

Aplicações de Mecânica Estatística à Psicologia Moral

Alex Kunze Susemihl
Instituto de Física
Universidade de São Paulo

13 de setembro de 2010

Agradecimentos

Foram mais de dois anos desde o meu ingresso no mestrado e a lista de agradecimentos é longa. Primeiramente quero agradecer à minha família: à minha esposa, pelo apoio, por acreditar e entender e por estar incondicionalmente ao meu lado; aos meus pais, que tanto se interessaram e tanto contribuíram com idéias e sugestões, sempre fazendo questão de fazer parte de meus projetos; e ao meu irmão, pelo apoio, pela prontidão e pelo companheirismo. Aos meus queridos avós agradeço pela força e inspiração que eles sempre me incitaram. Aos meus tios Corny e Cito agradeço o apoio e o interesse. Aos meus sogros e cunhados, a parte mais recente da minha família, agradeço o entusiasmo, a alegria e o apoio.

Dizem que os amigos são a família que escolhemos, de forma que continuo com esta. Ao meu grande amigo Domingos, aos meus amigos Marcos, Gabriel, Stefan e Catherine, agradeço pelos tão importantes momentos de descontração. Aos meus companheiros de mestrado Bruno e Leonardo agradeço por levar discussões científicas às madrugadas de sextas e sábados. Ao meu companheiro Igor agradeço pelas discussões acaloradas sobre filosofia e por questionar sempre.

Desde que cheguei ao instituto de física, vejo crescer gradualmente o número de pessoas que freqüentam os seminários de segundas-feiras do nosso grupo. Quero agradecer a todos que fizeram deste grupo um espaço de discussão e diversão científica. Nominalmente, Diogo, Edgar, Rafael, João, André, Prof. Marcos Vinícius, e mais muitos carregam minha gratidão.

No entanto, agradecimentos pessoais à parte, o maior agradecimento deste trabalho deve ser sem dúvida às duas pessoas que me assistiram em grande parte dos meus passos neste projeto. Agradeço especialmente ao meu orientador Nestor Caticha e ao professor Renato Vicente, pela paciência, pela disposição e pela atenção. Muito obrigado.

Sumário

Agradecimentos

I Aplicações de Mecânica Estatística à Psicologia Moral

1 Introdução	p. 7
1.1 Estrutura	p. 9
2 Psicologia Moral e Justificativa do Modelo	p. 10
2.1 Natureza do Modelo	p. 11
2.1.1 Objetivos da Modelagem	p. 12
2.2 Justificativa do Modelo	p. 13
2.2.1 <i>Moral Foundation Theory</i>	p. 13
2.2.2 Dados Neurocognitivos	p. 16
2.2.3 Diferentes Estratégias Cognitivas	p. 17
2.3 Análise Preliminar	p. 20
3 Descrição do Modelo e dos Métodos	p. 25
3.1 Modelo de Agentes	p. 26
3.2 Modelo Hamiltoniano	p. 27
3.2.1 Hipóteses Implícitas no Modelo	p. 28
3.2.2 Métodos de Monte Carlo e Cadeias de Markov	p. 31

3.2.3	Wang-Landau Methods	p. 33
3.3	Aproximação de Campo Médio	p. 35
3.3.1	Exemplo: Modelo de Ising	p. 36
3.3.2	Aproximação de Campo Médio para o Modelo de Fundações Morais	p. 38
4	Resultados	p. 41
4.1	Histogramas e Caracterização das Fases	p. 42
4.1.1	Aproximação de Campo Médio	p. 43
4.1.2	Modelo Hamiltoniano	p. 48
4.1.3	Modelo de Agentes	p. 49
4.2	Comparação com dados de psicologia quantitativa	p. 49
4.3	Diagramas de Fase	p. 53
4.3.1	<i>Assuntos</i> Ortogonais	p. 53
4.4	Perturbações	p. 60
4.4.1	Distância Euclidiana entre Histogramas	p. 61
4.4.2	Divergência de Kullback-Leibler	p. 63
5	Conclusões	p. 67
5.1	Discussão dos resultados	p. 68
5.1.1	Tratamento do modelo.	p. 68
5.1.2	Transições de Fase	p. 69
5.1.3	Perturbações	p. 69
5.2	Interpretação	p. 70

II Projetos Paralelos	71
6 Reconhecimento Invariante de Padrões	p. 73
6.1 Reconhecimento Invariante de Padrões	p. 74
6.1.1 Obtendo Representações Invariantes	p. 75
6.1.2 Implementação Computacional	p. 81
6.1.3 Além do <i>Grayscale</i>	p. 84
6.2 Métodos de Classificação	p. 85
6.2.1 Perceptron	p. 85
6.2.2 Learning Vector Quantization	p. 88
6.2.3 Superparamagnetic Clustering	p. 90
6.2.4 <i>Kernel-based</i> Methods	p. 91
6.3 <i>Features</i> Invariantes	p. 94
6.4 Conclusão	p. 97
7 Considerações Finais	p. 98
Referências Bibliográficas	p. 100

Parte I

*Aplicações de Mecânica Estatística à
Psicologia Moral*

1 Introdução

Em 1929, após os avanços trazidos pelo começo do século no campo da mecânica quântica, P. A. M. Dirac escreveu

The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble (1).

Esta noção de que a teoria por trás da química e da física (e para alguns de toda a ciência) estava acabada e restava aplicar a teoria a casos específicos ganhou bastante força. Cabe a cada um interpretar sua frase como correta ou não, mas ela apresenta uma questão cada vez mais presente na ciência, a do reducionismo no estudo de sistemas complexos. Em 1972, P. W. Anderson expressou sua opinião no trabalho *More is Different* (2), argumentando que ao considerarmos os aspectos de escala e complexidade em coleções de objetos simples, leis novas surgem, exigindo novas idéias e uma nova compreensão. Em suas palavras, se as entidades elementares do campo de pesquisa A obedecem às leis do campo B, isso não implica que *A é só B aplicado*. Em 1972 esta discussão estava viva dentro do campo da física, suscitando discussões entre física *fundamental* e física *aplicada*. Nos quase 40 anos que se passaram desde então, o reducionismo dos físicos tem escolhido novos alvos, entre eles neurônios, populações de seres vivos, sociedades humanas, populações de espécies sujeitas a pressão evolutiva e muitos outros. Mais que nunca o assunto parece atual, ao tentarmos apresentar a biólogos resultados de modelos simples que procuram explicar aquisição de traços morfológicos, ou ao tentarmos apresentar a sociólogos análises baseadas em teoria de grafos para sustentar uma teoria de influência entre vizinhos sociais. Apesar de repetidos sucessos de modelos quantitativos, é bastante comum vermos resistência em diversas áreas. Mas, como Anderson diz ao abrir seu artigo,

The reductionist hypothesis may still be a topic for controversy among philosophers, but among the great majority of active scientists I think it is accepted without question.(2)

ou seja, a hipótese reducionista não é questão de controvérsia para quem a vê funcionando. Talvez seja portanto mais produtivo aplicar um formalismo e produzir novo conhecimento, ao invés de prolongar-se em discussões acerca da validade do mesmo.

O assunto discutido neste trabalho é especialmente delicado por abordar um objeto considerado inerentemente humano: o juízo moral. Mesmo abordagens quantitativas em psicologia como a seguida por Haidt (3) são alvo de críticas veementes, então o que esperar de um trabalho que se propõe a produzir um modelo de agentes que reproduza o comportamento de uma sociedade? Muito pode ser dito a favor ou contra a proposta de usarmos um modelo reducionista ao estudarmos coleções de seres humanos, mas nenhum argumento justifica não tentarmos. Uma vez que um modelo exista, capaz de explicar determinados dados a partir de hipóteses sobre o comportamento humano, rejeitá-lo torna-se cada vez mais uma questão ideológica e não científica.

Este trabalho estuda um modelo de mecânica estatística de agentes em uma sociedade procurando consenso com seus vizinhos sociais a respeito de julgamentos morais. O modelo foi proposto para reproduzir dados de questionários cedidos por Jonathan Haidt (4) em um formalismo de mecânica estatística. Aqui procuramos analisar o modelo mais a fundo, procurando obter resultados para uma aproximação de campo médio, e quantificando seu comportamento em resposta a perturbações.

1.1 Estrutura

O restante desta tese será estruturado como descrito em seguida. No capítulo seguinte, discutirei o objeto do modelo descrito, abordando a teoria de fundações morais e alguns dados sobre diferenças cognitivas entre populações de afilições políticas diferentes. Elucidarei também a relação entre estes dados e as estratégias cognitivas representadas no nosso modelo. No capítulo 3 descreverei mais a fundo o modelo em cada um de seus paradigmas e mostrarei a derivação das equações de campo médio usadas no estudo do sistema. Neste capítulo serão apresentados também os métodos utilizados para estudar a resposta do sistema à perturbações. No capítulo 4 apresentarei os resultados de simulações nas diferentes versões do modelo, enfatizando os diagramas de fases resultantes e o comportamento do sistema em resposta a perturbações. No capítulo 5 discutirei os resultados sintetizando os resultados obtidos. No capítulo 6 descreverei brevemente projetos paralelos aos quais me dediquei no decorrer deste período. No capítulo 7 apresento considerações finais e agradecimentos.

2 *Psicologia Moral e Justificativa do Modelo*

*Zwei Dinge erfüllen das Gemüt mit
immer neuer und zunehmender
Bewunderung und Ehrfurcht, je öfter und
anhaltender sich das Nachdenken damit
beschäftigt: Der gestirnte Himmel über
mir und das moralische Gesetz in mir.^a*

^aDuas coisas me enchem a alma de crescente admiração e respeito, quanto mais intensa e frequentemente o pensamento delas se ocupa: o céu estrelado sobre mim e a lei moral dentro de mim.

Immanuel Kant

xNeste capítulo descreverei o objeto a ser modelado, a justificativa do modelo e finalmente como é feita a correspondência entre os parâmetros do modelo e comportamentos. Ao final do capítulo apresentarei uma análise quantitativa dos dados de questionários psicológicos, visando esclarecer a estrutura presente, que será posteriormente comparada com resultados do modelo estudado.

2.1 Natureza do Modelo

O debate sobre a influência da natureza e da criação na construção do caráter e personalidade do indivíduo é bastante antigo, tendo havido grandes contribuições para ambos os lados. Émile Durkheim foi um dos primeiros a atribuir autonomia aos fenômenos sociais desvinculados de motivações individuais, ressaltando a importância da criação. Os fenômenos sociais seriam emergentes, advindos da sociedade como um todo, não dependendo da natureza de cada indivíduo. O trabalho mais comentado em que Durkheim procurou aplicar estas idéias foi seu trabalho sobre as taxas de suicídio em diferentes populações (5). Ao invés de entrar no mérito de defender ou contestar o trabalho, vou mencionar somente o objetivo de Durkheim, qual seja, formalizar a sociologia como uma disciplina científica rigorosa, capaz de explicar estatísticas agregadas a partir de traços comportamentais. Claramente, descrever personalidades e pessoas apenas como resultado de sua criação e interação social é uma simplificação extrema no caso geral, como argumenta Pinker (6). No entanto, isso não exclui a possibilidade de diversos fenômenos sociais poderem ser explicados a partir do ponto de vista da criação. Ainda assim, para estudar exclusivamente o efeito do aprendizado e das interações sociais, escolhemos tomar o estado inicial dos indivíduos como variável aleatória, recorrendo portanto a um paradigma de *tabula rasa*.

Propriedades emergentes são propriedades que não podem ser atribuídas a nenhuma parte de um sistema, mas são características do sistema como um todo. Poderíamos pensar que o estudo de propriedades emergentes deveria ser por definição holístico, focando no sistema como um todo. No entanto, apesar de não podermos atribuir propriedades emergentes a partículas ou a indivíduos, muitas vezes podemos determinar quais as características das partículas ou interações que levam à emergência de certas propriedades. A mecânica estatística oferece um arcabouço privilegiado para o estudo de propriedades emergentes. No formalismo dos *ensem-*

bles o individual e o coletivo se confundem, pois temos fenômenos claramente coletivos, como a magnetização espontânea no modelo de Ising, podendo ser derivados da interação entre dois spins em sítios vizinhos. Vale notar que a distinção entre holismo e reducionismo não é clara neste caso, uma vez que uma característica como a intensidade de interação entre dois spins pode ser vista como individual por caracterizar um par específico, ou coletiva por ser a mesma em todos os pares. Ao estudarmos sistemas desordenados, como vidros de spin, esta diferenciação fica cada vez mais difícil. Claramente, diversas hipóteses tem de ser feitas para que possamos descrever um sistema neste formalismo, e em diversos sistemas esta hipóteses não se sustentam. Vale destacar a hipótese de que a energia de uma interação entre dois corpos (agentes, indivíduos) seja simétrica com relação à troca dos agentes, o que muitas vezes é falso. Especificamente, tentaremos abaixo modelar o processo de aprendizado moral, o que claramente implica em uma assimetria nos papéis do *professor* e do *aluno*, impossibilitando um tratamento direto pelo formalismo da física estatística. Veremos também como uma simplificação nos permite recuperar o formalismo.

No caso particular do trabalho aqui descrito, nosso fenômeno emergente a ser descrito é o aparecimento de consenso ou de discordância a respeito de opiniões sobre questões discutidas, e a variação da diversidade destas opiniões em ambos os casos. A característica dos blocos de construção do nosso sistema (agentes) que vamos procurar analisar será a estratégia de aprendizado ou a forma como este se comporta quando confrontado com consenso (corroboração) e dissenso (contradição).

2.1.1 Objetivos da Modelagem

Os modelos desenvolvidos no trabalho aqui descrito têm como principal objetivo caracterizar as conseqüências na escala social de diferentes estratégias cognitivas no nível individual. Grande atenção foi dada ao campo de dinâmica de opiniões, desde o modelo do votante majoritário até modelos mais refinados, como o modelo Sznajd com confiança limitada (7). No entanto, a forma extremamente simplificada como são descritas as interações sociais (no intuito de obter resultados analíticos), parece dificultar a comparação com dados reais. No trabalho de Vicente et al. (8), os autores tomam uma outra direção, procurando usar uma descrição mais complexa dos agentes e de suas interações com o preço de obterem resultados analíticos somente em alguns casos particulares e obtendo resultados numéricos para outros. Aqui foi

estendido o trabalho de Vicente e Caticha (4) sobre o problema de aprendizado moral, o que consistiu basicamente em uma análise de um espaço de parâmetros diferentes, e uma análise de novas propriedades do sistema, além de uma semântica nova para o sistema. O sistema foi também tratado numa descrição de mecânica estatística através de uma hamiltoniana que consiste no custo de discordância entre pares de agentes. Este modelo foi então estudado na tradição de mecânica estatística através de simulações de Monte-Carlo e através de uma aproximação de campo médio.

O objetivo deste trabalho é obter uma caracterização da influência de diferentes estratégias cognitivas no comportamento do sistema como um todo. Entre os comportamentos de interesse estão as transições ordem-desordem, e a resposta do sistema a perturbações nos assuntos discutidos. Para tal, serão utilizadas simulações tanto no paradigma de modelos baseados em agentes, quanto em mecânicas estatística, tanto em simulações de Monte-Carlo, como na aproximação de campo médio.

2.2 Justificativa do Modelo

Nesta seção embasarei o modelo proposto com base nas idéias da teoria de fundações morais de Jonathan Haidt (3) e dados de experimentos neurocognitivos, como os descritos em Amodio et al. (9) e Oxley et al. (10). Justificarei a escolha da representação dos estados internos dos agentes como vetores de *fundações morais* a partir do trabalho de psicologia moral de Haidt. Os resultados descritos por Amodio et al. (9), além de outros resultados citados, inspiram a escolha da regra de aprendizado e potencial de custo psicológico dos nossos agentes. Partindo destes resultados descreverei as diferentes estratégias cognitivas e como as associo com orientações políticas.

2.2.1 *Moral Foundation Theory*

O debate entre intuicionistas e racionalistas sobre a moral humana é dos debates mais antigos da filosofia. Apesar de grandes debates, dados recentes parecem estar empurrando para uma teoria da primazia intuicionista (mas não ditadura, como apontado por Haidt em (3)). Grande parte do julgamento moral é afetivo e não racional, de forma que em nossa modelagem do julgamento moral escolhemos tomar um viés do tipo caixa-preta em lugar de um viés de lógica

simbólica. As justificativas apontadas por sujeitos questionados sobre seus julgamentos morais são provavelmente geradas *a posteriori* na maioria destes julgamentos, apesar de haver casos em que pode se observar uma revisão do primeiro impulso afetivo que pode ser atribuído a um raciocínio moral (*moral reasoning*)¹. Alinhado com o paradigma do julgamento moral como caixa-preta, modelamos o julgamento moral de um assunto de acordo com o paradigma de redes neurais. Representamos a *matriz* moral² do agente como um vetor em um espaço real (J_i , para o agente i), e o assunto a ser julgado (aborto tardio, pena de morte, etc.) como outro vetor em um espaço real (S^μ , para o assunto μ). O julgamento de um agente sobre um assunto é então dado pelo produto escalar dos dois vetores ($h_i^\mu = J_i \cdot S^\mu$).

Um segundo ponto destacado na síntese da psicologia moral feita por Haidt é a existência de novas fundações morais. Historicamente se têm dado uma grande importância aos fatores associados à justiça (*fairness*) e à integridade física (*harm*) na psicologia moral. No entanto, em uma extensa análise estatística foi mostrado que há pelo menos mais três fundações conservadas ao longo de diversas culturas e regiões: lealdade ao grupo; respeito a autoridades; e respeito a pureza e santidade. Surpreendentemente conservadores autodeclarados e liberais autodeclarados demonstram utilização diferente das cinco fundações morais³. Como podemos ver na figura 2.1, liberais dão valor maior às fundações de justiça e integridade física, enquanto conservadores dão valor similar para as cinco fundações. A partir destes dados escolhemos a dimensão das matrizes morais de nossos agentes, isto é, na maioria de nossas simulações tomamos a dimensão dos vetores representando opiniões e assuntos discutidos como sendo igual a cinco.

Em um esforço para popularizar o conhecimento sobre as diversas fundações morais Haidt e seus colaboradores desenvolveram um portal na rede chamado *Yourmorals.org*, no qual se pode realizar uma série de testes para determinar a orientação moral de cada um. Anedoticamente, pode-se ver na figura 2.2 a orientação moral de um sujeito do questionário.

¹O exemplo citado por Haidt é o de pessoas que respondem afirmativamente à elaborada pergunta: *Você sufocaria seu próprio bebê se fosse a única forma de impedi-lo de revelar a posição de toda sua família a um exército determinado a matá-la?*

²O itálico é colocado para enfatizar que o termo matriz moral é originário do campo de psicologia social, não representando portanto uma matriz no sentido matemático mais comum, mas sim um conjunto de valores.

³Os termos liberal e conservador são usados na conotação americana, liberais à esquerda e conservadores à direita, com respeito a questões sociais (não necessariamente com respeito a questões econômicas). Doravante utilizaremos os termos conservadores como sinônimo de conservadores autodeclarados e analogamente com liberais.

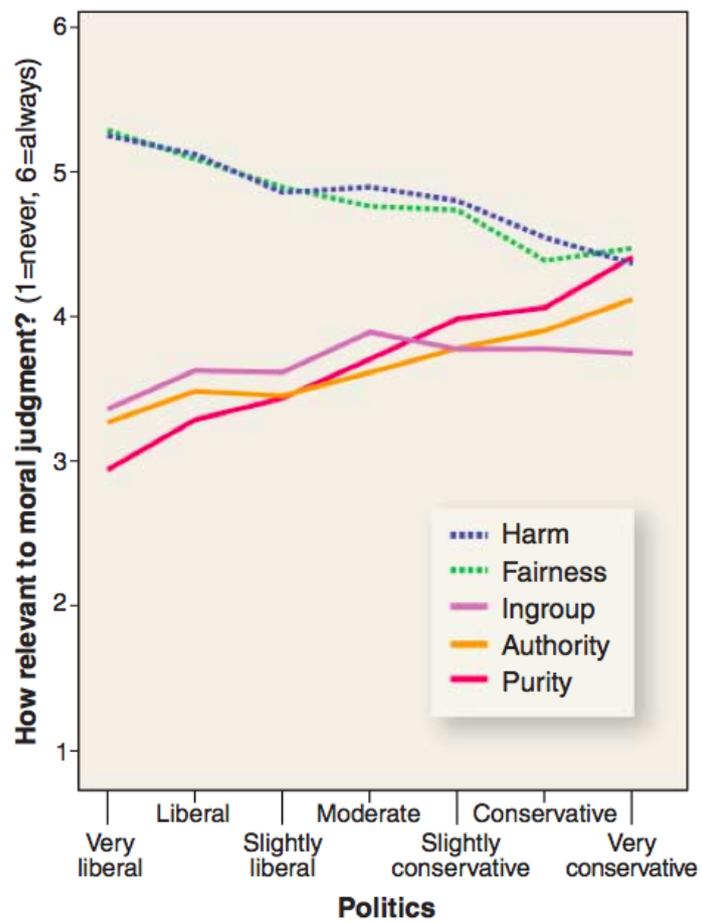


Figura 2.1: Frequência com a qual cada uma das fundações é importante para um julgamento moral em função da orientação política. Figura de (3).

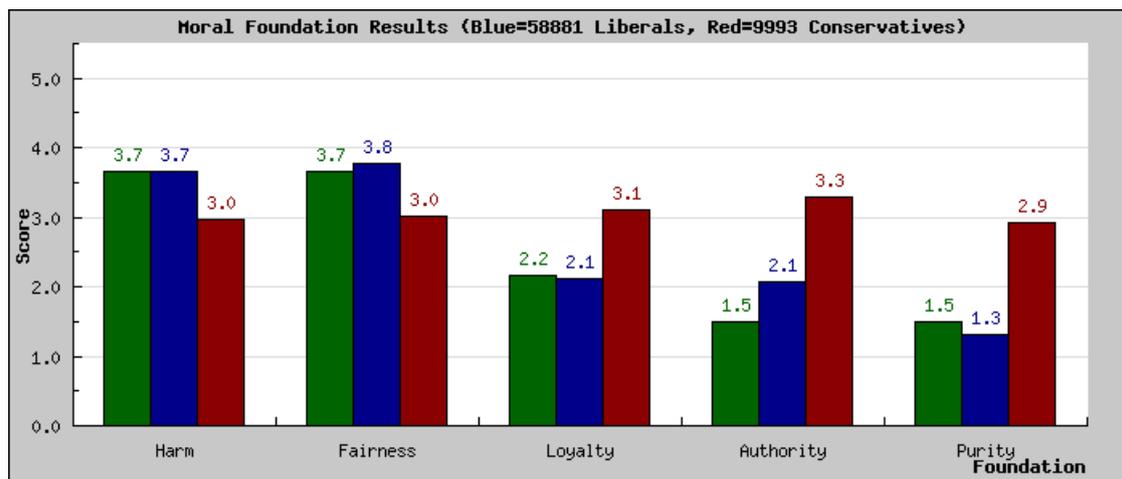


Figura 2.2: Matriz moral obtida através do questionário online em 2.2 para um sujeito do questionário (verde) em comparação com a média dos liberais (azul) e dos conservadores (vermelho).

2.2.2 Dados Neurocognitivos

Humanos são uma espécie social, dependendo do convívio para sobreviver, de forma que indivíduos excluídos de seu meio social dificilmente sobreviveriam em isolamento (11). Deve ter sido de suma importância para a espécie humana o desenvolvimento de meios para detectar e minimizar a exclusão social, para evitar o isolamento. "*Natural Selection*) works like a tinkerer.", disse François Jacob (12), contrapondo a ação da evolução à ação de um engenheiro e comparando-a ao ato da bricolagem, em que se procura utilizar de quaisquer elementos disponíveis para construir um resultado final. Desta forma, não surpreende que no curso da evolução o cérebro humano tenha recrutado áreas relacionadas à dor física para trabalhar os sentimentos de exclusão e dor social (13, 14). Experimentos de imagens funcionais por ressonância magnética (fMRI) mostram um forte envolvimento do córtex cingulado anterior dorsal (dACC) durante experimentos em que sujeitos são submetidos à reprovação e exclusão social (11, 15). O córtex cingulado anterior dorsal é conhecidamente envolvido quando pacientes são submetidos a dor física. Nossos agentes, simplificações do *Homo sapiens*, não procuram resolver conflitos através de esquemas elaborados como nós, mas simplesmente procuram aproximar suas opiniões das de seus pares (ver seção 3.1) ou minimizar um custo social (ver seções 3.2 e 3.3). Apesar de estarmos simplificando este comportamento, os dados apresentados repetidamente nos mostram que a dor social não é tão diferente da dor física quanto esperaríamos, de forma que agentes que

minimizam um custo psicológico parecem ser uma metáfora bastante razoável.

A ciência política e a psicologia frequentemente documentaram diferenças em estilos cognitivos e motivacionais entre indivíduos de perfil conservador e liberal⁴. No entanto, recentemente têm surgido evidências neurocientíficas de que estes resultados não são só verdadeiros no âmbito da psicologia ou sociologia, mas também no âmbito da neurociência. Em um trabalho de 2007, Amodio e colaboradores determinaram através de técnicas de eletroencefalografia diferenças nos padrões de atividade cerebral com respeito a orientação política (9). A tarefa executada, chamada de *Go/No-go*, é um simples jogo em que aos sujeitos são apresentados *slides* instruindo-os a apertarem ou não um botão. Após uma grande maioria de estímulos positivos (*Go*), são dados alguns estímulos negativos. Os dados do experimento são coletados quando o indivíduo erra em resposta a um estímulo negativo (aperta o botão em resposta a um *No-go*). Este evento tipicamente indica uma frustração por não estar em concordância com as instruções do jogo. Sujeitos conservadores demonstram uma ativação menor das áreas recrutadas para monitoramento de conflito nesta situação em comparação com sujeitos liberais. Interessantemente, entre as áreas ativadas durante experimentos de monitoramento de conflito está o córtex cingulado anterior, ativado durante situações associadas a dor social. Portanto, podemos deduzir que, dado que sujeitos conservadores têm uma ativação menor de áreas associadas a dor social em situações de monitoramento de conflito, o comportamento destes com relação a novidades deve ser menos controlado por minimização da dor social que o de liberais.

2.2.3 Diferentes Estratégias Cognitivas

Há inúmeras formas de escolher um modelo baseado na informação descrita acima. Nós nos concentramos em duas. Primeiramente podemos modelar uma estratégia de aprendizado moral a partir do perfil cognitivo traçado acima. Outra forma é modelar um custo social de dissenso e compreender como aprendizado social a minimização deste mesmo custo. Ao modelarmos uma estratégia de aprendizado, tomamos em consideração que um tipo de agente, os conservadores ideais, tratam toda informação da mesma forma, não diferenciando entre novidades e informações já conhecidas. O outro extremo seriam agentes ultra-liberais, que dariam uma importância infinitamente maior a novidades que a informações já conhecidas.

⁴Noto novamente que os termos são usados no sentido americano, e determinados por declaração própria de orientação política.

