UNIVERSIDADE DE SÃO PAULO
FACULDADE DE FILOSOFIA, CIÊNCIAS E LETRAS DE RIBEIRÃO PRETO

CRISTIANO ROBERTO FABRI GRANZOTTI

**Caminhadas com Memória em Meios Regulares e
Desordenados: Aspectos Estáticos e Dinâmicos**

Ribeirão Preto
2015

CRISTIANO ROBERTO FABRI GRANZOTTI

Caminhadas com Memória em Meios Regulares e Desordenados: Aspectos Estáticos e Dinâmicos

Dissertação apresentada à Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da Universidade de São Paulo como parte das exigências para a obtenção do título de Mestre em Ciências.

Área de Concentração:

Física Aplicada a Medicina e Biologia.

Orientador:

Alexandre Souto Martinez.

Versão corrigida

Versão original disponível na FFCLRP-USP

Ribeirão Preto

2015

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

FICHA CATALOGRÁFICA

Granzotti, Cristiano Roberto Fabri

Caminhadas com Memória em Meios Regulares e Desordenados: Aspectos Estáticos e Dinâmicos / Cristiano Roberto Fabri Granzotti; orientador: Alexandre Souto Martinez. - - Ribeirão Preto, 2015.

116 p. : il.

Dissertação (Mestrado) - - Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, 2015.

Inclui Bibliografia.

1. Caminhada Autorrepulsiva. 2. Processo Poissônico.
3. Estatística de Vizinhança. 4. Estatística de Distâncias. 5. Lei de Escala.

Nome: GRANZOTTI, Cristiano Roberto Fabri

Título: Caminhadas com Memória em Meios Regulares e Desordenados: Aspectos
Estáticos e Dinâmicos

Dissertação apresentada à Faculdade de Filosofia,
Ciências e Letras de Ribeirão Preto da Universi-
dade de São Paulo como parte das exigências para
a obtenção do título de Mestre em Ciências.

Aprovado em: ____/____/____.

Banca Examinadora

Prof(a). Dr(a). : _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof(a). Dr(a). : _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof(a). Dr(a). : _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Ao meu irmão José Maycon e à minha companheira
Lulu Wu.

Agradecimentos

Ao meu orientador Prof. Dr. Alexandre Souto Martinez, pela amizade e solicitude ao guiar-me desde a Iniciação Científica até o final do mestrado.

Ao meu segundo orientador, Prof. Dr. Marco Antônio Alves da Silva, seu convite permitiu minha participação no desenvolvimento do estudo sobre a caminhada aleatória autorrepulsiva presente nessa dissertação.

Aos colegas do Laboratório de Modelagem de Sistemas Complexos: Brenno Caetano Troca Cabella, Enock de Almeida Andrade Neto, Fabiano Lemes Ribeiro, Fernanda Miranda de Oliveira, Fernando Meloni, Gilberto Medeiros Nakamura, Juan Herbert Chuctaya Humari, Juliana Militão da Silva Berbert, Lindomar Soares dos Santos, Marcelo Alves Pereira, Natália Destefano, Olavo Henrique Menin, Rafael Fratucci, Rayner Montes Condori e Tiago José Arruda pela ajuda e apoio que recebi.

Aos demais colegas da pós graduação, especialmente à Hugo José Nogueira Pedroza Dias Mello e Diego Ronaldo Thomaz Sampaio, que sempre me acompanharam nos cafés da tarde.

Aos Docentes e Funcionários envolvidos direta e indiretamente com o programa FAMB.

Aos meus pais, José e Sueli, e ao meu irmão José Maycon, pelo apoio fornecido.

À minha companheira Lulu Wu, pela ajuda, paciência e prestatividade durante esses dois últimos anos.

À CAPES pelo suporte financeiro.

Life is the sum of trifling motions.

Joseph Brodsky

Resumo

GRANZOTTI, C. R. F. **Caminhadas com Memória em Meios Regulares e Desordenados: Aspectos Estáticos e Dinâmicos**. 2015. 116 p. Dissertação (Mestrado - Programa de Pós-Graduação em Física Aplicada a Medicina e Biologia) - Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2015.

Propomos o estudo do meio desordenado onde a caminhada determinista parcialmente autorrepulsiva (CDPA) é desenvolvida e o estudo da caminhada aleatória autorrepulsiva (SAW) em rede regular. O meio desordenado na CDPA, gerado por um processo Poissônico espacial, é caracterizado pela estatística de vizinhança e de distâncias. A estatística de vizinhança mede a probabilidade de um ponto ser m -ésimo vizinho mais próximo de seu n -ésimo vizinho mais próximo. A estatística de distâncias mede a distribuição de distância de um ponto ao seu k -ésimo vizinho mais próximo. No problema da estatística de distâncias, calculamos a função densidade de probabilidade (pdf) e estudamos os casos limites de alta ordem de vizinhança e alta dimensionalidade. Um caso particular dessa pdf pode verificar se um conjunto de pontos foi gerado por um processo Poissônico. Na SAW em rede regular, um caminhante escolhe aleatoriamente um sítio adjacente para ser visitado no próximo passo, mas é proibido visitar um sítio duas ou mais vezes. Desenvolvemos uma nova abordagem para estudar grandezas conformacionais por meio do produto escalar entre o vetor posição e vetor deslocamento no j -ésimo passo: $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$. Mostramos que para $j = N$ o produto escalar é igual ao comprimento de persistência (projeção do vetor posição na direção do primeiro passo) e que converge para uma constante. Calculamos a distância quadrática média ponta-a-ponta, $\langle \vec{R}_N^2 \rangle_N \sim N^{2\nu_0}$, como o somatório de $1 \leq j \leq N$ do produto escalar. Os dados gerados pelo algoritmo de simulação Monte Carlo, codificado em linguagem C e paralelizado em MPI, fornecem o expoente ν_0 da regra de escala $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N \sim j^{2\nu_0-1}$, para $1 \leq j \leq \Theta(N)$, próximo ao valor esperado. A partir de $\Theta(N) \approx N/2$ para rede quadrada e $\Theta(N) \approx N/3$ para rede cúbica, a caminhada torna-se mais flexível devido ao maior número de graus de liberdade disponível nos últimos passos.

Palavras-chave: 1. Caminhada Autorrepulsiva. 2. Processo Poissônico. 3. Estatística de Vizinhança. 4. Estatística de Distâncias. 5. Lei de Escala.

Abstract

GRANZOTTI, C. R. F. **Memory Walks in Regular and Disordered Media: Static and Dynamic Features.** 2015. 116 p. Dissertation (M.Sc. - Postgraduate program in Physics Applied to Medicine and Biology) - Faculty of Philosophy, Sciences and Letters, University of São Paulo, Ribeirão Preto, 2015.

We propose the study of disordered media where the deterministic partially self-avoiding walk (DPSW) is developed and the study of self-avoiding random walk (SAW) in regular lattices. The disordered media in the DPSW, generated by a spatial Poissonian process, is characterized by neighborhood and distance statistics. Neighborhood statistics quantifies the probability of a point to be the m th nearest neighbor of its n th nearest neighbor. Distance statistics quantifies the distance distribution of a given point to its k th nearest neighbor. For the distance statistics problem, we obtain the probability density function (pdf) and study the high dimensionality and high neighborhood order limits. A particular case of this pdf can verify if a points set is generated by a Poissonian process. In a SAW in regular lattice, the walker randomly chooses an adjacent site to be visited in the next step, but is forbidden to visit a site two or more times. We developed a new approach to study conformational quantities of SAW by means of the scalar product between the position vector and the displacement vector in the j th step: $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$. We show that for $j = N$ the scalar product is equal to the persistence length (projection of position vector in the direction of the first step) and that converges to a constant. We compute the square end-to-end distance, $\langle \vec{R}_N^2 \rangle_N \sim N^{2\nu_0}$, as the summation $1 \leq j \leq N$ of scalar product. The data generated by Monte Carlo simulation algorithm, coded in C language and parallelized in MPI, provides the exponent ν_0 of the scaling law $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N \sim j^{2\nu_0-1}$, for $1 \leq j \leq \Theta(N)$, close to the expected value. Starting from $\Theta(N) \approx N/2$ for square lattice and $\Theta(N) \approx N/3$ for cubic lattice, the walk becomes more flexible due to the large number of degrees of freedom available in the last steps.

Key-words: 1. Memory Walks. 2. Poisson Process. 3. Neighbourhood Statistics. 4. Distance Statistics. 5. Scaling Law.

Lista de Figuras

- 2.1 Meio desordenado bidimensional gerado pelo problema do ponto aleatório com densidade de pontos (a) ρ_1 , (b) $2\rho_1$ e (c) $3\rho_1$. Diferentemente de uma rede regular, onde os pontos são igualmente espaçados, há aqui pequenas subáreas com aglomeração ou vazio de pontos e a distância de um ponto aos demais está distribuída em torno de uma distância média. 9
- 2.2 Probabilidade de Cox em um meio bidimensional. Há i pontos na intersecção dos dois círculos, $n - i - 1$ pontos no crescente do ponto I e $m - i - 1$ no crescente do ponto J 10
- 2.3 Possíveis configurações onde o sítio I é o quarto vizinho mais próximo de J , enquanto J é o terceiro vizinho mais próximo de I 11
- 2.4 Comparação entre o valor exato do volume do crescente, Eq. 2.3, e o cálculo aproximado no limite de alta dimensionalidade, dado pela Eq. 2.12. Para $d \geq 10$ a aproximação se torna muito acurada. 13

- 2.5 Caminhada determinista parcialmente autorrepulsiva em um meio unidimensional e caminhante com $\mu = 1$. O caminhante parte do sítio s_0 e sempre vai ao sítio mais próximo, percorrendo os sítios de s_0 à s_4 em 4 passos, que compõem o transiente. Como o sítio s_4 e s_5 são mutualmente mais próximos (casal), o caminhante vai de $s_4 \rightarrow s_5$ e de $s_5 \rightarrow s_4$ indefinidamente, pois a memória $\mu = 1$ não permite que ele/ela visite outro sítio. Neste caso, o atrator é composto por 2 passos. A cada passo, a probabilidade do caminhante visitar um sítio que pertence a um casal é dada pela probabilidade de Cox, Eq. 2.6, para o caso $m = n = 1$. De posse da estatística de vizinhança, uma das características estáticas do meio desordenado, é possível calcular a distribuição do número de passos no transiente para $\mu = 1$, que é uma característica relacionada com a dinâmica de movimentação. . . 15

- 3.1 Comparação entre o resultado analítico gerado pela Eq. 3.1 (linhas cheias) até o quinto vizinho em um meio bidimensional. Simulação realizada com $\rho = 50000$, com condições periódicas de contorno, as barras de erro são equivalentes ou menores que o tamanho do ponto. As linhas cheias correspondem aos resultados analíticos dado pela Eq. 3.1. Note que o aumento da ordem de vizinhança recupera a simetria da distribuição. Gráfico adaptado da Ref. [25] 22

- 3.2 Assimetria da distribuição de distâncias. A simulação foi realizada com $d = 2$, $\rho = 65365$ e condições periódicas de contorno. A aproximação $\gamma_1 = 6\beta k^{-1/2}$ descreve bem o decaimento da assimetria em função da ordem de vizinhança, contudo o efeito de borda, tamanho finito e principalmente de baixa dimensionalidade fazem com que a simulação se distancie do valor real e do aproximado para $k \geq 10$ . . . 24

- 3.3 (a) Aproximação Gaussiana para a estatística de distância para $k \gg 1$ em um meio bidimensional. Os parâmetros da simulação são $d = 2$ e $\rho = 65365$ e $k = 10$. O ajuste pobre nas caudas é devido ao fato que o teorema central do limite garante convergência próximo à média, sendo a convergência da cauda mais lenta. Gráfico adaptado da Ref. [25]. (b) Razão σ/μ para a mesma simulação. O ajuste dado pela aproximação $c\beta k^{-1/2}$ descreve exatamente o comportamento em função da ordem de vizinhança, a simulação apresenta boa concordância com o esperado, contudo o efeito de borda, tamanho finito e a baixa dimensionalidade utilizada ($d = 2$) faz a simulação desviar do valor esperado. 26
- 4.1 Representação da cadeia de rotação livre por meio de uma caminhada com o ângulo de ligação fixo θ em qualquer orientação determinada pelo ângulo de rotação φ 32
- 4.2 O comprimento de contorno R_{max} é definido como o comprimento da cadeia quando totalmente esticada. Essa grandeza é fundamental, pois seu valor caracteriza o tamanho dos polímeros em experimentos Físicos/Químicos. Para a cadeia de rotação livre, $R_{max} = N\ell \cos(\theta/2)$. 34
- 5.1 Caminhadas aleatórias distintas, o losango aberto indica o início da caminhada e o fechado seu final. (a) Caminhada aleatória. (b) Caminhada aleatória não reversa, onde o caminhante que foi do sítio $\omega(i-1) \rightarrow \omega(i)$ no passo i não pode, no passo $i+1$, fazer o caminho inverso $\omega(i) \rightarrow \omega(i-1)$. (c) Caminhada autorrepulsiva (SAW). (d) Caminhada autorrepulsiva armadilhada, o caminhante não tem mais para onde ir e a caminhada é finalizada. 44

- 5.2 Testes realizados com os dados provenientes do nosso algoritmo. (a) Dados de acordo com a Eq. 5.9 para $d = 2$ e $d = 3$. Para $d = 2$ a oscilação no final do gráfico é devido ao baixo número de caminhadas que atingem $N > 120$ passos. (b) Constante de atrito, o ajuste não linear por meio da Eq. 5.8 fornece $\lambda = 0.12899(4)$ e $\lambda = 0.065298(4)$ para $d = 2$ e $d = 3$ respectivamente. Os valores da constante de atrito, calculadas a partir do valor de μ fornecido pela Ref. [16], são dados $\lambda = 0.128531205(1)$ e $\lambda = 0.0652762(28)$ para $d = 2$ e $d = 3$ respectivamente. 51
- 5.3 Abordagem utilizada para estudar as leis de escala para caminhadas aleatórias repulsivas. (a) O estudo da correlação entre os passos é dado pelo produto escalar $\langle \vec{u}_i \cdot \vec{u}_j \rangle_N$ e esse produto é a origem microscópica da regra de escala. Estudá-lo numericamente é complicado, haja vista que sua regra de escala é $s^{2\nu_0-2}$, onde s é a separação em número de passos entre os sítios i e j . (b) O estudo da SAW por meio da distância ponta-a-ponta quadrática considera apenas os pontos finais e iniciais. (c) Nossa abordagem para estudar a SAW é intermediária aos dois extremos anteriores e permite boa precisão na determinação de ν_0 53
- 5.4 (a) O ponto de máximo para a função correlação ocorre para $j \approx N/2$, a curva representa apenas uma guia para os olhos. (b) O colapso da função correlação é melhor quando utilizamos como fator de normalização $\langle \xi_{1,N/2} \rangle_N$, ou seja, o valor de $\langle \xi_{1,j} \rangle_N$ para o meio da cadeia. Isso indica que no meio da cadeia há uma mudança de comportamento da função de correlação angular. Os dados utilizados nesses gráficos são provenientes de enumeração exata. 57
- 5.5 Comprimento de persistência para rede quadrada e cúbica. O *inset* é o resíduo proveniente do ajuste por mínimos quadrados. Para $d = 2$: $\alpha_0 = 2.5254(36)$, $\alpha_1 = -2.319(25)$ e $\alpha_3 = +0.814(27)$ e os expoentes $w_1 = 0.5$ e $w_2 = 1$. Os coeficientes são: $\alpha_0 = 1.422(1)$, $\alpha_1 = -0.39(6)$ e $\alpha_2 = -0.022(5)$, os expoentes são $w_1 = 0.8248$ e $w_2 = 0.34$ 59

- 5.6 Ajuste dos dados do comprimento de persistência com a função $\langle x_N \rangle_N = \alpha_0 + \alpha_1 N^{-0.34(5)}$. Os parâmetros são $\alpha_0 = 2.664(3)$ e $\alpha_1 = -1.714(9)$ 60
- 5.7 Dados e colapso de dados para o produto escalar $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$ para a rede quadrada e cúbica. Os gráficos (a) e (b) mostram que $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$ é aproximadamente igual até um ponto $\Theta(N)$ para caminhadas com N distinto. O colapso de dados é obtido por meio da curva $\langle \vec{R}_j \cdot \vec{u}_j \rangle / \langle \vec{R}_j \cdot \vec{u}_j \rangle_{max} \times j/N$. O subíndice *max* indica o valor máximo do produto escalar para um dado N . (c) Produto escalar colapsado em escala linear e (d) em escala logarítmica para $d = 2$. (e) Produto escalar colapsado em escala linear e (f) em escala logarítmica para $d = 3$. Nos gráficos em escala logarítmica a inclinação da parte linear da curva é proporcional a $2\nu_0 - 1$ 62
- 5.8 Diferença entre produto escalar intermediário para duas caminhadas com número de passos distintos. (a) meio bidimensional caminhada com $N_1 = 40$ e $N_2 = 60$ passos; (b) meio bidimensional caminhada com $N_1 = 60$ e $N_2 = 90$ passos; (c) meio tridimensional com $N_1 = 60$ e $N_2 = 90$ passos; (d) meio tridimensional com $N_1 = 90$ e $N_2 = 108$. Os gráficos indicam o ponto ótimo para análise do expoente principal do produto escalar intermediário. Para $d = 2$, devemos usar os dados até $\Theta(N) \sim N/2$ e para $d = 3$, até $\Theta(N) \sim N/3$ 64
- 5.9 Produto escalar intermediário, no meio tridimensional os dados são provenientes de caminhadas com $N = 24$ até $N = 108$ e $\Delta N = 3$, no meio bidimensional $N = 18$ até $N = 60$ e $\Delta N = 6$. (a) Meio tridimensional sem correções de escala na Eq. 5.26 $\alpha_0 = 0.67871(32)$ $\omega = 0.20098(17)$ e $\tau = 0.7832(16)$ (b) Meio tridimensional com correção de escala $\Delta_1 = 0.5$, $\alpha_0 = 0.7203(23)$, $\alpha_1 = -0.0687(38)$, $\omega = 0.18778(71)$ e $\tau = 0.6387(81)$. (c) Meio bidimensional $\alpha_0 = 0.6618(17)$, $\omega = 0.50000(84)$ e $\tau = 0.4151(86)$ 66
- 5.10 Colapso de dados para a derivada de $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$. (a) Rede quadrada (b) Rede cúbica. 67

A.1	(a) Função erro. (b) Função erro complementar. Note que a soma das duas funções permanece constante, igual a 1, independentemente do valor de x	82
A.2	Gráfico da função gama A.7. A função é definida sob todo o plano complexo, exceto nos inteiros negativos, onde há divergência para os valores $+\infty$ ou $-\infty$	84
A.3	Comportamento da função beta em relação aos parâmetros a e b . Note que o aumento dos parâmetros leva a uma rápida diminuição do valor de $B(a, b)$	86
A.4	Comportamento da função digama. Assim como a função gama há divergência nos valores inteiros negativos.	87
B.5	Distribuição de Stacy com diversos valores dos parâmetros α e θ . (a) O aumento de α faz com que a assimetria da curva diminua. Note que com seu aumento a curva aproxima-se de uma normal. (b) Assim como na distribuição normal, um aumento na dispersão, θ , resulta em uma curva com maior variância.	92
B.6	Distribuição de Stacy em função do parâmetro τ . A medida que τ aumenta, o desvio padrão torna-se cada vez menor e a função aparentemente aproxima-se de uma sequência delta de Dirac.	93
B.7	Distribuição de log-gama em funções dos parâmetros α e λ . (a) O aumento de α diminui a dispersão em torno da média e modifica a forma da curva. (b) O aumento do parâmetro λ torna a curva mais dispersa.	98
C.8	Caminhada aleatória unidimensional. Os sítios são separados por uma distância l e a probabilidade de caminhar um passo ser dado para a esquerda ou direita vale, respectivamente, p e $q = 1 - p$	102

Lista de Tabelas

3.1	Resumo das distribuições de probabilidade para diferentes dimensionalidades e ordens de vizinhança. Aqui, o símbolo (-) significa valor arbitrário e ∞ é um valor muito grande e (*) significa distribuição na variável aleatória y , Eq 3.7.	27
4.1	Tabela resumo das principais grandezas conformacionais dos modelos de caminhada utilizados para representar cadeias poliméricas reais. Aqui, $f_1(\ell_p, R_{max}) = 1 - \exp(-R_{max}/\ell_p)$ e $f_2(\ell_p, R_{max}) = \ell_p^2(1 - 2\ell_p/R_{max}) + 2\ell_p^4/R_{max}^2(1 - \exp(-R_{max}/\ell_p))$	36
5.1	Tabela com os valores dos expoentes para duas e três dimensões. Tais expoentes foram obtidos de forma exata e/ou numérica. Para maiores detalhes, consulte a Ref. [16] para dados da rede quadrada e [28] para os dados da rede cúbica.	58
B.1	Tabela resumo das distribuições que são casos particulares da pdf Gama Generalizada/Stacy. O símbolo (-) indica valor arbitrário do parâmetro em questão.	99
D.2	Com essas seis funções principais do MPI é possível escrever um programa completo.	111
D.3	Equivalência entre os tipos de dado da linguagem C e os disponibilizados pelo MPI.	113
D.4	Operações básicas de redução, para mais detalhes, consulte a Ref. [82]. Quando os dados estão em um vetor, a redução ocorre elemento a elemento.	115

Lista de Abreviaturas e Siglas

CDPA	Caminhada determinista parcialmente autorrepulsiva.
CPU	<i>Central processing unit</i> (Unidade de processamento central).
FAMB	Programa de Física Aplicada à Medicina e Biologia.
FRC	<i>Freely rotating chain</i> (Cadeia de rotação livre).
MC	Monte Carlo.
MPI	<i>Message passing interface</i> (Interface de passagem de mensagens).
NUMA	<i>Non-uniform memory access</i> (Acesso não uniforme à memória).
pdf	<i>Probability density function</i> (Função densidade de probabilidade).
PPA	Problema do ponto aleatório.
RAM	<i>Random access memory</i> (Memória de acesso aleatório).
RW	<i>Random walk</i> (Caminhada aleatória).
SAW	Self-avoiding walk (Caminhada aleatória autorrepulsiva).
UMA	<i>Uniform memory access</i> (Acesso uniforme à memória).

Lista de Símbolos

Δ_i	O i -ésimo expoente de correção não analítico para a SAW.
ℓ_e	Comprimento de Kuhn.
ℓ_p	Comprimento de persistência de um passo.
γ	Constante de Euler-Mascheroni.
γ	Expoente entrópico para a SAW.
$\Gamma(a)$	Função Gama de a .
$\Gamma(a, b)$	Função gama incompleta complementar de a e b .
$\gamma(a, b)$	Função gama incompleta de a e b .
γ_1	Coefficiente de assimetria.
λ	Constante de atrito para a SAW.
$\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$	Produto escalar médio do vetor posição com o deslocamento no j -ésimo passo de uma caminhada com N passos.
$\langle x_N \rangle_N$	Comprimento de persistência de uma caminhada com N passos.
$\langle \vec{R}_g^2 \rangle_N$	Raio quadrático de giração médio de uma caminhada com N passos.
$B'_z(a, b)$	Função beta incompleta complementar de a e b .
$B(a, b)$	Função beta de a e b .
$B_z(a, b)$	Função beta incompleta de a e b .

$\operatorname{erfc}(z)$	Função erro complementar de z .
$\operatorname{erf}(z)$	Função erro de z .
$I_z(a, b)$	Função beta incompleta normalizada de a e b .
$P(a, b)$	Função gama incompleta normalizada de a e b .
$Q(a, b)$	Função gama incompleta complementar normalizada de a e b .
μ	Memória do caminhante na CDPA.
μ	Número de coordenação da SAW.
$\psi(x)$	Função digama de x .
$\psi^{(m)}(x)$	Função poligama de x .
ρ	Densidade de pontos do meio desordenado.
$\vec{R}_j$	Vetor posição do caminhante após j passos.
$\vec{u}_i$	Deslocamento do caminhante na SAW no i -ésimo passo.
C_∞	Raio Característico.
c_N	Número de SAWs com N passos.
d	Dimensão do meio Euclideano.
$D_{i,j}$	Distância entre os pontos i e j .
I_0	Integral Gaussiana de n -ésima ordem.
k	Ordem de vizinhança.
k_B	Constante de Boltzmann.
N_u	Número de pontos em uma esfera d -dimensional de raio unitário.
p_d	Volume relativo do crescente.
$P_{m,n}$	Probabilidade de Cox.
$V_d(l)$	Volume de uma hiperesfera de raio l .

Sumário

Lista de Figuras	xv
Lista de Tabelas	xxi
Lista de Abreviaturas e Siglas	xxiii
Lista de Símbolos	xxv
1 Introdução	1
2 Estatística de Vizinhança em Meios Desordenados	7
2.1 Processo Poissônico Espacial	8
2.2 A Fórmula de Cox	9
2.2.1 Limite de Alta Dimensionalidade	12
2.3 Aplicação à Caminhada Determinista Parcialmente Autorrepulsiva . .	14
2.4 Conclusão	15
3 Estatística de Distância em Meios Desordenados	17
3.1 Introdução e Revisão Bibliográfica	18
3.2 Solução Analítica	20
3.3 Casos Limite	23
3.3.1 Alta Dimensionalidade	23
3.3.2 Vizinho Distante	25
3.4 Aplicações	27
3.5 Conclusão	28

4	Caminhadas Aleatórias na Representação de Polímeros Ideais	29
4.1	Caminhada Aleatória	29
4.2	Cadeia de Rotação Livre	31
4.3	Comprimento de Persistência	36
4.4	O Modelo de Flory	37
4.5	Conclusão	39
5	Origem da Regra de Escala para a Caminhada Aleatória Autorre-	
	pulsiva	41
5.1	Introdução e Revisão Bibliográfica	42
5.2	Métodos	47
5.2.1	Grandezas Conformacionais	47
5.2.2	Algoritmo de Simulação Numérica	49
5.3	Produtos Escalares Intermediários	51
5.3.1	Relação com a Distância Quadrática	52
5.3.2	Média no <i>Ensemble</i> de Número de Passos	54
5.4	Comprimento de Persistência	55
5.5	Origem da Lei de Escala	61
5.6	Conclusões	65
6	Conclusão	69
	Referências	71
	Apêndice A - Funções Especiais	81
A.1	Erro	81
A.2	Gama	82
A.3	Beta	84
A.4	Digama e Poligama	86
	Apêndice B - Distribuições de Probabilidades	89
B.1	Distribuições Discretas	89
B.1.1	Bernoulli	89
B.1.2	Binomial	89

B.1.3	Poisson	90
B.2	Distribuições Contínuas	91
B.2.1	Stacy - Gama Generalizada	91
B.2.1.1	Exponencial	93
B.2.1.2	Gama	94
B.2.1.3	Weibull	94
B.2.1.4	Qui Quadrado - χ^2	95
B.2.1.5	Rayleigh	95
B.2.1.6	Maxwell-Boltzmann	96
B.2.2	Log-Gama	96
B.2.2.1	Normal - Gaussian	97
B.2.2.2	Gumbel Generalizada	98
Apêndice C - Caminhada Aleatória		101
C.1	Formulação do Modelo	101
C.2	Caso unidimensional	102
C.3	Caso Contínuo	103
C.4	Teorema Central do Limite a Partir da Caminhada Aleatória	105
Apêndice D - Computação em Paralelo Usando MPI		107
D.1	Surgimento da Computação Paralela	107
D.2	Arquitetura de Hardware	109
D.3	MPI	110
D.3.1	Principais Funções	111
D.3.2	Operações Coletivas de Comunicação	114
D.4	Métricas de Análise de Desempenho	116

xxx

Introdução

Na caminhada aleatória em redes regulares, um caminhante escolhe aleatoriamente e com igual probabilidade um sítio adjacente para ser visitado no próximo passo. Suas origens remontam ao início do século XX em um problema proposto por Karl Pearson [1]. Contudo, observações sobre fenômenos que ela pode descrever foram realizadas muito antes pelo botânico Brown (1828) [2], ao observar o movimento individual de grãos de pólen.

Tais caminhadas em redes regulares e suas variantes, com comprimento do passo dado por uma função densidade de probabilidade [3], são bem estudadas e utilizadas para descrever fenômenos Químicos [3], Biológicos [2] e Físicos [4] tratados, em geral, no contexto difusivo [5]. Menos comum são as variantes que dão origem a modelos de caminhadas determinista [6, 7] e autorrepulsivas [8, 9]. O caráter autorrepulsivo pode impedir ou não o caminhante de retornar a um sítio, ou conjunto de sítios, previamente visitado. O caminhante, no modelo da caminhada da rainha vermelha [6], segue a regra determinista de movimentação de ir a um dos sítios adjacentes menos degradado pela exploração prévia de seus recursos, ou que tenha se recuperado totalmente (os sítios recuperam o recurso com o passar do tempo). Note que a ausência de um sítio com todos os recursos disponíveis (ou totalmente recuperado) força o caminhante a visitar sítios que ele/ela, caso contrário, evitaria. Em outros modelos de caminhada autorrepulsiva o caminhante é impedido de retornar a um subconjunto dos sítios previamente visitados, tal como ocorre na movimentação de animais [9, 10], ou mesmo a todos os sítios visitados, assim como ocorre em modelos de caminhada que representam polímeros lineares em bom solvente [11].

O objetivo geral desta dissertação é abordar dois modelos de caminhada autorrepulsiva: o primeiro é o da caminhada determinista parcialmente autorrepulsiva (CDPA), onde o caminhante é impedido de visitar alguns sítios [8]; o segundo é o da caminhada aleatória autorrepulsiva (SAW), onde o caminhante é impedido de retornar a qualquer sítio previamente visitado [11].

Na CDPA, o grau de repulsão é regulado pelo parâmetro de memória μ . Nesse modelo, o substrato estático explorado pelo caminhante é gerado por um processo Poissoniano espacial, no qual a distância entre pares de sítios é dada pela métrica Euclideana. O caminhante segue a regra determinista de ir ao sítio mais próximo que não tenha sido visitado nos últimos μ passos. O deslocamento do caminhante pelo meio desordenado é dividido em duas grandezas: um tempo de transiente de t passos, onde o caminhante geralmente visita sítios distintos; e um período de atrator de p passos, onde o caminhante repete indefinidamente uma sequência de visitação imposta pela regra determinista de ir ao sítio mais próximo fora da janela de repulsão. Encontrar o atrator é o critério para finalizar a caminhada. Esse modelo foi proposto há pouco mais de uma década [8] e tem sido aplicado com sucesso no reconhecimento de padrões [12, 13], análise de textura em imagens [14], textos [15], entre outros.

Na (SAW), o substrato estático explorado é uma rede regular e o caminhante escolhe aleatoriamente e com igual probabilidade um sítio adjacente para ser visitado, o critério para o fim da caminhada é o caminhante tentar visitar qualquer um dos sítios previamente visitado. Nesse modelo de caminhada uma das grandezas fundamentais é o número c_N de configurações possíveis para uma trajetória com N passos. O deslocamento é caracterizado pelo expoente ν_0 da regra de escala da distância quadrática média entre o primeiro e o N -ésimo sítio em uma caminhada com N passos: $\langle \vec{R}_N^2 \rangle_N \sim N^{2\nu_0}$. Esse é um modelo de caminhada proposto há mais de meio século [16] e sua principal motivação é que ele pode descrever grandezas conformacionais de polímeros lineares em um bom solvente [17]. Além disso, o alto grau de complexidade imposto pela SAW motivou o desenvolvimento de técnicas combinatórias [18], computacionais baseadas em métodos de simulação Monte Carlo (MC) [19, 20] e de enumeração exata [21].

Especificamente, essa dissertação trata dois aspectos praticamente inexplo-

rados, um para cada modelo de caminhada. O estudo da SAW foi proposto pelo Prof. Dr. Marco Antônio Alves da Silva e realizado com sua colaboração.

Na CDPA foi mostrado na Ref. [22] que resultados da movimentação podem ser entendidos com base na estatística de vizinhança entre pares de pontos do meio desordenado. Essa estatística mede a probabilidade $P_{m,n}^{(d)}$ de um ponto I ser o m -ésimo vizinho mais próximo de seu n -ésimo vizinho mais próximo, o ponto J . Essa probabilidade foi calculada corretamente por Cox [23]. Esse problema foi revisitado recentemente na Ref. [24] por meio de uma descrição matemática conveniente que permite o cálculo da probabilidade $P_{m,n}^{(d)}$, assim como extensão para o caso limite de alta dimensionalidade $d \gg 1$. No Cap. 2 reproduzimos o cálculo da Ref. [24] e sua aplicação na descrição da movimentação do caminhante na CDPA. Nossa contribuição nesse assunto é estudar a estatística de distância, que mede como está distribuída a distância de um ponto ao k -ésimo vizinho mais próximo em um meio d -dimensional. Os novos resultados são apresentados no Cap. 3 e foram publicados recentemente na Ref. [25]. Em trabalhos futuros, pretendemos utilizar a estatística de distância para entender o processo difusivo gerado pelo caminhante na CDPA.

