uma posição possível para realização destes átomos.
- A realização dos átomos H_N^1 , N^i e C^i é dada, respectivamente, pelas cliques $\{H_N^1, N^1, C_\alpha^1, H_\alpha^1\}$, $\{N^{i-1}, C_\alpha^{i-1}, C^{i-1}, N^i\}$ e $\{C^{i-1}, N^i, C_\alpha^i, C^i\}$, que envolvem uma distância intervalar, não definem nenhuma quiralidade local e nem fazem parte de um mesmo plano peptídico. Assim, pela discussão apresentada na Seção 3.3, existem dois arcos de possibilidades para a realização de cada um destes átomos.

Como as três afirmações acima discorrem sobre todos os átomos da ordem hc , o resultado desta proposição fica demonstrado. \square

De fato, o resultado desta proposição nos mostra que informações de conhecimento prévio sobre as proteínas nos ajuda a reduzir o espaço de busca gerado na ordem *hc*. Para observar esta redução, a [Figura 46](#) mostra a estrutura da árvore quando levamos em conta o resultado desta proposição. Esta figura representa a mesma região do espaço de busca que observamos na [Figura 45](#).

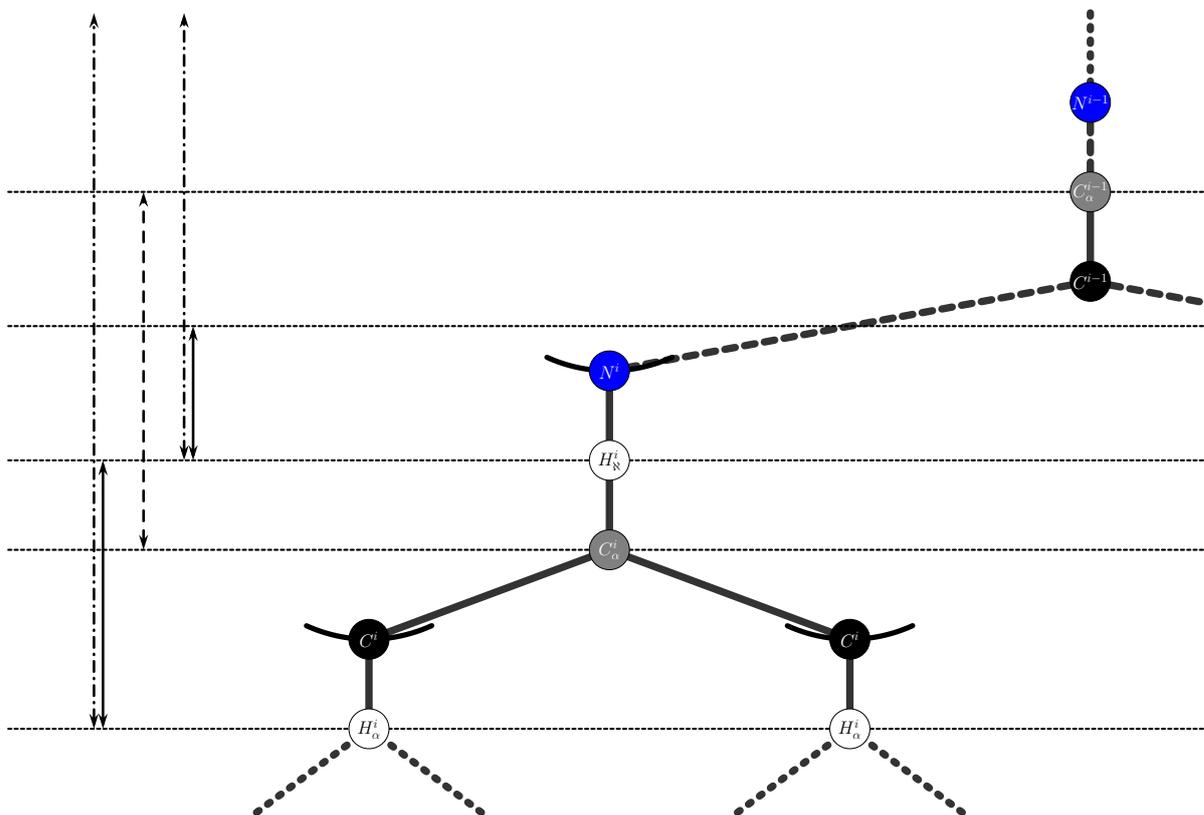


Figura 46 – Árvore da [Figura 45](#) considerando informações de quiralidade local e da planaridade imposta pela ligação peptídica.

Assim, para cada resíduo de aminoácido, o número de caminhos possíveis na árvore do espaço de busca é reduzido de 32 para 4. Como nos é intuitivo, esta redução no espaço de busca deve propiciar um impacto positivo no tempo computacional quando usamos o algoritmo *iBP* para encontrar uma solução deste problema. De fato um resultado que corrobora com esta intuição é apresentado em [GONÇALVES et al. \(2017\)](#) para uma outra ordem *hand-crafted*.

Como a maneira que apresentamos o algoritmo *iBP* na [Seção 3.3](#) é fortemente baseada em ângulos de torção, vamos considerar o seguinte corolário da [Proposição 5.2](#):

Corolário 5.1. Com exceção do primeiro ramo, os ângulos de torção associados a todas as outras ramificações na árvore de busca gerada pela aplicação do *iBP* na ordem *hc* são os ângulos ϕ_i ou ψ_i .

Demonstração. Pelo resultado da proposição anterior, sabemos que, com exceção da primeira, as ramificações na árvore descrita pela ordem *hc* são dadas nas cliques de átomos

$\{N^{i-1}, C_\alpha^{i-1}, C^{i-1}, N^i\}$ e $\{C^{i-1}, N^i, C_\alpha^i, C^i\}$. Como os ângulos de torção ψ_{i-1} e ϕ_i são definidos, respectivamente, nestas duas cliques de átomos, o resultado deste corolário fica demonstrado. \square

Estes corolário e a primeira proposição que apresentamos nesta seção são a base do algoritmo que propomos na próxima seção.

5.2 Um Novo Algoritmo do Tipo *iBP* Aplicado a Proteínas

A Proposição 5.1 nos garante que o algoritmo *iBP* pode ser utilizado para caracterizar a posição espacial dos átomos que são descritos pela ordem *hc*. Já o resultado da Proposição 5.2 pode ser levado em conta para aumentar a velocidade com que este algoritmo explora o espaço de busca. No entanto, como podemos observar na Figura 46, as informações de distâncias “extras”, que podemos utilizar para realizar podas na árvore, não estão associadas aos átomos em que a ramificação da árvore acontece. Assim, é necessário descer na árvore, ainda que um único nível, para verificar se, de fato, um ramo tomado é factível ou não. Ainda observamos que só existe uma distância intervalar associada aos átomos N^i e C^i , que são relacionados ao processo de ramificação da ordem *hc*. Desta maneira, não é possível utilizarmos uma estratégia de redução de arcos como apresentamos na Seção 3.3 para a imersão destes átomos. Com a finalidade de transpassar estas adversidades, propomos um novo algoritmo, baseado na estratégia do *iBP*, que leva em conta a geometria intrínseca das proteínas.

Como vimos no capítulo anterior, podemos definir relações geométricas entre os ângulos de torção definidos no *backbone* das proteínas e distâncias entre átomos de hidrogênio que ficam nas imediações destes ângulos. Dizemos que estas relações são dadas com base na geometria intrínseca das proteínas. Como vimos na Proposição 4.1, as distâncias do átomo de hidrogênio H_α^i a qualquer outro átomo de hidrogênio H_k^j , com $j \leq i$, pode ser escrita como um ângulo de torção ϕ_i . Uma conclusão semelhante acontece na Proposição 4.2, mas relacionando as distâncias entre os átomos de hidrogênio H_N^i e H_k^j com o ângulo de torção ψ_i . Estes dois resultados são a chave para o desenvolvimento deste novo algoritmo.

Da hipótese H7, temos que as distâncias entre átomos de hidrogênio são distâncias intervalares. Assim, devemos considerar uma adaptação nos resultados das Proposições 4.1 e 4.2 para que as distâncias intervalares correspondam a um intervalo angular para os ângulos ϕ_i e ψ_i . A parte principal deste processo já foi discutida na Seção 3.3, aqui só vamos apresentar as correções necessárias para descrever estes ângulos.

Primeiramente, por questões de formalização matemática, nos atentemos a seguinte definição:

Definição 5.1. A soma \boxplus entre um escalar $x \in \mathbb{R}$ e um intervalo $I = [a, b] \subset \mathbb{R}$ é dada por:

$$x \boxplus I = x \boxplus [a, b] \equiv [a + x, b + x].$$

Para seguir nesta construção, da Proposição 3.1, temos que:

$$\begin{aligned} \tau_{H_k^j, N^i, C_\alpha^i, H_\alpha^i}(\mathbf{d}_{H_k^j, H_\alpha^i}) &= \mathcal{T}_{H_k^j, H_\alpha^i} = \mathcal{T}_{H_k^j, H_\alpha^i}^- \cup \mathcal{T}_{H_k^j, H_\alpha^i}^+, \\ \tau_{H_k^j, C_\alpha^i, C^i, H_N^{i+1}}(\mathbf{d}_{H_k^j, H_N^{i+1}}) &= \mathcal{T}_{H_k^j, H_N^{i+1}} = \mathcal{T}_{H_k^j, H_N^{i+1}}^- \cup \mathcal{T}_{H_k^j, H_N^{i+1}}^+. \end{aligned}$$

Agora, definindo as quantidades:

$$\begin{cases} \Sigma_{i,j,k} \equiv \phi_{i,j,k} \boxplus \mathcal{T}_{H_k^j, H_\alpha^i} = \phi_{i,j,k} \boxplus \mathcal{T}_{H_k^j, H_\alpha^i}^- \cup \phi_{i,j,k} \boxplus \mathcal{T}_{H_k^j, H_\alpha^i}^+, \\ \Xi_{i,j,k} \equiv \psi_{i,j,k} \boxplus \mathcal{T}_{H_k^j, H_N^{i+1}} = \psi_{i,j,k} \boxplus \mathcal{T}_{H_k^j, H_N^{i+1}}^- \cup \psi_{i,j,k} \boxplus \mathcal{T}_{H_k^j, H_N^{i+1}}^+, \end{cases} \quad (5.2)$$

e as funções:

$$\begin{aligned} c_{\Phi_i}, c_{\Psi_i} &: [-540^\circ, 540^\circ] \longrightarrow (-180^\circ, 180^\circ], \\ t &\longmapsto \begin{aligned} c_{\Phi_i}(t) &= c_{\phi_i}(t) + t, \\ c_{\Psi_i}(t) &= c_{\psi_i}(t) + t, \end{aligned} \end{aligned} \quad (5.3)$$

onde $\phi_{i,j,k}$, c_{ϕ_i} e $\psi_{i,j,k}$, c_{ψ_i} são definidos, respectivamente, nas Proposições 4.1 e 4.2. Podemos escrever o seguinte lema:

Lema 5.1. Intervalos angulares para os ângulos ϕ_i e ψ_i podem ser escritos em função das distâncias intervalares $\mathbf{d}_{H_k^j, H_\alpha^i}$ e $\mathbf{d}_{H_k^j, H_N^{i+1}}$ da seguinte maneira:

$$\begin{cases} \Phi_{i,j,k} = c_{\Phi_i}(\phi_{i,j,k} \boxplus \tau_{H_k^j, N^i, C_\alpha^i, H_\alpha^i}(\mathbf{d}_{H_k^j, H_\alpha^i})) = c_{\Phi_i}(\phi_{i,j,k} \boxplus \mathcal{T}_{H_k^j, H_\alpha^i}) = c_{\Phi_i}(\Sigma_{i,j,k}), \\ \Psi_{i,j,k} = c_{\Psi_i}(\psi_{i,j,k} \boxplus \tau_{H_k^j, C_\alpha^i, C^i, H_N^{i+1}}(\mathbf{d}_{H_k^j, H_N^{i+1}})) = c_{\Psi_i}(\psi_{i,j,k} \boxplus \mathcal{T}_{H_k^j, H_N^{i+1}}) = c_{\Psi_i}(\Xi_{i,j,k}). \end{cases}$$

Demonstração. A construção destes intervalos angulares foi tomada observando as adequações matemáticas necessárias para que os mesmos representem as relações geométricas descritas das Proposições 4.1 e 4.2. Assim, a demonstração deste lema segue direto do resultado destas proposições. \square

Agora, se $U_{H_\alpha^i}$ e $U_{H_N^{i+1}}$ são, respectivamente, os conjuntos átomos de predecessores adjacentes de H_α^i e H_N^{i+1} , pela ordenação dos átomos na ordem *hc*, para todo $j < i$ observamos que:

- Se $H_k^j \in U_{H_\alpha^i}$, então $H_k^j < N^i < C_\alpha^i < C^i$.
- Se $H_k^j \in U_{H_N^i}$, então $H_k^j < C_\alpha^i < C^i < N^{i+1}$.

Se consideramos o algoritmo *iBP* aplicado a ordem *hc*, então:

- Pelo primeiro item, antes de encontrar a posição do átomo C^i podemos calcular, de maneira trivial, as distâncias $d(H_k^j, N^i)$ e $d(H_k^j, C_\alpha^i)$, pois os átomos H_k^j , N^i e C_α^i já foram realizados;
- Pelo segundo item, antes de realizar o átomo N^{i+1} , por um argumento similar ao anterior, as distâncias $d(H_k^j, C_\alpha^i)$ e $d(H_k^j, N^i)$ podem ser calculadas.

Das duas observações acima e do resultado do Lema 5.1, temos a seguinte proposição:

Proposição 5.3. Sejam $U_{H_\alpha^i}$ e $U_{H_N^{i+1}}$ o conjunto de predecessores adjacentes dos átomos H_α^i e H_N^{i+1} , respectivamente. Então, os intervalos angulares para os ângulos ϕ_i e ψ_i , descritos por distâncias entre átomos de hidrogênio, são dados por:

$$\Phi_i = \bigcap_{H_k^j \in U_{H_\alpha^i}} c_{\Phi_i}(\Sigma_{i,j,k}) \quad (5.4)$$

$$\Psi_i = \bigcap_{H_k^j \in U_{H_N^{i+1}}} c_{\Psi_i}(\Xi_{i,j,k}) \quad (5.5)$$

onde $\Sigma_{i,j,k}$, $\Xi_{i,j,k}$ e c_{Φ_i} , c_{Ψ_i} são definidos, respectivamente, em (5.2) e (5.3).

Demonstração. Como é conhecida a distância de cada átomo de hidrogênio de $U_{H_\alpha^i}$ para o átomo H_α^i , o Lema 5.1 nos garante um intervalo angular $\Phi_{i,j,k}$ do ângulo ϕ_i para cada uma destas distâncias. Assim, o intervalo angular Φ_i , que satisfaz todas os intervalos $\Phi_{i,j,k}$ é dado pela intersecção dos mesmos. A demonstração da relação para Ψ_i é feita com um argumento similar ao apresentado para Φ_i . \square

O resultado apresentado nesta proposição nos mostra uma maneira de considerar as restrições de distância entre átomos de hidrogênio nos ângulos de torção ϕ e ψ . Assim, se utilizamos as relações apresentadas na Proposição 5.3, podemos aferir as restrições de distâncias que envolvem o átomo H_N^i na imersão do átomo N^i . O mesmo vale para os átomos H_α^i e C^i . Desta maneira, podemos utilizar todas as restrições de distâncias antes de considerar cada ramificação na árvore do espaço de busca. A Figura 47 apresenta como fica a região da árvore descrita na Figura 46 considerando as relações descritas na Proposição 5.3.

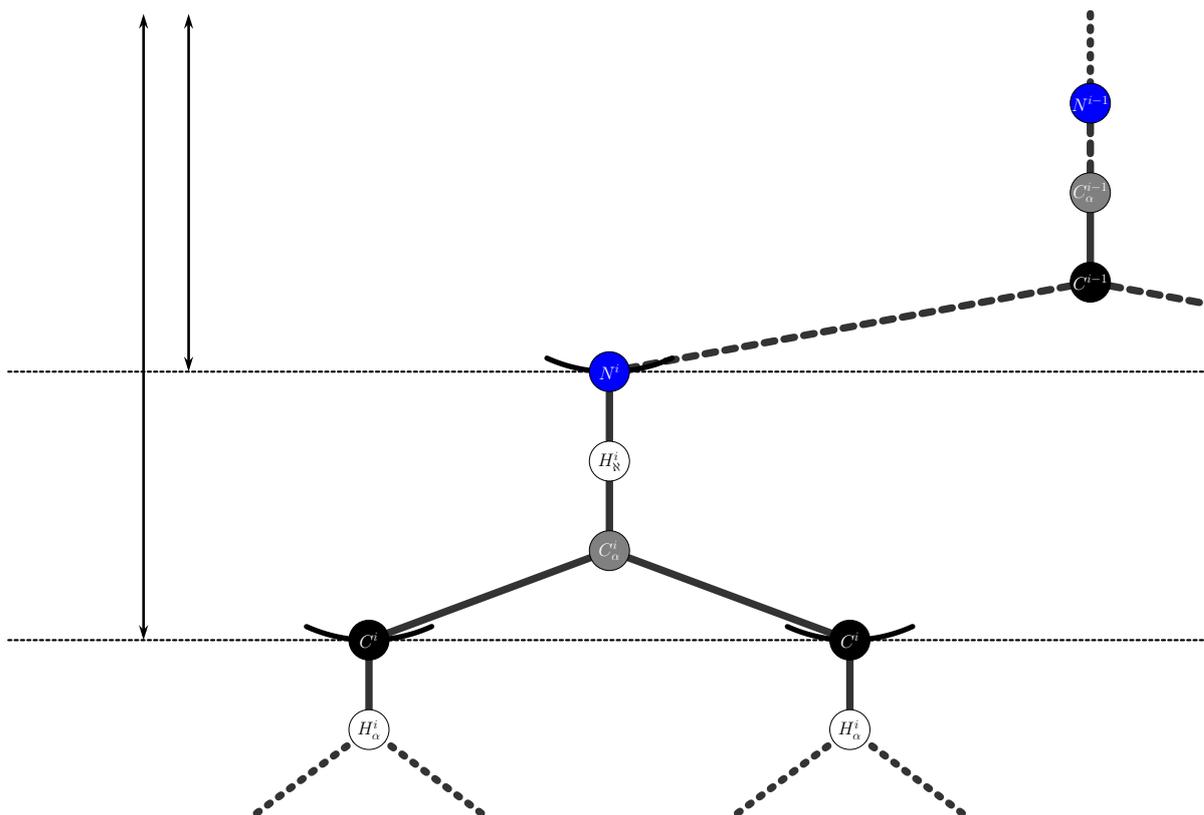


Figura 47 – Árvore da Figura 46 evidenciando a mudança das restrições de distância dos átomos H_N^i e H_α^i para os átomos N^i e C^i , respectivamente.

Do Corolário 5.1, temos que as ramificações da árvore de busca, com exceção do primeiro, são associadas aos ângulos de torção ϕ_i e ψ_i . Pela Proposição 5.3, podemos descrever estes ângulos utilizando todas as restrições de distâncias que envolvem os átomos de hidrogênio. Assim, os átomos de hidrogênio H_N^i e H_α^i podem ser descritos considerando somente as cliques apresentadas na Tabela 1, pois as restrições de distância para os átomos H_k^j já foram consideradas na imersão dos átomos N^i e C_i .

Com esta observação, propomos o algoritmo *protein interval Torsion angle Branch & Prune* (*iTBP*), que integra a geometria intrínseca das proteínas, descrita pelas relações geométricas que envolvem os ângulos de torção, com a estratégia do *iBP*. Sugerimos o pseudo-código do Algoritmo 3 como a estrutura geral do *iTBP*.

5.3 Experimentos Computacionais

Nesta seção, vamos apresentar experimentos computacionais para comparar os algoritmos *iBP* e *iTBP*. Primeiramente, vamos descrever a maneira que geramos as instâncias a partir de estruturas 3D de proteínas do *RCSB Protein Data Bank* (PDB) (BERMAN et al., 2000) utilizadas como entrada para estes algoritmos. Em seguida, vamos apresentar a estratégia que utilizamos para amostrar os intervalos angulares, e por fim, os resultados obtidos.

Algoritmo 3: Estrutura principal do algoritmo *i*TBP.