Primeiramente, descrevemos como quantificaremos o comportamento de agentes no paradigma de aprendizado on-line de agentes. O paradigma de aprendizado on-line em perceptrons pode ser resumido em uma equação (16):

$$J_{n+1} = J_n + \frac{\xi}{D} W(J_n, S_n, t_n) t_n S_n, \quad (2.1)$$

onde ξ é uma constante chamada de taxa de aprendizado e D é a dimensão do espaço vetorial de J_n e S_n . Os vetores J_n correspondem aos vetores de pesos do perceptron após n passos de treinamento. A função W é a função de modulação que determina a regra de aprendizado em questão. Os vetores S_n são os vetores que representam os exemplos e os valores t_n normalmente restritos aos valores ± 1 são a classe associada ao exemplo n . Como dito, o agente conservador ideal reagiria a qualquer tipo de informação da mesma forma. Esta é especificamente a prescrição do aprendizado hebbiano em perceptrons, em que a representação interna é corrigida a cada novo exemplo indiscriminadamente. A função de modulação correspondente ao aprendizado hebbiano é a função constante

$$W_H(J, S, t) = 1. \quad (2.2)$$

O agente ultra-liberal em oposição ao anterior só se preocupa com exemplos novos, ou seja, exemplos nos quais seu julgamento é diferente do julgamento do professor. Esta é especificamente a regra do perceptron, dada por

$$W_p(J, S, t) = \Theta(-J \cdot S t). \quad (2.3)$$

Na equação acima a função $\Theta(x)$ é a função de Heaviside. Claramente, nem todos nós nos encaixamos nos rótulos de ultra-liberais ou ultra-conservadores, de forma que estudaremos combinações lineares destas duas regras de aprendizados. As regras estudadas serão do tipo

$$W_\delta(J, S, t) = \delta W_H + (1 - \delta) W_p = \delta + (1 - \delta) \Theta(-J \cdot S t) \quad (2.4)$$

indexadas pelo parâmetro δ . Estudaremos os efeitos de diferentes valores de δ sobre diversas grandezas relacionadas à dinâmica do modelo de agentes.

A possibilidade seguinte seria não modelar o processo de aprendizado diretamente, mas modelar o custo da interação entre dois agentes. Podemos então amostrar estados do sistema a determinadas temperaturas, obtendo a termodinâmica com técnicas de Monte-Carlo. Outra possibilidade seria usarmos métodos como o de Wang-Landau (17), onde estimamos diretamente

a densidade de estados, obtendo uma descrição a qualquer temperatura do sistema. No caso de modelarmos a interação entre dois agentes, faremos a simplificação de assumir que ambos agentes têm a mesma estratégia cognitiva. Apesar de não ser estritamente necessário, isso simplifica bastante a construção da energia de interação. No caso de dois agentes serem ultra-liberais, a energia de interação entre os dois só seria não-nula no caso de ambos discordarem sobre um dado assunto. No caso de serem ultra-conservadores, teríamos uma contribuição de energia independente da concordância de ambos. Se tomarmos dois agentes com matrizes morais J e J' e um assunto S^5 , podemos, no caso de agentes ultra-conservadores, escrever a energia de interação como

$$V_{conservador} = -(J \cdot S)(J' \cdot S). \quad (2.5)$$

Podemos ver que a contribuição é negativa sempre que a opinião de ambos é igual e positiva caso contrário. Além disso podemos ver que a contribuição da interação para a energia total é tanto maior quanto mais alinhados estiverem os dois agentes em relação ao eixo dado pelo assunto S . No caso de agentes ultra-liberais podemos escrever

$$V_{liberal} = -\Theta(-(J \cdot S)(J' \cdot S))(J \cdot S)(J' \cdot S). \quad (2.6)$$

Assim como fizemos no parágrafo anterior, definimos um agente intermediário, dependendo de um parâmetro δ , para o qual podemos escrever a energia de interação como

$$V_\delta = \delta V_{conservador} + (1 - \delta)V_{liberal} = -\delta hh' + (1 - \delta)\Theta(-hh')hh', \quad (2.7)$$

onde passamos a usar a notação $h = J \cdot S$. Usando a relação $\Theta(x) = \frac{1 - \sigma(x)}{2}$, obtemos

$$V_\delta = -\frac{1 + \delta}{2}hh' + \frac{1 - \delta}{2}|hh'|. \quad (2.8)$$

A última hipótese que é necessário mencionar é a de que a energia do sistema é aditiva não só sobre todos os pares de agentes, mas também sobre todos os assuntos possivelmente em discussão. A energia total do sistema, associada a uma dada configuração das matrizes morais dos agentes, seria então dada por

$$H(\{J_i\}|\{S^\mu\}) = \frac{1}{P} \sum_{(i,j)} \sum_{\mu} \left(-\frac{1 + \delta}{2} h_i^\mu h_j^\mu + \frac{1 - \delta}{2} |h_i^\mu h_j^\mu| \right). \quad (2.9)$$

⁵Usarei, doravante o termo assunto ou *issue* para me referir à representação vetorial do assunto dado, assim como matriz moral se refere à representação vetorial da mesma

Acima, a primeira soma é sobre todos os pares de agentes interagentes e a segunda sobre todos os assuntos em discussão. Extendemos também a notação definida acima para dizer $h_j^\mu = J_j \cdot S^\mu$.

2.3 Análise Preliminar

Os dados fornecidos pela equipe de Jonathan Haidt são questionários voluntários realizados pela página *yourmorals.org*. Os dados consistem em identificadores de sexo e afiliação política⁶, respostas a 32 questões acerca de julgamento moral, sendo 16 questões do tipo *Quão importante você considera isso para um julgamento moral?* e 16 do tipo *Quanto você concorda com a frase abaixo?*, além dos valores das fundações morais obtidos a partir dos questionários. Para termos uma idéia da estrutura dos dados, podemos aplicar alguns métodos de visualização de dados e estatísticas clássicas. O método *Sorting Points into Neighborhoods* (18) procura permutar a ordem dos dados de forma a minimizar uma função da matriz de distâncias entre os pontos, ressaltando com isso características da distribuição geométrica dos dados. Desta forma, aglomerados distintos, formas unidimensionais, formas esféricas e formas toroidais resultam em assinaturas características na matriz de distância obtida ao final do método. Procuramos investigar então nos dados se podemos inferir alguma coisa sobre a organização geométrica deles.

Obtivemos de Jonathan Haidt dados a respeito de aproximadamente 14000 indivíduos, com informações sobre as respostas dadas às perguntas nos questionários e as matrizes morais obtidas destas. Primeiramente podemos, seguindo o que foi feito em (4), calcular histogramas da projeção de indivíduos de uma afiliação política sobre o vetor *Zeitgeist*, definido como a média dos vetores de conservadores. Vemos estes histogramas na figura 2.3 e podemos ver a clara diferenciação da distribuição de projeções para os conservadores, moderados e liberais. Para a comparação com o nosso modelo limitamo-nos às matrizes morais, e veremos aqui que estas de fato dão-nos essencialmente a mesma informação que os vetores de respostas. Nas figura 2.4(a), 2.4(b) e 2.4(c) podemos ver a análise estatística feita a partir dos vetores de respostas. Como elucidado no trabalho que introduz o método (18), a matriz obtida é característica de distribuições com uma direção de variação bastante pronunciada. Para verificar que esta variação está de fato associada à afiliação política declarada no questionário, temos logo abaixo a afiliação política plotada como função do índice da permutação obtida através do método.

⁶A afiliação política vai de 1 (*muito liberal*) até 7 (*muito conservador*).

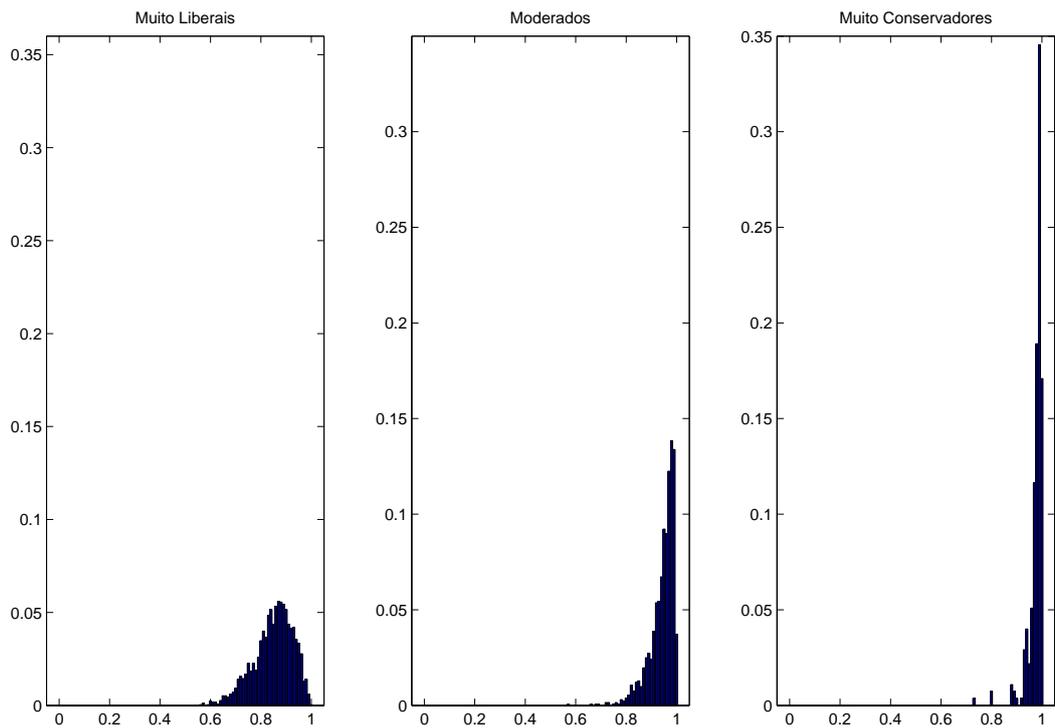
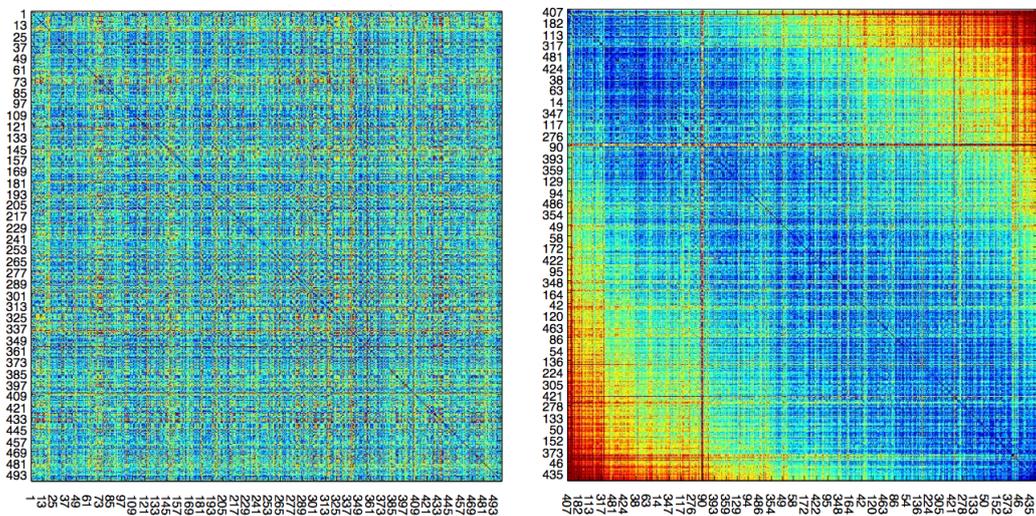
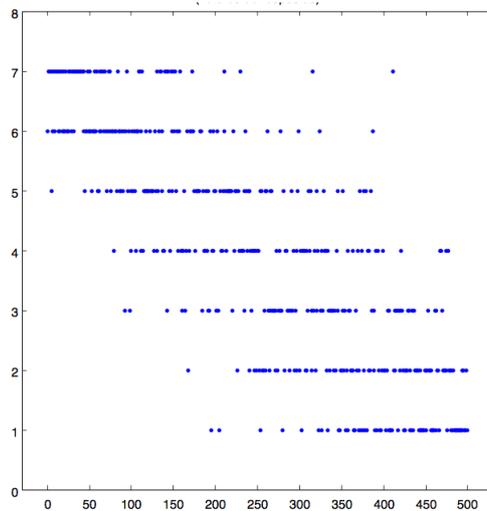


Figura 2.3: Vemos aqui histogramas de projeção de cada uma das afiliações políticas sobre o vetor *Zeitgeist* obtido como a média das matrizes morais conservadoras (ver (4)). Vemos que a afiliação política mais conservadora tem uma distribuição mais estreita em torno do máximo enquanto os liberais tem uma distribuição mais larga.



(a) Matriz de distâncias entre vetores de respostas.

(b) Matriz de distâncias após aplicação do método SPIN.



(c) Afiliação política como função da ordem de permutação do SPIN.

Figura 2.4: Temos aqui uma análise dos dados a partir dos vetores de respostas obtidos no questionário. Podemos ver uma clara correlação entre a ordenação obtida no SPIN e a afiliação política, permitindo-nos inferir que esta é um eixo de variação na população descrita.

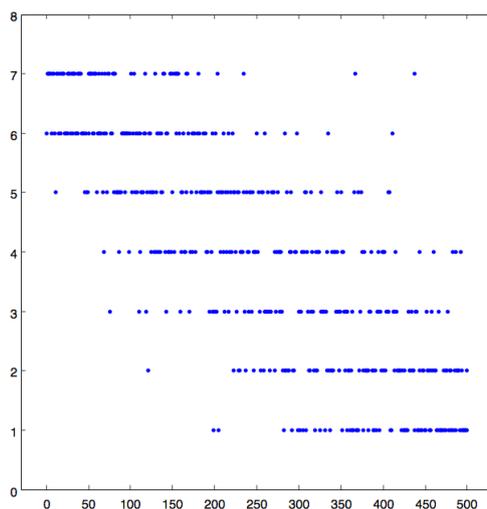
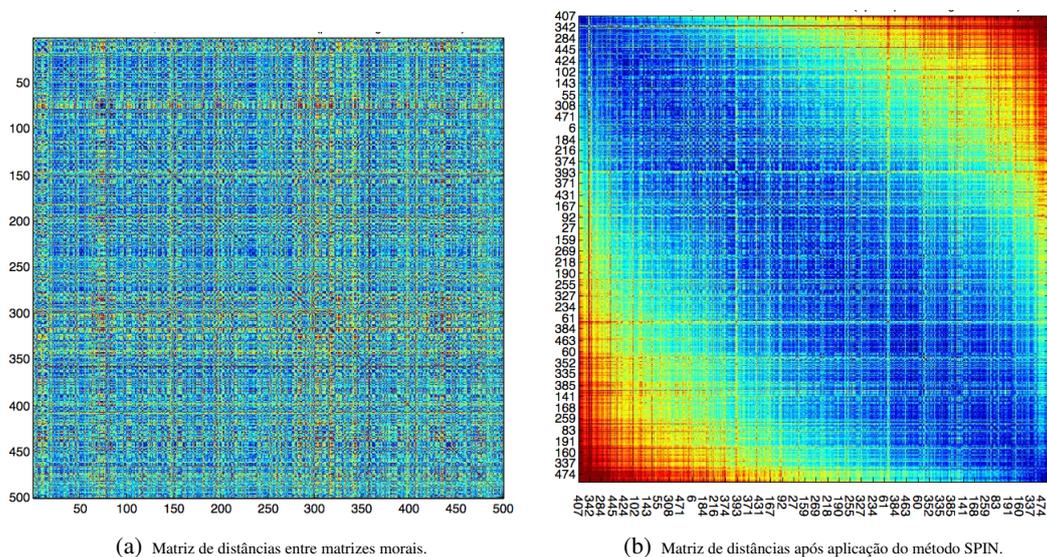


Figura 2.5: Acima temos a mesma análise aplicada aos mesmos dados, mas desta vez somente às matrizes morais. Os resultados obtidos são qualitativamente equivalentes, permitindo também inferir uma forte correlação entre a ordenação geométrica obtida e a afiliação política, confirmando que esta é importante para o estudo que pretendemos fazer.

Vemos claramente que a direção associada à variação principal deixa-se identificar com a afiliação política. Nas figuras seguintes (figuras 2.5(a), 2.5(b) e 2.5(c)) vemos a mesma análise aplicada ao caso das matrizes morais. Desta forma justificamos o uso das matrizes morais como representação do estado de cada indivíduo preferindo o uso de vetores de cinco dimensões doravante. A análise acima justifica também a associação da afiliação política com a dimensão principal de variação de comportamento no nosso modelo.

3 Descrição do Modelo e dos Métodos

Neste capítulo entrarei mais a fundo na definição de cada uma das versões do modelo. Primeiramente, detalharei o modelo baseado em agentes, e métodos de simulação usados para ele. Em seguida explicarei mais a fundo o modelo hamiltoniano, enunciando técnicas de simulação utilizadas durante o trabalho. Por último, apresentarei a aproximação de campo médio e deduzirei as equações autoconsistentes para o modelo.

3.1 Modelo de Agentes

O modelo de agentes, como dito acima, tira sua inspiração principalmente do estudo do aprendizado on-line de perceptrons. O objeto de estudo desta área, no entanto, é quase sempre a capacidade de generalização de um perceptron em um dado cenário, sob diversas regras de aprendizado, ou a otimização da capacidade de generalização do mesmo em alguns cenários (16). No nosso caso estamos, interessados no que acontece com uma sociedade de perceptrons, ligados por um grafo social, aprendendo uns dos outros. É interessante notar que, ao contrário do que acontece normalmente no estudo de perceptrons, não temos uma resposta *correta* para o problema de aprendizado, não há como nenhum perceptron obter a regra de classificação correta para o problema. Nosso interesse não está portanto em como os perceptrons se alinham a uma dada resposta *correta*, mas sim como eles se alinham uns com os outros. Para tal, estudaremos estatísticas como a projeção média entre agentes.

A dinâmica do modelo de agentes é dado por uma atualização assíncrona através da regra de update de aprendizado on-line usual introduzida acima. Teremos então a regra de atualização dada por

$$J_i^{novo} = J_i + \zeta \frac{\xi}{D} (\delta + (1 - \delta) \Theta(-J_i \cdot S^\mu J_j \cdot S^\mu)) \Theta(J_j \cdot S^\mu) S^\mu. \quad (3.1)$$

O agente i é o agente atualizado no instante dado, o agente j é um dos vizinhos do agente i sorteado uniformemente entre estes e o assunto S^μ é um assunto escolhido uniformemente entre os assuntos discutidos. Cada passo da dinâmica consiste portanto no sorteio de um agente a ser atualizado (o aluno), de um de seus vizinhos (o professor) e de um assunto a ser discutido. Começamos as simulações inicializando cada agente com uma matriz moral aleatória de módulo 1 e correndo a dinâmica por um tempo pré-determinado para atingir um estado descorrelacionado das condições iniciais. No caso ideal queremos encontrar o comportamento de equilíbrio deste sistema, mas não temos nenhum modo de dizer se o sistema dado de fato relaxa para al-

gum equilíbrio ou se ele entra em algum tipo de órbita caótica, periódica ou estocástica. Este é um dos motivos pelos quais nossa análise baseia-se principalmente na versão hamiltoniana do modelo, a ser descrita abaixo. A variável ζ é uma variável aleatória de Bernoulli com probabilidades $P(\zeta = 1) = 1 - \eta, P(\zeta = -1) = \eta$. Desta forma podemos adicionar um elemento de aleatoriedade ao modelo de agentes. Podemos então estudar a dependência da dinâmica deste valor de η em analogia com o parâmetro β do modelo hamiltoniano.

Outro assunto de grande interesse na literatura de sociofísica são as chamadas redes complexas. Uma grande diversidade de sistemas demonstra uma organização em redes chamadas complexas, que apresentam características que não se manifestam em redes regulares nem em redes aleatórias geradas da forma mais simples (sorteando pares de nós e conectando-os, como no modelo $G(n, p)$ de Erdős-Rényi). Dois tipos de redes são de especial interesse, redes *scale-free*, nas quais a distribuição de probabilidade dos graus de um nó são leis de potência, e redes *small-world*, modelos nos quais o caminho mínimo médio de uma rede com n nós cresce no máximo com $\log(n)$. O caminho mínimo médio de uma rede é a média sobre todos os pares de nós do comprimento do menor caminho levando de um nó ao outro. Muito esforço tem sido feito para detectar a influência da estrutura da rede sobre a dinâmica de um sistema, e isso também foi estudado brevemente, embora não tenhamos encontrado fortes indícios de que a rede social defina a dinâmica do sistema. Para tal interesse comparamos redes quadradas, redes quadradas com *rewiring* aleatório segundo o modelo de Watts-Strogatz (19) e redes geradas a partir do modelo de Barabasi-Alberts (20).

3.2 Modelo Hamiltoniano

O nosso modelo hamiltoniano é definido pela função hamiltoniana mencionada acima

$$H(\{J_i\}|\{S^\mu\}) = \frac{1}{P} \sum_{(i,j)} \sum_{\mu} -\frac{1+\delta}{2} h_i^\mu h_j^\mu + \frac{1-\delta}{2} |h_i^\mu| |h_j^\mu|. \quad (3.2)$$

A forma funcional desta hamiltoniana ressalta o fato de que esta função nos dá um valor para a energia da configuração associada às matrizes morais $\{J_i\}$ dados os assuntos em discussão $\{S^\mu\}$ ¹. Em estudos de sistemas estatísticos desordenados, como vidros de spin, freqüentemente é de interesse realizar um média sobre a desordem do sistema, no caso de vidros de spin a

¹A notação é usada em analogia com a notação de probabilidades condicionais $P(A|B)$.

desordem está nas energias de interação (21). Poderíamos procurar médias, como no caso do modelo de Sherrington-Kirkpatrick, do tipo *quenched*

$$\langle\langle \log(Z) \rangle\rangle = \int d\mu(S^\mu) \log \left(\int d\mu(J_i) \exp(-\beta H(\{J_i\}|\{S^\mu\})) \right). \quad (3.3)$$

No entanto, pode ocorrer de perdermos informação sobre transições fazendo esta média, poderia ocorrer de a transição ser descontínua para cada configuração específica de $\{S^\mu\}$ mas contínua quando tomamos a média sobre todos os conjuntos $\{S^\mu\}$. Além disso, não estamos interessados em tomar o limite termodinâmico no número de assuntos discutidos, pois este não é o caso de interesse para o nosso modelo. Desta forma concentramo-nos em estudar casos específicos de conjuntos de assuntos, como por exemplo, assuntos ortogonais, assuntos aleatórios, assuntos aleatórios em um dado domínio, etc. Uma vez especificado o conjunto de assuntos em discussão, usamos as estratégias clássicas de Monte-Carlo para obter aproximações de valores médios termodinâmicos como

$$\langle H \rangle = \frac{1}{Z} \int d\mu(\{J_i\}) H(\{J_i\}|\{S^\mu\}) e^{-\beta H}, \quad (3.4)$$

onde Z é a função de partição definida como

$$Z(\beta|\{S^\mu\}) = \int d\mu(\{J_i\}) e^{-\beta H(\{J_i\}|\{S^\mu\})}. \quad (3.5)$$

Podemos escrever a densidade de probabilidade de equilíbrio como

$$P(\{J_i\}|\{S^\mu\}) = \frac{e^{-\beta H(\{J_i\}|\{S^\mu\})}}{Z} \quad (3.6)$$

e considerar as médias dadas como médias sobre esta distribuição. Vamos abaixo discutir duas formas de obter estas médias, a primeira baseada na implementação de cadeias de Markov cuja distribuição de equilíbrio é a distribuição de Gibbs-Boltzmann do sistema e a segunda baseada em obter histogramas que aproximem a densidade de estados do sistema em função da energia e então calcular as médias a partir destes histogramas.

3.2.1 Hipóteses Implícitas no Modelo

Mencionarei rapidamente as hipóteses implícitas na utilização de um modelo hamiltoniano de mecânica estatística para o nosso sistema. Primeiramente discutirei o fato de estarmos utilizando mecânica estatística de equilíbrio para modelar um sistema. Em seguida mencionarei

o aparecimento do parâmetro β como consequência de uma maximização de entropia com um vínculo na energia média do sistema, e discutirei sua interpretação para o nosso modelo. Por último, discutirei a utilização de sociedades homogêneas em relação às estratégias cognitivas de seus agentes.

No trabalho que precedeu a esse (4), o tratamento dado ao sistema que estudamos foi o de perceptrons aprendendo *on-line*, focando na dinâmica do sistema, sem preocupações com estados de equilíbrio. No entanto, neste trabalho concentramos grande parte do esforço em obter resultados de mecânica estatística de equilíbrio. Justificamos a utilização de mecânica estatística de equilíbrio através da hipótese de que o valor médio do custo psicológico de discordância deve caracterizar o estado macroscópico do sistema. Uma vez que a informação disponível sobre o nosso sistema é incompleta, a ferramenta apropriada para tratar o sistema é o formalismo de probabilidades. Dados certos vínculos, podemos determinar a distribuição menos informativa que satisfaça estes vínculos. Vamos abaixo utilizar este método para encontrar a distribuição de estados, pois o processo da aproximação variacional que utilizaremos mais à frente se vale de métodos similares. Utilizando esta dedução, encontramos o parâmetro β , que surge como multiplicador de Lagrange na maximização da entropia, para forçar o vínculo de que $\langle H \rangle_\beta = E$.

A distribuição de equilíbrio pode ser obtida através da maximização do funcional de entropia dados alguns vínculos, como é canonicamente feito em textos de mecânica estatística. A entropia de uma distribuição segundo Shannon é dada por

$$S[P] = - \int d\mu(\{J_i\}) P(\{J_i\}|\{S^\mu\}) \log(P(\{J_i\}|\{S^\mu\})),$$

onde $\{J_i\}$ é o conjunto de vetores de opinião dos agentes, e $d\mu(\{J_i\})$ é a medida de integração sobre o espaço de conjuntos. Assumindo que o valor médio do hamiltoniano sobre a dada distribuição é conhecido $\langle H \rangle = E$, podemos então maximizar a entropia dado este vínculo e obter uma distribuição de probabilidades que usa o mínimo de hipóteses, ou seja, a menos informativa. Teremos então o funcional lagrangiano a ser maximizado

$$S' = S[P] - \mu \left(\int d\mu(\{J\}) P(\{J_i\}|\{S^\mu\}) - 1 \right) - \beta \left(\int d\mu(\{J\}) H(\{J_i\}|\{S^\mu\}) P(\{J_i\}|\{S^\mu\}) - E \right).$$

Tomando variações $P + \delta P$, obtemos expandindo em primeira ordem

$$\frac{\delta S'}{\delta P(\{J_i\}|\{S^\mu\})} = -1 - \log(P(\{J_i\}|\{S^\mu\})) - \mu - \beta H(\{J_i\}|\{S^\mu\}).$$

Impondo então que a distribuição deva ser de máxima entropia, teremos $\frac{\delta S'}{\delta P(\{J_i\}|\{S^\mu\})} = 0$ e portanto

$$P(\{J_i\}|\{S^\mu\}) \propto e^{-\beta H(\{J_i\}|\{S^\mu\})},$$

com uma constante de normalização adicional, que é precisamente a função de partição $Z(\beta)$. O valor do multiplicador de Lagrange β é determinado implicitamente através do requerimento que

$$\int d\mu(\{J\}) H(\{J_i\}|\{S^\mu\}) \frac{e^{-\beta H(\{J_i\}|\{S^\mu\})}}{Z} = E. \quad (3.7)$$

Em sistemas físicos podemos prontamente identificar o parâmetro β com o inverso da temperatura, no entanto, não sabemos o que seria uma temperatura numa sociedade de agentes. Em termos da mecânica estatística, a temperatura nos dá uma escala para determinar a importância de mudanças de energia no sistema. Podemos, no entanto, ler a equação 3.7 como uma equação que nos dá E como função de β . Então β nos diz quão alta vai ser a energia média do sistema. Especificamente, valores altos de β correspondem a energias médias baixas e valores baixos de β a energias médias altas. No caso do modelo idealizado aqui, a energia corresponde ao custo de discordância entre pares de agentes, de forma que podemos associar o parâmetro β com uma pressão por conformidade, ou no jargão da psicologia *peer pressure*. Um valor de β alto corresponde a uma forte pressão por conformidade, ou seja, um ambiente em que discordância não é tolerada, levando a uma minimização mais exigente do custo de discordância. No caso de β baixo, teríamos um ambiente mais tolerante, no qual certa quantidade de discordância é aceita.