Na SAW, os métodos empregados atualmente visam calcular com alta precisão, ou até mesmo exatamente, o expoente principal ν_0 e expoentes de correção para a distância quadrática média, o número de coordenação μ e expoente γ da expressão $c_N \sim \mu^N N^{\gamma-1}$, que conta o número de configurações possíveis de uma SAW com N passos. Do ponto de vista analítico, busca-se determinar expoentes exatamente por meio do mapeamento em modelos de teoria de campo, tal como o n -vetorial [26, 27], do ponto de vista computacional busca-se algoritmos de simulação Monte Carlo cada vez mais eficientes para gerar dados mais precisos, a partir dos quais será calculado o valor de ν_0 [28, 29]. Aqui, nossa contribuição consiste em estudar a SAW por meio de uma nova abordagem baseada na extração de informação da trajetória por meio de produtos escalares. Tal abordagem é semelhante àquela empregada no Cap. 4 para o estudo de caminhadas mais simples que podem representar polímeros lineares. Os novos resultados da SAW são apresentados no Cap. 5.

Até o momento, apresentamos as motivações, objetivos gerais e específicos. Esta dissertação é tangente a esses objetivos, assim como documentado nos capítulos e apêndices que a constituem. Sua organização é dada a seguir.

No Cap. 2, apresentamos uma descrição formal do meio desordenado onde a caminhada determinista parcialmente autorrepulsiva é desenvolvida. Nesse mesmo capítulo, reproduzimos os cálculos da estatística de vizinhança para meios d -dimensionais [23] e sua generalização para alta dimensionalidade [24]. Ao final do capítulo, destacamos a aplicação da estatística de vizinhança na caminhada determinista parcialmente autorrepulsiva.

No Cap. 3, determinamos a equação que descreve a função densidade de probabilidade da estatística de distância ao k -ésimo vizinho mais próximo para meios desordenados d -dimensionais [25]. A partir da função densidade de probabilidade para k e d arbitrários, obtemos os casos limite de alta ordem de vizinhança $k \gg 1$, alta dimensionalidade $d \gg 1$ e a combinação desses dois parâmetros. Ao final do capítulo, discutimos duas possíveis aplicações dos nossos resultados. A primeira consiste em verificar se uma distribuição de pontos é gerada por um processo Poissônico espacial. A segunda explora casos particulares da estatística de distância para gerar várias funções densidade de probabilidade, tal como gama, Weibul, qui-quadrado, etc.

No Cap. 4, apresentamos alguns modelos de caminhada que podem ser utilizados para descrever características conformacionais de polímeros lineares. Os modelos de caminhada apresentados são: caminhada aleatória, caminhada com ângulo de rotação livre e o modelo Kratky-Porod [30]. Ao final do capítulo, reproduzimos o cálculo de Flory que leva em conta o efeito de volume excluído, sendo a motivação para o estudo do modelo da caminhada aleatória autorrepulsiva SAW no capítulo seguinte.

No Cap. 5, tratamos a caminhada aleatória autorrepulsiva nas redes quadrada e cúbica. Nesse modelo, a distância quadrática média ponta-a-ponta após N passos escala como $\langle \vec{R}_N^2 \rangle_N \sim N^{2\nu_0}$, com expoente principal $\nu_0 > 1/2$, onde $\vec{R}_N = \vec{u}_1 + \vec{u}_2 + \dots + \vec{u}_N$ é o vetor posição e $\vec{u}_i$ é o vetor deslocamento no i -ésimo passo. Geralmente, os métodos utilizados para determinar esse expoente consistem no cálculo (via simulação ou enumeração exata) da distância quadrática média em função de N e então, por meio de ajustes numéricos encontrar ν_0 . Outra abordagem, menos comum, determina $\langle \vec{R}_N^2 \rangle_N$ por meio do produto escalar $\vec{u}_i \cdot \vec{u}_j$ com $1 \leq i, j \leq N$. Nosso método para estudar a SAW consiste em analisar o produto

escalar médio entre o vetor posição e o deslocamento no j -ésimo passo $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$ em uma caminhada com N passos.

Por meio de operação de simetria determinamos a relação não trivial entre o comprimento de persistência $\langle x_N \rangle_N$ (projeção do vetor posição na direção do primeiro passo) e distância quadrática média: $\langle \vec{R}_N^2 \rangle_N = \langle \vec{R}_{N-1}^2 \rangle_N + 2\langle x_N \rangle_N - 1$, onde $\langle \vec{R}_{N-1}^2 \rangle_N$ é a distância quadrática média ao penúltimo passo no *ensemble* de caminhadas com N passos. Analiticamente e com dados de simulação Monte Carlo, determinamos que $\langle x_N \rangle$ converge para um valor constante. Com dados de simulação MC para $1 \leq j \leq \Theta(N)$ obtemos a regra de escala $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N \sim 1/2 + \alpha_0(j - \tau)^{2\nu_0 - 1}$, onde τ é conhecida como constante de suavização e $\nu_0 > 1/2$ para rede quadrada e cúbica. No intervalo $\Theta(N) \leq j \leq N$ a regra de escala anterior não é suficiente para descrever os dados, pois a partir desse passo o produto escalar médio $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$ cresce mais lentamente até atingir um ponto de máximo, decrescendo monotonicamente até $j = N$. Numericamente determinamos $\Theta(N) \sim N/2$ e $\Theta(N) \sim N/3$ para as redes quadrada e cúbica, respectivamente. A partir desse valor de $\Theta(N)$ a caminhada torna-se mais flexível. Esse aumento da flexibilidade se deve à maior liberdade de movimentação média que o final da caminhada experimenta em relação à parte inicial.

As considerações finais, perspectiva e conclusão são apresentadas no Cap. 6. No Apêndice A, compilamos as principais funções especiais utilizadas nos cálculos da estatística de vizinhança e distância. No Apêndice B, reunimos as principais distribuições de probabilidade discretas e funções densidade de probabilidade. No Apêndice C, listamos alguns resultados relacionados com a caminhada aleatória. No Apêndice D, apresentamos uma breve introdução à computação em paralelo utilizando MPI.

Estatística de Vizinhança em Meios Desordenados

A estatística de vizinhança mede a probabilidade de um ponto I ser o m -ésimo vizinho mais próximo de seu n -ésimo vizinho mais próximo, o ponto J . Historicamente, esse problema de reciprocidade de ordem de vizinhança foi motivado pelo estudo de agregação de plantas. Do ponto de vista estatístico, os pontos são distribuídos aleatoriamente e com densidade constante segundo um processo Poissônico espacial em um meio d -dimensional. Nesse meio desordenado, a distância Euclideana é o critério de ordenamento da vizinhança entre pares de pontos.

Iniciamos o presente capítulo definindo um processo Poissônico espacial e sua possível representação computacional dada pelo problema do ponto aleatório (PPA). Por meio da distribuição de Poisson, reproduzimos os cálculos da estatística de vizinhança. Esse resultado é então reescrito de forma conveniente em função da distribuição multinomial e posteriormente generalizado para o caso particular de alta ordem de vizinhança. A caminhada determinista parcialmente autorrepulsiva é desenvolvida por um caminhante no meio desordenado onde a estatística de vizinhança é calculada. Apesar da caminhada ser dinâmica, apresentamos alguns resultados da estatística de vizinhança que permitem entender certos comportamentos do caminhante na CDPA. Essa aplicação é a nossa motivação para caracterização do meio desordenado através da estatística de vizinhança e de distâncias (esta última abordada no próximo capítulo).

Este capítulo organiza-se da seguinte maneira. Na Sec. 2.1, apresentamos o processo Poissônico espacial utilizado para gerar o meio desordenado. Na Sec. 2.2,

reproduzimos o cálculo da estatística de vizinhaça. Na Sec. 2.3 discutimos sua aplicação na caminhada determinista parcialmente autorrepulsiva. Ao final, na Sec. 2.4 apresentamos as conclusões.

2.1 Processo Poissônico Espacial

Em geral, os livros textos de Estatística apresentam eventos aleatórios que ocorrem no tempo para caracterizar um processo Poissônico [31], número de decaimentos de partículas radioativas ou chamadas que chegam a uma central telefônica [32], por exemplo. Nesse processo, o número de eventos em um intervalo de tempo depende apenas de sua magnitude, eventos em intervalos de tempos disjuntos são independentes e dois ou mais eventos não ocorrem simultaneamente. A probabilidade de k eventos ocorrerem em um tempo Δt é dada pela fórmula de Poisson Eq. B.34, detalhes sobre a distribuição de Poisson podem ser encontrados no Apêndice B.

No problema da estatística de vizinhaça, os eventos são os pontos¹ distribuídos ao longo de um meio d -dimensional ilimitado. O processo espacial deve seguir as mesmas propriedades do processo temporal, especificamente:

1. apenas um ponto ocupa uma dada posição do espaço;
2. os pontos são gerados aleatória e independentemente e
3. o número de pontos no interior de um volume é proporcional a esse volume.

Assim como destacado na Ref. [33], considerar tempo ou espaço pode alterar os resultados, pois ao contrário do tempo, o espaço apresenta ao menos dois sentidos (caso unidimensional). Além disso, os pontos podem ser posicionados em um espaço d -dimensional, até mesmo no limite $d \gg 1$.

Do ponto de vista computacional, o meio desordenado pode ser gerado pelo problema do ponto aleatório (PPA) [34]. Ele consiste em gerar as coordenadas de N pontos em cada aresta de um hipercubo d -dimensional, aleatória e independentemente, seguindo uma pdf uniforme. Além da dimensão, o outro parâmetro que

---

¹Embora o termo “sítios” esteja associado à redes regulares, utilizamos este termo como sinônimo do termo “pontos” para o meio desordenado.

caracteriza esse meio desordenado é a densidade de pontos, assim como ilustrado na Fig. 2.1. A distância entre quaisquer pares de pontos é obtida por meio da métrica Euclideana

$$D_{i,j} = [\sum_{k=1}^d (x_i^{(k)} - x_j^{(k)})^2]^{1/2}, \quad (2.1)$$

onde há a restrição de simetria $D_{i,j} = D_{j,i}$, desigualdade triangular $D_{i,j} + D_{j,k} \geq D_{i,k}$ e a distância de um ponto a ele mesmo é nula $D_{i,i} = 0$.

Figura 2.1 – Meio desordenado bidimensional gerado pelo problema do ponto aleatório com densidade de pontos (a) ρ_1 , (b) $2\rho_1$ e (c) $3\rho_1$. Diferentemente de uma rede regular, onde os pontos são igualmente espaçados, há aqui pequenas subáreas com aglomeração ou vazio de pontos e a distância de um ponto aos demais está distribuída em torno de uma distância média.

2.2 A Fórmula de Cox

Considere um meio d -dimensional, ilimitado, isotrópico e homogêneo onde os pontos são gerados pelo processo Poissônico espacial descrito anteriormente. A probabilidade de encontrar k pontos em um volume V_d é dada pela distribuição de Poisson: $P(k) = \lambda^k e^{-\lambda} / k!$, com $k = 1, 2, \dots, \infty$ e $\lambda = \rho V_d$, onde ρ é a densidade de pontos por unidade de volume.

Dado um par arbitrário de pontos I e J , estamos interessados em calcular a probabilidade do ponto I ser o m -ésimo vizinho mais próximo do seu n -ésimo vizinho mais próximo, o ponto J . A essa probabilidade denominamos probabilidade de Cox, que é ilustrada pela Fig. 2.2 e representada por $P_{m,n}^{(d)}$.

A probabilidade $P_{m,n}^{(d)}$ para $N \gg 1$ foi obtida por Clark e Evans [35] para o caso $m = n = 1$ e posteriormente generalizada por Clark para vizinhos recíprocos $m = n$ [36]. Dacey corrigiu a expressão obtida por Clark [37], que estava correta apenas para o caso $m = n = 1$. A estatística de vizinhaça foi generalizada por Cox para o caso $m \neq n$ [23] e interpretada em termos da distribuição multinomial por Terçariol *et al.* [24].

Figura 2.2 – Probabilidade de Cox em um meio bidimensional. Há i pontos na intersecção dos dois círculos, $n - i - 1$ pontos no crescente do ponto I e $m - i - 1$ no crescente do ponto J .

De acordo com a Fig. 2.2, os pontos I e J estão $D_{I,J} = l$ distantes um do outro. O volume da hipersfera centrada em I que passa por J é $V_d(l) = \pi^{d/2} l^d / \Gamma(d/2 + 1)$ e o volume da hipersfera centrada em J que passa por I é igual a $V_d(l)$.

No cálculo das probabilidades de Cox, as variáveis aleatórias são dadas pelas ordens de vizinhaça m e n , o meio é ilimitado, portanto a distância l entre pares de pontos pode variar de $[0, \infty)$ assim como o volume hipersfera centrado em cada ponto. Do ponto de vista matemático, é conveniente trabalhar com volumes relativos. O primeiro passo é calcular o volume da intersecção das hipersferas. Note que essa intersecção pode ser subdividida em duas hipercalotas [38] (linha tracejada da Fig. 2.2), seu volume é dado:

$$V_{\cap,d}(l) = \frac{\pi^{(d-1/2)}}{\Gamma[(d+1/2)]} \int_{1/4}^1 dt t^{1/2} (1-t)^{(d-1)/2}. \quad (2.2)$$

O volume relativo do crescente é definido como a razão entre o volume externo à intersecção das esferas e o volume total da esfera: $p_d = [V_d(l) - V_{\cap,d}(l)]/V_d(l)$. Manipulando a Eq. 2.2 obtemos a função beta incompleta normalizada A.27:

$$p_d = I_{1/4} \left(\frac{1}{2}, \frac{d+1}{2} \right). \quad (2.3)$$

Figura 2.3 – Possíveis configurações onde o sítio I é o quarto vizinho mais próximo de J , enquanto J é o terceiro vizinho mais próximo de I .

Determinada a razão dos volumes, é necessário impor as condições referentes ao número de pontos na intersecção e nos crescentes para calcular $P_{m,n}^{(d)}$. Tais condições são:

1. deve haver i pontos na intersecção das hiperesferas variando de 0 até $\min(m - 1, n - 1)$, do contrário a condição do m -ésimo vizinho do n -ésimo vizinho não é respeitada, veja Fig. 2.3. O número de pontos esperado na intersecção é $\mu(1 - p_d)$, onde $\mu = \rho V_d$;
2. deve haver $m - i - 1$ pontos no crescente de J , o valor esperado é μp_d ;
3. deve haver $n - i - 1$ pontos no crescente de I , o valor esperado é μp_d e
4. o número de pontos esperado em uma hiperesfera $\mu = \rho V_d$ pode assumir qualquer valor no intervalo $[0, \infty)$.

Obedecendo a estas condições e levando em conta o fato que $P_{m,n}^{(d)}$ deve ser igual a $P_{n,m}^{(d)}$ temos o seguinte resultado

$$P_{m,n}^{(d)} = \int_0^\infty d\mu \sum_{i=0}^{\min(m-1, n-1)} \frac{[\mu(1 - p_d)]^i e^{-\mu(1-p_d)}}{i!} \frac{(\mu p_d)^{m-i-1} e^{-\mu p_d}}{(m-i-1)!} \frac{(\mu p_d)^{n-i-1} e^{-\mu p_d}}{(n-i-1)!}. \quad (2.4)$$

Colocando os fatores que não dependem de μ fora da integral e trabalhando com a função gama, Eq. A.7, encontramos o resultado originalmente obtido por Cox [23], isto é,

$$P_{m,n}^{(d)} = \sum_{i=0}^{\min(m-1, n-1)} \frac{(m+n-i-2)!}{i!(m-i-1)!(n-i-1)!} \frac{(1-p_d)^i (p_d)^{m+n-2i-2}}{(1+p_d)^{m+n-i-1}}. \quad (2.5)$$

Um desenvolvimento mais aprofundado, presente na referência [24], resulta em

$$P_{m,n}^{(d)} = \frac{1}{1+p_d} \sum_{i=1}^{\min(m,n)} \frac{(m+n-i-1)!}{(i-1)!(m-i)!(n-1)!} \frac{(1-p_d)^{i-1}}{(1+p_d)^{i-1}} \left(\frac{p_d}{1+p_d} \right)^{m-i} \left(\frac{p_d}{1+p_d} \right)^{n-i}. \quad (2.6)$$

De acordo com a Eq. 2.6, fica evidente que o valor de $1/(1+p_d) = P_{1,1}^{(d)}$ é a probabilidade de dois pontos serem mutualmente mais próximos. É possível escrever a equação acima em função da distribuição discreta multinomial²

$$\frac{P_{m,n}^{(d)}}{P_{1,1}^{(d)}} = \sum_{i=1}^{\min(m,n)} \text{mult} \left(i-1, m-1, n-1; \frac{1-p_d}{1+p_d}, \frac{p_d}{1+p_d}, \frac{p_d}{1+p_d} \right), \quad (2.7)$$

note que a dimensionalidade do meio desordenado está implícita no fator p_d , que representa o volume relativo do crescente. Por meio da manipulação desse fator, reobtemos $P_{m,n}^{(d)}$ para o caso limite $d \gg 1$ a seguir.

2.2.1 Limite de Alta Dimensionalidade

Analisar $P_{m,n}^{(d)}$ no limite $d \gg 1$ consiste em entender como o volume do crescente se comporta em função da dimensão do meio. Numericamente, o valor de p_d é fornecido pela Eq. 2.3, que pode ser reescrita como

$$p_d = \frac{\int_0^{1/4} dt t^{-1/2} (1-t)^{(d-1)/2}}{B[1/2, (d+1)/2]}, \quad (2.8)$$

como $a = 1/2$ e $b = (d+1)/2$, a função beta da Eq. 2.8 pode ser escrita de acordo com a Eq. A.24, pois para $d \gg 1$ temos $b \gg a$:

$$p_d = \frac{b^a}{\Gamma(a)} \int_0^{1/4} dt t^{a-1} (1-t)^b, \quad (2.9)$$

podemos escrever $(1-t)^b = e^{b \ln(1-t)}$, como a variável t está limitada ao intervalo $0 < t < 1/4$ e o parâmetro $b = (d+1)/2 \gg 1$ a aproximação $e^{b \ln(1-t)} \approx e^{-bt}$ é válida

---

²A função de probabilidade que da distribuição multinomial é $P(Y_1 = n_1, Y_2 = n_2, \dots, Y_k = n_k) = \frac{n!}{n_1! n_2! \dots n_k!} p^{n_1} p^{n_2} \dots p^{n_k}$.

Figura 2.4 – Comparação entre o valor exato do volume do crescente, Eq. 2.3, e o cálculo aproximado no limite de alta dimensionalidade, dado pela Eq. 2.12. Para $d \geq 10$ a aproximação se torna muito acurada.

e a Eq. 2.9 torna-se

$$p_d = \frac{b^a}{\Gamma(a)} \int_0^{1/4} dt t^{a-1} e^{-bt}. \quad (2.10)$$

Realizando a substituição $x = bt$, na Eq. 2.10 obtemos que o volume relativo do crescente é escrito como a função gama complementar normalizada

$$p_d = \frac{1}{\Gamma(a)} \int_0^{b/4} dx x^{a-1} e^{-x} = \frac{\gamma(a, b/4)}{\Gamma(a)}. \quad (2.11)$$

Da Eq. A.15, a função $\gamma(a = 1/2, b/4)$ pode ser escrita de acordo com a função erro A.1 como $\gamma(1/2, b/4) = \sqrt{\pi} \operatorname{erf}(\sqrt{b/4})$. Dado que $\Gamma(1/2) = \sqrt{\pi}$ e $b \approx d/2$, a Eq. 2.11 torna-se $p_d \approx \operatorname{erf}(\sqrt{d/8})$. Em termos da função erro complementar Eq. A.2

$$p_d \approx 1 - \operatorname{erfc}\sqrt{\frac{d}{8}} = 1 - \alpha_d, \quad (2.12)$$

a comparação dessa aproximação com a Eq. 2.3 é ilustrada no gráfico da Fig. 2.4. No caso de alta dimensionalidade a aproximação $p_d \approx 1$ é razoável, pois o argumento da função erro complementar na Eq. 2.12 é muito maior que a unidade, isso implica de acordo com o gráfico da Fig. A.1b, que a função erfc é próximo de zero. Agora podemos reescrever as probabilidades de Cox no limite de alta dimensionalidade ao substituir $p_d \approx 1$ e $1 - p_d = \alpha_d$ na Eq. 2.5

$$P_{m,n}^{(d \gg 1)} = \sum_{i=0}^{\min(m-1, n-1)} \frac{(m+n-2-i)! \alpha_d^i}{i!(m-i-1)!(n-i-1)! 2^{m+n-i-1}}. \quad (2.13)$$

Na Eq. 2.13 considerar $i = 0$ é uma aproximação razoável, pois o volume da intersecção entre as duas hipersferas é praticamente nulo. A função já trabalhada em termos da Eq. 2.13:

$$\frac{P_{m,n}^{(d \gg 1)}}{P_{1,1}^{(d \gg 1)}} = \frac{1}{2^{m+n-2}} \frac{\Gamma(m+n-1)}{\Gamma(m)\Gamma(n)}, \quad (2.14)$$

onde a probabilidade de encontrar um casal (pontos mutualmente mais próximos) é $P_{1,1}^{(d \gg 1)} = 1/2$. A seguir, destacamos como alguns resultados da caminhada determinista parcialmente autorepulsiva CDPA podem ser compreendidos com a estatística de vizinhaça.

2.3 Aplicação à Caminhada Determinista Parcialmente Autorrepulsiva

Nessa seção, tratamos brevemente do uso da estatística de vizinhaça para descrever alguns resultados da CDPA, para uma revisão completa sobre essa caminhada recomendamos a Ref. [39]. O substrato estático³ onde a caminhada se desenvolve é o meio desordenado gerado por um processo Poissônico espacial (ver Fig. 2.1), onde a distância entre os pontos é dada pela métrica Euclideana, Eq. 2.1. O caminhante parte de um ponto s_0 (Fig. 2.5) e se movimenta de acordo com a regra determinista de ir ao ponto mais próximo que não tenha sido visitado nos últimos μ passos.

A trajetória que o caminhante percorre no meio é descrita por duas grandezas: o transiente de t passos, onde ele/ela passa por um conjunto de pontos sem seguir um padrão de visitaçaõ [8, 24] e o período de atrator de p passos, onde o mesmo conjunto de sítios é visitado sempre na mesma sequência de visitaçaõ. Identificar que o caminhante entrou no atrator é o critério de parada na CDPA.

Na CDPA a autorrepulsão é controlada pelo parâmetro de memória μ . Quando $\mu = 0$, o caminhante não conhece o ponto onde está, como $D_{i,i} = 0$ o caminhante fica preso em um atrator de $p = 1$. Para $\mu = 1$, o caminhante conhece o ponto onde se encontra, deve ir ao sítio mais próximo. A Fig. 2.5 ilustra esse

---

³Chamamos substrato estático ao meio desordenado pois este não sofre nenhuma mudança temporal, ou devido à movimentação do caminhante.

Figura 2.5 – Caminhada determinista parcialmente autorrepulsiva em um meio unidimensional e caminhante com $\mu = 1$. O caminhante parte do sítio s_0 e sempre vai ao sítio mais próximo, percorrendo os sítios de s_0 à s_4 em 4 passos, que compõem o transiente. Como o sítio s_4 e s_5 são mutualmente mais próximos (casal), o caminhante vai de $s_4 \rightarrow s_5$ e de $s_5 \rightarrow s_4$ indefinidamente, pois a memória $\mu = 1$ não permite que ele/ela visite outro sítio. Neste caso, o atrator é composto por 2 passos. A cada passo, a probabilidade do caminhante visitar um sítio que pertence a um casal é dada pela probabilidade de Cox, Eq. 2.6, para o caso $m = n = 1$. De posse da estatística de vizinhança, uma das características estáticas do meio desordenado, é possível calcular a distribuição do número de passos no transiente para $\mu = 1$, que é uma característica relacionada com a dinâmica de movimentação.

caso em um meio unidimensional. Para $\mu = 2$ o caminhante conhece o sítio onde se encontra e o que visitou anteriormente. Para um caminhante com memória $\mu = 1$, a caminhada sempre é finalizada em um atrator de período $p = 2$ [33]. A probabilidade do caminhante entrar em um atrator de período $p = 2$ logo no primeiro passo, é dada por $P_{1,1}^{(d)}$ [40], ou seja, a probabilidade de encontrar pontos mutualmente mais próximos (casal). No segundo passo, novamente a probabilidade do caminhante ficar preso em sítios mutuamente mais próximos é $P_{1,1}^{(d)}$. Como $P_{1,1}^{(d)} \geq 1/2 \forall d$, a probabilidade do caminhante explorar muitos pontos do meio é baixa e ele/ela encontra um atrator com $p = 2$ em poucos passos [41]. Para $\mu \geq 2$ o papel das probabilidades $P_{m,n}$ na movimentação do caminhante ainda não é bem entendido e carece de estudo.

2.4 Conclusão

Nesse capítulo caracterizamos o processo Poissônico espacial utilizado para gerar o meio desordenado onde a CDPA é desenvolvida. Nesse meio reproduzimos o cálculo da estatística de vizinhança, que mede a probabilidade de um ponto ser o m -ésimo vizinho mais próximo de seu n -ésimo vizinho mais próximo, ou simplesmente, probabilidade de Cox. O comportamento dessas probabilidades é modificado no limite de alta dimensionalidade devido ao fator p_d , assim como o cálculo aproximado

usando funções especiais mostrou. Ao final da seção discutimos a aplicação da fórmula de Cox no contexto da caminhada parcialmente autorrepulsiva CDPA. O próximo capítulo trata da estatística de distância, que mede como está distribuída a distância de um ponto ao seu k -ésimo vizinho mais próximo no meio desordenado.

Estatística de Distância em Meios Desordenados

A estatística de distância mede a distribuição de distâncias entre um ponto e seu k -ésimo vizinho mais próximo em um meio desordenado d -dimensional. Assim como na estatística de vizinhança, a desordem do meio é gerada por um processo Poissônico espacial. No presente capítulo, investigamos detalhadamente como essa estatística é afetada pelos parâmetros: densidade de pontos, dimensionalidade e ordem de vizinhança. Inicialmente, calculamos a distribuição de distâncias ao k -ésimo vizinho mais próximo em um meio com dimensionalidade e densidade de pontos arbitrárias. A pdf resultante que descreve a distribuição de distâncias a distribuição gama generalizada.

A partir da expressão geral para a distribuição de distâncias, exploramos os casos limite de alta dimensionalidade que leva à distribuição de Gumbel, alta ordem de vizinhança que leva à distribuição Gaussiana e a combinação desses dois últimos casos. O resultado para alta ordem de vizinhança foi obtido ao considerarmos uma expansão mais acurada da razão de funções gama: $\Gamma(z+x)/\Gamma(z) \approx z^x \exp(-x/2z + 3x^2/4z)$ para $z \gg x$. Essa expansão permitiu provar a conjectura de Cerf *et. al.* [42] com relação à distância média (primeiro momento) para $d \gg 1$, além de calcular os momentos de mais alta ordem.

O problema da estatística de distâncias não se aplica apenas ao deslocamento do caminhante na CDPA. A partir da distribuição gama generalizada, que descreve a estatística de distâncias, obtemos várias funções densidade de probabilidade variando a dimensão e ordem de vizinhança, tal como exponencial, gama,

Weibull, Rayleigh, normal, Nakagami, etc. Argumentamos que esse problema pode ser usado como motivação geométrica para ilustrar o surgimento dessas distribuições estatísticas. A última aplicação que destacamos é um teste para detectar se uma distribuição de pontos foi ou não gerada por um processo Poissônico espacial. Para realizar tal teste, é necessário escrever a distribuição (originalmente escrita na variável distância ao k -ésimo vizinho mais próximo) em uma variável proporcional ao volume compreendido entre um ponto e o k -ésimo vizinho mais próximo. Nesta última variável aleatória a distribuição obtida é uma χ^2 com $2k$ graus de liberdade. Conhecendo a densidade de pontos do meio se torna possível avaliar se o valor da distância média ao primeiro vizinho, segundo, ..., k -ésimo vizinho está de acordo com a hipótese Poissônica de distribuição dos pontos. Esse teste foi originalmente proposto por Thompson [43] para meio bidimensional e generalizado neste capítulo para dimensão arbitrária

Este capítulo organiza-se da seguinte maneira. Na Sec. 3.1 apresentamos a introdução e revisão bibliográfica. Na Sec. 3.2 calculamos a pdf da estatística de distâncias. Na Sec. 3.3 estudamos os casos limite de alta ordem de vizinhança e alta dimensionalidade, assim como a combinação destes dois. Na Sec. 3.4 apresentamos as aplicações da estatística de distâncias. Por último, na Sec. 3.4, expomos as conclusões referentes a esse capítulo. Os resultados desse capítulo foram publicados na Ref. [25].

3.1 Introdução e Revisão Bibliográfica

Considere um meio d -dimensional, ilimitado, isotrópico e homogêneo com perturbações geradas por um processo Poissônico espacial (pontos). O número de pontos esperado em um volume V_d é $\lambda = \rho V_d$, onde ρ é a densidade de pontos. Esse meio desordenado, apesar de ilimitado, pode ser representado computacionalmente como um hipercubo d -dimensional, que contém N coordenadas aleatoriamente distribuídas com função densidade de probabilidade uniforme em cada aresta (*random point problem*) [34]. Essa é uma maneira de construir o meio desordenado, na qual, as distâncias entre pares de pontos deixam de ser regulares e passam a variar estatisticamente. Nesse meio desordenado, é possível explorar duas estatísticas: vizinhança

e distâncias.

A estatística de distâncias mede como está distribuída a distância de um ponto ao seu k -ésimo vizinho mais próximo em um meio d -dimensional. Em Física e Biologia esse problema é aplicado no cálculo da separação média entre corpos celestes [44], determinação de agregação em comunidade de plantas [45, 43], trajetória ótima no problema do caixeiro viajante [42, 46], *Euclidean matching problem* [47, 48], caminhadas parcialmente autorepulsivas [8, 49], filmes finos [50], entre outros. Em Computação, o cálculo das distâncias aos primeiros vizinhos é empregado como classificador de padrões [51, 52], além de ser utilizado para o quantificar a distância entre terminais de rede [53].

Até o momento, o estudo concentra-se em duas frentes. A primeira, é o cálculo da distância média entre pontos [54, 55] e obtenção de seus momentos de mais alta ordem [56, 57] para diferentes configurações da distribuição de pontos. A segunda, é o cálculo da distribuição para baixa dimensionalidade, $d \leq 3$, para vizinho mais próximo [45, 50] e vizinhança arbitrária [43]. A distribuição de pontos ao n -ésimo vizinho em dimensão arbitrária foi obtida por Martin [53] no contexto da distribuição de distâncias entre terminais de acesso à *internet*. Apesar da expressão matemática ser conhecida [53, 58], a influência dos parâmetros são pouco exploradas, principalmente nos casos limite de alta dimensionalidade e alta ordem de vizinhança, que são pouco triviais.

Esses casos limite são pouco triviais devido à razão de funções gama $\Gamma(z+x)/\Gamma(z)$ para $z \gg x$. Se uma expansão mais simples dessa razão for considerada, inconsistências tal como os momentos centrais indefinidos ocorrem. O principal objetivo desse capítulo é corrigir essas inconsistências por meio de uma expansão mais acurada. Essa expansão mais acurada não permite apenas retirar as inconsistências, mas também a derivar os casos limite da estatística de distâncias e provar a conjectura de Cerf. *et. al.* [42].

As expressões obtidas para o caso de alta dimensionalidade confirmam a equivalência do modelo Euclidiano com o de ligações aleatórias (*random link*) [24] em alta dimensionalidade. O caso de alta ordem de vizinhança, $k \gg 1$, indica que a distribuição de distância converge para a Gaussiana. Além de descrever a distância até o k -ésimo vizinho, é possível detectar se a distribuição dos pontos segue um

processo Poissônico, assim como proposto por Thompson [43] e generalizado por nós. A partir da pdf geral para a estatística de distâncias compilamos as diferentes distribuições obtidas como casos especiais ao variar a dimensionalidade e ordem de vizinhança.

A seguir calculamos a pdf para estatística de distâncias de duas maneiras distintas. A primeira, é baseada em argumentação geométrica e a segunda na utilização de funções acumuladas, sendo a *pdf* resultante descrita pela distribuição gama generalizada [59, 60]. A partir dessa pdf reobtemos os momentos de mais alta ordem e a expansão mais acurada da razão das funções gama.

