UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS
DEPARTAMENTO DE FÍSICA E INFORMÁTICA

"ANÁLISE ESTATÍSTICA DO PROBLEMA DA PARTIÇÃO NUMÉRICA"

FERNANDO FAGUNDES FERREIRA

Tese apresentada ao Instituto de Física de São Carlos, Universidade de São Paulo, para obtenção do título de Doutor em Ciências "Física Básica"

USP/IFSC/SBI
8-2-001374

Orientador:
Prof. Dr. José Fernando Fontanari

Comissão Julgadora:
Prof. Dr. José Fernando Fontanari (IFSC-USP)
Prof. Dr. Roberto Nicolau Onody (IFSC-USP)
Prof. Dr. Roland Koberle (IFSC-USP)
Prof. Dr. Paulo Murilo de Castro Oliveira (UFF)
Prof. Dr. Domingos Humberto Urbano Marchetti (IF-USP)

São Carlos – São Paulo
2001

IFSC-USP SERVIÇO DE BIBLIOTECA E INFORMAÇÃO

Ferreira, Fernando Fagundes

"Análise estatística do problema da partição numérica"/ Fernando Fagundes Ferreira — São Carlos, 2001

Tese (Doutorado) — Instituto de Física de São Carlos da Universidade de São Paulo, 2001 páginas. 190
Área: Física Básica
Prof. Dr. José Fernando Fontanari

1. Otimização combinatorial, 2. Modelo de Ising, 3. Problema da partição numérica

I. Título

UNIVERSIDADE DE SÃO PAULO
Instituto de Física de São Carlos

Caixa Postal 369
13560-970 São Carlos, SP
Av. Trabalhador São-carlense, 400,
13566-590 São Carlos, SP

Fone/Fax 16 273 9777
www.if.sc.usp.br
wladerez@if.sc.usp.br

MEMBROS DA COMISSÃO JULGADORA DA TESE DE DOUTORADO DE FERNANDO FAGUNDES FERREIRA APRESENTADA AO INSTITUTO DE FÍSICA DE SÃO CARLOS, DA UNIVERSIDADE DE SÃO PAULO, EM 08 DE MARÇO DE 2001.

COMISSÃO JULGADORA:

Prof. Dr. José Fernando Fontanari/IFSC-USP

Prof. Dr. Roberto Nicolau Onody/IFSC-USP

Prof. Dr. Roland Koberle/IFSC-USP

Prof. Dr. Paulo Murilo de Castro Oliveira/UFF

Prof. Dr. Domingos Humberto Urbano Marchetti/IF-USP

IFSC-USP
SERVIÇO DE BIBLIOTECA E INFORMAÇÃO

Dedicatória

Com muita saudade, dedico esta tese a duas pessoas maravilhosas que foram muito importantes em minha vida, embora não estejam mais presentes, a lembrança dos bons momentos que tivemos junto e nossos sonhos permanecem em minha memória. Ao meu pai *Elpídio Aurelio Ferreira* (*in memoriam*) e à minha tia *Irani Ferreira* (*in memoriam*).

Agradecimentos

Tenho que agradecer a muita gente por toda a ajuda recebida tanto profissionalmente, como também emocional e socialmente. Foram muitos os momentos difíceis ao longo desses quatro anos, e contei com o apoio e com a ajuda de inúmeros amigos, que muito me orgulham e, portanto, presto aqui uma singela homenagem. Corro o risco de não citar pessoas amigas que marcaram esse período, mas espero que elas possam certamente compreender [lápsos de memória :-)].

- Seguindo a justa tradição, agradeço ao professor José Fernando Fontanari por inúmeros motivos: pela dedicação, pela postura e pela conduta exemplar, marcado pelo profissionalismo competente, sempre presente, pelo belíssimo tema, pelas condições de trabalho, pelo apoio. Acredito piamente ter aprendido muito durante esses quatro anos de doutoramento, e considero o professor Fontanari como um modelo a ser seguido.

- Agradeço de coração aos amigos de grupo, por terem me aturado durante tanto tempo, por terem proporcionado um excelente ambiente de trabalho e pesquisa combinando na medida certa amizade, profissionalismo e bom humor. Pela assistência computacional, pela promessa de esquecerem dos problemas causados pela BIOS. Vocês formam, indubitavelmente, um elenco muito especial, daí a minha admiração e meu carinho por todos vocês: Domingos Alves, Cláudia Pio, Paulo R. Campos, Viviane Moraes, Ana Teresa, Milton Sonoda, Fernando Crepaldi, Alexandre Colato e à minha versão mais moderna o Leonardo (vamos esperar para ver os ' bugs') :-).

- Agradeço aos amigos mais antigos que continuam presentes até hoje, sempre dedicados, carinhosos e me dando muito orgulho: Marcelo Ianela, Jose Alves Junior, Ricardo Pinto, Sandro Pinto, Marcelo Nakaema, Christian Petru.

- Agradeço também e com o mesmo ímpeto aos novos amigos, que tive o privilégio de conhecer durante a fase de doutorado: Fabrizio Mendes, Antonio Nojoso, Sidineco, Fabinho, Carlinho Bento, Carlão Boca, José Ricardo Mendonça, Adriano, Marcio, os irmãos T-sung.

- Agradeço aos meus amigos do IFSC, sem dúvida alguma pessoas muito queridas: Marcelinho (obrigado pela ajuda na conversão das figuras), Marcel (valeu pelas observações), Vivaldo, Nazareno, Paulo, Marcio Magini, Zé Paulo, Jabá, Fernandão, Leonardo, Cleber, Rafael, Rodrigo, Luís Araxá, Antonio Carlos, Nório, Felix, Fabrício, Reginaldo, João Vitor, Marletta, Max, Vander, Marcel, Vanius, André Muezeri.

- Aos amigos fundadores do projeto Cursinho Popular Palmares (sediado no Caaso) que muito contribuiram para a realização de um sonho: Dani e Sabinho, Márcio, Osvaldo (Batata), Junior, André e Andrezão, Hiroshi, Dinha, Marcos, José Luis, enfim, a todos os professores voluntários e alunos.
- à minha amada, doce e maravilhosa mãe Maria Ester Fagundes
- à minha querida namorada Elizangela
- Aos funcionários do IFSC, em especial à Cristina, à Wladerez, à Nilzeli, à Neusa e à Mara pela notória eficiência e pela amizade.
- Aos professores do IFSC que contribuiram para a minha formação, em especial ao professor Reginaldo Napolitano, pela confiança, pela amizade.
- Por fim, agradeço ao povo brasileiro, à **FAPESP** e ao **CNPq** pelo suporte financeiro.

Resumo

Nesta tese apresentamos a abordagem da Mecânica Estatística para o clássico problema de otimização denominado *problema da partição numérica* (**PPN**), que é definido como: Dada uma seqüência de N números reais positivos $\{a_1, a_2, ..., a_N\}$, o problema consiste em particioná-los em dois conjuntos complementares, A e A^c, tais que o valor absoluto da diferença da soma dos $a_j's$ nos dois conjuntos seja minimizada. No caso em que os a_j's são variáveis aleatórias estatisticamente independentes distribuídas uniformemente no intervalo unitário, este problema $\mathcal{NP}$-completo equivale ao problema de encontrar o estado fundamental de um modelo de Ising antiferromagnetico aleatório de alcance infinito. Conseqüentemente, a análise probabilística do **PPN** pode ser realizada com as ferramentas da Mecânica Estatística de sistemas desordenados. Neste trabalho empregamos a aproximação recozida ('annealed') para derivar uma expressão analítica para o limitante inferior do valor médio da diferença para partições tanto com vínculo de cardinalidade quanto sem vínculo para grandes valores de N. Além disso, calculamos analiticamente a fração de estados metaestáveis, isto é, estados que possuem a menor energia mediante todos os vizinhos (estados que diferem pela troca de um único spin). Concluimos a análise da abordagem direta, cujas instâncias $\{a_i\}$ são fixas e as partições são variáveis, com o estudo analítico da versão relaxada deste problema de programação inteira $\mathcal{NP}$-completo.

Na segunda parte desta tese propomos e exploramos uma abordagem inversa para o **PPN**, no qual as partições ótimas são fixas e as instâncias são variáveis. Especificamente, usando o método das réplicas estudamos analiticamente o espaço de configuração do problema da partição numérica. Mostramos que, independentemente da distribuição de entradas das instâncias, existe um limitante superior $\alpha_c N$ para o número de partições perfeitas aleatórias (isto é, partições com diferença igual a zero). Em particular, no caso em que os dois conjuntos têm a mesma cardinalidade (partições balanceadas) encontramos $\alpha_c = \frac{1}{2}$. Mais ainda, no caso de partições não balanceadas, mostramos que as partições perfeitas aleatórias existem somente se a diferença entre as cardinalidades dos dois conjuntos escala com $N^{1/2}$.

Abstract

In this thesis we present a statistical mechanics approach to a classical optimization problem called the *number partitioning problem* (**NPP**), which is stated as follows. Given a sequence of N positive real numbers $\{a_1, a_2, ..., a_N\}$, the number partitioning problem consists of partitioning them into two sets A and its complementary set A^c such that the absolute value of the difference of the sums of a_j over the two sets is minimized. In the case in which the a_j's are statistically independent random variables uniformly distributed in the unit interval, this $\mathcal{NP}$-complete problem is equivalent to the problem of finding the ground state of an infinite range, random antiferromagnetic Ising model. Hence the probabilistic analysis of the **NPP** can be carried out within the framework of the standard statistical mechanics of disordered systems. In this vein we employ the annealed approximation to derive analytical lower bounds to the average value of the difference for the best constrained and unconstrained partitions in the large N limit. Furthermore, we calculate analytically the fraction of metastable states, i.e. states that are stable against all single spin flips. We conclude the analysis of the so-called direct approach, in which the instances $\{a_i\}$ are fixed and the partitions are variable, with the analytical study of the linear programming relaxation of this $\mathcal{NP}$-complete integer programming. In the second part of this thesis we propose and explore an inverse approach to the **NPP**, in which the optimal partitions are fixed and the instances are variable. Specifically, using the replica framework we study analitically the instance space of the number partitioning problem. We show that, regardless of the distribution of the instance entries, there is an upper bound $\alpha_c N$ to the number of perfect random partitions (i.e. partitions for which that difference is zero). In particular, in the case where the two sets have the same cardinality (balanced partitions) we find $\alpha_c = \frac{1}{2}$. Moreover, in the case of unbalanced partitions, we show that perfect random partitions exist only if the difference between the cardinalities of the two sets scales like $mN^{1/2}$.

Conteúdo

1 Introdução **2**

2 A Mecânica Estatística do problema da Partição Numérica **18**

2.1 Aproximação 'Annealed' 20

2.1.1 O caso sem vínculo 21

2.1.2 O caso com vínculo 25

2.1.3 Energia Ótima E_m 29

2.2 Número médio de mínimos locais 33

3 Análise Estatística do Problema da Partição Numérica Relaxado **38**

3.1 Algoritmo de Relaxação 39

3.2 Cálculo da função $\hat{Z}_u$ 41

3.3 Cálculo da função $\hat{Z}_m$ 42

3.4 Energia média 46

3.5 Distribuição de probabilidade das cardinalidades 47

3.5.1 Momentos 50

3.6 Correlação 51

3.7 Cálculo da energia de integralização 56

4 O Espaço das Realizações do Problema da Partição Numérica **60**

4.1 Energia Livre 62

4.1.1 Energia livre com simetria de réplicas 67

4.2 Distribuição de probabilidade das entradas 76

5 Conclusão **82**

Apêndice **87**

A Complexidade Computacional **87**

A.0.1 Motivação 87

A.0.2 A medida da complexidade 88

A.0.3 Problemas tratáveis e intratáveis 91

A.1 Classes de Complexidade 96

A.1.1 Problemas de Decisão 96

A.1.2 As classes $\mathcal{P}$ e $\mathcal{NP}$ 103

A.1.3 $\mathcal{NP}$-completeza 106

A.1.4 Além de $\mathcal{NP}$ 111

B O cálculo exato **114**

C Heurísticas para a Partição Numérica **120**

C.0.5 A heurística do avarento 120

C.0.6 A heurística de Karmarkar & Karp 121

C.0.7 O algoritmo de Karmarkar & Karp Completo 123

D Análise de Estabilidade **127**

Bibliografia **136**

Lista de Figuras

2.1 Energia ótima média obtida através de busca exaustiva como função de $N^{1/2}$ 24

2.2 Razão entre o desvio padrão e o valor da energia ótima média obtida através de busca exaustiva como função de N 24

2.3 Energia ótima obtida através de busca exaustiva como função de N para partições pares ($m = 0$). 29

2.4 Razão entre o desvio padrão e o valor da energia ótima média obtida através de busca exaustiva como função de N para partições pares ($m = 0$). 30

2.5 Energia ótima média obtida através de busca exaustiva como função de m para $N = 24$. A curva contínua no gráfico principal é a estimativa teórica (2.66) enquanto a curva no gráfico menor é o limite inferior na aproximação 'annealed' (2.50). 32

3.1 A linha vermelha é o resultado do cálculo exato do problema inteiro derivado no capítulo anterior. A linha verde é o resultado teórico do problema relaxado. Os pontos foram obtidos da simulação utilizando o algoritmo da seção 1. 47

3.2 Distribuição do mínimos globais como função do vínculo de cardinalidade m para N fixo. Os pontos foram obtidos através de simulação. 49

3.3 (a)A média do módulo da cardinalidade m dos mínimos globais em função de N. (b) A razão entre o desvio padrão e a média da grandeza m. As linhas contínuas são estimativas teóricas. Os pontos em vermelho são o resultado dos experimentos computacionais via algoritmo da relaxação. Os pontos azuis foram obtidos aplicando-se o algoritmo da diferença de Karmakar e Karp (o algoritmo apresentado no apêndice C). Os pontos verdes são o resultado da busca exaustiva para o problema inteiro. . . . 51