Talvez a hipótese mais delicada deste modelo é a de tratarmos sociedades homogêneas com respeito a estratégias cognitivas. As redes sociais observadas em experimentos de sociologia são notórias por terem altos coeficientes de aglomeração e podemos, portanto, supor a existência de grupos entre os quais há pouco contato em nossa sociedade (como qualquer um pode atestar). No entanto, teríamos de justificar que esta divisão tem algo a ver com os traços cognitivos modelados em nosso sistema.

Poderíamos então estudar sociedades em que não temos uma distribuição uniforme de δ , mas ao menos alguma diversidade; no entanto, isto complica um pouco o tratamento do modelo.

Estudando o modelo com aprendizado *on-line*, poderíamos substituir a regra de aprendizado por uma regra em que δ é substituído pela versão individualizada δ_i , sem maiores problemas. Teríamos em analogia à equação 3.1 a seguinte regra de aprendizado

$$J_i^{novo} = J_i + \frac{\alpha}{D} (\delta_i + (1 - \delta_i) \Theta(-J_i \cdot S^\mu J_j \cdot S^\mu)) \Theta(J_j \cdot S^\mu) S^\mu. \quad (3.8)$$

Ainda que o estudo deste sistema não apresente nenhum desafio adicional, teríamos de determinar uma forma sistemática de estudar estas distribuições sobre δ s, o que é bastante trabalhoso. No caso do modelo hamiltoniano, as interações são simétricas, de forma que acrescentar δ s heterogêneos levaria a um $\delta_{(i,j)}$ para cada elo do grafo social. Para ver isso basta reescrevermos a contribuição de cada aresta do grafo social como

$$V_{ij} = \frac{1}{2} \left(-\frac{1 + \delta_i}{2} h_i h_j + \frac{1 - \delta_i}{2} |h_i h_j| - \frac{1 + \delta_j}{2} h_i h_j + \frac{1 - \delta_j}{2} |h_i h_j| \right). \quad (3.9)$$

No caso de $\delta_i = \delta_j$, claramente recuperamos o caso homogêneo. No entanto, considerando o caso acima teremos simplesmente a interação original com um novo $\delta_{ij} = \frac{1}{2}(\delta_i + \delta_j)$.

$$V_{ij} = -\frac{1 + \delta_{ij}}{2} h_i h_j + \frac{1 - \delta_{ij}}{2} |h_i h_j|.$$

Então ao invés de obter agentes com diferentes δ s, obtemos agentes que julgam diferentemente informações provenientes de diferentes agentes. Claramente isso não é o que queríamos estudar quando procuramos estudar sociedades heterogêneas. Desta forma, escolhemos estudar o caso de agentes com estratégias cognitivas homogeneamente distribuídas.

3.2.2 Métodos de Monte Carlo e Cadeias de Markov

Uma cadeia de Markov é um processo estocástico a tempo discreto sem memória, isto é, um processo tal que a probabilidade de um evento ocorrer num instante $n + 1$ só depende do estado do sistema no instante n . Escrevendo em linguagem de probabilidades condicionais, podemos escrever

$$P(S_{n+1} = X | S_n, S_{n-1}, S_{n-2}, \dots) = P(S_{n+1} = X | S_n). \quad (3.10)$$

Pode-se demonstrar que uma cadeia de Markov cuja probabilidade de transição satisfaz a condição de balanceamento detalhado para uma dada distribuição tem como probabilidade de equilíbrio esta mesma distribuição. Uma demonstração particularmente simples está no texto didático de

Binney (22). Uma cadeia de Markov é definida por uma densidade de probabilidade de transição (para sistemas cujos estados podem ser descritos por variáveis discretas seria uma probabilidade de transição) entre estados do sistema

$$w(\{J_i\} \rightarrow \{J'_i\}) = P(S_{n+1} = \{J'_i\} | S_n = \{J_i\}).$$

A densidade de probabilidade de transição deve satisfazer a condição de normalização

$$\int d\mu(\{J'_i\}) w(\{J_i\} \rightarrow \{J'_i\}) = 1, \quad (3.11)$$

ou seja o sistema tem de ir para algum lugar ao sair do estado $\{J_i\}$.

A condição de balanceamento detalhado ou microreversibilidade é dada por

$$\frac{w(\{J_j\} \rightarrow \{J'_j\})}{w(\{J'_j\} \rightarrow \{J_j\})} = \frac{P(\{J'_j\})}{P(\{J_j\})} = e^{-\beta(H(\{J'_j\}) - H(\{J_j\}))}, \quad (3.12)$$

onde a distribuição dada é a distribuição de equilíbrio descrita acima. Ou seja, escolhendo uma probabilidade de transição que satisfaça esta condição, podemos simular uma cadeia de Markov e a distribuição de equilíbrio desta cadeia será a distribuição de equilíbrio do nosso sistema. Desta forma, podemos após um tempo longo, chamado de termalização, assumir que amostrando os estados da cadeia de Markov estamos amostrando de uma distribuição muito próxima da distribuição de equilíbrio do nosso sistema.

A escolha canônica para simulações de Monte Carlo é o algoritmo de Metropolis (23). Este consiste em aceitar um novo estado sempre que este diminuir o valor de $H(\{J_i\})$, e aceitá-lo com probabilidade $e^{-\beta(H(\{J'_i\}) - H(\{J_i\}))}$ caso o novo valor de $H(\{J_i\})$ seja maior que o original. Verificamos rapidamente que a condição de microreversibilidade é satisfeita. Caso $H(\{J'_i\}) > H(\{J_i\})$, temos

$$\frac{w(\{J_j\} \rightarrow \{J'_j\})}{w(\{J'_j\} \rightarrow \{J_j\})} = \frac{e^{-\beta(H(\{J'_j\}) - H(\{J_j\}))}}{1} = e^{-\beta(H(\{J'_j\}) - H(\{J_j\}))}.$$

No caso de $H(\{J'_i\}) < H(\{J_i\})$, temos

$$\frac{w(\{J_j\} \rightarrow \{J'_j\})}{w(\{J'_j\} \rightarrow \{J_j\})} = \frac{1}{e^{-\beta(H(\{J_i\}) - H(\{J'_i\}))}} = e^{-\beta(H(\{J'_i\}) - H(\{J_j\}))}.$$

O algoritmo de Metropolis apresenta portanto uma possível cadeia de Markov para amostrarmos

a distribuição de Gibbs-Boltzmann de um dado sistema.

3.2.3 Wang-Landau Methods

Uma classe de métodos estatísticos que permite estudar regiões extensas do diagrama de fase a partir de uma única simulação é o método de Wang-Landau (17). A ideia inicial do método é reescrever a integral (no caso de variáveis discretas uma somatória) do valor médio das variáveis como uma integral sobre os valores da energia do sistema ponderado pela degenerescência de cada nível de energia. Este método já havia sido proposto anteriormente, no entanto o trabalho de Wang e Landau(17) propõe um método de Monte Carlo para obter a degenerescência dos estados numericamente. A equação 3.4 ficaria então

$$\langle H \rangle = \int d\mu(\{J_i\})P(\{J_i\})H(\{J_i\}) = \int dE\Omega(E)\frac{e^{-\beta E}}{Z}E, \quad (3.13)$$

onde a integral seria sobre todos os valores possíveis da energia, sendo agora uma integral em apenas uma dimensão. A probabilidade do valor do hamiltoniano de um estado ser E é dada por

$$P(E) = \Omega(E)\frac{e^{-\beta E}}{Z},$$

onde Z é a função de partição dada por

$$Z = \int d\mu(\{J_i\})e^{-\beta H(\{J_i\})}.$$

Podemos reescrever também Z usando $\Omega(E)$, obtendo

$$Z = \int dE\Omega(E)e^{-\beta E}.$$

Claramente, todas as grandezas da equação 3.13 são conhecidas, exceto a degenerescência de cada nível de energia $\Omega(E)$. Se conhecêssemos esta degenerescência, poderíamos calcular todas as grandezas termodinâmicas do sistema. Vale notar que neste caso só temos acesso à dependência da energia das grandezas. No caso do modelo de Ising, se quiséssemos considerar um campo não-nulo, teríamos também a dependência da magnetização, teríamos então que escrever a integral como

$$\langle H \rangle = \int dE dm\Omega(E, m)P(E, m)E, \quad (3.14)$$

ou para obter a magnetização média

$$\langle m \rangle = \int dE dm \Omega(E, m) P(E, m) m, \quad (3.15)$$

no entanto, isso não altera em nada o resto do argumento. O método de Wang-Landau consiste em uma técnica para calcular a degenerescência ou densidade de estados para um sistema dado. A ideia é que se escolhêssemos estados ao acaso com uma probabilidade $P_{WL}(\{J_i\}) \propto \frac{1}{\Omega(H(\{J_i\}))}$ o histograma das energias dos estados sorteados deveria ser uniforme, uma vez que

$$P_{WL}(H = E) = \int d\mu(\{J_i\}) \delta(H(\{J_i\}) - E) P_{WL}(\{J_i\}) = \Omega(E) \frac{1}{Z_{WL} \Omega(E)} = \frac{1}{Z_{WL}}.$$

No entanto, ainda não temos a degenerescência de nossos níveis de energia para poder sortear estados e aceitá-los de forma a gerar a distribuição dada. O método consiste portanto em um processo conjunto de elaborar histogramas e verificar se estes são uniformes, e ao mesmo tempo melhorar cada vez mais nossa aproximação de $\Omega(E)$.

Apesar de ter havido avanços em demonstrar a validade do algoritmo de Wang-Landau (WL) para aproximar a densidade de estados de um sistema (24), estas demonstrações estão fora do escopo deste trabalho. Vamos meramente mencionar o funcionamento do algoritmo. O algoritmo consiste em sortear estados e aceitá-los de acordo com a probabilidade de transição

$$P(\{J_{new}\}|\{J_{old}\}) = e^{g_{old} - g_{new}},$$

onde $g(\{J\})$ é uma aproximação para $\log(\Omega)$, inicialmente igual a 0. Estados são sorteados e cada vez que um nível de energia é visitado incrementamos $g(E)$ por um fator $\log(f)$, onde $f = e$ inicialmente. Vale notar que a probabilidade de transição procura sempre visitar energias com uma densidade de estados mais baixa, contrapondo-se ao efeito entrópico que empurra o sistema para energias com alta degenerescência. Depois de um número dado de passos, verifica-se o histograma dos níveis de energia visitados. Caso este satisfaça um critério de uniformidade, tomamos $f_{new} = \sqrt{f}$ e começamos a registrar um novo histograma de visitação dos níveis de energia. Este processo se repete até que $\log(f) < \epsilon$. Ao final do processo, e^g é usado como aproximação da densidade de estados Ω .

Modificações deste método surgem recorrentemente na literatura, adaptando o método a tarefas específicas. No caso do modelo estudado aqui, empregamos o algoritmo de Wang-Landau com janelas adaptativas (25). Basicamente, o método consiste em procurar o domínio

de energias acessíveis ao sistema, e amostrar os estados por etapas em janelas deste domínio para obter uma precisão maior.

3.3 Aproximação de Campo Médio

Aproximações de campo médio são uma classe geral de métodos que podem ser usados para aproximar distribuições de probabilidade em uma grande variedade de situações. No contexto mais geral podemos definir uma aproximação de campo médio como uma projeção em um subespaço de distribuições tratáveis de um problema estatístico não-tratável (26). No caso geral, gostaríamos de calcular médias sobre uma distribuição não-tratável $P(\sigma)$,

$$\langle f \rangle = \int d\mu(\sigma) f(\sigma) P(\sigma), \quad (3.16)$$

onde a integral não tem solução analítica. Normalmente isso se deve a termos de interação entre as variáveis, como no caso do modelo de Ising em duas dimensões. A ideia dos métodos de campo médio é tomar uma família de distribuições tratáveis, e dentre elas escolher a distribuição que melhor aproxima a distribuição não-tratável P . Isto pode ser feito de diversas maneiras, mas vamos nos ater ao método baseado na teoria de geometria da informação, no qual minimizamos a divergência de Kullback-Leibler (ou maximizamos a entropia relativa). A divergência de Kullback-Leibler entre duas distribuições de probabilidade P_1 e P_2 é dada por

$$D_{KL}(P_1 \| P_2) = \int d\mu(\sigma) P_1(\sigma) \log \left(\frac{P_1(\sigma)}{P_2(\sigma)} \right). \quad (3.17)$$

Dada uma família de distribuições parametrizada por uma série de parâmetros α , podemos então escrever a divergência de Kullback-Leibler entre uma distribuição da família tratável e a distribuição-alvo como uma média sobre a família de distribuições tratáveis

$$E(\alpha) = D_{KL}(P_\alpha \| P) = \int d\mu(\sigma) P_\alpha(\sigma) \log \left(\frac{P_\alpha(\sigma)}{P(\sigma)} \right). \quad (3.18)$$

A função $E(\alpha)$ pode agora ser minimizada normalmente de forma a obtermos uma distribuição sob a qual as integrais se tornam tratáveis. A sutileza e complexidade do método dependem da escolha da família de distribuições. No que se segue abaixo, no entanto, vamos usar argumentos variacionais para evitar a especificação do formato da família de distribuições a ser usada, que será deduzido da minimização da divergência KL .

3.3.1 Exemplo: Modelo de Ising

Como exemplo, podemos considerar o modelo de Ising. Temos a distribuição sobre variáveis de spin interagindo em uma rede quadrada bidimensional

$$P_0(\{\sigma_i\}) = \frac{e^{\beta J \sum \sigma_i \sigma_j}}{Z}. \quad (3.19)$$

As médias sobre a distribuição envolvem somas tediosas para redes finitas, e para redes infinitas têm de ser realizadas através de técnicas matemáticas mais refinadas, como o método de matrizes de transferência. No entanto, se a distribuição fosse da forma

$$P'(\{\sigma_i\}) = \prod_i P_i(\sigma_i), \quad (3.20)$$

as médias poderiam facilmente ser efetuadas, pois as variáveis estariam desacopladas. Vamos abaixo utilizar um argumento variacional para determinar a forma ótima da distribuição sobre sítios P_i para aproximar a distribuição P_0 . Temos

$$D_{KL}(P' \| P_0) = \sum_{\{\sigma_i\}} \prod_i P_i(\sigma_i) \log \left(\frac{\prod_i P_i(\sigma_i)}{P_0(\{\sigma_i\})} \right) = \quad (3.21)$$

$$- \sum_{(i,j)} \sum_{\sigma_i, \sigma_j} P_i(\sigma_i) P_j(\sigma_j) \beta J \sigma_i \sigma_j + \log(Z) + \sum_i \sum_{\sigma_i} P_i(\sigma_i) \log(P_i(\sigma_i)). \quad (3.22)$$

Tomando uma derivada funcional com relação ao valor da probabilidade do sítio i para o valor σ_i , teremos

$$\frac{\delta D_{KL}}{\delta(P_i(\sigma_i))} = - \sum_{j \in N(i)} \sum_{\sigma_j} P_j(\sigma_j) \beta J \sigma_i \sigma_j + 1 + \log(P_i(\sigma_i)). \quad (3.23)$$

Na equação acima, a expressão $j \in N(i)$ se refere a todos os sítios j que estão na vizinhança do sítio i . Claramente, só estes sítios tem uma contribuição não-nula para a derivada funcional. Igualando a expressão a zero teremos então, a menos de uma constante de normalização

$$P_i(\sigma_i) \propto \exp \left(\beta J \sigma_i \sum_{j \in N(i)} \langle \sigma_j \rangle \right). \quad (3.24)$$

A média $m_j = \langle \sigma_j \rangle$ é feita sob a distribuição P' , o que nos dá uma série de equações autoconsistentes do tipo

$$m_i = \sum_{\{\sigma\}} \sigma_i \prod_j P_j(\sigma_j) = \sum_{\sigma_i} \sigma_i \frac{e^{\beta J \sigma_i \sum_j m_j}}{Z'_i}. \quad (3.25)$$

Z'_i é a constante de normalização da distribuição sobre o sítio i , e a soma é sobre os vizinhos de i . Podemos resolver esta equação de duas formas, a primeira é considerar que a rede é homogênea e invariante por translações, de forma que todos os m_i devem ser iguais, teremos então uma única equação, que nos dá

$$m = \sum_{\sigma_i} \sigma_i \frac{e^{\beta J \sigma_i \sum m}}{Z'_i} = \tanh(\beta J N m), \quad (3.26)$$

que é a conhecida equação de campo médio para o modelo de Ising, onde N é o número de coordenação da rede. Podemos obter um pouco mais de informação se não fizermos esta suposição e procurarmos encontrar um conjunto de m_j que satisfaça o conjunto de equações 3.25. Reescrevendo-as, teríamos

$$m_i = \tanh(\beta J \sum_{N(i)} m_j), \quad (3.27)$$

que poderíamos agora resolver usando técnicas de ponto fixo. Isto nos permite também utilizar a técnica de campo médio para redes não-homogêneas, permitindo ampliar o espectro de possibilidades do método. Vale notar que diferentemente de métodos similares, não fizemos nenhuma hipótese *a priori* sobre a forma da distribuição de probabilidade, afora o fato desta ser fatorizável sobre os sítios da rede. Métodos similares deduzem a mesma solução de campo médio para o modelo de Ising usando hipóteses sobre a forma exata da hamiltoniana, ou mesmo assumindo uma forma sobre a distribuição de probabilidade. No caso de spins a diferença não é muito grande, pois a probabilidade é especificada por apenas um parâmetro ($P(\sigma) = 1 - P(-\sigma)$), mas para sistemas mais complexos, como a versão hamiltoniana do nosso modelo, é crucial poder fazer o mínimo de hipóteses sobre o formato da distribuição. Vale notar também que o processo que foi realizado acima equivale a encontrar um ponto em um subespaço que minimiza a distância para um ponto fora deste subespaço, como no caso de um plano e um ponto fora deste plano. O que permite realizar este processo para uma distribuição não-tratável é a assimetria da divergência KL que pode ser calculada usando-se médias apenas sobre as distribuições tratáveis.

Vale notar que no tratamento acima, diferentemente do tratamento dado ao modelo de Curie-Weiss, obtivemos uma aproximação para a distribuição de Gibbs (e consequentemente da termodinâmica) a partir de uma hipótese sobre o formato da distribuição de probabilidade do sistema. Apesar de obtermos essencialmente o mesmo resultado, podemos refinar nossa aproximação usando outras distribuições, como pares de sítios, triplas e por aí adiante. Além disso, a equação 3.25 nos permite analisar o modelo de Ising desprezando flutuações, pois os

spins interagem com o valor esperado dos spins vizinhos, mas não carece da hipótese de uma rede completamente conexa ou ainda invariante por translação. Desta forma podemos estudar o comportamento do modelo em redes afins sem usar métodos de Monte-Carlo.

3.3.2 Aproximação de Campo Médio para o Modelo de Fundações Morais

Vamos agora repetir o processo acima para o caso do nosso modelo da seção 3.2. Tomamos agora uma distribuição que seja fatorizável nos assuntos discutidos, além dos sítios. No entanto, isto não é uma hipótese adicional, uma vez que a probabilidade definida pelo hamiltoniano original já era fatorizável sobre os assuntos. A distribuição correta é dada por

$$P_0(\{J_i\}|\{S^\mu\}, \beta) = \frac{1}{Z} \exp \left(-\frac{\beta}{P} \sum_{\mu} \sum_{(i,j)} \frac{-1-\delta}{2} h_i^\mu h_j^\mu + \frac{1-\delta}{2} |h_i^\mu| |h_j^\mu| \right). \quad (3.28)$$

Em analogia aos parágrafos anteriores, vamos definir a distribuição fatorizável sobre sítios

$$P'(\{J_i\}) = \prod_i P_i(J_i). \quad (3.29)$$

Procedemos calculando a divergência de Kullback-Leibler, da qual omitimos o logaritmo da função de partição, que é irrelevante para a maximização, por não depender de $P_i(J_i)$,

$$D_{KL}(P' \| P_0) = \int d\mu(\{J_i\}) \prod_i P_i(J_i) \left(\sum_i \log(P_i(J_i)) + \frac{\beta}{P} \sum_{\mu} \sum_{(i,j)} \left(\frac{-1-\delta}{2} h_i^\mu h_j^\mu + \frac{1-\delta}{2} |h_i^\mu| |h_j^\mu| \right) \right). \quad (3.30)$$

A integral é feita sobre a casca hipersférica de dimensão 5, a medida $d\mu(\{J_i\})$ se refere a isto. Reescrevendo, temos então

$$\begin{aligned} D_{KL} &= \sum_i \int d\mu(J_i) P_i(J_i) \log(P_i(J_i)) \\ &+ \frac{\beta}{P} \sum_{\mu} \sum_{(i,j)} \int d\mu(J_i) d\mu(J_j) P_i(J_i) P_j(J_j) \left(\frac{-1-\delta}{2} h_i^\mu h_j^\mu + \frac{1-\delta}{2} |h_i^\mu| |h_j^\mu| \right). \end{aligned} \quad (3.31)$$

Tomando então a derivada funcional em relação ao valor de $P_i(J_i)$, temos

$$\begin{aligned} \frac{\delta \mathcal{D}_{KL}}{\delta(P_i(J_i))} &= 1 + \log(P_i(J_i)) \\ &+ \frac{\beta}{P} \sum_{\mu} \sum_{j \in N(i)} \int d\mu(J_j) P_j(J_j) \left(\frac{-1-\delta}{2} h_i^{\mu} h_j^{\mu} + \frac{1-\delta}{2} |h_i^{\mu}| |h_j^{\mu}| \right), \end{aligned} \quad (3.32)$$

onde $N(i)$ é o conjunto de nós vizinhos ao nó i . Aqui vemos que a expressão só depende de um conjunto de médias sobre a probabilidade $P_i(J_i)$, e esta distribuição é completamente definida por este conjunto de valores. Dizemos então que o estado deste sistema é completamente determinado por estes valores, que chamamos de parâmetros de ordem, por nos dizerem em que fase o sistema se encontra (ordenada ou desordenada). Podemos então definir estas grandezas como $m_j^{\mu} = \langle J_j \cdot S^{\mu} \rangle$ e $r_j^{\mu} = \langle |J_j \cdot S^{\mu}| \rangle$ e encontramos a expressão da distribuição como função destas grandezas

$$1 + \log(P_i(J_i)) = -\frac{\beta}{P} \sum_{\mu} \sum_{j \in N(i)} \left(\frac{-1-\delta}{2} h_i^{\mu} m_j^{\mu} + \frac{1-\delta}{2} |h_i^{\mu}| r_j^{\mu} \right). \quad (3.33)$$

O que nos dá

$$P_i(J_i) = \frac{1}{Z_i} \exp \left(-\frac{\beta}{P} \sum_{\mu} \sum_{j \in N(i)} \left(-\frac{1+\delta}{2} h_i^{\mu} m_j^{\mu} + \frac{1-\delta}{2} |h_i^{\mu}| r_j^{\mu} \right) \right). \quad (3.34)$$

Note que o método determina os parâmetros de ordem automaticamente, sem que tenhamos de intuir quais as grandezas importantes no modelo. Este procedimento é especialmente interessante em sistemas nos quais os parâmetros de ordem não podem ser prontamente determinados.

A partir destas probabilidades podemos escrever as equações análogas às equações 3.25 para o modelo de Ising. Teremos neste caso

$$m_i^{\mu} = \int d\mu(J_i) J_i \cdot S^{\mu} \frac{1}{Z_i} \exp \left(-\frac{\beta}{P} \sum_{\mu} \sum_{j \in N(i)} \left(-\frac{1+\delta}{2} h_i^{\mu} m_j^{\mu} + \frac{1-\delta}{2} |h_i^{\mu}| r_j^{\mu} \right) \right) \quad (3.35)$$

$$r_i^{\mu} = \int d\mu(J_i) |J_i \cdot S^{\mu}| \frac{1}{Z_i} \exp \left(-\frac{\beta}{P} \sum_{\mu} \sum_{j \in N(i)} \left(-\frac{1+\delta}{2} h_i^{\mu} m_j^{\mu} + \frac{1-\delta}{2} |h_i^{\mu}| r_j^{\mu} \right) \right) \quad (3.36)$$

Assim como no caso do modelo de Ising, aqui podemos tomar duas maneiras alternativas de resolver o conjunto de equações 3.35 e 3.36. Se escolhermos usar argumentos de homogeneidade

e invariância por translação, obtemos as equações autoconsistentes para o campo médio

$$m^\mu = \int d\mu(J) J \cdot S^\mu \frac{1}{Z_i} \exp\left(-\frac{\beta}{P} N \sum_\mu \left(-\frac{1+\delta}{2} h^\mu m^\mu + \frac{1-\delta}{2} |h^\mu| r^\mu\right)\right) \quad (3.37)$$

$$r^\mu = \int d\mu(J) |J \cdot S^\mu| \frac{1}{Z_i} \exp\left(-\frac{\beta}{P} N \sum_\mu \left(-\frac{1+\delta}{2} h^\mu m^\mu + \frac{1-\delta}{2} |h^\mu| r^\mu\right)\right). \quad (3.38)$$

Diferentemente do caso do modelo de Ising, aqui temos não apenas uma equação e um parâmetro de ordem, mas sim $2M$ equações, onde M é o número de assuntos sendo discutidos na nossa sociedade, e $2M$ parâmetros de ordem. Além disso, como temos uma integral e não uma soma sobre um spin somente, não podemos escrever uma forma analítica para as equações autoconsistentes, que terão de ser resolvidas através de integração numérica. No caso de termos uma sociedade numa rede com N nós, teríamos $2MN$ equações e parâmetros, o que dificulta a análise do problema para tamanhos de rede maiores, uma vez que a iteração das equações envolve uma integral numérica em quatro dimensões.

Podemos prontamente identificar uma interpretação para os parâmetros de ordem no nosso modelo. O produto escalar da matriz moral (J_i) de um agente com um assunto em discussão foi anteriormente interpretada como o julgamento moral de um agente sobre o dado assunto. Mantendo os vetores normalizados podemos manter este valor entre 1 (aprovação, aceitação) e -1 (desaprovação, rejeição), passando pela indiferença, quando o produto é igual a 0. Os parâmetros de ordem seriam então o julgamento moral médio de um agente sobre um assunto μ (m^μ) e o valor absoluto médio deste julgamento (r^μ). Podemos dizer então que o parâmetro m^μ indica se há ou não alguma polarização e o parâmetro r^μ indica com quanta certeza um julgamento moral é tomado sempre que ele é tomado.

4 Resultados

Investigaremos o comportamento do sistema obtendo diagramas de fase com os diferentes tratamentos desenvolvidos. Em seguida analisaremos o comportamento do sistema em relação a perturbações nos assuntos em discussão. Para o modelo de campo médio, será obtida uma expressão exata para a divergência de Kullback-Leibler entre dois estados diferentes. No modelo hamiltoniano, usaremos a distância euclidiana entre histogramas do parâmetro de ordem para medir a distância entre estados.

Começamos a seção discutindo como obtivemos os histogramas para cada um dos paradigmas considerados e os métodos que utilizamos para determinar transições de fase em cada um dos paradigmas. Discutimos alguns exemplos de histogramas para ilustrar cada um dos casos e as fases encontradas nos estudos. Em seguida discutimos os diagramas de fases em cada um dos paradigmas para uma variedade de parâmetros, para uma variedade de conjuntos de assuntos. Particular interesse é dedicado ao caso mais simples de apenas um vetor, por este nos dar uma posição privilegiada para discutir o efeito dos diversos parâmetros do sistema. Por último procuramos estudar como o sistema responde a perturbações nos assuntos após alcançar um estado de equilíbrio. Desenvolvemos um tratamento específico para estudar o caso da aproximação de campo médio, o que possibilita um tratamento mais rigoroso usando a distribuição de probabilidade completa dos vetores de opinião, ao invés da distribuição de probabilidade da projeção sobre os assuntos.