3.2 Solução Analítica

Nessa seção obtemos a expressão para a estatística de distância e a validação desta por meio de simulação Monte Carlo. Os momentos da distribuição são escritos de maneira mais simples em termos da expansão da razão $\Gamma(z+x)/\Gamma(z)$.

Considere um meio d -dimensional com densidade ρ , onde $\rho = \rho_1^d$ e ρ_1 é a densidade linear de pontos. Essa correção na densidade mantém a separação média entre pontos constante, o que permite comparar sistemas de diferentes dimensionalidades. O número de pontos esperado em uma hiperesfera de raio l é $\lambda = N_u l^d$ onde $N_u = \rho \pi^{d/2} / \Gamma(1 + d/2)$ é o número de pontos em uma esfera d -dimensional de raio unitário. A probabilidade de haver k pontos no interior de uma esfera de raio l é dado pela fórmula de Poisson, $P(k) = e^{-\lambda} \lambda^k / k!$.

O primeiro método para obter a estatística de distância é geométrico. A probabilidade de k vizinhos mais próximos caírem dentro de uma esfera de raio $l + dl$ é escrita como a probabilidade de uma esfera de raio l conter $k - 1$ vizinhos e uma fina casca esférica, de espessura dl , conter o k -ésimo $f_{\rho,d}^{(k)}(l)dl = P(k-1)P(1)$. Em uma casca esférica são esperados $d\lambda = dN_u l^{d-1}dl$ pontos de modo que

$$f_{\rho,d}^{(k)}(l)dl = \frac{e^{-N_u l^d} (N_u l^d)^{k-1}}{(k-1)!} e^{-dN_u l^{d-1}dl} dN_u l^{d-1}dl.$$

Como $dl \ll l$, a função de distribuição de probabilidade torna-se:

$$f_{\rho,d}^{(k)}(l) = \frac{dN_u l^{dk-1}}{\Gamma(k)} \exp(-N_u l^d), \quad (3.1)$$

onde k , é a ordem de vizinhança. A Eq. 3.1 é mapeada na distribuição gama generalizada, Eq. B.36, com o seguinte ajuste de parâmetros $\alpha = k$, $\tau = d$ e $\theta = N_u^{-\beta}$, onde $\beta = 1/d$ a partir desse ponto. Note que o parâmetro θ é afetado pela densidade de pontos e dimensionalidade do meio, além de ser afetado de maneira não trivial pela simetria do meio.

Na realização de uma simulação computacional, θ é afetado pelas bordas do meio por meio do parâmetro ρ . Se for considerada uma simulação em um hipercubo com aresta de comprimento $\mathcal{L}$ com N pontos, $\rho = N/\mathcal{L}^d$, caso a simetria do meio seja esférica $\rho = N\Gamma(1 + d/2)/\pi^{d/2}\mathcal{L}^d$.

A validação da Eq. 3.1 foi obtida por meio de simulação Monte Carlo desenvolvida em linguagem C e paralelizado em MPI. O meio utilizado tem simetria cúbica, número de pontos N , e densidade $\rho = N/\mathcal{L}^d$, onde $\mathcal{L}$ é o comprimento da aresta. Aplicar os resultados da Eq. 3.1 nesse meio limitado consiste em uma aproximação, devido ao efeito de borda, pois os pontos das extremidades tem menos vizinhos. Uma maneira de minimizar esse efeito é utilizar condições periódicas de contorno. O cenário simulado numericamente é apresentado no gráfico da Fig. 3.1. Pelo gráfico é possível notar que o aumento da ordem de vizinhança recupera a simetria da curva em torno do valor médio. A correção devido ao efeito de tamanho finito é da ordem de $1/N$ [54]. A Eq. 3.1, na variável $\lambda = N_u l^d$, que é o número de pontos esperado em uma esfera d -dimensional de raio l , leva ao colapso da distribuição: $f^{(k)}(\lambda) = \lambda^{k-1}e^{-\lambda}/\Gamma(k)$, que é a distribuição gama B.42.

O segundo método de dedução é baseado na utilização de funções acumuladas. Por simplicidade, considere o primeiro vizinho do ponto i em um meio bidimensional. Dado um ponto i , a probabilidade de não encontrar outro ponto em um raio l é $P(k = 0) = e^{-\rho\pi l^2}$. Considere a variável aleatória L , que descreve a distância até o ponto mais próximo, $L > l$ se, e somente se, não houver pontos na área πl^2 , portanto $P(L > l) = e^{-\rho\pi l^2}$. Dessa maneira, podemos encontrar a função acumulada da distribuição de L escrevendo $P(L \leq l) = 1 - P(L > l) = F_{\rho,d}^k(l)$. A pdf que descreve a distância ao primeiro vizinho é a derivada da função acumulada: $f_{\rho,d}^{(1)} = 2\rho\pi l e^{-\rho\pi l^2}$. O processo de dedução por meio de funções acumuladas estendido aos meios de dimensionalidade e ordem de vizinhança arbitrários leva à Eq. 3.1.

Figura 3.1 – Comparação entre o resultado analítico gerado pela Eq. 3.1 (linhas cheias) até o quinto vizinho em um meio bidimensional. Simulação realizada com $\rho = 50000$, com condições periódicas de contorno, as barras de erro são equivalentes ou menores que o tamanho do ponto. As linhas cheias correspondem aos resultados analíticos dado pela Eq. 3.1. Note que o aumento da ordem de vizinhança recupera a simetria da distribuição. Gráfico adaptado da Ref. [25]

A distância média de um ponto i ao seu k -ésimo vizinho vale

$$\langle l_{\rho,d}^{(k)} \rangle = N_u^{-\beta} \frac{\Gamma(k + \beta)}{\Gamma(k)}. \quad (3.2)$$

Assim, como no resultado obtido por Percus e Martin [54], há a fatorização entre o número de pontos, no caso a densidade, e a ordem de vizinhança do meio. A variância em torno da média é

$$\sigma^2(l)_{\rho,d,k} = N_u^{-2\beta} \left[\frac{\Gamma(k + 2\beta)}{\Gamma(k)} - \left(\frac{\Gamma(k + \beta)}{\Gamma(k)} \right)^2 \right], \quad (3.3)$$

sendo difícil de analisar, devido ao termo $\beta = 1/d$ no argumento da função gama. Quando $k \gg \beta$ uma expansão da razão $\Gamma(k + \beta)/\Gamma(k)$ como k^β ou $k^\beta e^{-\beta/k}$ leva à inconsistências nos momentos centrados, tal como variância e assimetria nulos. A razão das funções gama, para $z \gg x$, necessita de uma expansão com termos de

mais alta ordem:

$$\frac{\Gamma(z+x)}{\Gamma(z)} \approx z^x \exp\left(\frac{-x}{2z} + \frac{3x^2}{4z}\right). \quad (3.4)$$

De acordo com a Eq. 3.4 a média e variância, Eqs. 3.2 e 3.3, podem ser aproximadas¹ $\langle l_{\rho,d}^{(k)} \rangle \approx N_u^{-\beta} k^\beta$ e $\sigma(l)_{\rho,d,k} \approx c\beta N_u^{-\beta} k^{\beta-1/2}$, com $c = 3/2$, indicando que em alta dimensionalidade a média é pouco afetada pela ordem de vizinhança, enquanto a variância decai muito rapidamente. Esse efeito ocorre pois o volume de uma esfera está quase todo presente em uma casca esférica muito fina quando $d \gg 1$.

A assimetria de uma distribuição é definida como

$$\gamma_1 = \frac{E(X^3) - 3\mu\sigma^2 - \mu^3}{\sigma^3} \quad (3.5)$$

e para a pdf da estatística de distâncias é estabelecida em termos de uma relação não trivial entre α e β

$$\gamma_1 = \frac{2 - \Omega_1^2(k, \beta)/\Omega_2(k, \beta) + \Omega_1^3(k, \beta)/\Omega_3(k, \beta)}{(1 - \Omega_1^2(k, \beta)/\Omega_2(k, \beta))^{3/2}}, \quad (3.6)$$

onde $\Omega_n(k, \beta) = B(k, n\beta)/\Gamma(n\beta)$ e $B(a, b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$ é a função beta. Observe que a simetria da curva é modificada apenas pela ordem de vizinhança e dimensão, sendo independente das bordas do meio e da densidade.

A expressão obtida é complexa de ser analisada, entretanto, de forma aproximada usando a Eq. 3.4, obtemos $\gamma_1 \approx 6\beta k^{-1/2}$, assim como ilustrado no gráfico da Fig. 3.2. Essa simplificação descreve com exatidão o comportamento da assimetria em função da ordem de vizinhança, além de mostrar que há fatorização entre ordem de vizinhança e dimensionalidade.

3.3 Casos Limite

Nessa seção analisamos o comportamento da Eq. 3.1, primeiramente no limite $d \gg 1$, em seguida para $k \gg 1$ e finalmente ambos casos limite. Apesar da serem cálculos simples, há algumas condições/interpretações que serão destacadas.

3.3.1 Alta Dimensionalidade

A nova variável $y = (l - \langle l_{\rho,d}^{(1)} \rangle)/\sigma_{\rho,d,1}$ padroniza a distância por meio da separação média entre os pontos. Conforme $d \gg 1$, $\langle l_{\rho,d}^{(1)} \rangle \approx N_u^{-\beta}$ e $\sigma_{\rho,d,1}(l) \approx c\beta N_u^{-\beta}$

¹Como $x = 1/d, 2/d \ll k$ realizamos a expansão da função exponencial em série de Taylor.

Figura 3.2 – Assimetria da distribuição de distâncias. A simulação foi realizada com $d = 2$, $\rho = 65365$ e condições periódicas de contorno. A aproximação $\gamma_1 = 6\beta k^{-1/2}$ descreve bem o decaimento da assimetria em função da ordem de vizinhança, contudo o efeito de borda, tamanho finito e principalmente de baixa dimensionalidade fazem com que a simulação se distancie do valor real e do aproximado para $k \geq 10$.

e a distância entre os pontos pode ser escrita como

$$l = N_u^{-\beta}(1 + \beta cy) \quad (3.7)$$

com $c = 3/2$ e $\beta = 1/d$. Na variável y , a forma da *pdf* é obtida por meio da lei de transformação de probabilidades, utilizando a Eq. 3.1, com $k = 1$ e $d \rightarrow \infty$, encontramos a distribuição Gumbel Eq. B.64, com parâmetro $\bar{\lambda} = -1/c$,

$$g(y) = c \exp[cy - \exp(cy)], \quad (3.8)$$

como o parâmetro $\bar{\lambda}$ é negativo, ela descreve o mínimo desvio da média esperada: $\langle l_{\rho,d}^{(1)} \rangle = N_u^{-\beta}$. Para ordens de vizinhança superiores

$$g^{(k)}(y) = \frac{c}{\Gamma(k)} \exp[cky - \exp(cy)], \quad (3.9)$$

que é a distribuição log-gamma Eq. B.56, com $\alpha = k$, $\lambda = 1/c$ e $\nu = 0$. A separação média entre pontos é calculada em duas partes. O valor médio da pdf na Eq. 3.9 é $\langle y \rangle = \Psi(k)/c$, que é a função digama A.28, definida como a derivada do logaritmo natural da função gama com relação ao seu argumento [61], então a separação média

entre pontos é $\langle l_{\rho,d}^{(k)} \rangle = N_u^{-\beta}(1 + \beta \langle y \rangle)$. A ordem de vizinhança é um inteiro, que permite a representação especial $\Psi(k) = -\gamma + \sum_{i=1}^{k-1} i^{-1}$, reescrevendo i^{-1} como $(k - i)^{-1}$, torna a distância média em l :

$$\langle l_{\rho,d}^{(k)} \rangle = N_u^{-\beta} \left[1 + \beta \left(-\gamma + \sum_{i=1}^{k-1} \frac{1}{k-i} \right) \right], \quad (3.10)$$

onde $\gamma = 0.57721 \dots$ é a constante de Euler-Mascheroni. Para $k \gg 1$, $\Psi(k) \approx \ln(k)$ e $\langle l_{\rho,d}^{(k)} \rangle = N_u^{-\beta}(1 + \beta \ln(k))$, esse resultado foi obtido primeiramente por Cerf *et. al.* por meio da expansão $\Gamma(k + \beta)/\Gamma(k)$ da Eq. 3.2. Esse fator representa, em média, quanto a distância aumenta em função da ordem de vizinhança, quando a dimensão é mantida fixa. Além disso, a Eq. 3.9 permite o cálculo da variância e momentos de mais alta ordem.

A variância vale $\sigma_y^2 = \Psi^{(1)}(k)$, onde $\Psi^{(1)}(k)$ é a função trigama de k . Lembrando que $\sigma^2(a + bx) = b^2 \sigma^2(x)$, o desvio padrão na variável l é escrito como

$$\sigma(l)_{\rho,d,k} = \frac{\beta N_u^{-\beta}}{\sqrt{k}}, \quad (3.11)$$

onde utiliza-se a aproximação $\Psi^{(1)}(k) \approx 1/k$ para $k \gg 1$. Na variável l a média é muito pouco afetada pela ordem de vizinhança, além da variância cair muito rapidamente. Esse efeito ocorre pois, em alta dimensionalidade, um pequeno incremento no raio leva a um grande aumento de volume e, quanto maior o raio, menor é o incremento para gerar o mesmo aumento de volume. Por isso, quanto maior o *rank* de vizinhança, menor é aumento do raio e menor é a variância em torno do valor médio. Desse modo a distribuição de distância na variável l pode ser descrita como uma sequência delta.

3.3.2 Vizinho Distante

O segundo caso limite é a distribuição de distâncias para alta ordem de vizinhança. O desvio padrão com respeito à ordem de vizinhança cai com $k^{\beta-1/2}$ para $k \gg \beta$. De acordo com a Ref. [48], a soma S , de N variáveis aleatórias independentes e identicamente distribuídas, deve apresentar o desvio padrão relativo, $\sigma_r = \sigma/\langle S \rangle$ e assimetria decaindo com $1/\sqrt{N}$ para indicar convergência para a distribuição Gaussiana. No limite $k \gg 1$, para qualquer dimensão, o desvio padrão

relativo e a assimetria decaem² com $1/\sqrt{k}$. Isso indica que, além de recuperar a simetria da curva, o aumento da ordem de vizinhança faz com que a pdf da Eq. 3.1 se aproxime da distribuição normal. Esse comportamento é obtido por meio de simulação numérica e ilustrado nos gráficos das Figs. 3.3 e 3.1.

A convergência para a Gaussiana pode ser entendida como a soma volumes. O volume necessário para encontrar o k -ésimo vizinho é, em geral, kV_1 , onde V_1 é o volume necessário para encontrar um ponto. A espessura dl da casca esférica que tem, em média, um volume V_1 é uma variável aleatória. A distância de um ponto até seu k -ésimo vizinho é um somatório dessas variáveis aleatórias, o que indica, para $k \gg 1$, sua convergência para a pdf Gaussiana devido ao teorema central do limite.

Figura 3.3 – (a) Aproximação Gaussiana para a estatística de distância para $k \gg 1$ em um meio bidimensional. Os parâmetros da simulação são $d = 2$ e $\rho = 65365$ e $k = 10$. O ajuste pobre nas caudas é devido ao fato que o teorema central do limite garante convergência próximo à média, sendo a convergência da cauda mais lenta. Gráfico adaptado da Ref. [25]. (b) Razão σ/μ para a mesma simulação. O ajuste dado pela aproximação $c\beta k^{-1/2}$ descreve exatamente o comportamento em função da ordem de vizinhança, a simulação apresenta boa concordância com o esperado, contudo o efeito de borda, tamanho finito e a baixa dimensionalidade utilizada ($d = 2$) faz a simulação desviar do valor esperado.

²Esses cálculos são realizados a partir do uso da expansão da Eq. 3.4 no cálculo do desvio padrão e assimetria. A aproximação, Eq. 3.4, obtida para $k \gg 1$ e $d \gg 1$ ainda é válida quando a dimensão é baixa e $k \gg 1$, entretanto, ela é menos acurada.

3.4 Aplicações

Nesta seção, discutimos possíveis aplicações da estatística de distâncias no contexto de geração de números pseudo aleatórios, testes que detectam violação da hipótese Poissônica de distribuição dos pontos. A Tab. 3.1 enumera várias funções densidade de probabilidade que surgem como casos particulares da Eq. 3.1 ao variar a dimensão e ordem de vizinhança.

Devido ao grande número de casos especiais uma possível aplicação é utilizar a pdf da estatística de distâncias como um gerador de números pseudo aleatórios muito geral. Esse gerador, apesar de não ser eficiente em termos de tempo, permite visualizar como as funções densidade de probabilidade surgem a partir de medidas de distância em meios aleatórios.

Tabela 3.1 – Resumo das distribuições de probabilidade para diferentes dimensionalidades e ordens de vizinhança. Aqui, o símbolo (-) significa valor arbitrário e ∞ é um valor muito grande e (*) significa distribuição na variável aleatória y , Eq 3.7.

d	k	Distribuição
1	1	Exponencial
1	-	Gama
1	∞	Normal
2	1	Rayleigh
2	-	Nakagami
3	-	Wilson-Hilferty
-	1	Weibull
-	-	Stacy
-	∞	Normal
∞	1	*Gumbel
∞	-	*Log-Gama
∞	∞	*Normal

Outra possível aplicação da estatística de distâncias é avaliar se as distâncias entre pontos varia da hipótese Poissônica. Essa avaliação foi empregada inicialmente por Thompson [43], no contexto de distribuição de distância entre árvores em um ambiente bidimensional. Uma forma de avaliar desvios da hipótese Poissônica de distribuição dos pontos é realizar um teste de significância para a distância média até o k -ésimo vizinho. O teste utiliza os limites dados pela própria pdf da estatística de distância, quando ela é transformada em uma distribuição χ^2 (qui quadrado). Como generalização do resultado obtido por Thompson, propomos o mesmo teste em um ambiente de dimensionalidade arbitrária. A Eq. 3.1 escrita em termos da variável $x_n = 2N_u l^d$ torna-se:

$$f^{(k)}(x_n) = \frac{1}{2\Gamma(k)} \left(\frac{x_n}{2}\right)^{k-1} \exp(-x_n/2), \quad (3.12)$$

que é a distribuição do χ^2 B.47, com $2k$ graus de liberdade. Uma vez conhecida a densidade de pontos do meio, é possível aplicar o teste e detectar desvios da hipótese Poissônica para qualquer ordem de vizinhança.

3.5 Conclusão

Neste capítulo, usando apenas o processo Poissônico espacial nós calculamos a distribuição de distância de um ponto ao seu k -ésimo vizinho mais próximo em um ambiente d -dimensional. Nossos resultados foram validados com simulação numérica Monte Carlo por meio de um código desenvolvido em linguagem C e paralelizado em MPI. Iniciando com a Eq. 3.1 calculamos os casos limite de alta dimensionalidade e alta ordem de vizinhança. A estatística de distância em alta dimensionalidade torna-se uma sequência delta em torno da média, resultado que foi primeiro conjecturado por by Cerf *et. al.* [42]. A estatística de distância para alta ordem de vizinhança converge para uma distribuição Gaussiana. A pdf geral para $d < 3$ e vizinhança arbitrária consiste de casos especiais que recuperam distribuições bem conhecidas, tal como Gama, Weibull, etc. A estatística de distância pode detectar desvios da hipótese Poissônica de distribuição dos pontos, como proposto por Thomson para $d = 2$, e generalizado pela Eq. 3.12, o que permite analisar distribuição de células em imagens em 3 dimensões.

Caminhadas Aleatórias na Representação de Polímeros Ideais

Neste capítulo, utilizamos modelos de caminhada aleatória para introduzir grandezas que podem representar propriedades conformacionais de polímeros lineares, mais especificamente, distância ponta-a-ponta quadrática, raio quadrático de giração, comprimento de contorno e comprimento de persistência. Os modelos de caminhada utilizados para representar polímeros ideais não levam em conta interação entre um monômero e os monômeros não adjacentes e entre monômero-solvente. Nosso principal objetivo é calcular como as grandezas conformacionais escalam em função do número de passos ao considerar tipos diferentes de caminhadas.

Esse capítulo está organizado da seguinte maneira. Na Sec. 4.1 apresentamos a caminhada aleatória. Na Sec. 4.2 descrevemos a caminhada que representa o modelo de cadeia de rotação livre. Na Sec. 4.3 apresentamos as grandezas homônimas relacionadas e conhecidas como comprimento de persistência. Na Sec. 4.4 calculamos de maneira simplificada os resultados o cálculo do expoente de Flory. Ao final, na Sec. 4.5, apresentamos as conclusões e motivações para estudar a caminhada aleatória autorrepulsiva (SAW) no próximo capítulo.

4.1 Caminhada Aleatória

Em sua forma mais simples, a caminhada aleatória, do inglês *random walk* (RW), em rede regular de espaçamento ℓ consiste de um caminhante que em cada passo escolhe, aleatoriamente e com igual probabilidade, um sítio adjacente

para ser visitado. Nesse modelo de caminhada, não há restrição quanto ao número e frequência de visitas a qualquer sítio da rede. Neste capítulo, as grandezas conformacionais são derivadas com base em propriedades geométricas, uma abordagem mais sofisticada em termos de função características, transformada de Fourier, dentre outras é apresentada no apêndice C.

Após N passos, um caminhante percorre $\vec{R}_N = \vec{u}_1 + \vec{u}_2 + \cdots + \vec{u}_N$, a distância quadrática ponta-a-ponta pode ser escrita como:

$$\vec{R}_N^2 = (\vec{R}_{N-1} + \vec{u}_N)^2, \quad (4.1)$$

onde $\vec{u}_N$ é o vetor percorrido no N -ésimo passo e $|\vec{u}_N| = \ell$. Antes de prosseguir, definiremos a notação utilizada, neste capítulo e no próximo, para expressar a média das grandezas conformacionais estudadas.

Há várias configurações distintas para uma caminhada aleatória com N passos. Em cada configuração uma grandeza conformacional $\mathcal{O}_j$ assume um valor, em um grande número de repetições de uma caminhada com N passos é conveniente tratar $\langle \mathcal{O}_j \rangle_N$, conhecida como média configuracional. O subíndice N de $\langle \rangle_N$ indica o *ensemble* de configurações relativo ao número de passos onde $\mathcal{O}_j$ é amostrado, enquanto o subíndice j de $\mathcal{O}_j$ indica em que passo da caminhada $\mathcal{O}_j$ é medido. Esse último subíndice pode variar de 1 até N . Em termos dessa notação, a média configuracional da distância quadrática média pode ser escrita em termos da expansão da Eq. 4.1 e vale:

$$\langle \vec{R}_N^2 \rangle_N = \langle \vec{R}_{N-1}^2 \rangle_N + 2\langle \vec{R}_{N-1} \cdot \vec{u}_N \rangle_N + \ell^2. \quad (4.2)$$

Na caminhada aleatória, qualquer passo u_i é independente dos demais e todas as direções e sentidos são igualmente prováveis, assim $\langle \vec{R}_{N-1} \cdot \vec{u}_N \rangle_N = \langle \vec{R}_{N-1} \rangle_N \cdot \langle \vec{u}_N \rangle_N$, que é a definição de variáveis aleatórias independentes. O produto entre $\langle \vec{R}_{N-1} \rangle_N \cdot \langle \vec{u}_N \rangle_N = 0$ pois, em média, a projeção de um passo i na direção do passo j é nula $\langle \vec{u}_i \cdot \vec{u}_j \rangle_N = 0$ para $i \neq j$.

Como não há restrição de visitação, medir a distância quadrática ao penúltimo passo $\vec{R}_{N-1}^2$ é invariante em relação ao *ensemble* número de passos, N ou $N-1$, assim vale a igualdade $\langle \vec{R}_{N-1}^2 \rangle_{N-1} = \langle \vec{R}_{N-1}^2 \rangle_N$. A Eq. 4.2 pode ser reescrita como:

$$\langle \vec{R}_N^2 \rangle_N = \langle \vec{R}_{N-1}^2 \rangle_{N-1} + \ell^2. \quad (4.3)$$

Esse processo pode ser repetido para $\langle \vec{R}_{N-1}^2 \rangle_{N-1}$ e assim por diante, de modo que a distância ponta-a-ponta quadrática média vale

$$\langle \vec{R}_N^2 \rangle_N = N\ell^2. \quad (4.4)$$

O raio de giração quadrático é definido como o somatório ao quadrado da distância entre as coordenadas do sítio i e do centro de massa:

$$\langle \vec{R}_g^2 \rangle_N = \frac{1}{N} \sum_{i=1}^N \langle (\vec{R}_i - \vec{R}_{cm})^2 \rangle_N = \frac{1}{2N^2} \sum_{i=1}^N \sum_{j=1}^N \langle (\vec{R}_i - \vec{R}_j)^2 \rangle_N, \quad (4.5)$$

onde $\vec{R}_{cm} = N^{-1} \sum_{i=1}^N \vec{u}_i$ é a posição do centro de massa. Para a RW o raio de giração vale

$$\langle \vec{R}_g^2 \rangle_N = \frac{(N+2)\langle \vec{R}_N^2 \rangle_N}{6(N+1)}. \quad (4.6)$$

A caminhada aleatória é o modelo mais simples para a representação de um polímero linear em conformação *random coil*. Entretanto, a ausência de restrições quanto ao ângulo entre os passos, por exemplo, a tornam pouco robusta, pois cadeias reais podem apresentar esse tipo de restrição [62].

4.2 Cadeia de Rotação Livre

Esse modelo é uma variante da caminhada aleatória que leva em conta a restrição do ângulo entre passos adjacentes. Esse ângulo é fixo e vale θ . Um caminhante pode adicionalmente rotacionar o novo passo em relação ao anterior por qualquer ângulo φ , assim como ilustrado na Fig. 4.1. Apesar do comprimento do passo ser discreto, a região na qual a caminhada ocorre é contínua, devido ao ângulo de rotação $0 \leq \varphi \leq 2\pi$. A média configuracional da distância quadrática ponta-a-ponta é expressa por meio do duplo somatório:

$$\langle \vec{R}_N^2 \rangle_N = \sum_{i=1}^N \sum_{j=1}^N \langle \vec{u}_i \cdot \vec{u}_j \rangle = N\ell^2 + 2\ell^2 \sum_{i=1}^N \sum_{j>i}^N \langle \cos(\theta_{ij}) \rangle, \quad (4.7)$$

onde θ_{ij} é o ângulo entre os passos i e j . O cálculo de $\langle \cos(\theta_{ij}) \rangle$ consiste em determinar o produto escalar entre o passo $\vec{u}_j$ e o passo $\vec{u}_i$. Para $j = i+1$, $\langle \vec{u}_i \cdot \vec{u}_j \rangle = \ell^2 \cos(\theta)$, dado que a componente transversal introduzida pelo ângulo de rotação φ se cancela

Figura 4.1 – Representação da cadeia de rotação livre por meio de uma caminhada com o ângulo de ligação fixo θ em qualquer orientação determinada pelo ângulo de rotação φ .

(ver Fig. 4.1). Para $j = i + 2$ devemos calcular primeiramente a projeção em $i + 1$: $\langle \vec{u}_{i+1} \cdot \vec{u}_{i+2} \rangle = \ell \cos(\theta)$ e depois o produto escalar entre $\vec{u}_i$ com essa projeção, o que resulta em $\langle \vec{u}_i \cdot \vec{u}_{i+2} \rangle = \ell^2 \cos^2(\theta)$. Pelo processo de indução temos $\cos(\theta_{ij}) = \cos^{j-i}(\theta) = \alpha^{j-i}$, onde $\alpha = \cos(\theta)$, e a distância quadrática ponta-a-ponta é escrita como

$$\langle \vec{R}_N^2 \rangle_N = N\ell^2 + 2\ell^2 \sum_{i=1}^N \sum_{j>i}^N \alpha^{j-i}. \quad (4.8)$$

Para valores pequenos de N , o somatório da Eq. 4.8 é facilmente calculado, para $N = 2$: α , para $N = 3$: $2\alpha + \alpha^2$, para $N = 4$: $3\alpha + 2\alpha^2 + \alpha^3$ e assim por diante. O somatório da Eq. 4.8 pode ser reescrito, sem perda de generalidade, de modo que

$$\langle \vec{R}_N^2 \rangle = N\ell^2 + 2\ell^2 \sum_{k=1}^{N-1} (N-k)\alpha^k = N\ell^2 \left[\frac{1+\alpha}{1-\alpha} - \frac{2\alpha(1-\alpha^N)}{N(1-\alpha)^2} \right]. \quad (4.9)$$

Para valores grandes de N , a Eq. 4.9 pode ser reescrita como

$$\langle \vec{R}_N^2 \rangle = N\ell^2 \left[\frac{1+\cos(\theta)}{1-\cos(\theta)} \right]. \quad (4.10)$$

As duas parcelas que mais contribuem com o valor do raio de giração são:

$$\langle \vec{R}_g^2 \rangle_N = \frac{\langle \vec{R}_N^2 \rangle_N (N+2)[1+\cos(\theta)]}{6(N+1)[1-\cos(\theta)]} - \frac{\langle \vec{R}_N^2 \rangle_N \cos(\theta)}{(N+1)[1-\cos(\theta)]^2}, \quad (4.11)$$

as demais parcelas da Eq. 4.11 podem ser encontradas na Ref. [63].

A Eq. 4.10 apresenta o mesmo crescimento da caminhada aleatória, Eq. 4.4, a não ser pela razão de funções cosseno. No contexto da Física de Polímeros, essa razão que multiplica $N\ell^2$ é conhecida como raio característico, que é definido como

$$C_\infty = \lim_{N \rightarrow \infty} (\langle \vec{R}_N^2 \rangle_N / N\ell^2). \quad (4.12)$$

Para caminhada aleatória, $C_\infty = 1$, enquanto na caminhada de rotação livre ele vale $C_\infty = [1 + \cos(\theta)]/[1 - \cos(\theta)]$.

Na caminhada aleatória o produto escalar $\langle \vec{u}_i \cdot \vec{u}_j \rangle = 0$ para $i \neq j$, pois os passos estão aleatoriamente orientados de acordo com uma função densidade de probabilidade uniforme. As variantes da caminhada aleatória utilizadas para representar um polímero ideal¹ têm em comum o comportamento $\langle \vec{u}_i \cdot \vec{u}_j \rangle \rightarrow 0$ para $|i - j| \gg 1$ com valor de $C_\infty > 1$ e finito [30]. Tal comportamento é resultante de restrições no ângulo entre os passos, assim como na caminhada de rotação livre.

É possível representar os modelos de caminhada, com C_∞ finito, como uma caminhada aleatória com número de passos e comprimento do passo distintos da caminhada original, porém com as mesmas medidas conformacionais, tal como a distância quadrática ponta-a-ponta e comprimento de contorno. Nessa caminhada aleatória, que representa os modelos de caminhada com C_∞ finito, há N_e passos com comprimento ℓ_e formado pelo agrupamento dos N passos da caminhada original de comprimento ℓ . Para determinar o valor de ℓ_e e N_e em função de ℓ e N da caminhada original é necessário estabelecer duas relações entre as grandezas conformacionais.

A primeira relação é entre as distâncias quadráticas médias. Na caminhada original, quando o valor de $\langle \vec{R}_N^2 \rangle_N / N\ell^2$ atinge o valor do raio característico C_∞ , temos $\langle \vec{R}_N^2 \rangle_N = C_\infty N\ell^2$, enquanto que na caminhada aleatória $\langle \vec{R}_{N_e}^2 \rangle_{N_e} = N_e \ell_e^2$. Vale, portanto, a igualdade $N_e \ell_e^2 = C_\infty N\ell^2$. A segunda relação é proveniente de uma grandeza conformacional conhecida como comprimento de contorno R_{max} . Essa grandeza parametriza o tamanho das cadeias poliméricas em experimentos [64, 65]. Ela é definida como o tamanho do polímero ao ser distendido até seu comprimento máximo [66] e a Fig. 4.2 ilustra essa grandeza. Na caminhada aleatória o comprimento máximo entre o início e final é $N_e \ell_e$, enquanto na caminhada original é R_{max} . Vale, portanto, a igualdade $R_{max} = N_e \ell_e$. Das duas igualdades, distância quadrática e comprimento de contorno, determinamos o valor de N_e e ℓ_e , Eqs. 4.14 e 4.13,

$$\ell_e = \frac{C_\infty N \ell^2}{R_{max}}, \quad (4.13)$$

$$N_e = \frac{R_{max}^2}{C_\infty N \ell^2}, \quad (4.14)$$

---

¹Nas cadeias ideais, não é levada em conta a interação entre monômeros não adjacentes e nem dos monômeros com o solvente.