Entrada: $G = (V, E, \mathbf{d})$, $X = \{x_1, x_2, x_3\}$, $i = 4$, $A = hc$ definida em (5.1),
 $\mathcal{T}_v = [\underline{\tau}_v, \bar{\tau}_v]$, $b_v = 1$ e $N_v \geq 1 \forall v \in V$

Saída: $X = \{x_v \in \mathbb{R}^3 \mid v \in V\}$

```

1 Função iTBP( $G, X, i, A_i, \mathcal{T}_i, b_i, N_i$ )
2   se ( $A_i \notin \{N, C, H_\alpha^1\}$ ) então
3     compute as quantidades  $\mathbf{a}_i(i_2, i_1, i)$ ,  $\mathbf{b}_i(i_3, i_2, i_1, i)$ ,  $\mathbf{c}_i(i_3, i_2, i_1, i)$ 
      definidas em (3.41);
4      $X \leftarrow X \cup \{\mathbf{a}_i(i_2, i_1, i) + \mathbf{b}_i(i_3, i_2, i_1, i) \cos \mathcal{T}_i + \mathbf{c}_i(i_3, i_2, i_1, i) \sin \mathcal{T}_i\}$ ;
5     se ( $i = |V|$ ) então
6       | retorne  $X$ ;
7     senão
8       | iTBP( $G, X, i + 1, A_{i+1}, \mathcal{T}_{i+1}, b_{i+1}, N_{i+1}$ );
9     fim
10  senão
11    se ( $b_i = 1$ ) então
12      se ( $A_i = H_\alpha^1$ ) então
13        | compute a quantidade  $\Theta_i$  definida em (3.57);
14        |  $\mathcal{T}_i \leftarrow \Theta_i$ ;
15      senão se ( $A_i = N$ ) então
16        | compute a quantidade  $\Psi_i$  definida em (5.5);
17        |  $\mathcal{T}_i \leftarrow \mathcal{T}_i \cap \Psi_i$ ;
18      senão se ( $A_i = C$ ) então
19        | compute a quantidade  $\Phi_i$  definida em (5.4);
20        |  $\mathcal{T}_i \leftarrow \mathcal{T}_i \cap \Phi_i$ ;
21      fim
22      se ( $\mathcal{T}_i \neq \emptyset$ ) então
23        | compute  $\mathbf{a}(i_2, i_1, i)$ ,  $\mathbf{b}(i_3, i_2, i_1, i)$ ,  $\mathbf{c}(i_3, i_2, i_1, i)$  definida em
          (3.41);
24        | considere  $T_i = \{\tau_1, \dots, \tau_{N_i}\}$  uma amostra do intervalo  $\mathcal{T}_i$  com
           $N_i$  pontos;
25        senão
26          |  $T_i \leftarrow \emptyset$ ;
27        fim
28      fim
29      se ( $T_i \neq \emptyset$ ) então
30        |  $X \leftarrow X \cup \{\mathbf{a}(i_2, i_1, i) + \mathbf{b}(i_3, i_2, i_1, i) \cos \tau_{b_i} + \mathbf{c}(i_3, i_2, i_1, i) \sin \tau_{b_i}\}$ ;
31        se ( $i = |V|$ ) então
32          | retorne  $X$ ;
33        senão
34          | iTBP( $G, X, i + 1, A_{i+1}, \mathcal{T}_{i+1}, b_{i+1}, N_{i+1}$ );
35        fim
36      senão
37        | retroceda para o primeiro  $v_j < v_i$  tal que  $b_j < N_j$ ;
38        |  $b_j \leftarrow b_j + 1$ ;
39        |  $X \leftarrow X \setminus \{x_j, x_{j+1}, \dots, x_{i-1}\}$ ;
40        |  $b_{j+1}, b_{j+2}, \dots, b_{i-1} \leftarrow 1$ ;
41        | iTBP( $G, X, j, A_j, \mathcal{T}_j, b_j, N_j$ );
42      fim
43    fim
44  fim

```

Geração das Instâncias

Os arquivos do PDB geralmente apresentam mais de uma conformação estrutural para a molécula descrita nestes arquivos. Estas distintas conformações são classificadas como modelos, onde cada um destes modelos descreve a estrutura quaternária das moléculas de proteínas. Assim:

- Para todos os arquivos que apresentam mais de um modelo estrutural para a molécula, tomamos sempre a estrutura descrita no primeiro modelo.
- Para toda molécula de proteína que é composta por mais de uma cadeia, consideramos que esta molécula é dada por sua cadeia A.

Para cada estrutura do PDB, consideramos conhecidas as restrições de distâncias descritas nas hipóteses **H2**, **H3**, **H5** e **H7** apresentadas na Seção 4.1. Para todos os átomos i e j em que uma distância intervalar é definida pela hipótese **H7**, consideramos que $\mathbf{d}_{i,j} = [\underline{d}_{i,j}, \bar{d}_{i,j}]$, com $\mathcal{E}_{d_{i,j}} = \bar{d}_{i,j} - \underline{d}_{i,j}$, cujos extremos são definidos por:

$$\underline{d}_{i,j} = \max \{d_{i,j}^* - \mathcal{E}_{d_{i,j}}/2, 0\} \text{ \AA} \quad \text{e} \quad \bar{d}_{i,j} = \min \{d_{i,j}^* + \mathcal{E}_{d_{i,j}}/2, 5\} \text{ \AA},$$

onde $d_{i,j}^* \sim \mathcal{N}(d_{i,j}, \mathcal{E}_{d_{i,j}}/8)$, sendo $d_{i,j}$ a distância entre os átomos i e j da estrutura considerada. A estratégia utilizada para definir estes intervalos foi inspirada no processo descrito em **BISWAS; TOH; YE (2008)**. No entanto, nosso procedimento leva em conta um desvio padrão de $\mathcal{E}_{d_{i,j}}/8$ na distribuição normal, o que nos garante, estatisticamente, que o valor de $d_{i,j}^*$ é tomado em um intervalo de comprimento $\mathcal{E}_{d_{i,j}}$.

Pelas hipóteses **H8** e **H9**, supomos que um intervalo angular para todo ângulo de torção ψ_i e ϕ_i é conhecido. Estes intervalos foram tomados por uma estratégia análoga a apresentada para distâncias intervalares, ou seja, se $\mathcal{T}_i = [\underline{\tau}_i, \bar{\tau}_i]$, onde $\mathcal{E}_{\tau_i} = \bar{\tau}_i - \underline{\tau}_i$, é um intervalo angular, então:

$$\bar{\tau}_i = \tau_i^* - \mathcal{E}_{\tau_i}/2 \quad \text{e} \quad \underline{\tau}_i = \tau_i^* + \mathcal{E}_{\tau_i}/2,$$

onde $\tau_i^* \sim \mathcal{N}(\tau_i, \mathcal{E}_{\tau_i}/8)$, sendo τ_i o ângulo de torção dado na estrutura considerada. Observando que os termos da relação (3.42) podem ser rearranjados com o objetivo de escrever a distância entre os átomos que definem o ângulo de torção em função deste próprio ângulo, para cada intervalo angular definido por um ângulo de torção, consideramos uma distância intervalar entre estes dois átomos calculada por este processo.

A hipótese **H4** nos garante que a quiralidade local em alguns pontos da proteína é conhecida. Assim, para as cliques de átomos onde esta quiralidade é definida, consideramos como conhecido o ângulo de torção dado nesta clique.

O valor de $\mathcal{E}_{d_{i,j}}$ foi tomado com base em GONÇALVES et al. (2017) da seguinte maneira:

$$\mathcal{E}_{d_{i,j}} = \begin{cases} 0,5 \text{ \AA} & \text{se os átomos } i, j \text{ pertencem ao mesmo resíduo} \\ & \text{de aminoácido ou a resíduos adjacentes,} \\ 1,0 \text{ \AA} & \text{caso contrário.} \end{cases}$$

Já para o comprimento dos intervalos angulares, consideramos que $\mathcal{E}_\tau = 40^\circ$. Valor escolhido em consequência das discussões apresentadas em MALLIAVIN et al. (2019).

Amostragem dos Intervalos

Como discutimos na Seção 3.3, em cada passo do *i*BP definimos um intervalo angular no átomo i que deve ser amostrado. Para a amostragem destes intervalos, primeiramente, consideramos todas as restrições de distâncias intervalares que envolvem este átomo e utilizamos as relações descritas em (3.57), (5.4) ou (5.5), a depender do átomo a ser realizado, para calcular um intervalo angular que leva em conta todas estas restrições de distâncias. Conforme vimos na Seção 3.3, este intervalo angular resultante pode ser descrito por um único intervalo, pela união de dois intervalos disjuntos ou até mesmo pela união de mais de dois intervalos disjuntos. Se o número destes intervalos é maior que dois, consideramos somente dois deles, conjuntura que é explicada mais à frente nesta subseção. Desta maneira, amostramos cada um dos intervalos que compõem o intervalo resultante em D pontos equidistantes, estabelecemos uma ordem nestes D pontos e os tomamos um-a-um na estratégia do *i*BP.

Para a escolha dos intervalos que compõem o intervalo angular resultante e para a ordenação dos D pontos da amostra, afirmações feitas no parágrafo anterior, adotamos a seguinte estratégia.

1. Consideramos que cada distância intervalar $\mathbf{d}_{i,\ell}$ é representada por uma gaussiana $g(A_{i,\ell}, \mu_{i,\ell}, \sigma_{i,\ell})$, onde

$$g(A, \mu, \sigma)(x) \equiv A \exp \left[-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right],$$

com amplitude $A_{i,\ell} = 1$, média $\mu_{i,\ell} = \frac{d_{i,\ell} + \bar{d}_{i,\ell}}{2}$ e desvio padrão $\sigma_{i,\ell} = \frac{\mathcal{E}_{d_{i,\ell}}}{8}$.

2. As distâncias intervalares são convertidas em intervalos angulares com base na Proposição 3.1, que também são descritos pelas gaussianas $g(1, -\mu_{\tau_{\ell,i}}, \sigma_{\tau_{\ell,i}})$ e $g(1, \mu_{\tau_{\ell,i}}, \sigma_{\tau_{\ell,i}})$, onde:

$$\mu_{\tau_{\ell,i}} = \frac{\tau_{i_\ell, i_2, i_1, i} + \tau_{i_\ell, i_2, i_1, i}}{2} \quad \text{e} \quad \sigma_{\tau_{\ell,i}} = \frac{|\tau_{i_\ell, i_2, i_1, i} - \tau_{i_\ell, i_2, i_1, i}|}{8}.$$

Para seguir o raciocínio, vamos observar o seguinte lema:

Lema 5.2. Se $g_1 = g(A_1, \mu_1, \sigma_1)$ e $g_2 = g(A_2, \mu_2, \sigma_2)$ são duas gaussianas, temos que $\hat{g} = g_1 \cdot g_2$ também é uma gaussiana e é descrita por $\hat{g} = g(\hat{A}, \hat{\mu}, \hat{\sigma})$, onde

$$\hat{\mu} = \frac{\mu_1 \sigma_2^2 + \mu_2 \sigma_1^2}{\sigma_1^2 + \sigma_2^2}, \quad \hat{\sigma} = \sqrt{\frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2}} \quad \text{e} \quad \hat{A} = A_1 A_2 \exp \left[-\frac{1}{2} \frac{(\mu_1 - \mu_2)^2}{\sigma_1^2 + \sigma_2^2} \right].$$

Demonstração. A demonstração deste resultado pode ser verificada sem muita dificuldade, então a deixamos como exercício para o leitor. \square

Seguindo com a estratégia empregada:

3. Para a relação (3.57), consideramos que:

$$g_{\mathcal{T}_i} = \prod_{\ell=3}^{k_i} \left[g \left(1, \varphi_{i_3, i_2, i_1, i_\ell} \oplus (-\mu_{\tau_{\ell, i}}), \sigma_{\tau_{\ell, i}} \right) + g \left(1, \varphi_{i_3, i_2, i_1, i_\ell} \oplus \mu_{\tau_{\ell, i}}, \sigma_{\tau_{\ell, i}} \right) \right].$$

4. Para a relação (5.4), consideramos que:

$$g_{\Phi_i} = \prod_{H_k^j \in U_{H_\alpha^i}} \left[g \left(1, c_{\phi_i}(\sigma_{i, j, k}^-) + \sigma_{i, j, k}^-, \sigma_{\tau_{H_k^j, H_\alpha^i}} \right) + g \left(1, c_{\phi_i}(\sigma_{i, j, k}^+) + \sigma_{i, j, k}^+, \sigma_{\tau_{H_k^j, H_\alpha^i}} \right) \right],$$

$$\text{onde } \sigma_{i, j, k}^\pm = \phi_{i, j, k} \pm \mu_{\tau_{H_k^j, H_\alpha^i}}.$$

5. Para a relação (5.5), consideramos que:

$$g_{\Psi_i} = \prod_{H_k^j \in U_{H_\mathbb{N}^{i+1}}} \left[g \left(1, c_{\psi_i}(\xi_{i, j, k}^-) + \xi_{i, j, k}^-, \sigma_{\tau_{H_k^j, H_\mathbb{N}^{i+1}}} \right) + g \left(1, c_{\psi_i}(\xi_{i, j, k}^+) + \xi_{i, j, k}^+, \sigma_{\tau_{H_k^j, H_\mathbb{N}^{i+1}}} \right) \right],$$

$$\text{onde } \xi_{i, j, k}^\pm = \psi_{i, j, k} \pm \mu_{\tau_{H_k^j, H_\mathbb{N}^{i+1}}}.$$

Portanto, para as funções definidas em cada um dos três itens anteriores, observamos que $g_{\mathcal{T}_i}(\mathcal{T}_i)$, $g_{\Phi_i}(\Phi_i)$ e $g_{\Psi_i}(\Psi_i)$ são, respectivamente, as imagens dos intervalos \mathcal{T}_i , Φ_i , Ψ_i por estas funções. Assim, conseguimos associar uma amplitude para cada intervalo que descreve os intervalos resultantes \mathcal{T}_i , Φ_i , Ψ_i , pois cada um destes intervalos é descrito pela soma de gaussianas. Desta maneira, podemos ordenar todos estes intervalos com base na sua amplitude. Então, caso \mathcal{T}_i , Φ_i ou Ψ_i sejam compostos pela união de mais de dois intervalos, com base em nossa experiência computacional, tomamos os dois de maior amplitude associada. Nossos resultados computacionais apontam que os intervalos com menor amplitude associada são menos prováveis conter um intervalo angular factível para o problema a ser resolvido.

Para a ordenação dos D pontos de cada amostra, ordenamos o valor da imagem destes pontos pela função g associada e tomamos esta ordenação para os pontos amostrados.

Como supomos que as distâncias intervalares e os intervalos angulares são descritos por uma distribuição normal, esta estratégia é uma primeira tentativa de levar em conta a estrutura em que estes intervalos foram definidos, isto é, uma maneira de considerarmos que alguns pontos destes intervalos são mais prováveis que outros. Como a multiplicação de duas gaussianas é uma gaussiana, este processo ainda é respeitado quando ordenamos as imagens das funções $g_{\mathcal{T}_i}$, g_{Φ_i} ou g_{Ψ_i} tomadas nos intervalos correspondentes.

Resultados Numéricos

Nesta seção, para a apresentação dos resultados obtidos, utilizamos as métricas *Mean and Largest Distance Error* (MDE, LDE) (BRUGLIERI; CORDONE; LIBERTI, 2022) e *Root Mean Square Deviation* (RMSD) (MAIOROV; CRIPPEN, 1994) para avaliar as soluções encontradas. Considerando $G = (V, E, \mathbf{d})$ o grafo associado ao PDGD $_i$, X é uma solução deste problema e X^* é a estrutura que originou G , neste caso X e X^* possuem o mesmo centro de massa. Então estas métricas são definidas da seguinte maneira:

$$\text{MDE}(G, X) = \frac{1}{|E|} \sum_{\{v_i, v_j\} \in E} \max \{ |d_{i,j} - \|x_i - x_j\||, |\bar{d}_{i,j} - \|x_i - x_j\|| \},$$

$$\text{LDE}(G, X) = \max_{\{v_i, v_j\} \in E} \left\{ \max \{ |d_{i,j} - \|x_i - x_j\||, |\bar{d}_{i,j} - \|x_i - x_j\|| \} \right\},$$

$$\text{RMSD}(X, X^*) = \min_{Q \in O(3)} \|X^* - XQ\|_F, \text{ onde } \|\cdot\|_F \text{ é a norma de Frobenius e } O(3)$$

é o grupo das matrizes ortogonais 3×3 .

As métricas MDE e LDE descrevem a qualidade da solução perante a instância do problema. Já a métrica RMSD, retrata a semelhança da solução encontrada com a estrutura que originou a instância resolvida; optamos por apresentar uma descrição matricial desta métrica conforme explicitado em LIBERTI et al. (2014). Considera-se que duas proteínas são semelhantes se o RMSD das duas estruturas for menor que 3 Å (REVA; FINKELSTEIN; SKOLNICK, 1998).

Os experimentos numéricos foram processados em um *notebook* Acer Aspire 5, Intel® Core™ i5-1035G1 CPU @ 1.00 GHz com 8GB de RAM e sistema operacional Pop!_OS 21.04. Os Algoritmos 2 e 3 foram implementados em MATLAB®, versão R2017b.

As Tabelas 2 e 3 mostram a comparação dos algoritmos i BP e i TBP. Para cada instância gerada da estrutura do PBD (PDB id), são apresentados o número de aminoácidos (naa), átomos ($|V|$) e de distâncias consideradas ($|E|$). O tempo computacional considerado, em segundos, para encontrar uma solução da instância é o tempo médio \pm desvio padrão de 10 testes computacionais. Os algoritmos foram processados com um tempo máximo de 60 segundos, assim o carácter ∞ significa que nenhuma solução da instância foi encontrada neste limite de tempo. Os valores de referência para as métricas apresentadas anteriormente são dados pela aplicação das mesmas na estrutura do PDB. Os resultados da Tabela 2

foram construídos considerando que todos os intervalos angulares foram amostrados com três pontos ($D = 3$). Já para os resultados apresentados na [Tabela 3](#), tomamos $D = 5$.

Dos resultados apresentados nas Tabelas [2](#) e [3](#) observamos que:

1. O algoritmo *i*TBP supera o *i*BP nos quesitos: tempo de processamento e número de instâncias resolvidas em menos de 60 segundos. Dos 42 testes apresentados, o *i*TBP encontrou uma solução para 32 deles. Já o *i*BP conseguiu resolver apenas 14 destes testes. Em todas as instâncias que os dois algoritmos foram capazes de resolver, o *i*TBP realizou esta tarefa em menos tempo.
2. Nos 10 testes que o *i*TBP não resolveu, 9 deles foram por conta de ter atingido o tempo limite estabelecido. Já para o *i*BP, de 28 casos onde este algoritmo não encontrou uma solução, em 14 deles o limite de tempo foi atingido.
3. Os valores de MDE e LDE das soluções encontradas pelos dois algoritmos são praticamente equivalentes.
4. Algumas das soluções encontradas pelos dois algoritmos possuem considerável dissimilaridade pela estrutura do PDB.
5. Algumas instâncias que não foram resolvidas quando consideramos uma amostragem dos intervalos angulares com três pontos passaram a ser resolvidas quando tomamos amostras com cinco pontos. Este caso aconteceu quase que majoritariamente no algoritmo *i*BP.

Como descrito em [LAVOR et al. \(2021b\)](#), a utilização de uma estratégia de redução de arcos consegue restringir os erros provenientes das medidas experimentais, o que promove uma melhoria no algoritmo *i*BP. No entanto, a ordenação dos átomos que tomamos não possibilita a utilização desta estratégia para o algoritmo *i*BP, pois nesta ordenação as ramificações acontecem em átomos associados aos ângulos de torção ϕ e ψ que não possuem distâncias “extras”. Contudo, pelas duas primeiras observações feitas acima, podemos dizer que o algoritmo proposto, *i*TBP, assimila com sucesso esta estratégia de redução de arcos. O que pode ser creditado as relações geométricas sobre os ângulos de torção que descrevemos no [Capítulo 4](#), pois elas associam distâncias de hidrogênio de longo alcance com estes ângulos do *backbone*.

A observação 3 pode ser justificada pelo fato dos algoritmos *i*BP e *i*TBP utilizarem uma mesma estratégia na descida da árvore que modela o espaço de busca.

Para a observação 4, de acordo com [D’AMBROSIO et al. \(2017\)](#) a métrica RMSD não tem muito significado prático para a aferição de estruturas 3D de proteínas, uma vez que ela não consegue distinguir reflexões parciais na estrutura.

Tabela 2 – Resultados numéricos obtidos em instâncias artificiais geradas a partir de estruturas do PDB. No processamento dos algoritmos, consideramos que cada intervalo foi amostrado com três pontos ($D = 3$). Os valores de referência foram tomados na estrutura do PDB.