3.4 A curva continua mede a correlação entre dois estados s_i e s_j com energia mínima (ótimo global) com a mesma cardinalidade (M fixo) obtida analiticamente. Os pontos resultam da aplicação do algoritmo da relaxação para vários valores de N. 56

3.5 Energia minima obtida com a heurística da integralização em função de N. A linha cheia é o resultado teórico e os pontos foram obtidos via simulações. 59

4.1 Limitante superior para o número de partições perfeitas aleatórias α_c como função do parâmetro m. Para partições balanceada ($m = 0$) encontramos $\alpha_c = \frac{1}{2}$. 73

4.2 A densidade de energia ε normalizada por Q em função de α. A curva em verde é o resultado para $m = 0$ equanto a curva preta é um resultado assintótico (obtida para grandes valores de m). As curvas intermediárias correspondem aos valores: $m = 2, 4, 7, 10, 20$, de baixo para cima. 75

4.3 O parâmetro q contra α obtido resolvendo-se numericamente as equações de ponto de sela. Da direita para a esquerda as linhas correspondem aos seguintes valores de m: 0, 0.5, 1, 2, 5, 10 76

4.4 Distribuição de probabilidade acumulada de entradas de realizações com energia zero para partições balanceadas para $\alpha = 0.01$, 0.25, 0.36, 0.42, 0.48, 0.49 *e* 0.499, vistos de baixo para cima. A curva em vermelho corresponde à distribuição acumulada quando $\alpha = \alpha_c$. 81

A.1 Um grafo valorado e sua árvore geradora mínima (arestas coloridas). 92

A.2 Mesma instância, diferentes problemas: Uma configuração válida do PROBLEMA DO CAIXEIRO VIAJANTE (esquerda) e o problema DESIGNAÇÃO (direita). Enquanto DESIGNAÇÃO pode ser resolvido em tempo polinomial, o PCV é intratável. 95

A.3 As sete pontes de of Königsberg, desenhada como no artigo de Euler de 1736 [88] (esquerda) e representado como um grafo (direita). No grafo, a terra e as ilhas são condensadas em pontos (vértices) e cada ponte é desenhada como uma linha (arestas). 97

A.4 O Jogo do Icosaedro de Hamilton: Encontre uma rota ao longo das arestas de um dodecaedro (esquerda), passando em cada vértice exatamente uma vez e retornando ao ponto inicial. Uma solução está indicada no grafo planar (arestas sombreadas) que é isomorfo ao dodecaedro (direita). 100

A.5 Exemplo da execução da história de um algoritmo não determinístico. 103

A.6 Um primeiro mapa para a complexidade. Todos problems indicados estão definidos dentro do texto. Problemas com um (D) são as versões decisão dos problemas de otimização. 106

A.7 O mapa da complexidade revisado. 109

A.8 Três tentativas de mapas $\mathcal{NP}$. É muito provável (mas não com certeza) que (a) seja o mapa correto. 111

B.1 Contribuição térmica da entropia $\tilde{S}$ contra $\frac{\Delta a}{T}$ 118

C.1 Árvore representativa da construção da melhor partição dada pela heurística KK . . . 122

C.2 Logaritmo da função custo do PPN encontrado pela heurística KK em função do quadrado do logaritmo do tamanho do problema N. A linha contínua foi obtida por regressão linear e a inclinação fornece o valor da constante $\alpha = 0.07$. 123

C.3 Árvore gerada pelo algoritmo KKC quando aplicado a uma lista formada pelos números 8,7,6,5,4. Ramos esquerdos: troca-se os dois maiores elementos pela diferença entre eles. Ramos direitos: troca-se os dois maiores elementos pela sua soma. As linhas pontilhadas representam os nós cortados pelo algoritmo durante a busca. 125

C.4 C é a função custo (diferença) mínima e N é o tamanho do problema. Todos os pontos foram obtidos fazendo-se a média sobre 1000 experimentos independentes. Os valores inteiros a_i a serem particionados foram gerados aleatoriamente usando uma distribuição uniforme no intervalo $(0, 10^7]$. As linhas vermelha, verde e preta correspondem à heurística do avarento, à heurística KK e ao algoritmo KKC, respectivamente. 126

C.5 O número médio de nós gerados durante a busca da solução ótima pelo KKC como função do tamanho do problema. Os parâmetros do experimento são os mesmos da figura anterior. 126

D.1 O valor da grandeza W em função do parâmetro q na região abaixo de α_c, para vários valores da variável m. 135

Lista de Tabelas

A.1 A ordem de complexidade versus o tempo de processamento. A unidade, quando não especificada está em segundos. 90

Capítulo 1

Introdução

Esta tese apresenta uma abordagem Física de um problema oriundo da Ciência da Computação teórica chamado de *Problema da Partição Numérica* (**PPN**). Primeiramente vamos contextualizar este problema segundo uma perspectiva que o torna relevante, apesar de corrermos o risco de nos afastar do viés físico característico deste trabalho. Por se tratar de um problema notadamente interdisciplinar, vamos iniciar nosso estudo com uma digressão sobre a teoria da complexidade computacional: essa perspectiva vai nos dar uma idéia da existência de um mundo muito diverso do qual estamos acostumados em Física, mas que possui com esta uma interface significativa que vem sendo explorada nos últimos anos.

A teoria da Complexidade Computacional dedica-se à análise e à classificação de algoritmos associados a problemas de otimização. Algoritmos são procedimentos passo-a-passo para resolver um problema. De forma concreta, podemos pensá-lo como sendo simplesmente o programa de computação escrito em uma linguagem de programação específica. Dizemos que um algoritmo resolve um problema Π se ele puder ser aplicado a qualquer realização de Π garantindo produzir sempre uma solução. Em geral, estamos interessados em encontrar algoritmos 'eficientes' para resolver o problema. No sentido mais amplo, a noção de eficiência envolve vários aspectos dos recursos computacionais disponíveis. Normalmente o algoritmo mais eficiente significa simplesmente o mais rápido. A medida mais rigorosa da eficiência é feita por meio da *função complexidade temporal* que depende somente do tamanho do problema.

Há uma diversidade enorme de algoritmos que diferem pelas suas funções complexidade temporal, e

a caracterização de quais deles são eficientes o bastante ou quase ineficientes dependeria a princípio dos recursos computacionais. Contudo, os cientistas da computação formularam um critério robusto para distingüí -los e fornecer uma resposta para essa questão. Foi introduzida a noção de algoritmo que processa em um tempo polinomial e outros que gastam um tempo exponencial. Um algoritmo polinomial é aquele cuja função de complexidade é da ordem $\mathcal{O}(p(n))$, onde $p(n)$ é um polinômio de grau n (tamanho do problema). Um algoritmo cuja função complexidade temporal não pode ser limitada por um polinômio é dito ser um algoritmo de complexidade temporal exponencial. Essa noção desempenha um papel crucial na classificação dos problemas e independe da capacidade de processamento das máquinas ou dos detalhes de implementação do programa. Essa distinção leva ao surgimento de duas famílias de problemas, os denominados tratáveis, para os quais se conhece um algoritmo polinomial, e os denominados intratáveis que são tão difíceis que nenhum algoritmo polinomial pode resolvê-los.

Essas idéias já foram discutidas em 1964 por Cobham [19] e por Edmonds [20]. O tempo de complexidade é definido como a medida do *pior* caso, de forma que, por exemplo, um algoritmo com tempo de complexidade 2^n significa somente que pelo menos uma instância de tamanho n pode requerer tanto tempo para ser resolvido. A solução da maioria das instâncias realmente requer muito menos tempo. Como exemplo citamos o algoritmo simplex desenvolvido para resolver problemas de programação linear que foi mostrado ter complexidade exponencial [Klee e Minty, 1972], mas que na prática funciona impressionantemente rápido. Este é também o caso do algoritmo Branch-and-Bound quando aplicado ao problema da Mochila [6, 7]. Infelizmente, exemplos como esses são raros.

Problemas NP-completos

A busca de métodos para investigar os problemas intratáveis e para entender a relação entre eles levou à descoberta de uma família de problemas especiais. Em 1960 Dantzig [21] reduziu vários problemas de otimização para um problema geral de programação inteira do tipo zero-um. Edmonds [22], em

1962, estudando problemas de teoria dos grafos como o problema da cobertura de vértices e o problema de encontrar o máximo conjunto independente de vértices, consegui reduzí-los a um problema mais geral: problema do conjunto de cobertura. Muitos outros trabalhos isolados que seguiram este caminho resultaram na teoria da $\mathcal{NP}$-completeza. A fundamentação teórica foi lançada no artigo de Stephen Cook [23] em 1971, que contém muitos resultados importantes. Primeiro, Cook enfatizou a importância da *redução polinomial*. Se conhecermos a redução polinomial de um problema para outro, isto assegura que qualquer algoritmo polinomial para o segundo problema pode ser convertido em um algoritmo também polinomial para o primeiro. Segundo, ele chamou a atenção para uma classe de problemas de decisão que poderiam ser resolvidos em tempo polinomial por um algoritmo não-determinístico (apêndice A) que foi denominada de classe $\mathcal{NP}$ (não-determinístico polinomial). Um problema de decisão é aquele cuja solução é sim ou não. Todos os problemas intratáveis quando enunciados como problemas de decisão pertencem a essa classe. Terceiro, ele provou que um particular problema em $\mathcal{NP}$ conhecido como o problema da satisfabilidade tem a propriedade de que todos os problemas em $\mathcal{NP}$ podem ser reduzidos polinomialmente para ele. Se o problema da satisfabilidade for resolvido em tempo polinomial então todos os problemas na classe $\mathcal{NP}$ também poderão ser resolvidos por um algoritmo polinomial, acabando assim com a noção de intratabilidade. O problema da satisfabilidade passou então a ser o problema mais difícil nessa classe.

Subseqüentemente, Richard Karp [24] apresentou em 1972 uma coleção de resultados provando que existem outros problemas de decisão, como o problema do caixeiro viajante, que são tão difíceis quanto o problema da satisfabilidade e dotados da mesma propriedade quanto à redutibilidade. Esses resultados passaram a pertencer a uma nova classe que recebeu o nome de classe $\mathcal{NP}$-completa. A prova de que $\mathcal{P} = \mathcal{NP}$ ou $\mathcal{P} \neq \mathcal{NP}$ está entre os mais fascinantes e desafiadores problemas em aberto da atualidade. Uma descrição introdutória e detalhada dos tópicos discutidos acima pode ser encontrada no apêndice A.

Problemas de otimização

A otimização preocupa-se com a computação e a caracterização dos mínimos (ou máximos) globais de funções lineares ou não-lineares. Graças ao interesse atual na modelagem matemática de sistemas do mundo real, problemas de otimizaç ão global estão sendo usados em uma gama cada vez maior de aplica ções. Tais aplicações incluem economias de escala, finanças, problemas de alocação e de localização, pesquisa operacional, estatística, problemas de redes e de transporte, problemas de bancos de dados, além de muitos outros problemas que aparecem em muitas áreas científicas e tecnológicas como na engenharia química e na biologia molecular, para mencionar apenas algumas. Com o desenvolvimento e a implementação de algoritmos de otimização genéricos, cada vez mais e mais cientistas em diversas disciplinas têm usado essas técnicas para resolver seus problemas específicos. A otimização está desempenhando um papel muito importante na ciência e na indústria, daí a necessidade de compreendermos todos os aspectos possíveis envolvendo a dificuldade de elaboração de novas ferramentas práticas (algoritmos) e teóricas (matemática) de otimização que sejam mais eficientes para resolver os problemas ainda em aberto.

Os problemas de otimização parecem se dividir naturalmente em duas categorias: os formulados em termos de variáveis contínuas e aqueles formulados com variáveis discretas, que recebem o nome de problemas combinatoriais. Os problemas contínuos envolvem geralmente a busca por um conjunto de números reais ou até mesmo uma função, enquanto os problemas combinatoriais envolvem a busca de objetos que sejam finitos ou enumeráveis, tipicamente um conjunto de números inteiros ou ainda um grafo. Esses dois tipos de problemas têm características bastante diferentes e os métodos para resolvê-los têm se tornado cada vez mais divergentes. Um problema de programação inteira ou problema de otimização combinatorial pode ser enunciado da seguinte forma:

$$\text{minimizar} \qquad E(x) = \sum_i a_i x_i \tag{1.1}$$

sujeito a

$$g_k(x) = \sum_{j=1}^{m} c_{ij}^k x_i \leq b_k \tag{1.2}$$

$$x_i \geq 0. \tag{1.3}$$

A função $E(x)$ é a chamada função custo ou função objetiva a ser otimizada enquanto g representa um conjunto de equações que devem satisfazer um vínculo de capacidade b_k do problema. A partir das restrições (1.2) e (1.3) obtemos um conjunto de soluções factíveis $\{x_i\} \in Z$, das quais uma ou mais minimizarão a função custo. Para estudar esses problemas vamos emprestar da Mecânica Estatística um conjunto de idéias e técnicas descritas sucintamente nos parágrafos a seguir.