Figura 4.2 – O comprimento de contorno R_{max} é definido como o comprimento da cadeia quando totalmente esticada. Essa grandeza é fundamental, pois seu valor caracteriza o tamanho dos polímeros em experimentos Físicos/Químicos. Para a cadeia de rotação livre, $R_{max} = N\ell \cos(\theta/2)$.

onde ℓ_e é comumente conhecido como raio de Kuhn [63], uma grandeza importante utilizada na teoria de polímeros.

Além da medida dispersiva, distância ponta-a-ponta quadrática, há a medida de correlação entre dois passos na trajetória do caminhante separados por $s = |j - i|$ passos. Ao contrário da caminhada aleatória, onde não há correlação entre os passos, a caminhada que representa a cadeia de rotação livre tem seus passos correlacionados:

$$[\cos(\theta)]^{|j-i|} = \exp[|j-i| \ln[\cos(\theta)]] = \exp\left[\frac{-|j-i|}{s_p}\right]. \quad (4.15)$$

A constante característica do decaimento dessa correlação é

$$s_p = \frac{-1}{\ln[\cos(\theta)]}. \quad (4.16)$$

como $0 \leq \cos(\theta) \leq 1$, $s_p > 0$. A grandeza $\ell_p = \ell s_p$ é conhecida como *comprimento de persistência do passo*. A Eq. 4.15 pode ser escrita em função de ℓ_p como $\exp[-|j-i|\ell/\ell_p]$.

A última caminhada que pode representar um polímero ideal tratada aqui é aquela onde os passos adjacentes apontam aproximadamente para uma mesma direção, isso equivale a uma cadeia de rotação livre com baixo ângulo de ligação (ângulo θ pequeno). Esse modelo é conhecido como *worm like chain* ou Kratky-Porod [30, 67] e é amplamente utilizado para caracterizar polímeros reais [64, 65].

Na caminhada com o ângulo de rotação livre, o pequeno valor de θ permite realizar a aproximação $\cos(\theta) \approx 1 - \theta^2/2$. Como $\ln(1 - \theta^2/2) \approx -\theta^2/2$, o decaimento da correlação entre dois passos da caminhada é $s_p = 2/\theta^2$, de modo que o projeção

do comprimento ℓ de um passo j na direção do passo i vale

$$\ell_p = \ell_{s_p} = \frac{2\ell}{\theta^2}. \quad (4.17)$$

O raio característico vale

$$C_\infty = \frac{1 + \cos(\theta)}{1 - \cos(\theta)} \approx \frac{2 - \theta^2/2}{\theta^2/2} \approx \frac{4}{\theta^2}. \quad (4.18)$$

Para a caminhada que representa o modelo da cadeia de rotação livre, o comprimento de contorno vale $R_{max} = N\ell \cos(\theta/2) \approx N\ell$ (ver Fig. 4.2 e Ref. [30]). Como o valor de raio característico e comprimento de contorno são conhecidos, a substituição de $R_{max} \approx N\ell$ e $C_\infty \approx 4/\theta^2$ na Eq. 4.13 fornece o comprimento de Kuhn:

$$\ell_e \approx 2\ell_p. \quad (4.19)$$

Note que a combinação ℓ/θ^2 está presente no comprimento de persistência do passo e no comprimento de Kuhn. O modelo Kratky-Porod é definido no limite $\ell \rightarrow 0$ e $\theta \rightarrow 0$ com ℓ_p constante e comprimento de contorno R_{max} constante [30]. A distância quadrática média pode ser calculada usando o decaimento exponencial da correlação entre versores tangentes a pontos ao longo da caminhada e vale

$$\langle \vec{R}^2 \rangle = 2\ell_p R_{max} - 2\ell_p^2 \left[1 - \exp\left(\frac{-R_{max}}{\ell_p}\right) \right]. \quad (4.20)$$

Há dois casos limite da Eq. 4.20 que merecem atenção. O limite de cadeia ideal, representado por uma caminhada aleatória, é recuperado quando $R_{max} \gg \ell_p$ e a Eq. 4.20 reduz-se a $\langle \vec{R}^2 \rangle \approx 2\ell_p R_{max} = \ell_e R_{max}$. O caso limite do bastão rígido² é recuperado quando $R_{max} \ll \ell_p$, expandindo a exponencial da Eq. 4.20 em série de Taylor leva à $\langle \vec{R}^2 \rangle \approx R_{max}^2$. O raio de giracão vale [68]:

$$\langle \vec{R}_g^2 \rangle = \frac{\ell_p R_{max}}{3} - \ell_p^2 + \frac{2\ell_p^3}{R_{max}} - \frac{2\ell_p^4}{R_{max}^2} \left[1 - \exp\left(\frac{-R_{max}}{\ell_p}\right) \right] \quad (4.21)$$

Para resumir os resultados apresentamos a Tab. 4.1 que contém o valor das principais grandezas conformacionais para os modelos de caminhada aqui apresentados.

Nos modelos de caminhada descritos nesse capítulo, foi definido a grandeza *comprimento de persistência do passo* a constante de decaimento da correlação entre dois passos i e j arbitrários. Na próxima seção, apresentamos uma curta revisão sobre a ocorrência/utilização do termo comprimento de persistência, que é empregado para designar grandezas distintas, mas que estão relacionadas entre si.

²A configuração do tipo bastão rígido é definida quando $\langle \vec{R}_N^2 \rangle \propto R_{max}^2$, no caso da caminhada aleatória isso equivale à $\langle \vec{R}_N^2 \rangle_N \propto N^2$.

Tabela 4.1 – Tabela resumo das principais grandezas conformacionais dos modelos de caminhada utilizados para representar cadeias poliméricas reais. Aqui, $f_1(\ell_p, R_{max}) = 1 - \exp(-R_{max}/\ell_p)$ e $f_2(\ell_p, R_{max}) = \ell_p^2(1 - 2\ell_p/R_{max}) + 2\ell_p^4/R_{max}^2(1 - \exp(-R_{max}/\ell_p))$.

Modelo	RW	FRC	Kratky-Porod
$\langle \vec{R}^2 \rangle$	$N\ell^2$	$N\ell^2[1 + \cos(\theta)]/[1 - \cos(\theta)]$	$2\ell_p R_{max} - 2\ell_p^2 f_1(\ell_p, R_{max})$
C_∞	1	$[1 + \cos(\theta)]/[1 - \cos(\theta)]$	$4/\theta^2$
s_p	0	$-\ln^{-1}[\cos(\theta)]$	$2/\theta^2$
$\langle \vec{R}^2 \rangle$	$\langle \vec{R}^2 \rangle/6$	$\langle \vec{R}^2 \rangle[1 + \cos(\theta)]/6[1 - \cos(\theta)]$	$\ell_p R_{max}/3 - f_2(\ell_p, R_{max})$

4.3 Comprimento de Persistência

Por definição [62, 69], o *comprimento de persistência* $\langle x_N \rangle_N$ é a projeção do vetor deslocamento na direção do primeiro passo, no limite $N \gg 1$, matematicamente

$$\langle x_N \rangle_N = \sum_{i=1}^N \langle \hat{u}_1 \cdot \vec{u}_i \rangle_N, \quad (4.22)$$

onde $\hat{u}_1 = \vec{u}_1/|\vec{u}_1|$ é o versor na direção do primeiro passo. A Eq. 4.22 é a definição matemática de comprimento de persistência. A ligação entre a Eq. 4.22 e o *comprimento de persistência do passo* é feita por meio da relação $\langle x_N \rangle_N = \sum_{n=1}^N \langle \vec{u}_1 \cdot \vec{u}_n \rangle_N$ [69], onde a correlação entre os sítios i e j é ponderada pela Eq. 4.15 no caso da caminhada que representa o modelo da cadeia de rotação livre. Para a caminhada aleatória $\langle x_N \rangle_N = \ell$ [62] e para a cadeia de rotação livre $\langle x_N \rangle_N = \ell/[1 - \cos(\theta)]$ [63].

Há uma definição homônima associada ao modelo de Kratky-Porod [30, 65], que é muito utilizado na descrição de resultados de experimentos, com DNA [65], por exemplo. No modelo de Kratky-Porod a correlação entre dois versores tangentes a dois pontos separados pelo comprimento de contorno s decai como $\langle \vec{u}(l) \cdot \vec{u}(l+s) \rangle \propto \langle \cos[\theta(s)] \rangle = \exp(-s/l_p)$, onde l_p é definido como o comprimento de persistência, que é conflitante com a definição de comprimento de persistência dada pela Eq. 4.22. A notação l_p representa a distância característica do decaimento da correlação entre versores tangentes, separados por um comprimento de contorno s ao longo do polímero.

4.4 O Modelo de Flory

Os dois modelos apresentados para a representação do polímero, caminhada aleatória e cadeia de rotação livre, apresentam o comportamento $\langle \vec{R}_N^2 \rangle_N \sim N^{2\nu_0}$, onde $\nu_0 = 1/2$. Experimentos com polímeros em conformação *random coil* mostraram que o tamanho do polímero em função do número de resíduos comporta-se com um expoente $\nu > 1/2$, em duas e três dimensões [11]. Tal comportamento capturado por Flory por meio de sua teoria.

O modelo de Flory considera dois componentes básicos: entropia e probabilidade de interação por exclusão de volume dos monômeros. A entropia $S(\vec{R}_N)$ associada com todas as configurações de uma caminhada começando no sítio $\omega(0)$ e terminando no sítio $\omega(N)$, tal que $|\omega(N) - \omega(0)| = |\vec{R}_N|$, é proporcional ao logaritmo do número de caminhadas distintas $\Gamma_N(\vec{R}_N)$:

$$S(\vec{R}) = k_B \ln[\Gamma_N(\vec{R}_N)], \quad (4.23)$$

onde $k_B = 1.38 \times 10^{23} JK^{-1}$ é a constante de Boltzmann. Se um sítio tem z vizinhos adjacentes, o número de escolhas para o próximo passo é z . Para uma caminhada aleatória em rede regular d -dimensional há $2d$ sítios que podem ser visitados pelo caminhante no próximo passo, de modo que $\sum_{\vec{R}} \Gamma_N(\vec{R}_N) = (2d)^N$.

O vetor ponta-a-ponta $\vec{R}_N$ é a soma de N componentes: $\vec{R}_N = \sum_{n=1}^N \vec{u}_n$. A probabilidade de encontrar um caminhante a uma distância $\vec{R}_N$ após N passos vale:

$$p(\vec{R}_N) = \frac{\Gamma_N(\vec{R}_N)}{\sum_{\vec{R}_N} \Gamma_N(\vec{R}_N)}. \quad (4.24)$$

Para caminhadas aleatórias não viciadas em d dimensões [62] a probabilidade de encontrar um caminhante em um ponto com coordenadas x, y, z é:

$$p(x, y, z) = \left(\frac{d}{2\pi N \ell^2} \right)^{d/2} \exp \left(-\frac{dR_N^2}{2N \ell^2} \right). \quad (4.25)$$

Para simplificar, chamamos $c = (d/2\pi N \ell^2)^{d/2}$. O número de caminhadas com N passos que terminam a uma dada distância $\vec{R}_N$ da origem é

$$\Gamma_N(\vec{R}_N) = c[2d]^N \exp \left(\frac{-dR_N^2}{2N \ell^2} \right), \quad (4.26)$$

a entropia

$$S(\vec{R}_N) = S(0) - k_B \frac{dR^2}{2N\ell^2} \quad (4.27)$$

e a energia livre de Helmholtz

$$F(\vec{R}_N) = F(0) + \frac{dT R_N^2}{2N\ell^2}, \quad (4.28)$$

onde $F(0) = U(0) - TS(0)$. A distância ponta-a-ponta da cadeia, na situação de equilíbrio, é resultado da competição entre o efeito de volume excluído, que tende a expandir a cadeia, e o efeito de perda configuracional que tende a contrair uma cadeia, pois para um mesmo N , se $\langle \vec{R}_N \rangle_N$ for muito grande poucas configurações são acessíveis.

A contribuição energética devido ao efeito de volume excluído custa, por interação $k_B T$. Para um dado monômero, a probabilidade de haver uma interação por exclusão de volume com outro monômero é $v_{exc} N / R^d$, onde N / R^d é densidade média de monômeros e v_{exc} é o volume ocupado por um dado monômero. Para N monômeros a contribuição de exclusão vale:

$$F_{exc} \approx k_B T v_{exc} \frac{N^2}{R_N^d}. \quad (4.29)$$

Aqui está a aproximação de campo médio. Todos os monômeros têm, em média, a mesma probabilidade de interação com os demais. Essa aproximação superestima as interações por exclusão de volume. A contribuição devido perda conformacional, F_{ent} :

$$F_{ent} \approx k_B T \frac{R_N^2}{N\ell^2}. \quad (4.30)$$

Estamos omitindo constantes multiplicativas, pois seu efeito prático no cálculo do expoente é irrelevante. A energia livre total da cadeia real, segundo Flory, é escrita como a soma de F_{exc} e F_{ent} :

$$F_T \approx k_B T \left(v_{exc} \frac{N^2}{R^d} + \frac{R^2}{N\ell^2} \right). \quad (4.31)$$

Ao atingir o tamanho na situação de equilíbrio, a variação da energia livre é nula, $\partial F_T / \partial R_N = 0$, de modo que $R_N \propto N^{3/(d+2)}$, ou seja, o expoente de Flory:

$$\nu_{Flory} = \frac{3}{d+2}. \quad (4.32)$$

Para $2d$, $\nu_{Flory} = 3/4$ e para $3d$, $\nu_{Flory} = 3/5$ que são bons resultados [11], no entanto a teoria de Flory está fundamentada em duas contribuições superestimadas. O efeito da exclusão de volume é superestimado, pois a probabilidade de haver uma interação desse tipo deve decair com o aumento da distância entre monômeros, ao invés de ser constante. Ao mesmo tempo, a contribuição entrópica é superestimada, pois todos os sítios da rede foram assumidos como passíveis de visitação mais de uma vez.

Flory, em sua argumentação, busca um modo simplificado de introduzir o efeito de volume excluído com o finalidade de descrever melhor cadeias poliméricas reais. Baseado nessa argumentação, é possível construir um modelo simplificado de caminhada, conhecido como *self-avoiding walk* (SAW), que captura a essência das ideias propostas por Flory ao levar em conta a exclusão de volume.

4.5 Conclusão

Neste capítulo, introduzimos modelos simples de caminhadas aleatórias motivados pela representação conformacional de polímeros ideais. Iniciamos o capítulo com o modelo mais simples, caminhada aleatória, evoluindo para a caminhada com ângulo de ligação fixos e ângulos de rotação livre, e o caso particular deste último modelo no qual os ângulos de ligação são pequenos. Por último, apresentamos a abordagem tipo campo médio usada por Flory para incluir o efeito de volume excluído. Essa abordagem é a motivação para o próximo capítulo, cujo objeto de estudo é a *self-avoiding walk* (SAW).

Origem da Regra de Escala para a Caminhada Aleatória Autorrepulsiva

---

A caminhada aleatória autorrepulsiva, em inglês *self-avoiding walk* (SAW), consiste de uma caminhada aleatória, com N passos, que não intercepta a própria trajetória. No presente capítulo, investigamos o comportamento em função de N de grandezas conformacionais conhecidas, tais como a distância ponta-a-ponta quadrática média, que mede a distância ao quadrado da origem até o final da caminhada e o comprimento de persistência $\langle x_N \rangle_N$, definido como a projeção do vetor posição na direção do primeiro passo.

Nossa contribuição consiste em estudar $\langle \vec{R}_N^2 \rangle_N$ e $\langle x_N \rangle_N$ de forma inovadora, onde essas grandezas são escritas em função de $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$, o produto escalar do vetor posição com o vetor deslocamento no j -ésimo passo¹. Demonstramos analiticamente que o comprimento de persistência é o caso particular de $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$ quando $j = N$. Por meio da análise numérica de $\langle \vec{R}_N \cdot \vec{u}_N \rangle_N$, determinamos que o comprimento de persistência converge para uma constante nas redes quadrada e cúbica.

Escrevemos $\langle \vec{R}_N^2 \rangle_N$ como somatório de $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$. Nossa análise dos dados de simulação numérica Monte Carlo (MC) mostram dois comportamentos distintos do produto escalar intermediário. Para $1 \leq j \leq \Theta(N)$ determinamos a regra de escala $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N \sim 1/2 + \alpha_0(j - \tau)^{2\nu_0 - 1}$, onde τ é conhecida como constante de suavização e $\nu_0 > 1/2$ para rede quadrada e cúbica. No intervalo $\Theta(N) \leq j \leq N$ a regra de

---

¹No decorrer do texto nos referimos à $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$ como *produto escalar intermediário*.

escala anterior não é suficiente para descrever os dados, pois a partir desse passo o produto escalar médio $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$ cresce mais lentamente até atingir um ponto de máximo, decrescendo monotonamente até $j = N$. Em princípio são necessárias mais parcelas, termos de correção, para descrever o comportamento do produto escalar intermediário nessa última região.

No conjunto completo de caminhadas com N passos determinamos $\Theta(N) \sim N/2$ e $\Theta(N) \sim N/3$ para as redes quadrada e cúbica, respectivamente. A partir desse valor de $\Theta(N)$ a caminhada torna-se mais flexível. Qualitativamente esse aumento da flexibilidade se deve à maior liberdade de movimentação média que o final da caminhada experimenta em relação à parte inicial, pois há mais graus de liberdade acessíveis para os últimos passos.

Essa análise do produto escalar intermediário é muito robusta, mesmo para caminhadas com poucos passos, $N \leq 100$, já obtivemos convergência dos resultados, o que permitiu a determinação do expoente principal ν_0 com boa precisão e o colapso de dados de $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$.

Esse capítulo está organizado da seguinte forma. Na Sec. 5.1 apresentamos a introdução e revisão bibliográfica. Na Sec. 5.2 apresentamos as principais grandezas conformacionais estudadas e os métodos numéricos de simulação Monte Carlo para a SAW. Na Sec. 5.3 descrevemos a abordagem via produtos escalares intermediários. Na Sec. 5.4 abordamos o comprimento de persistência. Na Sec. 5.5 calculamos a regra de escala associada ao produto escalar intermediário. Por último, na Sec. 5.6 apresentamos as conclusões.

5.1 Introdução e Revisão Bibliográfica

No contexto de caminhadas aleatórias, a representação mais simples de um polímero linear em conformação *random coil* é dada por uma caminhada aleatória (RW) [2, 3] em redes regulares. A simplicidade dessa caminhada permite a existência do limite de escala [70], o movimento Browniano. Contudo, a falta de restrição na visitação permite um sítio ser visitado diversas vezes, assim como ilustrado na Fig. 5.1a. A ausência de restrição implica que, na representação do polímero, mais de um monômero pode estar presente no mesmo local. Para corrigir esse problema, um

modelo conhecido como caminhada aleatória autorrepulsiva, em inglês *self-avoiding walk* (SAW), pode ser utilizado [64, 71, 72]. Nesse modelo, as caminhadas podem ser geradas com a mesma dinâmica da caminhada aleatória, contudo o caminhante não pode visitar um dado sítio da rede mais que uma vez.

Uma SAW em rede regular, ilustrada na Fig. 5.1c, pode ser representada por um grafo direcionado onde os sítios da rede equivalem aos vértices e as ligações entre sítios adjacentes às arestas. Nessa representação a SAW pode ser descrita por uma coleção de vértices distintos $\omega = \omega_0, \omega_1, \omega_2, \dots, \omega_N$ ligados por arestas. O número de arestas que chegam ou saem de um vértice nunca pode ser maior que 1, pois isso viola a condição de autorrepulsividade da caminhada. O número de passos da caminhada é determinado pela quantidade de arestas N .

Apesar da simplicidade, assim como a caminhada aleatória, poucos resultados analíticos da SAW são demonstrados rigorosamente. Em parte, a dificuldade em obtê-los é inerente à característica não-Markoviana² da evolução da caminhada, dado que o caminhante deve lembrar todos os passos anteriores para que no próximo passo a condição de autorrepulsividade seja satisfeita. Técnicas numéricas, em particular simulação de Monte Carlo, são os principais meios de investigação da SAW. Além disso o mapeamento da caminhada em modelos de cadeia de *spin* [11, 72], tal como o n -vetorial com $n = 0$, e a dinâmica estocástica de Laeuer com $\kappa = 8/3$ [73], também permite o estudo de expoentes críticos e amplitudes universais. Enfatizamos que está fora do escopo dessa dissertação descrever a SAW no contexto de modelos de cadeia de *spin*, tal como o n -vetorial, ou da dinâmica estocástica de Laeuer.

No que concerne às técnicas numéricas, o estudo de séries, tal como a função geradora da caminhada³, permite estimativas precisas de expoentes críticos e constantes universais. Contudo, elas necessitam do conhecimento de c_N , o número total de caminhadas distintas com N passos, ou a distribuição das posições finais das caminhadas com um determinado número de passos. Essas técnicas são susceptíveis à enumeração exata. Atualmente o número c_N de caminhadas distintas com N passos é conhecido até $N = 79$ [74] para rede quadrada e $N = 36$ para a rede cúbica [75].

Um dos métodos de simulação Monte Carlo mais simples para gerar uma

²A evolução da caminhada aleatória pode ser descrita por um processo Markoviano, pois o próximo sítio a ser visitado não é afetado pelo histórico de visitação do caminhante.

³A função geradora é resultado do mapeamento da SAW no modelo n -vetorial.

SAW é por meio da caminhada aleatória não reversa, Fig. 5.1b. O caminhante parte de um sítio $\omega(0)$ e escolhe aleatoriamente um sítio adjacente $\omega(1)$ para ser visitado. Esse processo é repetido, contudo é proibido ao caminhante ir do sítio atual, $\omega(1)$, para o anterior $\omega(0)$. A dinâmica evolui segundo essa regra até que o caminhante tente ir a um sítio previamente visitado, o que implica impossibilidade de avanço, a caminhada é então finalizada para que a condição de autorrepulsão exista. O problema com a dinâmica de movimentação acima é que a probabilidade do caminhante não interceptar sua própria trajetória decai exponencialmente em função do número de passos N .

Figura 5.1 – Caminhadas aleatórias distintas, o losango aberto indica o início da caminhada e o fechado seu final. (a) Caminhada aleatória. (b) Caminhada aleatória não reversa, onde o caminhante que foi do sítio $\omega(i - 1) \rightarrow \omega(i)$ no passo i não pode, no passo $i + 1$, fazer o caminho inverso $\omega(i) \rightarrow \omega(i - 1)$. (c) Caminhada autorrepulsiva (SAW). (d) Caminhada autorrepulsiva armadilhada, o caminhante não tem mais para onde ir e a caminhada é finalizada.

Métodos de simulação MC mais sofisticados do que o mencionado anteriormente consistem em técnicas computacionais que permitem amostrar caminhadas mais longas. Dentre as técnicas destaca-se o algoritmo de pivotamento [17, 19]. Esse algoritmo consiste em gerar novas caminhadas a partir de rotações e reflexões em partes de caminhadas já existentes. Atenção deve ser dispendida nessa construção, pois há correlações entre as diversas caminhadas geradas por operações diedrais sucessivas. Essa abordagem permite aumentar o número de passos ($N \sim 10^7$ [18, 28]) quando associada à estruturas de dados eficientes e consequentemente complexas do ponto de vista computacional.

O estudo da lei de escala para grandezas conformacionais via simulação numérica MC baseia-se em duas abordagens. A primeira estuda o comportamento da função correlação $\langle \vec{u}_i \cdot \vec{u}_j \rangle_N$ entre os passos i e j [76, 77]. A segunda consiste em usar apenas o sítio final e inicial [28, 69], tal como na distância ponta-a-ponta quadrática.

Essas abordagens constituem dois extremos. Por um lado, $\langle \vec{u}_i \cdot \vec{u}_j \rangle_N$ é a origem microscópica da regra de escala, mas é difícil de ser estudada pois a convergência dos expoentes, principal e de correção, é lenta. Por outro lado, $\langle \vec{R}_N^2 \rangle_N = \langle |\omega(N) - \omega(0)|^2 \rangle_N$ considera apenas o sítio inicial e final, deixando de lado os detalhes da trajetória do caminhante.

Nossa proposta é tratar as grandezas conformacionais, distância ponta-a-ponta quadrática e comprimento de persistência, de uma maneira distinta do que é abordado na literatura (que conhecemos) e consiste em estudar essas grandezas em função dos produtos escalares intermediários médios $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$ entre o vetor posição e o vetor deslocamento no j -ésimo passo. Por meio da grandeza $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$ relacionamos, simultaneamente, a distância ponta-a-ponta quadrática média com o comprimento de persistência. Nossos resultados analíticos são validados por meio de simulação numérica MC. O foco desse estudo não é o algoritmo de simulação numérica MC, mas a nova maneira de analisar as grandezas conformacionais da SAW.

O ponto de partida é calcular a distância ponta-a-ponta quadrática média como função dos produtos escalares intermediários $\langle \vec{R}_N^2 \rangle_N = 2 \sum_{j=1}^N \langle \vec{R}_j \cdot \vec{u}_j \rangle_N - N$. O produto escalar intermediário traz a regra de escala para $\langle \vec{R}_N^2 \rangle_N$. Além disso, vamos fixar o primeiro passo da caminhada em uma direção específica do espaço, de modo que a projeção do vetor posição nessa direção represente o comprimento de persistência, $\langle x_N \rangle_N$.

O caso mais simples é o produto escalar $\langle \vec{R}_N \cdot \vec{u}_N \rangle_N$. Sua relação com a distância ponta-a-ponta quadrática é dada por meio da expressão $\langle \vec{R}_N^2 \rangle_N = \langle \vec{R}_{N-1}^2 \rangle_N + 2\langle \vec{R}_N \cdot \vec{u}_N \rangle_N - 1$. Quando o primeiro passo é fixo na direção do versor $+\hat{x}$, obtemos o novo resultado $\langle x_N \rangle_N = \langle \vec{R}_N \cdot \vec{u}_N \rangle_N$ e a distância quadrática média⁴, $\langle \vec{R}_N^2 \rangle_N = \langle \vec{R}_{N-1}^2 \rangle_N + 2\langle x_N \rangle_N - 1$, torna-se função do comprimento de persistência.

Ao que sabemos, o primeiro estudo do comprimento de persistência para SAW foi conduzido por Grassberger [69], que encontrou a regra de escala $\langle x_N \rangle_N \sim N^{\delta_0}$, com $\delta_0 = 0.063(10)$. Esse resultado é questionado por Redner e Privman [78], que encontraram $\langle x_N \rangle_N \sim \ln(N)$. Há cerca de uma década, Eisenberg e Baram [76] argu-

---

⁴Para facilitar a leitura usamos o termo *distância quadrática média* ao invés de distância ponta-a-ponta quadrática média.

mentaram que o comprimento de persistência não se comporta como lei de potência ou logaritmo, mas converge para uma constante. No entanto, nesse estudo os autores não calculam o valor da constante para o qual $\langle x_N \rangle_N$ converge, além disso, seu estudo é puramente numérico (simulação MC e enumeração exata) e não é estabelecida relação do comprimento de persistência com outras grandezas conformacionais.

Para o conjunto de caminhadas com N passos, na rede quadrada e cúbica, determinamos a regra de escala $\langle x_N \rangle_N \sim a_0 + a_1 N^{-w_0} + a_1 N^{-w_1}$, ou seja, o comprimento de persistência converge para uma constante com correções de escala. Essas correções são oriundas da série que representa a distância quadrática média. O ajuste dos nossos dados de simulação MC revela que $a_0 \approx 5/2$ para rede quadrada e $a_0 \approx \sqrt{2}$, para a rede cúbica.

Na parte inicial da caminhada a lei de escala $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N \sim 1/2 + \alpha_0(j - \tau)^{2\nu_0 - 1}$ vale até um determinado valor $\Theta(N)$ que depende não apenas de N , mas também da dimensão (aqui α_0 e τ são constantes⁵). A partir de $\Theta(N)$ o produto escalar intermediário incrementa mais lentamente que $\sim 1/2 + \alpha_0(j - \tau)^{2\nu_0 - 1}$, atinge um ponto de máximo e depois decai monotonamente até $j = N$ onde converge para uma constante. O comportamento de $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$ entre $\Theta(N) < j < N$ é devido ao aumento da flexibilidade da caminhada na porção final em relação à porção inicial. Qualitativamente, o aumento da flexibilidade é consequência da maior liberdade de movimentação (mais graus de liberdade) que o final da caminhada experimenta.

Empiricamente, por meio da análise de $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$, determinamos que o ponto ótimo para medir o expoente principal ν_0 é até o meio da caminhada, $\Theta(N) \approx N/2$, para rede quadrada, e para rede cúbica um terço, $\Theta(N) \approx N/3$. A implicação desse resultado é que a análise da convergência do expoente dominante ν_0 deve ser realizada até o passo $\Theta(N)$ onde a caminhada começa tornar-se mais flexível. A partir desse passo é possível ver uma clara tendência de separação entre as curvas de $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$, pois na caminhada de menor tamanho o produto escalar intermediário incrementa mais lentamente devido ao aumento da flexibilidade.

Por meio da análise da regra de escala de $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$ mostramos que é viável medir o valor de ν_0 com boa precisão, mesmo para caminhadas com poucos passos,

---

⁵A constante τ é conhecida como constante de suavização. É utilizada para diminuir efeitos de oscilação par-ímpar, veja gráfico da Fig. 5.4b.

pois já obtivemos convergência dos resultados. A seguir, apresentamos as principais grandezas conformacionais utilizadas e o método de simulação numérica MC empregado para validar nossos resultados analíticos.

5.2 Métodos

Nesta seção, descrevemos as principais grandezas conformacionais e expoentes estudados, além do algoritmo de simulação de MC utilizado na validação dos resultados analíticos.

5.2.1 Grandezas Conformacionais

Uma das grandezas fundamentais no estudo da SAW é o número c_N de caminhadas distintas com N passos. Para caminhadas curtas, tal como dois ou três passos é fácil calcular c_N . Contudo, para caminhadas longas o caráter autorrepulsivo torna a tarefa impraticável, como exemplo, considere a SAW da Fig. 5.1c com $N = 12$. Essa conformação é uma das $c_{12} = 324932$ possíveis para uma caminhada com 12 passos [79]. Todos os possíveis caminhos de tamanho $n + m$ podem ser gerados concatenando caminhadas com n e m passos, mas nem todas as caminhadas com $m + n$ passos são autorrepulsivas, de modo que $c_{m+n} < c_n c_m$, o que implica na existência do limite

$$\mu = \lim_{N \rightarrow \infty} c_N^{1/N}. \quad (5.1)$$

A grandeza μ é conhecida como constante de conectividade e depende da estrutura da rede. Para a rede hipercúbica o intervalo a que pertence o número de caminhadas é $d^N \leq c_N \leq 2d(2d - 1)^{N-1}$. O limite inferior é dado por caminhadas que evoluem em apenas um sentido (em cada passo). Já o limite superior leva em conta que o caminhante tem todas as direções e sentidos acessíveis, exceto o sentido imediatamente reverso (de onde o caminhante veio no último passo). O valor da constante de conectividade para rede regular quadrada é $\mu \approx 2.638158$ e para cúbica $\mu \approx 4.684043$. A grandeza c_N possui a regra de escala:

$$c_N \sim A\mu^N N^{\gamma-1}. \quad (5.2)$$

O expoente γ é igual a 1 para a caminhada aleatória, pois há exatamente $c_N = \mu^N$, com $\mu = 2d$, caminhadas possíveis. Acredita-se que o valor exato para rede quadrada

é $\gamma = 43/32$ [16], enquanto na rede cúbica $\gamma \approx 1.1575$ [16]. A Eq. 5.2 ainda apresenta correções analíticas, do tipo a_i/N^i com $i = 1, 2, \dots, i$ e correções não analíticas caracterizadas pelos expoentes Δ_1 e Δ_2 :

$$c_N = A\mu^N N^{\gamma-1} \left[1 + \frac{a_1}{N} + \frac{a_2}{N^2} + \dots + \frac{b_1}{N^{\Delta_1}} + \frac{b_2}{N^{\Delta_1+1}} + \dots + \frac{c_1}{N^{\Delta_2}} + \frac{c_2}{N^{\Delta_2+1}} + \dots \right], \quad (5.3)$$

onde a_i , b_i e c_i são constantes.

A conjectura para os expoentes de correção são provenientes dos mapeamentos da SAW em modelos de cadeia de *spin* [11] e teoria de grupo de renormalização. Não é objetivo do nosso estudo utilizar a teoria de fenômenos críticos para estudar a SAW, mas utilizamos o comportamento das grandezas conformacionais, dado que são amplamente aceitos na comunidade científica. Em termos de resultados de enumeração exata, Δ_1 e Δ_2 são estimados por meio da função geradora da caminhada⁶ $C(x) = \sum c_N x^N$ [26, 80]. Esse método para determinar o expoente principal ν_0 e os expoentes de correção de escala usa informação da caminhada como um todo, ou seja, está susceptível à enumeração exata.