PDB id	Instância		iBP				$iTBP$				Valores de Referência			
	naa	$ V $	$ E $	Tempo	MDE	LDE	RMSD	Tempo	MDE	LDE	RMSD	MDE	LDE	RMSD
1tos	10	52	215	0,07 ± 0,01	0,10	0,50	8,07	0,05 ± 0,04	0,10	0,50	6,95	0,09	0,40	0,00
1uao	10	52	229	1,03 ± 0,04	--	--	--	2,64 ± 0,13	0,11	0,50	2,89	0,10	0,38	0,00
1kum	10	52	235	0,24 ± 0,00	--	--	--	0,18 ± 0,00	0,12	0,50	1,63	0,11	0,39	0,00
1id6	15	77	345	0,09 ± 0,01	--	--	--	0,49 ± 0,01	0,11	0,50	3,61	0,10	0,41	0,00
1dng	15	77	353	0,58 ± 0,01	--	--	--	0,07 ± 0,00	0,12	0,50	2,92	0,11	0,42	0,00
1o53	15	77	355	0,09 ± 0,00	--	--	--	0,04 ± 0,00	0,11	0,50	4,23	0,11	0,39	0,00
1du1	20	102	447	0,16 ± 0,01	--	--	--	0,43 ± 0,01	--	--	--	0,10	0,42	0,00
1dpk	20	102	457	0,50 ± 0,02	--	--	--	0,30 ± 0,03	0,11	0,50	6,54	0,10	0,43	0,00
1ho7	20	102	485	0,08 ± 0,01	--	--	--	0,17 ± 0,05	0,12	0,50	6,46	0,11	0,40	0,00
1ckz	25	127	547	0,02 ± 0,00	--	--	--	0,07 ± 0,00	0,10	0,50	15,72	0,10	0,46	0,00
1lfc	25	127	562	∞	--	--	--	∞	--	--	--	0,10	0,43	0,00
1a11	25	127	606	0,91 ± 0,03	--	--	--	0,17 ± 0,01	0,13	0,50	8,20	0,11	0,44	0,00
1ho0	30	152	646	∞	--	--	--	0,11 ± 0,00	0,09	0,49	17,73	0,09	0,40	0,00
1mmc	30	152	702	2,91 ± 1,21	--	--	--	∞	--	--	--	0,10	0,49	0,00
1d0r	30	152	708	0,21 ± 0,02	--	--	--	0,31 ± 0,02	0,11	0,50	8,64	0,11	0,44	0,00
1zwd	35	177	799	2,74 ± 0,59	--	--	--	0,25 ± 0,00	0,11	0,50	14,24	0,10	0,45	0,00
1d1h	35	177	804	∞	--	--	--	∞	--	--	--	0,10	0,42	0,00
1spf	35	177	836	37,11 ± 3,55	--	--	--	0,15 ± 0,01	0,12	0,50	15,27	0,11	0,46	0,00
1aml	40	202	906	∞	--	--	--	0,40 ± 0,03	0,11	0,50	18,20	0,10	0,46	0,00
1ba4	40	202	940	∞	--	--	--	3,56 ± 0,85	0,12	0,50	14,83	0,11	0,42	0,00
1c56	40	202	942	∞	--	--	--	∞	--	--	--	0,11	0,42	0,00

Tabela 3 – Resultados numéricos obtidos em instâncias artificiais geradas a partir de estruturas do PDB. No processamento dos algoritmos, consideramos que cada intervalo foi amostrado com três pontos ($D = 5$). Os valores de referência foram tomados na estrutura do PDB.

PDB id	Instância			iBP			$iTBP$			Valores de Referência				
	naa	$ V $	$ E $	Tempo	MDE	LDE	RMSD	Tempo	MDE	LDE	RMSD	MDE	LDE	RMSD
	1tos	10	52	215	0,10 ± 0,01	0,10	0,47	2,53	0,05 ± 0,05	0,10	0,48	1,30	0,09	0,40
1uao	10	52	229	5,73 ± 0,22	0,11	0,49	1,27	1,32 ± 0,08	0,11	0,49	1,55	0,10	0,38	0,00
1kum	10	52	235	0,52 ± 0,01	0,11	0,50	2,25	0,06 ± 0,00	0,11	0,49	3,11	0,11	0,39	0,00
1id6	15	77	345	∞	--	--	--	0,35 ± 0,00	0,11	0,50	6,53	0,10	0,41	0,00
1dng	15	77	353	0,34 ± 0,00	0,12	0,49	2,60	0,09 ± 0,00	0,12	0,50	3,66	0,11	0,42	0,00
1o53	15	77	355	0,38 ± 0,00	0,12	0,50	2,89	0,04 ± 0,00	0,11	0,47	3,89	0,11	0,39	0,00
1du1	20	102	447	0,78 ± 0,02	0,10	0,48	7,97	0,12 ± 0,00	0,10	0,50	10,36	0,10	0,42	0,00
1dpk	20	102	457	∞	--	--	--	0,21 ± 0,00	0,11	0,50	7,34	0,10	0,43	0,00
1ho7	20	102	485	26,23 ± 2,66	--	--	--	0,08 ± 0,00	0,13	0,50	3,98	0,11	0,40	0,00
1ckz	25	127	547	0,11 ± 0,00	0,10	0,47	19,33	0,07 ± 0,01	0,10	0,50	12,27	0,10	0,46	0,00
1lfc	25	127	562	∞	--	--	--	∞	--	--	--	0,10	0,43	0,00
1a11	25	127	606	0,44 ± 0,04	0,13	0,49	5,43	0,08 ± 0,00	0,12	0,50	5,19	0,11	0,44	0,00
1ho0	30	152	646	0,19 ± 0,01	0,09	0,46	25,17	0,09 ± 0,00	0,09	0,50	17,91	0,09	0,40	0,00
1mmc	30	152	702	∞	--	--	--	∞	--	--	--	0,10	0,49	0,00
1d0r	30	152	708	13,01 ± 1,56	--	--	--	0,67 ± 0,05	0,12	0,50	14,94	0,11	0,44	0,00
1zwd	35	177	799	0,67 ± 0,09	0,11	0,50	26,63	0,20 ± 0,01	0,11	0,50	15,71	0,10	0,45	0,00
1d1h	35	177	804	∞	--	--	--	∞	--	--	--	0,10	0,42	0,00
1spf	35	177	836	∞	--	--	--	0,38 ± 0,01	0,12	0,50	9,40	0,11	0,46	0,00
1aml	40	202	906	0,87 ± 0,04	0,11	0,50	18,96	0,15 ± 0,01	0,11	0,50	22,53	0,10	0,46	0,00
1ba4	40	202	940	∞	--	--	--	∞	--	--	--	0,11	0,42	0,00
1c56	40	202	942	∞	--	--	--	∞	--	--	--	0,11	0,42	0,00

A observação 5 já foi amplamente discutida na literatura, como por exemplo pelo trabalho que citamos anteriormente. Desta maneira, sabe-se que quando os intervalos são amostrados em muitos pontos, o espaço de busca aumenta exponencialmente e, para uma amostra de poucos pontos, uma solução pode não ser encontrada.

Os resultados computacionais e a discussão apresentada evidenciam um ganho computacional do *i*TBP sobre o *i*BP, sem contar que o *i*TBP é baseado nos ângulos de torção ϕ e ψ , o que facilita a integração de qualquer informação mais específica sobre estes ângulos no algoritmo. Para exemplificar esta última afirmação, vale ressaltar que o conjunto de hipóteses que consideramos para a definição das restrições de distâncias e intervalos angulares leva em conta um conhecimento mais geral sobre as proteínas. No entanto, existem conhecimentos experimentais e teóricos mais restritos, como por exemplo, estimativas para os ângulos de torção ϕ e ψ em sequências de resíduos que envolvem os resíduos de aminoácidos glicina e prolina (GUNASEKARAN et al., 1998; JACOB; DUCLOHIER; CAFISO, 1999; HO; BRASSEUR, 2005).

Para finalizar esta seção de resultados, os arquivos das instâncias utilizadas e os códigos do *i*BP e do *i*TBP, em MATLAB, que escrevemos, podem ser consultados neste *link* <https://github.com/wagneRocha/MATLABcodes>.

CAPÍTULO

6 Conclusões

Nesta tese, trabalhamos a parte conceitual do Problema Discretizável de Geometria de Distâncias (PDGD) e todo ferramental matemático utilizado para tentar solucionar este problema utilizando uma abordagem combinatória. Como este trabalho versa sobre um problema de natureza interdisciplinar, os conceitos apresentados podem ser de interesse de pessoas que não são da grande área de Matemática ou Física. Desta maneira, desafiamos-nos a apresentar todas as discussões da forma mais simples e intuitiva possível, mas sempre atentos à formalização matemática necessária. Para a parte computacional, apresentamos pseudocódigos com um certo grau de detalhes e disponibilizamos os códigos que escrevemos para consulta.

No [Capítulo 2](#), apresentamos de maneira rápida os conceitos da estrutura tridimensional de proteínas e do experimento de Ressonância Magnética Nuclear, visando somente os aspectos necessários para a modelagem do problema que abordamos nesta tese.

Apresentamos toda parte teórica do PDGD no [Capítulo 3](#). Nesta parte do texto, apresentamos o Problema de Geometria de Distâncias (PGD), o PDGD e o Problema Discretizável de Geometria de Distâncias Intervalares (PDGD i). Para construção deste capítulo, seguimos uma linha em que o nível de hipóteses assumidas nos conceitos teóricos apresentados aumenta gradativamente. Esta estratégia nos possibilitou a construção do algoritmo *interval Branch & Prune* (*iBP*) para o PDGD i em \mathbb{R}^3 de uma maneira inédita. A descrição que apresentamos para o PDGD nos auxiliou na generalização de um processo que consegue restringir os erros provenientes das medidas experimentais para o PDGD i . O que só tinha sido descrito para uma versão mais restrita do PDGD, o Problema Discretizável de Geometria de Distâncias Moleculares (PDGDM).

No [Capítulo 4](#), apresentamos, de maneira inédita, relações geométricas que descrevem os ângulos de torção ϕ e ψ do *backbone* das proteínas com relação a distâncias entre átomos de hidrogênio nas imediações dos átomos que definem estes ângulos. Estas relações foram construídas observando a geometria intrínseca das proteínas. Consideramos que as relações

que envolvem átomos de hidrogênio do mesmo resíduo de aminoácido ou de resíduos adjacentes podem ser de interesse de pesquisadores da área de estrutura de proteínas no geral, pois elas não necessitam de nenhum aspecto teórico do campo da Geometria de Distâncias. Já as relações mais abrangentes, que envolvem distâncias entre átomos de hidrogênio de longo alcance, isto é, entre átomos de resíduos não adjacentes, para serem definidas, necessitam que a estrutura da proteína seja construída átomo por átomo, o que já é um limitante maior. No entanto, estas relações podem ser utilizadas se caracterizarmos a estrutura 3D de uma proteína utilizando a estratégia geral do algoritmo *iBP*.

Acreditamos que a parte mais frutífera deste trabalho é apresentada no [Capítulo 5](#). Neste ponto da tese, propomos um novo algoritmo, que denominamos *protein interval Torsion angle Branch & Prune (iTBP)*. Ele integra uma ordenação dos átomos das proteínas, que é fundamentada nos ângulos de torção ϕ e ψ , e as relações geométricas descritas no capítulo 4 com uma estratégia baseada no algoritmo *iBP* para encontrar a posição espacial dos átomos do *backbone* de uma proteína. Este algoritmo, assim como o *iBP*, é aplicado a uma instância do PDGD*i*, que pela modelagem empregada, é dada por uma lista de restrições intervalares de distâncias e ângulos entre os átomos das proteínas. Os resultados dos experimentos computacionais apresentados nos mostram que o algoritmo proposto é mais eficiente computacionalmente que o *iBP*.

Ainda no [Capítulo 5](#), apresentamos uma nova maneira de se amostrar os intervalos angulares definidos em qualquer estratégia *iBP*. Ela leva em conta um aspecto da Teoria Fuzzy (PAPA; KLIR; FOLGER, 1988), pois considera que alguns pontos dos intervalos são mais prováveis que outros. Baseado no que foi apresentado neste capítulo, afirmamos que a principal contribuição desta tese foi o desenvolvimento de um algoritmo para caracterizar a posição espacial dos átomos do *backbone* das proteínas que integra conhecimentos geométricos intrínsecos das proteínas no processo deste algoritmo. Se, por um lado, a especificidade deste algoritmo o impede de ser utilizado em outras aplicações, por outro, os resultados dos nossos experimentos numéricos mostram que esta abordagem mais direcionada ao problema resulta em ganhos computacionais, o que é suficiente para considerarmos esta perspectiva.

Se, de fato, é possível utilizar a Teoria Fuzzy de maneira formal e quais seriam as implicações desta abordagem na modelagem das distâncias e ângulos intervalares é uma pergunta que não foi respondida por este trabalho. Outro questionamento, que segue direto das discussões apresentadas, seria qual ordem PDGD*i hand-crafted* nos átomos considerados levaria a um melhor resultado computacional do *iBP* e se, neste caso, ainda conseguiríamos constatar uma superioridade computacional do *iTBP*. Todos estes questionamentos são perguntas abertas que podem ser objeto de estudos futuros.



Referências

- ALIPANAHI, B.; KRISLOCK, N.; GHODSI, A.; WOLKOWICZ, H.; DONALDSON, L.; LI, M. Determining protein structures from noesy distance constraints by semidefinite programming. *Journal of computational biology : a journal of computational molecular cell biology*, Mary Ann Liebert, Inc., v. 20, n. 4, p. 296–310, 2013.
- ALVES, R.; LAVOR, C.; SOUZA, C.; SOUZA, M. Clifford algebra and discretizable distance geometry. *Mathematical Methods in the Applied Sciences*, v. 41, n. 11, p. 4063–4073, 2018.
- AMABIS, J. M.; MARTHO, G. R. *CRC Handbook of Chemistry and Physics, 93rd Edition*. [S.l.]: Atualidades Biológicas - Editora Moderna, 1997.
- ANDRICOPULO, A. D.; SALUM, L. B.; ABRAHAM, D. J. Structure-based drug design strategies in medicinal chemistry. *Current Topics in Medicinal Chemistry*, v. 9, n. 9, p. 771–790, 2009.
- ANFENSEN, C. B.; HABER, E.; SELA, M.; JR., F. H. W. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proceedings of the National Academy of Sciences of the United States of America*, v. 47, n. 9, p. 1309–1314, 1961.
- APOSTOL, T. *Mathematical Analysis*. [S.l.]: Addison-Wesley, 1974. (Addison-Wesley series in mathematics).
- BAJAJ, C. The algebraic degree of geometric optimization problems. *Discrete & Computational Geometry*, v. 3, n. 2, p. 177–191, Jun 1988.
- BERMAN, H. M.; WESTBROOK, J.; FENG, Z.; GILLILAND, G.; BHAT, T. N.; WEISSIG, H.; SHINDYALOV, I. N.; BOURNE, P. E. The protein data bank. *Nucleic Acids Research*, v. 28, n. 1, p. 235–242, 2000.
- BISWAS, P.; TOH, K.-C.; YE, Y. A distributed sdp approach for large-scale noisy anchor-free graph realization with applications to molecular conformation. *SIAM Journal on Scientific Computing*, v. 30, n. 3, p. 1251–1277, 2008.
- BLUMENTHAL, L. *Theory and applications of distance geometry*. [S.l.]: Oxford University Press, 1953.

- BODLAENDER, H. L.; FOMIN, F. V.; KOSTER, A. M. C. A.; KRATSCH, D.; THILIKOS, D. M. A note on exact algorithms for vertex ordering problems on graphs. *Theory of Computing Systems*, v. 50, n. 3, p. 420–432, Apr 2012.
- BROWN, A. C. Xliv.-on the theory of isomeric compounds. *Transactions of the Royal Society of Edinburgh*, Royal Society of Edinburgh Scotland Foundation, v. 23, n. 3, p. 707–719, 1864.
- BRUGLIERI, M.; CORDONE, R.; LIBERTI, L. Maximum feasible subsystems of distance geometry constraints. *Journal of Global Optimization*, v. 83, n. 1, p. 29–47, 2022.
- BRÜNGER, A. X-ray crystallography and nmr reveal complementary views of structure and dynamics. *Nature structural biology*, v. 4 Suppl, p. 862–865, 1997.
- COJOCARI, D. *pKa Data: CRC Handbook of Chemistry and Physics, v.2010*. 2010. Princess Margaret Cancer Centre, University of Toronto. Licensed under CC BY-SA 3.0. Disponível em: <https://commons.wikimedia.org/wiki/File:Amino_Acids.svg>. Acesso em: 30 mai 2022.
- CORMEN, T.; LEISERSON, C.; RIVEST, R.; C., S. *Introduction to Algorithms*. [S.l.]: MIT Press, 2009. (Introduction to Algorithms).
- COSTA, V.; MUCHERINO, A.; LAVOR, C.; CASSIOLI, A.; CARVALHO, L. M.; MACULAN, N. Discretization orders for protein side chains. *Journal of Global Optimization*, v. 60, n. 2, p. 333–349, Oct 2014.
- CREIGHTON, T. *The Biophysical Chemistry of Nucleic Acids & Proteins*. [S.l.]: Helvetian Press, 2010.
- CRIPPEN, G.; HAVEL, T. *Distance geometry and molecular conformation*. [S.l.]: Research Studies Press, 1988.
- D'AMBROSIO, C.; VU, K.; LAVOR, C.; LIBERTI, L.; MACULAN, N. New error measures and methods for realizing protein graphs from distance data. *Discrete & Computational Geometry*, v. 57, n. 2, p. 371–418, 2017.
- DOBSON, C. M. Protein folding and misfolding. *Nature*, v. 426, n. 6968, p. 884–890, 2003.
- DONALD, B.; ISTRAIL, S.; PEVZNER, P. *Algorithms in Structural Molecular Biology*. [S.l.]: MIT Press, 2011. (Computational Molecular Biology).
- DRENTH, J.; MESTERS, J. *Principles of Protein X-Ray Crystallography*. [S.l.]: Springer, 2007. (Springer Advanced Texts in Che).
- DUGUNDJI, J. *Topology*. [S.l.]: Allyn and Bacon, 1966. (Allyn and Bacon series in advanced mathematics).
- ENGH, R. A.; HUBER, R. Accurate bond and angle parameters for x-ray protein structure refinement. *Acta Crystallographica Section A*, v. 47, n. 4, p. 392–400, 1991.
- EREN, T.; GOLDENBERG, O. K.; WHITELEY, W.; YANG, Y. R.; MORSE, A. S.; ANDERSON, B. D. O.; BELHUMEUR, P. N. Rigidity, computation, and randomization in network localization. In: *IEEE INFOCOM 2004*. [S.l.: s.n.], 2004. v. 4, p. 2673–2684.

FRANK, J. *Single-particle Cryo-electron Microscopy: The Path Toward Atomic Resolution: Selected Papers of Joachim Frank with Commentaries*. [S.l.]: World Scientific Publishing Company Pte Limited, 2018. (Series in structural biology).

GONÇALVES, D. S.; MUCHERINO, A. Discretization orders and efficient computation of cartesian coordinates for distance geometry. *Optimization Letters*, v. 8, n. 7, p. 2111–2125, 2014.

GONÇALVES, D. S.; MUCHERINO, A.; LAVOR, C.; LIBERTI, L. Recent advances on the interval distance geometry problem. *Journal of Global Optimization*, v. 69, n. 3, p. 525–545, 2017.

GUNASEKARAN, K.; NAGARAJARAM, H.; RAMAKRISHNAN, C.; BALARAM, P. Stereochemical punctuation marks in protein structures: glycine and proline containing helix stop signals. *Journal of Molecular Biology*, v. 275, n. 5, p. 917–932, 1998.

GÜNTERT, P. Structure calculation of biological macromolecules from nmr data. *Quarterly Reviews of Biophysics*, Cambridge University Press, v. 31, n. 2, p. 145–237, 1998.

GÜNTHER, H. *NMR Spectroscopy: Basic Principles, Concepts and Applications in Chemistry*. [S.l.]: Wiley, 2013.

HALL, R. M.; COLLIS, C. M. Mobile gene cassettes and integrons: Capture and spread of genes by site-specific recombination. *Molecular Microbiology*, v. 15, n. 4, p. 593–600, 1995.

HENNEBERG, L. *Statik der Starren Systeme*. [S.l.]: Bergstraesser, Darmstadt, 1886.

HESTENES, D. *New Foundations for Classical Mechanics*. [S.l.]: Springer Netherlands, 1999. (Fundamental Theories of Physics).

HO, B. K.; BRASSEUR, R. The ramachandran plots of glycine and pre-proline. *BMC Structural Biology*, v. 5, n. 1, p. 14, Aug 2005.

HOFFMAN, K.; KUNZE, R. *Linear Algebra*. [S.l.]: Prentice-Hall, 1971.

JACOB, J.; DUCLOHIER, H.; CAFISO, D. S. The role of proline and glycine in determining the backbone flexibility of a channel-forming peptide. *Biophysical journal*, v. 76, n. 3, p. 1367–1376, Mar 1999.

KEELER, J. *Understanding NMR Spectroscopy*. [S.l.]: Wiley, 2011.

LAVOR, C.; LEE, J.; JOHN, A. L.-S.; LIBERTI, L.; MUCHERINO, A.; SVIRIDENKO, M. Discretization orders for distance geometry problems. *Optimization Letters*, v. 6, p. 783–796, 2012.

LAVOR, C.; LIBERTI, L.; DONALD, B.; WORLEY, B.; BARDIAUX, B.; MALLIAVIN, T. E.; NILGES, M. Minimal nmr distance information for rigidity of protein graphs. *Discrete Applied Mathematics*, v. 256, p. 91–104, 2019. Distance Geometry Theory and Applications (DGTA 16).

LAVOR, C.; LIBERTI, L.; MACULAN, N. Computational experience with the molecular distance geometry problem. In: _____. *Global Optimization: Scientific and Engineering Case Studies*. Boston, MA: Springer US, 2006. p. 213–225.

LAVOR, C.; LIBERTI, L.; MACULAN, N.; MUCHERINO, A. The discretizable molecular distance geometry problem. *Computational Optimization and Applications*, v. 52, n. 1, p. 115–146, May 2012.

_____. Recent advances on the discretizable molecular distance geometry problem. *European Journal of Operational Research*, v. 219, n. 3, p. 698–706, 2012.