A Mecânica Estatística

O objetivo da Mecânica Estatística é predizer as propriedades dos sistemas compostos por um número muito grande de partículas em termos de suas propriedades mecânicas individuais e das forças entre elas. Neste momento já podemos perguntar quão grande é 'muito grande'. Um grama de matéria consiste em aproximadamente 10^{23} átomos. O estado de um sistema é especificado classicamente pela posição, velocidade e alguns graus de liberdade internos de cada um desses átomos. As equações que descrevem a evolução temporal desses microestados, em princípio, são conhecidas, mas é completamente insano ter esperanças de resolvê-las para as 10^{23} partículas. Da experiência do dia-a-dia sabemos que a temperatura e a pressão de um gás num recipiente isolado não mudam no tempo, muito embora a posição e a velocidade de todos os átomos mudem constantemente. Conseqüentemente, as propriedades macroscópicas do sistema não são sensíveis ao seu estado microscópico particular. Esta é a variante física da lei dos grandes números e constitui o ponto de partida da Mecânica Estatística. Agora, ao invés de determinarmos o valor exato de uma quantidade macroscópica em um sistema singular, calculamos o seu valor médio, sendo a média tomado sobre um conjunto adequado de sistemas similares (ensemble canônico). Note que isto funciona apenas para um grande número de partículas. O resultado da Mecânica Estatística é válido somente para $N \to \infty$, chamado limite termodinâmico. O uso da letra maiúscula N é proposital e certamente nos fará lembrar deste fato.

Com o uso das ferramentas e idéias desenvolvidas e consolidadas na Mecânica Estatística podemos calcular várias grandezas de interesse em muitas áreas em que esse tipo de abordagem possa ser aplicada. Existe um procedimento padrão para quem quer realizar um cálculo de grandezas termodinâmicas médias. Primeiro calculamos a função de partição, definida como

$$Z = \sum_{\{s\}} e^{-\beta H(s)}, \tag{1.4}$$

onde s denota os microestados, $\beta = \frac{1}{T}$ é o inverso da temperatura (fizemos a constante de Boltzmann $k_B = 1$) e H é a Hamiltoniana do sistema, que leva em conta de forma explícita a interação entre as partículas do sistema. A seguir calculamos uma grandeza muito importante que é a chamada energia livre, definida como

$$F = -T \ln Z. \tag{1.5}$$

da qual podemos derivar todas as demais relações termodinâmicas [44,45].

Vidros de spin

Os vidros de spin estão entre os mais fascinantes tópicos da Mecânica Estatística desenvolvidos essencialmente a partir da segunda metade da década de 70. O desenvolvimento desse ramo de pesquisa baseia-se em três diferentes abordagens: teoria, experimento e simulação computacional. Vidros de spin são uma coleção de spins, ou seja, momentos magnéticos, cujos estados a baixa temperatura estão congelados e desordenados ao invés de seguirem algum tipo de padrão uniforme ou periódico encontrados em magnetos convencionais. Para produzir estes estados dois ingredientes são necessários:

(i) deve existir uma competição das diferentes interações entre os momentos magnéticos, no sentido de que nenhuma configuração simples de spins seja unicamente favorecida por todas as interações.

(ii) Essas interações devem ser, pelo menos parcialmente, aleatórias.

Esses fatos sugerem que o estado da matéria do tipo vidros de spin é intrinsicamente diferente de outras formas convencionais e requer novos conceitos para descrevê-lo.

Experimentalmente não parece ser muito difícil encontrar vidros de spin. O primeiro tipo de sistema estudado foi a solução de impurezas constituídas de metais de transição magnéticos diluídos em metais nobres (hospedeiro). Devido à localização aleatória das impurezas algumas interações são positivas (favorecendo o alinhamento paralelo dos momentos) enquanto outras são negativas (favorecendo o alinhamento anti-paralelo dos momentos). Não é difícil imaginar que teremos competição entre as interações que por sua vez estão distribuídas de forma aleatória (frustração e desordem). Os vidros de spin também foram encontrados em isolantes magnéticos e anéis magnéticos amorfos, onde a dependência da interação com a distância é completamente diferente daquela observada nos sistemas metálicos cristalinos.

Esses estados congelados desordenados sugerem a existência de estados com magnetização espontânea local $m_i = \langle s_i \rangle$ não nula, embora, por outro lado, a magnetização total média $M = \frac{1}{N}\sum_i m_i$ seja nula. A fim de estudar as mais diversas características fenomenológicas dos vidros de spin, precisamos lançar mão de modelos que capturem a física essencial desses sistemas.

Modelos

As características dos vidros de spin são prontamente visualizadas em um modelo Hamiltoniano simples apropriado para uma idealização das interações magnéticas entre átomos com momentos magnéticos bem definidos. O modelo mais simples já proposto foi o modelo de Ising, cujo Hamiltoniano é dado por

$$H = -\sum_{i,j} J_{ij} s_i s_j, \tag{1.6}$$

onde $s_i = \pm 1$ e J_{ij} mede a intensidade da interação entre os pares de spin (ij). As variáveis deste modelo são os $\{s_i\}$, pois os $\{J_{ij}\}$ são vínculos congelados. Em experimentos convencionais, a localização dos spins $\vec{R}_i$ é aleatória e J_{ij} é uma função da distância $\left|\vec{R}_i - \vec{R}_j\right|$. O modelo de Edwards e Anderson [3] supõe que os J_{ij} obedeçam a seguinte função densidade de probabilidade (i.i.d.)

$$P(J_{ij}) = \left(\frac{1}{2\pi\Delta_{ij}}\right)^{\frac{1}{2}} \exp\left(-\frac{J_{ij}^2}{2\Delta_{ij}}\right), \tag{1.7}$$

com $\Delta_{ij} = \left|\vec{R}_i - \vec{R}_j\right|$. Esse é um modelo com interações de curto alcance e com invariância translacional. No modelo de Sherrington e Kirkpatrick (SK) [57], os J_{ij} são parâmetros congelados escolhidos aleatoriamente segundo uma distribuição de probabilidade, de forma que os primeiros momentos sejam:

$$\langle J_{ij}\rangle = \frac{1}{N}J_0 \qquad e \qquad \langle J_{ij}^2\rangle = \frac{1}{N}J^2. \tag{1.8}$$

Este é um modelo de alcance infinito pois J_{ij} não depende da distância. Usualmente a distribuição é a Gaussiana com média $J_0 = 0$ e variância $J = 1$.

O método das réplicas

Na prática, é impossível calcular a energia livre para uma configuração arbitrária de impurezas. Para simplificar o problema, vamos supor que o sistema possa ser dividido em um grande número de subsistemas com dimensões superiores aos comprimentos de correlação. Dessa forma, construímos um conjunto de sistemas estatisticamentes independentes, cada um com uma configuração de impurezas gerada pela mesma distribuição. A energia livre média por sítio fica escrita como

$$f = -\frac{1}{\beta N}\langle \ln Z\rangle . \tag{1.9}$$

O cálculo dessa média $\langle ... \rangle$ é um problema difícil. Descreveremos brevemente um dos mais intrigantes e populares métodos para calcular a média (1.9), conhecido como método das réplicas. Este método faz o uso formal da identidade

$$\ln Z = \lim_{n \to 0} \frac{Z^n - 1}{n} \tag{1.10}$$

onde

$$Z^n = Tr_{S_{i,\alpha}} \exp\{-\beta \sum_{\alpha=1}^{n} H[s_{i,\alpha}]\}. \tag{1.11}$$

Alguns pontos que merecem ser ressaltados. Note que começamos com n inteiro para o cálculo de Z^n e quando tomamos o limite $n \to 0$ estamos considerando-o como real. Esse é um passo delicado no procedimento e poderia levar a resultados não físicos no regime de baixa temperatura. A maior justificativa para o uso deste método tem sido o resultado das simulações que, em geral, concordam muito bem com a teoria, embora haja alguns resultados matemáticos rigorosos que justifiquem esse procedimento [4].

A aleatoriedade (desordem) introduz características especiais em problemas teóricos em Mecânica Estatística. Primeiro temos o problema fundamental de que não conhecemos todos os parâmetros da Hamiltoniana do sistema que tentamos estudar; conhecemos somente os parâmetros de um ensemble estatístico de Hamiltonianas. Além disso, não calculamos (e nem sabemos calcular) as propriedades de um conjunto particular de interações aleatórias J_{ij}. Felizmente estamos interessados em médias sobres sistemas grandes. Sabemos da Mecânica Estatística ordinária que as flutuações relativas da energia em torno do seu valor médio são da ordem $\mathcal{O}(N^{-\frac{1}{2}})$ e assim, esperamos que essas flutuações amostrais tendem a zero no limite termodinâmico (esse argumento é válido para expoentes menores do que 1). Uma grandeza aleatória com essa propriedade é dita ser *automediante*. Se soubermos de antemão que uma grandeza de interesse é automediante, então podemos esperar que não somente os resultados de diferentes amostras em um experimento sejam os mesmos, como também esperar que o cálculo teórico do valor médio dessa grandeza sobre todo o ensemble dê a mesma resposta que os experimentos. Portanto, a estratégia é sempre tentar se concentrar em grandezas automediantes e cal-

cular seu valor médio sobre a distribuição aleatória de um ensemble estatístico. Sabemos que muitas grandezas não são automediantes; porém, felizmente muitas grandezas que pretendemos medir ou calcular são somas ou integrais sobre todo o volume da amostra, sendo então denominadas grandezas extensivas e apresentando flutuações estatísticas que se tornam muito pequenas para grandes amostras.

Conexão entre Mecânica Estatística e Otimização

Problemas de otimização combinatorial podem ser formalmente considerados como modelos de Mecânica Estatística e vice-versa. A função custo é renomeada Hamiltoniana e o estado fundamental do modelo formal corresponde às soluções do problema de otimização. Em geral, a função custo é mapeada em uma Hamiltoniana de vidros de spin, onde os a_j seriam equivalentes aos acoplamentos J_{ij} e o vetor x seria isomorfo ao vetor de spin s. Desta forma, os métodos desenvolvidos na Mecânica Estatística de sistemas desordenados tornam-se ferramentas poderosas para a análise probabilística de problemas de otimização [25]. De fato, estes métodos vêm sendo aplicados a inúmeros problemas de otimização como por exemplo, o **problema do caixeiro viajante** [26, 27], o **problema do particionamento de grafos** [28] e o **problema da k -sat** [29, 30], para mencionar alguns. Um exemplo particular deste tipo de abordagem que merece destaque aqui é a análise do **problema da designação** : Dada uma matriz $N \times N$ com elementos não negativos $a_{i,j} \geq 0$, encontre

$$E_N^* = \min_{\sigma} \sum_{i=1}^{N} a_{i,\sigma(i)} \tag{1.12}$$

onde o mínimo é tomado sobre todas as permutações σ de $(1, \ldots, N)$. Para entender melhor esse problema, imagine que $a_{i,\sigma(i)}$ é uma matriz aleatória. Devemos atribuir uma linha a cada coluna e no espaçøde possibilidades (gerado pelas permutações), devemos ainda escolher aquela que minimiza a função custo. Portanto, a soma 1.12 não é simplesmente uma soma de variáveis aleatórias trivial, pois não conhecemos a distribuição de probabilidades das atribuição ótimas.

A análise probabilística visa calcular as propriedades médias de um ensemble de instâncias aleatórias,

onde as variáveis aleatórias $a_{i,j}$ são estatisticamente independentementes extraídas de uma função densidade de probabilidade comum $\rho(a)$ (note que a assume valores reais). Usando o método das réplicas, Mézard e Parisi [31] encontraram (entre outras coisas) que

$$\lim_{N\to\infty} \langle E_N^* \rangle = \frac{\pi^2}{6}, \tag{1.13}$$

onde $\langle ... \rangle$ denota a média sobre os $a_{i,j}$. Alguns anos depois Parisi reconheceu que para uma matriz com elementos distribuídos exponencialmente, ou seja, $\rho(a) = e^{-a}$, a energia média ótima para $N = 1$ e $N = 2$ é, respectivamente,

$$\langle E_1^* \rangle = 1 \qquad \langle E_2^* \rangle = 1 + \frac{1}{2^2}. \tag{1.14}$$

Esta observação e o fato de que o resultado do cálculo usando o método das réplicas para $N \to \infty$ pode ser escrito como

$$\frac{\pi^2}{6} = \sum_{k=1}^{\infty} \frac{1}{k^2}, \tag{1.15}$$

levou-o a conjecturar que o valor médio ótimo para sistemas finitos é [32]

$$\langle E_N^* \rangle = \sum_{k=1}^{N} \frac{1}{k^2}. \tag{1.16}$$

A prova da conjectura de Parisi foi apresentada recentemente [33, 34] e representa um dos raros casos onde o resultado do método das réplicas foi confirmado por um método rigoroso. Note que as equações (1.14) e (1.16) são válidas somente para $\rho(a) = e^{-a}$, enquanto a equação (1.13) é válida para todas as densidade com $\rho(a \to 0) = 1$. Para a densidade de probabilidade uniforme no intervalo unitário $[0, 1)$ os primeiros termos são

$$\langle E_1^* \rangle = \frac{1}{2} \qquad \langle E_2^* \rangle = \frac{23}{30}, \tag{1.17}$$

e agora fica difícil adivinhar o resultado geral para N finito.

Transições de fase em complexidade computacional

A teoria da complexidade computacional é baseada inteiramente na análise do *pior* caso. Pode acontecer que um algoritmo requeira um tempo exponencial para resolver um problema em somente uma instância patológica. Um exemplo famoso, mencionado anteriormente é o método Simplex para resolver problemas de **programação linear**. Apesar de sua complexidade ser exponencial, hoje ele é utilizado em muitas aplicações para resolver muitos problemas. Aparentemente as instâncias que necessitam de um tempo exponencial de processamento não aparecem em condições práticas. Tomemos o problema 3-SAT como um exemplo (veja o apêndice A para a definição deste problema). Em realizações geradas aleatoriamente com N variáveis e M cláusulas, quando executadas por um algoritmo especializado, mas baseado em busca exaustiva, observa-se um tempo polinomial de execução, a menos que a razão M/N seja ajustada cuidadosamente. Se a razão M/N for pequena, o problema é dito *subvinculado*, isto é, tem muitas soluções que satisfazem o problema. Quando a razão M/N é muito grande o problema se torna *supervinculado*, isto é, existe um grande número de vínculos contraditórios. Instâncias realmente muito difícies são geradas para razão $\alpha = M/N$ próximo de um valor crítico α_c [35].