4.1 Histogramas e Caracterização das Fases

A determinação da fase em que se encontra um determinado sistema normalmente se baseia em grandezas estatísticas, cujos valores caracterizam uma fase ou outra. O exemplo canônico é como sempre o modelo de Ising, em que as fases ferromagnética e paramagnética são determinadas pelo valor da magnetização a campo nulo (nulo na fase paramagnética e não-nulo na fase ferromagnética). Muitas outras grandezas e critérios podem ser considerados, como uma quebra de analiticidade no calor específico ou algum potencial termodinâmico, mas o método usado especificamente em cada caso depende do problema em questão.

Temos três paradigmas nos quais gostaríamos de estabelecer diagramas de fase e possivelmente linhas de transição de fase, no entanto, os três paradigmas pedem abordagens diferentes. O paradigma no qual possuímos a maior informação sobre o sistema é certamente o de campo

médio, em que podemos calcular explicitamente a distribuição de probabilidade dos vetores de opinião a partir das equações auto-consistentes para os parâmetros de ordem. Analisaremos então o caso do campo-médio primeiramente. No caso do modelo hamiltoniano, a obtenção da forma analítica da distribuição não é possível, pois envolve a integração sobre um número muito grande de variáveis. No entanto, podemos utilizar o método de Wang-Landau descrito acima para obter um retrato fiel da termodinâmica do sistema, a partir da densidade de estados da energia. A partir desta, podemos calcular grandezas como o calor específico, e a partir de médias confiáveis destas podemos determinar temperaturas de transição de fase e, por conseguinte, determinar um diagrama de fases. O paradigma que coloca o maior desafio no sentido da determinação das fases é o modelo baseado em agentes, pois sequer podemos certeza da existência de um regime estacionário neste caso. Vamos investigar a existência de uma fase estacionária e verificar que, para uma variedade de parâmetros, o sistema não relaxa para uma distribuição estável.

4.1.1 Aproximação de Campo Médio

Na aproximação de campo médio, usamos as equações autoconsistentes 3.35 e 3.36 para calcular os parâmetros de ordem. Especificamente, consideramos as equações como relações de recorrência,

$$m^\mu = \int d\mu(J) J \cdot S^\mu \frac{1}{Z_i} \exp\left(-\frac{\beta}{P} N \sum_\mu \left(-\frac{1+\delta}{2} h^\mu m^\mu + \frac{1-\delta}{2} |h^\mu| r^\mu\right)\right) = f^\mu(\{m^\nu, r^\nu\})$$

$$r^\mu = \int d\mu(J) |J \cdot S^\mu| \frac{1}{Z_i} \exp\left(-\frac{\beta}{P} N \sum_\mu \left(-\frac{1+\delta}{2} h^\mu m^\mu + \frac{1-\delta}{2} |h^\mu| r^\mu\right)\right) = g^\mu(\{m^\nu, r^\nu\})$$

e consideramos que o valor dos parâmetros de ordem para um dado conjunto de parâmetros $\delta, \beta, \{S^\mu\}$ é o ponto fixo destas equações:

$$m^\mu(\delta, \beta, \{S^\nu\}) = f(\{m^\sigma(\delta, \beta, \{S^\nu\}), r^\sigma(\delta, \beta, \{S^\nu\})\}) \quad (4.1)$$

$$r^\mu(\delta, \beta, \{S^\nu\}) = g(\{m^\sigma(\delta, \beta, \{S^\nu\}), r^\sigma(\delta, \beta, \{S^\nu\})\}). \quad (4.2)$$

Podemos então inicializar os parâmetros de ordem em um valor arbitrário e usar as equações acima como relações de recorrência, criando uma seqüência de parâmetros de ordem até que estes converjam. Os valores de convergência serão então os parâmetros de ordem corretos para

os parâmetros dados. Fazemos então

$$m_n^\mu = f(\{m_{n-1}^\sigma, r_{n-1}^\sigma\}) \quad (4.3)$$

$$r_n^\mu = g(\{m_{n-1}^\sigma, r_{n-1}^\sigma\}). \quad (4.4)$$

Podemos ver na figura abaixo 4.1 um exemplo do valor dos parâmetros de ordem para um conjunto de parâmetros ($\delta = 0.25, P = 1$).

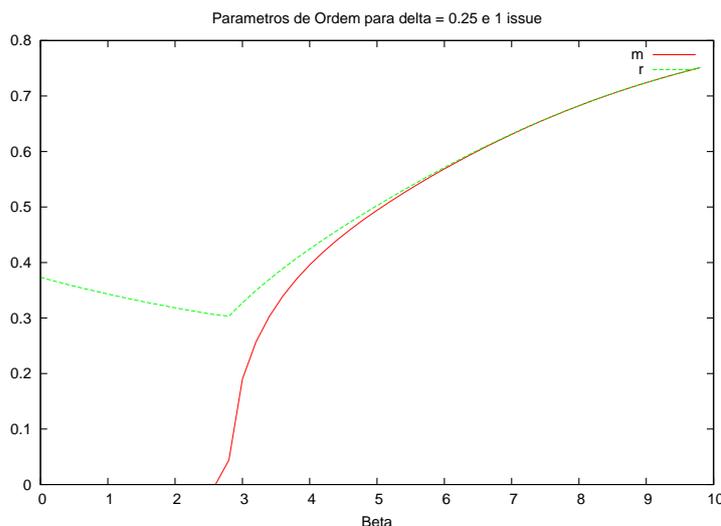


Figura 4.1: Vemos acima os parâmetros m e r para um sistema com apenas 1 assunto na aproximação de campo médio.

Como comparação mostramos abaixo também os parâmetros de ordem para um sistema análogo mas com dois assuntos ortogonais (4.2).

Em ambos os casos podemos ver uma mudança de comportamento em ambos os parâmetros m e r , os valores de m passam de valores nulos a valores positivos crescentes com a temperatura e os valores de r tem um ponto de mínimo. Para verificar que estas mudanças representam de fato uma mudança na fase do sistema, investigamos a distribuição de probabilidades dos vetores ao longo dos assuntos. Podemos verificar, ao investigar os histogramas, uma quebra de simetria ao longo do eixo definido pelo vetor assunto no caso da figura 4.1. Vemos que para pressões de pares mais baixas (temperaturas altas) temos um histograma perfeitamente simétrico, enquanto para pressões mais altas (temperaturas mais baixas), temos um viés para o quadrante de projeção

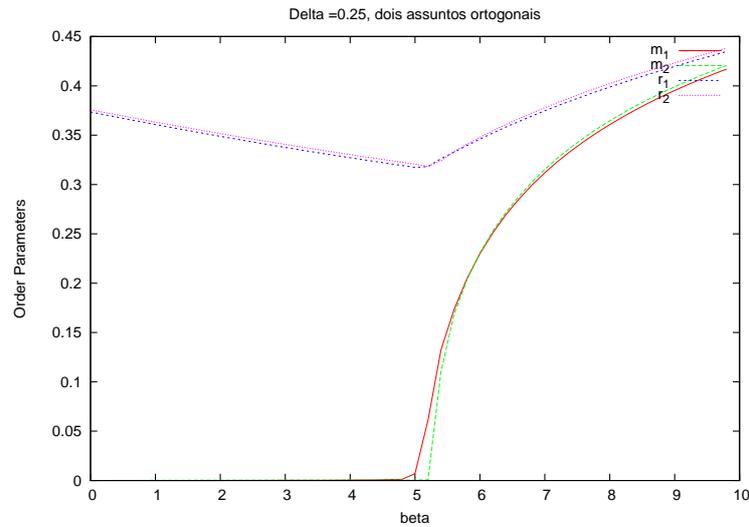


Figura 4.2: Vemos acima os parâmetros m e r para um sistema com 2 assuntos ortogonais na aproximação de campo médio.

positiva no eixo dado.

Claramente, valores não-nulos dos parâmetros de ordem indicam uma projeção média não-nula nos assuntos, e portanto uma ordenação do sistema. Seguindo este raciocínio, usamos o valor dos parâmetros de ordem m como indicadores da fase em que o sistema se encontra. Partindo deste raciocínio podemos simplesmente usar um método de bissecção para determinar a linha de transição no plano $\delta \times \beta$. Vemos abaixo um exemplo de linhas de diagramas de fases para um assunto (figura 4.4).

Implementação Computacional da Integração

As integrais que precisamos calcular para obter os parâmetros de ordem são todas integrais sobre a hipersfera de raio 1 em cinco dimensões. Usando o formalismo geral, podemos

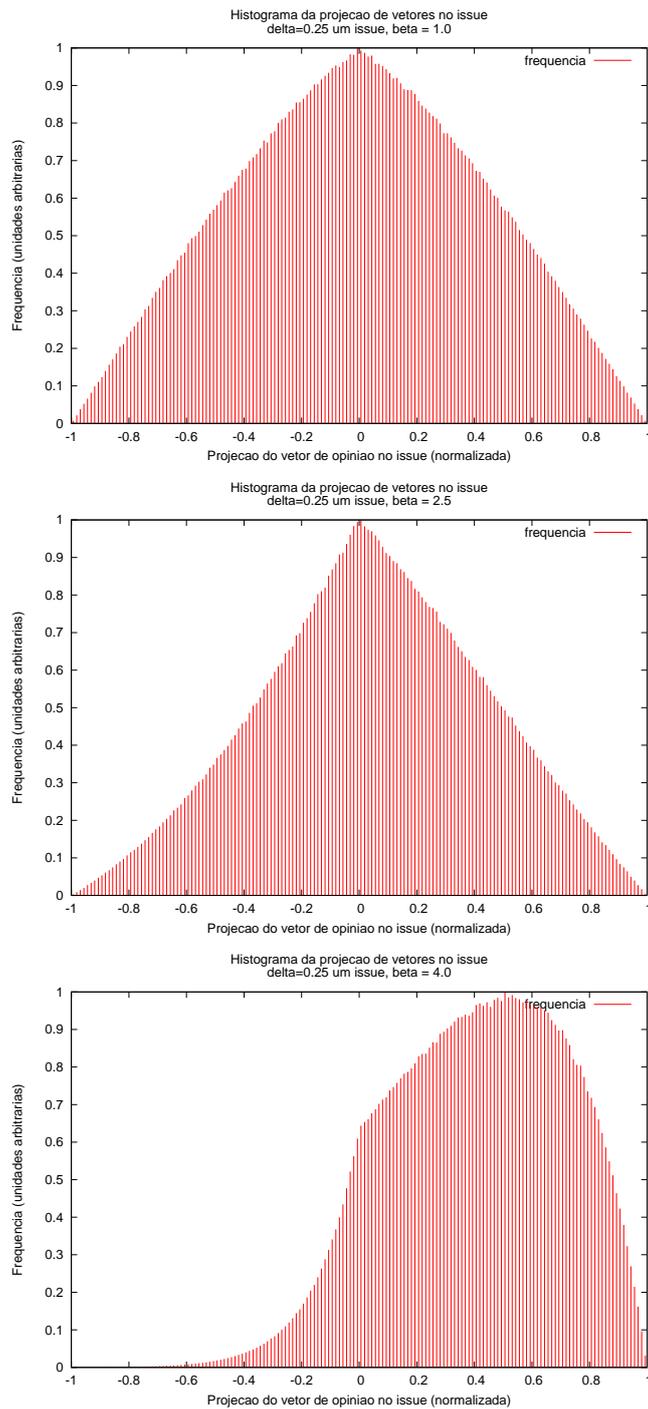


Figura 4.3: Podemos ver a quebra de simetria induzida pelo aumento da pressão de pares no sistema observando os histogramas da projeção no assunto. O comportamento do sistema se altera fundamentalmente.

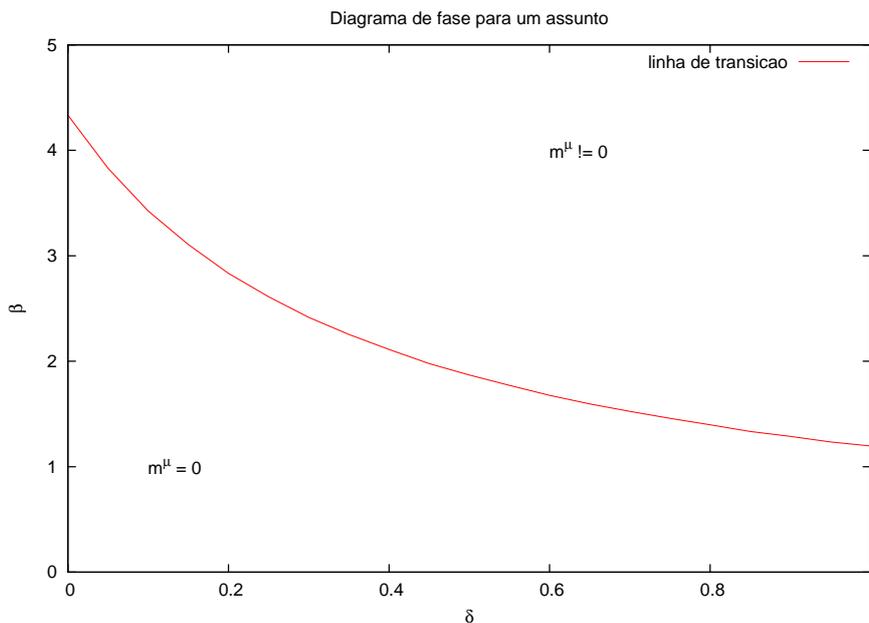


Figura 4.4: Linhas de transição para o modelo de campo médio com um assunto.

parametrizar a hipersfera como

$$x_1 = \cos(\alpha_1), \quad (4.5)$$

$$x_2 = \sin(\alpha_1)\cos(\alpha_2), \quad (4.6)$$

$$x_3 = \sin(\alpha_1)\sin(\alpha_2)\cos(\alpha_3), \quad (4.7)$$

$$x_4 = \sin(\alpha_1)\sin(\alpha_2)\sin(\alpha_3)\cos(\alpha_4), \quad (4.8)$$

$$x_5 = \sin(\alpha_1)\sin(\alpha_2)\sin(\alpha_3)\sin(\alpha_4), \quad (4.9)$$

com $\alpha_1, \alpha_2, \alpha_3 \in [0, \pi]$ e $\alpha_4 \in [0, 2\pi]$. Claramente todos estes pontos obedecem $x \cdot x = 1$, e podemos também verificar que todas as soluções de $x \cdot x = 1$ em \mathbf{R}^5 são deste tipo. O elemento de integração se torna então $\sin(\alpha_1)^3 \sin(\alpha_2)^2 \sin(\alpha_3) d\alpha_1 d\alpha_2 d\alpha_3 d\alpha_4$. As integrais se tornam então

$$\int_0^\pi d\alpha_1 \int_0^\pi d\alpha_2 \int_0^\pi d\alpha_3 \int_0^{2\pi} d\alpha_4 \sin(\alpha_1)^3 \sin(\alpha_2)^2 \sin(\alpha_3) f(x(\alpha_1, \alpha_2, \alpha_3, \alpha_4)). \quad (4.10)$$

Após escrevermos a integral como uma integração sobre um domínio retangular, podemos facilmente discretizar a integral de acordo com a regra do trapézio, e aplicar o método de Romberg,

como explicitado mais à frente na seção 6.1.2 em outro contexto.

4.1.2 Modelo Hamiltoniano

No caso do modelo hamiltoniano, não temos como calcular os parâmetros de ordem explicitamente; temos, portanto, de obter estimativas de grandezas a partir de métodos de amostragem. O método mais clássico para obter estados é o que se chama comumente de métodos de Monte-Carlo, onde geramos uma cadeia de Markov a partir de uma regra de transição entre estados e amostramos estados desta cadeia de Markov. A partir de um conjunto de amostras desta cadeia, tiramos uma média das grandezas de interesse e obtemos assim alguma informação sobre o sistema. Vamos descrever aqui outro método, baseado no método de amostragem de Wang-Landau (17). Obtendo a densidade de estados em função da energia, podemos obter estimativas do calor específico do sistema a qualquer temperatura, utilizando a relação de flutuação-dissipação, temos

$$C(\beta) = \frac{1}{\beta^2} (\langle H^2 \rangle - \langle H \rangle^2), \quad (4.11)$$

e as médias abaixo podem ser escritas como

$$\langle H(\beta) \rangle = \frac{1}{Z} \int_{E_{min}}^{E_{max}} H \Omega(H) e^{-\beta H} dH \quad (4.12)$$

$$\langle H(\beta)^2 \rangle = \frac{1}{Z} \int_{E_{min}}^{E_{max}} H^2 \Omega(H) e^{-\beta H} dH \quad (4.13)$$

$$Z(\beta) = \int_{E_{min}}^{E_{max}} \Omega(H) e^{-\beta H} dH \quad (4.14)$$

Desta forma, podemos obter, a partir da densidade de estados, um retrato de toda a termodinâmica do sistema a δ fixo. Transições de fase tipicamente se manifestam como quebras de analiticidade em algum potencial termodinâmico, sendo a ordem da transição dada pela derivada da energia livre mais baixa que apresenta uma descontinuidade na transição. Portanto, transições de primeira ordem são associadas a descontinuidades nas primeiras derivadas da energia livre (como a transição induzida por uma inversão do campo magnético num modelo de Ising, sendo a descontinuidade na magnetização). No caso de transições de segunda ordem, a descontinuidade é encontrada em derivadas segundas da energia livre, como a suscetibilidade magnética ou o calor específico. Para sistemas finitos, não há quebra de analiticidade, e não há transições fase no sentido rigoroso da palavra. No entanto, podemos observar mudanças bruscas de comportamento do sistema mesmo em sistemas finitos, e estes se manifestam como máximos na

suscetibilidade ou no calor específico. No nosso caso especificamente, determinaremos a temperatura de transição para o modelo através do máximo no valor do calor específico. Podemos simular o sistema para diversos valores de δ através do algoritmo de Wang-Landau, e obter desta forma um diagrama de fases para nosso modelo.

4.1.3 Modelo de Agentes

O modelo de agentes é o modelo que exige mais cuidado ao tentarmos determinar um diagrama de fases para o mesmo. Isto se deve ao fato de que, diferentemente dos outros dois paradigmas, o modelo foi definido a partir de uma regra dinâmica de atualização dos estados e não a partir de um funcional de energia ou uma distribuição. Devido a isso não podemos ter certeza da existência de um regime estacionário, e por conseguinte de uma distribuição de probabilidades sobre os estados acessíveis. Neste tipo de sistema freqüentemente temos estados absorventes, dos quais outros estados não são mais acessíveis. Em um modelo de agentes verificamos que a convergência para uma distribuição estável depende bastante dos valores dos parâmetros da simulação. Para um ruído muito baixo ou ausente ($\eta \approx 0$), o sistema relaxa rapidamente para uma ou duas facções, dependendo das condições iniciais. Já para valores maiores de η , temos flutuações muito grandes e, apesar de assemelharem-se aos histogramas obtidos no caso de campo médio e do modelo de agentes, flutuações grandes tornam o modelo pouco prático para este tipo de análise. Concentramo-nos, portanto, no paradigma hamiltoniano e de campo médio na seção seguinte.

4.2 Comparação com dados de psicologia quantitativa

De posse de um modelo, podemos procurar compará-lo com os dados obtidos em questionários para investigar sua capacidade descritiva. Seguindo o que foi feito em Vicente et al. (4), vamos analisar os histogramas da projeção das matrizes morais sobre a média dos vetores de assuntos, que convencionamos chamar de *Zeitgeist*. No caso de nosso modelo, investigamos o caso de campo médio para compará-lo com os dados e obter uma comparação com o que foi feito em Vicente et al. (4) para o caso do modelo hamiltoniano (ver figura 4.6). Na figura 4.5, podemos ver um paralelo entre os histogramas obtidos a partir dos questionários cedidos por J. Haidt e os histogramas obtidos através de uma busca de parâmetros no modelo de campo-médio.

No caso de estudarmos o modelo de campo médio temos acesso explícito aos vetores e podemos calcular a projeção das matrizes morais no *Zeitgeist* diretamente. No entanto, quando lidamos com os dados dos questionários, não temos acesso direto aos vetores que representam os assuntos. Ainda assim podemos verificar que as matrizes morais são todas concentradas ao redor da matriz média, assim como no caso do nosso modelo as matrizes são concentradas em volta dos vetores de assunto. Desta forma, tomamos como a média dos vetores de assunto a média das matrizes morais dos agentes conservadores, pois este grupo apresenta a menor variabilidade, obtendo com isso uma estimativa razoável para a média dos assuntos. Assim, podemos calcular a projeção de todos os vetores neste vetor, obtendo assim os histogramas abaixo.

Vemos que o modelo reproduz relativamente bem o comportamento estatístico da projeção das matrizes morais sobre a média dos vetores de assuntos. Ao compararmos os dois modelos, no entanto, vemos que, como esperado, o modelo de campo médio facilita o ordenamento do sistema, levando a distribuições mais estreitas em torno do valor máximo da projeção. O procedimento empregado para acharmos a correspondência entre o modelo e os histogramas obtidos dos dados foi uma simples busca de parâmetros sobre os histogramas. De posse dos histogramas dos dados e de histogramas para o modelo de campo médio para um conjunto de parâmetros, podemos calcular a distância quadrática entre os histogramas, como feito na seção 4.17, e encontrar para cada afiliação política o valor de δ e β que minimizem esta distância entre os dois histogramas. Desta forma obtivemos a figura 4.6.

A correspondência entre os dados psicológicos e os dados de nosso modelo são bastante compatíveis, permitindo-nos afirmar que a estrutura estatística das matrizes morais em subpopulações divididas por afiliação política é bem reproduzida pelo modelo proposto. Parece que o ingrediente essencial para este sistema é a minimização de um custo social entre vizinhos sociais, fenômeno presente em sociedades reais. Para verificar a fundo esta hipótese precisaríamos, no entanto, verificar tanto do ponto de vista teórico quanto do ponto de vista experimental a validade destas afirmações. Do ponto de vista teórico, teríamos de verificar se a estrutura encontrada para outros modelos similares e em outras situações mantêm os mesmos traços fundamentais. Do lado da psicologia, seria necessário verificar a hipótese de que seres humanos minimizam um custo psicológico ao interagir em sociedades (ainda que apenas em algumas situações).

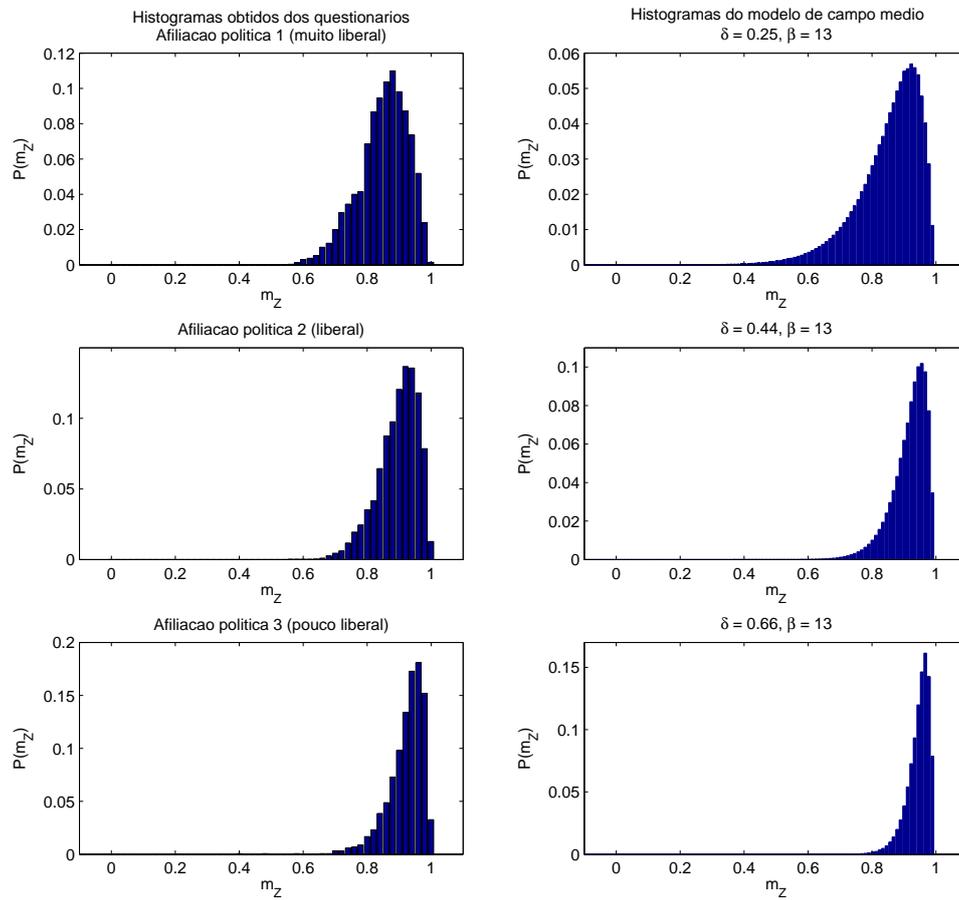


Figura 4.5: Comparação entre os histogramas do modelo de campo médio e os histogramas obtidos com os dados de Haidt. Os histogramas à esquerda correspondem às afiliações políticas 1, 2 e 3 (muito liberal, liberal e pouco liberal, respectivamente).

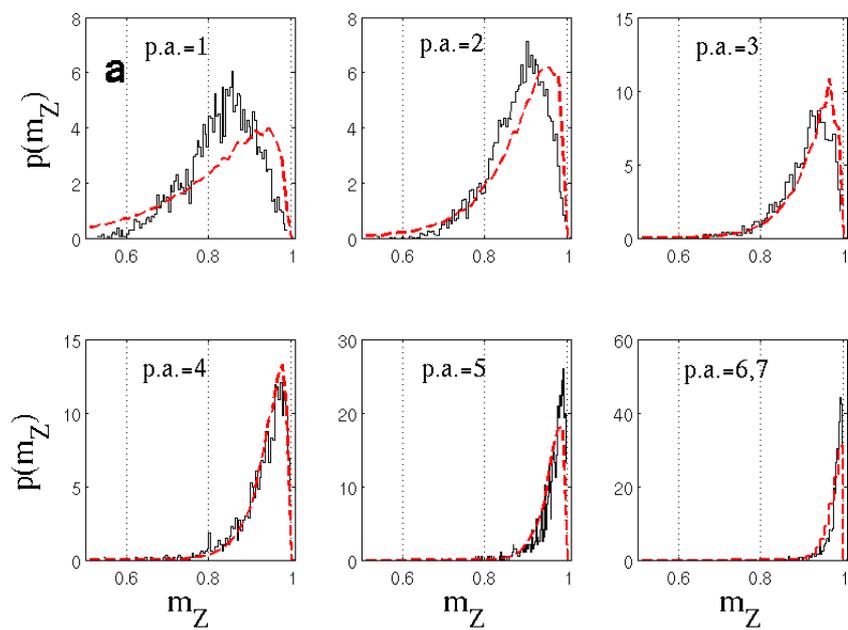


Figura 4.6: Comparação entre os histogramas da projeção das matrizes morais nos dados e os histogramas obtidos por amostragem com o algoritmo de Metropolis do modelo hamiltoniano.

4.3 Diagramas de Fase

No trabalho de Renato Vicente et al. (8), foram estudadas situações nas quais agentes procurando o consenso com seus pares numa escala microscópica podem levar à formação de facções discordantes numa escala macroscópica. Podemos observar a dependência deste tipo de comportamento analisando os histogramas obtidos para tempos grandes nestas simulações. Essencialmente, temos agentes dispostos sobre uma rede complexa aprendendo de seus vizinhos na rede de acordo com a regra descrita para o modelo de agentes. Na figura 4.3 podemos ver claramente a formação de facções opostas em relação a um assunto dado. No caso mostrado estamos considerando assuntos ortogonais com uma dinâmica de aprendizado on-line. Este comportamento de formação de facções é bastante robusto, mantendo-se mesmo para tempos de simulação muito longos. No entanto, é difícil falar de estado de equilíbrio no caso de um modelo não-hamiltoniano. Variando os parâmetros do modelo podemos obter outros comportamentos, como uma sociedade completamente ordenada em torno de um assunto ou uma sociedade sem nenhuma polarização como podemos ver na figura 4.3.