LAVOR, C.; LIBERTI, L.; MUCHERINO, A. The interval branch-and-prune algorithm for the discretizable molecular distance geometry problem with inexact distances. *Journal of Global Optimization*, v. 56, n. 3, p. 855–871, 2013.

LAVOR, C.; MACULAN, N.; SOUZA, M.; ALVES, R. *Álgebra e Geometria no Cálculo de Estrutura Molecular*. [S.l.]: Instituto Nacional de Matemática Pura e Aplicada - IMPA, 2017. (31° Colóquio Brasileiro de Matemática).

LAVOR, C.; MUCHERINO, A.; LIBERTI, L.; MACULAN, N. On the computation of protein backbones by using artificial backbones of hydrogens. *Journal of Global Optimization*, v. 50, n. 2, p. 329–344, 2011.

LAVOR, C.; OLIVEIRA, A.; ROCHA, W.; SOUZA, M. On the optimality of finding DMDGP symmetries. *Computational and Applied Mathematics*, v. 40, n. 3, p. 98, 2021.

LAVOR, C.; SOUZA, M.; CARVALHO, L. M.; GONÇALVES, D. S.; MUCHERINO, A. Improving the sampling process in the interval branch-and-prune algorithm for the discretizable molecular distance geometry problem. *Applied Mathematics and Computation*, v. 389, p. 125586, 2021. ISSN 0096-3003.

LIBERTI, L.; LAVOR, C.; MACULAN, N. A branch-and-prune algorithm for the molecular distance geometry problem. *International Transactions in Operational Research*, v. 15, n. 1, p. 1–17, 2008.

LIBERTI, L.; LAVOR, C.; MACULAN, N.; MUCHERINO, A. Euclidean distance geometry and applications. *SIAM Review*, v. 56, n. 1, p. 3–69, 2014.

LUCAS, É. *Récréations mathématiques*. [S.l.]: Gauthier-Villars, 1893. (Récréations mathématiques).

MADDOX, B. The double helix and the ‘wronged heroine’. *Nature*, v. 421, n. 6921, p. 407–408, 2003.

MAIOLI, D. S.; LAVOR, C.; GONÇALVES, D. S. A note on computing the intersection of spheres in r^n . *The ANZIAM Journal*, v. 59, n. 2, p. 271–279, 2017.

MAIOROV, V. N.; CRIPPEN, G. M. Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins. *Journal of Molecular Biology*, v. 235, n. 2, p. 625–634, 1994.

MALLIAVIN, T. E.; MUCHERINO, A.; LAVOR, C.; LIBERTI, L. Systematic exploration of protein conformational space using a distance geometry approach. *Journal of Chemical Information and Modeling*, v. 59, n. 10, p. 4486–4503, 2019.

MATTHEWS, B. W. How planar are planar peptide bonds? *Protein science : a publication of the Protein Society*, John Wiley and Sons Inc., v. 25, n. 4, p. 776–777, 2016.

- MENGER, K. Untersuchungen über allgemeine metrik. *Mathematische Annalen*, v. 100, p. 75–163, 1928.
- MUCHERINO, A. A pseudo de bruijn graph representation for discretization orders for distance geometry. In: _____. *Proceedings of the 3rd International Work-Conference on Bioinformatics and Biomedical Engineering (IWBBIO15), Part I, Lecture Notes in Bioinformatics*. Granada, Spain: Springer International Publishing, 2015. v. 9043, p. 514–523.
- MUCHERINO, A.; LAVOR, C.; LIBERTI, L. The discretizable distance geometry problem. *Optimization Letters*, v. 6, n. 8, p. 1671–1686, 2012.
- _____. Exploiting symmetry properties of the discretizable molecular distance geometry problem. *Journal of Bioinformatics and Computational Biology*, v. 10, n. 03, p. 1242009, 2012.
- MULDER, G. J. Over proteine en hare verbindingen en ontledingsproducten. *Natuuren scheikundig Archief*, v. 1, p. 104, 1838.
- _____. Sur le composition de quelques substances animales. *Bulletin des sciences physiques et naturelles en Néerlande*, v. 6, p. 87–162, 1838.
- NATURE. Due credit. *A cell for all seasons*, v. 496, n. 7446, p. 270, Abril 2013.
- NELSON, D.; COX, M. *Lehninger Principles of Biochemistry*. Quarta. [S.l.]: W. H. Freeman, 2012.
- NEUMAIER, A. Molecular modeling of proteins and mathematical prediction of protein structure. *SIAM Review*, v. 39, n. 3, p. 407–460, 1997.
- NOBELPRIZE. *The Nobel Prize in Physiology or Medicine 1962*. 1962. Nobel Prize Outreach AB 2022. Disponível em: <<https://www.nobelprize.org/prizes/medicine/1962/summary/>>. Acesso em: 31 mai 2022.
- _____. *The Nobel Prize in Chemistry 1972*. 1972. Nobel Prize Outreach AB 2022. Disponível em: <<https://www.nobelprize.org/prizes/chemistry/1972/summary/>>. Acesso em: 31 mai 2022.
- _____. *The Nobel Prize in Chemistry 2002*. 2002. Nobel Prize Outreach AB 2022. Disponível em: <<https://www.nobelprize.org/prizes/chemistry/2002/summary/>>. Acesso em: 31 mai 2022.
- PAPA, C.; KLIR, G.; FOLGER, T. *Fuzzy Sets, Uncertainty, and Information*. [S.l.]: Prentice Hall, 1988. (Prentice-Hall International editions).
- PAPERS, J. W. Double helix, 50th anniversary collection - reprint of vol. 421, no. 6921. *Nature*, 2003.
- PETITJEAN, M. Sphere unions and intersections and some of their applications in molecular modeling. *Distance Geometry: Theory, Methods, and Applications*, Springer, New York, p. 61–83, 2013.
- POGORELOV, A. *Geometry*. Moscow: Mir, 1987.

- RAMACHANDRAN, G.; RAMAKRISHNAN, C.; SASISEKHARAN, V. Stereochemistry of polypeptide chain configurations. *Journal of Molecular Biology*, v. 7, n. 1, p. 95–99, 1963.
- REVA, B. A.; FINKELSTEIN, A. V.; SKOLNICK, J. What is the probability of a chance prediction of a protein structure with an rmsd of 6 Å? *Folding and Design*, v. 3, n. 2, p. 141–147, 1998.
- ROSSI, F.; BEEK, P. van; WALSH, T. *Handbook of Constraint Programming*. [S.l.]: Elsevier Science, 2006. (Foundations of Artificial Intelligence).
- SAXE, J. *Embeddability of Weighted Graphs in K-space is Strongly NP-hard*. [S.l.]: Carnegie-Mellon University, Department of Computer Science, 1980. (CMU-CS-80-102).
- SCHLICK, T. *Molecular Modeling and Simulation: An Interdisciplinary Guide*. [S.l.]: Springer New York, 2013. (Interdisciplinary Applied Mathematics).
- SHEN, Y.; BAX, A. Protein backbone and sidechain torsion angles predicted from nmr chemical shifts using artificial neural networks. *Journal of Biomolecular NMR*, v. 56, n. 3, p. 227–241, 2013.
- STRANG, G. *Introduction to Linear Algebra*. [S.l.]: Wellesley-Cambridge Press, 2016.
- THORNTON, S.; MARION, J. *Classical Dynamics of Particles and Systems*. [S.l.]: Brooks/Cole, 2004.
- VICKERY, H. B. The origin of the word protein. *The Yale journal of biology and medicine*, v. 22, n. 5, p. 387–393, 1950.
- VÖGELI, B.; OLSSON, S.; GÜNTERT, P.; RIEK, R. The exact noe as an alternative in ensemble structure determination. *Biophysical Journal*, v. 110, n. 1, p. 113–126, 2016.
- WATSON, J. D.; CRICK, F. H. C. Molecular structure of nucleic acids: A structure for deoxyribose nucleic acid. *Nature*, v. 171, n. 4356, p. 737–738, 1953.
- WEXLER, C. *Analytic Geometry: A Vector Approach*. [S.l.]: Addison-Wesley Publishing Company, 1962. (Addison-Wesley series in introductory mathematics).
- WILKINS, M. H. F.; STOKES, A. R.; WILSON, H. R. Molecular structure of nucleic acids: Molecular structure of deoxypentose nucleic acids. *Nature*, v. 171, n. 4356, p. 738–740, 1953.
- WÜTHRICH, K. *NMR of Proteins and Nucleic Acids*. [S.l.]: Wiley, 1986. (A Wiley-interscience publication).
- _____. Protein structure determination in solution by nuclear magnetic resonance spectroscopy. *Science*, American Association for the Advancement of Science, v. 243, n. 4887, p. 45–50, 1989.

APÊNDICE

A Descrições Geométricas Utilizadas na Seção 4.3

Utilizamos as duas seções seguintes para desenvolver o aspectos geométricos que utilizamos para descrever a relação dos ângulos de torção ψ_i , τ_{N^i, H_k^j} , $\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}}$ e $\tau_{H_k^j, H_{\mathbb{N}}^{i+1}}$ que é apresentada na Seção 4.3.

A.1 Situação (i): $\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}} \geq 0^\circ$

Caso 1: $0^\circ \leq \tau_{N^i, H_k^j} \leq \tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}}$

Para este caso, a [Figura 48](#) evidencia a disposição dos átomos H_k^j , N^i , C^i , N^{i+1} e $H_{\mathbb{N}}^{i+1}$ para as seguintes situações: $\psi_i = 0^\circ$, $|\psi_i| = 180^\circ$, $\psi_i = {}^1\psi_i^{*+}$ e $\psi_i = {}^1\psi_i^{*-}$. Portanto, observa-se que os valores ${}^1\psi_i^{*\pm}$ podem ser descritos em função da quantidade $\psi_{i,j,k}$, que é descrita por (4.29).

Sobre as quantidades ${}^1\psi_i^{*-} = \psi_{i,j,k}$ e ${}^1\psi_i^{*+} = {}^1\psi_i^{*-} + 180^\circ$, observamos que:

$$\begin{aligned}
 0^\circ &\leq \tau_{N^i, H_k^j} \leq \tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}} && \Leftrightarrow \\
 -\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}} &\leq \psi_{i,j,k} \leq 0^\circ && \Leftrightarrow \\
 -\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}} &\leq {}^1\psi_i^{*-} \leq 0^\circ && \Leftrightarrow \\
 180^\circ - \tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}} &\leq {}^1\psi_i^{*+} \leq 180^\circ.
 \end{aligned} \tag{A.1}$$

Agora, utilizando a intuição do desenho geométrico, na [Figura 49](#) sistematizamos as relações que descrevem a dinâmica entre os ângulos de torção ψ_i , τ_{N^i, H_k^j} , $\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}}$ e $\tau_{H_k^j, H_{\mathbb{N}}^{i+1}}$. Os casos apresentados se distinguem com relação a orientação e ao intervalo considerado do ângulo $\tau_{H_k^j, H_{\mathbb{N}}^{i+1}}$. Estes intervalos são dados com relação as quantidades descritas na [Figura 48](#). Desta maneira, considerando as relações apresentadas na [Figura 49](#), podemos escrever o seguinte lema:

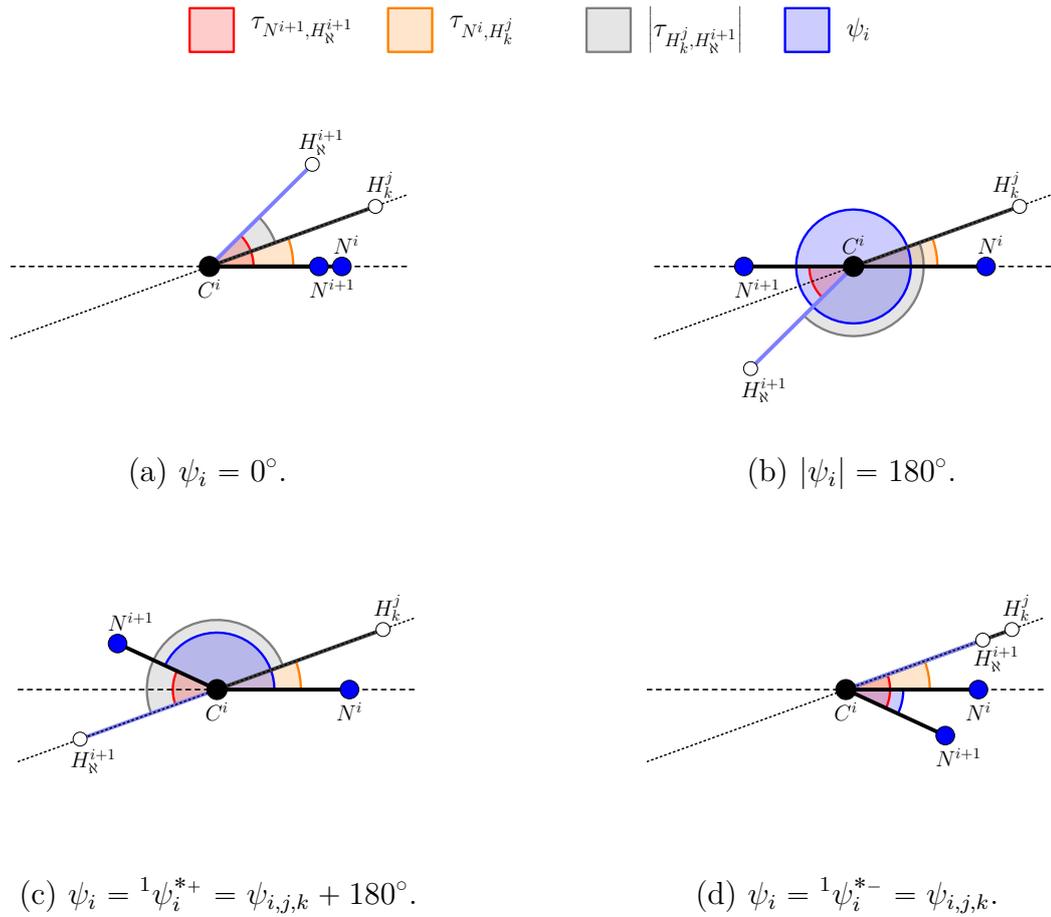


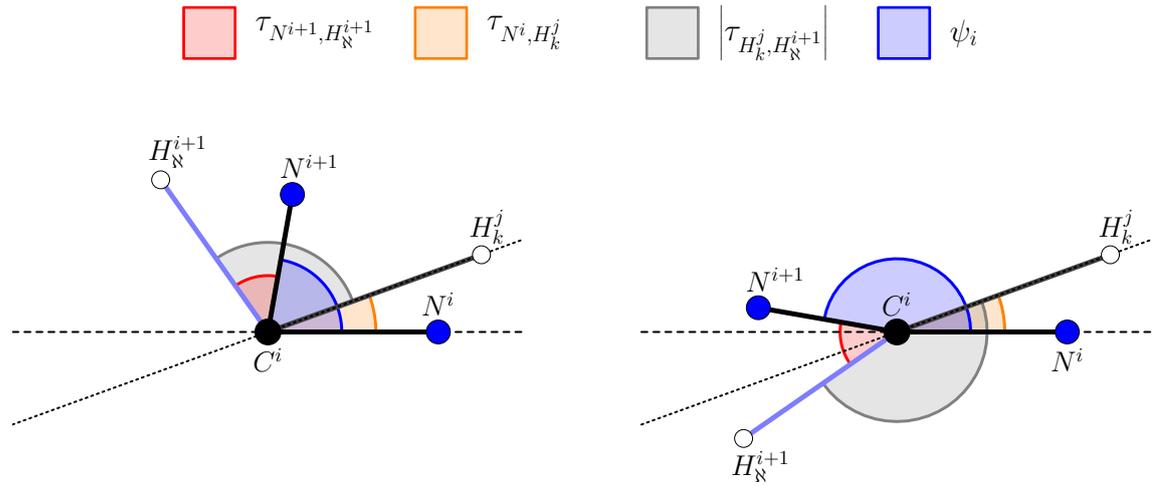
Figura 48 – Disposição no plano da projeção ortogonal dos átomos H_k^j , N^i , C^i , N^{i+1} e H_N^{i+1} , com $\psi_i = 0^\circ$, $|\psi_i| = 180^\circ$, $\psi_i = {}^1\psi_i^{*+}$ e $\psi_i = {}^1\psi_i^{*-}$, para o Caso 1.

Lema A.1. Sejam ψ_i , τ_{N^i, H_k^j} , $\tau_{N^{i+1}, H_N^{i+1}}$, $\tau_{H_k^j, H_N^{i+1}}$ os ângulos apresentados na Figura 49. Considere que os valores de τ_{N^i, H_k^j} e $\tau_{N^{i+1}, H_N^{i+1}}$ são conhecidos de modo que $0^\circ \leq \tau_{N^i, H_k^j} \leq \tau_{N^{i+1}, H_N^{i+1}}$ e o valor absoluto do ângulo $\tau_{H_k^j, H_N^{i+1}}$ é conhecido. Então, o ângulo $\psi_i \in (-180^\circ, 180^\circ]$ pode ser escrito de duas formas, dadas por ψ_i^+ e ψ_i^- , em função dos outros três ângulos dados, da seguinte maneira:

$$\psi_i^+ = \psi_{i,j,k} + \left| \tau_{H_k^j, H_N^{i+1}} \right|.$$

$$\psi_i^- = \begin{cases} \psi_{i,j,k} - \left| \tau_{H_k^j, H_N^{i+1}} \right| & \text{se } \left| \psi_{i,j,k} - \left| \tau_{H_k^j, H_N^{i+1}} \right| \right| < 180^\circ, \\ \psi_{i,j,k} - \left| \tau_{H_k^j, H_N^{i+1}} \right| + 360^\circ & \text{se } \psi_{i,j,k} - \left| \tau_{H_k^j, H_N^{i+1}} \right| \leq -180^\circ. \end{cases}$$

Demonstração. A demonstração deste lema pode ser feita de maneira análoga a do Lema 4.1. No entanto, neste caso, considerando as desigualdades de (A.1) para observar o valor de $\psi_{i,j,k} \pm \left| \tau_{H_k^j, H_N^{i+1}} \right|$. \square



(a₁) $0^\circ \leq \psi_i \leq {}^1\psi_i^{*+}$

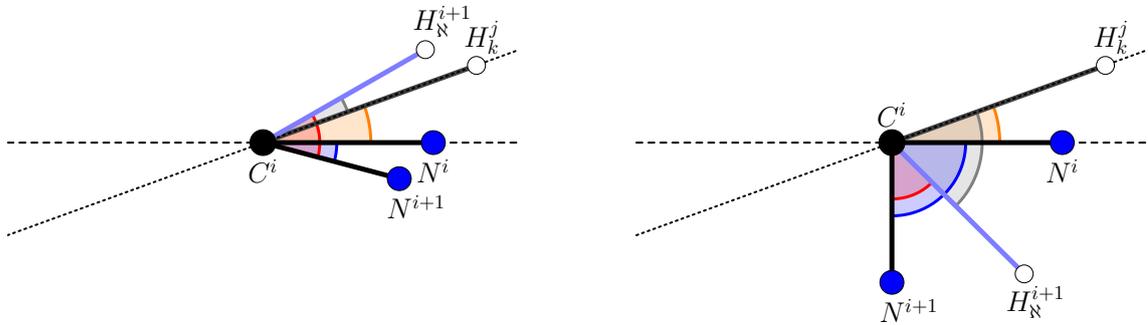
(b₁) ${}^1\psi_i^{*+} \leq \psi_i \leq 180^\circ$

(a₂) $-{}^1\psi_i^{*-} \leq |\tau_{H_k^j, H_N^{i+1}}| \leq 180^\circ$

(b₂) ${}^1\psi_i^{*+} \leq |\tau_{H_k^j, H_N^{i+1}}| \leq 180^\circ$

(a₃) $\psi_i = \psi_{i,j,k} + |\tau_{H_k^j, H_N^{i+1}}|$

(b₃) $\psi_i = \psi_{i,j,k} - |\tau_{H_k^j, H_N^{i+1}}| + 360^\circ$



(c₁) ${}^1\psi_i^{*-} \leq \psi_i \leq 0^\circ$

(d₁) $-180^\circ \leq \psi_i \leq {}^1\psi_i^{*-}$

(c₂) $0 \leq |\tau_{H_k^j, H_N^{i+1}}| \leq -{}^1\psi_i^{*-}$

(d₂) $0 \leq |\tau_{H_k^j, H_N^{i+1}}| \leq {}^1\psi_i^{*+}$

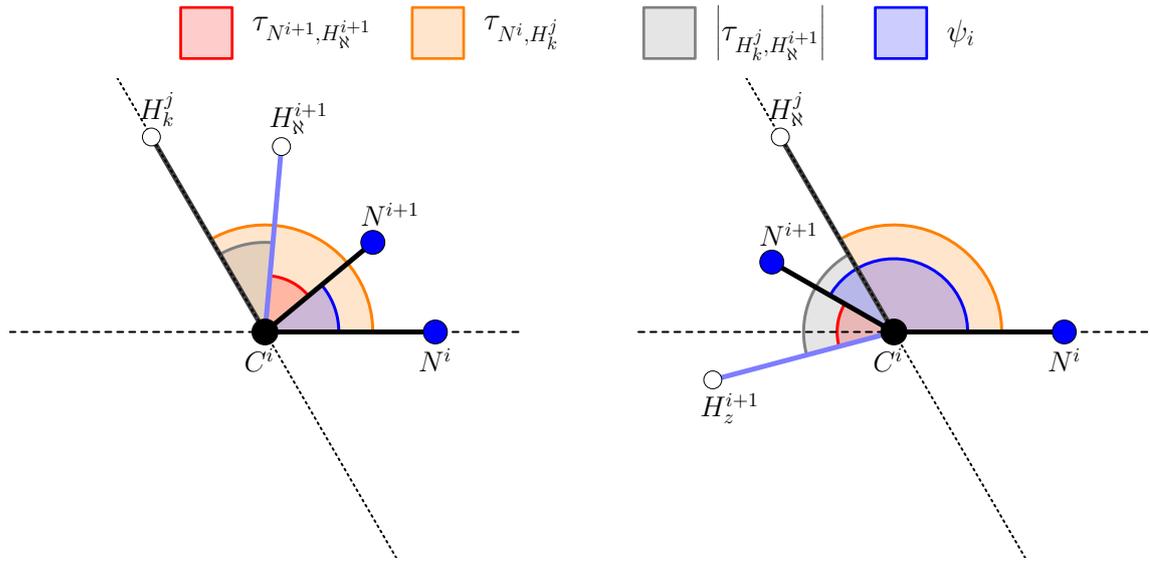
(c₃) $\psi_i = \psi_{i,j,k} + |\tau_{H_k^j, H_N^{i+1}}|$

(d₃) $\psi_i = \psi_{i,j,k} - |\tau_{H_k^j, H_N^{i+1}}|$

Figura 49 – Dinâmica do ângulo ψ_i com relação aos ângulos τ_{N^i, H_k^j} , $\tau_{N^{i+1}, H_N^{i+1}}$ e $|\tau_{H_k^j, H_N^{i+1}}|$ para o Caso 1.