A transição de fórmulas lógicas subvinculadas para supervinculadas no 3-SAT apresenta os sinais característicos de uma transição de fase em sistemas físicos. O parâmetro de controle é α, e o parâmetro de ordem é a probabilidade da fórmula lógica ser satisfatível. Uma transição de fase semelhante foi descoberta e analisada com os métodos da Mecânica Estatística em outros problema computacionais como o problema da cobertura de vértices [36] e o problema da partição numérica [37].

O problema da partição numérica (**PPN**)

O **PPN** é um problema clássico da otimização combinatorial que, apesar de possuir uma formulação

extremamente simples, fornece resultados não triviais. Ele é definido como segue: Dada uma lista $a_1, a_2, \ldots, a_N$ de números positivos, encontre uma partição, isto é, um subconjunto $\mathcal{A} \subset \{1, \ldots, N\}$ tal que a *diferença entre as partições*

$$E(\mathcal{A}) = \left| \sum_{i \in \mathcal{A}} a_i - \sum_{i \notin \mathcal{A}} a_i \right|, \tag{1.18}$$

seja minimizada. Podemos inserir um vínculo especial tal que a diferença de cardinalidade entre os subconjuntos (partições) $\mathcal{A}$ com N_1 elementos e seu complementar com N_2 elementos ,

$$M = |N_1 - N_2|, \tag{1.19}$$

seja mantida fixa. Um caso especial é o problema da *partição numérica balanceado*, obtido impondo -se $M = 0$. Alternativamente podemos mapear o **PPN** em um modelo de vidro de spin: associando a cada entrada a_i uma variável s_i de forma que se $a_i \in A$ então $s_i = 1$, caso contrário $s_i = -1$. Agora podemos procurar no espaço de configuração de spins $s_1, s_2, ..., s_N$ que minimiza a função custo (ou a energia)

$$E = \left| \sum_{i=1}^{N} s_i a_i \right|. \tag{1.20}$$

Elevando a expressão (1.20) ao quadrado

$$E_{SG} = \sum_{ij} J_{ij} s_i s_j \tag{1.21}$$

com $J_{ij} = a_i a_j$, podemos ver que (1.21) é uma Hamiltoniana de um vidro de spin de alcance infinito. O vínculo de cardinalidade fica sendo a magnetização

$$m = \frac{1}{N} \left| \sum_{i=1}^{N} s_i \right|. \tag{1.22}$$

Encontrar a solução ótima para o problema da partição numérica consiste em encontrar a configuração de spin do estado fundamental de um problema de vidro de spin.

O **PPN** é de interesse tanto prático quanto teórico. Em particular, ele é um dos seis problemas básicos

NP-completos, segundo Garey e Johnson, que estão situados no coração da teoria da NP-completeza [1]. Entre as aplicações práticas encontramos o escalonamento de tarefas [38,39].

O **PPN** requer entradas $\{a_1, a_2, ..., a_N\}$ de tamanho arbitrário com precisão ilimitada. Para estudar as propriedades típicas deste problema as instâncias são usualmente consideradas como sendo números aleatórios estatisticamente independentes. Sob estas hipóteses probabilísticas a partição ótima E^* é uma variável estocástica. Quando as entradas são reais e limitadas entre 0 e 1 (precisão infinita), Karmakar e colaboradores [40] provaram que o valor médio de E^* é de ordem $\mathcal{O}(\sqrt{N}2^{-N})$ no caso sem vínculo de cardinalidade e é da ordem $\mathcal{O}(N2^{-N})$ para o caso balanceado. Entretanto, através de simulações numérica mostramos que a largura relativa da distribuição de E^*, definida como

$$r = \frac{\sqrt{\langle E^{*2}\rangle - \langle E^*\rangle^2}}{\langle E^*\rangle}, \tag{1.23}$$

onde $\langle ... \rangle$ denota a média sobre os a_j, tende para o valor 1 no limite $N \to \infty$ tanto para a versão sem vínculo quanto para a versão balanceada. Isto significa que as flutuações típicas de E^* são da mesma ordem de grandeza de seu valor esperado.

Outra característica do **PPN** é o péssimo desempenho das heurísticas propostas para tentar resolvê-lo [16]. Outro fato que chama a atenção é que ao tomarmos os números aleatórios a_j codificados por b-bits, o tempo para encontrarmos a solução cresce exponencialmente com N para $N < b$ e, polinomialmente para $N \gg b$. As instâncias ou realizações em que $N \sim b$ requerem o maior tempo para se encontrar a solução. Essa transição da fase computacionalmente difícil para uma fase fácil é muito semelhante às transições de fase observadas em sistemas físicos. Como já dissemos anteriormente, transições de fase deste tipo já foram observadas em inúmeros problemas NP-completos e são freqüentemente analisados quantitativamente dentro do esquema da Mecância Estatística. Esta análise é surpreendentemente mais simples para o caso do **PPN** do que para a maioria dos outros problemas.

Organização da tese

No capítulo 2 empregamos a aproximação recozida para derivar limitantes inferiores analíticos para a energia ótima esperada do **PPN** , tanto para o caso sem vínculo quanto para o caso com vínculo de cardinalidade. Para o vínculo m acima de um certo valor crítico (caso supervinculado), calculamos a energia ótima em função da diferença de cardinalidade. Este é o regime em que a energia é extensiva e usamos a propriedade de automediância para obter esse resultado. Calculamos também o número de mínimos locais e a energia típica desses mínimos. Um mínimo local é o estado que possui a menor energia numa vizinhança de N estados que diferem pelo sinal de apenas um spin. Todos os resultados teóricos foram confrontados com simulações. A solução exata da termodinâmica do **PPN** encontra-se no apêndice B.

No capítulo 3 investigamos a versão relaxada do **PPN** , ou seja, permitimos que a variável s_i possa assumir qualquer valor real. Começamos pela derivação de um algoritmo de minimização que será usado para fazermos experimentos computacionais gerando os valores médios de grandezas que serão confrontados com a teoria. Em seguida calculamos a função de partição Z_u e Z_m para o caso sem vínculo e com vínculo respectivamente. Essas funções permitiram o cálculo exato da energia livre. Ainda calculamos a distribuição de probabilidade de uma partição ótima ter diferença de cardinalidade igual a m. Com essa distribuição derivamos facilmente os momentos e provamos que m não é uma grandeza automediante. A seguir, exibimos o cálculo da função de correlação entre 2 mínimos. Finalmente, concluímos o capítulo com uma análise de desempenho de uma heurística que denominamos heurística de integralização. Lembramos que heurísticas são algoritmos que não garantem que a solução encontrada seja ótima.

No capítulo 4 invertemos o problema. Ao invés de procurarmos pela configuração ótima dada uma realização, agora perguntamos se existe uma realização capaz de satisfazer P partições perfeitas. Na primeira seção usamos o método das réplicas para calcular a densidade de energia livre no limite termodinâmico e derivamos os parâmetros de ordem na hipótese de simetria de réplicas que descreve as propriedades do espaço de realização. Mostramos que há um valor crítico P_c que limita o número de partições perfeitas. Calculamos também a densidade de probabilidade para uma entrada a_k assumir um valor y. A análise de estabilidade que garante que nossos parâmetros de ordem obtidos no ansatz

de simetria de réplicas são confiáveis é apresentada no apêndice D.

Para uma introdução à complexidade computacional podemos recorrer ao apêndice A. Nele apresentamos alguns problemas da família $\mathcal{P}$ (polinomial) e $\mathcal{NP}$ (não-determinístico polinomial) de forma detalhada. O apêndice C dedica-se ao estudo da heurística de Karmarkar e Karp.

Capítulo 2

A Mecânica Estatística do problema da Partição Numérica

O problema da partição numérica é um problema clássico em otimização combinatorial: dado um conjunto $\mathcal{A}$ de números inteiros (ou reais), ele consiste em encontrar dois subconjuntos disjuntos que minimizem a diferença da soma dos elementos nos subconjuntos. Esse problema é muito interessante pois, apesar de sua formulação extremamente simples, é capaz de reproduzir bem a dificuldade de um problema de otimização combinatorial tanto do ponto de vista teórico quanto prático. Até mesmo pequenas instâncias desse problema apresentam essa dificuldade estimulando o interesse e a motivação para investigar potenciais métodos de solução. Apresentamos aqui uma abordagem do ponto de vista da física do problema da partição numérica. Formulamos o problema como um modelo de vidros de spin ou, mais corretamente, como um antiferromagneto desordenado. Encontramos expressões analíticas para a energia ótima média e o número de mínimos locais na aproximação 'annealed'. No apêndice B encontra-se o cálculo exato devido a Mertens [37]. Formalmente podemos enunciar o **PPN** da seguinte forma: dado uma sequência de números reais $A = \{a_1, a_2, ..., a_N\}$, esse problema consiste em dividí-los em dois subconjuntos A_1 e A_2 tal que a diferença

$$\left| \sum_{a_j \in A_1} a_j - \sum_{a_j \in A_2} a_j \right|, \tag{2.1}$$

seja minimizada. Alternativamente, podemos mapear este problema em um modelo de Ising e procurar a configuração s que minimiza a energia ou função custo

$$E(\mathbf{s}) = \left| \sum_{j=1}^{N} a_j s_j \right|, \tag{2.2}$$

onde $s_j = 1$ se $a_j \in A_1$ e $s_j = -1$ se $a_j \in A_2$. Podemos também considerar o problema de partições com vínculos, isto é, aquelas em que a cardinalidade dos conjuntos A_1 e A_2 diferem por uma constante m:

$$m = \frac{1}{N} \left| \sum_{j=1}^{N} s_j \right|. \tag{2.3}$$

No caso mais geral podemos dividir o conjunto A em **K** grupos. A energia é então definida como

$$f(a_1, a_2, ..., a_N) = \sum_{k=1}^{\mathbf{K}} \mid s_k - \overline{s} \mid \tag{2.4}$$

onde

$$s_k = \sum_{i \in A_k} a_i, \qquad \overline{s} = \frac{1}{\mathbf{K}} \sum_{k=1}^{\mathbf{K}} s_k. \tag{2.5}$$

A respeito de sua complexidade, o PPN pertence a classe NP-completo, portanto não é conhecido um algoritmo determinístico que resolve qualquer instância deste problema em um tempo polinomial [1]. O problema de minimizar a energia torna-se equivalente ao de encontrar o estado fundamental da Hamiltoniana de Ising antiferromagnética desordenada [5]. Ele também pode ser visto como o problema de encontrar a maneira mais justa de dividir um conjunto de N objetos j= $1, 2, ..., N$, cada um com valor a_j, entre duas pessoas. O **PPN** ainda aparece como um problema de programação da produção. Especificamente, dadas duas máquinas idênticas, um conjunto de tarefas, e o tempo requerido por cada máquina para a realização de cada tarefa, o problema é designar cada tarefa a uma máquina de modo a minimizar o tempo total de execução do trabalho. Em outras palavras, temos que dividir o tempo de processamento da tarefas em dois subconjuntos de modo que as somas dos tempos em cada um sejam aproximadamente iguais. Naturalmente, como já exposto anteriormente, este problema pode ser estendido para **K** máquinas.

Na seção 2 calculamos a energia ótima média na aproximação 'annealed' primeiro para o caso sem

vínculo e em seguida para o caso com vínculo de cardinalidade. Ainda nesta seção mostramos que existe um valor para o vínculo m onde a energia pode ser calculada exatamente. Na seção 3 calculamos o número médio de mínimos locais na aproximação 'annealed' e estimamos a energia típica de um mínimo. Todos os resultados apresentados aqui foram comparados com simulações numéricas.

2.1 Aproximação 'Annealed'

Nesta seção será desenvolvido de forma detalhada o procedimento usado para a obtenção de um limitante inferior para a energia ótima média do problema da partição numérica a partir de conceitos simples. No formalismo da Mecânica Estatística usando o ensemble canônico, o valor médio da energia ótima para partições sem vínculo de cardinalidade é dada por

$$\langle E \rangle = \lim_{T \to 0} F(T) = - \lim_{T \to 0} T \langle \ln Z \rangle , \tag{2.6}$$

onde $F(T)$ é a energia livre média e $Z(T)$ é a função de partição

$$Z(T) = \sum_{s} e^{-\frac{E(s)}{T}} \tag{2.7}$$

Aqui $\langle ... \rangle$ representa uma média sobre as variáveis aleatórias a_j. O limite $T \to 0$ garante que somente os estados de mais baixa energia contribuem para Z [1]. A entropia média $S(T) = -\frac{dF}{dT}$ de um sistema de spins de Ising é positiva para qualquer temperatura, portanto F deve ser uma função decrescente de T, o que torna válida a seguinte relação

$$\langle E \rangle = F(0) \geq F(T), \quad \text{para todo T} \tag{2.8}$$

Definindo a energia livre 'annealed' por

$$F^a(T) = -T \ln \langle Z(T) \rangle , \tag{2.9}$$

[1]Naturalmente, já que T não possui o significado físico de temperatura, tomamos a constante de Boltzmann como sendo igual a 1.

e usando a desigualdade de Jensen [8],

$$\ln \langle Z(T) \rangle \geq \langle \ln Z(T) \rangle \tag{2.10}$$

chegamos a seguinte desigualdade:

$$F^a(T) \leq F(T) \leq \langle E \rangle . \tag{2.11}$$

Portanto, a energia livre 'annealed' calculada para uma temperatura qualquer T fornece um limite inferior rigoroso para $\langle E \rangle$ [9, 10]. Claramente, o menor limite é dado por

$$\langle E^a \rangle = F^a(T^*), \tag{2.12}$$

onde T^* é a temperatura que maximiza $F^a(T)$, isto é,

$$\frac{dF^a}{dT} |_{T^*} = 0. \tag{2.13}$$

Este será o procedimento para obtermos soluções aproximadas para o PPN sem e com vínculo de cardinalidade.