Motivado pela discussão do modelo proposto em Vicente et al. (8), procuramos então obter uma caracterização análoga para os modelos de campo-médio e hamiltoniano, caracterizando a dependência das fases nos parâmetros de aprendizado, no nível de ruído (temperatura) e na distribuição dos assuntos. Podemos claramente notar a transição entre desordem e polarização como mostrado nas figuras 4.4 para o modelo de campo médio, por outro lado, a formação de facções não é tão fácil de obter nestes modelos. No entanto, a transição entre desordem e polarização mostra-se bastante confiável e pôde ser estudada extensivamente. Para determinar a fase no modelo de campo médio simplesmente determinamos os parâmetros de ordem com uma certa precisão e então calculamos o valor da polarização média do estado, $\sum_{\mu} m^{\mu}$. Estabelecemos uma tolerância para esta grandeza, nas simulações mostradas tomamos uma tolerância de 10^{-2} .

4.3.1 Assuntos Ortogonais

Vimos acima nas figuras 4.1 e 4.2 o comportamento do sistema para um dado valor de δ ao longo de um contínuo de valores de β . Na figura 4.8 podemos ver as linhas de transição para 1 e 2 assuntos ortogonais com a linha sobre a qual foram calculados os parâmetros de ordem das figuras anteriores. Observamos que quanto menor o valor de δ mais alto é o valor de β para o

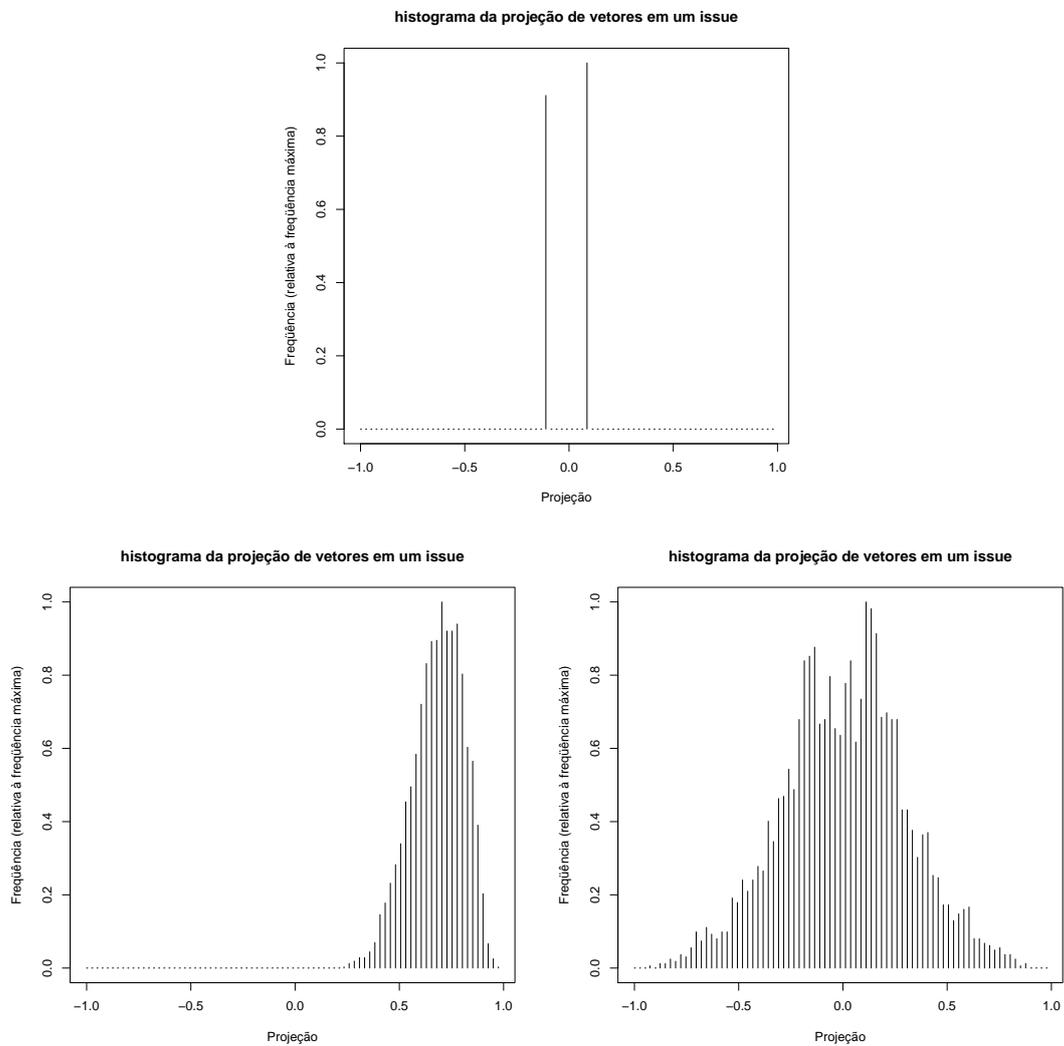


Figura 4.7: Histograma da projeção dos vetores morais em um dos assuntos, mostrando os diferentes tipos de comportamento encontrados para o modelo de agetnes. A primeira figura mostra a formação de facções ($\delta = 0.0, \eta = 0.2, 2 \times 10^6$ passos de MC). A segunda figura mostra a formação de um consenso com relação a um assunto ($\delta = 0.5, \eta = 0.0, 3 \times 10^6$ passos de MC). Na terceira figura vemos a ausência de polarização devido ao alto ruído na comunicação ($\delta = 0.5, \eta = 0.4, 3 \times 10^6$ passos de MC).

qual temos uma transição para a fase ordenada do sistema.

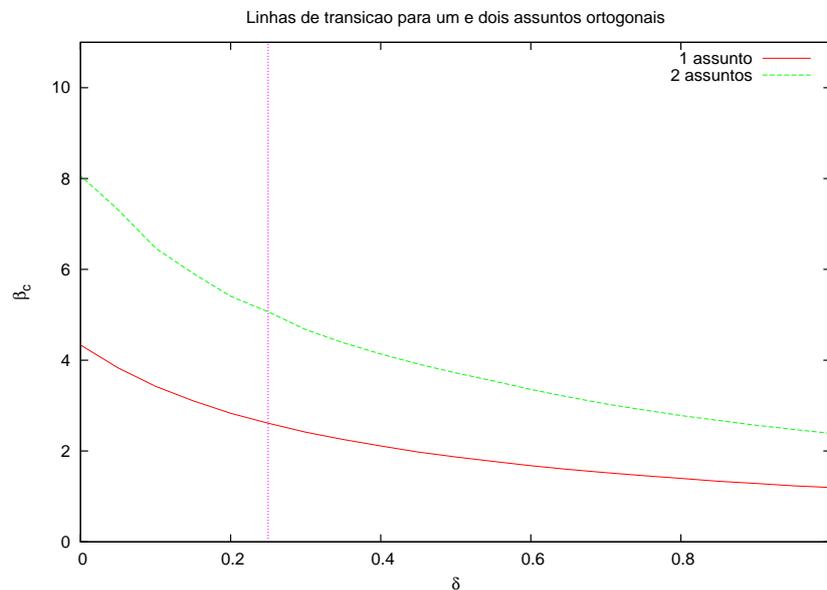


Figura 4.8: Vemos o diagrama fase para 1 assunto e 2 assuntos ortogonais. A linha vertical representa a linha sobre a qual obtivemos os parâmetros de ordem das figuras 4.1 e 4.2.

Outro fato interessante é o fato de que um aumento no número de assuntos eleva também o valor de β crítico, como podemos ver na figura 4.9. No entanto, não podemos concluir uma mudança de fato no comportamento do sistema devido ao fator $\frac{1}{P}$ presente no hamiltoniano. Este termo foi incluído no hamiltoniano para que tenhamos uma correspondência entre uma versão do sistema com três assuntos muito próximos e uma versão com apenas um assunto. No entanto podemos verificar este efeito verificando como seria o sistema se não tivéssemos um fator $\frac{1}{P}$ em frente ao hamiltoniano. Podemos reescrever o hamiltoniano regraduando o valor de β por um fator de $\frac{1}{P}$. O expoente do fator de Boltzmann se tornaria então

$$-\frac{\beta}{P}H = -\frac{\beta P}{P}H = -\beta'H.$$

Podemos ver esta regraduação na figura 4.10, e verificamos que apesar de termos um efeito relacionado ao número de assuntos discutidos, o efeito na linha de transição é aproximadamente multiplicativo, especialmente para valores de δ próximos de 1.

Podemos ver abaixo o diagrama de fases gerado para o modelo hamiltoniano através de

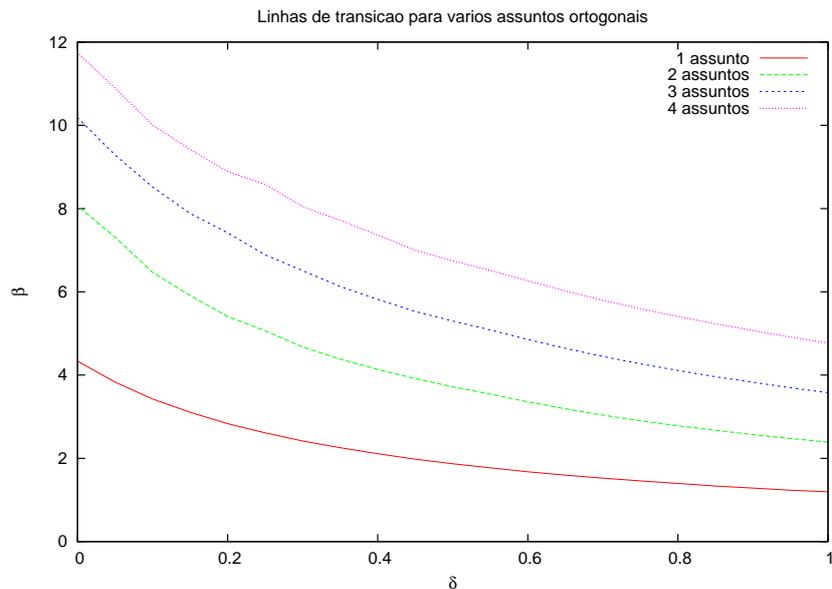


Figura 4.9: Vemos o diagrama fase para 1 assunto, 2 e 3 assuntos ortogonais. Podemos ver o efeito do número crescente de assuntos discutidos, elevando o valor crítico de β .

amostragem de Wang-Landau. Na figura 4.11 podemos ver as curvas de calor específico para alguns valores de δ . O pico pronunciado no calor específico indica uma transição de fase de primeira ordem, especialmente visível para valores altos de δ . Para valores mais baixos de δ , podemos ver um segundo pico no calor específico, no entanto, não podemos identificá-lo com uma transição de fase com certeza. O método de amostragem de Wang-Landau por vezes não captura toda a geometria do espaço de fase em temperaturas mais baixas, de forma que desprezamos o segundo pico no calor específico. A partir destes dados podemos calcular um diagrama de fases, assim como fizemos para o caso do modelo de campo médio. Na figura 4.13 podemos ver este diagrama de fases. As barras de erro na temperatura de transição são calculadas a partir de diversas realizações da amostragem para o sistema dado.

Comparando o diagrama de fases com o diagrama de campo médio, vemos uma característica comum de métodos de campo médio que é o fato de eles superestimarem a fase ordenada para o sistema. Como é característico, vemos nas figuras 4.14 e 4.15 a linha de transição para o modelo de campo médio em valores menores de β que o de Wang-Landau.

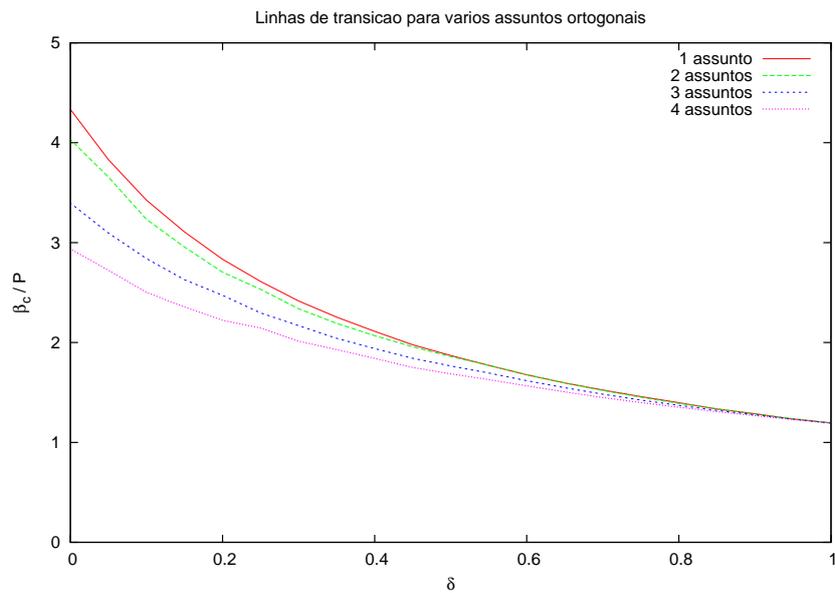


Figura 4.10: Regraduando a linha de transição com o número de assuntos discutidos, vemos a influência deste na geometria das linhas de transição. A influência de assuntos ortogonais é essencialmente a de multiplicar o valor de β crítico pelo número de assuntos, ainda que para valores de δ próximo de zero, o número crescente de assuntos discutidos parece resultar em uma diminuição de β crítico.

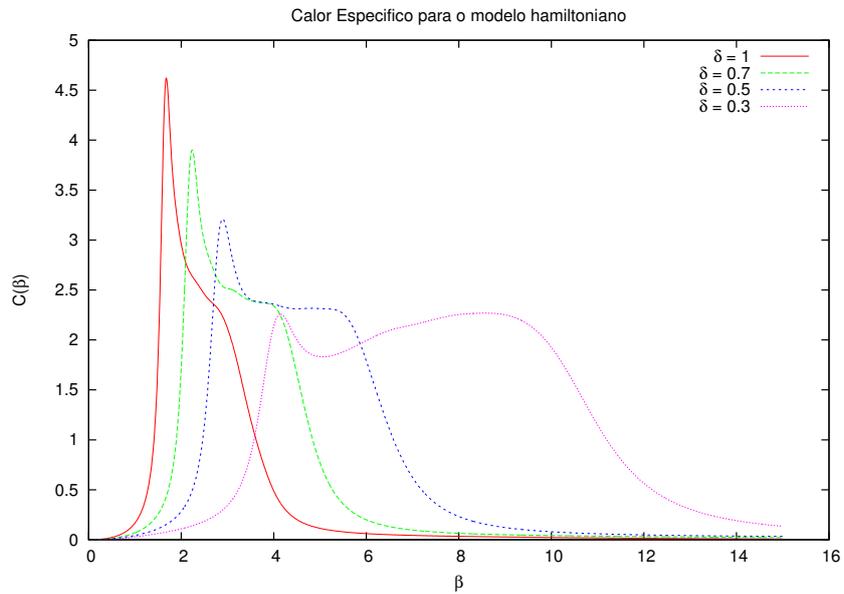


Figura 4.11: Calor específico para alguns valores de δ obtido através de amostragem de Wang-Landau.

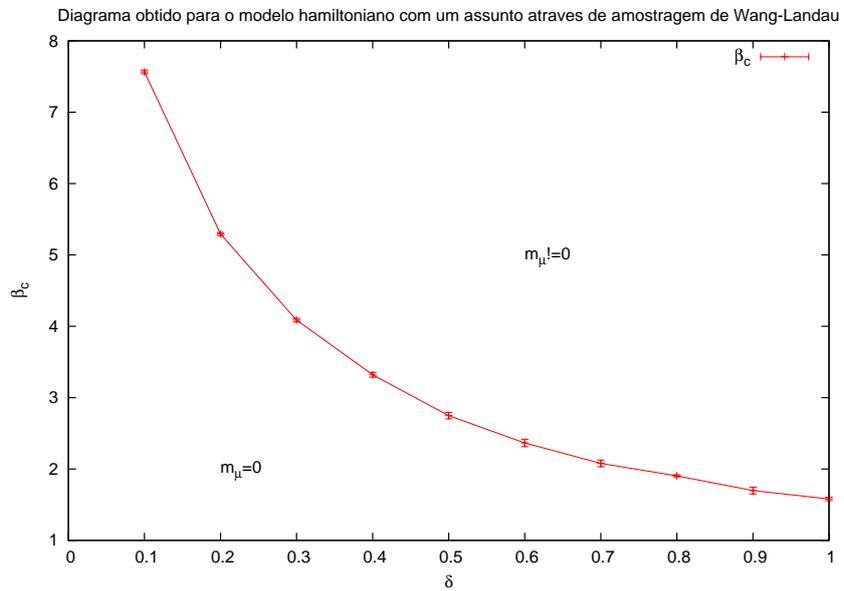


Figura 4.12: Diagrama de fases obtido através do algoritmo de Wang-Landau.

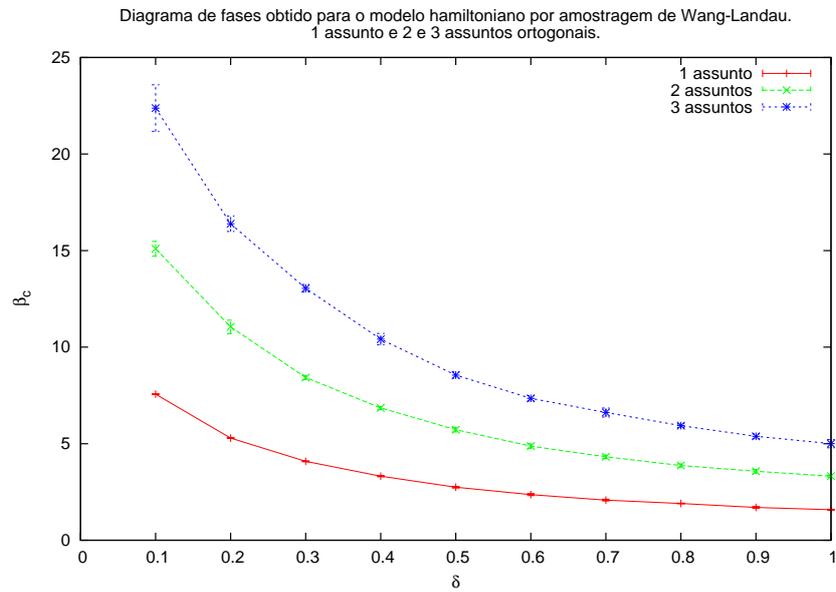


Figura 4.13: Diagrama de fases obtido através do algoritmo de Wang-Landau para um assunto e dois e três assuntos ortogonais.

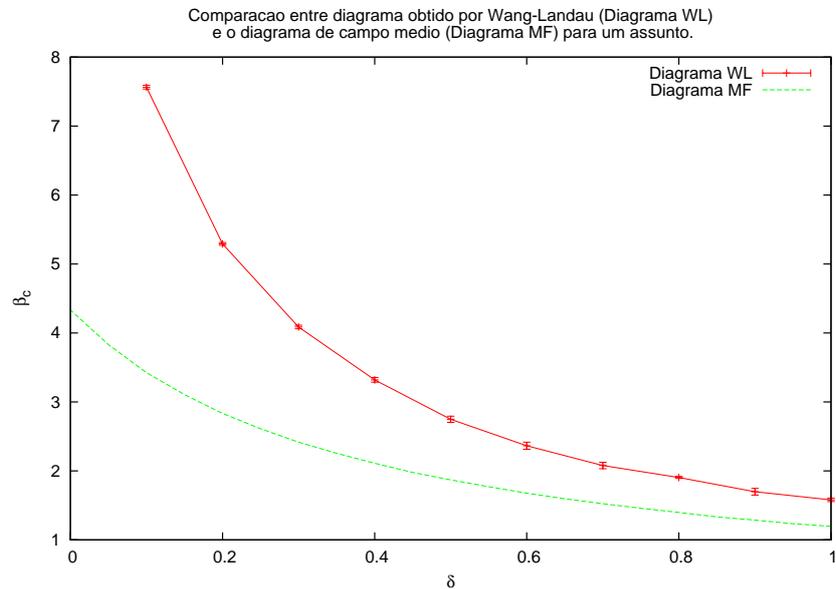


Figura 4.14: Diagrama de fases para um assunto em campo médio e para o modelo hamiltoniano.

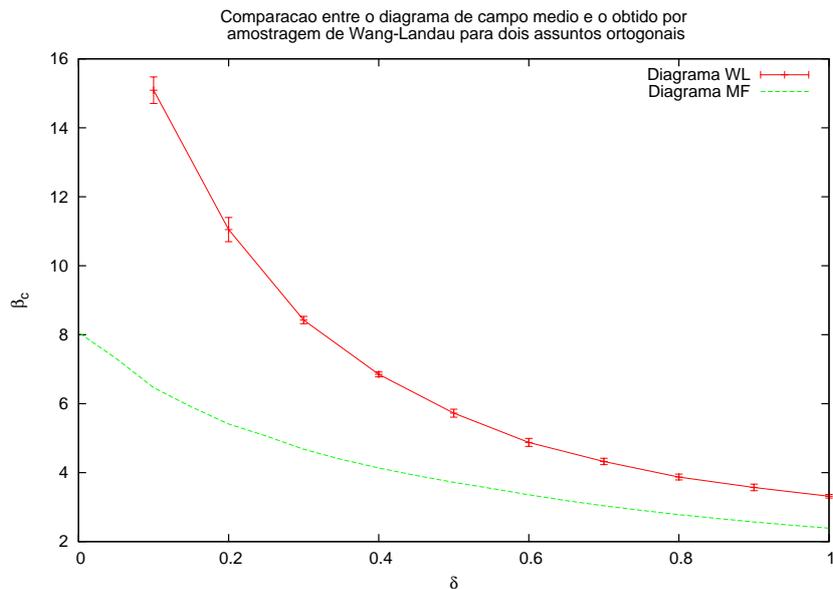


Figura 4.15: Diagrama de fases para dois assuntos ortogonais para o modelo de campo médio e o modelo hamiltoniano.

4.4 Perturbações

Outro aspecto de interesse no nosso estudo é a capacidade de adaptação dos sistemas para diferentes valores dos parâmetros. A adaptabilidade a mudanças repentinas nos assuntos discutidos pode ter tido algum valor evolutivo na determinação deste tipo de comportamento em indivíduos humanos. Passamos então a estudar o relaxamento de um sistema em equilíbrio após alterarmos os vetores associados aos *issues* de alguma forma. A forma mais simples de perturbar estes vetores é aplicar uma rotação por um ângulo ζ e verificar a evolução do sistema após esta alteração. Gostaríamos de verificar a discrepância entre o estado anterior à perturbação e os estados gerados após a perturbação. Uma forma de representar um estado de um sistema é obter histogramas de grandezas de interesse. Podemos então usar critérios de comparação de distribuições, como por exemplo a divergência de Kullback-Leibler.

No caso da aproximação do campo médio podemos deduzir uma expressão para a divergência de Kullback-Leibler entre dois estados que depende somente dos parâmetros de ordem do estado dado. No entanto, como é comum em formulações de mecânica estatística não há um tempo explicitamente definido no modelo de campo-médio. A formulação de modelos de spins

em mecânica estatística também era desprovida de uma interpretação dinâmica até o trabalho de Glauber sobre o modelo de Ising (27). A solução usada muitas vezes no caso de termos um conjunto de equações autoconsistentes é considerar as relações de recorrência geradas por estas equações como um tipo de dinâmica. No caso do modelo hamiltoniano podemos considerar a dinâmica de amostragem de Monte Carlo como uma dinâmica do sistema como é frequentemente feito neste tipo de estudo.

4.4.1 Distância Euclidiana entre Histogramas

Uma forma intuitiva de obter alguma informação sobre um sistema é obter histogramas das grandezas de interesse. No nosso caso, os parâmetros de ordem determinados na análise de campo médio são as projeções médias ao longo de cada um dos assuntos e o valor absoluto médio destas projeções. Para determinar portanto o relaxamento do sistema após uma perturbação na orientação dos assuntos podemos determinar o histograma no equilíbrio, então perturbar o sistema e tomar histogramas em passos subseqüentes computando a distância euclidiana entre ambos os histogramas. Um problema que se coloca ao tomarmos este *ansatz* é a comparabilidade entre o tempo de relaxamento e o tempo de amostragem dos histogramas. Temos que garantir que o tempo que usamos para amostrar os histogramas seja significativamente menor que o tempo característico de relaxamento do sistema. Não temos nenhuma forma de descobrir isto *a priori*, porém podemos sempre tomar sistemas de dimensões maiores de forma que os histogramas possam ser amostrados em tempos mais curtos. Nosso objetivo é obter uma medida da rapidez do relaxamento do sistema em função dos parâmetros deste.

Esta metodologia se mostra particularmente interessante em simulações de Monte-Carlo, essencialmente desenvolvidas para amostragem do espaço de estados. Procuramos analisar o relaxamento do sistema através de simulações usando o algoritmo de Metropolis. Primeiramente devemos termalizar o sistema, para ter certeza que estamos amostrando da distribuição de equilíbrio. Após um certo número de passos de MC, efetuamos uma rotação nos assuntos e passamos a registrar histogramas a intervalos regulares, calculando a distância euclidiana entre estes e o histograma do estado de equilíbrio. Para facilitar a análise restringimo-nos ao caso de um único assunto. Na figura abaixo podemos ver o relaxamento do sistema com apenas um assunto, com um decaimento característico de uma função exponencial. No caso mostrado abaixo, obtivemos um decaimento exponencial $f(t) = ae^{-bt}$, com valores $a = 0.0589 \pm 0.0013$

e $b = 0.0220 \pm 0.0007$ através de um simples método de mínimos quadrados.

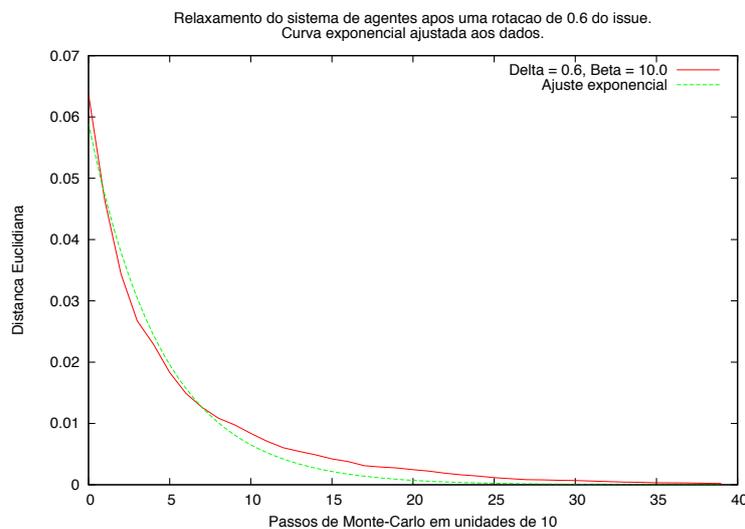


Figura 4.16: Acima temos o decaimento da distância para o histograma no equilíbrio em função do tempo de simulação.

Podemos então seguir a prescrição acima e obter o comportamento para uma gama de parâmetros de interesse. Podemos ver abaixo um comparativo para alguns valores de δ para $\beta = 10$. O gráfico em escala logarítmica evidencia a natureza exponencial do decaimento. Vemos que para δ menor o decaimento é mais rápido e quando chegamos perto da transição o próprio valor inicial da distância diminui. Isso pode ser compreendido observando-se que na fase totalmente desordenada a perturbação não deveria ter efeito nenhum, uma vez que a distribuição dos vetores de opinião é isotrópica. A distribuição na fase desordenada não é necessariamente isotrópica, muito pelo contrário, esperamos um alinhamento no eixo do assunto em discussão, mas em relação a este assunto a distribuição é simétrica, logo perturbações devem ter um efeito menor.