Do ponto de vista da Física de Polímeros, e também na nossa abordagem em termos de caminhadas, as grandezas mais interessantes estão relacionadas com valores métricos, tal como a distância ponta-a-ponta quadrática:

$$\vec{R}_N^2 = |\omega(N) - \omega(0)|^2, \quad (5.4)$$

onde $\omega(0)$ é o sítio inicial e $\omega(N)$ o sítio final. O raio quadrático de giração:

$$\vec{R}_g^2 = \frac{1}{2(N+1)^2} \sum_{i,j=0}^N |\omega(i) - \omega(j)|^2. \quad (5.5)$$

O valor médio dessas duas últimas grandezas apresenta o mesmo comportamento assintótico: $\langle R_x^2 \rangle_N \sim D_x N^{2\nu_0}$, $x \in N, g$. O expoente ν_0 é universal, ou seja, depende da dimensão da rede, mas não da simetria. Assim como c_N , as Eqs. 5.4 e 5.5 apresentam as mesmas correções de escala que aquelas na Eq. 5.2:

$$\langle R_x^2 \rangle_N = N^{2\nu_0} \left[a_0 + \frac{a_1}{N} + \frac{a_2}{N^2} + \dots + \frac{b_1}{N^{\Delta_1}} + \frac{b_2}{N^{\Delta_1+1}} + \dots + \frac{c_1}{N^{\Delta_2}} + \frac{c_2}{N^{\Delta_2+1}} + \dots \right]. \quad (5.6)$$

---

⁶ $C(x)$ é conhecido como susceptibilidade da caminhada.

Para qualquer observável, no limite de uma caminhada longa [27], há uma regra de escala como aquela presente na Eq. 5.6. O que caracteriza cada um deles é seu expoente principal, ν_0 para distância quadrática, contudo os expoentes de correção de escala são sempre os mesmos, não importa o observável em questão.

A última grandeza conformacional tratada aqui é o comprimento de persistência⁷. Essa grandeza é definida como a projeção do vetor posição na direção do primeiro passo, matematicamente, já em termos da média:

$$\langle x_N \rangle_N = \langle \hat{u}_1 \cdot \vec{R}_N \rangle_N, \quad (5.7)$$

onde $\hat{u}_1 = \vec{u}_1/|\vec{u}_1|$ e $\vec{u}_1$ é vetor que representa o primeiro passo. No capítulo anterior estão disponíveis mais detalhes sobre o comprimento de persistência.

5.2.2 Algoritmo de Simulação Numérica

Numericamente, o modo mais simples de gerar uma caminhada aleatória autorrepulsiva com N passos é gerar caminhadas aleatórias com N passos e descartar as que não forem autorrepulsivas. O problema com esse método é que a probabilidade p_N de obter uma SAW com N passos é baixo, pois

$$p_N = \frac{c_N}{(2d)^N} \sim \left(\frac{\mu}{2d}\right)^N N^{\gamma-1} \sim e^{-\lambda N} N^{\gamma-1}, \quad (5.8)$$

onde $\lambda = \ln(2d/\mu)$ é chamada de constante de atrito. Para N grande, a probabilidade da caminhada continuar sendo uma SAW cai exponencialmente com N . Essa ineficiência no algoritmo é conhecida como problema de atrito [81]. Todos os métodos de MC para simular a caminhada autorrepulsiva têm por objetivo minimizar a constante de atrito [20].

Uma forma de melhorar esse algoritmo é utilizar caminhadas aleatórias com memória da direção e sentido do último passo (caminhada aleatória não reversa). Nessa dinâmica de movimentação, o caminhante que foi do sítio $\omega(i-1) \rightarrow \omega(i)$ no i -ésimo passo não pode, no $(i+1)$ -ésimo passo ir do sítio $\omega(i) \rightarrow \omega(i-1)$.

Dessa forma na caminhada aleatória não reversa há $2d(2d-1)^{N-1}$ configurações para uma caminhada com N passos e $\lambda = \ln[(2d-1)/\mu]$. Isso é melhor que utilizar uma caminhada aleatória pura, contudo, o problema de atrito permanece.

---

⁷Sugerimos a leitura da Sec. 4.3 para maiores detalhes

Os códigos mais rápidos são baseados em algoritmos de pivotamento [19] e variantes que usam representações de dados mais complexas [18, 28, 29] do ponto de vista computacional.

O modelo implementado é a caminhada aleatória não reversa. Nosso algoritmo de simulação numérica foi codificado em linguagem C e paralelizado com a interface de passagem de mensagem, em inglês *message passing interface* MPI [82], por meio da biblioteca OpenMPI. A paralelização consiste em vários caminhantes saindo de um mesmo ponto, ao mesmo tempo.

Um critério utilizado para validar o funcionamento correto do nosso algoritmo foi verificar que os nossos dados de simulação numérica MC estão de acordo com dados de numeração exata, para o comprimento de persistência [69, 78], deslocamento quadrático [75, 79] e com outros dados de simulação Monte Carlo [28, 80] da literatura.

Além de comparar os valores médios, testamos o momento centrado na distância quadrática média, $\langle \vec{R}_N^2 \rangle_N$, definido como

$$\Delta R_N^2 = \frac{\sqrt{\langle \vec{R}_N^4 \rangle_N - \langle \vec{R}_N^2 \rangle_N^2}}{\langle \vec{R}_N^2 \rangle_N}. \quad (5.9)$$

Para a caminhada aleatória, onde a distribuição de distância do caminhante em relação à origem é Gaussiana, a Ref. [81] fornece o resultado $\langle \vec{R}_N^4 \rangle_N = \langle \vec{R}_N^2 \rangle_N^2 (d+2)/d$ no limite $N \gg 1$, de modo que a Eq. 5.9 resulta em uma constante. Para a SAW, a distribuição da distância do caminhante em relação à origem não é Gaussiana. Portanto, existe uma outra constante de proporcionalidade para a razão ΔR_N^2 que numericamente, é ilustrada no gráfico da Fig. 5.2a, para a rede quadrada as caminhadas com $N > 120$ são pouco frequentes, por isso há oscilação mais acentuada a partir desse ponto. Note que tanto para a rede quadrada, quanto para rede cúbica, há a convergência da Eq. 5.9 para um valor constante.

O segundo teste numérico é o cálculo da constante de conectividade μ , que está ilustrado no gráfico da Fig. 5.2b. O valor de μ é fornecido pelo ajuste da probabilidade de obter uma SAW com N passos⁸, Eq. 5.8. Nos ajustes numéricos que seguem, os dados de simulação não estão com as barras de erro (dado que são

⁸A construção desse gráfico é simples, basta contar o número de SAWs que atingiram N passos, então, dividir esse número pelo total de caminhadas geradas.

Figura 5.2 – Testes realizados com os dados provenientes do nosso algoritmo. (a) Dados de acordo com a Eq. 5.9 para $d = 2$ e $d = 3$. Para $d = 2$ a oscilação no final do gráfico é devido ao baixo número de caminhadas que atingem $N > 120$ passos. (b) Constante de atrito, o ajuste não linear por meio da Eq. 5.8 fornece $\lambda = 0.12899(4)$ e $\lambda = 0.065298(4)$ para $d = 2$ e $d = 3$ respectivamente. Os valores da constante de atrito, calculadas a partir do valor de μ fornecido pela Ref. [16], são dados $\lambda = 0.128531205(1)$ e $\lambda = 0.0652762(28)$ para $d = 2$ e $d = 3$ respectivamente.

pequenas). Para uma grandeza conformacional qualquer $A = \{\vec{R}_j \cdot \vec{u}_j, \vec{R}_N^2\}$, medimos o erro padrão: $\epsilon^2 = (\langle A^2 \rangle_N - \langle A \rangle_N^2)/M$, onde M é o número de amostras. O valor obtido para ϵ não retrata a realidade, pois a distribuição do vetor posição ($\vec{R}_j$ ou $\vec{R}_N$) não é Gaussiana [81]. Além disso, o nosso algoritmo de simulação MC se torna cada vez mais ineficiente em obter uma boa estatística para as SAWs conforme o número de passos aumenta. Isso ocorre porque enquanto o número c_N de configurações aumenta, Eq. 5.2, a probabilidade do algoritmo utilizado obter uma SAW com N passos decai exponencialmente devido ao problema de atrito, Eq. 5.8. A ineficiência é portanto oriunda dessa situação, onde a contagem de caminhadas longas é cada vez menor enquanto o número c_N de SAWs aumenta com aumento de N . Utilizamos os dados de $N < 85$ e $N < 110$ para rede quadrada e cúbica, respectivamente, pois a oscilação do valor médio $\langle A \rangle_N$ não prejudica nossos ajustes numéricos.

5.3 Produtos Escalares Intermediários

Nessa seção apresentamos nossa abordagem do problema de caminhadas aleatórias autorrepulsivas. Ela consiste em estudar, detalhadamente, a média do produto

escalar entre o vetor posição com o versor deslocamento no j -ésimo passo: $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$. Como consequência apresentamos dois resultados analíticos relacionados ao produto escalar no final da caminhada, ou seja, o caso $j = N$. O primeiro resultado está relacionado com cálculo da distância quadrática média no *ensemble* adequado. O segundo consiste em demonstrar a igualdade $\langle x_N \rangle_N = \langle \vec{R}_N \cdot \vec{u}_N \rangle_N$, onde $\langle x_N \rangle_N$ é o comprimento de persistência.

5.3.1 Relação com a Distância Quadrática

Anteriormente, escrevemos a distância quadrática média, Eq. 5.6, como o quadrado do módulo da distância entre o sítio inicial $\omega(0)$ e final $\omega(N)$. Nesta forma os detalhes da evolução da caminhada até o N -ésimo passo são perdidos, pois olhamos apenas dois pontos, $\omega(0)$ e $\omega(N)$. Uma outra forma de escrever a distância quadrática é usar a função de correlação passo-a-passo:

$$\langle \vec{R}_N^2 \rangle_N = N + 2 \sum_{i=1}^N \sum_{j=i+1}^N \langle \vec{u}_i \cdot \vec{u}_j \rangle_N. \quad (5.10)$$

Na caminhada aleatória, $\langle \vec{u}_i \cdot \vec{u}_j \rangle_N = 0$ para $i \neq j$ e recuperamos o resultado $\langle \vec{R}_N^2 \rangle_N = N$. Ao contrário da caminhada aleatória, para a SAW a parcela $\langle \vec{u}_i \cdot \vec{u}_j \rangle_N \neq 0$ é a origem microscópica da regra de escala $\langle \vec{R}_N^2 \rangle_N \sim N^{2\nu_0}$ com $\nu_0 \geq 1/2$. Devido à importância do decaimento da função correlação, diversos trabalhos tratam do assunto [77, 83, 84] e buscam uma equação analítica para descrever tal comportamento. Entretanto, os dados de simulação e enumeração exata são escassos devido à complexidade (tempo computacional e correções de escala) envolvida na determinação da função correlação entre os passos. Para uma caminhada com comprimento do passo fixo, a distância quadrática do caminhante em relação à origem após j passos:

$$\vec{R}_j^2 = \vec{R}_{j-1}^2 + 2\vec{R}_{j-1} \cdot \vec{u}_j + u_j^2, \quad (5.11)$$

sem perda de generalidade assumimos $|u_j| = 1$. Realizando o somatório

$$\sum_{j=1}^N \vec{R}_j^2 = \sum_{j=1}^N \vec{R}_{j-1}^2 + \sum_{j=1}^N 2\vec{R}_{j-1} \cdot \vec{u}_j + N, \quad (5.12)$$

Figura 5.3 – Abordagem utilizada para estudar as leis de escala para caminhadas aleatórias repulsivas. (a) O estudo da correlação entre os passos é dado pelo produto escalar $\langle \vec{u}_i \cdot \vec{u}_j \rangle_N$ e esse produto é a origem microscópica da regra de escala. Estudá-lo numericamente é complicado, haja vista que sua regra de escala é $s^{2\nu_0-2}$, onde s é a separação em número de passos entre os sítios i e j . (b) O estudo da SAW por meio da distância ponta-a-ponta quadrática considera apenas os pontos finais e iniciais. (c) Nossa abordagem para estudar a SAW é intermediária aos dois extremos anteriores e permite boa precisão na determinação de ν_0 .

como $\vec{R}_0 = \vec{0}$ vale $\sum_{j=1}^N \vec{R}_{j-1}^2 = \sum_{j=1}^{N-1} \vec{R}_j^2$, e portanto a distância ponta-a-ponta quadrática média fica

$$\langle \vec{R}_N^2 \rangle_N = 2 \sum_{j=1}^N \langle \vec{R}_j \cdot \vec{u}_j \rangle_N - N. \quad (5.13)$$

A Eq. 5.13 mostra que a origem da lei de escala para a SAW está nas parcelas $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$. Essa nova forma de analisar $\langle \vec{R}_N^2 \rangle_N$ não é detalhista a ponto de estudar a caminhada passo-a-passo por meio da função correlação e tampouco generalista ao ponto de olhar apenas a distância quadrática ponta-a-ponta. Nossa forma de abordar a SAW é intermediária a esses dois extremos como ilustrado na Fig. 5.3. Isso possibilitou obter novos resultados, tal como o comprimento de persistência que será estudado analiticamente a seguir.

5.3.2 Média no *Ensemble* de Número de Passos

Para uma caminhada com N passos em uma rede regular qualquer, a distância ponta-a-ponta quadrática média é caso particular da Eq. 5.13:

$$\langle \vec{R}_N^2 \rangle_N - \langle \vec{R}_{N-1}^2 \rangle_N = 2\langle \vec{R}_N \cdot \vec{u}_N \rangle_N - 1, \quad (5.14)$$

onde $\vec{u}_N$ é o versor do último passo e $|\vec{u}_N| = 1$. Uma forma alternativa de escrever a Eq. 5.14 é $\langle \vec{R}_N^2 \rangle_N = \langle (\vec{R}_{N-1} + \vec{u}_N)^2 \rangle_N = \langle \vec{R}_{N-1}^2 \rangle_N + 2\langle \vec{R}_{N-1} \cdot \vec{u}_N \rangle_N + 1$. Trocando o valor de $\langle \vec{R}_{N-1} \cdot \vec{u}_N \rangle_N = \langle \vec{R}_N \cdot \vec{u}_N \rangle_N - 1$, recuperamos a Eq. 5.14.

Para a caminhada aleatória $\langle \vec{R}_N \cdot \vec{u}_N \rangle_N = 1$. Como não há restrição de visitaço, $\langle \vec{R}_{N-1}^2 \rangle_N = \langle \vec{R}_{N-1}^2 \rangle_{N-1}$, de modo que $\langle \vec{R}_N^2 \rangle_N = \langle \vec{R}_{N-1}^2 \rangle_{N-1} + 1$. Repetindo essa operao $(N - 1)$ vezes temos $\langle \vec{R}_N^2 \rangle_N = N$. Esse o comportamento de escala para caminhadas aleatrias. Diferentemente da caminhada aleatria, o valor de $\langle \vec{R}_N \cdot \vec{u}_N \rangle_N > 1$ para SAWs, devido ao carter autorrepulsivo. Essa caracterstica impede que stios sejam revisitados, um caminhante pode ficar at mesmo sem opes de visitao no ltimo passo (aprisionado), veja Fig. 5.1d.

A parcela esquerda na Eq. 5.14 parece a derivada com respeito ao nmero de passos, contudo isso no verdade. A derivada deve ser calculada utilizando o *ensemble* nmero de passos adequado, ou seja, $\langle \vec{R}_{N-1}^2 \rangle$ deve ser calculado no *ensemble* de caminhadas com comprimento $N - 1$ passos. A forma mais clara de ver que $\langle \vec{R}_{N-1}^2 \rangle_N \neq \langle \vec{R}_{N-1}^2 \rangle_{N-1}$ notar que nem todas as caminhadas autorrepulsivas com $N - 1$ passos so possveis no N -simo passo, veja Fig. 5.1d.

A ligao de $\langle \vec{R}_N \cdot \vec{u}_N \rangle_N$ com o comprimento de persistncia obtida por meio de operaes de translao de base e inverso no conjunto completo de caminhadas com N passos. Antes de mais nada devemos definir o conjunto que estamos trabalhando. O conjunto A_1 diz respeito a todas as c_N configuraes de uma SAW com N passos com incio no stio $\omega(0)$ (base comum). Essas caminhadas terminam em um conjunto de stios $\omega(N)$ arbitrrio. O segundo conjunto A_2 corresponde a todas as SAWs com origem nos stios $\omega(N)$ com um stio final comum $\omega(0)$. Note que cada caminhada agora tem uma origem distinta (base no comum).

Referenciamos as grandezas no conjunto A_2 com o *superscript* ($'$). Sendo assim, a relao entre os produtos escalares nos dois conjunto surge de uma operao

de inversão e mudança de base:

$$\langle \vec{R}_N \cdot \vec{u}_N \rangle_N = \langle -\vec{R}_N \cdot -\vec{u}_N \rangle_N \stackrel{t}{=} \langle \vec{R}'_N \cdot \vec{u}'_1 \rangle_N \quad (5.15)$$

onde $\stackrel{t}{=}$ indica a operação de translação de base. Nesses conjuntos há duas propriedades que devem ser válidas:

1. $\{\vec{R}_N\} \stackrel{t}{=} \{\vec{R}'_N\}$, ou seja, o conjunto de vetores que representam as SAWs são invariantes perante operação de inversão e mudança de base. Dessa propriedade decorre;
2. $\{\vec{u}_N\} \stackrel{t}{=} \{\vec{u}'_1\}$, ou seja, o conjunto de versores finais da caminhada com N passos dever ser o mesmo conjunto de versores iniciais.

Note que de acordo essas propriedades, a Eq. 5.15 pode ser reescrita apenas trocando a notação, sem perda de generalidade, como $\langle \vec{R}_N \cdot \vec{u}_N \rangle_N = \langle \vec{R}_N \cdot \vec{u}_1 \rangle_N$. Ao fixar o primeiro passo $\vec{u}_1$ reobtemos a própria definição de comprimento de persistência: $\langle x_N \rangle_N = \langle \vec{R}_N \cdot \vec{u}_1 \rangle_N$. A distância quadrática ponta-a-ponta pode ser reescrita como:

$$\langle \vec{R}_N^2 \rangle_N = \langle \vec{R}_{N-1}^2 \rangle_N + 2\langle x_N \rangle_N - 1. \quad (5.16)$$

Note que esta equação está relacionando medidas de dispersão, tal como $\langle \vec{R}_N^2 \rangle_N$ com o comprimento de persistência, que é uma medida de correlação. Até onde sabemos esse cálculo envolvendo o comprimento de persistência com $\langle \vec{R}_N \cdot \vec{u}_N \rangle_N$ é inédito. A Eq. 5.16 é obtida no contexto da SAW, contudo ela se aplica a outros modelos de caminhada, tal como a caminhada aleatória e a caminhada que representa o modelo da cadeia de rotação livre⁹. Na próxima seção estudaremos com detalhes o valor de $\langle \vec{R}_N \cdot \vec{u}_N \rangle_N$ no contexto do comprimento de persistência.

5.4 Comprimento de Persistência

Nessa seção mostramos que o comprimento de persistência converge para uma constante, acrescida de termos de correção do tipo N^{-w} , onde w é uma combinação linear dos expoentes ν_0 , Δ_i e as correções analíticas, dadas pelo expoente i

⁹É possível calcular $\langle x_N \rangle_N$ para a RW e a FRC, apresentados na Sec. 4.3, por meio da Eq. 5.16.

inteiro. O comprimento de persistência, para o meio bidimensional, foi inicialmente estudado por Grassberger [69], que encontrou o comportamento $\langle x_N \rangle_N \sim N^{\delta_0}$, onde $\delta_0 = 0.063(10)$. Como $\delta_0 \ll 1$, o comprimento de persistência pode ser bem ajustado pela função logaritmo, $\langle x_N \rangle_N \sim \ln(N)$, assim como encontrado por Privmann [78]. Recentemente conjecturou-se que ele é constante [76] para caminhadas bidimensionais. Medir o comprimento de persistência é uma tarefa complicada, pois determinar se uma curva converge para uma constante, diverge logaritmicamente ou cresce como lei de potência (com $\delta_0 \ll 1$) não é simples quando há poucos dados disponíveis.

Antes de prosseguirmos com a análise dos dados de $\langle \vec{R}_N \cdot \vec{u}_N \rangle_N$, vamos situar nossa conjectura acerca de $\langle x_N \rangle_N$ com o estudo desenvolvido por Eisenberg e Baram [76]. Os autores iniciam com a definição do comportamento da função correlação, $\xi_{1,j}(N)$, entre o primeiro e o j -ésimo passo em uma caminhada com N passos:

$$\langle \xi_{1,j} \rangle_N = \frac{\sum_{k=1}^{c_N} \vec{u}_1 \cdot \vec{u}_j(k)}{c_N}, \quad (5.17)$$

onde c_N é o número de configurações total de uma caminhada com N passos, Eq. 5.2, e $\vec{u}_1 \cdot \vec{u}_j(k)$ é o produto escalar entre o primeiro e j -ésimo passo na k -ésima configuração de uma caminhada com N passos. Se o primeiro passo for fixo na direção $+\hat{x}$, o valor de $\langle \xi_{1,j} \rangle_N$ é igual ao número médio de passos nessa direção, a contribuição em $\hat{y}$ não conta pois $\cos(\pi/2) = 0$. Assim, o comprimento de persistência para uma caminhada com N passos é dado por

$$\langle x_N \rangle_N = \sum_{j=1}^N \langle \xi_{1,j} \rangle_N. \quad (5.18)$$

De acordo com a Ref. [76], o decaimento da função correlação entre passos é $\langle \xi_{1,j} \rangle_N = c_0 j^{-1.34(5)}$. Portanto, da Eq. 5.18 concluímos que $\langle x_N \rangle_N$ converge para uma constante. Os autores não calculam o valor de $\langle x_N \rangle_N$, assim como o caso tridimensional não é investigado. Eisenberger e Baram apontam que $\langle \xi_{1,j} \rangle_N$ atinge seu valor máximo próximo ao passo $j \approx N/2$, decaindo a partir desse ponto, assim como indicado no gráfico da Fig. 5.4a.

Com dados de enumeração exata¹⁰ para $N \leq 24$, verificamos o comportamento descrito por Eisenberg e Baram. Adicionalmente, colapsamos a função

---

¹⁰O código foi desenvolvido pelo Prof. Marco A. A. Silva.

de correlação entre passos em termos das variáveis $\langle \xi_{1,j} \rangle_N / \langle \xi_{1,N/2} \rangle_N \times j/N$, onde $\langle \xi_{1,N/2} \rangle_N$ é a função de correlação para o meio da caminhada (visualmente o colapso é melhor ao usar a função de correlação no ponto central da caminhada com N passos). O gráfico da Fig. 5.4b mostra o comportamento dos dados no colapso. Note que mesmo para N pequeno, o colapso dos dados é muito bom.

Figura 5.4 – (a) O ponto de máximo para a função correlação ocorre para $j \approx N/2$, a curva representa apenas uma guia para os olhos. (b) O colapso da função correlação é melhor quando utilizamos como fator de normalização $\langle \xi_{1,N/2} \rangle_N$, ou seja, o valor de $\langle \xi_{1,j} \rangle_N$ para o meio da cadeia. Isso indica que no meio da cadeia há uma mudança de comportamento da função de correlação angular. Os dados utilizados nesses gráficos são provenientes de enumeração exata.

A partir do resultado da Ref. [76] a função de correlação angular decai como $\langle \xi_{1,j} \rangle_N \sim 0.6j^{-1.34(5)}$. Ao realizar o somatório dessa função obtemos o comprimento de persistência $\langle x_N \rangle_N \sim \alpha_0 - 1.7N^{-0.34(5)}$, onde α_0 é a constante de integração. Enquanto isso, da Eq. 5.16 esperamos que $\langle \vec{R}_{N-1}^2 \rangle_N \sim a'_0 N^{2\nu_0} + a'_1 N^{2\nu_0-1} + \dots + b'_0 N^{2\nu_0-\Delta_1} + b'_1 N^{2\nu_0-\Delta_1-1} \dots + c'$. Do mesmo modo $\langle \vec{R}_N^2 \rangle_N \sim a_0 N^{2\nu_0} + a_1 N^{2\nu_0-1} + \dots + b_0 N^{2\nu_0-\Delta_1} + b_1 N^{2\nu_0-\Delta_1-1} \dots + c$ e o comprimento de persistência pode ser escrito como

$$2\langle x_N \rangle_N = (a_0 - a'_0)N^{2\nu_0} + (a_1 - a'_1)N^{2\nu_0-1} + \dots + (b_0 - b'_0)N^{2\nu_0-\Delta_1} + (b_1 - b'_1)N^{2\nu_0-\Delta_1-1} + \dots + 1 + c - c'. \quad (5.19)$$

O comprimento de persistência converge se a Eq. 5.19 convergir. Na rede

quadrada, por exemplo, se $a_0 \neq a'_0$ ou $a_1 \neq a'_1$, $\langle x_N \rangle_N$ vai divergir com $N^{3/2}$ ou $N^{1/2}$ respectivamente, ao invés de $N^{0.063}$ ou $\ln(N)$ como Grasberger e Privmann sugerem. Do gráfico da Fig. 5.5a é evidente que o comprimento de persistência não diverge como $N^{3/2}$ ou $N^{1/2}$, além disso $\langle x_N \rangle_N$ parece ser convergente. Na rede cúbica a mesma argumentação pode ser aplicada e a conclusão é análoga, nesse caso os expoentes ν_0 e Δ_1 são diferentes e estão presentes na Tab. 5.1.

As parcelas referentes à divergência na Eq. 5.19 devem ser canceladas, para que um comportamento do tipo $\langle x_N \rangle_N \sim \alpha_0 + \alpha_1 N^{-0.34(5)}$ exista. Conjecturamos que apenas alguns termos referentes às correções de escala para $\langle \vec{R}_N^2 \rangle_N$ descrevem o comportamento de $\langle x_N \rangle_N$. O comprimento de persistência pode ser escrito como:

$$\langle x_N \rangle_N = \alpha_0 + \alpha_1 N^{-w_1} + \alpha_2 N^{-w_2} + \dots, \quad (5.20)$$

onde os expoentes w_i são combinações lineares do expoente dominante ν_0 com os expoentes de correção analítica, do tipo N^{-i} , e as não-analítica dado por $N^{-(\Delta_i+i)}$. Os valores dos expoentes, provenientes de cálculos exatos e simulação Monte Carlo, estão presentes na Tab. 5.1.

Tabela 5.1 – Tabela com os valores dos expoentes para duas e três dimensões. Tais expoentes foram obtidos de forma exata e/ou numérica. Para maiores detalhes, consulte a Ref. [16] para dados da rede quadrada e [28] para os dados da rede cúbica.

d	2	3
ν_0	3/4	0.587597(7)
Δ_1	3/2	0.52(12)
$2\nu_0 - 1$	1/2	0.1751
$2\nu_0 - 2$	-1/2	-0.8248
$2\nu_0 - \Delta_1$	0	0.65
$2\nu_0 - \Delta_1 - 1$	-1	-0.34

O ajuste pelo método estatístico dos mínimos quadrados, para rede quadrada e cúbica, confirma o resultado da Eq. 5.20. Além disso, nota-se que os dois primeiros expoentes da Eq. 5.20 são provenientes da correção de segunda ordem analítica e

não-analítica presentes na Eq. 5.6. A Fig. 5.5 apresenta os gráficos do comprimento de persistência.

Figura 5.5 – Comprimento de persistência para rede quadrada e cúbica. O *inset* é o resíduo proveniente do ajuste por mínimos quadrados. Para $d = 2$: $\alpha_0 = 2.5254(36)$, $\alpha_1 = -2.319(25)$ e $\alpha_3 = +0.814(27)$ e os expoentes $w_1 = 0.5$ e $w_2 = 1$. Os coeficientes são: $\alpha_0 = 1.422(1)$, $\alpha_1 = -0.39(6)$ e $\alpha_2 = -0.022(5)$, os expoentes são $w_1 = 0.8248$ e $w_2 = 0.34$.

A equação para o comprimento de persistência obtida por meio do ajuste dos dados numéricos de simulação Monte Carlo para rede quadrada:

$$\langle x_N \rangle_N = 2.525(4) - \frac{2.32(3)}{N^{0.5}} + \frac{0.81(3)}{N}, \quad (5.21)$$

para rede cúbica:

$$\langle x_N \rangle_N = 1.422(1) - \frac{0.39(6)}{N^{0.8248}} - \frac{0.022(5)}{N^{0.34}}. \quad (5.22)$$

A partir do resultado da Ref. [76] calculamos a forma assintótica $\langle x_N \rangle_N \sim \alpha_0 + \alpha_1 j^{-0.34(5)}$, onde $\alpha_1 \approx -1.7$. Realizamos o ajuste dos nossos dados de simulação MC com essa função, assim como ilustrado no gráfico da Fig. 5.6. Os parâmetros de ajuste:

$$\langle x_N \rangle_N = 2.664(3) - \frac{1.714(9)}{N^{0.34}}, \quad (5.23)$$

o valor da constante $\alpha_0 = 2.664(3)$ dado pela Eq. 5.23 é muito próximo ao $\alpha_0 = 2.525(4)$ obtido com a nossa função de ajuste Eq. 5.21. Note que o valor $\alpha_1 = -1.714(9)$ da Eq. 5.23, está próximo do esperado $\alpha_1 \approx -1.7$ dado pela integração de $\langle \xi_{1,j} \rangle_N \sim 0.6j^{-1.34(5)}$.

Figura 5.6 – Ajuste dos dados do comprimento de persistência com a função $\langle x_N \rangle_N = \alpha_0 + \alpha_1 N^{-0.34(5)}$. Os parâmetros são $\alpha_0 = 2.664(3)$ e $\alpha_1 = -1.714(9)$.

Demostramos a concordância da constante α_0 das Eqs. 5.21 e Eq. 5.23 para a qual o comprimento de persistência converge. Agora, utilizamos um método que consiste em calcular o expoente efetivo que relaciona as parcelas dependentes do número de passos da Eq. 5.21 com a parcela $N^{-0.34(5)}$ da Eq. 5.23. De acordo com Caracciolo *et. al.* [80] uma equação do tipo $\alpha_0 + \alpha_1/N^{w_1} + \alpha_2/N^{w_2}$ pode ser escrita como $a_0 + a_1/N^{w_{eff}(N)}$ onde $w_{eff}(N)$ é um expoente efetivo dependente das constantes α_1 e α_2 e expoentes w_1 e w_2 da seguinte maneira:

$$w_{eff}(N) = \frac{\alpha_1 w_1^2 N^{-w_1} + \alpha_2 w_2^2 N^{-w_2}}{\alpha_1 w_1 N^{-w_1} + \alpha_2 w_2 N^{-w_2}}, \quad (5.24)$$

para os valores de coeficientes dados pela Eq. 5.21, obtemos $w_{eff}(80) = 0.45(1)$, próximo ao resultado¹¹ $w_{eff} = 0.34(5)$, dado por Eisenberg e Baram [76]. Nossos dados são poucos para obter resultado mais preciso, além disso, não sabemos quantos termos mais entram na série da Eq. 5.21, usada para descrever o comprimento de persistência. Para o caso tridimensional, não encontramos estudos publicados sobre

¹¹O coeficiente $w_{eff}(N)$ é calculado para um valor de número de passos, no caso escolhemos o valor $N = 80$ pois a curva de ajuste da Eq. 5.21 descreve bem o valor do comprimento de persistência neste ponto, gráfico da Fig. 5.5.

o comprimento de persistência, por isso apenas o ajuste dos dados de simulação é apresentado.

5.5 Origem da Lei de Escala

Na seção anterior, escrevemos a distância ponta-a-ponta quadrática como $\langle \vec{R}_N^2 \rangle_N = \langle \vec{R}_{N-1}^2 \rangle_N + 2\langle \vec{R}_N \cdot \vec{u}_N \rangle_N - 1$. Essa forma de escrever $\langle \vec{R}_N^2 \rangle_N$ nos levou a análise detalhada do produto escalar $\langle \vec{R}_N \cdot \vec{u}_N \rangle_N$. Sabemos que $\langle \vec{R}_N^2 \rangle_N \sim N^{2\nu_0}$ com $\nu_0 > 1/2$ para $d = 2$ e $d = 3$ [26]. Como $\langle \vec{R}_N^2 \rangle_N = 2 \sum_{j=1}^N \langle \vec{R}_j \cdot \vec{u}_j \rangle_N - N$ esperamos que $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$ escale como $j^{2\nu_0-1}$, contudo na análise do comprimento de persistência determinamos que $\langle \vec{R}_N \cdot \vec{u}_N \rangle_N$ converge para uma constante. Essa constatação motivou o estudo do produto escalar $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$ para $1 \leq j \leq N$ e para vários valores de N .