Caso 2: $\tau_{N^{i+1}, H_N^{i+1}} < \tau_{N^i, H_k^j} \leq 180^\circ$

Para este caso, a Figura 50 evidencia a disposição dos átomos H_k^j , N^i , C^i , N^{i+1} e H_N^{i+1} para as seguintes situações: $\psi_i = 0^\circ$, $|\psi_i| = 180^\circ$, $\psi_i = {}^2\psi_i^{*+}$ e $\psi_i = {}^2\psi_i^{*-}$.



(a₁) $0^\circ \leq \psi_i \leq \psi_i^{*+}$

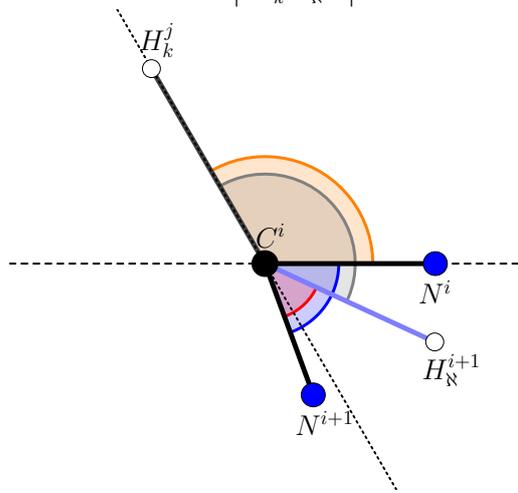
(a₂) $0^\circ \leq \left| \tau_{H_k^j, H_N^{i+1}} \right| \leq \psi_i^{*+}$

(a₃) $\psi_i = \psi_{i,j,k} - \left| \tau_{H_k^j, H_N^{i+1}} \right|$

(b₁) $\psi_i^{*+} \leq \psi_i \leq 180^\circ$

(b₂) $0^\circ \leq \left| \tau_{H_k^j, H_N^{i+1}} \right| \leq -\psi_i^{*-}$

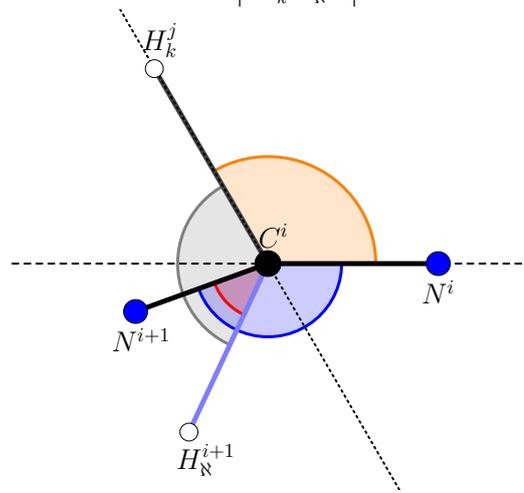
(b₃) $\psi_i = \psi_{i,j,k} + \left| \tau_{H_k^j, H_N^{i+1}} \right|$



(c₁) $\psi_i^{*-} \leq \psi_i \leq 0^\circ$

(c₂) $\psi_i^{*+} \leq \left| \tau_{H_k^j, H_N^{i+1}} \right| \leq 180^\circ$

(c₃) $\psi_i = \psi_{i,j,k} - \left| \tau_{H_k^j, H_N^{i+1}} \right|$



(d₁) $-180^\circ \leq \psi_i \leq \psi_i^{*-}$

(d₂) $-\psi_i^{*-} \leq \left| \tau_{H_k^j, H_N^{i+1}} \right| \leq 180^\circ$

(d₃) $\psi_i = \psi_{i,j,k} + \left| \tau_{H_k^j, H_N^{i+1}} \right| - 360^\circ$

Figura 51 – Dinâmica do ângulo ψ_i com relação aos ângulos τ_{N^i, H_k^j} , $\tau_{N^{i+1}, H_N^{i+1}}$ e $\left| \tau_{H_k^j, H_N^{i+1}} \right|$ para o Caso 2.

$(-180^\circ, 180^\circ]$ pode ser escrito de duas formas, dadas por ψ_i^+ e ψ_i^- , em função dos outros

três ângulos dados, da seguinte maneira:

$$\psi_i^+ = \begin{cases} \psi_{i,j,k} + \left| \tau_{H_k^j, H_N^{i+1}} \right| & \text{se } \left| \psi_{i,j,k} + \left| \tau_{H_k^j, H_N^{i+1}} \right| \right| \leq 180^\circ, \\ \psi_{i,j,k} + \left| \tau_{H_k^j, H_N^{i+1}} \right| - 360^\circ & \text{se } \left| \psi_{i,j,k} + \left| \tau_{H_k^j, H_N^{i+1}} \right| \right| > 180^\circ. \end{cases}$$

$$\psi_i^- = \psi_{i,j,k} - \left| \tau_{H_k^j, H_N^{i+1}} \right|.$$

Demonstração. A demonstração deste lema pode ser feita de maneira análoga a do Lema 4.1. No entanto, neste caso, consideramos as desigualdades de (A.2) para observar o valor de $\psi_{i,j,k} \pm \left| \tau_{H_k^j, H_N^{i+1}} \right|$. \square

Caso 3: $-180^\circ + \tau_{N^{i+1}, H_N^{i+1}} \leq \tau_{N^i, H_k^j} < 0^\circ$

Para este caso, a Figura 52 evidencia a disposição dos átomos H_k^j , N^i , C^i , N^{i+1} e H_N^{i+1} para as seguintes situações: $\psi_i = 0^\circ$, $|\psi_i| = 180^\circ$, $\psi_i = {}^3\psi_i^{*+}$ e $\psi_i = {}^3\psi_i^{*-}$.

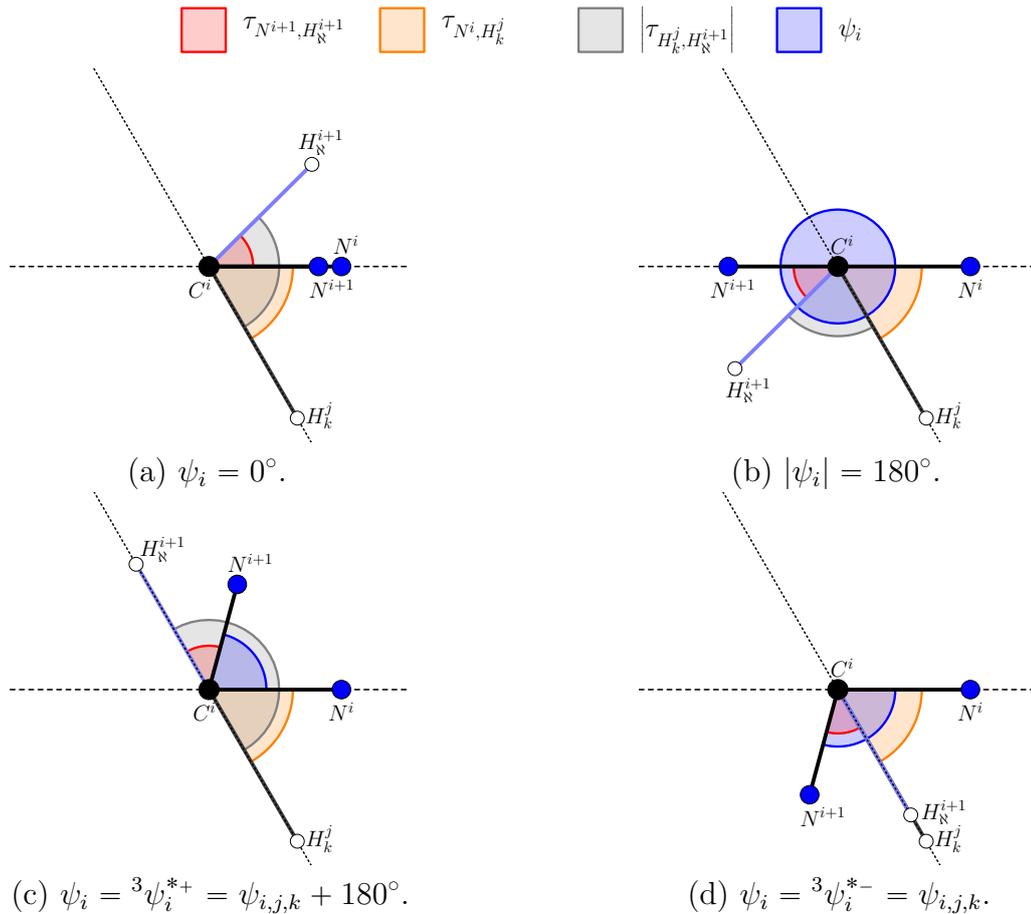


Figura 52 – Disposição no plano da projeção ortogonal dos átomos H_k^j , N^i , C^i , N^{i+1} e H_N^{i+1} , com $\psi_i = 0^\circ$, $|\psi_i| = 180^\circ$, $\psi_i = {}^3\psi_i^{*+}$ e $\psi_i = {}^3\psi_i^{*-}$, para o Caso 3.

Sobre as quantidades ${}^3\psi_i^{*-} = \psi_{i,j,k}$ e ${}^3\psi_i^{*+} = {}^3\psi_i^{*-} + 180^\circ$, observamos que:

$$\begin{aligned}
 -180^\circ + \tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}} &\leq \tau_{N^i, H_k^j} < 0^\circ && \Leftrightarrow \\
 -180^\circ &\leq \psi_{i,j,k} < -\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}} && \Leftrightarrow \\
 -180^\circ &\leq {}^3\psi_i^{*-} < -\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}} && \Leftrightarrow \\
 0^\circ &\leq {}^3\psi_i^{*+} < 180^\circ - \tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}}. &&
 \end{aligned} \tag{A.3}$$

Como no caso anterior, pela intuição do desenho geométrico, na [Figura 53](#) sistematizamos as relações que descrevem a dinâmica entre os ângulos de torção ψ_i , τ_{N^i, H_k^j} , $\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}}$ e $\tau_{H_k^j, H_{\mathbb{N}}^{i+1}}$.

Assim, considerando as relações apresentadas na [Figura 53](#), podemos escrever o seguinte lema:

Lema A.3. Sejam ψ_i , τ_{N^i, H_k^j} , $\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}}$, $\tau_{H_k^j, H_{\mathbb{N}}^{i+1}}$ os ângulos apresentados na [Figura 53](#). Considere que os valores de τ_{N^i, H_k^j} e $\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}}$ são conhecidos de modo que $\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}} \geq 0^\circ$ e $-180^\circ + \tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}} \leq \tau_{N^i, H_k^j} < 0^\circ$; e o valor absoluto do ângulo $\tau_{H_k^j, H_{\mathbb{N}}^{i+1}}$ é conhecido. Então, o ângulo $\psi_i \in (-180^\circ, 180^\circ]$ pode ser escrito de duas formas, dadas por ψ_i^+ e ψ_i^- , em função dos outros três ângulos dados, da seguinte maneira:

$$\begin{aligned}
 \psi_i^+ &= \psi_{i,j,k} + \left| \tau_{H_k^j, H_{\mathbb{N}}^{i+1}} \right|. \\
 \psi_i^- &= \begin{cases} \psi_{i,j,k} - \left| \tau_{H_k^j, H_{\mathbb{N}}^{i+1}} \right| & \text{se } \left| \psi_{i,j,k} - \left| \tau_{H_k^j, H_{\mathbb{N}}^{i+1}} \right| \right| < 180^\circ, \\ \psi_{i,j,k} - \left| \tau_{H_k^j, H_{\mathbb{N}}^{i+1}} \right| + 360^\circ & \text{se } \psi_{i,j,k} - \left| \tau_{H_k^j, H_{\mathbb{N}}^{i+1}} \right| \leq -180^\circ. \end{cases}
 \end{aligned}$$

Demonstração. A demonstração deste lema pode ser feita de maneira análoga a do [Lema 4.1](#). No entanto, neste caso, consideramos as desigualdades de [\(A.3\)](#) para observar o valor de $\psi_{i,j,k} \pm \left| \tau_{H_k^j, H_{\mathbb{N}}^{i+1}} \right|$. \square

Pelo resultado observado nos [Lemas A.1](#) e [A.3](#), notamos que a relação entre os ângulos de torção analisados é a mesma se $-180^\circ + \tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}} \leq \tau_{N^i, H_k^j} \leq \tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}}$. A justificativa desta observação reside no fato de que os valores ${}^1\psi_i^{*\pm}$ e ${}^3\psi_i^{*\pm}$ são escritos em função da quantidade $\psi_{i,j,k}$ da mesma maneira e que a orientação do ângulo $\tau_{H_k^j, H_{\mathbb{N}}^{i+1}}$ é a mesma se $\psi_i \in [{}^m\psi_i^{*-}, {}^m\psi_i^{*+}]$ ou $\psi_i \in [-180^\circ, {}^m\psi_i^{*-}] \cup [{}^m\psi_i^{*+}, 180^\circ]$, com $m = 1, 3$.

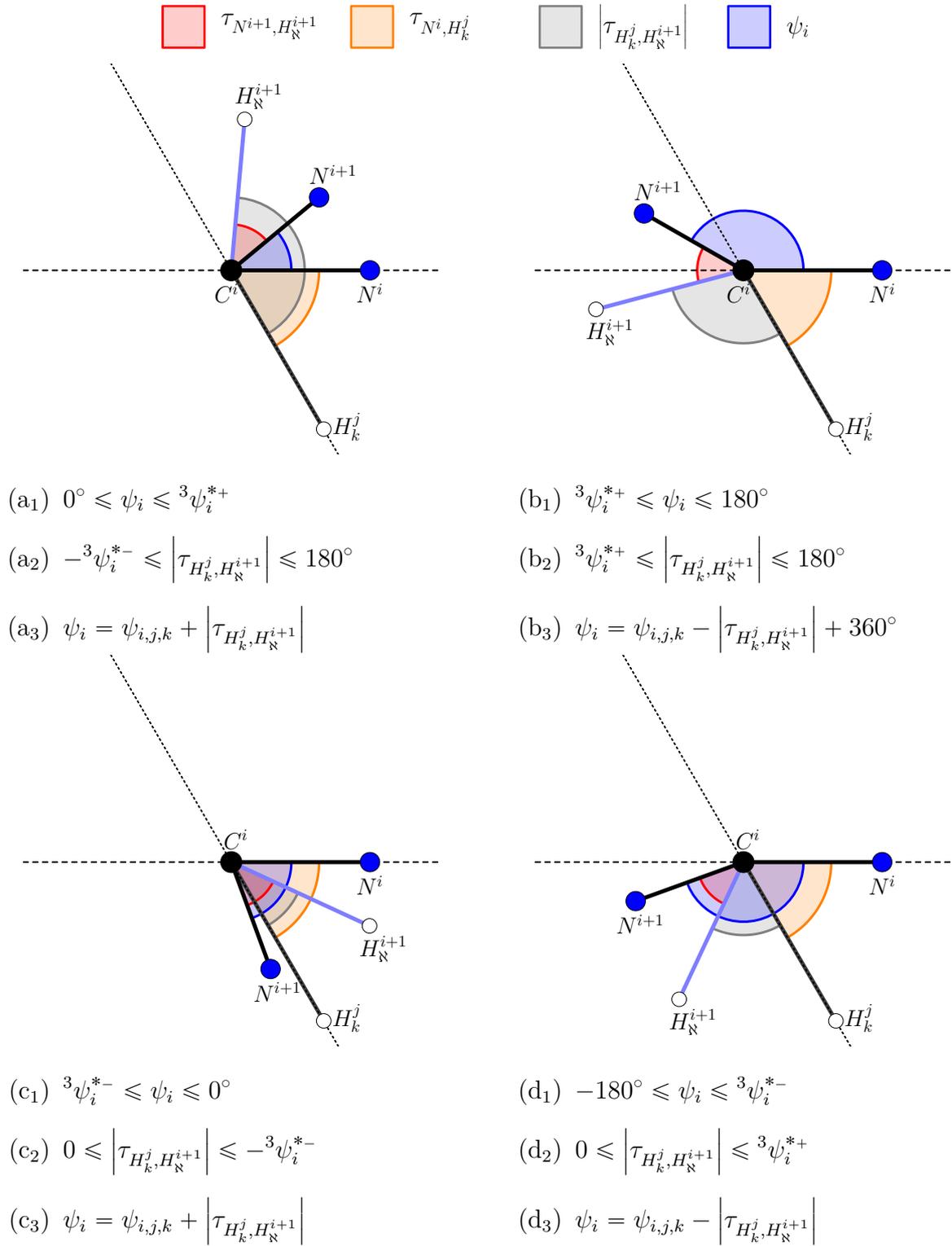


Figura 53 – Dinâmica do ângulo ψ_i com relação aos ângulos τ_{N^i, H_k^j} , $\tau_{N^{i+1}, H_N^{i+1}}$ e $\left| \tau_{H_k^j, H_N^{i+1}} \right|$ para o Caso 3.

Caso 4: $-180^\circ < \tau_{N^i, H_k^j} < -180^\circ + \tau_{N^{i+1}, H_N^{i+1}}$

Para este caso, a Figura 54 evidencia a disposição dos átomos H_k^j , N^i , C^i , N^{i+1} e H_N^{i+1} para as seguintes situações: $\psi_i = 0^\circ$, $|\psi_i| = 180^\circ$, $\psi_i = {}^4\psi_i^{*+}$ e $\psi_i = {}^4\psi_i^{*-}$.

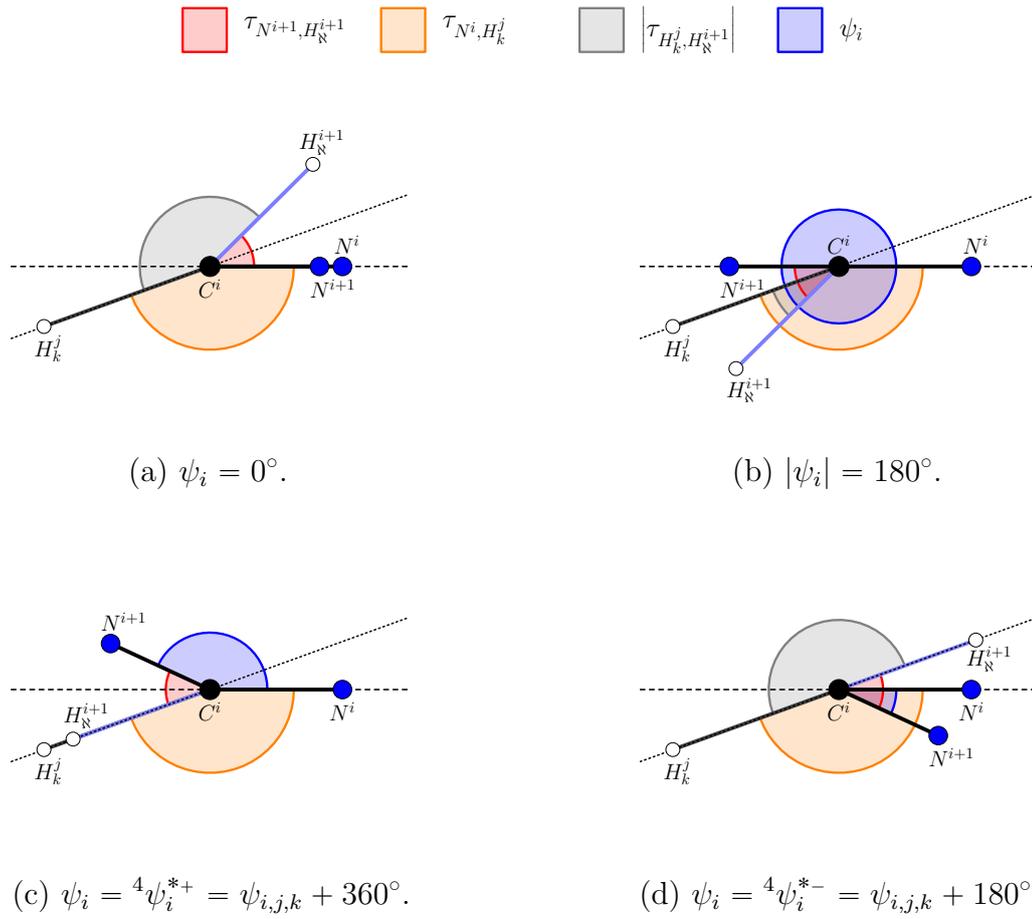


Figura 54 – Disposição no plano da projeção ortogonal dos átomos H_k^j , N^i , C^i , N^{i+1} e H_N^{i+1} , com $\psi_i = 0^\circ$, $|\psi_i| = 180^\circ$, $\psi_i = {}^4\psi_i^{*+}$ e $\psi_i = {}^4\psi_i^{*-}$, para o Caso 4.