2.1.1 O caso sem vínculo

Para obtermos a energia ótima na aproximação 'annealed', temos primeiro que calcular a função de partição Z_u definida pela equação (2.7), com a energia $E(s)$ dada pela equação (2.2). Agora vamos calcular explicitamente a energia $\langle E_u^a \rangle$. A função de partição neste caso escrita como

$$Z_u = \sum_s e^{-\frac{1}{T}|\sum_j a_j s_j|}. \tag{2.14}$$

Note que, se não houvesse o módulo no expoente, esta soma poderia ser calculada de forma muito simples, a saber,

$$\begin{aligned}
\sum_s e^{-\frac{1}{T}\sum_j a_j s_j} &= \sum_{s_j}\prod_{j=1}^{N} e^{-\frac{1}{T}a_j s_j} \\
&= \sum_{s_1=\pm 1} e^{-\frac{1}{T}a_1 s_1}\cdot \sum_{s_2=\pm 1} e^{-\frac{1}{T}a_2 s_2}\cdot \ldots \cdot \sum_{s_N=\pm 1} e^{-\frac{1}{T}a_N s_N} \\
&= 2\cosh\frac{a_1}{T}\cdot 2\cosh\frac{a_2}{T}\cdot \ldots \cdot 2\cosh\frac{a_N}{T} \\
&= 2^N\prod_{j=1}^{N}\cosh\frac{a_j}{T}. \qquad (2.15)
\end{aligned}$$

A questão é: como podemos tirar o módulo do expoente? O truque padrão em Mecânica Estatística para remover objetos não lineares como esse se dá por meio do uso criativo da *função delta*. Introduzida por P. A. M. Dirac no contexto de Mecânca Quântica, ela hoje possui uma descrição matemática exata [11]. Aqui utilizamos uma definição da *função delta* na representação integral,

$$\delta(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} d\hat{x}\, e^{\imath x\hat{x}}, \qquad (2.16)$$

de modo que $\delta(x)$ é 0 para $x \neq 0$ e ∞ para $x = 0$. O pico em $x = 0$ é perfeitamente calibrado para dar

$$\int_{-\infty}^{\infty} dx f(x)\delta(x-c) = f(c), \qquad (2.17)$$

para qualquer função f razoavelmente comportada. A função delta nos ajuda a separar o valor absoluto da soma sobre a variável s_j:

$$\begin{aligned}
Z_u &= \sum_{s_j}\int_{-\infty}^{\infty} dx\, e^{-|x|}\delta(x - \frac{1}{T}\sum_{j=1}^{N} a_j s_j) \\
&= \int_{-\infty}^{\infty} dx\, e^{-|x|}\frac{1}{2\pi}\int_{-\infty}^{\infty} d\hat{x}\, e^{-\imath x\hat{x}}\sum_{s_j} e^{-\imath\frac{\hat{x}}{T}\sum_j a_j s_j}. \qquad (2.18)
\end{aligned}$$

Agora podemos realizar a soma sobre s_j como na equação 2.15

$$Z_u = 2^N\int_{-\infty}^{\infty}\frac{d\hat{x}}{2\pi}\prod_{j=1}^{N}\cos(\frac{a_j}{T}\hat{x})\int_{-\infty}^{\infty} dx\, e^{-|x|+\imath\hat{x}x}. \qquad (2.19)$$

A integral sobre x é simplesmente

$$\int_{-\infty}^{\infty} dx\, e^{-|x|+\imath\hat{x}x} = \frac{2}{1+\hat{x}^2}. \tag{2.20}$$

Os a_j seguem uma distribuição uniforme no intervalo unitário $[0,1)$. Após efetuarmos a média sobre a_j a função Z_u fica

$$\begin{aligned}\langle Z_u\rangle &= 2^N \int_{-\infty}^{\infty} \frac{d\hat{x}}{2\pi}\frac{2}{1+\hat{x}^2}\prod_{j=1}^{N}\int_0^1 da_j \cos(\frac{a_j\tilde{x}}{T}) \\ &= 2^N \int_{-\infty}^{\infty} \frac{d\hat{x}}{2\pi}\frac{2}{1+\hat{x}^2}\left(\frac{\sin(\hat{x}/T)}{\hat{x}/T}\right)^N. \end{aligned} \tag{2.21}$$

Por simplicidade, realizamos uma mudança de variável ($\hat{x} = yT$), daí

$$\langle Z_u\rangle = 2^N \int_{-\infty}^{\infty} \frac{dy}{\pi}\frac{T}{1+(yT)^2} e^{NG_u(y)} \tag{2.22}$$

onde

$$G_u(T) = \ln\left[\frac{\sin(y)}{y}\right]. \tag{2.23}$$

A integral (2.22) pode ser efetuada assintoticamente para grandes valores de N usando o *método de ponto de sela* [13]: a idéia geral é que a integral é dominada pelas contribuições do máximo de $G(y)$. Se $G(y)$ tem um máximo em $y = y_{ps}$, então

$$\int e^{NG(y)}dx \approx e^{NG(y_{ps})}\int e^{-\frac{N}{2}G''(y_{ps})(y-y_{ps})^2}dy = e^{NG(y_{ps})}\sqrt{\frac{2\pi}{NG''(y_{ps})}} \tag{2.24}$$

para N grande. Na expressão (2.22) o ponto de sela é $y_{ps} = 0$. O resultado final é simplesmente

$$\langle Z_u(T)\rangle = 2^N T\sqrt{\frac{6}{\pi N}}, \tag{2.25}$$

A energia média na aproximação 'annealed' é então

$$\langle E_u^a\rangle = 2^{-N}\sqrt{\frac{\pi N}{6\,e^2}} \approx 0.266\, 2^{-N}\sqrt{N}. \tag{2.26}$$

Na figura 2.1 apresentamos a estimativa numérica para a energia média ótima $\langle E_u\rangle$ obtida através de uma busca exaustiva para N$\leq$ 24. Os dados foram muito bem ajustados pela curva $\langle E_u\rangle$ = 1.12

$2^{-N}\sqrt{N}$. Na figura 2.2 mostramos a razão $r_u = \sqrt{\sigma^2}/\langle E_u \rangle$ como função de N. Vemos que esta razão tende para 1 à medida em que N aumenta, indicando que $\langle E_u \rangle$ não é automediante.[2] Em todos os experimentos discutidos nesta seção, os pontos resultam de 10^4 realizações do conjunto $\{a_j\}$. A barra de erro foi calculada medindo-se o desvio padrão da energia média ótima obtida em 25 experimentos, cada uma envolvendo 400 realizações do conjunto $\{a_j\}$.

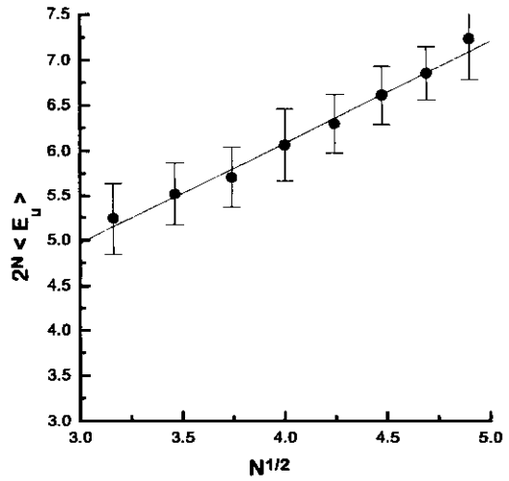


Figura 2.1: Energia ótima média obtida através de busca exaustiva como função de $N^{1/2}$ para partições sem vínculo. A curva contínua é um ajuste por regressão linear.

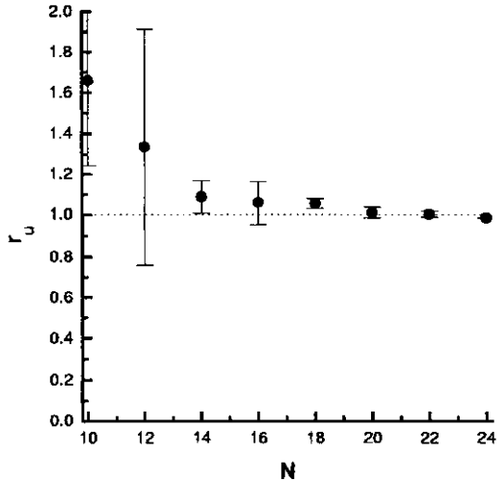


Figura 2.2: Razão entre o desvio padrão e o valor da energia ótima média obtida através de busca exaustiva como função de N

[2]Por definição, uma grandeza física extensiva é dita automediante se o erro entre o valor de uma observação e o seu valor médio cresce mais lentamente do que o tamanho do sistema N.

2.1.2 O caso com vínculo

Analisaremos agora o caso com vínculo de cardinalidade. Embora o cálculo seja um pouco mais complicado, o procedimento é o mesmo da subseção anterior. Novamente partimos da função de partição do PPN considerando o vínculo de cardinalidade:

$$Z_m = \sum_s \delta\left(Nm, \sum_{j=1}^{N} s_j\right) \exp\left[-\frac{E(s)}{T}\right], \tag{2.27}$$

com m $= -1, -1 + \frac{2}{N}, ..., 1 - \frac{2}{N}, 1$. A soma é sobre os 2^N estados s, $\delta(n, m)$ é a delta de Kronecker e T é a temperatura. Usando a representação integral da função delta de Kronecker, dada por

$$\delta_{n,m} = \int_{-\pi}^{\pi} \frac{d\hat{m}}{2\pi} e^{i\hat{m}(n-m)} = \{ \begin{array}{ll} 1 & \text{se } m = n \\ 0 & \text{se } m \neq n \end{array} \tag{2.28}$$

escrevemos

$$Z_m(T) = \sum_{\{s\}} \int_{-\infty}^{\infty} \frac{dx d\tilde{x}}{2\pi} e^{[i\tilde{x}x - \frac{|x|}{T}]} \int_{-\pi}^{\pi} \frac{d\tilde{m}}{2\pi} e^{i\tilde{m}Nm} e^{[-i\sum_j s_j(\tilde{x}a_j + \tilde{m})]}. \tag{2.29}$$

A soma sobre s fornece

$$Z_m(T) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{dx d\tilde{x}}{2\pi} e^{i\tilde{x}x - \frac{|x|}{T}} \int_{-\pi}^{\pi} \frac{d\tilde{m}}{2\pi} e^{i\tilde{m}Nm} \prod_j \{e^{[i(\tilde{x}a_j + \tilde{m})]} + e^{[-i(\tilde{x}a_j + \tilde{m})]}\}. \tag{2.30}$$

Agora podemos calcular $\langle Z_m(T) \rangle$ efetuando a média sobre a_j

$$\langle Z_m(T) \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{dx d\tilde{x}}{2\pi} e^{i\tilde{x}x - \frac{|x|}{T}} \int_{-\pi}^{\pi} \frac{d\tilde{m}}{2\pi} e^{iNm\tilde{m}} \tag{2.31}$$

$$\prod_j \int_0^1 da_j 2\{e^{[i(\tilde{x}a_j + \tilde{m})]} + e^{[-i(\tilde{x}a_j + \tilde{m})]}\}. \tag{2.32}$$

Como a integral sobre a_j é

$$\int_0^1 da_j 2\cos(a_j\tilde{x}+\tilde{m}) = \frac{\sin(\tilde{x}/2)}{\tilde{x}/2}\{\exp(i\tilde{m}+i\tilde{x}/2)+\exp(-i\tilde{m}-i\tilde{x}/2)\}, \tag{2.33}$$

e a integral sobre x já é conhecida,

$$\int_{-\infty}^{\infty} dx\, e^{i\tilde{x}x-\frac{|x|}{T}} = \frac{2T}{1+(T\tilde{x})^2}, \tag{2.34}$$

obtemos

$$\langle Z_m(T)\rangle = \int_{-\infty}^{\infty}\frac{d\tilde{x}}{2\pi}\frac{2T}{1+(T\tilde{x})^2}\left[\frac{\sin(\frac{\tilde{x}}{2})}{\frac{\tilde{x}}{2}}\right]^N\int_{-\pi}^{\pi}\frac{d\tilde{m}}{2\pi}e^{iNm\tilde{m}}[e^{i\tilde{m}+i\tilde{x}/2}+e^{-i\tilde{m}-i\tilde{x}/2}]^N. \tag{2.35}$$

Para continuar, precisamos recorrer ao teorema binomial para extrair a integral sobre $\tilde{m}$,

$$\left(e^{i\tilde{m}+i\tilde{x}/2}+e^{-i\tilde{m}-i\tilde{x}/2}\right)^N = \sum_{n=0}^{\infty}\begin{pmatrix} N \\ n \end{pmatrix} e^{i\tilde{m}(2n-N)}e^{i\tilde{x}(2n-N)/2}. \tag{2.36}$$

Daí substituindo (2.36) em (2.35) a integral sobre $\tilde{m}$ fica

$$\int_{-\pi}^{\pi}\frac{d\tilde{m}}{2\pi}e^{iNm\tilde{m}}e^{i\tilde{m}(2n-N)} = \delta(n, N\frac{1-m}{2}). \tag{2.37}$$