O comportamento de $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$ está ilustrado no gráfico da Fig. 5.7a para $d = 2$, e no gráfico da Fig. 5.7b, para $d = 3$.

O colapso de dados para diversos valores de N mostra o comportamento comum de $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$ quando este é normalizado pelo produto escalar máximo. Esse comportamento é revelado mesmo para cadeias muito pequenas, como ilustrado nos gráficos das Figs. 5.7c e 5.7d para $d = 2$ e nos gráficos das Figs. 5.7e e 5.7f para $d = 3$.

Dos gráficos em escala $\log \times \log$, Figs. 5.7d e 5.7f, notamos dois comportamentos distintos de $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$. A região inicial do gráfico apresenta um crescimento linear, com coeficiente angular $2\nu_0 - 1$, enquanto a região final apresenta uma queda acentuada.

Esse declínio acentuado caracteriza a maior flexibilidade da parte final da caminhada em relação à porção inicial. Qualitativamente esse aumento da flexibilidade deve-se à maior liberdade de movimentação que o final da caminhada experimenta em relação à parte inicial, pois há mais graus de liberdade acessíveis para os últimos passos.

Indiretamente a queda de $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$ é reportada em estudos recentes na forma

Figura 5.7 – Dados e colapso de dados para o produto escalar $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$ para a rede quadrada e cúbica. Os gráficos (a) e (b) mostram que $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$ é aproximadamente igual até um ponto $\Theta(N)$ para caminhadas com N distinto. O colapso de dados é obtido por meio da curva $\langle \vec{R}_j \cdot \vec{u}_j \rangle / \langle \vec{R}_j \cdot \vec{u}_j \rangle_{\max} \times j/N$. O subíndice $_{\max}$ indica o valor máximo do produto escalar para um dado N . (c) Produto escalar colapsado em escala linear e (d) em escala logarítmica para $d = 2$. (e) Produto escalar colapsado em escala linear e (f) em escala logarítmica para $d = 3$. Nos gráficos em escala logarítmica a inclinação da parte linear da curva é proporcional a $2\nu_0 - 1$.

da correlação entre dois passos¹² i e j separados por outros s passos que comporta-se como $\langle \vec{u}_i \cdot \vec{u}_j \rangle_N \sim s^{2\nu_0-2}$ [85], tanto para a rede quadrada [86], quanto para a rede cúbica [87, 88]. Desses estudos é evidente¹³ que a partir de determinado passo $\Theta(N)$ a função correlação entre passos começa a decair mais rapidamente que $s^{2\nu_0-2}$. Em tais estudos, o passo $\Theta(N)$ a partir do qual a caminhada torna-se mais flexível não é investigado. Em nossa abordagem, determinar o valor $\Theta(N)$ a partir do qual a caminhada torna-se mais flexível é fundamental para saber até qual passo devemos considerar $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$ no ajuste do expoente ν_0 .

Para determinar o valor de $\Theta(N)$, propomos uma função de diferença do produto escalar intermediário entre *ensembles*:

$$\Delta \vec{R}_j(N_1, N_2) = \langle \vec{R}_j \cdot \vec{u}_j \rangle_{N_2} - \langle \vec{R}_j \cdot \vec{u}_j \rangle_{N_1}, \quad (5.25)$$

onde $N_2 > N_1$. Os gráficos associados à Eq. 5.25 para o meio bidimensional e tridimensional são mostrados na Fig. 5.8. Para o meio bidimensional, o produto escalar intermediário é igual até o meio da caminhada com menos passos, $j \sim N_1/2$. Para o meio tridimensional, eles são iguais até um terço da caminhada de menor tamanho, ou seja, até $j \sim N_1/3$. Esse resultado nos diz qual valor de j em função do número de passos deve ser utilizado para analisar o produto escalar intermediário. Para o meio bidimensional, esse valor é $\Theta(N) \approx N/2$ e para $d = 3$ é $\Theta(N) \approx N/3$.

A determinação do valor $\Theta(N)$ ótimo para análise de $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$ permite estudar a lei de escala associada ao seu crescimento. Na Eq. 5.13, a não existência da parcela linear N é relatada [79, 89], em nossas análises não encontramos indícios de sua existência. Essa parcela parece não existir em $\langle \vec{R}_N^2 \rangle_N$. A forma mais apropriada do termo linear ser cancelado é que o somatório de $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$ o cancele e ainda traga a regra de escala. Essa constatação nos impele a escrever a regra de escala para $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$ como:

$$\langle \vec{R}_j \cdot \vec{u}_j \rangle_N = \frac{1}{2} + \alpha_0(j - \tau)^\omega + \dots, \quad (5.26)$$

é esperado que o termo $\omega = 2\nu_0 - 1$, pois o somatório da Eq. 5.13 deve escalar como $N^{2\nu_0}$, onde τ é conhecida como constante de suavização. Numericamente, τ

---

¹²O produto escalar intermediário pode ser escrito como $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N = \sum_{i=1}^j \langle \vec{u}_i \cdot \vec{u}_j \rangle_N$.

¹³Veja o gráfico (a) da Fig.2 na Ref. [88].

Figura 5.8 – Diferença entre produto escalar intermediário para duas caminhadas com número de passos distintos. (a) meio bidimensional caminhada com $N_1 = 40$ e $N_2 = 60$ passos; (b) meio bidimensional caminhada com $N_1 = 60$ e $N_2 = 90$ passos; (c) meio tridimensional com $N_1 = 60$ e $N_2 = 90$ passos; (d) meio tridimensional com $N_1 = 90$ e $N_2 = 108$. Os gráficos indicam o ponto ótimo para análise do expoente principal do produto escalar intermediário. Para $d = 2$, devemos usar os dados até $\Theta(N) \sim N/2$ e para $d = 3$, até $\Theta(N) \sim N/3$.

representa um pequeno deslocamento em N , com o intuito de minimizar oscilações par-ímpar no número de sítios da caminhada, além disso ela pode carregar parte das correções analíticas [90], dado que para $j \gg 1$ a parcela $(j - \tau)^\omega \sim j^\omega(1 + f_1(\tau, \omega)/j + f_2(\tau, \omega)/j^2 \dots)$.

Os gráficos que mostram o ajuste do produto escalar à Eq. 5.26 estão na Fig. 5.9. Para o meio bidimensional $\omega = 0.50000(84)$, o que implica que ao realizar a soma da Eq. 5.13, encontramos a lei de escala $\langle \vec{R}_N^2 \rangle_N \sim N^{2\nu_0}$. Assim como esperado, o mesmo vale para 3 dimensões, só que $\omega = 0.20098(17)$, o mesmo resultado obtido

a partir do cálculo de Flory. Usando a Eq. 5.26, acrescida da primeira correção não analítica¹⁴, temos $\omega = 0.18778(71)$.

O resultado para $d = 3$ parece ruim quando comparado com o valor esperado de $2\nu_0 - 1 = 0.17518$, contudo esse valor está de acordo com o obtido pela teoria de Flory. Note que, ao adicionarmos um termo de correção não-analítico, fez o valor de ν_0 mais próximo do esperado. A dificuldade de determinar $\omega = 2\nu_0 - 1$ é devida a 3 fatores:

1. o valor de $\omega = 2\nu_0 - 1$ é muito pequeno;
2. podemos utilizar apenas um terço dos dados, sendo assim, em uma caminhada com 90 passos, apenas os 30 primeiros servem para o ajuste de $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$;
3. as correções não analíticas, dadas pelos termos b_i/N^{Δ_1+i} com $i \in \mathbb{N}$, são mais relevantes, pois Δ_1 é menor para $d = 3$ quando comparado com $d = 2$. Acredita-se que $\Delta_1 \approx 0.5$ para $d = 3$, enquanto $\Delta_1 = 3/2$ para $d = 2$.

Em geral, os ajustes para encontrar o expoente dominante, ν_0 , na Eq. 5.6 utilizam a informação de toda a caminhada, ou seja, a distância ponta-a-ponta [28, 80], nesse caso é necessário calcular $\langle \vec{R}_N^2 \rangle_N$ para diversos valores de N . Com a análise de $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$, em cada *ensemble* de N passos é possível determinar um valor $2\nu_0 - 1$, ou seja, podemos calcular uma média de ν_0 ou traçar uma curva de extrapolação de ν_0 em função de N .

Os gráficos das Figs. 5.7 indicam claramente que o produto escalar intermediário máximo desloca-se com uma regra bem definida em função do número de passos. Investigamos esse deslocamento por meio da derivada de $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$. O resultado está ilustrado nos gráficos da Fig. 5.10. O gráfico para rede quadrada é mais refinado que para rede cúbica, aparentemente o deslocamento do ponto de máximo obedece a lei $\langle \vec{R}_j \cdot \vec{u}_j \rangle_{N,max} \sim N\nu_0$.

5.6 Conclusões

Neste capítulo propomos o estudo do produto escalar médio entre o vetor posição e deslocamento no j -ésimo passo $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$ na SAW. Essa abordagem é o

---

¹⁴Acredita-se que $\Delta_1 \approx 1/2$ [28, 16].

Figura 5.9 – Produto escalar intermediário, no meio tridimensional os dados são provenientes de caminhadas com $N = 24$ até $N = 108$ e $\Delta N = 3$, no meio bidimensional $N = 18$ até $N = 60$ e $\Delta N = 6$. (a) Meio tridimensional sem correções de escala na Eq. 5.26 $\alpha_0 = 0.67871(32)$ $\omega = 0.20098(17)$ e $\tau = 0.7832(16)$ (b) Meio tridimensional com correção de escala $\Delta_1 = 0.5$, $\alpha_0 = 0.7203(23)$, $\alpha_1 = -0.0687(38)$, $\omega = 0.18778(71)$ e $\tau = 0.6387(81)$. (c) Meio bidimensional $\alpha_0 = 0.6618(17)$, $\omega = 0.50000(84)$ e $\tau = 0.4151(86)$.

nível intermediário entre estudar a correlação passo-a-passo ou apenas considerando o ponto inicial e final da caminhada, tal como é feito na distância ponta-a-ponta quadrática. Por meio de operações de simetria nas caminhadas estabelecemos o vínculo entre o comprimento de persistência e a distância ponta-a-ponta quadrática. Estudamos em detalhes o comportamento do comprimento de persistência e concluímos que ele converge para uma constante acrescida de correção de escala. Finalmente, estudamos o comportamento do produto escalar entre o vetor $\vec{R}_j$ e o

Figura 5.10 – Colapso de dados para a derivada de $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$. (a) Rede quadrada
(b) Rede cúbica.

passo intermediário $\vec{u}_j$ e encontramos a origem da regra de escala da SAW, mesmo para caminhadas com poucos passos.

Conclusão

O estudo desenvolvido nessa dissertação concerne ao aspecto estático e dinâmico relacionados à modelos de caminhada autorrepulsiva.

No modelo da caminhada determinista parcialmente autorrepulsiva (CDPA), estudamos o meio desordenado explorado pelo caminhante por meio da estatística de vizinhança e de distâncias. A estatística de vizinhança fornece a probabilidade de um ponto ser o m -ésimo vizinho mais próximo de seu n -ésimo vizinho mais próximo. Essas probabilidades são dadas pela fórmula de Cox, cujo cálculo foi reproduzido no Cap. 2. A extensão da fórmula de Cox foi reobtida para o caso de alta dimensionalidade por meio do uso adequado das funções especiais apresentadas no Apêndice A. Esses resultados foram empregados na descrição do deslocamento do caminhante com memória unitária em um meio desordenado com dimensionalidade arbitrária.

A estatística de vizinhança, informa a reciprocidade da ordem de vizinhança por meio do critério de proximidade entre um par de pontos. Note que a informação acerca da distância entre os pontos é vaga e consiste num caráter quantitativo que traz pouca informação. Suprimos essa falta de informação no Cap. 3 com o estudo da estatística de distâncias, que mede como está distribuída a distância de um ponto ao seu k -ésimo vizinho mais próximo em um meio d -dimensional, gerado por um processo Poissônico espacial. A função densidade de probabilidade obtida é a distribuição gama generalizada e por meio de combinações de k e d obtemos várias distribuições particulares representadas pelas pdfs do Apêndice D. Exploramos os casos limite de alta dimensionalidade e alta ordem de vizinhança e propomos um teste para verificar se um conjunto de pontos foi gerado por um processo Poissônico

espacial. Os resultados foram publicados e pretendemos utilizá-los no futuro para descrever o processo difusivo gerado pelo movimentação do caminhante na CDPA.

O aspecto dinâmico apresentado nessa dissertação é oriundo da caminhada aleatória autorrepulsiva (SAW), um modelo que pode ser utilizado para descrever polímeros reais. No Cap. 5, estudamos esse modelo por meio de regras de simetria e produtos vetoriais, em suma inspirados pelo tratamento utilizado em modelos de caminhadas mais simples, como a caminhada aleatória e outras apresentadas no Cap. 4 para representar polímeros ideais. Nosso estudo é inovador e consiste da análise do produto escalar entre o vetor posição e o vetor deslocamento no j -ésimo passo $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$. A vantagem com esse método é que não somos detalhistas a ponto de utilizar a informação de cada passo, fornecida pela função de correlação entre passos $\langle \vec{u}_i \cdot \vec{u}_j \rangle_N$. Tampouco somos generalistas ao usar a distância quadrática entre o ponto inicial $\omega(0)$ e ponto final $\omega(N)$. Nosso código de simulação Monte Carlo foi codificado em linguagem C e paralelizado com MPI. Para $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$ os dados de simulação suportam 3 novos resultados. O primeiro é o caso particular $j = N$, no qual o produto escalar coincide com o comprimento de persistência, numericamente determinamos que o comprimento de persistência converge para uma constante. O segundo resultado é que é possível determinar o expoente principal ν_0 com boa precisão a partir de $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$ utilizando os dados de $1 \leq j \leq \Theta(N)$. O terceiro resultado refere-se ao passo $\Theta(N)$, que foi determinado quantitativamente para rede quadrada e cúbica, qualitativamente interpretamos que a partir desse passo ocorre um aumento da flexibilidade da caminhada de N passos.

Todo esse conjunto de resultados ainda nos mostra que há pontos em aberto que carecem de uma investigação futura: Determinação da fórmula de Cox para sistemas de tamanho finito. Utilização da estatística de distâncias para descrever a distância percorrida pelo turista após N passos. Melhorar os dados de simulação Monte Carlo da SAW por meio da implementação do algoritmo de pivotamento. Provar analiticamente que o comprimento de persistência para a SAW converge para uma constante. Estudar e entender o comportamento do produto escalar intermediário, $\langle \vec{R}_j \cdot \vec{u}_j \rangle_N$, para $\Theta(N) \leq j \leq N$, dentre outros.

Referências¹

- [1] K. Pearson, “The problem of the random walk”, *Nature*, vol. 72, p. 342, 1905. [1](#), [101](#)
- [2] E. A. Codling, M. J. Plank e S. Benhamou, “Random walk models in Biology”, *Journal of The Royal Society Interface*, vol. 5, no. 25, pp. 813–834, 2008. [1](#), [42](#), [103](#)
- [3] J. Klafter e I. M. Sokolov, *First Steps in Random Walks: From Tools to Applications*. Oxford Scholarship Online, 2013. [1](#), [42](#)
- [4] G. M. Akselrod, P. B. Deotare, N. J. Thompson, J. Lee, W. A. Tisdale, M. A. Baldo, V. M. Menon e V. Bulović, “Visualization of exciton transport in ordered and disordered molecular solids”, *Nature Communications*, vol. 5, no. 3646, pp. 1–8, 2014. [1](#)
- [5] H. E. Stanley, “Statistical physics: The salesman and the tourist”, *Nature*, vol. 413, pp. 373–374, 2001. [1](#)
- [6] H. Freund e P. Grassberger, “The red queen’s walk”, *Physica A: Statistical Mechanics and its Applications*, vol. 190, no. 3-4, pp. 218–237, 1992. [1](#)
- [7] L. A. Bunimovich, “Walks in rigid environments”, *Physica A: Statistical Mechanics and its Applications*, vol. 279, no. 1-4, pp. 169–179, 2000. [1](#)
- [8] G. F. Lima, A. S. Martinez e O. Kinouchi, “Deterministic walks in random media”, *Phys. Rev. Lett.*, vol. 87, p. 010603, Jun 2001. [1](#), [2](#), [14](#), [19](#)

---

¹De acordo com o estilo LMSC (baseado no estilo IEEE Transactions - Electrical and Electronics Engineers).

-
- [9] J. M. Berbert e W. F. Fagan, “How the interplay between individual spatial memory and landscape persistence can generate population distribution patterns”, *Ecological Complexity*, vol. 12, no. 0, pp. 1–12, 2012. [1](#)
- [10] W. F. Fagan, M. A. Lewis, M. Auger-Méthé, T. Avgar, S. Benhamou, G. Breed, L. LaDage, U. E. Schlägel, W.-W. Tang, Y. P. Papastamatiou, J. Forester e T. Mueller, “Spatial memory and animal movement”, *Ecology Letters*, vol. 16, no. 10, pp. 1316–1329, 2013. [1](#)
- [11] P. de Gennes, *Scaling concepts in polymer physics*. Cornell Univ. Pr., 1979. [1](#), [2](#), [37](#), [39](#), [43](#), [48](#)
- [12] M. G. Campiteli, P. D. Batista, O. Kinouchi e A. S. Martinez, “Deterministic walks as an algorithm of pattern recognition”, *Phys. Rev. E*, vol. 74, p. 026703, Aug 2006. [2](#)
- [13] T. C. Silva e L. Zhao, “High-level pattern-based classification via tourist walks in networks”, *Information Sciences*, vol. 294, no. 0, pp. 109–126, 2015. Innovative Applications of Artificial Neural Networks in Engineering. [2](#)
- [14] A. R. Backes, W. N. Gonçalves, A. S. Martinez e O. M. Bruno, “Texture analysis and classification using deterministic tourist walk”, *Pattern Recognition*, vol. 43, no. 3, pp. 685–694, 2010. [2](#)
- [15] O. Kinouchi, A. Martinez, G. Lima, G. Lourenço e S. Risau-Gusman, “Deterministic walks in random networks: an application to thesaurus graphs”, *Physica A: Statistical Mechanics and its Applications*, vol. 315, no. 3-4, pp. 665–676, 2002. [2](#)
- [16] E. J. J. van Rensburg, “Monte Carlo methods for the self-avoiding walk”, *Journal of Physics A: Mathematical and Theoretical*, vol. 42, no. 32, p. 323001, 2009. [xviii](#), [xxi](#), [2](#), [48](#), [51](#), [58](#), [65](#)
- [17] A. D. Sokal, “Monte Carlo methods for the self-avoiding walk”, *Nuclear Physics B - Proceedings Supplements*, vol. 47, no. 1-3, pp. 172 – 179, 1996. [2](#), [44](#)

-
- [18] N. Clisby, “Endless self-avoiding walks”, *Journal of Physics A: Mathematical and Theoretical*, vol. 46, no. 23, p. 235001, 2013. [2](#), [44](#), [50](#)
- [19] N. Madras e A. Sokal, “The pivot algorithm: A highly efficient Monte Carlo method for the self-avoiding walk”, *Journal of Statistical Physics*, vol. 50, no. 1-2, pp. 109–186, 1988. [2](#), [44](#), [50](#)
- [20] A. D. Sokal, “Monte Carlo methods for the self-avoiding walk”, *arXiv:hep-lat/9405016v1*, 1994. [2](#), [49](#)
- [21] I. Jensen, “Enumeration of self-avoiding walks on the square lattice”, *Journal of Physics A: Mathematical and General*, vol. 37, no. 21, p. 5503, 2004. [2](#)
- [22] C. A. S. Terçariol e A. S. Martinez, “Analytical results for the statistical distribution related to a memoryless deterministic walk: Dimensionality effect and mean-field models”, *Phys. Rev. E*, vol. 72, p. 021103, 2005. [3](#)
- [23] T. F. Cox, “Reflexive nearest neighbours”, *Biometrics*, vol. 37, no. 2, pp. pp. 367–369, 1981. [3](#), [4](#), [10](#), [11](#)
- [24] C. A. S. Terçariol, F. D. M. Kiipper e A. S. Martinez, “An analytical calculation of neighbourhood order probabilities for high dimensional poissonian processes and mean field models”, *J. Phys. A: Math. Theor*, vol. 40, no. 9, p. 1981, 2007. [3](#), [4](#), [10](#), [12](#), [14](#), [19](#)
- [25] C. R. F. Granzotti e A. S. Martinez, “Distance statistics in random media”, *The European Physical Journal B*, vol. 87, no. 4, pp. 1–5, 2014. [xvi](#), [xvii](#), [3](#), [4](#), [18](#), [22](#), [26](#)
- [26] B. Li, N. Madras e A. D. Sokal, “Critical exponents, hyperscaling, and universal amplitude ratios for two- and three-dimensional self-avoiding walks”, *Journal of Statistical Physics*, vol. 80, no. 3-4, pp. 661–754, 1995. [3](#), [48](#), [61](#)
- [27] P. Belohorec, *Renormalization group calculation of the universal critical exponents of a polymer molecule*. Tese de Doutorado, The University of Guelph, 1997. [3](#), [49](#)

-
- [28] N. Clisby, “Accurate estimate of the critical exponent ν for self-avoiding walks via a fast implementation of the pivot algorithm”, *Physical Review Letters*, vol. 104, no. 5, 2010. [xxi](#), [3](#), [44](#), [50](#), [58](#), [65](#)
- [29] N. Clisby, “Efficient implementation of the pivot algorithm for self-avoiding walks”, *Journal of Statistical Physics*, vol. 140, no. 2, pp. 349–392, 2010. [3](#), [50](#)
- [30] M. Rubinstein e R. Colby, *Polymer physics*. OUP Oxford, 2003. [4](#), [33](#), [34](#), [35](#), [36](#)
- [31] W. Feller, *An introduction to probability theory and its applications*, vol. 1. Wiley, 1968. [8](#)
- [32] P. Meyer, *Probabilidade: aplicações à estatística*. Ltc, 1984. [8](#)
- [33] J. M. S. Berbert, “Transição entre os comportamentos estendido e localizado em caminhadas estocásticas parcialmente auto-repulsivas em sistemas desordenados unidimensionais”, Dissertação de Mestrado, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2009. [8](#), [15](#)
- [34] C. A. S. Terçariol e A. S. Martinez, “An efficient algorithm to generate random uncorrelated Euclidean distances: the random link model”, *Brazilian Journal of Physics*, vol. 36, pp. 232–236, 2006. [8](#), [18](#)
- [35] P. J. Clark e F. C. Evans, “On some aspects of spatial pattern in biological populations”, *Science*, vol. 121, no. 3142, pp. 397–398, 1955. [10](#)
- [36] P. J. Clark, “Grouping in spatial distributions”, *Science*, vol. 123, no. 3192, pp. 373–374, 1956. [10](#)
- [37] M. F. Dacey, “Proportion of reflexive n -th order neighbors in spatial distribution”, *Geographical Analysis*, vol. 1, no. 4, pp. 385–388, 1969. [10](#)
- [38] C. A. S. Terçariol, *Caminhadas deterministas parcialmente auto-repulsivas: resultados analíticos para o efeito da memória do turista na exploração de meios desordenados*. Tese de Doutorado, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, University of São Paulo, Ribeirão Preto, 2008. [10](#), [82](#)

- [39] R. S. González, J. M. a. Berbert, C. A. S. Terçariol, W. T. R. Oliveira e A. S. Martinez, “Deterministic tourist walk”, *Journal of Computational Interdisciplinary Sciences*, vol. 1, no. 2, p. 119, 2009. [14](#)
- [40] C. A. S. Terçariol, R. S. González, W. T. R. Oliveira e A. S. Martinez, “Deterministic and random partially self-avoiding walks in random media”, *Physica A: Statistical Mechanics and its Applications*, vol. 386, no. 2, pp. 678–680, 2007. Disorder and Complexity Proceedings of the Pan American Scientific Institute (PASI) Conference “Disorder and Complexity”, Mar del Plata, Argentina, 11–20 December 2006 Pan American Scientific Institute (PASI) Conference “Disorder and Complexity”. [15](#)
- [41] R. S. González, *Funções generalizadas, modelos de crescimento contínuos e discretos e caminhadas estocásticas em meios desordenados*. Tese de Doutorado, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2011. [15](#)
- [42] N. J. Cerf, J. Boutet de Monvel, O. Bohigas, O. C. Martin e A. G. Percus, “The random link approximation for the Euclidean traveling salesman problem”, *J. Phys. I France*, vol. 7, no. 1, pp. 117–136, 1997. [17](#), [19](#), [28](#)
- [43] H. R. Thompson, “Distribution of distance to n -th neighbour in a population of randomly distributed individuals”, *Ecology*, vol. 37, no. 2, pp. 391–394, 1956. [18](#), [19](#), [20](#), [28](#)
- [44] S. Chandrasekhar, “Stochastic problems in physics and astronomy”, *Rev. Mod. Phys.*, vol. 15, pp. 1–89, Jan 1943. [19](#)
- [45] P. J. Clark e F. C. Evans, “Distance to nearest neighbor as a measure of spatial relationships in populations”, *Ecology*, vol. 35, no. 4, pp. 445–453, 1954. [19](#)
- [46] A. Chakraborti, “The Euclidean travelling salesman problem: frequency distribution of neighbors for small-size systems”, *Int. J. Mod. Phys. C*, vol. 12, no. 06, pp. 857–863, 2001. [19](#)
- [47] M. Mezard e G. Parisi, “The Euclidean matching problem”, *Europhysics Letters*, vol. 49, no. 12, pp. 2019–2025, 1988. [19](#)

-
- [48] J. Houdayer, J. Boutet de Monvel e O. Martin, “Comparing mean field and Euclidean matching problems”, *Eur. Phys. J. B*, vol. 6, no. 3, pp. 383–393, 1998. [19](#), [25](#)
- [49] C. A. S. Terçariol, R. S. González e A. S. Martinez, “Analytical calculation for the percolation crossover in deterministic partially self-avoiding walks in one-dimensional random media”, *Phys. Rev. E*, vol. 75, p. 061117, 2007. [19](#)
- [50] A. Tewari e A. Gokhale, “Nearest-neighbor distributions in thin films, sheets, and plates”, *Acta Materialia*, vol. 54, no. 7, pp. 1957–1963, 2006. [19](#)
- [51] T. Cover e P. Hart, “Nearest neighbor pattern classification”, *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, 1967. [19](#)
- [52] S. Singh, J. Haddon e M. Markou, “Nearest-neighbour classifiers in natural scene analysis”, *Pattern Recognition*, vol. 34, no. 8, pp. 1601–1612, 2001. [19](#)
- [53] M. Haenggi, “On distances in uniformly random networks”, *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3584–3586, 2005. [19](#)
- [54] A. G. Percus e O. C. Martin, “Scaling universalities of k -th nearest neighbor distances on closed manifolds”, *Adv. Appl. Math*, vol. 21, no. 3, pp. 424–436, 1998. [19](#), [21](#), [22](#)
- [55] P. Bhattacharyya e B. K. Chakrabarti, “The mean distance to the n -th neighbour in a uniform distribution of random points: an application of probability theory”, *Eur. J. Phys.*, vol. 29, no. 3, p. 639, 2008. [19](#)
- [56] D. Evans, A. J. Jones e W. M. Schmidt, “Asymptotic moments of near neighbour distance distributions”, *Proc. R. Soc. A*, vol. 458, no. 2028, pp. 2839–2849, 2002. [19](#)
- [57] D. Evans, “A law of large numbers for nearest neighbour statistics”, *Proc. R. Soc. A*, vol. 464, no. 2100, pp. 3175–3192, 2008. [19](#)
- [58] D. Moltchanov, “Distance distributions in random networks”, *Ad Hoc Networks*, vol. 10, no. 6, pp. 1146–1166, 2012. [19](#)

-
- [59] E. Stacy, “A generalization of the gamma distribution.”, *Annals Mathematical Statistics*, vol. 33, pp. 1187–1192, 1962. [20](#), [91](#)
- [60] C. G. E., “The Amoroso distribution”, *arXiv:1005.3274*, 2010. [20](#), [89](#), [91](#), [95](#)
- [61] M. Abramowitz e I. A. Stegun, *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. Dover books on mathematics, Dover, 1972. [24](#), [82](#), [85](#), [86](#), [87](#)
- [62] C. Cantor e P. Schimmel, *Biophysical chemistry: Part III: The behavior of biological macromolecules*. Biophysical Chemistry, W. H. Freeman, 1980. [31](#), [36](#), [37](#)
- [63] J. Mark, *Physical properties of polymers handbook*. Springer, 2007. [32](#), [34](#), [36](#)
- [64] G. Witz, K. Rechendorff, J. Adamcik e G. Dietler, “Conformation of circular DNA in two dimensions”, *Phys. Rev. Lett.*, vol. 101, p. 148103, 2008. [33](#), [34](#), [43](#)
- [65] K. Rechendorff, G. Witz, J. Adamcik e G. Dietler, “Persistence length and scaling properties of single-stranded DNA adsorbed on modified graphite”, *The Journal of Chemical Physics*, vol. 131, no. 9, p. 095103, 2009. [33](#), [34](#), [36](#)
- [66] I. Teraoka, *Polymer solutions: An introduction to physical properties*. John Wiley & Sons, 2002. [33](#)
- [67] H.-P. Hsu, W. Paul e K. Binder, “Polymer chain stiffness vs. excluded volume: A Monte Carlo study of the crossover towards the worm-like chain model”, *EPL (Europhysics Letters)*, vol. 92, no. 2, p. 28003, 2010. [34](#)
- [68] P. D. Kilanowski, *On the Kratky-Porod model for semi-flexible polymers in an external force field*. Tese de Doutorado, Ohio State University, 2010. [35](#)
- [69] P. Grassberger, “On persistency in self-avoiding random walks”, *Physics Letters A*, vol. 89, no. 8, pp. 381–384, 1982. [36](#), [44](#), [45](#), [50](#), [56](#)
- [70] J. Silva e J. Lima, “Quatro abordagens para o movimento Browniano”, *Revista Brasileira de Ensino de Física*, vol. 29, no. 1, pp. 25–35, 2007. [42](#), [103](#)

-
- [71] F. Valle, M. Favre, P. De Los Rios, A. Rosa e G. Dietler, “Scaling exponents and probability distributions of DNA end-to-end distance”, *Phys. Rev. Lett.*, vol. 95, p. 158105, 2005. [43](#)
- [72] L. Peliti e L. Pietronero, “Random walks with memory”, *La Rivista del Nuovo Cimento*, vol. 10, no. 6, pp. 1–33, 1987. [43](#)
- [73] G. F. Lawler, O. Schramm e W. Werner, “On the scaling limit of planar self-avoiding walk”, *arXiv preprint math/0204277*, 2002. [43](#)
- [74] I. Jensen, “A new transfer-matrix algorithm for exact enumerations: self-avoiding walks on the square lattice”, *arXiv preprint arXiv:1309.6709*, 2013. [43](#)
- [75] R. D. Schram, G. T. Barkema e R. H. Bisseling, “Exact enumeration of self-avoiding walks”, *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2011, no. 06, p. P06019, 2011. [43](#), [50](#)
- [76] E. Eisenberg e A. Baram, “The persistence length of two-dimensional self-avoiding random walks”, *Journal of Physics A: Mathematical and General*, vol. 36, no. 8, p. L121, 2003. [44](#), [45](#), [56](#), [57](#), [59](#), [60](#)
- [77] C. Domb e F. T. Hioe, “Correlations in a self-avoiding walk”, *The Journal of Chemical Physics*, vol. 51, no. 5, pp. 1920–1928, 1969. [44](#), [52](#)
- [78] S. Redner e V. Privman, “Persistency of two-dimensional self-avoiding walks”, *Journal of Physics A: Mathematical and General*, vol. 20, no. 13, p. L857, 1987. [45](#), [50](#), [56](#)
- [79] N. Madras e G. Slade, *The Self-Avoiding Walk*. Modern Birkhäuser Classics, Springer, 2012. [47](#), [50](#), [63](#)
- [80] S. Caracciolo, A. J. Guttmann, I. Jensen, A. Pelissetto, A. N. Rogers e A. D. Sokal, “Correction-to-scaling exponents for two-dimensional self-avoiding walks”, *Journal of Statistical Physics*, vol. 120, no. 5-6, pp. 1037–1100, 2005. [48](#), [50](#), [60](#), [65](#)

-
- [81] K. Binder e D. Heermann, *Monte Carlo Simulation in statistical physics: An introduction*. Graduate Texts in Physics, Springer-Verlag, 2010. [49](#), [50](#), [51](#)
- [82] G. A. Kumar Vipin, Karpis George e G. Ananth, *Introduction to Parallel Computing*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2nd ed., 2002. [xxi](#), [50](#), [109](#), [111](#), [115](#), [116](#)
- [83] A. Baumgärtner, “Statistics of self-avoiding ring polymers”, *The Journal of Chemical Physics*, vol. 76, no. 8, pp. 4275–4280, 1982. [52](#)
- [84] C. A. Croxton, “Statistics of self-avoiding ring polymers”, *Journal of Physics A: Mathematical and General*, vol. 18, no. 6, p. 995, 1985. [52](#)
- [85] L. Schafer, A. Ostendorf e J. Hager, “Scaling of the correlations among segment directions of a self-repelling polymer chain”, *Journal of Physics A: Mathematical and General*, vol. 32, no. 45, p. 7875, 1999. [63](#)
- [86] H.-P. Hsu, W. Paul e K. Binder, “Breakdown of the Kratky-Porod wormlike chain model for semiflexible polymers in two dimensions”, *EPL (Europhysics Letters)*, vol. 95, no. 6, p. 68004, 2011. [63](#)
- [87] H.-P. Hsu, W. Paul e K. Binder, “Estimation of persistence lengths of semiflexible polymers: insight from simulations”, *Polymer Science Series C*, vol. 55, no. 1, pp. 39–59, 2013. [63](#)
- [88] H.-P. Hsu, W. Paul e K. Binder, “Standard definitions of persistence length do not describe the local “intrinsic” stiffness of real polymer chains”, *Macromolecules*, vol. 43, no. 6, pp. 3094–3102, 2010. [63](#)
- [89] Y.-B. Chan e A. Rechnitzer, “A Monte Carlo study of non-trapped self-avoiding walks”, *Journal of Physics A: Mathematical and Theoretical*, vol. 45, no. 40, p. 405004, 2012. [63](#)
- [90] N. Clisby, R. Liang e G. Slade, “Self-avoiding walk enumeration via the lace expansion”, *Journal of Physics A: Mathematical and Theoretical*, vol. 40, no. 36, p. 10973, 2007. [64](#)

-
- [91] G. B. Arfken, H. J. Weber e F. E. Harris, *Mathematical methods for physicists*. Academic Press, 6 ed., 2005. [82](#), [85](#)
- [92] S. Ross, *A first course in probability*. Pearson Prentice Hall, 2010. [85](#), [93](#), [94](#), [103](#)
- [93] H. Gould e J. Tobochnik, *An introduction to computer simulation methods: applications to physical systems*. No. v. 1-2 em Addison-Wesley series in physics, Addison-Wesley, 1988. [103](#)
- [94] P. García-Risueño e P. E. Ibáñez, “A review of high performance computing foundations for scientists”, *International Journal of Modern Physics C*, vol. 23, no. 07, p. 1230001, 2012. [108](#), [109](#)
- [95] H. Jin, D. Jespersen, P. Mehrotra, R. Biswas, L. Huang e B. Chapman, “High performance computing using MPI and OpenMP on multi-core parallel systems”, *Parallel Computing*, vol. 37, no. 9, pp. 562 – 575, 2011. [109](#)
- [96] W. D. Gropp, “Learning from the success of MPI”, in *High Performance Computing - HiPC 2001* (B. Monien, V. Prasanna e S. Vajapeyam, eds.), vol. 2228 de *Lecture Notes in Computer Science*, pp. 81–92, Springer Berlin Heidelberg, 2001. [110](#)

Funções Especiais

Neste apêndice apresentamos o estudo de funções especiais que emergem em muitos problemas de Estatística e Física Matemática. Estas funções também surgem naturalmente nos cálculos da estatística de vizinhança e de distâncias apresentadas no Cap. 2 e Cap. 3.