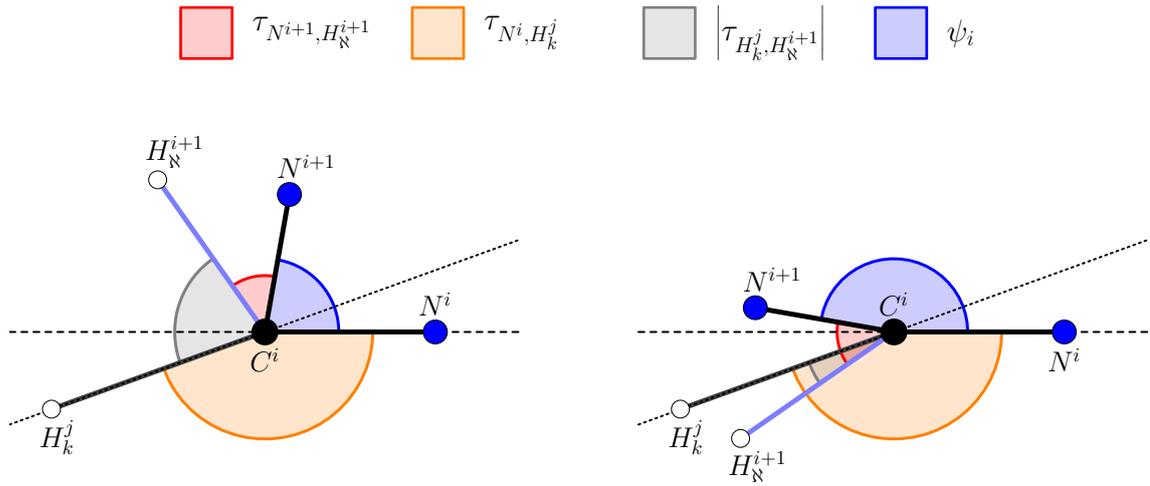
Sobre as quantidades ${}^4\psi_i^{*-} = \psi_{i,j,k} + 180^\circ$ e ${}^4\psi_i^{*+} = {}^4\psi_i^{*-} + 180^\circ$, observamos que:

$$\begin{aligned}
 -180^\circ &< \tau_{N^i, H_k^j} < -180^\circ + \tau_{N^{i+1}, H_N^{i+1}} && \Leftrightarrow \\
 -180^\circ - \tau_{N^{i+1}, H_N^{i+1}} &< \psi_{i,j,k} < -180^\circ && \Leftrightarrow \\
 -\tau_{N^{i+1}, H_N^{i+1}} &< {}^4\psi_i^{*-} < 0^\circ && \Leftrightarrow \\
 180^\circ - \tau_{N^{i+1}, H_N^{i+1}} &< {}^4\psi_i^{*+} < 180^\circ. &&
 \end{aligned} \tag{A.4}$$

Como no caso anterior, pela intuição do desenho geométrico, na Figura 55 sistematizamos as relações que descrevem a dinâmica entre os ângulos de torção ψ_i , τ_{N^i, H_k^j} , $\tau_{N^{i+1}, H_N^{i+1}}$ e $\tau_{H_k^j, H_N^{i+1}}$.

Assim, considerando as relações apresentadas na Figura 55, podemos escrever o seguinte lema:

Lema A.4. Sejam ψ_i , τ_{N^i, H_k^j} , $\tau_{N^{i+1}, H_N^{i+1}}$, $\tau_{H_k^j, H_N^{i+1}}$ os ângulos apresentados na Figura 55. Considere que os valores de τ_{N^i, H_k^j} e $\tau_{N^{i+1}, H_N^{i+1}}$ são conhecidos de modo que $\tau_{N^{i+1}, H_N^{i+1}} \geq 0^\circ$



(a₁) $0^\circ \leq \psi_i \leq {}^4\psi_i^{*+}$

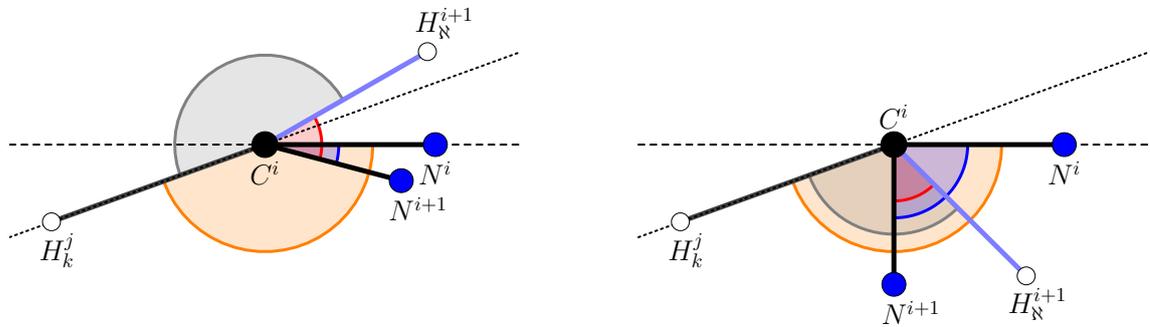
(a₂) $0^\circ \leq \left| \tau_{H_k^j, H_N^{i+1}} \right| \leq {}^4\psi_i^{*+}$

(a₃) $\psi_i = \psi_{i,j,k} - \left| \tau_{H_k^j, H_N^{i+1}} \right| + 360^\circ$

(b₁) ${}^4\psi_i^{*+} \leq \psi_i \leq 180^\circ$

(b₂) $0^\circ \leq \left| \tau_{H_k^j, H_N^{i+1}} \right| \leq -{}^4\psi_i^{*-}$

(b₃) $\psi_i = \psi_{i,j,k} + \left| \tau_{H_k^j, H_N^{i+1}} \right| + 360^\circ$



(c₁) ${}^4\psi_i^{*-} \leq \psi_i \leq 0^\circ$

(c₂) ${}^4\psi_i^{*+} \leq \left| \tau_{H_k^j, H_N^{i+1}} \right| \leq 180^\circ$

(c₃) $\psi_i = \psi_{i,j,k} - \left| \tau_{H_k^j, H_N^{i+1}} \right| + 360^\circ$

(d₁) $-180^\circ \leq \psi_i \leq {}^4\psi_i^{*-}$

(d₂) $-{}^4\psi_i^{*-} \leq \left| \tau_{H_k^j, H_N^{i+1}} \right| \leq 180^\circ$

(d₃) $\psi_i = \psi_{i,j,k} + \left| \tau_{H_k^j, H_N^{i+1}} \right|$

Figura 55 – Dinâmica do ângulo ψ_i com relação aos ângulos τ_{N^i, H_k^j} , $\tau_{N^{i+1}, H_N^{i+1}}$ e $\left| \tau_{H_k^j, H_N^{i+1}} \right|$ para o Caso 4.

e $-180^\circ < \tau_{N^i, H_k^j} < -180^\circ + \tau_{N^{i+1}, H_N^{i+1}}$; e o valor absoluto do ângulo $\tau_{H_k^j, H_N^{i+1}}$ é conhecido. Então, o ângulo $\psi_i \in (-180^\circ, 180^\circ]$ pode ser escrito de duas formas, dadas por ψ_i^+ e ψ_i^- ,

em função dos outros três ângulos dados, da seguinte maneira:

$$\psi_i^+ = \begin{cases} \psi_{i,j,k} + \left| \tau_{H_k^j, H_N^{i+1}} \right| & \text{se } \left| \psi_{i,j,k} + \left| \tau_{H_k^j, H_N^{i+1}} \right| \right| < 180^\circ, \\ \psi_{i,j,k} + \left| \tau_{H_k^j, H_N^{i+1}} \right| + 360^\circ & \text{se } \psi_{i,j,k} + \left| \tau_{H_k^j, H_N^{i+1}} \right| \leq -180^\circ. \end{cases}$$

$$\psi_i^- = \psi_{i,j,k} - \left| \tau_{H_k^j, H_N^{i+1}} \right| + 360^\circ.$$

Demonstração. A demonstração deste lema pode ser feita de maneira análoga a do Lema 4.1. No entanto, neste caso, consideramos as desigualdades de (A.4) para observar o valor de $\psi_{i,j,k} \pm \left| \tau_{H_k^j, H_N^{i+1}} \right|$. \square

Agora, para sistematizar os quatro casos considerados nesta seção, a Proposição A.1 apresenta uma relação geral que descreve o ângulo ψ_i em função dos ângulos de torção τ_{N^i, H_k^j} , $\tau_{N^{i+1}, H_N^{i+1}}$ e $\tau_{H_k^j, H_N^{i+1}}$.

Proposição A.1. Sejam R_j , R_i e R_{i+1} resíduos de aminoácido que compõe uma proteína, com $j \leq i$, $\psi_i \equiv \{N^i, C_\alpha^i, C^i, N^{i+1}\}$, $\tau_{N^i, H_k^j} \equiv \{N^i, C_\alpha^i, C^i, H_k^j\}$, $\tau_{N^{i+1}, H_N^{i+1}} \equiv \{N^{i+1}, C_\alpha^i, C^i, H_N^{i+1}\}$ e $\tau_{H_k^j, H_N^{i+1}} \equiv \{H_k^j, C_\alpha^i, C^i, H_N^{i+1}\}$ ângulos de torção definidos em átomos destes resíduos. Considere que os valores de $\tau_{N^{i+1}, H_N^{i+1}}$ e τ_{N^i, H_k^j} são conhecidos de modo que $\tau_{N^{i+1}, H_N^{i+1}} \geq 0^\circ$ e o valor absoluto do ângulo $\tau_{H_k^j, H_N^{i+1}}$ é conhecido. Então, existem duas possibilidades para se escrever o ângulo $\psi_i \in (-180^\circ, 180^\circ]$, dadas por ψ_i^+ e ψ_i^- , em função dos outros três ângulos de torção, que são descritas da seguinte maneira:

$$\psi_i^+ = c_{\psi_i}(\xi_{i,j,k}^+) + \xi_{i,j,k}^+ \quad \text{e} \quad \psi_i^- = c_{\psi_i}(\xi_{i,j,k}^-) + \xi_{i,j,k}^-, \quad (\text{A.5})$$

onde,

$$\xi_{i,j,k}^\pm \equiv \psi_{i,j,k} \pm \left| \tau_{H_k^j, H_N^{i+1}} \right|, \quad \psi_{i,j,k} \equiv \tau_{N^i, H_k^j} - \tau_{N^{i+1}, H_N^{i+1}},$$

$$c_{\psi_i}(\xi) \equiv \begin{cases} 0 & \text{se } |\xi| < 180^\circ, \\ 360^\circ \cdot (-1)^{m+1} & \text{se } |\xi| > 180^\circ, \\ 180^\circ - \xi & \text{se } |\xi| = 180^\circ, \end{cases}$$

e

$$m \equiv \begin{cases} 1 & \text{se } -180^\circ + \tau_{N^{i+1}, H_N^{i+1}} \leq \tau_{N^i, H_k^j} \leq \tau_{N^{i+1}, H_N^{i+1}}, \\ 2 & \text{se } \tau_{N^{i+1}, H_N^{i+1}} < \tau_{N^i, H_k^j} \leq 180^\circ, \\ 3 & \text{se } -180 < \tau_{N^i, H_k^j} < -180^\circ + \tau_{N^{i+1}, H_N^{i+1}}, \end{cases}$$

sendo $\text{mod}(m, 2)$ o resto da divisão inteira de m por 2.

Demonstração. Este resultado foi construído diretamente dos quatro casos descritos anteriormente nesta seção. No entanto, neste caso, condensamos a notação utilizada. Desta forma, sua demonstração segue direto dos Lemas A.1, A.2, A.3 e A.4. \square

A.2 Situação (ii): $\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}} < 0^\circ$

Caso 1: $0^\circ \leq \tau_{N^i, H_k^j} \leq 180^\circ + \tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}}$

Para este caso, a Figura 56 evidencia a disposição dos átomos H_k^j , N^i , C^i , N^{i+1} e $H_{\mathbb{N}}^{i+1}$ para as seguintes situações: $\psi_i = 0^\circ$, $|\psi_i| = 180^\circ$, $\psi_i = {}^1\psi_i^{*+}$ e $\psi_i = {}^1\psi_i^{*-}$. Portanto, observa-se que os valores ${}^1\psi_i^{*\pm}$ podem ser descritos em função da quantidade $\psi_{i,j,k}$, que é descrita por (4.29).

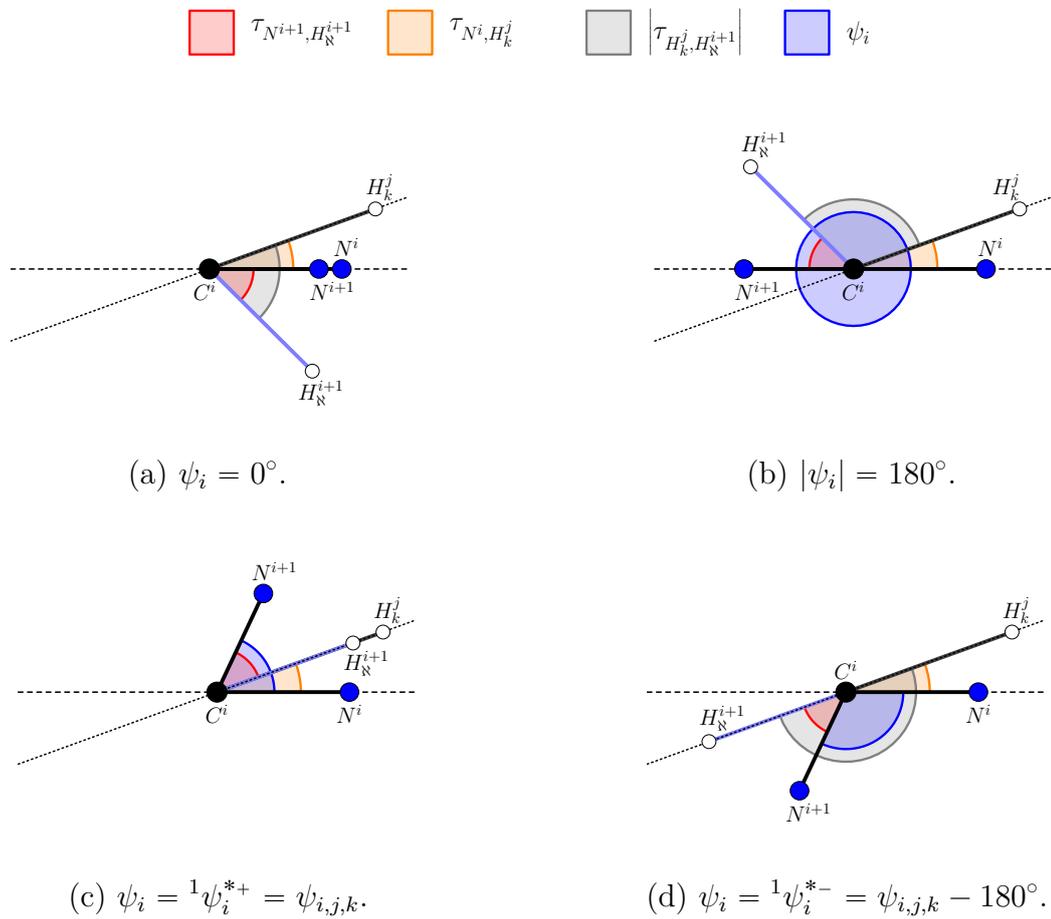


Figura 56 – Disposição no plano da projeção ortogonal dos átomos H_k^j , N^i , C^i , N^{i+1} e $H_{\mathbb{N}}^{i+1}$, com $\psi_i = 0^\circ$, $|\psi_i| = 180^\circ$, $\psi_i = {}^1\psi_i^{*+}$ e $\psi_i = {}^1\psi_i^{*-}$, para o Caso 1.

Sobre as quantidades ${}^1\psi_i^{*-} = \psi_{i,j,k} - 180^\circ$ e ${}^1\psi_i^{*+} = {}^1\psi_i^{*-} + 180^\circ$, observamos que:

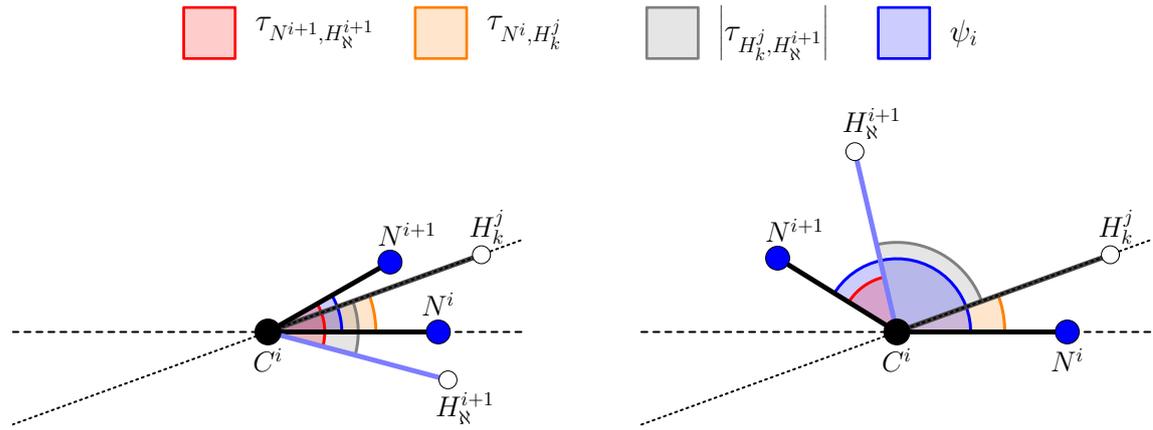
$$\begin{aligned}
 0^\circ &\leq \tau_{N^i, H_k^j} \leq 180^\circ + \tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}} && \Leftrightarrow \\
 -\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}} &\leq \psi_{i,j,k} \leq 180^\circ && \Leftrightarrow \\
 -180^\circ - \tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}} &\leq {}^1\psi_i^{*-} \leq 0^\circ && \Leftrightarrow \\
 -\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}} &\leq {}^1\psi_i^{*+} \leq 180^\circ. &&
 \end{aligned} \tag{A.6}$$

Agora, utilizando a intuição do desenho geométrico, na [Figura 57](#) sistematizamos as relações que descrevem a dinâmica entre os ângulos de torção ψ_i , τ_{N^i, H_k^j} , $\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}}$ e $\tau_{H_k^j, H_{\mathbb{N}}^{i+1}}$. Os casos apresentados se distinguem com relação a orientação e ao intervalo considerado do ângulo $\tau_{H_k^j, H_{\mathbb{N}}^{i+1}}$. Estes intervalos são dados com relação as quantidades descritas na [Figura 56](#). Desta maneira, considerando as relações apresentadas na [Figura 57](#), podemos escrever o seguinte lema:

Lema A.5. Sejam ψ_i , τ_{N^i, H_k^j} , $\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}}$, $\tau_{H_k^j, H_{\mathbb{N}}^{i+1}}$ os ângulos apresentados na [Figura 57](#). Considere que os valores de $\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}}$ e τ_{N^i, H_k^j} são conhecidos de modo que $\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}} < 0^\circ$ e $0^\circ \leq \tau_{N^i, H_k^j} \leq 180^\circ + \tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}}$; e o valor absoluto do ângulo $\tau_{H_k^j, H_{\mathbb{N}}^{i+1}}$ é conhecido. Então, o ângulo $\psi_i \in (-180^\circ, 180^\circ]$ pode ser escrito de duas formas, dadas por ψ_i^+ e ψ_i^- , em função dos outros três ângulos dados, da seguinte maneira:

$$\begin{aligned}
 \psi_i^+ &= \begin{cases} \psi_{i,j,k} + \left| \tau_{H_k^j, H_{\mathbb{N}}^{i+1}} \right| & \text{se } \left| \psi_{i,j,k} + \left| \tau_{H_k^j, H_{\mathbb{N}}^{i+1}} \right| \right| \leq 180^\circ, \\ \psi_{i,j,k} + \left| \tau_{H_k^j, H_{\mathbb{N}}^{i+1}} \right| - 360^\circ & \text{se } \left| \psi_{i,j,k} + \left| \tau_{H_k^j, H_{\mathbb{N}}^{i+1}} \right| \right| > 180^\circ. \end{cases} \\
 \psi_i^- &= \psi_{i,j,k} - \left| \tau_{H_k^j, H_{\mathbb{N}}^{i+1}} \right|.
 \end{aligned}$$