Essa delta de Kronecker facilitará nosso cálculo, uma vez que eliminaremos o somatório sobre n. Reescrevendo (2.35) como

$$\langle Z_m(T)\rangle = \sum_{n=0}^{\infty}\begin{pmatrix} N \\ n \end{pmatrix}\int_{-\infty}^{\infty}\frac{d\tilde{x}}{2\pi}\frac{2T}{1+(T\tilde{x})^2}e^{i\tilde{x}(2n-N)/2}\delta(n, N\frac{1-m}{2})\, e^{[N\ln(\frac{\sin(\frac{\tilde{x}}{2})}{\frac{\tilde{x}}{2}})]}, \tag{2.38}$$

e introduzindo a variável $y = \frac{\tilde{x}}{2}$, temos

$$\langle Z_m(T)\rangle = \begin{pmatrix} N \\ n^* \end{pmatrix}\int_{-\infty}^{\infty}\frac{dy}{\pi}\frac{2T}{1+(2Ty)^2}e^{NG_m(y)} \tag{2.39}$$

onde

$$n^* = N\frac{1-m}{2}, \tag{2.40}$$

$$G_m(y) = imy + \ln\left(\frac{\sin y}{y}\right). \tag{2.41}$$

No limite em que N é grande, a integral sobre y pode ser novamente avaliada usando-se o método de ponto de sela. A equação de ponto de sela tem como solução um imaginário puro $y_{ps} = i\varsigma$, onde ς é a solução real da equação

$$\frac{dG_m(y)}{dy}\,|_{i\varsigma} = 0, \tag{2.42}$$

ou seja,

$$m - \coth\varsigma + \frac{1}{\varsigma} = 0, \tag{2.43}$$

e, conseqüentemente, a função $G_m(y_{ps}) = G_m$ dada por

$$G_m = -m\varsigma + \ln\frac{\sinh\varsigma}{\varsigma}, \tag{2.44}$$

é real. Finalmente, com o auxílio da fórmula de Stirling para o coeficiente binomial, reescrevemos a equação 2.39 no limite em que N é grande, como

$$\langle Z_m(T)\rangle = \frac{2}{\pi N}\sqrt{\frac{1}{(1-m^2)\mid G''\mid}}\frac{2T}{1-(2T\varsigma)^2}e^{Ng_m} \tag{2.45}$$

onde

$$g_m = G_m - \frac{1+m}{2}\ln\frac{1+m}{2} - \frac{1-m}{2}\ln\frac{1-m}{2} \tag{2.46}$$

e

$$G'' = \frac{d^2G_m}{dy}\,|_{i\varsigma} = -\frac{1}{\varsigma^2} - 1 + \coth^2\varsigma = -1 + m^2 + \frac{2m}{\varsigma}. \tag{2.47}$$

IFSC-USP SERVIÇO DE BIBLIOTECA E INFORMAÇÃO

Neste estágio podemos calcular a temperatura que maximiza a energia livre 'annealed'. De fato, a equação (2.13) pode ser escrita como

$$\ln\langle Z_m(T_m^*)\rangle + \frac{1+(2T_m^*\varsigma)^2}{1-(2T_m^*\varsigma)^2} = 0. \tag{2.48}$$

Vamos considerar primeiro o regime no qual $\langle Z_m(T_m^*)\rangle$ é da ordem de 1. Neste caso, a equação (2.45) implica que T_m^* é muito pequeno e, então, a equação (2.48) reduz-se a

$$\langle Z_m(T_m^*)\rangle = e^{-1}. \tag{2.49}$$

Inserindo este resultado na equação (2.9) encontramos que $\langle E_m^a\rangle = T_m^*$ e, conseqüentemente,

$$\langle E_m^a\rangle = \frac{\pi N}{4}\sqrt{(1-m^2)\mid G''\mid}e^{-1-Ng_m}, \tag{2.50}$$

que é consistente com a hipótese de que T_m^* é pequeno para grandes valores de N, uma vez que $g_m > 0$. Desde que g_m decresce monotonicamente com m, de $g_0 = \ln 2$ para $g_1 = -\infty$, esta hipótese deixa de ser válida para $\mid m \mid > 0.560$ onde g_m é negativo. Doravante iremos supor que $m \geq 0$.

É instrutivo considerar em detalhes o caso de partições pares ou balanceadas ($m = 0$). Encontramos $\varsigma = 0$, $g_0 = \ln 2$ e $G''_0 = -\frac{1}{3}$, e a função de partição e a energia 'annealed' neste caso são dadas por

$$\langle Z_0(T)\rangle = 2^N T\frac{4\sqrt{3}}{\pi N}, \tag{2.51}$$

e

$$\langle E_0^a\rangle = 2^{-N}\frac{\pi N}{4e\sqrt{3}} \approx 0.167\, 2^{-N}N, \tag{2.52}$$

respectivamente. Nas figuras 2.3 e 2.4 apresentamos os resultados dos experimentos numéricos para estimar a energia do mínimo global para partições pares obtidas por meio de uma busca exaustiva no espaço de configurações para N $\leq$ 24. Nestes experimentos enfocamos a dependência em N da energia ótima média $\langle E_m\rangle$, e a razão $r_m = \sqrt{\sigma_m^2}/\langle E_m\rangle$, onde $\sigma_m^2 = \langle E_m^2\rangle - \langle E_m\rangle^2$ é a variância da variável aleatória E_m. Na figura 2.3, mostramos $\langle E_0\rangle$ como função de N. A linha contínua mostra a

curva obtida de $\langle E_0 \rangle = 0.80\, 2^{-N} N$. Portanto, embora o limite 'annealead' $\langle E_0^a \rangle$ escale corretamente com N, ele é cinco vezes menor do que nossa estimativa para $\langle E_m \rangle$. Na figura 2.4 mostramos a razão r_0 como função de N. O fato interessante é que essa razão tende para 1 para grandes valores de N, mostrando que a energia ótima E_0 não é automediante.

Consideremos agora o regime no qual $\langle Z_m(T_m^*) \rangle$ é da ordem de e^N. Encontramos $\ln\langle Z_m(T_m^*) \rangle \approx N g_m$ e $T_m^* \approx 1/2\varsigma$, então

$$\langle E_m^a \rangle = -N \frac{g_m}{2\varsigma}. \tag{2.53}$$

Naturalmente, esta solução é válida somente para $m > 0.560$ onde g_m é negativo. Notamos que (2.53) dá um limitante inferior muito pobre para $\langle E_m \rangle$. Em particular, para $m = 1$, temos $\langle E_1 \rangle = N/2$ enquanto o limite 'annealed' fornece $\langle E_1^a \rangle = 0$.

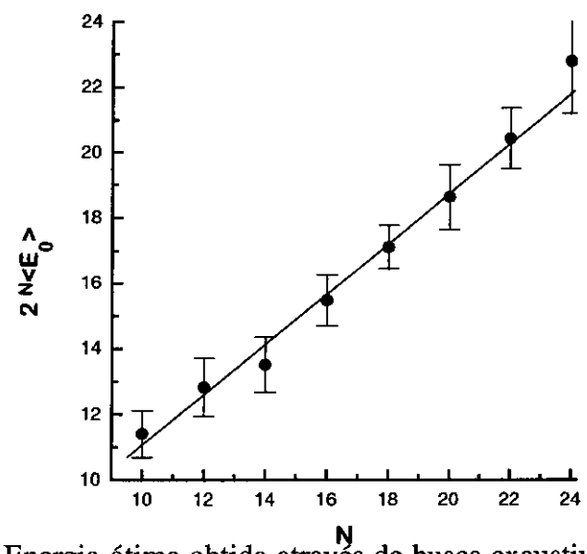


Figura 2.3: Energia ótima obtida através de busca exaustiva como função de N para partições pares ($m = 0$).

2.1.3 Energia Ótima E_m

Felizmente, no regime de energia extensiva podemos usar a propriedade de automediância da densidade de energia livre para calcular E_m exatamente para grandes valores de N. Assim podemos calcu-

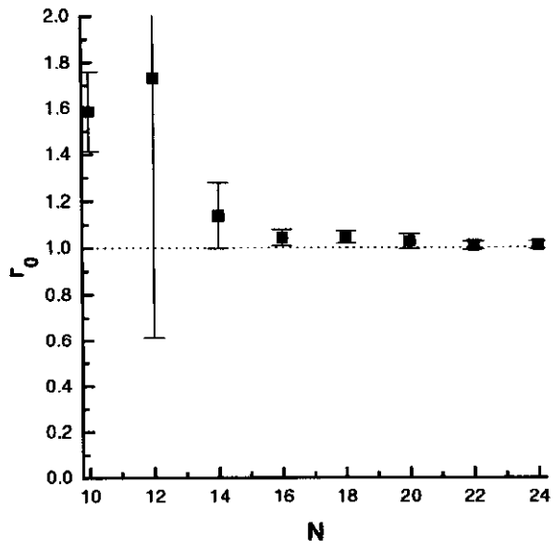


Figura 2.4: Razão entre o desvio padrão e o valor da energia ótima média obtida através de busca exaustiva como função de N para partições pares ($m = 0$).

equação (2.32) escrevemos a função de partição definida em (2.27) como

$$Z_m(T) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{dx d\tilde{x}}{2\pi} e^{ix\tilde{x}-\beta|x|} \int_{-\pi}^{\pi} \frac{d\tilde{m}}{2\pi} e^{iNm\tilde{m}} \prod_j \sum_{s_j=\pm 1} \exp[-is_j(a_j\tilde{x}+\tilde{m})], \tag{2.54}$$

onde $\beta = \frac{1}{T}$ é o inverso da temperatura. Como na aproximação 'annealed', a soma sobre s_j fornece

$$\begin{aligned} Z_m(T) &= N\beta^2 \int_{-\infty}^{\infty} dx \int_{-i\infty}^{i\infty} \frac{d\tilde{x}}{2\pi i} \int_{-i\pi/\beta}^{i\pi/\beta} \frac{d\tilde{m}}{2\pi i} e^{iNm\tilde{m}} \exp[-N\beta(x\tilde{x}+ \mid x \mid +m\tilde{m})] \\ &\times \exp[N\int_0^1 da \ln 2\cosh\beta(\tilde{x}a+\tilde{m})], \end{aligned} \tag{2.55}$$

onde usamos a propriedade de automediância

$$\frac{1}{N}\sum_j \ln 2\cosh(\tilde{x}a_j+\tilde{m}) = \int_0^1 da \ln 2\cosh(\tilde{x}a+\tilde{m}), \tag{2.56}$$

que é exata para N$\rightarrow \infty$. Neste limite podemos resolver a integral usando o método de ponto de sela. Assim, obtemos a seguinte equação para a densidade média da energia livre $\bar{f}_m = F_m/N$:

$$\bar{f}_m = x\tilde{x} + \mid x \mid + m\tilde{m} - \frac{1}{\beta}\int_0^1 da \ln 2\cos\beta(\tilde{x}a + \tilde{m}). \tag{2.57}$$

As equações de ponto de sela para o expoente acima são determinadas derivando-se em relação a $\tilde{x}$, $\tilde{m}$ e x:

$$x - \int_0^1 a \tanh\beta(\tilde{x}a + \tilde{m})da = 0, \tag{2.58}$$

$$m - \int_0^1 \tanh\beta(\tilde{x}a + \tilde{m})da = 0, \tag{2.59}$$

$$\tilde{x} = -1. \tag{2.60}$$

No limite de temperatura zero ($\beta \to \infty$), temos $\tanh\beta(-a + \tilde{m}) = 1$, se $a < \tilde{m}$ e $\tanh\beta(-a + \tilde{m}) = -1$, se $a > \tilde{m}$. Daí,

$$x = \int_0^1 a \tanh\beta(-a + \tilde{m})da = \int_0^{\tilde{m}} ada - \int_{\tilde{m}}^1 ada = \tilde{m}^2 - 1/2, \tag{2.61}$$

$$m = \int_0^1 \tanh\beta(-a + \tilde{m})da = \int_0^{\tilde{m}} da - \int_{\tilde{m}}^1 da = 2\tilde{m} - 1. \tag{2.62}$$

Resumindo, os pontos de sela encontrados são

$$\tilde{x} = -1, \tag{2.63}$$

$$\tilde{m} = (1 + m)/2 \quad , \tag{2.64}$$

$$x = \frac{(1+m)^2}{4} - \frac{1}{2}, \tag{2.65}$$

onde supomos $x \geq 0$ e $m \geq 0$. A energia média ótima é obtida tomando-se o limite de temperatura zero na equação (2.57), a qual fornece $\bar{f}_m \to \langle E_m \rangle / N = \mid x \mid$. O resultado final é simplesmente

$$\langle E_m \rangle = \frac{N}{2}\left[\frac{(1+m)^2}{2} - 1\right], \tag{2.66}$$

válido para $m \geq \sqrt{2} - 1 \approx 0.414$. Podemos observar que o limite inferior 'annealed' é ainda muito pobre para

$0.414 < m < 0.560$, já que nesta região $\langle E_m^a \rangle$ decresce exponecialmente com N, enquanto $\langle E_m \rangle$ aumenta linearmente com N.

Para melhor apreciarmos as diferenças qualitativas entre os regimes que escalam distintamente com N, apresentamos na figura (2.5) a estimativa numérica de $\langle E_m \rangle$ com função de m para $N = 24$. Elas confirmam a existência de dois regimes diferentes de escala com N, bem como a concordância muito boa com as predições teóricas para m > 0.414. Uma característica a ressaltar em nossa estimativa numérica para $\langle E_m \rangle$, em contraste com o limite inferior 'annealed' (2.50), é que as partições pares ($m = 0$) não dão a menor energia. Verificamos que este resultado também vale para pequenos valores de N. Além do mais, parece ocorrer uma transição abrupta para $m \approx 0.25$ como indicado pelas grandes barras de erros e pela mudança em três ordens de magnitude em $\langle E_m \rangle$.