Podemos então usar os dados obtidos nas simulações acima para obter uma medida da velocidade de adaptação para o sistema em função de δ . Para tal podemos calcular o tempo característico usando uma regressão com a família de funções $f(x; a, \tau) = ae^{-\frac{x}{\tau}}$ e determinar os valores de a e τ que minimizam alguma medida de erro. Na figura 4.18 podemos ver os valores de τ para os valores de δ e β indicados acima. Vemos claramente uma adaptação mais rápida no ponto de transição de fases, um resultado bastante surpreendente. No entanto, interpretando este fato à luz dos resultados da teoria de fenômenos críticos podemos interpretar isso de uma

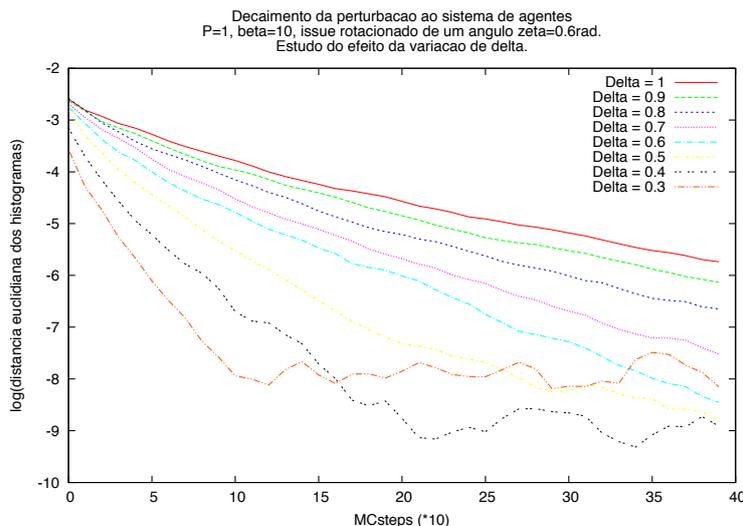


Figura 4.17: O decaimento em escala logarítmica explicita a variação da velocidade do decaimento em função de δ .

forma simples.

4.4.2 Divergência de Kullback-Leibler

No caso do modelo hamiltoniano considerado acima, não temos acesso à distribuição de probabilidades, portanto restringimo-nos à amostragem de histogramas. No entanto, poderíamos ter escolhido montar um histograma da frequência com a qual surge um determinado vetor, e calcular a distância euclidiana entre estes histogramas. Claramente, se temos vetores unitários em 5 dimensões, teríamos que amostrar um histograma em quatro dimensões, procedimento muito mais trabalhoso e ineficiente do que amostrar um histograma unidimensional. Por isso, no caso do modelo hamiltoniano, a opção escolhida foi a de amostrar somente a projeção dos vetores sobre o *issue*. Na aproximação de campo-médio, no entanto, temos a expressão exata da distribuição dos vetores, podendo calcular expressões exatas. Poderíamos calcular a distribuição da projeção sobre os *issues* diretamente como funções dos parâmetros de ordem. Para um *issue* teríamos

$$P(J \cdot S = m) = \int d\mu(J) \delta(J \cdot S - m) P(J | \{m^\mu, r^\mu\}, S) \quad (4.15)$$

$$= \frac{1}{Z} \int d\mu(J) d\zeta \exp(i\zeta(J \cdot S - m) - \beta(am^\mu J \cdot S - br^\mu |J \cdot S|)). \quad (4.16)$$

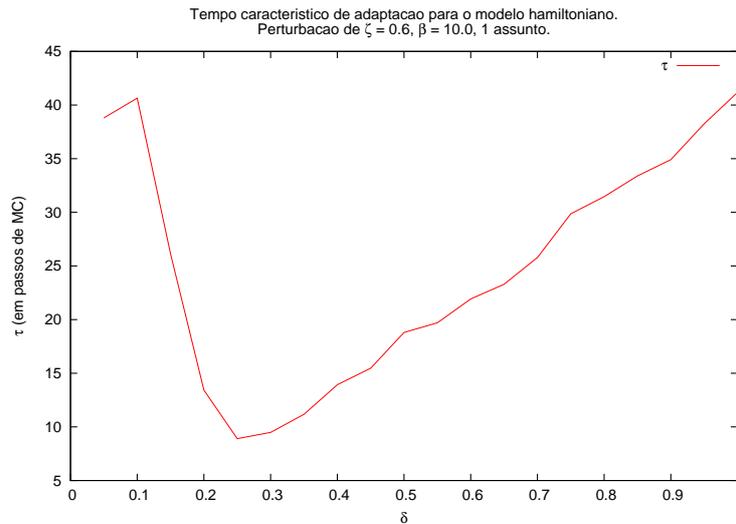


Figura 4.18: Vemos o efeito de diferentes estratégias cognitivas sobre a adaptabilidade do sistema.

A expressão acima é um tanto incômoda apesar de parecer simples, devido ao fato de estarmos integrando sobre a reta na variável da transformada de Fourier, mas sobre a hipersfera unitária em J , o que complica a resolução desta expressão já para um *issue*, para não dizer de mais *issues*. No entanto, nesta aproximação, estamos de posse da expressão analítica da distribuição de probabilidades, de forma que podemos calcular analiticamente a distância euclidiana ou a divergência KL. Vamos focar na divergência KL nesta seção.

A divergência de Kullback-Leibler para dois conjuntos de parâmetros de ordem diferente pode ser escrita como

$$D_{KL}(\{m^\mu, r^\mu\} | \{m'^\mu, r'^\mu\}) = \int d\mu(J) P(J) \log \left(\frac{P(J)}{P'(J)} \right), \quad (4.17)$$

onde $P(J)$ denota a distribuição em relação aos parâmetros $\{m^\mu, r^\mu\}$ e $P'(J)$ denota a distribuição em relação aos parâmetros $\{m'^\mu, r'^\mu\}$. Assumindo que ambas distribuições se referem ao mesmo

conjunto de *issues* teremos então

$$\begin{aligned}
D_{KL}(\{m^\mu, r^\mu\}|\{m'^\mu, r'^\mu\}) &= \\
\int d\mu(J) \frac{e^{-\frac{\beta N}{P}(\sum_\mu a m^\mu h^\mu - b r^\mu |h^\mu|)}}{Z} &\left(\log\left(\frac{Z'}{Z}\right) - \frac{\beta N}{P} \sum_\mu (a(m^\mu - m'^\mu)h^\mu - b(r^\mu - r'^\mu)|h^\mu|) \right) \\
&= \log\left(\frac{Z'}{Z}\right) - \frac{\beta N}{P} \left(\sum_\mu (a(m^\mu - m'^\mu) \langle h^\mu \rangle - b(r^\mu - r'^\mu) \langle |h^\mu| \rangle) \right) \\
&= \log\left(\frac{Z'}{Z}\right) - \frac{\beta N}{P} \left(\sum_\mu (a(m^\mu - m'^\mu)m^\mu - b(r^\mu - r'^\mu)r^\mu) \right). \tag{4.18}
\end{aligned}$$

A partir desta expressão podemos calcular a distância entre distribuições definidas por parâmetros de ordem perturbados e acompanhar a evolução da divergência KL usando a recorrência como dinâmica. Primeiramente, dado β e δ encontramos os parâmetro de ordem no equilíbrio usando as equações autoconsistentes. A partir deste processo definimos:

$$\begin{aligned}
m_0^\mu &= \int d\mu(J) J \cdot S'^\mu P(J), \\
r_0^\mu &= \int d\mu(J) |J \cdot S'^\mu| P(J).
\end{aligned} \tag{4.19}$$

Estes são os parâmetros de ordem se os *issues* fossem alterados em uma escala de tempo muito mais rápida que a escala de tempo característica da dinâmica do sistema. A partir daí consideramos o relaxamento do sistema de volta para o equilíbrio de acordo com

$$m_i^\mu = \int d\mu(J) J \cdot S^\mu P(J|\{m_{i-1}^\mu, r_{i-1}^\mu\}) \tag{4.20}$$

$$r_i^\mu = \int d\mu(J) |J \cdot S^\mu| P(J|\{m_{i-1}^\mu, r_{i-1}^\mu\}), \tag{4.21}$$

e calculamos a divergência KL entre a distribuição de equilíbrio e a distribuição perturbada, obtendo

$$D_{KL}(\{m_i^\mu, r_i^\mu\}|\{m_{eq}^\mu, r_{eq}^\mu\}) = \log\left(\frac{Z_{eq}}{Z_i}\right) - \frac{\beta N}{P} \left(\sum_\mu (a(m_i^\mu - m_{eq}^\mu)m_i^\mu - b(r_i^\mu - r_{eq}^\mu)r_i^\mu) \right). \tag{4.22}$$

Abaixo podemos ver um exemplo para $\delta = 0.24$ e $\beta = 3.1$.

É importante notar um abuso que cometemos no procedimento acima. O procedimento

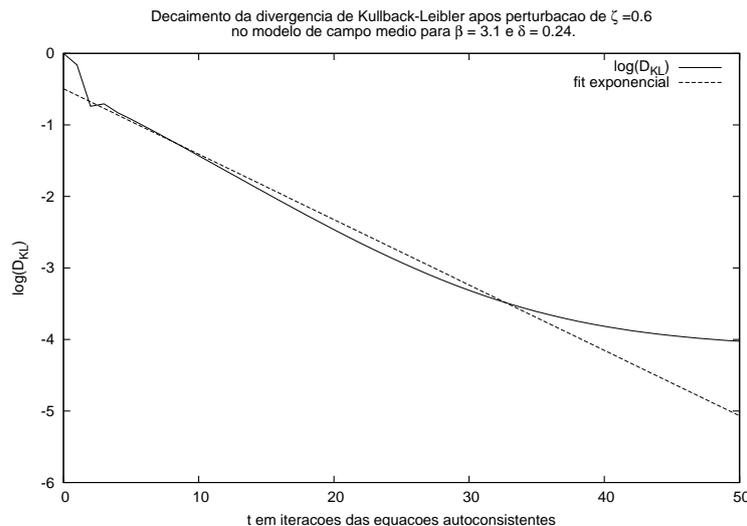


Figura 4.19: Podemos ver o decaimento exponencial da divergência KL, até notarmos uma saturação em 10^{-4} a partir da qual a aproximação exponencial não é mais tão precisa.

variacional através do qual determinamos a forma da distribuição de campo médio assume que estamos no equilíbrio. Na dedução acima extrapolamos esta distribuição para o caso em que os parâmetros de ordem não estão em seu valor de equilíbrio. Se fôssemos considerar rigorosamente, não temos como determinar a distribuição para J 's fora do equilíbrio. No entanto, o tratamento acima define uma família de distribuições que podemos usar de qualquer forma. Apesar de não podermos garantir que a distribuição dada é a distribuição de equilíbrio para algum conjunto de parâmetros, podemos tratar a dinâmica do sistema nesta aproximação.

Os resultados usando divergência KL na aproximação de campo médio e utilizando a distância euclidiana entre histogramas nas simulações de metropolis nos dão resultados compatíveis, o que permite inferir que o comportamento do sistema em relação a perturbações é relativamente robusto com relação a diferentes modelagens.

5 *Conclusões*

O objetivo que nos colocamos neste projeto de pesquisa foi caracterizar a influência de diferentes estratégias cognitivas em um modelo de aprendizado cultural. Isso foi feito utilizando-se três paradigmas, modelagem de agentes, simulações de montecarlo e uma aproximação de campo médio. Uma caracterização ampla foi obtida de diversas grandezas do sistema, especialmente da dependência da linha de transição entre as fases ordenada e desordenada da estratégia cognitiva e do ambiente de aprendizado (a distribuição de assuntos). Pudemos também estudar em dois dos paradigmas como o sistema se adapta a mudanças bruscas no ambiente de aprendizado (que convencionamos chamar de *Zeitgeist*). Nesta seção vamos sumarizar os resultados obtidos e discutir a relevância e impacto destes.

Vamos abaixo apresentar uma síntese dos resultados obtidos neste trabalho e interpretá-los à luz do objeto que pretendemos modelar.

5.1 Discussão dos resultados

Abaixo vou rapidamente comentar os resultados obtidos ao longo deste trabalho.

5.1.1 Tratamento do modelo.

Obtivemos uma aproximação de campo médio para o modelo estudado em (4) e a partir de equações autoconsistentes calculamos diagramas de fase, valores para os parâmetros de ordem e deduzimos uma expressão analítica para a divergência de Kullback-Leibler. Apesar de o método empregado ser bastante comum e ter sido bastante estudado no campo de teoria de campos, vale notar que ao definir a distribuição aproximadora não fizemos nenhuma hipótese sobre a forma funcional além de que esta seja fatorizável sobre os sítios. Apesar de que, como apresentado, em casos mais simples isso não faça grande diferença, é importante notar que a possibilidade de deixar o modelo definir a forma da distribuição é bastante desejável.

No caso do modelo hamiltoniano, apesar de não termos desenvolvido resultados analíticos novos, o estudo do diagrama de fase através de amostragem de Wang-Landau foi um tratamento novo para o modelo. Além disso o tratamento de perturbações do modelo usando a distância euclidiana entre histogramas também apresenta um resultado novo, mostrando dados interessantes sobre a adaptabilidade do sistema em questão.

5.1.2 Transições de Fase

Pudemos verificar a existência de duas fases para o nosso modelo de sociedade, uma em que apesar da tentativa dos agentes de individualmente conformar-se com seus vizinhos sociais, não observamos a emergência de um consenso, e outra em que a procura por conformidade com os vizinhos leva a um consenso entre os agentes (ainda que apenas parcial).

No caso de assuntos ortogonais, verificamos que a inclusão de novos assuntos discutidos no modelo tem um efeito aproximadamente aditivo no valor de β_c , dificultando a formação de um consenso no sistema. Para valores mais baixos de δ a aditividade da inclusão de assuntos novos se quebra e temos subaditividade. Os resultados são ao menos qualitativamente correspondentes no caso da aproximação de campo médio e no modelo hamiltoniano, mesmo que a natureza estatística dos resultados não permita verificar com tanta precisão esta correspondência.

5.1.3 Perturbações

A adaptabilidade do sistema estudado foi estudado em dois paradigmas, no modelo hamiltoniano e no modelo de campo médio. No modelo hamiltoniano, pudemos verificar um decaimento exponencial da distância entre os histogramas de equilíbrio e os histogramas amostrados após a perturbação. Estudando a velocidade de decaimento desta distância em função do parâmetro δ pudemos verificar que a adaptabilidade deste sistema tem um pico na região de transição entre as fases ordenada e desordenada. A rápida adaptação de uma sociedade a mudanças bruscas no ambiente de discussão é uma característica desejável para sociedades, uma vez que mesmo que diversidade tenha um papel importante, a emergência de um consenso que possibilita um diálogo é fundamental para o convívio humano. No entanto, observações sobre dados obtidos de Jonathan Haidt (4) parecem colocar grupos humanos dentro da fase ordenada, longe da linha de transição. Podemos atribuir isso à necessidade de, além de adaptabilidade, da existência de um capital social, manifesto na existência de algum consenso na sociedade. Estes dois elementos poderiam empurrar um estado de equilíbrio evolutivo para a fase ordenada do diagrama.

Para o modelo de campo médio apesar de podermos obter uma expressão exata para a divergência de Kullback-Leibler entre dois estados diferentes, a dinâmica inerente nas relações de recorrência parece não nos oferecer a mesma informação contida na dinâmica de Monte Carlo do modelo hamiltoniano. No entanto pudemos obter um decaimento grosseiramente exponencial

e obter estimativas para as velocidades de adaptação ainda que menos confiáveis. A estrutura encontrada no modelo hamiltoniano parece ser conservada.

5.2 Interpretação

O modelo apresentado e estudado neste trabalho se propõe a modelar o comportamento de sociedades com relação ao julgamento moral. Como indicado no trabalho de Caticha e Vicente (4), este modelo se mostra apropriado para reproduzir estatísticas obtidas a partir de questionários de psicologia quantitativa como função da orientação política dos questionados. Segundo esta analogia podemos atribuir a diferença de uso de fundações morais de conservadores e liberais a uma diferença nas estratégias cognitivas entre os dois. Especificamente podemos associar estratégias cognitivas de δ baixo a comportamento liberal e estratégias de δ alto a comportamento conservador. Segundo nosso estudo, podemos além disso identificar alguns outros padrões no comportamento do modelo.

Primeiramente podemos identificar que a discussão de vários assuntos dificulta a formação de qualquer consenso dentro da sociedade. Verificamos que a modulação de um parâmetro que identificamos com a temperatura inversa (β) pode reverter o sistema de um estado de dissenso para consenso. Este parâmetro controla o valor esperado do custo de discordância entre pares, de forma que convencionamos chamá-lo de pressão de pares ou *peer pressure*. Concluimos também que a mudança abrupta no ambiente de aprendizado tem um efeito mais transitório para estratégias cognitivas que identificamos com comportamento liberal do que em estratégias identificadas com comportamento conservador.

Parte II

Projetos Paralelos

Ao longo do período ocupado por este projeto de mestrado dediquei-me também a projetos de pesquisa paralelos. Inicialmente o projeto de mestrado previa a continuação de meu projeto de iniciação científica. Neste, dediquei-me a procurar um algoritmo de reconhecimento invariante de padrões baseado em integrais de grupo. Abaixo vou descrever os resultados obtidos brevemente e dar uma idéia geral do progresso do projeto.

Enfatizo no entanto, que trata-se de trabalhos em progresso que não foram levados a cabo como projetos de pesquisa, de forma que sua apresentação manterá um tom menos profundo que o restante do trabalho.

6 *Reconhecimento Invariante de Padrões*

6.1 Reconhecimento Invariante de Padrões

Nesta seção discutiremos uma variante do método descrito em (28) para o reconhecimento invariante de imagens. O problema de reconhecimento invariante de imagens consiste em determinar uma forma eficiente de classificar imagens levando em conta simetrias contínuas predefinidas pelo usuário. Apesar de estas simetrias poderem ser de diversas naturezas, a ferramenta matemática preferida para tratar simetrias é a teoria de grupos. O desenvolvimento de reconhecimento invariante de imagens tem uma vasta gama de aplicações em diversas áreas. Citamos como exemplos o caso de reconhecimento de microorganismos para propósitos diagnósticos e o reconhecimento de falhas em produção em escala de peças.

Primeiramente mencionaremos algumas definições básicas da teoria de grupos. Um grupo é um conjunto G de objetos $g \in G$ e uma operação $\cdot : G \times G \rightarrow G$ satisfazendo os seguintes axiomas:

$$\forall g, h \in G, \quad g \cdot h \in G, \quad (6.1)$$

$$\forall a, b, c \in G, \quad (a \cdot b) \cdot c = a \cdot (b \cdot c), \quad (6.2)$$

$$\exists \mathbf{1} \in G \quad | \quad \forall g \in G, \mathbf{1} \cdot g = g \cdot \mathbf{1} = g, \quad (6.3)$$

$$\forall g \in G, \exists g^{-1} \quad | \quad g g^{-1} = \mathbf{1}. \quad (6.4)$$

Estes quatro axiomas são comumente chamados de axiomas de grupo. Apesar de estes quatro axiomas já conterem muita informação, o nosso interesse está em um tipo mais específico de grupos, sendo estes os grupos de transformações de simetria. Consideraremos portanto grupos de transformações de um dado espaço X , por exemplo o espaço vetorial Y de vetores $\vec{y} \in Y$. Os elementos do grupo são então definidos como transformações $g : X \rightarrow X$. Este grupo define um conjunto de classes de equivalência no espaço X . Dizemos que dois elementos x, y de X têm uma relação de similaridade pelo grupo G se existe um elemento g tal que $gx = y$, ou seja

$$x \sim y \iff \exists g \in G \quad | \quad gx = y. \quad (6.5)$$

A partir do conceito de similaridade definimos a classe de equivalência de x como o conjunto de

elementos de X similares a x , isto é

$$[x] = \{y \in X \mid y \sim x\}. \quad (6.6)$$

O problema clássico de reconhecimento de padrões é associar um conjunto de dados $\{X_i\}$ (vetores, palavras, imagens) a classes $\{t_i\}$ (categorias, cores, sim/não) usando um método consistente, para obter uma generalização sobre a estrutura das classes dadas. No nosso caso, soma-se ao desejo de classificar objetos o desejo de fazê-lo levando em conta algumas simetrias, ou seja, objetos que são similares (e portanto estão contidos na mesma classe de equivalência), devem ser associados à mesma classe. Podemos dizer então que estamos fazendo um reconhecimento de padrões não sobre o conjunto de objetos $\{X_i\}$, mas sobre as classes de equivalência de objetos $\{[X_i]\}$. Há diversas formas de obter algoritmos que sejam invariantes, no entanto, é interessante que o algoritmo seja eficiente, flexível, e que mostre uma boa robustez com respeito a erros.

6.1.1 Obtendo Representações Invariantes

Como foi dito acima, repito que queremos um algoritmo de classificação sobre o espaço de classes de equivalência e não sobre os pontos em si. Poderíamos para tal usar algoritmos tradicionais de reconhecimento de padrões e procurar treiná-los com conjuntos de dados que contivessem pontos característicos de uma classe de equivalência dada associados a uma categoria. Esta é entretanto uma possibilidade muito pouco prática, pois desta forma dependeríamos de um comportamento de convergência do algoritmo em um conjunto de dados sobre o qual não temos grande informação. Além disso, teríamos de ter informação sobre propriedades de generalização de algoritmos de aprendizado sobre o classes de equivalência e não sobre o conjunto de treinamento. Podemos por exemplo agrupar dados pertencentes à mesma classe e treinar um algoritmo bastante complexo, para que ele infira a estrutura das classes de equivalência. Não temos nenhuma garantia no entanto que a geometria inferida pelo algoritmo reflita a geometria correta. Podemos desta forma usar a informação que temos sobre os grupos de transformações que geram as classes de equivalência e criar um algoritmo mais poderoso. Uma forma ingênua de fazê-lo é gerar diversas *cópias* transformando nossos dados através das transformações de simetrias e incluí-las no conjunto de treinamento. Apesar de possível em essência, o método aumenta fortemente o custo computacional do treinamento, que passa a depender do número de cópias dos nossos dados que decidimos usar para representar suas classes, e esta dependência

torna esta idéia muito pouco prática em situações que temos grupos contínuos ou de dimensão maior.

Como elicitado, é melhor para a tarefa em questão desenvolver um algoritmo de classificação que leve em conta explicitamente as simetrias desejadas para a nossa classificação. Porém, isso ainda não especifica o nosso *ansatz*, como pode-se ver pela grande gama de métodos estudados para a tarefa (29), (30), (31) (32). A obtenção da invariância pode ser feita em duas etapas do processo de classificação. Primeiramente podemos escolher de alguma forma uma representação invariante para os nossos dados e valer-mo-nos de métodos tradicionais de aprendizado e classificação para extrair uma regra de classificação a partir das representações invariantes encontradas. Uma outra possibilidade é escolher um algoritmo que já contenha a informação das simetrias desejadas (como *kernels* invariantes (33)) ou então treinar algoritmos para que eles se adaptem às simetrias do problema (como descrito em (30)). Existe claramente um compromisso entre estas duas possibilidades, uma vez que treinar um algoritmo simples leva muito menos tempo computacional que treinar um algoritmo mais complexo (como um *support vector machine* por exemplo), porém obter representações invariantes envolve cálculos bastante custosos enquanto simplesmente usar os dados sem nenhum tratamento não exige custo adicional nenhum. Esta questão é bastante discutida no artigo de Burkhart (33) em que ele introduz métodos baseados em *kernels* invariantes para a mesma tarefa. Neste trabalho no entanto, favorecemos a primeira opção, de utilizar um pré-processamento dos dados para obter representações invariantes dos dados e utilizar métodos de classificação mais simples. Tomamos esta decisão por dois motivos. Primeiramente, baseamo-nos no resultado de que dadas algumas hipóteses sobre a estrutura do espaço de dados, pode-se demonstrar a existência de uma representação completa invariante para um grupo dado de transformações (34), isto é, uma representação tal que as representações invariantes de dois pontos no espaço de dados sejam iguais se e só se os pontos pertencerem à mesma classe de equivalência. Em segundo lugar, concentrando-nos na extração de *features* invariantes podemos adaptar nossa representação invariante para a tarefa em questão, desta forma maximizando a performance e minimizando a dimensão da representação a ser usada. A utilização de ferramentas mais poderosas inevitavelmente nos leva a um paradigma de *caixa-preta*, no qual não temos tanta flexibilidade para investigar possibilidades mais eficientes e baratas computacionalmente. E finalmente, escolhemos este *ansatz* por ele permitir-nos estudar o comportamento de diversas quantidades, como erros de generalização e velocidade do aprendizado em função do número de dimensões, distribuição de *features* e outros.

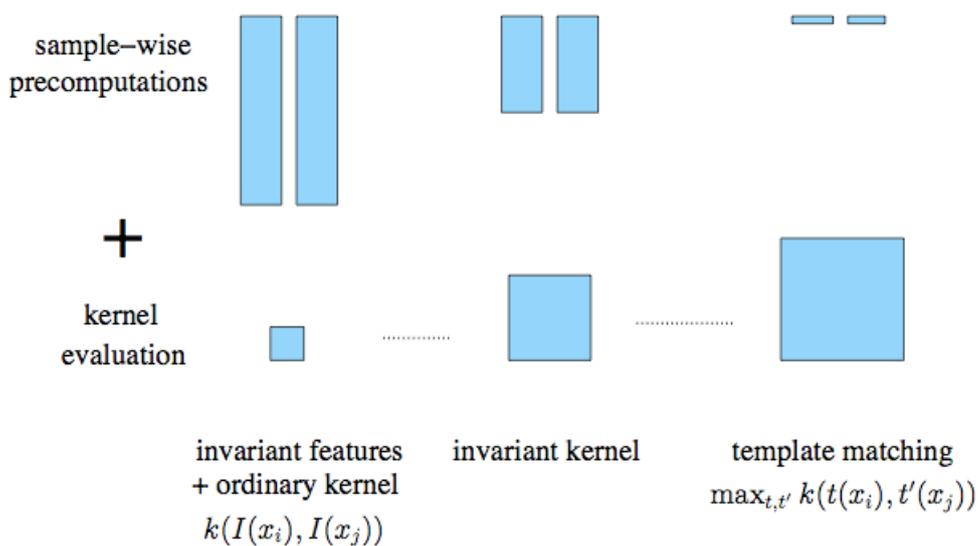


Figura 6.1: Esta figura descreve a escolha entre usar um pré-processamento pesado (*feature extraction*) e um algoritmo de aprendizado simples ou usar pouco pré-processamento e classificadores mais complexos como *Support Vector Machines* (figura de (33))

Resta-nos então a decisão de como representar imagens e como calcular representações invariantes das mesmas. Denotamos no que segue abaixo o grupo a ser integrado por G e o espaço de dados por S , sendo que as imagens são pontos $X \in S$. Podemos caracterizar uma imagem idealmente por uma função de um retângulo do plano num intervalo dos reais. Por exemplo uma imagem de N por M pixels em escala de cinza pode ser representada por uma função $X : [0, N] \times [0, M] \rightarrow [0, 1]$, de forma que o valor de $X(a, b)$ será o valor de cinza da imagem no ponto (a, b) . Usando a definição clássica de distância entre funções definidas em um intervalo podemos escrever uma distância entre duas imagens,

$$d(X, Y) = \int_A dw (X(\vec{w}) - Y(\vec{w}))^2,$$

onde definimos o domínio de integração $A = [0, N] \times [0, M]$. Esta definição, entretanto, claramente não apresenta qualquer invariância a rotações e translações. Podemos estender nossa distância efetuando uma minimização da distância sobre as classes de equivalência,

$$d_1(X, Y) = \min_{g \in G} \int_A dw (gX(\vec{w}) - Y(\vec{w}))^2.$$

A distância $d_1(\cdot, \cdot)$ já apresenta a desejável propriedade de que $d_1(X, gX) = 0$, como pode ser prontamente observado. Computá-la, porém, envolve a minimização em relação às transformações do grupo, que é uma tarefa pouco eficiente. Outra tentativa possível seria

$$d_2(X, Y) = \left(\int_G dg \quad gX(\vec{0}) - gY(\vec{0}) \right)^2,$$

que apresenta a mesma propriedade de d_1 . Denotamos por $\langle f \rangle_G$ a integral de f sobre todo o grupo de transformações, temos

$$\langle f \rangle_G = \frac{1}{\|G\|} \int_G gf(\vec{a})dg,$$

com o módulo do grupo G definido por

$$\|G\| = \int_G dg.$$

Dividindo pela raiz do módulo do grupo, teríamos então a distância

$$d'_2(X, Y) = \frac{1}{\sqrt{\|G\|}} \left(\int_G dg \quad gX(\vec{0}) - gY(\vec{0}) \right)^2 = (\langle X(\vec{0}) - Y(\vec{0}) \rangle_G)^2.$$

Uma rápida análise nos mostra no entanto que a função d_2 não representa as propriedades desejadas, uma vez que

$$d_2(X, Y) \propto (\langle X \rangle_G - \langle Y \rangle_G)^2,$$

e nos dá uma distância nula para imagens diferentes que tenham um valor de cinza médio igual. Claramente a integral $\langle X \rangle_G$ não depende do ponto em que começamos a integração sobre o grupo, uma vez que a integral sobre G inevitavelmente integrará sobre todo o domínio A . Vemos que para obter uma separação entre imagens diferentes pelo caminho sugerido teríamos de considerar outras funções. De forma geral podemos escrever uma distância como

$$d_I(X, Y) = \sum_i (\langle T_i[X] \rangle_G - \langle T_i[Y] \rangle_G)^2,$$

em que $T_i[\]$ são funcionais. A distância claramente tem sua origem no mapa $T : S \rightarrow F = \mathcal{R}^N$ definido por

$$T[X] = (\langle T_1[X] \rangle_G, \langle T_2[X] \rangle_G, \dots, \langle T_N[X] \rangle_G)^t.$$

A questão fundamental a ser respondida neste trabalho é em que condições um número finito de funcionais $T_i[\]$ é suficiente para separar padrões em imagens e como otimizar o número de

padrões escolhendo os funcionais da melhor forma possível.