Demonstração. A demonstração deste lema pode ser feita de maneira análoga a do [Lema 4.1](#). No entanto, neste caso consideramos as desigualdades de [\(A.6\)](#) para observar o valor de $\psi_{i,j,k} \pm \left| \tau_{H_k^j, H_{\mathbb{N}}^{i+1}} \right|$. \square



$$(a_1) \quad 0^\circ \leq \psi_i \leq {}^1\psi_i^{*+}$$

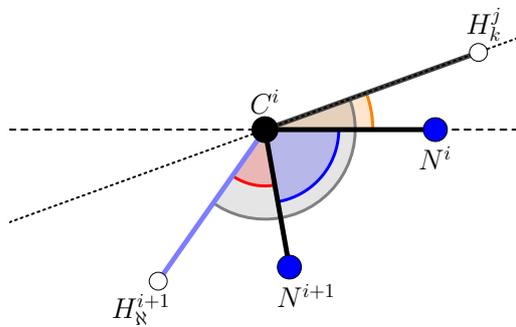
$$(a_2) \quad 0^\circ \leq \left| \tau_{H_k^j, H_N^{i+1}} \right| \leq {}^1\psi_i^{*+}$$

$$(a_3) \quad \psi_i = \psi_{i,j,k} - \left| \tau_{H_k^j, H_N^{i+1}} \right|$$

$$(b_1) \quad {}^1\psi_i^{*+} \leq \psi_i \leq 180^\circ$$

$$(b_2) \quad 0^\circ \leq \left| \tau_{H_k^j, H_N^{i+1}} \right| \leq -{}^1\psi_i^{*-}$$

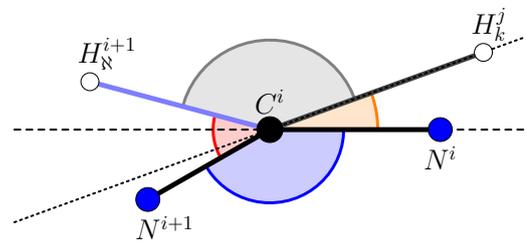
$$(b_3) \quad \psi_i = \psi_{i,j,k} + \left| \tau_{H_k^j, H_N^{i+1}} \right|$$



$$(c_1) \quad {}^1\psi_i^{*-} \leq \psi_i \leq 0^\circ$$

$$(c_2) \quad {}^1\psi_i^{*+} \leq \left| \tau_{H_k^j, H_N^{i+1}} \right| \leq 180^\circ$$

$$(c_3) \quad \psi_i = \psi_{i,j,k} - \left| \tau_{H_k^j, H_N^{i+1}} \right|$$



$$(d_1) \quad -180^\circ \leq \psi_i \leq {}^1\psi_i^{*-}$$

$$(d_2) \quad -{}^1\psi_i^{*-} \leq \left| \tau_{H_k^j, H_N^{i+1}} \right| \leq 180^\circ$$

$$(d_3) \quad \psi_i = \psi_{i,j,k} + \left| \tau_{H_k^j, H_N^{i+1}} \right| - 360^\circ$$

Figura 57 – Dinâmica do ângulo ψ_i com relação aos ângulos τ_{N^i, H_k^j} , $\tau_{N^{i+1}, H_N^{i+1}}$ e $\left| \tau_{H_k^j, H_N^{i+1}} \right|$ para o Caso 1.

Caso 2: $180^\circ + \tau_{N^{i+1}, H_N^{i+1}} < \tau_{N^i, H_k^j} \leq 180^\circ$

Para este caso, a Figura 58 evidencia a disposição dos átomos H_k^j , N^i , C^i , N^{i+1} e H_N^{i+1} para as seguintes situações: $\psi_i = 0^\circ$, $|\psi_i| = 180^\circ$, $\psi_i = {}^2\psi_i^{*+}$ e $\psi_i = {}^2\psi_i^{*-}$.

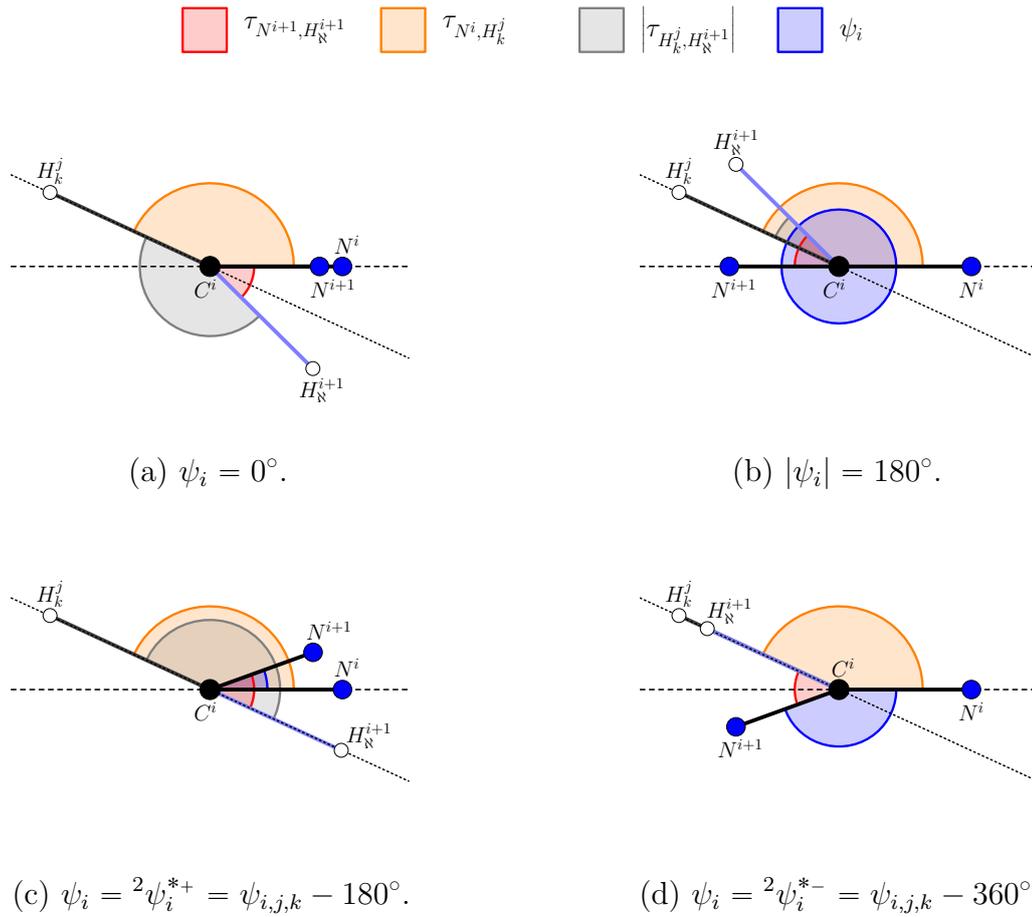


Figura 58 – Disposição no plano da projeção ortogonal dos átomos H_k^j , N^i , C^i , N^{i+1} e H_N^{i+1} , com $\psi_i = 0^\circ$, $|\psi_i| = 180^\circ$, $\psi_i = {}^2\psi_i^{*+}$ e $\psi_i = {}^2\psi_i^{*-}$, para o Caso 2.

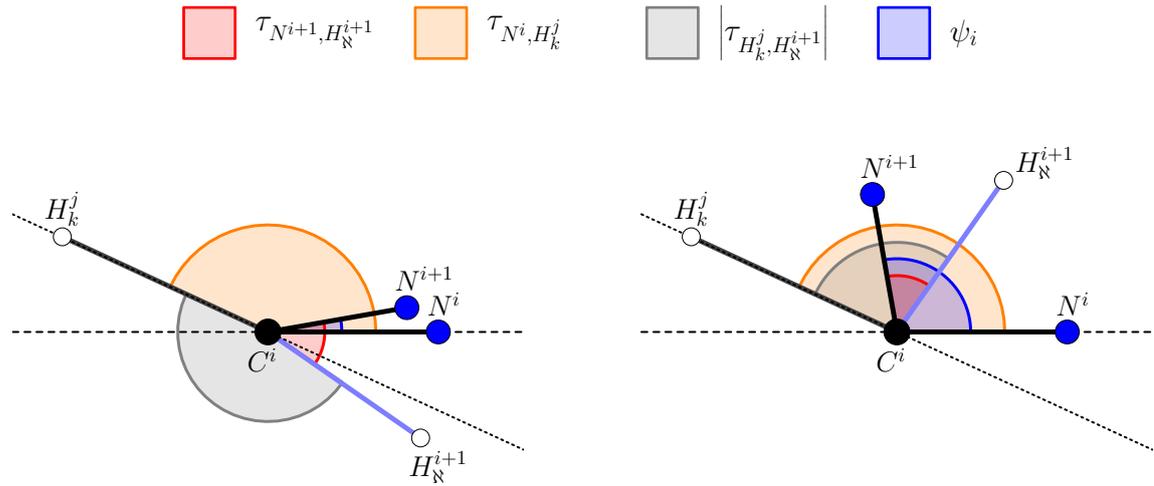
Sobre as quantidades ${}^2\psi_i^{*-} = \psi_{i,j,k} - 360^\circ$ e ${}^2\psi_i^{*+} = {}^2\psi_i^{*-} + 180^\circ$, observamos que:

$$\begin{aligned}
 180^\circ + \tau_{N^{i+1}, H_N^{i+1}} &< \tau_{N^i, H_k^j} \leq 180^\circ && \Leftrightarrow \\
 180^\circ &< \psi_{i,j,k} \leq 180^\circ - \tau_{N^{i+1}, H_N^{i+1}} && \Leftrightarrow \\
 -180^\circ &< {}^2\psi_i^{*-} \leq -180^\circ - \tau_{N^{i+1}, H_N^{i+1}} && \Leftrightarrow \\
 0^\circ &< {}^2\psi_i^{*+} \leq -\tau_{N^{i+1}, H_N^{i+1}}. &&
 \end{aligned} \tag{A.7}$$

Como no caso anterior, pela intuição do desenho geométrico, na [Figura 59](#) sistematizamos as relações que descrevem a dinâmica entre os ângulos de torção ψ_i , τ_{N^i, H_k^j} , $\tau_{N^{i+1}, H_N^{i+1}}$ e $\tau_{H_k^j, H_N^{i+1}}$.

Assim, considerando as relações apresentadas na [Figura 59](#), podemos escrever o seguinte lema:

Lema A.6. Sejam ψ_i , τ_{N^i, H_k^j} , $\tau_{N^{i+1}, H_N^{i+1}}$, $\tau_{H_k^j, H_N^{i+1}}$ os ângulos apresentados na [Figura 59](#). Considere que os valores de $\tau_{N^{i+1}, H_N^{i+1}}$ e τ_{N^i, H_k^j} são conhecidos de modo que $\tau_{N^{i+1}, H_N^{i+1}} < 0^\circ$ e $180^\circ + \tau_{N^{i+1}, H_N^{i+1}} < \tau_{N^i, H_k^j} \leq 180^\circ$ e o valor absoluto do ângulo $\tau_{H_k^j, H_N^{i+1}}$ é conhecido.



(a₁) $0^\circ \leq \psi_i \leq \psi_i^{*+}$

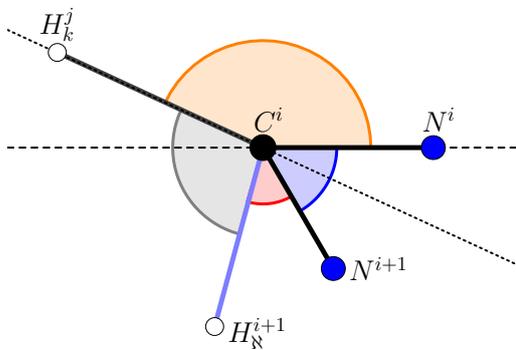
(a₂) $0^\circ \leq |\tau_{H_k^j, H_N^{i+1}}| \leq \psi_i^{*+}$

(a₃) $\psi_i = \psi_{i,j,k} + |\tau_{H_k^j, H_N^{i+1}}| - 360^\circ$

(b₁) $\psi_i^{*+} \leq \psi_i \leq 180^\circ$

(b₂) $0^\circ \leq |\tau_{H_k^j, H_N^{i+1}}| \leq -\psi_i^{*-}$

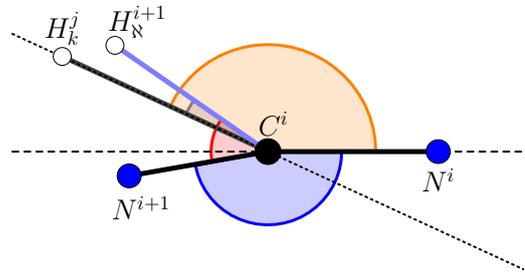
(b₃) $\psi_i = \psi_{i,j,k} - |\tau_{H_k^j, H_N^{i+1}}|$



(c₁) $\psi_i^{*-} \leq \psi_i \leq 0^\circ$

(c₂) $\psi_i^{*+} \leq |\tau_{H_k^j, H_N^{i+1}}| \leq 180^\circ$

(c₃) $\psi_i = \psi_{i,j,k} + |\tau_{H_k^j, H_N^{i+1}}| - 360^\circ$



(d₁) $-180^\circ \leq \psi_i \leq \psi_i^{*-}$

(d₂) $-\psi_i^{*-} \leq |\tau_{H_k^j, H_N^{i+1}}| \leq 180^\circ$

(d₃) $\psi_i = \psi_{i,j,k} - |\tau_{H_k^j, H_N^{i+1}}| - 360^\circ$

Figura 59 – Dinâmica do ângulo ψ_i com relação aos ângulos τ_{N^i, H_k^j} , $\tau_{N^{i+1}, H_N^{i+1}}$ e $|\tau_{H_k^j, H_N^{i+1}}|$ para o Caso 2.

Então, o ângulo $\psi_i \in (-180^\circ, 180^\circ]$ pode ser escrito de duas formas, dadas por ψ_i^+ e ψ_i^- ,

em função dos outros três ângulos dados, da seguinte maneira:

$$\begin{aligned} \psi_i^+ &= \psi_{i,j,k} + \left| \tau_{H_k^j, H_{\mathbb{N}}^{i+1}} \right| - 360^\circ. \\ \psi_i^- &= \begin{cases} \psi_{i,j,k} - \left| \tau_{H_k^j, H_{\mathbb{N}}^{i+1}} \right| & \text{se } \left| \psi_{i,j,k} - \left| \tau_{H_k^j, H_{\mathbb{N}}^{i+1}} \right| \right| \leq 180^\circ, \\ \psi_{i,j,k} - \left| \tau_{H_k^j, H_{\mathbb{N}}^{i+1}} \right| - 360^\circ & \text{se } \left| \psi_{i,j,k} - \left| \tau_{H_k^j, H_{\mathbb{N}}^{i+1}} \right| \right| > 180^\circ. \end{cases} \end{aligned}$$

Demonstração. A demonstração deste lema pode ser feita de maneira análoga a do Lema 4.1. No entanto, neste caso, consideramos as desigualdades de (A.7) para observar o valor de $\psi_{i,j,k} \pm \left| \tau_{H_k^j, H_{\mathbb{N}}^{i+1}} \right|$. \square

Caso 3: $\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}} \leq \tau_{N^i, H_k^j} < 0^\circ$

Para este caso, a Figura 60 evidencia a disposição dos átomos H_k^j , N^i , C^i , N^{i+1} e $H_{\mathbb{N}}^{i+1}$ para as seguintes situações: $\psi_i = 0^\circ$, $|\psi_i| = 180^\circ$, $\psi_i = {}^3\psi_i^{*+}$ e $\psi_i = {}^3\psi_i^{*-}$.

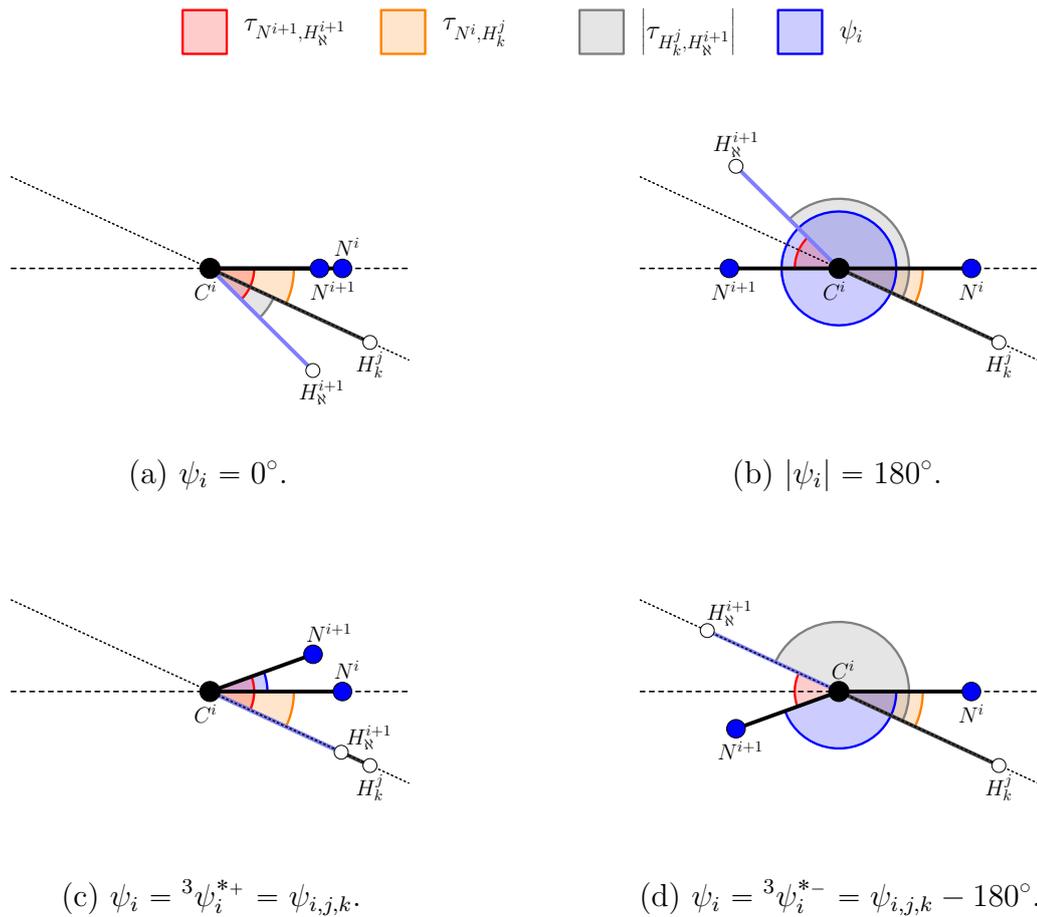


Figura 60 – Disposição no plano da projeção ortogonal dos átomos H_k^j , N^i , C^i , N^{i+1} e $H_{\mathbb{N}}^{i+1}$, com $\psi_i = 0^\circ$, $|\psi_i| = 180^\circ$, $\psi_i = {}^3\psi_i^{*+}$ e $\psi_i = {}^3\psi_i^{*-}$, para o Caso 3.

Sobre as quantidades ${}^3\psi_i^{*-} = \psi_{i,j,k} - 180^\circ$ e ${}^3\psi_i^{*+} = {}^3\psi_i^{*-} + 180^\circ$, observamos que:

$$\begin{aligned}
 \tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}} &\leq \tau_{N^i, H_k^j} < 0^\circ && \Leftrightarrow \\
 0^\circ &\leq \psi_{i,j,k} < -\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}} && \Leftrightarrow \\
 -180^\circ &\leq {}^3\psi_i^{*-} < -180^\circ - \tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}} && \Leftrightarrow \\
 0^\circ &\leq {}^3\psi_i^{*+} < \tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}}. &&
 \end{aligned} \tag{A.8}$$

Como no caso anterior, pela intuição do desenho geométrico, na [Figura 61](#) sistematizamos as relações que descrevem a dinâmica entre os ângulos de torção ψ_i , τ_{N^i, H_k^j} , $\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}}$ e $\tau_{H_k^j, H_{\mathbb{N}}^{i+1}}$.