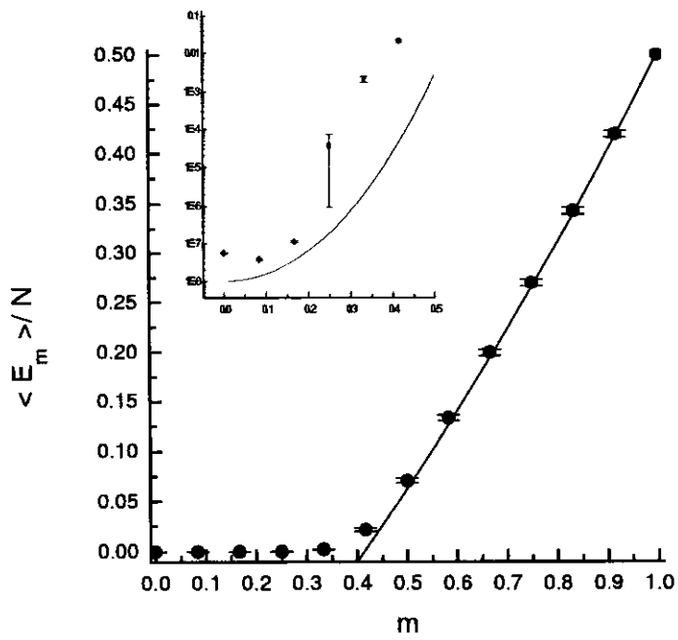


Figura 2.5: Energia ótima média obtida através de busca exaustiva como função de m para $N = 24$. A curva contínua no gráfico principal é a estimativa teórica (2.66) enquanto a curva no gráfico menor é o limite inferior na aproximação 'annealed' (2.50).

2.2 Número médio de mínimos locais

Nesta seção vamos desenvolver uma expressão que determina o número médio de estados meta-estáveis e permite estimar na aproximação 'annealed' a energia típica de um mínimo local. Um estado meta-estável é identificado como aquele que tem a menor energia dentre os N estados que diferem dele somente por um único spin [16]. A análise seguinte está restrita a partições sem vínculo, uma vez que para partições com vínculo teríamos que considerar a mudança simultânea de dois spins para satisfazer a cardinalidade do vínculo. O número médio de mínimos locais com energia $E = \mid t \mid$ é definido por

$$M(t) = \sum_{\{s\}} \delta\left(t - \sum_j s_j a_j\right) \prod_j \Theta\left(\mid t - 2s_j a_j \mid - \mid t \mid\right) \tag{2.67}$$

onde $\delta(x)$ é a função delta de Dirac e $\Theta(x) = 1$, se $x \geq 0$ e $\Theta(x) = 0$, se $x < 0$. Usando a representação integral da função delta, obtemos

$$M(t) = \sum_{\{s\}} \int_{-\infty}^{\infty} \frac{d\tilde{t}}{2\pi} e^{i\tilde{t}(t - \sum s_j a_j)} \prod_j \Theta\left(\mid t - 2s_j a_j \mid - \mid t \mid\right) \tag{2.68}$$

Efetuando a soma sobre s, temos

$$M(t) = \int_{-\infty}^{\infty} \frac{d\tilde{t}}{2\pi} e^{i\tilde{t}t} \prod_j \left[e^{-i\tilde{t}a_j}\Theta\left(\mid t - 2a_j \mid - \mid t \mid\right) + e^{i\tilde{t}a_j}\Theta\left(\mid t + 2a_j \mid - \mid t \mid\right)\right]. \tag{2.69}$$

Devemos ter em mente que a média é sobre as realizações a_j, daí

$$\langle M(t) \rangle = \int_{-\infty}^{\infty} \frac{d\tilde{t}}{2\pi} e^{i\tilde{t}t} \prod_j \int_0^1 da_j \left[e^{-i\tilde{t}a_j}\Theta\left(x_-\right) + e^{i\tilde{t}a_j}\Theta\left(x_+\right)\right], \tag{2.70}$$

onde $x_- = \mid t - 2a_j \mid - \mid t \mid$ e $x_+ = \mid t + 2a_j \mid - \mid t \mid$. Para calcular a integral em a_j é preciso analisar os argumentos da função Θ para determinarmos os limites corretos, ou seja, aqueles que garantem que $\Theta(x_\pm) = 1$. Como este cálculo é um pouco delicado, adotamos a seguinte estratégia:

$$x_- = |t - 2a_j| - |t| > 0 \tag{2.71}$$

se $t < 0$ e $t - 2a_j > 0$, então $\Theta(x_-) = 1$ para $-\infty < t < 0$,

se $t > 0$ e $t - 2a_j > 0$, então $\Theta(x_-) = 1$ para $0 < t < a_j$,

Podemos então resolver a integral em a_j com o integrando $\Theta(x_-)$,

$$\int_t^1 da_j e^{-i\tilde{t}a_j} = \frac{e^{-i\tilde{t}t} - e^{-i\tilde{t}}}{i\tilde{t}} \qquad 0 < t < \tag{2.72}$$

$$\int_0^1 da_j e^{-i\tilde{t}a_j} = \frac{1 - e^{-i\tilde{t}}}{i\tilde{t}} \qquad t < 0. \tag{2.73}$$

Por outro lado,

$$x_+ = |t + 2a_j| - |t| > 0 \tag{2.74}$$

se $t < 0$ e $t + 2a_j > 0$, então $\Theta(x_+) = 1$ para $-a_j < t < 0$,

se $t > 0$ e $t + 2a_j > 0$, então $\Theta(x_+) = 1$ para $0 < t < \infty$,

Neste caso a integral em a_j fica

$$\int_{-t}^1 da_j e^{i\tilde{t}a_j} = \frac{e^{i\tilde{t}} - e^{-i\tilde{t}t}}{i\tilde{t}} \qquad -1 < t < 0 \tag{2.75}$$

$$\int_0^1 da_j e^{i\tilde{t}a_j} = \frac{e^{i\tilde{t}} - 1}{i\tilde{t}} \qquad t > 0. \tag{2.76}$$

Resta agora resolver a integral em $\tilde{t}$, para isto devemos observar que fazendo $t \to -t$ e $\tilde{t} \to -\tilde{t}$ nas integrais (2.72), (2.73) e (2.75), (2.76) podemos resumí-las em apenas duas. Portanto

$$\langle M(t) \rangle = \int_{-\infty}^{\infty} \frac{d\tilde{t}}{2\pi} e^{it\tilde{t}} \left(\frac{e^{-it\tilde{t}} - e^{-i\tilde{t}} + e^{i\tilde{t}} - 1}{i\tilde{t}} \right)^N \qquad \text{se} \quad |t| < 1, \tag{2.77}$$

e

$$\langle M(t)\rangle = \int_{-\infty}^{\infty} \frac{d\tilde{t}}{2\pi} e^{it\tilde{t}} \left(\frac{e^{i\tilde{t}}-1}{i\tilde{t}}\right)^N = 0 \qquad \text{se} \qquad |\, t \,| > 1. \tag{2.78}$$

Usamos o resultado que a integral na equação (2.78) anula-se para todo N [17]. Por isso não existe nenhum mínimo local com $E \geq 1$. Como é usual, supomos N grande na integral da equação (2.77), podemos efetuá-la via integração de ponto de sela:

$$< M(t) >= \int_{-\infty}^{\infty} \frac{d\tilde{t}}{2\pi} e^{it\tilde{t}} \left[\frac{2\sin\tilde{t}}{\tilde{t}} - 2e^{-i\tilde{t}t/2}\frac{\sin(t\tilde{t}/2)}{\tilde{t}}\right]^N = \int_{-\infty}^{\infty} \frac{d\tilde{t}}{2\pi} e^{it\tilde{t}+NH(\tilde{t})}, \tag{2.79}$$

de modo que o resultado final é simplesmente

$$\langle M(t)\rangle = \sqrt{\frac{1}{2\pi N \mid H''(\xi) \mid}} e^{NH(\xi)} \tag{2.80}$$

onde

$$H(\xi) = \ln 2 + \ln\left[\frac{\sinh\xi}{\xi} - e^{-t\xi/2}\frac{\sinh(t\xi/2)}{\xi}\right]. \tag{2.81}$$

Aqui, ξ é solução de

$$\frac{2}{\xi} - \frac{2\cosh\xi - e^{-t\xi}t}{\sinh\xi - e^{-t\xi/2}\sinh(t\xi/2)} = 0. \tag{2.82}$$

A função H(ξ) é uma função monotonicamente decrescente de $E = \mid t \mid$. Em particular, ela decresce de ln2 quando E = 0 ($\xi = 0$) para -∞ quando $E = 1$ ($\xi = -\infty$). Ela se anula para $E \approx 0.851$, portanto o número de mínimos locais com energia maior do que esta, decresce exponencialmente com N.

Desconsiderando a energia, o número médio de mínimos locais é definido por

$$< M >= \int_{-\infty}^{\infty} dt < M(t) > . \tag{2.83}$$

Somente na vizinhança $t = 0$ ocorre uma contribuição para esta integral, por isso podemos expandir o integrando de (2.77) em série de potência de t e $\tilde{t}$ e manter os termos de menor ordem,

$$\frac{e^{-i\tilde{t}t} - e^{-i\tilde{t}} + e^{t} - 1}{i\tilde{t}} \approx 2 - t - \frac{1}{3}\tilde{t}^2. \tag{2.84}$$

Logo,

$$H = \ln 2 - \frac{t}{2} - \frac{1}{6}\tilde{t}^2, \tag{2.85}$$

e a integral (2.83) fica

$$\begin{aligned} \langle M \rangle &= 2^N \int_{-\infty}^{\infty} \frac{d\tilde{t}}{2\pi} \int_0^1 dt \, \exp(-N\frac{t}{2} - N\frac{\tilde{t}^2}{6} + +it\tilde{t}) && (2.86) \\ &= 2^N \left(\frac{3}{2\pi N}\right)^{\frac{1}{2}} \int_0^1 dt \, \exp(-\frac{Nt}{2}), && (2.87) \end{aligned}$$

o que resulta em

$$\langle M \rangle = \sqrt{\frac{24}{\pi}} \frac{2^N}{N^{3/2}} \approx 2.764 \frac{2^N}{N^{3/2}}. \tag{2.88}$$

É interessante estimar a dependência em N da energia típica de um mínimo local. Esta quantidade, denotada por E_t, é definida por

$$E_t = \left\langle \frac{\int dt \, t \, M(t)}{\int dt \, M(t)} \right\rangle. \tag{2.89}$$

Na aproximação 'annealed' [18], reescrevemos

$$E_t \approx \frac{\int dt \, t \, \langle M(t) \rangle}{\langle M \rangle} \tag{2.90}$$

O procedimento para avaliar (2.90) é idêntico ao usado para avaliar (2.83) e fornece

$$E_t \approx \frac{2}{N}. \tag{2.91}$$

Essas grandezas foram estimadas numericamente: para cada valor de N, variando de 100 a 3000, geramos 10^5 estados aleatórios s e contamos a fração daqueles que são mínimos locais e medimos sua energia. Os dados numéricos foram muito bem ajustados pelas equações $\langle M \rangle \approx (2.81 \pm 0.02) \frac{2^N}{N^{3/2}}$ e $E_t \approx \frac{(1.76 \pm 0.04)}{N}$, que concordam muito bem com as predições teóricas. Note que enquanto a equação (2.88) dá exatamente o termo de ordem do número médio de mínimos locais, a equação (2.91) é uma aproximação para a energia típica dos mínimos. A precisão dos números aleatórios são irrelevantes quando queremos simplesmente estimar o número de mínimos locais.

Capítulo 3

Análise Estatística do Problema da Partição Numérica Relaxado

Neste capítulo vamos utilizar a Mecânica Estatística para estudar analiticamente a versão relaxada do problema da partição numérica. O termo relaxado refere-se à perda da condição de integralidade das componentes s_j da partição, que originalmente assumem os valores ± 1. Agora s_j é uma variável real, podendo assumir qualquer valor no intervalo $(-\infty, \infty)$. As versões relaxadas de problemas de programação inteira são úteis por fornecerem limitantes inferiores para a função custo. Vamos derivar uma expressão que descreve a distribuição de probabilidade dos vínculos de cardinalidade m associados às partições ótimas e, a partir desse resultado, derivamos os seus momentos. Todos os resultados obtidos analiticamente são confrontados com simulações computacionais e a primeira seção deste capítulo é dedicada à derivação de um algoritmo para realizarmos essas simulações. Além disso, vamos estudar o desempenho de uma heurística que denominamos de heurística de integralização. Esse estudo é interessante dentro do contexto de análise de algoritmos. Quando alguém desenvolve um novo algoritmo para resolver um problema, o que se busca saber é quão bom é o seu desempenho e compará-lo com outros algoritmos existentes para resolver a mesma tarefa. Somente então podemos avaliar as vantagens deste algoritmo em relação aos demais. A última seção é dedicada a este tópico e mostramos que a energia por ela obtida é idêntica à energia típica de um estado (partição) escolhido aleatoriamente.