O problema de determinar se existe um espaço de *features* para as classes de equivalência associadas a um grupo de simetrias em um problema de reconhecimento de padrões foi abordado em um artigo (34) em que o autor determina critérios suficientes para que exista um espaço de *features* completo. Um espaço de *features* completo é um espaço tal que para todas os padrões do espaço dos sinais S tenhamos

$$T[X] = T[Y] \iff \exists g \in G, gX = Y.$$

Ou seja, um espaço de *features* completo para um dado conjunto de classes de equivalência é um espaço que separe as classes de equivalência. Para que tenhamos um espaço completo de *features* basta que o conjunto de polinômios invariantes seja finitamente gerado e que o espaço de sinais S seja separável. A primeira condição pode ser assegurada pelo fato de que todas as representações do grupo em questão sejam completamente redutíveis. A segunda condição, entretanto, é mais sutil, no sentido de que assumir que os dados são separáveis é assumir que o problema é solúvel, que é em parte o que queremos demonstrar. Entretanto, dada a alta dimensionalidade do espaço em que os dados estão inseridos no caso de reconhecimento de imagens, esta hipótese pode ser razoável. Caso ambas as hipóteses sejam satisfeitas, podemos tomar como conjunto de *features* simplesmente a base do conjunto de polinômios invariantes. Em aplicações, entretanto, se torna pouco prático encontrar a base do conjunto de polinômios invariantes, uma vez que estamos trabalhando num espaço de dimensão alta.

Vale notar que a condição de que exista um espaço completo e invariante de *features* não garante eficiência do algoritmo de classificação. Poderíamos perfeitamente obter uma representação invariante completa que resulta em uma representação muito pouco eficiente. Podemos ter casos onde pequenas alterações no espaço de sinais acarretam grandes mudanças no espaço invariante, de forma que a sensibilidade da representação a ruído a torne inútil. Seria desejável portanto que a representação seja uma função contínua de variações no espaço de sinais. Para propósitos de classificação é interessante ainda que a separação entre as classes seja grande o suficiente para que possamos discernir entre as classes com facilidade.

O caso mais discutido de separabilidade de dados é sem dúvida o caso de dados linearmente separáveis. Caso um conjunto de dados seja linearmente separável, o algoritmo de treinamento do perceptron converge para um classificador correto em tempo finito. Entretanto é complicado

deduzir limites para o tempo de convergência. Uma solução para este problema foi proposta por Nabutovsky e Domany (35) em um artigo de 1991. O algoritmo proposto por eles consiste em uma regra de aprendizado que além de guardar o valor dos pesos sinápticos da rede em cada passo de tempo, guarda o valor de uma variável d apelidada de *desespero*. Pode-se demonstrar que caso os dados sejam linearmente separáveis, a variável d fica abaixo de um limitante superior. Desta forma podemos verificar explicitamente a separabilidade dos dados. Para casos em que temos mais de duas classes ou que a separabilidade linear não vale, temos de nos valer de outros métodos para verificar a separabilidade dos dados. Uma forma muito popular é a utilização de *support vector machines*, que utilizam *kernels* para determinar um conjunto de vetores do conjunto de dados a partir dos quais podemos determinar uma superfície de decisão. Outro conjunto de métodos bastante populares são os métodos de *clustering* como o método de *superparamagnetic clustering* (36) ou o método de *vector quantization* (37), que procuram separar o conjunto de dados em aglomerados que partilhem de características próximas. Ambos têm a vantagem em relação às *SVM*'s de terem um custo computacional relativamente baixo.

No mesmo sentido, mas tomando um método mais generativo, podemos estudar o comportamento de *features* construídos a partir de funções arbitrárias e estudar o comportamento da separabilidade a partir de métodos de classificação em função do número de *features* usados. Para simplificar o método, limitamo-nos neste trabalho a funções polinomiais dos pixels do tipo

$$T_i[X] = X(\vec{a}_1)^{p_1} X(\vec{a}_2)^{p_2} \dots X(\vec{a}_n)^{p_n}.$$

Aonde o conjunto de pontos $\{\vec{a}_1, \vec{a}_2, \dots, \vec{a}_n\}$ e as potências $\{p_1, \dots, p_n\}$ são as quantidades que definem a função $T_i[\cdot]$.

O algoritmo desenvolvido no decorrer do projeto consiste em escolher aleatoriamente conjuntos de inteiros $\{a_1, a_2, \dots, a_n\}$ e $\{p_1, \dots, p_n\}$ e utilizar estes funcionais como *features* para discriminação de imagens. Apesar da pouca praticidade em escolher aleatoriamente os *features*, podemos desta forma analisar o comportamento da discriminação entre padrões em função do número de *features*. Para aplicações surge a possibilidade de utilizarmos algum algoritmo de otimização para determinar os funcionais que se adaptam da melhor forma aos dados do problema em questão. Uma possibilidade de implementação surgiu no estudo dos algoritmos chamados *Learning Vector Quantization*(38) que será descrita abaixo. Uma possibilidade que apresenta sérias limitações práticas, mas que seria de grande valor para um estudo teórico seria

definir um update bayesiano sobre uma distribuição de probabilidades sobre os funcionais T_i , usando a taxa de erro sobre o conjunto de treinamento como verossimilhança.

Baseado no estudo acima, tomamos então como ponto de partida para o estudo de *features* invariantes a seguinte idéia: tomamos conjuntos de n vetores $A^i = \{\vec{a}_1^i, \vec{a}_2^i, \dots, \vec{a}_n^i\}$ e $P^i = \{p_1^i, \dots, p_n^i\}$, escolhidos aleatoriamente de uma distribuição *ad hoc* e trabalhamos com estes funcionais $T_i = T(A^i, P^i)$ como *features* a serem examinados. Uma implementação eficiente da extração destes *features* está descrita mais abaixo.

O grupo de simetrias de interesse no caso que estudamos é o grupo das rotações e translações em duas dimensões. Para calcular as integrais necessárias temos então de parametrizar o grupo. Tomando as translações

$$S_{\vec{a}}X(\vec{z}) = X(\vec{z} + \vec{a})$$

e as rotações em torno da origem

$$R_{\alpha}X(\vec{z}) = X(M_{\alpha}\vec{z})$$

aonde denotamos por M_{α} a matriz de rotação em duas dimensões por um ângulo α :

$$M_{\alpha} = \begin{pmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{pmatrix}$$

podemos então escrever as integrais sobre o grupo como

$$\begin{aligned} \langle T_i[X] \rangle_G &= \int_A d\vec{a} \int_0^{2\pi} d\alpha T_i[S_{\vec{a}}R_{\alpha}X], \\ \langle T_i[X] \rangle_G &= \int_A \int_0^{2\pi} d\vec{a} d\alpha X(M_{\alpha}\vec{a}_1^i + \vec{a})^{p_1} \dots X(M_{\alpha}\vec{a}_n^i + \vec{a})^{p_n}. \end{aligned}$$

Neste formato podemos facilmente discretizar a integral e aplicar o método de integração de Romberg para calcular o valor numérico da integral.

6.1.2 Implementação Computacional

Para obter um ganho maior na precisão de integrais numéricas sem ter de diminuir muito o passo de integração usado podemos usar o esquema de integração de Romberg baseado no método de extrapolação de Richardson. Para uma integral de uma função de um intervalo da reta dos reais nos reais podemos descrever o esquema de Romberg como segue abaixo. Seja a

integral a ser aproximada

$$I = \int_a^b f(x)dx$$

e seja a regra do trapézio de passo de integração $h_n = \frac{b-a}{2^n}$

$$T_{0,n} = \frac{b-a}{2^n} \left(\frac{f(a)}{2} + \sum_{i=1}^{2^n-1} f(a+ih) + \frac{f(b)}{2} \right),$$

que tem um erro assintótico da ordem do intervalo de integração elevado ao quadrado, temos

$$|I - T_{0,n}| = O\left(\left(\frac{b-a}{2^n}\right)^2\right) = O(h_n^2).$$

Partindo disto definimos então o esquema de recursão de Romberg como

$$T_{m,k} = \frac{4^m T_{m-1,k+1} - T_{m-1,k}}{4^m - 1}.$$

Os elementos $T_{m,n}$ construídos desta forma terão um erro em relação a integral exata que será $O\left(\left(\frac{b-a}{2^n}\right)^{2m+2}\right)$. Normalmente se toma uma série de aproximações trapezoidais e utiliza-se o elemento $T_{m,m}$ com o maior m possível como aproximação da integral. Uma generalização para este esquema para integrais multidimensionais foi proposto em (39) simplesmente redefinindo a regra do trapézio com a mesma regra de recursão. Para a integral multidimensional

$$I' = \int_{a_1}^{b_1} dx_1 \int_{a_2}^{b_2} dx_2 \dots \int_{a_n}^{b_n} dx_n f(x_1, x_2, \dots, x_n)$$

teremos a regra do *hiperparalelograma*

$$T_{0,k} = \prod_{i=1}^n \frac{b_i - a_i}{2^{k+1}} \sum_{\lambda_1=0}^{2^k-1} \dots \sum_{\lambda_n=0}^{2^k-1} (f(a_1 + \lambda_1 h_1, \dots, a_n + \lambda_n h_n) + f(a_1 + (\lambda_1 + 1)h_1, \dots, a_n + (\lambda_n + 1)h_n))$$

e o mesmo esquema de recursão. Assim asseguramos que o erro do elemento $T_{k,0}$ é da ordem de h^{2m+2} . A definição de convergência de erros para este esquema é um pouco diferente do usual, dizemos que uma função $f(h_1, h_2, \dots, h_n) = O(h^m)$ se existe uma constante C tal que

$$\|f(h_1, \dots, h_n)\| < C h_1^{\beta_1} h_2^{\beta_2} \dots h_n^{\beta_n}$$

quando os $h_i \rightarrow 0$ mantendo as razões $\frac{h_i}{h_j}$ limitadas para todos os i, j e a ordem de convergência m é dada por

$$m = \sum \beta_i.$$

Aplicando estes princípios para a integral de grupo mencionada acima temos, escrevendo como

$$g(\alpha, a, b) = f(X(\cos(\alpha)x + a, \sin(\alpha)y + b))$$

em que a função f é a função a ser integrada sobre o grupo. Temos então

$$I = \int_0^M \int_0^N \int_0^{2\pi} da db d\alpha X(\cos(\alpha)x_1 + a, \sin(\alpha)y_1 + b)^{p_1} \dots X(\cos(\alpha)x_n + a, \sin(\alpha)y_n + b)^{p_n},$$

ou simplificando com a definição acima

$$I = \int_0^M \int_0^N \int_0^{2\pi} da db d\alpha g(\alpha, a, b).$$

Temos então a regra do trapézio

$$T_{0,k} = \frac{ab2\pi}{2^{3k+3}} \sum_{\lambda_1=0}^{2^k-1} \sum_{\lambda_2=0}^{2^k-1} \sum_{\lambda_3=0}^{2^k-1} (g(\lambda_1, \lambda_2, \lambda_3) + g(\lambda_1 + 1, \lambda_2 + 1, \lambda_3 + 1)).$$

Utilizando-nos do esquema acima podemos então calcular as integrais mencionadas com uma precisão melhor no mesmo intervalo de tempo.

Apesar do ganho substancial obtido pela utilização do método de integração de Romberg, o cálculo das quantidades envolvidas ainda é bastante intenso no que diz respeito a tempo computacional. A estrutura do algoritmo entretanto se mostra extremamente propícia para uma implementação paralela. Para tirar proveito deste fato, o algoritmo foi paralelizado com a biblioteca MPI(40). Este programa foi então otimizado para a arquitetura do cluster de processadores do grupo de física estatística do instituto de Física da USP, de 32 nós.

O aprendizado de técnicas de programação paralela foi bastante proveitoso e pôde ser reutilizado em outros momentos do projeto, como no desenvolvimento de simulações de Monte Carlo para o modelo de Ising e nas simulações para determinar o comportamento do aprendizado do perceptron em função do número de *features*.

Quando do desenvolvimento dos primeiros circuitos integrados a *IBM* investigou quais seriam as limitações de velocidade para programas executados em paralelo. Isto levou à formulação das leis de Amdahl, que afirmam que o ganho de velocidade de processamento em computadores altamente paralelos é fortemente limitado pela comunicação entre os nós do cluster(41). Na época, isso foi motivo para a *IBM* abandonar a pesquisa em programação paralela, devido ao alto custo e à baixa velocidade das redes na época. O advento das redes atuais trouxe

um novo incentivo para os clusters. Entretanto, após um curto período de hegemonia no rama de supercomputadores acessíveis, os clusters estão novamente perdendo espaço para computadores baseados em unidades de processamento gráfico (*GPU*) que têm uma arquitetura e bibliotecas dedicadas a efetuar processamento fortemente paralelo. De qualquer forma, independentemente da realização física, a programação paralela continua sendo uma das grandes tendências em computação de alta performance.

6.1.3 Além do *Grayscale*

Claramente a restrição de usar somente imagens em tons de cinza restringe bastante o tipo de imagens a serem usadas nas tarefas propostas. Podemos facilmente estender o formalismo para o caso de imagens em escala *RGB* (vermelho, verde e azul), representando-as como uma tripla de funções do domínio A em $[0, 1]$. Desta forma, teríamos que uma imagem X seria representada por

$$X = (X^r, X^g, X^b)$$

sendo X^r, X^g, X^b as componentes referentes a cada uma das cores. Esta modificação aumenta mais uma vez significativamente o número de funcionais que podemos escolher para definir os *features* e nos deixa com a tarefa de decidir se desejamos considerar algum tipo de simetria adicional sobre a tripla X . O caso de escala de cinza corresponde a adicionar às simetrias de rotação e translação a simetria de permutações das cores. Desta forma somamos sobre todas as permutações das cores, gerando representações independentes da ordem destas e dependentes só das invariantes por permutações

$$\phi_1 = X^r + X^g + X^b,$$

$$\phi_2 = X^r X^g + X^r X^b + X^g X^b,$$

$$\phi_3 = X^r X^g X^b$$

6.2 Métodos de Classificação

Vamos brevemente comentar os métodos de classificação usados no resto deste capítulo. Como o presente trabalho intenciona situar-se no campo de trabalho de reconhecimento de padrões e aprendizado de máquina, houve um grande trabalho na direção de compreender os diversos métodos utilizados em tarefas de classificação atualmente.

O exemplo mais básico de um algoritmo de classificação é o perceptron. Proposto por Rosenblatt em 1957, hoje ele está presente em inúmeras aplicações cotidianas. Apesar de bem simples, sua análise resultou em resultados bastante interessantes sobre capacidade de redes, e ele se mantém como o bloco de construção de grande parte do que é feito em redes neurais. O perceptron só é capaz de aprender problemas linearmente separáveis, o que o torna sua aplicação pouco prática em uma variedade de casos. Entretanto, redes neurais de arquitetura mais complexa podem facilmente superar este problema e a regra de aprendizado de retropropagação de erros fornece uma forma eficiente de calcular os pesos sinápticos ótimos para um problema dado. Outros métodos também foram estudados, como os métodos baseados em kernels e algoritmos de clustering.

Apesar de diversos trabalhos terem sido dedicados a encontrar métodos de kernel invariantes, resultando em algoritmos bastante satisfatórios(33)(42), métodos como estes têm um custo computacional relativamente grande mesmo após o treinamento do classificador. Métodos mais simples como redes neurais multicamada, têm por vezes um custo de treinamento mais alto, com um tempo de convergência mais lento, entretanto o ganho vigoroso de velocidade após o treinamento nos levou a escolher estes métodos como foco.

Outro tipo de método de classificação que tem sido utilizado com grande sucesso são os métodos de clustering, entre os quais destacam-se o método de *K-nearest neighbours*(43), o algoritmo de *Learning Vector Quantization*(38) e alguns métodos não-paramétricos, como o *Superparamagnetic Clustering*(36) e o *Sorting Points into Neighbourhoods*(18).

6.2.1 Perceptron

Apesar de bastante ingênuo, a regra de aprendizado do perceptron nos dá uma forma simples e eficiente de verificar a separabilidade de um conjunto de dados. O perceptron é uma rede

neural artificial de uma camada e tem uma regra de aprendizado bastante simples, que foi a primeira regra de aprendizado *on-line* a ser proposta. Dados vetores ξ^i associados a duas classes, indexadas pelas variáveis $t^i = \pm 1$, o problema de aprendizado consiste em encontrar um vetor de pesos w tal que

$$\text{sign}(w \cdot \xi^i) = t^i, \forall i.$$

Acima consideramos $\text{sign}(x) = \frac{x}{\|x\|}$ a função sinal e $w \cdot \xi^i$ o produto escalar euclidiano. A regra do percéptron consiste em atualizar os pesos w iterativamente, em um laço que percorre repetidamente os exemplos fazendo

$$w^{n+1} = w^n + t^i \xi^i, \text{ se } \text{sign}(w \cdot \xi^i) t^i = -1,$$

e deixando w inalterado caso contrário. Esta regra tem convergência garantida caso o problema seja linearmente separável.

Podemos então utilizar um percéptron como teste de separabilidade para nossos dados. Para tal tomamos imagens de caracteres sujeitos a rotações e translações (impondo condições periódicas de contorno), e treinamos um percéptron para separá-los. Para acessar a separabilidade dos dados em função do número de *features* gerados, podemos gerar um conjunto de dados com diversos números de *features* e calcular o número de iterações da regra do percéptron necessárias para a convergência do algoritmo. Desta forma veremos a separabilidade dos dados em função do número de *features*. Claramente para tarefas envolvendo mais de duas classes de equivalência uma análise mais complexa se faz necessária para determinar o número apropriado de *features*, entretanto, separabilidade dois a dois já nos permite inferir separabilidade dos dados em geral. Podemos ver no gráfico abaixo o tempo médio do convergência para o algoritmo do percéptron em função do número de *features*. Vemos claramente mais uma vez o comportamento evidenciado pela análise geométrica da sessão anterior. Temos um forte ganho de representatividade para um certo número de *features*, a partir do qual a inclusão de um número maior de *features* passa a não trazer grandes benefícios para os propósitos de classificação.

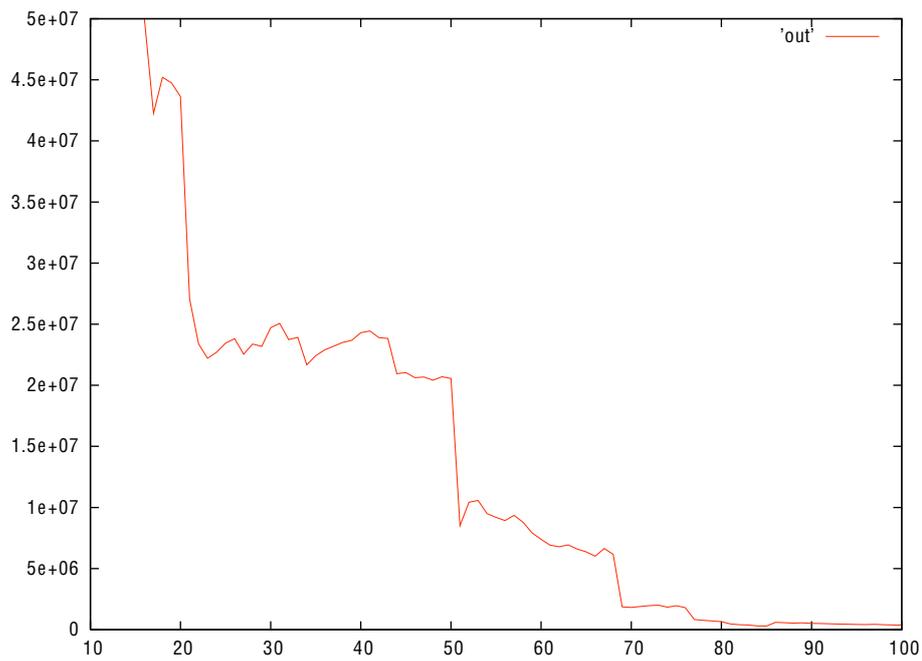


Figura 6.2: Tempo médio de convergência da regra de aprendizado do perceptron em função do número de *features*. O eixo vertical está em iterações da regra do perceptron sobre o conjunto de aprendizado até a convergência. O eixo horizontal apresenta o número de *features* usados. Podemos ver claramente o ganho para a separabilidade em espaços de maior dimensão, apesar da aparente saturação após certa dimensão.

6.2.2 Learning Vector Quantization

O formalismo dos algoritmos conhecidos por *Learning Vector Quantization*(38) (37) (44) fornece uma alternativa intuitiva e simples a métodos mais complexos de classificação. A idéia do algoritmo consiste em determinar um ou mais protótipos para cada classe do problema em questão de forma que o espaço seja dividido em classes de acordo com as distâncias aos protótipos, isto é, dados protótipos $\{w_1, \dots, w_n\}$ associados a classes $\{c_1, \dots, c_n\}$ o algoritmo (em sua versão mais simples) associa um vetor x a uma classe c se e só se

$$\exists w_j \mid \forall k \parallel x - w_k \parallel \geq \parallel x - w_j \parallel, e c_j = c.$$

O esquema mais simples de lvq, chamado de *winner takes all* é o descrito acima. Podemos entretanto utilizar as distâncias dos protótipos para treinar modelos generativos e obter superfícies de decisão por critérios mais refinados. A regra de aprendizado mais clássica para LVQ é simplesmente corrigir o protótipo mais próximo de cada vetor x_i do conjunto de treinamento pertencente à classe y^i segundo a regra

$$w_j^{n+1} = \left\{ \begin{array}{ll} w_j^n + \varepsilon(x_i - w_j^n) & \text{se } y^i = c^j \\ w_j^n - \varepsilon(x_i - w_j^n) & \text{caso contrário,} \end{array} \right\}$$

onde ε é uma taxa de aprendizado.

Várias outras variantes podem ser definidas, desde regras que corrigem todos os protótipos a cada iteração até regras mais simples que corrigem o protótipo correto mais próximo e o incorreto mais próximo. Entretanto uma modificação mais interessante nestes algoritmos é introduzir um aprendizado sobre a métrica incluindo algumas restrições. Desta forma podemos executar o que é chamado de *relevance learning*, ou seja aprender a relevância de cada dimensão do problema(44). Desta forma podemos também *podar* dimensões do problema, desconsiderando dimensões irrelevantes, que não adicionam nenhuma qualidade à classificação.

A idéia descrita em (44) é entretanto bastante restritiva, no sentido de só permitir métricas diagonais. Uma outra forma de usar os princípios descritos é permitir que a métrica assuma uma forma arbitrária, permitindo incluir nela possíveis correlações entre dimensões do problema e outras possibilidades. Esta generalização tem a desvantagem de não mais permitir a poda de dimensões irrelevantes, uma vez que a relevância de um vetor da base não poderá mais ser esti-

mada somente pelo termo diagonal da métrica. Entretanto, poderíamos facilmente desenvolver outros critérios para a eliminação de dimensões do problema. O método descrito em (44) consiste em uma descida no gradiente com uma função erro do tipo

$$E = \sum_{i=1}^m \text{sgd}\left(\frac{d_j^i - d_k^i}{d_j^i + d_k^i}\right),$$

com $\text{sgd}(x) = (1 + e^{-x})^{-1}$ e d_j^i (d_k^i) a distância do vetor i até o protótipo correto (incorreto) mais próximo. A inclusão de métricas alternativas é feita simplesmente pela substituição da distância euclidiana por uma distância generalizada dada por uma métrica g_{lk}

$$\|x_i - w_j\|_{g_{lk}} = g_{lk}(x_i^l - w_j^l)(x_i^k - w_j^k),$$

onde estamos adotando a convenção de Einstein em que índices repetidos são somados. Podemos então executar uma descida de gradiente não só nas variáveis w_j mas também na métrica g_{lk} . Temos que tomar cuidado no entanto para não sairmos do domínio das métricas válidas, isto é temos que ter certeza que a distância definida por esta métrica satisfaça os axiomas que definem uma distância. Que a distância entre um ponto e ele mesmo seja sempre 0 é satisfeito imediatamente, entretanto a condição de que

$$g_{lk}x^l x^k \geq 0, \forall x$$

tem que ser verificada caso a caso. No algoritmo de GRLVQ(44) isso é garantido tomando uma taxa de aprendizado pequena para a descida no gradiente nos parâmetros da métrica e normalizando os valores a cada passo. Uma outra forma é tomar a descida do gradiente com um multiplicador de Lagrange que imponha que a matriz g_{lk} seja positiva definida. O problema é que a condição de positividade não é uma condição apropriada para usarmos os multiplicadores de Lagrange, pois implementa um vínculo sobre todos os pontos do espaço. Uma possibilidade para tratar este empecilho é usar a condição de positividade só sobre os protótipos, possibilitando assim a implementação do algoritmo para métricas mais gerais. Outra possibilidade seria aplicar a *rank approximation* sobre a decomposição SVD da métrica e jogar fora os menores autovalores da métrica. Desta forma poderíamos obter uma representação mais simples e de dimensão mais baixa para os dados.