A.1 Erro

A função erro é definida como:

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z dt e^{-t^2}, \quad (\text{A.1})$$

em estatística, essa expressão surge no cálculo da função de distribuição acumulada da pdf Gaussiana, Eq. B.63. A função erro complementar é definida como:

$$\operatorname{erfc}(z) = \frac{2}{\sqrt{\pi}} \int_z^\infty dt e^{-t^2}. \quad (\text{A.2})$$

A soma das Eqs. A.1 e A.2 é a integral $\int_0^\infty dt e^{-t^2}$, que representa a área sob metade da curva de uma distribuição Gaussiana padrão. Essa área pode ser calculada a partir da integral Gaussiana de ordem zero

$$I_0 = \int_{-\infty}^\infty dx e^{-x^2}, \quad (\text{A.3})$$

o segredo para calcular essa última expressão é elevá-la ao quadrado:

$$I_0^2 = \int_{-\infty}^\infty dx e^{-x^2} \int_{-\infty}^\infty dy e^{-y^2} = \int_{-\infty}^\infty \int_{-\infty}^\infty dx dy e^{-(x^2+y^2)}, \quad (\text{A.4})$$

identificar que $r^2 = x^2 + y^2$ e que a Eq. A.4 pode ser escrita em função de coordenadas polares

$$I_0^2 = \int_{-\infty}^{\infty} \int_0^{2\pi} dr d\phi r e^{-r^2} = \pi. \quad (\text{A.5})$$

Por completeza, a integral Gaussiana pode ser estendida para n -ésima ordem: $I_n = \int_0^{\infty} dr r^n e^{-r^2} = \Gamma[(n+1)/2]/2$ [38], onde $\Gamma(\cdot)$ é a função gama dada pela Eq. A.7. Das Eqs. A.1, A.2 e A.5 obtemos a propriedade de soma da função erro com sua complementar.

$$\text{erf}(z) + \text{erfc}(z) = 1. \quad (\text{A.6})$$

A Eq. A.6 pode ser verificada graficamente, como ilustrado nos gráficos da Fig. A.1.

Figura A.1 – (a) Função erro. (b) Função erro complementar. Note que a soma das duas funções permanece constante, igual a 1, independentemente do valor de x .

A.2 Gama

A integral Gaussiana e a função erro estão associadas não apenas entre si, mas também com a função gama, cuja definição é dada pela integral [61, 91]

$$\Gamma(a) = \int_0^{\infty} dt t^{a-1} e^{-t}, \quad (\text{A.7})$$

onde $a > 0$. Esta função pode ser escrita em termos do logaritmo natural, por meio da substituição: $x = e^{-t}$. Os novos limites de integração são $x_0 = 1$ e $x_1 = 0$ e por sua vez $dx = -e^{-t} dt$, com isso a Eq. A.7 é reescrita:

$$\Gamma(a) = \int_0^1 dx \ln^{a-1} \left(\frac{1}{x} \right), \quad (\text{A.8})$$

realizando a substituição $y = \sqrt{x}$, temos $x = y^2$ e o diferencial $dx = 2ydy$, os limites de integração da Eq. A.8 continuam iguais, de modo que:

$$\Gamma(a) = 2 \int_0^1 dy y [-2 \ln(y)]^{a-1}. \quad (\text{A.9})$$

Para encontrar a relação com a integral Gaussiana ainda é necessária a substituição $y = e^{-z^2/2}$ que implica $dy = -ze^{-z^2/2}dz$ e $z^2 = -2 \ln(y)$ e a Eq. A.9 é reescrita

$$\Gamma(a) = 2 \int_0^\infty dz z^{2a-1} e^{-z^2}, \quad (\text{A.10})$$

cujas ligação com a Eq. A.3 é dada pelo caso especial $a = 1/2$, de modo que $\Gamma(1/2) = \sqrt{\pi}$. Integrando a Eq. A.7 por partes

$$\Gamma(a) = -[t^{a-1}e^{-t}]_0^\infty + \int_0^\infty dt (a-1)t^{a-2}e^{-t} = (a-1)\Gamma(a-1), \quad (\text{A.11})$$

integrando mais uma vez: $\Gamma(a) = (a-1)(a-2)\Gamma(a-2)$ e assim por diante. Se a for inteiro positivo então a função gama se reduz ao fatorial

$$\Gamma(a) = (a-1)!. \quad (\text{A.12})$$

Nessa dissertação a principal aplicação da função gama é o cálculo do volume de uma hipersfera d -dimensional $V_d = \pi^{d/2}r^d/\Gamma(1+d/2)$. O gráfico da Fig. A.2 ilustra o comportamento da função gama.

A relação da função erro com a função gama é obtida por meio de um caso especial dessa última, conhecido como função gama incompleta, definida como

$$\gamma(a, b) = \int_0^b dt t^{a-1} e^{-t}, \quad (\text{A.13})$$

substituindo $x = \sqrt{t}$ na Eq. A.13, resulta em $t = x^2$ e $dt = 2xdx$ e a função gama incompleta torna-se

$$\gamma(a, b) = 2 \int_0^{\sqrt{b}} dx x^{2a-1} e^{-x^2}. \quad (\text{A.14})$$

A relação com a função erro surge como o caso especial onde $a = 1/2$

$$\gamma(1/2, b) = 2 \int_0^{\sqrt{b}} dx e^{-x^2} = \sqrt{\pi} \text{erf}(\sqrt{b}). \quad (\text{A.15})$$

De modo análogo à função erro complementar, é possível definir a função gama incompleta complementar:

$$\Gamma(a, b) = \int_b^\infty dt t^{a-1} e^{-t}. \quad (\text{A.16})$$

Figura A.2 – Gráfico da função gama A.7. A função é definida sob todo o plano complexo, exceto nos inteiros negativos, onde há divergência para os valores $+\infty$ ou $-\infty$.

Pela propriedade de aditividade das integrais é claro que $\gamma(a, b) + \Gamma(a, b) = \Gamma(a)$. A função gama incompleta normalizada é definida como:

$$P(a, b) = \frac{\gamma(a, b)}{\Gamma(a)} \quad (\text{A.17})$$

e a função gama incompleta complementar normalizada

$$Q(a, b) = \frac{\Gamma(a, b)}{\Gamma(a)}. \quad (\text{A.18})$$

A relação entre essas duas últimas funções é $P(a, b) + Q(a, b) = [\gamma(a, b) + \Gamma(a, b)]/\Gamma(a) = 1$. As funções do tipo gama incompleta surgem no cálculo da função densidade de probabilidade acumulada para a distribuição gama generalizada ou Stacy, Eq. B.39.

A.3 Beta

Assim como a integral Gaussiana, a função beta está relacionada com a função gama. Partindo da definição dada pela Eq. A.7, o produto de duas funções gama pode ser escrito como

$$\Gamma(a)\Gamma(b) = \int_0^\infty \int_0^\infty dt dw e^{-(t+w)} t^{a-1} w^{b-1}, \quad (\text{A.19})$$

trocando as variáveis $t = xy$ e $w = x(1 - y)$ escrevemos o jacobiano

$$J = \frac{\partial(t, w)}{\partial(x, y)} = -x,$$

como $x > 0$, concluímos que $dt ds = |J| dx dy = x dx dy$. Como $t + w = x$, os limites de integração são $0 < x < \infty$ e $0 < y < 1$ e a Eq. A.19 pode ser escrita como

$$\Gamma(a)\Gamma(b) = \int_0^\infty dx e^{-x} x^{a+b-1} \int_0^1 dy y^{a-1} (1-y)^{b-1} \quad (\text{A.20})$$

A segunda integral da Eq. A.20 é a função Beta [61, 91]:

$$B(a, b) = \int_0^1 dt t^{a-1} (1-t)^{b-1}. \quad (\text{A.21})$$

A primeira integral da Eq. A.20 é a função gama, Eq. A.7, com argumento $a + b$, de forma que a função beta pode ser escrita como uma razão de funções gama

$$B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}, \quad (\text{A.22})$$

essa última representação é utilizada em estatística como o fator de normalização da distribuição beta [92]. Para a e b inteiros e não nulos a relação da função gama com o fatorial, Eq. A.12, permite escrever a Eq. A.22 em função do coeficiente binomial²

$$\frac{1}{B(a, b)} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} = \frac{(a+b-1)!}{(a-1)!(b-1)!} = (a+b-1) \binom{a+b-2}{a-1}. \quad (\text{A.23})$$

Caso $b \gg a$ a Eq. A.23 pode ser aproximada usando a fórmula de Stirling³

$$B(a, b) \approx \frac{\Gamma(a)}{b^a}. \quad (\text{A.24})$$

Graficamente, a função beta está ilustrada na Fig. A.3. Se a integral da Eq. A.21 não variar por todo intervalo de integração é necessário definir a função beta incompleta

$$B_z(a, b) = \int_0^z dt t^{a-1} (1-t)^{b-1}, \quad (\text{A.25})$$

assim como sua complementar

$$B'_z(a, b) = \int_z^1 dt t^{a-1} (1-t)^{b-1} \quad (\text{A.26})$$

---

²O coeficiente binomial é $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

³A fórmula de Stirling para o fatorial $n! \sim \sqrt{2\pi n} n^n e^{-n}$.

Figura A.3 – Comportamento da função beta em relação aos parâmetros a e b . Note que o aumento dos parâmetros leva a uma rápida diminuição do valor de $B(a, b)$.

conhecida como função beta incompleta complementar. A razão da Eq. A.26 pela Eq. A.21 é função beta incompleta normalizada

$$I_z(a, b) = \frac{B_z(a, b)}{B(a, b)}. \quad (\text{A.27})$$

que representa a função de distribuição acumulada da função densidade de probabilidade beta.

A.4 Digama e Poligama

As funções digama e poligama [61] surgem no cálculo da média e variância da distribuição log-gama, Eq. B.56. A função digama, é definida como a derivada do logaritmo natural da função gama:

$$\psi(x) = \frac{d}{dx} \ln \Gamma(x), \quad (\text{A.28})$$

sua representação integral para reais positivos é

$$\psi(x) = \int_0^\infty dx \left[\frac{e^{-t}}{t} - \frac{e^{-xt}}{1 - e^{-t}} \right]. \quad (\text{A.29})$$

Uma representação especial para x inteiro positivo é dada por

$$\psi(x) = -\gamma + \sum_{k=1}^{x-1} \frac{1}{k} = -\gamma + H_{x-1}, \quad (\text{A.30})$$

onde $\gamma = 0.5722 \dots$ é a constante de Euler-Mascheroni e H_n é o n -ésimo número harmônico. Para $n \gg 1$, $H_n \sim \gamma + \ln(n)$, desse modo a Eq. A.28 converge para $\psi(x) \sim \ln(x)$. Assim como a função gama, a função digama pode ser calculada sob todo o plano complexo, exceto para inteiros negativos, seu gráfico é ilustrado na Fig. A.4.

Figura A.4 – Comportamento da função digama. Assim como a função gama há divergência nos valores inteiros negativos.

A Eq. A.28 é generalizada para função Poligama [61], de ordem m ao aplicar a $(m + 1)$ -ésima derivada no logaritmo natural da função gama

$$\psi^{(m)}(x) = \frac{d^m}{dx^m} \psi(x) = \frac{d^{m+1}}{dx^{m+1}} \ln[\Gamma(x)], \quad (\text{A.31})$$

para $m = 0$ reobtemos a função digama, Eq. A.28. A função $\psi^{(1)}(x)$ é conhecida como função Trigama, para o caso especial onde $x \gg 1$ é inteiro $\psi^{(1)}(x) \sim 1/x$.

Distribuições de Probabilidades

Neste apêndice, apresentamos uma breve revisão das principais ferramentas estatísticas utilizadas nessa dissertação. Essa revisão é dividida em 2 partes: distribuições discretas e contínuas. A revisão sobre distribuições discretas é concisa e limita-se a descrever a distribuição de Poisson (utilizada na estatística de distâncias e vizinhanças) como limite da distribuição binomial. Nesta dissertação nos concentramos na segunda parte, onde é apresentada a distribuição de Amoroso [60], a partir da qual obtemos diversas outras distribuições como casos especiais.

B.1 Distribuições Discretas

B.1.1 Bernoulli

A distribuição discreta mais simples é aquela onde o espaço amostral é composto por dois resultados, tal como o lançamento de uma moeda que resulta em cara ou coroa. A probabilidade de sucesso está ligada à variável aleatória $x = 1$, dada por $P(x = 1) = p$ e de fracasso à variável aleatória $x = 0$ com probabilidade $P(x = 0) = 1 - p$. A média e variância valem, respectivamente $E(X) = p$ e $Var(X) = p(1 - p)$.

B.1.2 Binomial

É resultado da realização de n ensaios de Bernoulli. Os ensaios são independentes e a probabilidade de sucesso, p , permanece constante, sendo $1 - p$ a probabilidade de fracasso. O número de sucessos é representado por k , portanto

ocorrem $n - k$ fracassos, tal como a sequência a seguir:

$$pp...(k \text{ vezes})(1 - p)(1 - p) \cdots (n - k \text{ vezes}). \quad (\text{B.32})$$

Como as repetições são independentes, a probabilidade da sequência, na ordem específica da Eq. B.32 ser sorteada é $p^k(1 - p)^{n-k}$, porém essa mesma probabilidade é associada a qualquer outro resultado no qual ocorre k sucessos. O número total de combinações onde há k sucessos dentre n realizações é dado pelo coeficiente binomial⁴. A probabilidade de ocorrer k sucessos é o produto da Eq. B.32 e o coeficiente binomial

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad (\text{B.33})$$

com $k = 0, 1, 2, 3, \dots, n$. O valor esperado e a variância valem, respectivamente, $E(X) = np$ e $Var(X) = npq$. Os valores da esperança e da variância são os mesmos que na distribuição de Bernoulli, multiplicados pelo número de realizações do experimento.

B.1.3 Poisson

É o caso limite da distribuição Binomial que ocorre quando $n \rightarrow \infty$ e $p \rightarrow 0$ com o produto $\lambda = np$ constante. A função distribuição de probabilidade de Poisson é dada pela equação

$$P(x = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad (\text{B.34})$$

o parâmetro $\lambda = np$ é o valor esperado, a variância vale $\sigma^2(x) = \lambda$.

A motivação para entender o surgimento da distribuição de Poisson é considerar o exemplo: Um grande centro de atendimento ao consumidor recebe *e-mails* e, em um período de três horas, foram recebidos 300 *e-mails*. Suponha que a tarefa é calcular a probabilidade de serem recebidos $0, 1, 2, \dots$ *e-mails* nos próximos 5 minutos. Ao considerar o fenômeno da chegada de *e-mails* concluímos que qualquer instante é tão provável quanto os outros para a chegada de um *e-mail*. Assim a probabilidade de recebimento permanece constante em cada instante de tempo. Uma forma de tratar o problema é dividir o intervalo de 5 minutos em 10 intervalos de 30 segundos. Cada um desses intervalos pode ser tratado como uma prova de

---

⁴O coeficiente binomial $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

Bernoulli, durante a qual um *e-mail* pode ou não chegar. Entretanto observe que desconsideramos a probabilidade da chegada de mais de um *e-mail* nestes 30 segundos. A forma de calcular essa probabilidade é realizar a aproximação $n \rightarrow \infty$, $t \rightarrow 0$, na Eq. B.33, onde n representa o número de intervalos de tempo e t representa a duração do intervalo.

B.2 Distribuições Contínuas

A distribuição de Amoroso é representada por uma função densidade de probabilidade univariada, ou seja, é a pdf de apenas uma variável aleatória. A partir dessa distribuição são geradas outras famílias de pdfs como combinação, ou casos limite, dos quatro parâmetros que descrevem a distribuição de Amoroso, cuja pdf é

$$\text{Amoroso}(x|a, \theta, \alpha, \tau) = \frac{1}{\Gamma(\alpha)} \left| \frac{\tau}{\theta} \right| \left(\frac{x-a}{\theta} \right)^{\alpha\tau-1} \exp \left[- \left(\frac{x-a}{\theta} \right)^\tau \right]. \quad (\text{B.35})$$

A Eq. B.35 apresenta quatro parâmetros reais, sendo que o parâmetro a é o parâmetro de localização, θ é de dispersão, α e τ são os parâmetros de forma. Note que a grandeza $Z = (x-a)/\theta$ tem o mesmo papel que um escore Z em uma pdf Gaussiana. No texto que segue, obtemos as pdfs oriundas da Eq. B.35 que surgem no problema da estatística de distâncias. Para uma revisão mais detalhada, recomendamos a leitura da Ref. [60].

B.2.1 Stacy - Gama Generalizada

A distribuição de Stacy é o caso particular da Eq. B.35 no qual o parâmetro de localização a é nulo e todos os outros parâmetros são positivos⁵ Ref. [59]. A pdf resultante desta transformação é a distribuição de Stacy:

$$\text{Stacy}(x; \theta, \alpha, \tau) = \frac{1}{\Gamma(\alpha)} \left(\frac{\tau}{\theta} \right) \left(\frac{x}{\theta} \right)^{\alpha\tau-1} \exp \left[- \left(\frac{x}{\theta} \right)^\tau \right], \quad (\text{B.36})$$

onde $x > 0$. O primeiro e segundo momentos, em relação a origem, valem respectivamente:

$$\mu_1 = \frac{1}{\Gamma(\alpha)} \left(\frac{\tau}{\theta} \right) \frac{1}{(\theta)^{\alpha\tau-1}} \int_0^\infty dx x^{\alpha\tau} \exp \left[- \left(\frac{x}{\theta} \right)^\tau \right] = \frac{\theta}{\Gamma(\alpha)} \Gamma \left(\alpha + \frac{1}{\tau} \right),$$

⁵Existe a possibilidade do parâmetro τ , Eq. B.36, ser negativo. Para mais detalhes sobre este caso consulte a Ref. [60].

$$\mu_2 = \frac{1}{\Gamma(\alpha)} \left(\frac{\tau}{\theta} \right) \frac{1}{(\theta)^{\alpha\tau-1}} \int_0^\infty dx x^{\alpha\tau+1} \exp \left[- \left(\frac{x}{\theta} \right)^\tau \right] = \frac{\theta^2}{\Gamma(\alpha)} \Gamma \left(\alpha + \frac{2}{\tau} \right).$$

Dos dois primeiros momentos calculamos a variância da distribuição de Stacy

$$\sigma^2 = \frac{\theta^2}{\Gamma(\alpha)} \left[\Gamma \left(\alpha + \frac{2}{\tau} \right) - \frac{1}{\Gamma(\alpha)} \Gamma^2 \left(\alpha + \frac{1}{\tau} \right) \right]. \quad (\text{B.37})$$

Os momentos de ordem superior valem

$$\mu_n = \frac{\theta^n}{\Gamma(\alpha)} \Gamma \left(\alpha + \frac{n}{\tau} \right). \quad (\text{B.38})$$

A influência de cada parâmetro no formato da distribuição está ilustrada nos gráficos das Figs. B.5 e B.6.

Figura B.5 – Distribuição de Stacy com diversos valores dos parâmetros α e θ . (a) O aumento de α faz com que a assimetria da curva diminua. Note que com seu aumento a curva aproxima-se de uma normal. (b) Assim como na distribuição normal, um aumento na dispersão, θ , resulta em uma curva com maior variância.

Note que com o aumento de α há uma diminuição da assimetria da curva. Empiricamente, observamos que para α grande, a distribuição aparenta convergir para a normal. Ao aumentar θ , o desvio padrão é aumentado. A variância diminui quando o valor de τ é incrementado, aparentemente a distribuição torna-se uma sequência delta.

O valor da função distribuição acumulada pode ser calculado, integrando a Eq. B.36 até um valor arbitrário z , isto é,

$$F(z) = 1 - \frac{\Gamma(\alpha, (z/\theta)^\tau)}{\Gamma(\alpha)} = 1 - Q \left[\alpha, \left(\frac{z}{\theta} \right)^\tau \right], \quad (\text{B.39})$$

Figura B.6 – Distribuição de Stacy em função do parâmetro τ . A medida que τ aumenta, o desvio padrão torna-se cada vez menor e a função aparentemente aproxima-se de uma sequência delta de Dirac.

onde $Q(a, b)$ é a função gama incompleta complementar normalizada Eq. A.18, que guarda a seguinte relação com a função gama incompleta normalizada Eq. A.17: $Q(a, b) = 1 - P(a, b)$. A função acumulada, Eq. B.39, pode ser reescrita como

$$F(z) = P \left[\alpha, \left(\frac{z}{\theta} \right)^\tau \right]. \quad (\text{B.40})$$

B.2.1.1 Exponencial

A distribuição exponencial é o caso particular da distribuição de Stacy, Eq. B.36, no qual $\alpha = \tau = 1$. O surgimento da distribuição exponencial ocorre naturalmente na descrição do tempo ou espaço entre ocorrências de dois eventos gerados por um processo Poissônico, tal como o tempo entre decaimentos radiativos consecutivos e chamadas telefônicas consecutivas [92]. A pdf da distribuição exponencial é

$$\exp(x) = S(x; \theta, 1, 1) = \frac{1}{\theta} \exp \left[\frac{-x}{\theta} \right], \quad (\text{B.41})$$

onde $x \geq 0$ e $\theta > 0$. O valor esperado e a variância da distribuição exponencial são, $\mu = \theta$ e $\sigma^2 = \theta^2$, respectivamente. A função de distribuição acumulada vale

$$F(x) = 1 - e^{-\theta x}.$$

B.2.1.2 Gama

A distribuição gama é o caso particular da distribuição de Stacy, Eq. B.36, no qual $\tau = 1$. Quando α é inteiro, essa distribuição descreve o tempo ou espaço entre ocorrências de eventos não consecutivos gerados por um processo Poissônico. A pdf da distribuição gama é

$$\text{Gama}(x) = S(x; \theta, \alpha, 1) = \frac{x^{\alpha-1}}{\Gamma(\alpha)\theta^\alpha} \exp\left[\frac{-x}{\theta}\right] \quad (\text{B.42})$$

onde $x > 0$, $\theta > 0$ e $\alpha > 0$. A média e a variância da distribuição gama são dadas por: $\mu = \alpha\theta$ e $\sigma^2 = \alpha\theta^2$, respectivamente. A função de distribuição acumulada vale

$$F(x) = \frac{\gamma(\alpha, x/\theta)}{\Gamma(\alpha)} = P[\alpha, z], \quad (\text{B.43})$$

onde $P(a, b)$ é a função gama incompleta normalizada, Eq. A.17.

B.2.1.3 Weibull

A distribuição de Weibull é o caso especial de Stacy, Eq. B.36, no qual $\alpha = 1$. A pdf da distribuição é dada por

$$\text{Weibull}(x) = S(x; \theta, 1, \tau) = \frac{\tau}{\theta^\tau} x^{\tau-1} \exp\left[-\left(\frac{x}{\theta}\right)^\tau\right], \quad (\text{B.44})$$

onde $x \geq 0$, $\tau > 0$ e $\theta > 0$. Considere um objeto composto de muitas partes, suponha que esse objeto falha quando um de seus componentes falha, nessas circunstâncias a distribuição do tempo de vida de tal objeto é dada pela distribuição de Weibull [92]. A média e variância valem respectivamente,

$$\mu = \theta \left(1 + \frac{1}{\tau}\right), \quad (\text{B.45})$$

e

$$\sigma^2 = \theta^2 \left[\Gamma\left(1 + \frac{2}{\tau}\right) - \Gamma^2\left(1 + \frac{1}{\tau}\right) \right]. \quad (\text{B.46})$$

A função de distribuição acumulada vale $F(x) = 1 - e^{-(x/\theta)^\tau}$.

B.2.1.4 Qui Quadrado - χ^2

A distribuição do qui quadrado é o caso especial da distribuição de Stacy, Eq. B.36, no qual $\theta = 2$, $\alpha = k/2$ e $\tau = 1$. Se $X_1, X_2, X_3, \dots, X_k$ são variáveis aleatórias independentes normalmente distribuídas, com média $\mu = 0$ e variância unitária, a variável $Y = X_1^2 + X_2^2 + X_3^2 + \dots + X_k^2$ segue a distribuição do χ^2 com k graus de liberdade. A pdf da distribuição do χ^2 é dada por

$$\chi^2 = S(x; 2, k/2, 1) = \frac{1}{2\Gamma(k/2)} \left(\frac{x}{2}\right)^{k/2-1} \exp\left[-\frac{x}{2}\right], \quad (\text{B.47})$$

onde $x > 0$ e $k > 0$. A média e a variância da distribuição valem respectivamente $\mu = k$ e $\sigma^2 = 2k$. A função de distribuição acumulada até um valor z pode ser escrita como

$$F(x) = \frac{\gamma(k/2, z/2)}{\Gamma(k/2)} = P\left(\frac{k}{2}, \frac{z}{2}\right), \quad (\text{B.48})$$

onde $P(a, b)$ é a função gama incompleta normalizada, Eq. A.17. A pdf dada pela Eq. B.47 é caso particular da distribuição gama B.42 com $\alpha = k/2$ e $\theta = 2$. Para $k \gg 1$ a distribuição converge para a pdf Gaussiana devido ao teorema central do limite.

B.2.1.5 Rayleigh

A distribuição de Rayleigh é o caso especial da distribuição de Stacy, Eq. B.36, na qual $\theta = \sqrt{2}\sigma^2$, $\alpha = 1$ e $\tau = 2$. A variável aleatória $R = \sqrt{X^2 + Y^2}$, onde $X \sim N(0, \sigma^2)$ e $Y \sim N(0, \sigma^2)$, segue a distribuição de Rayleigh caso X seja independente de Y .

A distribuição de velocidade do vento apresenta uma distribuição muito próxima à distribuição de Rayleigh, dado que as componentes horizontais seguem uma distribuição aproximadamente normal com média nula, enquanto as componentes verticais podem ser desprezadas [60]. A pdf da distribuição é dada por

$$\text{Rayleigh}(x) = S(x; \sqrt{2}\sigma^2, 1, 2) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right), \quad (\text{B.49})$$

onde $0 \leq x < \infty$ e $\sigma > 0$. A média e a variância da distribuição valem, respectivamente,

$$\mu = \sigma\sqrt{\frac{\pi}{2}}, \quad (\text{B.50})$$

e

$$s^2 = \frac{4 - \pi}{2} \sigma^2. \quad (\text{B.51})$$

A função de distribuição acumulada é escrita como $F(x) = 1 - e^{-x^2/2\sigma^2}$.

B.2.1.6 Maxwell-Boltzmann

A distribuição de Maxwell-Boltzmann é o caso especial da distribuição de Stacy, Eq. B.36, na qual $\theta = \sqrt{2\sigma^2}$, $\alpha = 3/2$ e $\tau = 2$. A variável aleatória $R = \sqrt{X_1^2 + X_2^2 + X_3^2}$, onde $X_i \sim N(0, \sigma^2)$, segue a distribuição de Maxwell-Boltzmann. Esta distribuição descreve as velocidades das moléculas de um gás ideal em equilíbrio termodinâmico. A pdf da distribuição Maxwell-Boltzmann é

$$\text{Maxwell}(x) = S(x; \sqrt{2\sigma^2}, 3/2, 2) = \frac{\sqrt{2}}{\sqrt{\pi}\sigma^3} x^2 \exp\left(-\frac{x^2}{2\sigma^2}\right), \quad (\text{B.52})$$

onde $0 \leq x < \infty$ e $\sigma > 0$. A média e a variância valem, respectivamente,

$$\mu = 2\sigma\sqrt{\frac{2}{\pi}}, \quad (\text{B.53})$$

e

$$s^2 = \frac{\sigma^2(3\pi - 8)}{\pi}. \quad (\text{B.54})$$

A função de distribuição acumulada dada por

$$F(x) = \text{erf}\left(\frac{x}{\sqrt{2}\sigma}\right) - \frac{2}{\pi} \frac{x}{\sigma} e^{-x^2/2\sigma^2}, \quad (\text{B.55})$$

onde $\text{erf}(\cdot)$ é a função erro definida pela Eq. A.1.

B.2.2 Log-Gama

A distribuição Log-Gama é uma distribuição contínua, composta de 3 parâmetros reais e univariada, com domínio $-\infty < x < \infty$. Sua definição matemática é

$$\text{LogGama}(x; \nu, \lambda, \alpha) = \frac{1}{|\lambda|\Gamma(\alpha)} \exp\left[\alpha\left(\frac{x - \nu}{\lambda}\right) - \exp\left(\frac{x - \nu}{\lambda}\right)\right]. \quad (\text{B.56})$$

Os três parâmetros reais consistem de um parâmetro de localização ν , um parâmetro de dispersão λ e um parâmetro de forma α . A distribuição Log-Gama padrão é o caso especial da Eq. B.56, onde $\nu = 0$ e $\lambda = 1$

$$\text{StdLogGama}(x; \alpha) = \frac{1}{\Gamma(\alpha)} \exp[\alpha x - \exp(x)], \quad (\text{B.57})$$

a denominação Log-Gama surge pois a Eq. B.57 é a transformação logaritmica da distribuição gama padrão:

$$\text{StdLogGama}(\alpha) \approx \ln[\text{StdGama}], \text{ onde } \text{StdGama} = \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x}.$$

Assim como a distribuição gama generalizada, a distribuição Log-Gama é obtida diretamente da distribuição de Amoroso sendo o limite da Eq. B.35 no qual:

$$\lim_{\tau \rightarrow \infty} \text{Amoroso}_{(\nu-\tau\lambda, \tau\lambda, \alpha, \tau)}(x) = \text{LogGama}_{(\nu, \lambda, \alpha)}(x).$$

A média e a variância da Eq. B.56 valem respectivamente

$$\mu = \nu + \lambda\psi(\alpha) \quad (\text{B.58})$$

$$\sigma^2 = \lambda^2\psi^{(1)}(\alpha), \quad (\text{B.59})$$

onde a média é escrita em termos da função Digama, Eq. A.28, e a variância é escrita em termos da função Poligama, Eq. A.31.