Assim, considerando as relações apresentadas na [Figura 61](#), podemos escrever o seguinte lema:

Lema A.7. Sejam ψ_i , τ_{N^i, H_k^j} , $\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}}$, $\tau_{H_k^j, H_{\mathbb{N}}^{i+1}}$ os ângulos apresentados na [Figura 61](#). Considere que os valores de $\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}}$ e τ_{N^i, H_k^j} são conhecidos de modo que $\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}} \leq \tau_{N^i, H_k^j} < 0^\circ$ e o valor absoluto do ângulo $\tau_{H_k^j, H_{\mathbb{N}}^{i+1}}$ é conhecido. Então, o ângulo $\psi_i \in (-180^\circ, 180^\circ]$ pode ser escrito de duas formas, dadas por ψ_i^+ e ψ_i^- , em função dos outros três ângulos dados, da seguinte maneira:

$$\begin{aligned}
 \psi_i^+ &= \begin{cases} \psi_{i,j,k} + \left| \tau_{H_k^j, H_{\mathbb{N}}^{i+1}} \right| & \text{se } \left| \psi_{i,j,k} + \left| \tau_{H_k^j, H_{\mathbb{N}}^{i+1}} \right| \right| \leq 180^\circ, \\ \psi_{i,j,k} + \left| \tau_{H_k^j, H_{\mathbb{N}}^{i+1}} \right| - 360^\circ & \text{se } \left| \psi_{i,j,k} + \left| \tau_{H_k^j, H_{\mathbb{N}}^{i+1}} \right| \right| > 180^\circ. \end{cases} \\
 \psi_i^- &= \psi_{i,j,k} - \left| \tau_{H_k^j, H_{\mathbb{N}}^{i+1}} \right|.
 \end{aligned}$$

Demonstração. A demonstração deste lema pode ser feita de maneira análoga a do [Lema 4.1](#). No entanto, neste caso consideramos as desigualdades de [\(A.8\)](#) para observar o valor de $\psi_{i,j,k} \pm \left| \tau_{H_k^j, H_{\mathbb{N}}^{i+1}} \right|$. \square

Pelo resultado observado nos [Lemas A.5](#) e [A.7](#), notamos que a relação entre os ângulos de torção analisados é a mesma se $\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}} \leq \tau_{N^i, H_k^j} \leq 180^\circ + \tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}}$. A justificativa desta observação reside no fato de que os valores ${}^1\psi_i^{*\pm}$ e ${}^3\psi_i^{*\pm}$ são escritos em função da quantidade $\psi_{i,j,k}$ da mesma maneira e que a orientação do ângulo $\tau_{H_k^j, H_{\mathbb{N}}^{i+1}}$ é a mesma se $\psi_i \in [{}^m\psi_i^{*-}, {}^m\psi_i^{*+}]$ ou $\psi_i \in [-180^\circ, {}^m\psi_i^{*-}] \cup [{}^m\psi_i^{*+}, 180^\circ]$, com $m = 1, 3$.

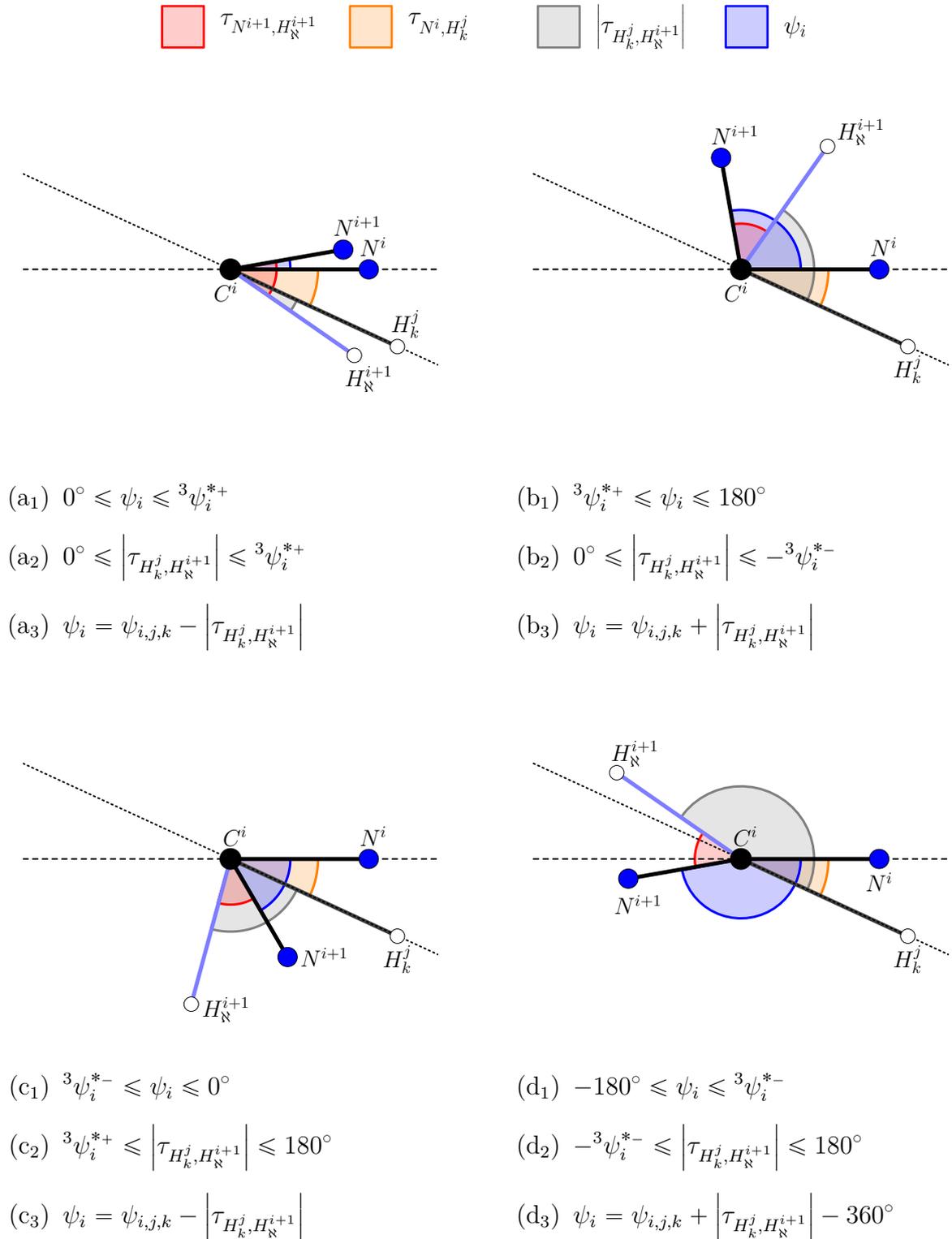


Figura 61 – Dinâmica do ângulo ψ_i com relação aos ângulos τ_{N^i, H_k^j} , $\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}}$ e $|\tau_{H_k^j, H_{\mathbb{N}}^{i+1}}|$ para o Caso 3.

Caso 4: $-180^\circ < \tau_{N^i, H_k^j} < \tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}}$

Para este caso, a [Figura 62](#) evidencia a disposição dos átomos H_k^j , N^i , C^i , N^{i+1} e $H_{\mathbb{N}}^{i+1}$ para as seguintes situações: $\psi_i = 0^\circ$, $|\psi_i| = 180^\circ$, $\psi_i = {}^4\psi_i^{*+}$ e $\psi_i = {}^4\psi_i^{*-}$.

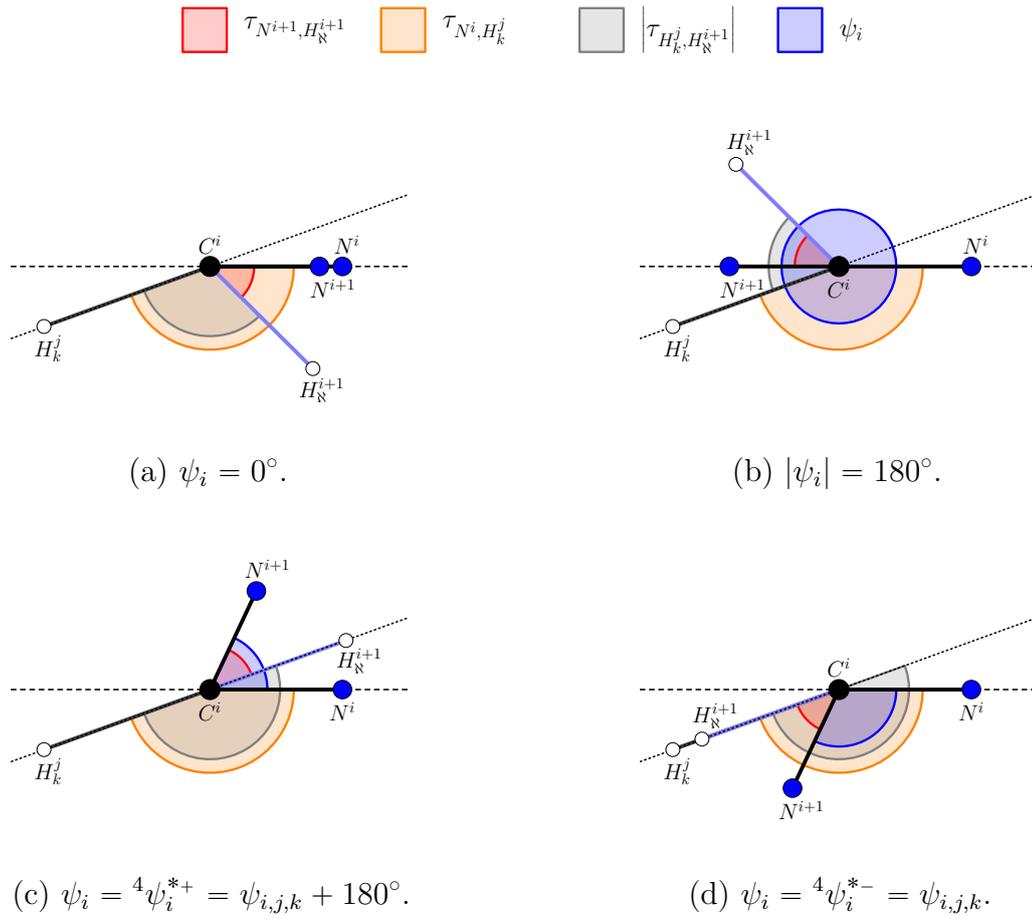


Figura 62 – Disposição no plano da projeção ortogonal dos átomos H_k^j , N^i , C^i , N^{i+1} e H_N^{i+1} , com $\psi_i = 0^\circ$, $|\psi_i| = 180^\circ$, $\psi_i = {}^4\psi_i^{*+}$ e $\psi_i = {}^4\psi_i^{*-}$, para o Caso 4.

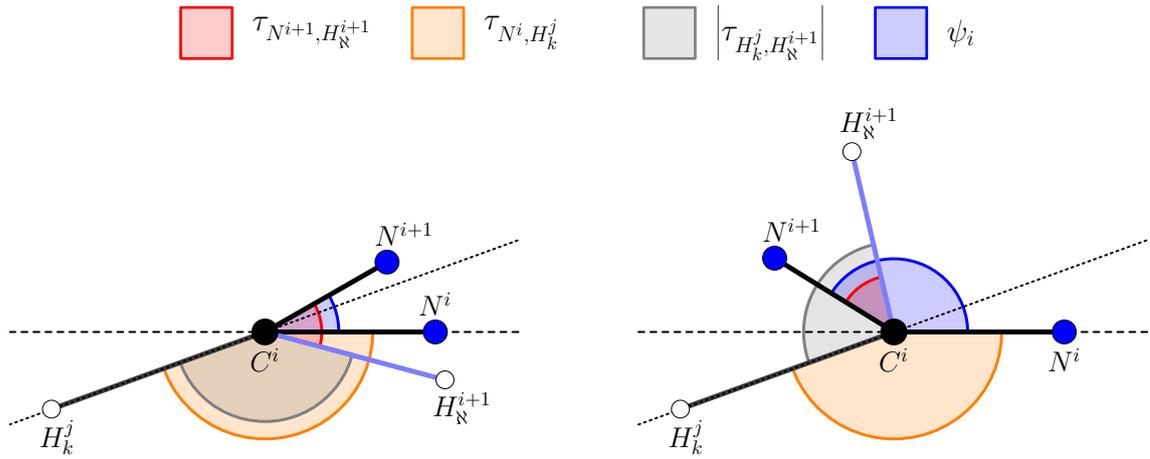
Sobre as quantidades ${}^4\psi_i^{*-} = \psi_{i,j,k}$ e ${}^4\psi_i^{*+} = {}^4\psi_i^{*-} + 180^\circ$, observamos que:

$$\begin{aligned}
 -180^\circ &< \tau_{N^i, H_k^j} < \tau_{N^{i+1}, H_N^{i+1}} && \Leftrightarrow \\
 -180^\circ - \tau_{N^{i+1}, H_N^{i+1}} &< \psi_{i,j,k} \leq 0^\circ && \Leftrightarrow \\
 -180^\circ - \tau_{N^{i+1}, H_N^{i+1}} &< {}^4\psi_i^{*-} \leq 0^\circ && \Leftrightarrow \\
 \tau_{N^{i+1}, H_N^{i+1}} &< {}^4\psi_i^{*+} \leq 180^\circ. &&
 \end{aligned} \tag{A.9}$$

Como no caso anterior, pela intuição do desenho geométrico, na Figura 63 sistematizamos as relações que descrevem a dinâmica entre os ângulos de torção ψ_i , τ_{N^i, H_k^j} , $\tau_{N^{i+1}, H_N^{i+1}}$ e $\tau_{H_k^j, H_N^{i+1}}$.

Assim, considerando as relações apresentadas na Figura 63, podemos escrever o seguinte lema:

Lema A.8. Sejam ψ_i , τ_{N^i, H_k^j} , $\tau_{N^{i+1}, H_N^{i+1}}$, $\tau_{H_k^j, H_N^{i+1}}$ os ângulos apresentados na Figura 63. Considere que os valores de $\tau_{N^{i+1}, H_N^{i+1}}$ e τ_{N^i, H_k^j} são conhecidos de modo que $-180^\circ <$



(a₁) $0^\circ \leq \psi_i \leq {}^4\psi_i^{*+}$

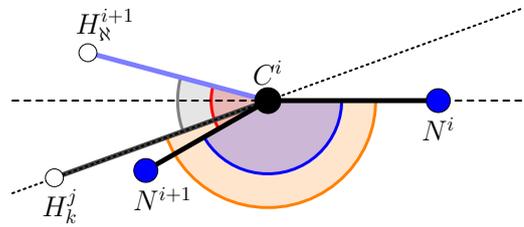
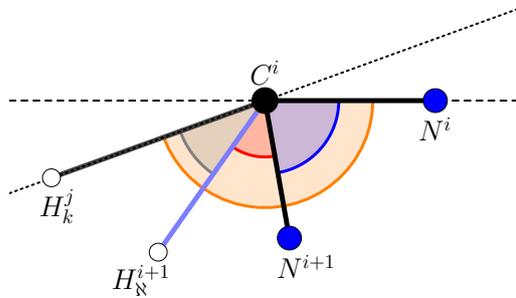
(b₁) ${}^4\psi_i^{*+} \leq \psi_i \leq 180^\circ$

(a₂) ${}^4\psi_i^{*+} \leq \left| \tau_{H_k^j, H_N^{i+1}} \right| \leq 180^\circ$

(b₂) $-{}^4\psi_i^{*-} \leq \left| \tau_{H_k^j, H_N^{i+1}} \right| \leq 180^\circ$

(a₃) $\psi_i = \psi_{i,j,k} + \left| \tau_{H_k^j, H_N^{i+1}} \right|$

(b₃) $\psi_i = \psi_{i,j,k} - \left| \tau_{H_k^j, H_N^{i+1}} \right| + 360^\circ$



(c₁) ${}^4\psi_i^{*-} \leq \psi_i \leq 0^\circ$

(d₁) $-180^\circ \leq \psi_i \leq {}^4\psi_i^{*-}$

(c₂) $0^\circ \leq \left| \tau_{H_k^j, H_N^{i+1}} \right| \leq {}^4\psi_i^{*+}$

(d₂) $0^\circ \leq \left| \tau_{H_k^j, H_N^{i+1}} \right| \leq -{}^4\psi_i^{*-}$

(c₃) $\psi_i = \psi_{i,j,k} + \left| \tau_{H_k^j, H_N^{i+1}} \right|$

(d₃) $\psi_i = \psi_{i,j,k} - \left| \tau_{H_k^j, H_N^{i+1}} \right|$

Figura 63 – Dinâmica do ângulo ψ_i com relação aos ângulos τ_{N^i, H_k^j} , $\tau_{N^{i+1}, H_N^{i+1}}$ e $\left| \tau_{H_k^j, H_N^{i+1}} \right|$ para o Caso 4.

$\tau_{N^i, H_k^j} < \tau_{N^{i+1}, H_N^{i+1}} < 0^\circ$ e o valor absoluto do ângulo $\tau_{H_k^j, H_N^{i+1}}$ é conhecido. Então, o ângulo $\psi_i \in (-180^\circ, 180^\circ]$ pode ser escrito de duas formas, dadas por ψ_i^+ e ψ_i^- , em função dos

outros três ângulos dados, da seguinte maneira:

$$\begin{aligned}\psi_i^+ &= \psi_{i,j,k} + \left| \tau_{H_k^j, H_{\mathbb{N}}^{i+1}} \right| \\ \psi_i^- &= \begin{cases} \psi_{i,j,k} - \left| \tau_{H_k^j, H_{\mathbb{N}}^{i+1}} \right| & \text{se } \left| \psi_{i,j,k} - \left| \tau_{H_k^j, H_{\mathbb{N}}^{i+1}} \right| \right| < 180^\circ, \\ \psi_{i,j,k} - \left| \tau_{H_k^j, H_{\mathbb{N}}^{i+1}} \right| + 360^\circ & \text{se } \psi_{i,j,k} - \left| \tau_{H_k^j, H_{\mathbb{N}}^{i+1}} \right| \leq -180^\circ. \end{cases}\end{aligned}$$

Demonstração. A demonstração deste lema pode ser feita de maneira análoga a do Lema 4.1. No entanto, neste caso, consideramos as desigualdades de (A.9) para observar o valor de $\psi_{i,j,k} \pm \left| \tau_{H_k^j, H_{\mathbb{N}}^{i+1}} \right|$. \square

Agora, para sistematizar os quatro casos considerados nesta seção, a Proposição A.2 apresenta uma relação geral que descreve o ângulo ψ_i em função dos ângulos de torção τ_{N^i, H_k^j} , $\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}}$ e $\tau_{H_k^j, H_{\mathbb{N}}^{i+1}}$.

Proposição A.2. Sejam R_j , R_i e R_{i+1} resíduos de aminoácido que compõe uma proteína, com $j \leq i$, $\psi_i \equiv \{N^i, C_\alpha^i, C^i, N^{i+1}\}$, $\tau_{N^i, H_k^j} \equiv \{N^i, C_\alpha^i, C^i, H_k^j\}$, $\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}} \equiv \{N^{i+1}, C_\alpha^i, C^i, H_{\mathbb{N}}^{i+1}\}$ e $\tau_{H_k^j, H_{\mathbb{N}}^{i+1}} \equiv \{H_k^j, C_\alpha^i, C^i, H_{\mathbb{N}}^{i+1}\}$ ângulos de torção definidos em átomos destes resíduos. Considere que os valores de $\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}}$ e τ_{N^i, H_k^j} são conhecidos de modo que $\tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}} < 0^\circ$ e o valor absoluto do ângulo $\tau_{H_k^j, H_{\mathbb{N}}^{i+1}}$ é conhecido. Então, existem duas possibilidades para se escrever o ângulo $\psi_i \in (-180^\circ, 180^\circ]$, dadas por ψ_i^+ e ψ_i^- , em função dos outros três ângulos de torção, que são descritas da seguinte maneira:

$$\psi_i^+ = c_{\psi_i}(\xi_{i,j,k}^+) + \xi_{i,j,k}^+ \quad \text{e} \quad \psi_i^- = c_{\psi_i}(\xi_{i,j,k}^-) + \xi_{i,j,k}^-, \quad (\text{A.10})$$

onde,

$$\begin{aligned}\xi_{i,j,k}^\pm &\equiv \psi_{i,j,k} \pm \left| \tau_{H_k^j, H_{\mathbb{N}}^{i+1}} \right|, \quad \psi_{i,j,k} \equiv \tau_{N^i, H_k^j} - \tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}}, \\ c_{\psi_i}(\xi) &\equiv \begin{cases} 0 & \text{se } |\xi| < 180^\circ, \\ 360^\circ \cdot (-1)^{m+1+\text{mod}(m,3)} & \text{se } |\xi| > 180^\circ, \\ 180^\circ - \xi & \text{se } |\xi| = 180^\circ, \end{cases}\end{aligned}$$

e

$$m \equiv \begin{cases} 1 & \text{se } \tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}} \leq \tau_{N^i, H_k^j} \leq 180^\circ + \tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}}, \\ 2 & \text{se } 180^\circ + \tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}} < \tau_{N^i, H_k^j} \leq 180^\circ, \\ 3 & \text{se } -180 < \tau_{N^i, H_k^j} < \tau_{N^{i+1}, H_{\mathbb{N}}^{i+1}}, \end{cases}$$

sendo $\text{mod}(m, 2)$ o resto da divisão inteira de m por 2.

Demonstração. Este resultado foi construído diretamente dos quatro casos descritos anteriormente nesta seção. No entanto, neste caso, condensamos a notação utilizada. Desta forma, sua demonstração segue direto dos Lemas [A.5](#), [A.6](#), [A.7](#) e [A.8](#). \square