6.2.3 Superparamagnetic Clustering

Métodos de *clustering* não são necessariamente algoritmos de classificação, apesar de podermos utilizá-los como tais na maioria dos casos. No entanto, podemos no caso de *clustering* deixar em aberto o número de classes a serem usadas, e não especificar a estrutura dos aglomerados (*clusters*), no que é referido como método não-paramétrico de *clustering*. Um exemplo que foi inspirado no estudo de física do estado sólido é o método de *superparamagnetic clustering*, que se vale de simulações de Monte Carlo de um sistema de spins de Potts para estimar a correlação entre os pontos e então separa-os em classes de acordo com uma regra de limiar sobre a correlação entre pontos. As variáveis de um sistema de spins de Potts são dadas por inteiros de 0 a Q , aonde Q é um valor que define o modelo estudado. Cada partícula pode ter um *spin* de 0 a Q e as interações se dão somente entre partículas de mesmo spin de forma a baixar a energia. Temos então

$$\mathcal{H} = - \sum_{(i,j)} J_{ij} \delta_{\sigma_i, \sigma_j},$$

onde os pares (i, j) são os vizinhos da rede, J_{ij} é a intensidade da interação entre os vizinhos, $\delta_{i,j}$ é o delta de Kronecker e σ_i é o valor do spin da partícula i . Podemos então efetuar simulações de Monte Carlo deste sistema e estimar as correlações $\langle \delta_{\sigma_i, \sigma_j} \rangle$ a partir da amostragem dos estados. De posse podemos definir a partir de um limiar θ que i e j estão no mesmo cluster sempre que $\langle \delta_{\sigma_i, \sigma_j} \rangle > \theta$. A utilização deste sistema como um método de aglomeração reside na escolha das intensidades das interações de acordo com a discrepância entre os dados. Podemos, no caso de termos dados vetoriais, tomar simplesmente $J_{ij} = \exp(-\|S_i - S_j\|)$. Além disso, no artigo (36) em que o método é apresentado as interações ainda são restritas aos N vizinhos mais próximos, parâmetro do qual os resultados não dependem sensivelmente. Resta no entanto a questão de escolher a temperatura na qual fazemos esta simulação, que é a origem do termo *superparamagnetic* no nome do método. A escolha da temperatura consiste em encontrar a fase superparamagnética do sistema, que é uma fase intermediária entre a fase ferromagnética e a fase paramagnética, em que não temos magnetização total mas temos a formação de domínios de larga escala. Isso é feito percorrendo o domínio das temperaturas e verificando-se a fase pelos picos do calor específico do sistema.

6.2.4 Kernel-based Methods

Apesar da grande diversidade que há no campo de aprendizado computacional, houve nos últimos anos uma forte investida na direção dos métodos baseados em *kernels*. A idéia que embasa todo o trabalho feito em métodos de *kernels* reside na observação de que apesar de treinarmos redes neurais com pesos sinápticos e vetores de pesos, a classificação pode sempre ser reescrita como uma combinação linear de produtos internos entre os exemplos e os dados a serem classificados. No caso do aprendizado do perceptron podemos observar isso notando que todas correções aos vetores de peso são proporcionais aos vetores exemplo, de forma que o vetor de pesos pode sempre ser reescrito como uma combinação linear dos exemplos. Podemos então reescrever o problema de aprendizado para não encontrar o vetor de pesos explicitamente, mas encontrar os pesos da combinação linear de pesos que nos dá o vetor correto. A classificação não seria mais escrita como

$$\sigma(X) = \text{sign}(J \cdot X),$$

mas como

$$\sigma(X) = \text{sign}\left(\sum_i \alpha_i S_i \cdot X\right) = \text{sign}\left(\sum_i \alpha_i k(S_i, X)\right).$$

Na última equação introduzimos convenientemente a notação $S_i \cdot X = k(S_i, X)$, para introduzir a idéia de *kernel*. Kernels são generalizações de produtos escalares, no sentido que eles são funções de duas variáveis nos reais com a propriedade de que $k(X, X) > 0$ (o caso de kernels semidefinidos positivos em que $k(X, X) \geq 0$, também é útil, mas não será abordado). Os métodos de kernels como os *Support Vector Machines* são métodos nos quais procuramos obter os coeficientes α_i que satisfaçam devidos vínculos, que podem requerer que a classificação seja correta por exemplo. A vantagem deste formalismo é que ele nos permite estender a analogia do kernel como produto escalar a funções arbitrárias, resultando em métodos de classificação mais poderosos. Apesar de os exemplos descritos se basearem em dados vetoriais, existem extensões para textos e dados de bioinformática, entre outros.

O problema de classificação linear é o de encontrar um hiperplano separador. Dado um conjunto de dados $\mathcal{D} = \{x_i, t_i\}$, onde $x_i \in \mathbb{R}^N$, $t_i = \pm 1$, podemos descrever o problema como a tarefa de encontrar um vetor w tal que para todo par (x_i, t_i) do conjunto de treinamento

$$t_i(w \cdot x_i) > 0,$$

podemos ainda adicionar um viés, obtendo

$$t_i(w \cdot x_i + b) > 0.$$

No entanto, como o problema é linear, qualquer vetor w e viés b que satisfaçam as condições podem gerar todo um subespaço de soluções $(\alpha w, \alpha b)$. Para superar este problema, vamos mudar as condições, utilizando-nos desta liberdade, reescrevendo as equações dos vínculos como

$$t_i(w \cdot x_i + b) \geq 1.$$

Desta forma podemos então escolher um w particular minimizando o módulo do vetor. Temos então o problema de minimizar $f(w) = \frac{1}{2} \|w\|^2$ sujeito aos vínculos dados. Incluindo multiplicadores de Lagrange para cada vínculo teremos então

$$L(w, a, b) = \frac{1}{2} \|w\|^2 - \sum a_i \{t_i(w \cdot x_i + b) - 1\}.$$

Igualando as derivadas em relação a w e b a zero, temos as condições de mínimo

$$w = \sum a_i t_i x_i \tag{6.7}$$

$$0 = \sum a_i t_i.$$

Substituindo estas expressões na lagrangiana anterior obtemos o problema de otimização dual, agora uma função somente dos a_i .

$$\tilde{L}(a) = \sum a_i - \frac{1}{2} \sum_i \sum_j a_i a_j t_i t_j x_i \cdot x_j,$$

onde a está sujeito aos vínculos

$$a_i \geq 0, \quad \sum a_i t_i = 0.$$

Podemos agora definir $k(x_i, x_j) = x_i \cdot x_j$, e teremos

$$\tilde{L}(a) = \sum a_i - \frac{1}{2} \sum_i \sum_j a_i a_j t_i t_j k(x_i, x_j). \tag{6.8}$$

Vemos portanto que no caso geral o problema de classificação de vetores pode ser escrito exclusivamente como função dos produtos escalares entre os vetores do conjunto de treinamento e das classificações destes vetores. Para classificar novos vetores com o classificador obtido pela

minimização de 6.8 usamos a relação 6.7 para escrever

$$y(x) = w \cdot x + b = \sum a_i t_i x_i \cdot x + b = \sum a_i t_i k(x_i, x) + b.$$

Podemos então usando o formalismo de otimização quadrática tratar um problema de aprendizado em espaço de alta dimensão desde que saibamos calcular produtos escalares neste espaço. Desta forma generalizamos o método e podemos utilizar qualquer escolha de funções $k(\cdot, \cdot)$, desde que estas satisfaçam as condições de um produto interno. Nominalmente, devemos exigir que a função k seja positiva definida ($k(x, x) \geq 0, \forall x$), para garantir que $k(x, y)$ represente um produto escalar em algum espaço. Chamamos as funções $k(\cdot, \cdot)$ satisfazendo estas relações de *kernels* e os classificadores obtidos por este processo são chamados de classificadores de margem máxima¹ devido ao fato de a distância entre os vetores mais próximos do hiperplano separador ser dada por $\frac{2}{\|w\|}$.

Os lemas de Mercer permitem estender o número de *kernels* disponíveis através de uma série de relações. Entre exemplos conhecidos temos por exemplo as *radial basis functions*, como por exemplo,

$$k_{rbf}(x, y) = e^{-\|x-y\|^2}$$

ou os kernels polinomiais

$$k_d(x, y) = (1 + x \cdot y)^d,$$

que representam o produto escalar num espaço que tem como coordenadas todos os produtos de coordenadas dos vetores x até ordem d .

No artigo de Burkhart (33) os autores discutem uma variedade de técnicas que poderiam ser usadas para obter-se kernels invariantes por um grupo de transformações predefinido. Particularmente eles focam em dois meios de obter kernels invariantes, a integração de transformações e a substituição de distância invariante. O primeiro meio consiste em simplesmente tomar um kernel não-invariante e integrar sobre todas as transformações de grupo nos dois argumentos obtendo desta forma um kernel invariante. Teríamos então

$$k_{TI}(x, y) = \int dg \int dg' k(gx, g'y).$$

No caso de substituição de distância invariante teríamos de tomar um kernel que dependa so-

¹*Maximum margin classifiers*

mente da distância entre seus argumentos e substituir esta distância por uma distância invariante, por exemplo $\|x - y\|_I = \min_g \|x - gy\|$. Temos então

$$k_{IDS}(x, y) = e^{-\|x - y\|_I^2}.$$

O kernel que procuramos estudar pertence à primeira classe. Procuramos usar o kernel

$$k_d(X, Y) = \int dg dg' (1 + X(gx_0)Y(g'x_0))^d,$$

onde X e Y são representações de imagens como funções do plano na reta como descrito anteriormente. Podemos reescrever este kernel de forma a explicitar algumas relações

$$k_d(X, Y) = 1 + \sum_{n=1}^d \int \prod_{i=1}^n dg_i dg'_i \prod_{i=1}^n X(g_i x_0) Y(g'_i x_0).$$

Vemos então que de certa forma, todos os possíveis *features* gerados por produto de pixels até d pixels estão representados no *kernel* integrados sobre as distâncias e ângulos entre eles. Poderíamos esperar então que a classificação advinda deste *kernel* representasse a melhor classificação que podemos obter usando *features* como descritos acima. Resultados preliminares mostram no entanto que o *kernel* descrito não apresenta bons resultados.

6.3 Features Invariantes

Para verificar qualitativamente a representatividade do espaço invariante gerado pelos *features* calculamos a projeção

$$\phi = \frac{T[X] \cdot T[Y]}{\|T[X]\| \|T[Y]\|}$$

entre os vetores de *features* gerados por diferentes imagens em função do número de *features*. Para tal geramos um número N de features repetidamente e calculamos a projeção ϕ entre as imagens da mesma classe e entre classes diferentes para diferentes valores de N . No gráfico abaixo podemos ver um gráfico das projeções entre quatro imagens, duas pertencentes a cada classe. Podemos ver claramente que o alinhamento entre vetores de *features* da mesma classe de equivalência é maior que o alinhamento entre vetores de classes distintas.

Outra observação interessante que pode ser feita é a saturação da separação a partir de um certo número de *features*. Vemos claramente nos gráficos que acima de aproximadamente 20

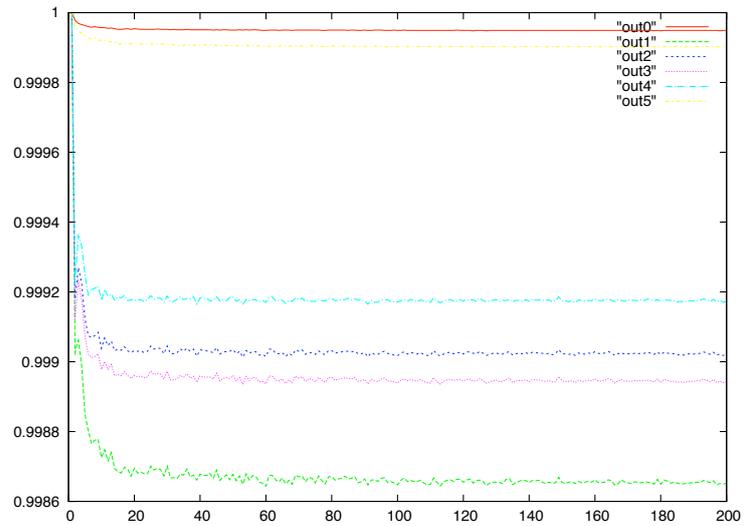


Figura 6.3: Gráfico do ângulo entre um vetor de *features* e outro entre classes iguais ("out0" e "out5") e entre classes diferentes ("out1" a "out4"). Apesar da pouca diferença do ângulo entre vetores da mesma classe e de classes diferentes, podemos claramente ver a separação. Podemos ver a separação melhor se subtraímos a média dos vetores, mostrado abaixo na figura 6.4.

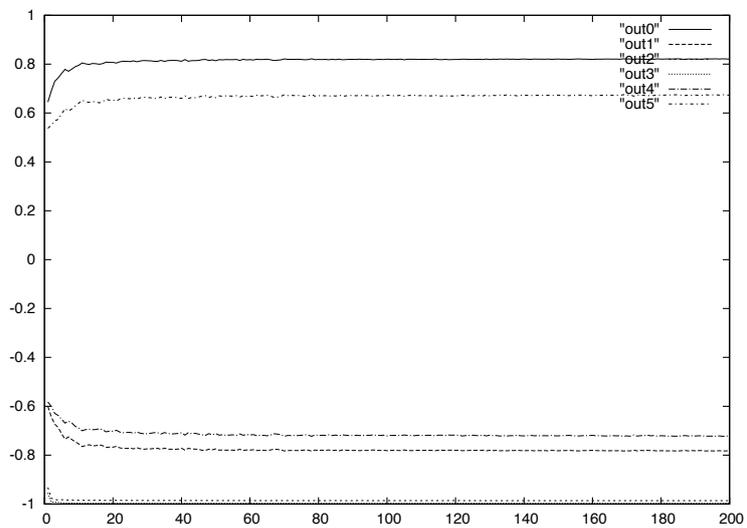


Figura 6.4: Gráfico do ângulo entre um vetor de *features* e outro entre classes iguais ("out0" e "out5") e entre classes diferentes ("out1" a "out4") após a subtração da média dos vetores. Vemos claramente a separação entre as classes.

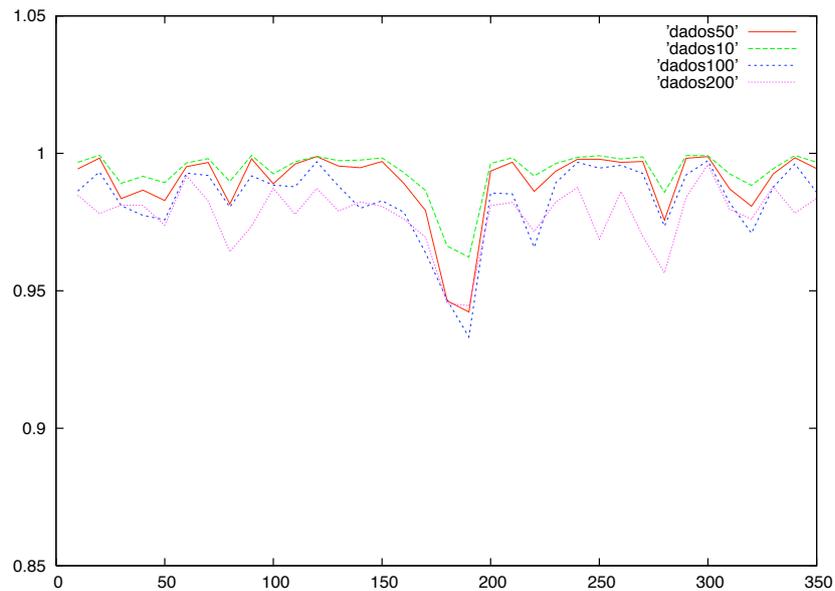


Figura 6.5: Vemos o ângulo ϕ entre o vetor gerado por uma imagem rotacionada por um ângulo α e a imagem original. São apresentados resultados para 10 (verde), 50 (vermelho), 100 (azul) e 200 *features* (roxo).

features não temos mais nenhum ganho significativo.

Abaixo temos ainda um gráfico analisando a projeção entre um vetor associado a uma imagem e um vetor associado à mesma imagem rotacionada de um ângulo α para diversos números de *features*.

Como esperado vemos uma leve alteração na projeção entre os vetores para o aumento do ângulo de rotação até um máximo em 180 graus e a partir daí temos o movimento inverso novamente. Aqui vemos a outra face da moeda do aumento do número de dimensões do problema, vemos que quanto maior o número de *features* menor a projeção entre os vetores. Isto se deve ao fato do ruído inerente a aproximações numéricas ter influência maior na projeção entre dois vetores em espaços de alta dimensão. Logo, indo para espaços de dimensão mais alta temos uma influência cada vez maior de um erro numérico.

6.4 Conclusão

Após uma investigação do método proposto de extração de *features* concluímos que o método proposto, apesar de viável, não apresenta uma melhoria sobre os métodos anteriores propostos na literatura. Mesmo a utilização de métodos mais sofisticados de classificação (que contraria o espírito inicial do projeto), não apresentou melhoria substancial quando usado o método proposto. Procuramos uma implementação de *kernel* inspirado no algoritmo sugerido acima, mas este também não mostrou-se proveitoso. O valor didático do trabalho foi muito grande, permitindo-me o envolvimento com um número de métodos muito aplicados e estudados.

7 Considerações Finais

O intercâmbio entre sociologia, psicologia e física ainda está engatinhando. No entanto, me parece que o recente crescimento da ciência baseada em dados (*data-driven science*) representa um passo na direção de um intercâmbio mais forte entre estes campos. Primordialmente, é importante embasar firmemente estes esforços em dados reais, para que possamos garantir a relevância de modelos sociais e psicológicos. A natureza dos fenômenos psicológicos e sociais é um grande desafio ao paradigma de conhecimento de um físico, e grandes obstáculos ainda encontram-se no caminho, mas já podemos vislumbrar a possibilidade de análises quantitativas baseadas em modelos assistindo políticas públicas.

O que procuramos apresentar neste trabalho é a possibilidade de utilizarmos uma grande variedade de métodos de física estatística para construir e analisar um modelo de um fenômeno sociopsicológico. O modelo sugere que talvez a assinatura estatística encontrada em dados de questionários morais seja devido a um aprendizado moral ocorrendo dentro das sociedades humanas. Comprovar esta afirmação não está no escopo do trabalho, e seria um dos obstáculos mencionados acima, mas ainda que o objetivo continue distante, cada vez mais parece que o caminho está delineado.

Referências Bibliográficas

- 1 DIRAC, P. A. M. Quantum mechanics of many-electron systems. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, The Royal Society, v. 123, n. 792, p. 714, 1929. ISSN 09501207. Disponível em: <<http://www.jstor.org/stable/95222>>.
- 2 ANDERSON, P. W. More Is Different. *Science*, v. 177, n. 4047, p. 393, 1972. Disponível em: <<http://www.sciencemag.org>>.
- 3 HAIDT, J. The new synthesis in moral psychology. *Science*, v. 316, n. 5827, p. 998–1002, May 2007.
- 4 CATICHA, N.; VICENTE, R. Signatures of the neurocognitive basis of culture wars found in moral psychology data. march 2010. Disponível em: <<http://arxiv.org/pdf/1005.5718v1>>.
- 5 DURKHEIM, E. Book. *Suicide : a study in sociology*. [S.l.]: Routledge & Kegan, London, 1952. 404 p. ; p. ISBN 0710033117.
- 6 PINKER, S. Book. *The blank slate : the modern denial of human nature*. [S.l.]: Allen Lane, London, 2002. xvi, 509 p. : p. ISBN 0713992565 0713996722 0713992565 0713992565.
- 7 CASTELLANO, C.; FORTUNATO, S.; LORETO, V. Statistical physics of social dynamics. *Rev. Mod. Phys.*, American Physical Society, v. 81, n. 2, p. 591–646, May 2009.
- 8 VICENTE, R.; MARTINS, A. C. R.; CATICHA, N. Opinion dynamics of learning agents: does seeking consensus lead to disagreement? *Journal of Statistical Mechanics: Theory and Experiment*, v. 2009, n. 03, p. P03015 (14pp), 2009. Disponível em: <<http://stacks.iop.org/1742-5468/2009/P03015>>.
- 9 AMODIO, D. M. et al. Neurocognitive correlates of liberalism and conservatism. *Nat Neurosci*, Nature Publishing Group, v. 10, n. 10, p. 1246–1247, 2007. Disponível em: <<http://dx.doi.org/10.1038/nn1979>>.
- 10 OXLEY, D. R. et al. Political attitudes vary with physiological traits. *Science*, v. 321, n. 5896, september 2008.
- 11 SOMERVILLE, L. H.; HEATHERTON, T. F.; KELLEY, W. M. Anterior cingulate cortex responds differentially to expectancy violation and social rejection. *Nat Neurosci*, v. 9, n. 8, p. 1007–1008, 2006. Disponível em: <<http://dx.doi.org/10.1038/nn1728>>.

- 12 JACOB, F. Evolution and tinkering. *Science*, v. 196, n. 4295, p. 1163, 1977. Disponível em: <<http://www.sciencemag.org>>.
- 13 MASTEN, C. L. et al. Neural correlates of social exclusion during adolescence: understanding the distress of peer rejection. *Social Cognitive and Affective Neuroscience*, v. 4, n. 2, p. 143–157, 2009. Disponível em: <<http://scan.oxfordjournals.org/cgi/content/abstract/4/2/143>>.
- 14 RAINVILLE, P. et al. Pain affect encoded in human anterior cingulate but not somatosensory cortex. *Science*, v. 277, n. 5328, p. 968–971, 1997. Disponível em: <<http://www.sciencemag.org/cgi/content/abstract/277/5328/968>>.
- 15 EISENBERGER, N. I.; LIEBERMAN, M. D.; WILLIAMS, K. D. Does rejection hurt? an fmri study of social exclusion. *Science*, v. 302, n. 5643, p. 290–292, 2003. Disponível em: <<http://www.sciencemag.org/cgi/content/abstract/302/5643/290>>.
- 16 KINOUCI, O.; CATICHA, N. Optimal generalization in perceptions. *Journal of Physics A: Mathematical and General*, v. 25, n. 23, p. 6243–6250, 1992. Disponível em: <<http://stacks.iop.org/0305-4470/25/6243>>.
- 17 WANG, F.; LANDAU, D. P. Determining the density of states for classical statistical models: A random walk algorithm to produce a flat histogram. *Phys. Rev. E*, American Physical Society, v. 64, n. 5, p. 056101, Oct 2001.
- 18 TSAFRIR, D. et al. Sorting points into neighborhoods (spin): data analysis and visualization by ordering distance matrices. *Bioinformatics*, Oxford University Press, Oxford, UK, v. 21, n. 10, p. 2301–2308, 2005. ISSN 1367-4803.
- 19 STROGATZ, S. H.; WATTS, D. J. Collective dynamics of 'small-world' networks. *Nature*, v. 393, n. 6684, p. 440–442, June 1998.
- 20 BARABÁSI, A.-L.; ALBERT, R. Emergence of Scaling in Random Networks. *Science*, v. 286, n. 5439, p. 509–512, 1999. Disponível em: <<http://www.sciencemag.org/cgi/content/abstract/286/5439/509>>.
- 21 SHERRINGTON, D.; KIRKPATRICK, S. Solvable model of a spin-glass. *Phys. Rev. Lett.*, American Physical Society, v. 35, n. 26, p. 1792–1796, Dec 1975.
- 22 BINNEY N. J. DOWRICK, A. J. F. J. J.; NEWMAN, M. E. J. *The Theory of Critical Phenomena: An Introduction to the Renormalization Group*. [S.l.]: Oxford University Press, 1992.
- 23 METROPOLIS, N. et al. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, AIP, v. 21, n. 6, p. 1087–1092, 1953. Disponível em: <<http://link.aip.org/link/?JCP/21/1087/1>>.
- 24 ZHOU, C.; BHATT, R. N. Understanding and improving the wang-landau algorithm. *Phys. Rev. E*, American Physical Society, v. 72, n. 2, p. 025701, Aug 2005.

- 25 CUNHA-NETTO, A. G. et al. Improving wang-landau sampling with adaptive windows. *Phys. Rev. E*, American Physical Society, v. 78, n. 5, p. 055701, Nov 2008.
- 26 OPPER, M.; SAAD, D. M. *Advanced Mean Field Methods: Theory and Practice*. [S.l.]: MIT Press, 2001.
- 27 GLAUBER, R. J. Time-dependent statistics of the ising model. *Journal of Mathematical Physics*, AIP, v. 4, n. 2, p. 294–307, 1963. Disponível em: <<http://link.aip.org/link/?JMP/4/294/1>>.
- 28 BURKHARDT, H.; SIGGELKOW, S. Invariant features in pattern recognition - fundamentals and applications. In: _____. *Nonlinear Model-Based Image/Video Processing and Analysis*. [S.l.]: John Wiley and Sons, 2001. p. 269–307.
- 29 FLUSSER, J.; SUK, T. Pattern recognition by affine moment invariants. *Pattern Recognition*, v. 26, n. 1, p. 167 – 174, 1993. ISSN 0031-3203.
- 30 PERANTONIS, S.; LISBOA, P. Translation, rotation, and scale invariant pattern recognition by high-order neural networks and moment classifiers. *Neural Networks, IEEE Transactions on*, v. 3, n. 2, p. 241–251, Mar 1992. ISSN 1045-9227.
- 31 WOOD, J. Invariant pattern recognition: A review. *Pattern Recognition*, v. 29, n. 1, p. 1 – 17, 1996. ISSN 0031-3203.
- 32 HU, M.-K. Visual pattern recognition by moment invariants. *Information Theory, IRE Transactions on*, v. 8, n. 2, p. 179–187, February 1962. ISSN 0096-1000.
- 33 HAASDONK, B.; BURKHARDT, H. Invariant kernel functions for pattern analysis and machine learning. *Mach. Learn.*, Kluwer Academic Publishers, Hingham, MA, USA, v. 68, n. 1, p. 35–61, 2007. ISSN 0885-6125.
- 34 SCHULZ-MIRBACH, H. On the existence of complete invariant feature spaces in pattern recognition. *Pattern Recognition, 1992. Vol.II. Conference B: Pattern Recognition Methodology and Systems, Proceedings., 11th IAPR International Conference on*, p. 178–182, Aug-3 Sep 1992.
- 35 NABUTOVSKY, D.; DOMANY, E. Learning the unlearnable. *Neural Comput.*, MIT Press, Cambridge, MA, USA, v. 3, n. 4, p. 604–616, 1991. ISSN 0899-7667.
- 36 BLATT, M.; WISEMAN, S.; DOMANY, E. Superparamagnetic clustering of data. *Phys. Rev. Lett.*, American Physical Society, v. 76, n. 18, p. 3251–3254, Apr 1996.
- 37 KOHONEN, T. The handbook of brain theory and neural networks. In: _____. [S.l.]: MITPress, 1995. p. 537–540.
- 38 BIEHL, M.; GHOSH, A.; HAMMER, B. Dynamics and generalization ability of lvq algorithms. *J. Mach. Learn. Res.*, MIT Press, Cambridge, MA, USA, v. 8, p. 323–360, 2007. ISSN 1533-7928.

- 39 ANDERS, E. B. An extension of romberg integration procedures to n-variables. *J. ACM*, v. 13, n. 4, p. 505–510, 1966.
- 40 MESSAGE-PASSING Interface. <http://www-unix.mcs.anl.gov/mpi/>.
- 41 AMDAHL, G. M. Validity of the single processor approach te achieving large scale computing capabilities. In: *Proc. AFIPS Spring Joint Computer Conf.* [S.l.: s.n.], 1967. v. 30, p. 483–485.
- 42 WALDER, C.; CHAPELLE, O. Learning with transformation invariant kernels. In: PLATT, J. et al. (Ed.). *Advances in Neural Information Processing Systems 20*. Cambridge, MA: MIT Press, 2008. p. 1561–1568.
- 43 BISHOP, C. M. *Pattern Recognition and Machine Learning*. [S.l.]: Springer, 2006. 124ff p.
- 44 HAMMER, B.; VILLMANN, T. Generalized relevance learning vector quantization. *Neural Networks*, Elsevier, v. 15, p. 1059–1068, 2002.