A função de distribuição acumulada da Eq. B.56 vale

$$F(x) = \begin{cases} 1 - Q(\alpha, e^{x-\nu/\lambda}) & \text{para } \lambda > 0, \\ Q(\alpha, e^{x-\nu/\lambda}) & \text{para } \lambda < 0, \end{cases} \quad (\text{B.60})$$

onde $Q(a, b) = \Gamma(a, b)/\Gamma(a)$ é a função gama incompleta complementar normalizada, Eq. A.17. O comportamento da distribuição Log-Gama em relação aos parâmetros α e λ está ilustrado nos gráficos da Fig. B.7. A distribuição para qual a Log-Gama converge quando $\alpha \rightarrow \infty$ na Eq. B.56 é uma Normal, i. e.,

$$\text{Normal}(x) = \lim_{\alpha \rightarrow \infty} \text{LogGama}_{(\mu-\sigma\sqrt{\alpha}\ln(\alpha), \sigma\sqrt{\alpha}, \alpha)}(x). \quad (\text{B.61})$$

B.2.2.1 Normal - Gaussian

Além de ser o limite da distribuição log-gama ela descreve uma série de fenômenos físicos/sociais, além de apresentar aplicações na inferência estatística. A distribuição de altura dos indivíduos de uma população segue a curva Gaussiana, por exemplo. A pdf da distribuição normal é

$$\text{Normal}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(x-\mu)^2}{2\sigma^2} \right], \quad (\text{B.62})$$

Figura B.7 – Distribuição de log-gama em funções dos parâmetros α e λ . (a) O aumento de α diminui a dispersão em torno da média e modifica a forma da curva. (b) O aumento do parâmetro λ torna a curva mais dispersa.

onde $-\infty < x < \infty$ e $\sigma > 0$, μ é a média e σ desvio padrão. Existe uma mudança de variável que transforma qualquer distribuição normal em uma normal padrão, essa transformação leva de X para Z via $Z = (X - \mu)/\sigma$ que substituída na Eq. B.62 resulta:

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{z^2}{2} \right]. \quad (\text{B.63})$$

B.2.2.2 Gumbel Generalizada

A distribuição de Gumbel generalizada é o caso especial da distribuição log-gama Eq. B.56, na qual os parâmetros $\nu = u + \lambda \ln n$, $\lambda = -\lambda$ e $\alpha = n$. Esta distribuição surge para descrever estatísticas de extremos, quando o parâmetro λ é positivo há a descrição de valores máximos, se o λ for negativo há a distribuição de valores mínimos. A pdf da distribuição é

$$\text{GenGumbel}(x) = S(x; u + \lambda \ln n, -\lambda, n) = \frac{n^n}{|\lambda| \Gamma(n)} \exp \left[-n \left(\frac{x - \nu}{\lambda} \right) - n \exp \left(\frac{x - \nu}{\lambda} \right) \right]. \quad (\text{B.64})$$

A média e a variância da distribuição Gumbel generalizada valem, respectivamente, $\mu = u + \lambda \ln(n) - \lambda \psi(n)$ e $\sigma^2 = \lambda \psi_1(n)$. Quando $n = 1$ a Eq. B.64 torna-se a distribuição de Gumbel.

Para finalizar esse apêndice, apresentamos a Tab. B.1 que resume das distribuições que são casos particulares da distribuição de Stacy. A pdf da distribuição log-gama não se encontra na tabela por ser caso limite da distribuição de Amoroso.

Tabela B.1 – Tabela resumo das distribuições que são casos particulares da pdf Gama Generalizada/Stacy. O símbolo (-) indica valor arbitrário do parâmetro em questão.

Distribuição	α	τ	θ
Exponencial	1	1	-
Gama	-	1	-
Weibull	1	-	-
Qui quadrado	$k/2$	1	2
Rayleigh	1	2	$\sqrt{2}$
Maxwell	$3/2$	2	$\sqrt{2\sigma^2}$

Caminhada Aleatória

Neste apêndice, apresentamos resultados mais aprofundados referentes à caminhada aleatória. No Cap. 4, que aborda modelos de caminhada, apenas a distância quadrática média foi calculada. Nesse compêndio, visitamos a caminhada de um ponto de vista mais geral por meio do uso de ferramentas matemáticas, tal como transformada de Fourier e teoremas relacionados. O uso dessas ferramentas possibilita, não apenas calcular grandezas da caminhada, mas também a derivar teoremas fundamentais, tal como o Teorema Central do Limite de Gauss.

C.1 Formulação do Modelo

Nosso ponto de partida é o termo caminhada aleatória, que surgiu quando Karl Pearson lançou a seguinte questão na revista *Nature* [1]: “*Can any of your readers refer me to a work wherein I should find a solution of the following problem, or failing the knowledge of any existing solution provide me with an original one? I should be extremely grateful for aid in the matter. A man starts from a point O and walks l yards in a straight line; he then turns through any angle whatever and walks another l yards in a second straight line. He repeats this process n times. I require the probability that after n stretches he is at a distance between r and $r + \delta r$ from his starting point O .*”

Essa questão tinha como motivação o movimento de animais, contudo suas aplicações em física e química vão muito além disso: difusão simples, reações químicas, distância ponta-a-ponta de cadeias poliméricas e assim por diante. Na mesma edição da revista *Nature* [1] Rayleigh responde à Pearson e a expressão no limite

$N \gg 1$ é a distribuição Gaussiana, como veremos ao decorrer da seção.

C.2 Caso unidimensional

Considere um caminhante partindo do ponto 0 em uma reta que contém $2N + 1$ sítios igualmente espaçados por uma distância l , assim como ilustrado na Fig. C.8. A probabilidade desse caminhante ir para direita ou esquerda vale, respectivamente p e $q = 1 - p$. Considere a expressão $pe^{i\theta} + qe^{-i\theta}$, o coeficiente precedendo cada termo exponencial é a probabilidade de um passo ser dado para direita ou esquerda, respectivamente. Elevando a expressão ao quadrado temos $p^2e^{2i\theta} + 2pq + q^2e^{-2i\theta}$, note que os coeficientes dão a probabilidade do caminhante dar dois passos para a direita, uma para direita e outro para esquerda e dois para esquerda. Esse resultado

Figura C.8 – Caminhada aleatória unidimensional. Os sítios são separados por uma distância l e a probabilidade de caminhar um passo ser dado para a esquerda ou direita vale, respectivamente, p e $q = 1 - p$.

pode ser generalizado para qualquer valor de n e encontramos que a probabilidade de um caminhante estar em um sítio j é dada pelo coeficiente do termo $e^{i\theta j}$ da expansão de $(pe^{i\theta} + qe^{-i\theta})^n$. Essa probabilidade pode ser obtida pela da transformada inversa de Fourier:

$$P_n(j) = \frac{1}{2\pi} \int_{-\pi}^{\pi} (pe^{i\theta} + qe^{-i\theta})^n e^{-i\theta j} d\theta. \quad (\text{C.65})$$

A expressão $\lambda(\theta) = e^{i\theta} + qe^{-i\theta}$ é a função característica do deslocamento do caminhante em um passo. Como os passos são aleatórios, independentes e identicamente distribuídos, a função característica da soma de n passos é o produto das funções características individuais, ou seja, $\lambda^n(\theta)$. Para o caso $p = q = 1/2$ a probabilidade de encontrar o caminhante no sítio j vale:

$$P_n(j) = \frac{1}{2^{n+1}} \int_{-\pi}^{\pi} (e^{i\theta} + e^{-i\theta})^n e^{-i\theta j} d\theta = \frac{1 + (-1)^{n+j}}{2^{n+1}} \binom{n}{\frac{n+j}{2}}. \quad (\text{C.66})$$

Solução análoga, e usualmente abordada na literatura [93], é obtida por meio da distribuição binomial, onde o sucesso é associado a um passo para direita com probabilidade p e o fracasso é associado com um passo para a esquerda, com probabilidade $q = 1 - p$. Neste caso, o resultado obtido é a distribuição do número de passos para a esquerda ou direita, ao invés da posição como na Eq. C.66.

C.3 Caso Contínuo

No caso contínuo o caminhante escolhe o deslocamento em cada passo a partir de uma função densidade de probabilidade (pdf) $f(x)$. A caminhada no meio contínuo é conhecida como movimento Browniano [70], em homenagem ao botânico Brown, que em 1828 observou o movimento errático dos grãos de pólen [2]. Em nossa concepção atual, associamos a caminhada aleatória no substrato contínuo como o limite de escala do meio discreto. Por limite de escala entende-se que o tamanho do passo da rede regular torna-se tão pequeno quanto se queira, de modo que o discreto se confunde com o contínuo. Agora, a função característica do deslocamento para cada passo dado é representada por uma integral:

$$\lambda(k) = \langle e^{ikx} \rangle = \int_{-\infty}^{\infty} dx e^{ikx} f(x), \quad (\text{C.67})$$

onde k é a frequência espacial. Assim como no caso discreto, a Eq. C.67 é a equação transformada de Fourier de $f(x)$. A pdf do deslocamento no primeiro passo é dada exatamente por $P_1(x) = p(x)$. Se cada passo for independente dos demais, a seguinte relação vale [92]:

$$P_n(x) = \int_{-\infty}^{\infty} dy P_n(y) p(x - y) \quad (\text{C.68})$$

que é a probabilidade caminhante ir, no n -ésimo passo, da posição y , atingida no $n - 1$ -ésimo passo, para a posição x . A Eq. C.68 é a soma de duas variáveis aleatórias independentes, ou seja, é uma convolução. A transformada de Fourier da Eq. C.68 vale $P_n(k) = P_{n-1}(k)\lambda(k) = P_{n-2}(k)\lambda^2(k) \cdots \lambda^n(k)$. Como $P_n(k)$ é a transformada de Fourier da pdf do deslocamento após n passos, $P_n(x)$ é obtido via a transformada inversa:

$$P_n(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk \lambda^n(k) e^{-ikx}. \quad (\text{C.69})$$

Como exemplo, considere o caso onde o comprimento do passo dado é distribuído segundo a pdf Gaussiana, com $\mu = 0$ e variância σ^2 . A primeira etapa para encontrar a distribuição de distância do caminhante, em relação à origem, é calcular a função característica:

$$P_n(k) = \lambda^n(k) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} dx e^{ikx - x^2/2\sigma^2}, \quad (\text{C.70})$$

que é resolvida completando o expoente da seguinte forma: $ikx - x^2/2\sigma^2 = (x/\sqrt{2\sigma^2} - ik\sqrt{2\sigma^2}) - 2k^2\sigma^2$ e $\lambda(k) = e^{-\sigma^2 k^2/2}$. Nesse caso, $P_n(k) = \lambda^n(k) = e^{-n\sigma^2 k^2/2}$, de maneira que a distribuição de distâncias vale (por meio da transformada inversa de Fourier):

$$P_n(x) = \frac{1}{\sqrt{2\pi n\sigma^2}} \exp\left(\frac{-x^2}{2n\sigma^2}\right) \quad (\text{C.71})$$

A função característica da distribuição também é a função geradora de momentos. O m -ésimo momento de uma distribuição $f(x)$ é dado por:

$$\mu_m = \langle x^m \rangle = \int_{-\infty}^{\infty} dx x^m f(x), \quad (\text{C.72})$$

expandindo a exponencial da definição da função característica em série de Taylor:

$$\lambda(k) = \int_{-\infty}^{\infty} dx f(x) \left[1 + ikx - \frac{k^2 x^2}{2} + \frac{ik^3 x^3}{6} + \dots \right] = \sum_{n=0}^{\infty} i^n \frac{\mu_n}{n!} k^n, \quad (\text{C.73})$$

de modo que, se o m -ésimo momento existir ele será dado por:

$$\mu_m = \langle x^m \rangle = (-i)^m \left. \frac{d^m \lambda(k)}{dk^m} \right|_{k=0}. \quad (\text{C.74})$$

Outra forma de notar que $P_n(x) = \lambda^n(k)$ é pela própria definição da função característica. O deslocamento do caminhante no primeiro, segundo, ... n -ésimo passo é a soma dos deslocamentos $X_n = x_1 + x_2 + \dots + x_n$, de forma que a função característica é dada por:

$$P_n(k) = \langle e^{ik(x_1 + x_2 + \dots + x_n)} \rangle = \langle e^{ikx_1} \rangle \langle e^{ikx_2} \rangle \dots \langle e^{ikx_n} \rangle = \lambda^n(k), \quad (\text{C.75})$$

pois cada deslocamento é independente dos demais, assim obtemos uma nova derivação da Eq. C.68.

C.4 Teorema Central do Limite a Partir da Caminhada Aleatória

A função densidade de probabilidade do deslocamento do caminhante após n passos é dado pela transformada inversa de Fourier da função característica:

$$P_n(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk P_n(k) e^{-ikx} = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk [\lambda(k)]^n e^{-ikx}, \quad (\text{C.76})$$

onde o comportamento de $P_n(x)$ para grandes valores de x é dado pelo comportamento de $P_n(k)$ em valores pequenos de k . Como exemplo, considere o caso discreto onde $\lambda(k) = \cos(k)$. Em uma expansão até segunda ordem $\cos(k) \approx 1 - k^2/2 \dots$, tem-se:

$$P_n(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk \left(1 - \frac{k^2}{2}\right)^n e^{-ikx} = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk \exp \left[n \ln \left(1 - \frac{k^2}{2}\right) \right] e^{-ikx}, \quad (\text{C.77})$$

a série para o logaritmo é dada por:

$$\log(1 - x) = \sum_{n=1}^{\infty} \frac{x^n}{n}, \quad \text{para } |x| < 1, \quad (\text{C.78})$$

de forma que a função de distribuição do deslocamento do caminhante vale:

$$P_n(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk \exp \left(\frac{-nk^2}{2} \right) e^{-ikx} = \frac{1}{\sqrt{2\pi n}} e^{-x^2/2n} \quad (\text{C.79})$$

Considere que a distribuição de distância em cada passo é dada pela pdf de Laplace: $f(x) = (a/2)e^{-a|x|}$. A função característica vale:

$$\lambda_k = \frac{a}{2} \int_{-\infty}^{\infty} e^{a|x|+ikx} = \frac{a^2}{a^2 + k^2}, \quad (\text{C.80})$$

Para valores pequenos de k , $\lambda(k) \approx 1 - (k/a)^2 + \dots$. Fazendo a transformada inversa de Fourier, obtêm-se:

$$P_n(x) \approx \frac{a}{\sqrt{2\pi n}} e^{-a^2 x^2 / 2n}. \quad (\text{C.81})$$

Sempre que o segundo momento existir em uma caminhada e a função característica for escrita como $\lambda_k = 1 - k\sigma^2/2 + \dots$, a função densidade de probabilidade que representa o deslocamento do caminhante após n passos (n grande) converge para a distribuição Gaussiana.

Computação em Paralelo Usando MPI

Neste apêndice apresentamos uma breve introdução à computação paralela utilizando MPI. Utilizamos o modelo de paralelismo inerente ao MPI para paralelizar o algoritmo de simulação Monte Carlo da SAW. Inicialmente, as principais motivações históricas que levaram ao desenvolvimento da computação paralela são apresentadas. Na sequência as características de *hardware*, tal como memória principal e CPU, essenciais para entender como um computador paralelizável opera são tratadas. Com as motivações históricas e os conceitos de hardware, avançamos no estudo da interface de passagem de mensagens (MPI), apresentando as principais funções necessárias para construir um código paralelo. Ao final do apêndice, abordamos algumas métricas para análise do desempenho de um programa paralelo. Salientamos que este apêndice não constitui um texto de referência na área de computação paralela, mas um guia que apresenta os principais conceitos e ferramentas utilizadas nessa dissertação.

D.1 Surgimento da Computação Paralela

Em Física, há duas abordagens tradicionais para o tratamento de um problema. A primeira delas é a abordagem teórica, que usa principalmente ferramentas matemáticas para modelar sistemas reais, ou que tenham sentido físico. A segunda é utilizar observação/experimentação para descrever fenômenos que acontecem naturalmente ou por meio da ação humana. O desenvolvimento dessas duas vertentes, em conjunto com o aperfeiçoamento da metodologia científica, propiciou a criação

de equipamentos cada vez mais complexos, tal como o computador digital em 1943. O advento da computação digital remonta ao final da segunda guerra mundial, pois o esforço de guerra favoreceu o desenvolvimento de computadores para fins militares. Essas máquinas primitivas usavam circuitos com válvulas eletrônicas, ocupavam grandes espaços e dissipavam muita energia. Com a invenção do transistor, em 1947, e com a criação de circuitos integrados, que empacotam estes transistores em um único *chip*, o desenvolvimento de computadores cada vez mais rápidos e de custo acessível tornou-se possível. Nos dias atuais um simples *smartphone* é capaz de executar *softwares* complexos que exigem grande poder de processamento.

Esse incremento da capacidade computacional trouxe uma nova ferramenta para a ciência que permite realizar cálculos em um curto intervalo de tempo, o que alavancou avanços consideráveis em Física/Química com a possibilidade de simular um experimento por meio de um modelo matemático simplificado ou muito realista. Nesse contexto, há ao menos três razões para utilizar simulação computacional [94] quando:

- realizar o experimento é inviável devido à aspectos relacionados ao custo, inviabilidade de tecnologia, perigo, etc. Nesses casos, simulações podem gerar resultados satisfatórios capazes de prever ou descrever fenômenos;
- caso o experimento seja executado, mas os resultados não são completamente entendidos. Nesse caso simulações, que são modelos do experimento, podem auxiliar na compreensão dos resultados;
- a simulação é realizada em paralelo ao experimento, para diminuir o espaço de busca ou guiar o experimento de maneira mais adequada.

O uso de simulações computacionais é viável quando duas regras são obedecidas: os dados que representam os resultados são acurados e precisos e o tempo da simulação é praticável. A última condição é de especial importância, pois contribuiu para o impulso de uma área da Ciência da Computação conhecida como *high performance computing* (HPC) [94]. Apesar de máquinas vendidas para o grande público serem capazes de realizar 10^{10} operações por segundo, para propósitos científicos, máquinas com maior performance são exigidas.

Uma das características primordiais da supercomputação é utilizar *paralelismo*. Seu uso é desafiador para os desenvolvedores de *software* científico, dado que este deve explicitar o paralelismo por meio de uma interface de programação. No caso de CPUs as interfaces mais conhecidas são MPI e OpenMp [95]. Neste apêndice, tratamos de alguns aspectos gerais, ligados ao hardware, antes de avançar para o estudo de MPI.

D.2 Arquitetura de Hardware

Antes de entender MPI é necessário estudar conceitos que envolvem arquitetura de memória e organização de uma máquina paralela no nível de *hardware*.

Do ponto de vista arquitetural, a memória RAM⁶, do inglês *random access memory*, é a memória volátil onde as variáveis e estruturas de dados de um programa são temporariamente armazenadas [94]. Em um computador paralelo a memória RAM pode ser compartilhada ou distribuída. Um computador paralelo com memória compartilhada é definido como um sistema no qual um número de CPUs opera em um espaço de endereçamento comum. Há duas configurações para esse tipo de acesso à memória:

- **UMA** (*uniform memory access*): o tempo de acesso dos processadores à memória principal é essencialmente o mesmo;
- **NUMA** (*non-uniform memory access*): a memória principal está fisicamente distribuída através dos processadores, contudo circuitos de integração garantem que o conjunto de memórias seja vista como um único banco de memória. O tempo de acesso aos dados pode ser variável, dependendo da posição do banco de memória. Se for adjacente ao processador, o tempo é menor.

Um computador paralelo com memória distribuída é aquele onde as CPUs possuem espaços de endereçamento distintos [82]. Um exemplo típico dessa configuração é um *cluster beowulf*, onde computadores individuais são interconectados entre si formando uma rede, com um computador atuando como mestre. Cada computador da rede possui seu próprio banco de memória, o qual não é compartilhado com os demais.

---

⁶A memória RAM também é conhecida como memória principal.

Além da memória, alguns aspectos ligados às CPUs e sua organização física e lógica são relevantes. Em termos muito simplificados, uma CPU é composta por 2 unidades: a unidade de controle, que controla o fluxo de dados e as operações que serão realizadas e a unidade lógica e aritmética, que realiza operações lógicas e matemáticas.

Um conjunto de CPUs no mesmo circuito integrado é base dos processadores modernos, os quais contam com muitas CPUs empacotadas no mesmo *chip*⁷. Esses processadores compartilham a mesma memória RAM, ou seja, trabalham em um espaço de endereçamento de memória compartilhado. Os diferentes programas que são executados em uma mesma máquina são chamados de *processos*. Quando um mesmo programa é decomposto em várias linhas de execução paralelas, diz-se que o programa é composto por várias *linhas de execução* ou *threads*.

A equivalência entre a divisão lógica de um programa e de um *chip* muitas vezes não é 1:1, uma CPU pode executar mais de um processo ou mais de uma *thread* ao mesmo tempo, nesse caso a CPU “alterna” entre um processo ou *thread* e os outros.

D.3 MPI

A principal dificuldade de programar em paralelo é dividir a tarefa em diferentes partes que podem ser executadas simultaneamente em processos distintos. MPI é uma biblioteca de rotinas que supre as funcionalidades necessárias para que as partes constituintes da tarefa possam se comunicar. A sigla MPI é abreviação para *message passing interface* (interface de passagem de mensagens), assim como sugerido pelo nome, a comunicação entre processos é realizada por meio de troca de mensagens. O papel do programador é especificar como a tarefa será dividida e como será a troca de mensagens. O MPI apresenta boa performance e portabilidade, pode ser implementado em ambas arquiteturas de memória (distribuída e compartilhada), o que permite a utilização em *clusters* de computadores ou em um computador caseiro [96]. Apesar da interface ser definida (nome e comportamento das funções rigorosamente definidos), há várias implementações de bibliotecas que

---

⁷Atualmente há de 2 – 8 CPUs no mesmo *chip*.

permitem utilizar o MPI, as principais são *Openmpi* e *mpich*, ambas disponibilizam o MPI para a linguagem C e FORTRAN. Nesse apêndice, não focamos em métodos de divisão da tarefa em diferentes partes consulte a Ref. [82]. A seguir apresentamos uma coletânea de funções simples que são o cerne da programação paralelo em MPI.

D.3.1 Principais Funções

Para utilizar as funções e variáveis internas do MPI é necessário a inclusão da biblioteca `mpi`. Na linguagem C basta inserir `#include <mpi.h>` no início do código. É possível escrever um código funcional em MPI usando apenas seis rotinas listadas na Tab. D.2.

Tabela D.2 – Com essas seis funções principais do MPI é possível escrever um programa completo.

Função	Resultado
<code>MPI_Init</code>	Inicia o MPI
<code>MPI_Finalize</code>	Finaliza o MPI
<code>MPI_Comm_rank</code>	Identificador do processo
<code>MPI_Comm_size</code>	Número de processos
<code>MPI_Send</code>	Envio de mensagem
<code>MPI_Recv</code>	Recebimento de mensagem

Todo programa que usa MPI deve chamar inicialmente `MPI_Init`. Essa chamada deve ser feita uma única vez, em qualquer ponto do código. Ela deve ser chamada da seguinte maneira `MPI_Init(&argc,&argv)`, de modo que todos os processos iniciados recebam os argumentos da linha de comando: `&argc` e `&argv`. A função `MPI_Finalize` deve ser a última chamada de uma função MPI presente no código. É utilizada para terminar o ambiente de comunicação paralela, liberar memória, etc. Essas duas funções retornam um valor inteiro 0, se a execução foi bem sucedida, do contrário o programa é abortado. As funções `MPI_Comm_size` e `MPI_Comm_rank` retornam respectivamente, o número de processos e o identificador de cada processo (*rank*). A cada processo é atribuído um número ou identificador que varia de 0 até

$N - 1$, onde N é o número de processos criados⁸. Com `MPI_Comm_rank` é possível recuperar o identificador do processo e determinar seu papel nos cálculos que serão realizados, assim como no envio e recepção de mensagens. A chamada dessas funções é:

```
MPI_Comm_rank(MPI_Comm comm, int &rank)
MPI_Comm_size(MPI_Comm comm, int &size).
```

Essas funções aceitam dois argumentos, o primeiro é referente ao comunicador, que é definido como o conjunto de processos que podem se comunicar. O tipo de dado do comunicador é `MPI_Comm` e no caso dos comandos anteriores `comm` é a variável. Por padrão, ao encontrar o `MPI_Init` é criado um comunicador global chamado `MPI_COMM_WORLD`, em geral, apenas este comunicador criado pelo MPI é suficiente, pois nele todos os processos podem se comunicar⁹. A variável `size` recebe o número de processos, enquanto `rank` armazena o identificador de cada processo. Os últimos dois comandos dos seis da Tab. D.2 são envio e recebimento de dados. A sintaxe deles:

```
int MPI_Send(void *buf, int count, MPI_Datatype datatype, int dest,
int tag, MPI_Comm comm);
int MPI_Recv(void *buf, int count, MPI_Datatype datatype, int source,
int tag, MPI_Comm comm, MPI_Status *status);
```

O parâmetro `void *buf` é o espaço de memória dos dados que serão enviados ou recebidos, o tipo `void` pode ser trocado por qualquer tipo de dado válido na biblioteca MPI. Para as comunicações deve ser especificado o tipo de dado por meio do argumento `MPI_Datatype`, cuja equivalência com as tipos da linguagem C é dada pela Tab. D.3. O argumento `int count` especifica o número de elementos do tipo `MPI_Datatype` no *buffer* de envio, que no caso de um vetor é o número de elementos. Especificar o número de elementos enviados torna o código portátil, pois

---

⁸A quantidade de processos criados é especificada no momento da execução do programa.

⁹A *grosso modo*, um comunicador define um grupo de processos que podem se comunicar, o comunicador é parâmetro para todas as funções de envio/recepção de dados.

o tamanho, em *bytes*, pode variar em função da arquitetura da máquina utilizada. Os argumentos **dest**, **source** denotam o destino e fonte da mensagem, enquanto que o argumento **tag** indica um identificador escolhido pelo usuário para a mensagem que está sendo transmitida. Depois que os dados são recebidos, a variável **status** guarda informação sobre a mensagem recebida. Essa variável é uma *struct* com um campo para o processo fonte dos dados, para a **tag** da comunicação e para variável de erro.

```
struct MPI_Status{
    int MPI_SOURCE;
    int MPI_TAG;
    int MPI_ERROR;
}
```

Tabela D.3 – Equivalência entre os tipos de dado da linguagem C e os disponibilizados pelo MPI.

Tipo MPI	Tipo C
MPI_CHAR	char
MPI_SHORT	short int
MPI_INT	int
MPI_LONG	long
MPI_UNSIGNED_CHAR	unsigned char
MPI_UNSIGNED_SHORT	unsigned short
MPI_UNSIGNED	unsigned int
MPI_UNSIGNED_LONG	unsigned long
MPI_FLOAT	float
MPI_DOUBLE	double
MPI_LONG_DOUBLE	unsigned double

D.3.2 Operações Coletivas de Comunicação

As operações de comunicação coletiva garantem o envio e recebimento de dados por todos os processos em um dado comunicador, `MPI_COMM_WORLD` por exemplo. Essas funções facilitam a comunicação por meio de recursos, tais como envio de dados de um processo a todos os demais, redução, juntar dados de todos os processos em apenas um, etc. A seguir estão listadas as principais operações coletivas, assim como a sintaxe do comando utilizado.

Barrier: A sincronização por meio de barreira apenas retorna quando todos os processos no grupo de comunicação `comm` atingem o mesmo ponto no código:

```
int MPI_Barrier(MPI_Comm comm);
```

Broadcast: Envia os dados guardados no vetor `buf` do processo fonte para todos os processos no grupo de comunicação. Os dados dos outros processos são armazenados no vetor `buf`. O campo `count` especifica a quantidade de dados do tipo `MPI_Datatype` que o processo `target` recebe depois da redução.

```
int MPI_Bcast(void *buf, int count, MPI_Datatype datatype,  
int source, MPI_Comm comm)
```

Reduce: A combinação de cada processo é armazenada no `void *recvbuf` após passar por uma operação determinada pelo argumento `op`. O MPI especifica várias operações de redução do tipo `MPI_Op` assim como apresentado na Tab. D.4. O campo `count` é a quantidade de dados do tipo `MPI_Datatype` que serão enviados. No caso de um vetor, a redução ocorre elemento-a-elemento, ou seja, a i -ésima posição no vetor `void *recvbuf` equivale à operação de redução aplicada ao i -ésimo elemento do vetor `void *sendbuf` em cada processo.

```
int MPI_Reduce(void *sendbuf, void *recvbuf, int count,  
MPI_Datatype, MPI_Op op, int target, MPI_Comm comm)
```

Scan: O resultado desse comando é realizar a redução dos dados em `sendbuf` e

Tabela D.4 – Operações básicas de redução, para mais detalhes, consulte a Ref. [82]. Quando os dados estão em um vetor, a redução ocorre elemento a elemento.

Operação	Redução	Tipo de dado
MPI_MAX	máximo	int,float
MPI_MIN	mínimo	int,float
MPI_SUM	soma	int,float
MPI_PROD	produto	int,float

devolver o resultado para cada processo no vetor `recvbuf`, cada processo recebe um valor distinto. Essa diferença no recebimento ocorre pois a redução é parcial, o processo de *rank* *i* recebe dados da operação de redução dos processos com *rank* de zero até *i*. Assim como no `MPI_Reduce`, se o `void *sendbuf` for um vetor, a operação `op` é aplicada elemento-a-elemento.

```
int MPI_Scan(void *sendbuf, void *recvbuf, int count,
MPI_Datatype datatype, MPI_Op op, MPI_Comm comm);
```

Gather: Todos os processos enviam os dados armazenados no vetor `sendbuf` para o processo `target`, sendo guardados no vetor `recvbuf`, na posição inicial `i*sendcount`, onde *i* é o *rank* do processo que envia os dados. A variável `recvcount` é o número de elementos do vetor `recvbuf`, em geral, é o produto do número de processos pela quantidade de elementos do vetor `sendbuf`. Há uma variante dessa função chamada `MPI_Allgather`, que ao invés de enviar os dados para um processo alvo, envia os dados de todos os processos para os demais processos. Os argumentos de envio são os mesmos que na função `MPI_Gather`.

```
int MPI_Gather(void *sendbuf, int sendcount, int MPI_Datatype senddata,
void *recvbuf, int recvcount, MPI_Datatype recvdats, int target,
MPI_Comm comm);
```

Scatter: O processo `source` envia partes distintas dos dados do vetor `sendbuf`

para os demais processos, incluindo a si mesmo. Os dados recebidos são armazenados no vetor `recvbuf`. O processo i recebe `sendcount` elementos contíguos do tipo `sendbuf` começando da posição $i \times \text{sendcount}$.

```
int MPI_Scatter(void *sendbuf, int sendcount, MPI_Datatype senddatatype,
void *recvbuf, int recvcount, MPI_Datatype recvdatatype, int source,
MPI_Comm comm);
```

D.4 Métricas de Análise de Desempenho

O tempo de execução serial é definido como o tempo entre início e término da execução do programa em um computador serial. O tempo de execução em paralelo é o intervalo entre início e término da execução do último processo. Esses tempos são denotados, respectivamente, por T_s e T_p . A função de sobrecarga de um sistema paralelo, T_0 , é a diferença do tempo total em paralelo subtraída pelo tempo total serial Eq. D.82.

$$T_0 = pT_p - T_s. \quad (\text{D.82})$$

O *speedup* S é definido como a razão do tempo serial (do melhor algoritmo) e o tempo em paralelo para resolver o problema idêntico em p processadores. Em teoria, o ganho de tempo não pode ser maior que o número de processadores, contudo há casos especiais tais como de busca em árvores onde o *speedup* é maior que p , esse efeito é conhecido como *speedup superlinear*. Esses casos especiais estão geralmente associados ao aumento da taxa de encontro das variáveis na memória *cache* ou exploração em árvores binárias [82].

A *eficiência* é uma medida da fração de tempo na qual os processadores são empregados de maneira útil. Ela é definida como a razão do *speedup* pelo número de elementos processadores, Eq. D.83.

$$E = \frac{S}{p}. \quad (\text{D.83})$$

A habilidade de um conjunto paralelo manter eficiência em um valor fixo, incrementando simultaneamente o número de processadores e o tamanho do problema é chamada *escalabilidade*. Tais sistemas paralelos tem a habilidade de usar recursos incrementais de maneira produtiva.