3.1 Algoritmo de Relaxação

O problema da partição numérica com vínculo de cardinalidade já foi enunciado no capítulo anterior. Basicamente ele consiste em minimizar a energia ou função custo E definida por

$$E = \left| \sum_{i=1}^{N} s_i a_i \right| \tag{3.1}$$

sujeito ao vínculo de cardinalidade

$$\frac{1}{N} \sum_{i=1}^{N} s_i = m. \tag{3.2}$$

Podemos ainda introduzir uma norma

$$\sum_{i=1}^{N} s_i^2 = N. \tag{3.3}$$

A equação (3.1) é a energia correspondente a uma divisão ótima de um conjunto de N objetos em dois subconjuntos, sujeitos a duas condições: a primeira é que a diferença das cardinalidades de cada subconjunto seja igual a um número fixo $M = Nm$, conforme equação (3.2), e a segunda é a imposição de um vínculo esférico para a solução $s = (s_1, s_2, ..., s_N)$, equação (3.3), simplesmente para garantir que s assuma valores finitos (como $s_i = \pm 1$ esse vínculo é sempre satisfeito no problema original).

Vamos permitir que as soluções sejam contínuas, isto é, $s_i \in \Re$. Uma forma de resolvermos este problema de otimização é utilizar o método dos multiplicadores de Lagrange. Neste caso devemos minimizar a seguinte energia

$$E_L = (\sum_i s_i a_i)^2 + \lambda_1 \sum_i s_i^2 + \lambda_2 \sum_i s_i, \tag{3.4}$$

cuja evolução temporal é expressa por

$$\frac{\partial E_L}{\partial t} = \frac{\partial E_L}{\partial s_i} \frac{\partial s_i}{\partial t}. \tag{3.5}$$

Impondo a seguinte dinâmica para as variáveis de spin

$$\frac{\partial s_i}{\partial t} = -\eta \frac{\partial E_L}{\partial s_i} \tag{3.6}$$

com $\eta \geq 0$, garantimos que E_L seja minimizado, isto é, $\frac{\partial E_L}{\partial t} \leq 0$. Usando

$$\frac{\partial E_L}{\partial s_i} = 2a_i \sum_k s_k a_k + 2\lambda_1 s_i + \lambda_2 \tag{3.7}$$

podemos determinar λ_1 e λ_2 substituindo (3.7) na equação (3.6) . A seguir multiplicamos $\frac{\partial s_i}{\partial t}$ por s_i e somamos sobre i, resultando

$$\frac{1}{2}\frac{\partial}{\partial t}\sum s_i^2 = -\eta \left[2(\sum_k s_k a_k)^2 + 2\lambda_1 N + Nm\lambda_2\right] = 0. \tag{3.8}$$

Por outro lado, o vínculo (3.2) leva a

$$\frac{\partial}{\partial t}\sum s_i = 0 \tag{3.9}$$

de modo que

$$2\sum_i a_i \sum_k s_k a_k + 2\lambda_1 Nm + N\lambda_2 = 0. \tag{3.10}$$

Resolvendo as equações (3.8) e (3.10) obtemos fórmulas explícitas para os multiplicadores de Lagrange:

$$\lambda_1 = \frac{(\sum_k s_k a_k)}{N(1-m^2)}\left[m\sum_k a_k - \sum_k s_k a_k\right], \tag{3.11}$$

$$\lambda_2 = \frac{2\sum_k s_k a_k}{N(1-m^2)}\left[m\sum_k s_k a_k - \sum_k a_k\right]. \tag{3.12}$$

Finalmente, a equação dinâmica (3.6) é reescrita como

$$\begin{aligned}\frac{\partial s_i}{\partial t} &= -\eta\{2a_i\left(\sum_k s_k a_k\right) + 2s_i\frac{\sum s_k a_k}{N(1-m^2)}\left(m\sum_k a_k - \sum_k s_k a_k\right) \\ &+ \frac{2\sum_k s_k a_k}{N(1-m^2)}\left[m\sum_k s_k a_k - \sum_k a_k\right]\},\end{aligned} \tag{3.13}$$

que pode ser discretizada para fins de implementação numérica resultando em

$$s_i^{t+1} = s_i^t - 2\eta \sum_k s_k^t a_k \left[a_i + \frac{\sum_k a_k(ms_i^t - 1) + \sum_k s_k^t a_k(m - s_i^t)}{N(1-m^2)} \right]. \tag{3.14}$$

Utilizaremos a equação (3.14) na simulação numérica da dinâmica , simulando uma amostragem de realização a_k iniciais, para verificar os resultados teóricos das seções posteriores. Observe que η deve ser determinado ao longo da execução do programa, de sorte que s_i fique estável, ou seja, que ele possa convergir para um determinado valor. Uma vez obtida a solução, podemos medir o módulo da magnetização média, sua variância, bem como a energia mínima para vários valores de $distribuiçãode|\ m\ |$.

No caso sem vínculo de cardinalidade, repetimos o procedimento anterior (lembrando que λ_2 é igual a zero) e obtemos simplesmente

$$\frac{\partial s_i}{\partial t} = -\eta \sum_k s_k a_k (a_i - \frac{\sum_k s_k a_k}{N} s_i), \qquad i = 1, ...N. \tag{3.15}$$

As simulações serão realizadas para instâncias a_j estatisticamente independentes que obedecem uma distribuição uniforme no intervalo unitário.

3.2 Cálculo da função $\hat{Z}_u$

Para principiar do modelo mais simples, vamos tomar como ponto de partida o cálculo da função de partição para o caso sem vínculo de cardinalidade. Primeiro, escrevemos a energia,

$$E = (\sum_j s_j a_j)^2, \tag{3.16}$$

e, em seguida, escrevemos a função de partição, impondo apenas o vínculo da normalização (3.3),

$$\hat{Z}_u = \int_{-\infty}^{\infty} \prod_j ds_j \delta(N - \sum_j s_j^2) \int_{-\infty}^{\infty} dW\, e^{-\beta W^2} \delta(W - \sum_j s_j a_j). \tag{3.17}$$

Substituindo a delta de Dirac pela sua representação integral obtemos

$$\hat{Z}_u = \int_{-\infty}^{\infty} \frac{d\tilde{Q}d\tilde{R}dW}{4\pi^2} \int_{-\infty}^{\infty} \left[\prod_j ds_j\right] e^{i\tilde{Q}(N-\sum_j s_j^2)+i\tilde{R}(W-\sum_j s_j a_j)-\beta W^2}. \tag{3.18}$$

As integrais em s_j, W e $\tilde{R}$ são todas do tipo gaussianas, portanto ao efetuarmos essas integrais chegamos à expressão:

$$\hat{Z}_u = \frac{\pi^{N/2}}{2i\pi\left(\beta\sum_j a_j^2\right)^{1/2}} \int_{-i\infty}^{i\infty} dx\, e^{N\left(x-\frac{1}{2}\ln(x)\right)}\sqrt{x}, \tag{3.19}$$

onde fizemos a mudança de variável $x = i\tilde{Q}$. Utilizando o método de ponto de sela para avaliar a integral acima, a função de partição, no limite N$\rightarrow \infty$ e $\beta >> 1$ é escrita como

$$\hat{Z}_u = (2\pi e)^{\frac{N}{2}}\sqrt{\frac{1}{8\pi N\beta\sum_j a_j^2}}. \tag{3.20}$$

3.3 Cálculo da função $\hat{Z}_m$

Consideremos agora o modelo mais complicado em que a diferença de cardinalidade dos dois conjuntos é fixa $M = Nm$. A função de partição para este caso é escrita como

$$\hat{Z}_m = \int_{-\infty}^{\infty} \left[\prod_j ds_j\right] \delta(N-\sum_j s_j^2)\delta(M-\sum_j s_j) \int_{-\infty}^{\infty} dW\, \delta(W-\sum_j s_j a_j)\, e^{-\beta W^2} \tag{3.21}$$

Note que a primeira delta refere-se ao vínculo esférico e a segunda ao vínculo de cardinalidade. Em seguida, escrevemos as deltas de Dirac na forma integral,

$$\hat{Z}_m = \int_{-\infty}^{\infty} \frac{d\tilde{Q}d\tilde{P}d\tilde{R}dW}{8\pi^3} \prod_j ds_j\, e^{i\tilde{Q}(N-\sum_j s_j^2)+i\tilde{P}(M-\sum_j s_j)+i\tilde{R}(W-\sum_j s_j a_j)-\beta W^2}. \tag{3.22}$$

Novamente não exibiremos os detalhes das integrais gaussinas em s_j, W e $\tilde{P}$. O resultado dessas integrais leva-nos a seguinte expressão

$$\hat{Z}_m = \sqrt{\frac{4}{N\beta}} \int_{-\infty}^{\infty} \frac{d\widetilde{Q}d\widetilde{R}}{8\pi^2}\, e^{-\frac{\widetilde{R}^2}{4\beta}+i\frac{N\widetilde{R}^2\langle a^2\rangle}{4\widetilde{Q}}-i\frac{N}{4\widetilde{Q}}\widetilde{R}^2\langle a\rangle^2-iNm\widetilde{R}\langle a\rangle-iNm^2\widetilde{Q}+i\widetilde{Q}N}\sqrt{i\widetilde{Q}}\left(\frac{\pi}{i\widetilde{Q}}\right)^{\frac{N}{2}}. \tag{3.23}$$

onde definimos $\langle a\rangle = \frac{\sum_j a_j}{N}$ e $\langle a^2\rangle = \frac{\sum_j a_j^2}{N}$. A integral em $\widetilde{R}$ é calculada de forma simples,

$$\int_{-\infty}^{\infty} d\widetilde{R}\, e^{-\frac{\widetilde{R}^2}{4\beta}+i\frac{N\widetilde{R}^2\langle a^2\rangle}{4\widetilde{Q}}-i\frac{N}{4\widetilde{Q}}\widetilde{R}^2\langle a\rangle^2-iNm\widetilde{R}\langle a\rangle} = \int_{-\infty}^{\infty} d\widetilde{R}\, e^{-\frac{\widetilde{R}^2}{4\beta}-\frac{N\widetilde{R}^2\sigma^2}{4x}-iNm\widetilde{R}\langle a\rangle} \tag{3.24}$$

onde introduzimos a notação $\sigma^2 = \langle a^2\rangle - \langle a\rangle^2$ e a nova variável $x = i\widetilde{Q}$. Daí

$$\int_{-\infty}^{\infty} d\widetilde{R}\, e^{-\gamma\widetilde{R}^2-iNm\widetilde{R}\langle a\rangle} = \sqrt{\frac{\pi}{\gamma}}\, e^{-\frac{N^2m^2\langle a\rangle^2}{4\gamma}} \tag{3.25}$$

com $\gamma = \frac{1}{4}\left(\frac{1}{\beta}+\frac{N}{x}\sigma^2\right)$. Finalmente escrevemos

$$\hat{Z}_m = \sqrt{\frac{\pi}{\beta N}}\frac{\pi^{\frac{N}{2}}}{4i\pi^2}\int_{-i\infty}^{i\infty} dx\, e^{-Nm^2x+Nx-\frac{N^2m^2\langle a\rangle^2}{4\gamma}-\frac{N}{2}\ln(x)}\sqrt{\frac{x}{\gamma}}. \tag{3.26}$$

Para calcular esta última integral, vamos recorrer novamente ao método de ponto de sela. Assim vamos simplesmente reescrever a integral na forma padrão e apresentar o resultado final. A solução geral é dada por

$$\hat{Z}_m = \sqrt{\frac{\pi}{\beta N}}\frac{\pi^{\frac{N}{2}}}{4i\pi^2}\int_{-i\infty}^{i\infty} dx\, e^{NF(x)}g(x) \tag{3.27}$$

$$= \sqrt{\frac{\pi}{\beta N}}\frac{\pi^{\frac{N}{2}}}{4i\pi^2}\sqrt{\frac{-2\pi}{NF''(x_{ps})}}\, e^{NF(x_{ps})}g(x_{ps}), \tag{3.28}$$

onde

$$F(x) = -m^2x + Nx - \frac{Nm^2\langle a\rangle^2}{\left(\frac{1}{\beta}+\frac{N}{x}\sigma^2\right)} - \frac{1}{2}\ln(x), \tag{3.29}$$

$$g(x) = \sqrt{\frac{x}{\frac{1}{4}\left(\frac{1}{\beta} + \frac{N}{x}\sigma^2\right)}} \tag{3.30}$$

e x_{ps} é tal que maximiza $F(x)$. Os cálculos devem ser feitos no limite $\beta \to \infty$. Observe que podemos desprezar o termo $\frac{1}{\beta}$ se $x_{ps} \ll \beta$, caso contrário temos que levar em conta esta contribuição. Vamos considerar a seguir esses dois casos possíveis separadamente.

Caso I: $\frac{x_{ps}}{\beta} \to 0$

Neste caso as equações (3.29) e (3.30) ficam escritas como

$$F(x_{ps}) = x_{ps}(1 - m^2) - x_{ps}\frac{m^2 \langle a \rangle^2}{\sigma^2} - \ln(x_{ps}) \tag{3.31}$$

$$g(x_{ps}) = \frac{2x_{ps}}{\sqrt{N\sigma^2}}, \tag{3.32}$$

onde o ponto de sela x_{ps} é obtido resolvendo-se a equação

$$(1 - m^2) - \frac{m^2 \langle a \rangle^2}{\sigma^2} - \frac{1}{2x_{ps}} = 0. \tag{3.33}$$

Logo,

$$x_{ps} = \frac{1}{2[1 - m^2(1 + \frac{\langle a \rangle^2}{\sigma^2})]}. \tag{3.34}$$

Uma vez determinado x_{ps}, podemos agora calcular explicitamente $F(x_{ps})$ e $F''(x_{ps})$, obtendo

$$F(x_{ps}) = \frac{1}{2}\{1 + \ln 2 + \ln\left[1 - m^2(1 + \frac{\langle a \rangle^2}{\sigma^2})\right]\}, \tag{3.35}$$

$$F''(x_{ps}) = 2\left[1 - m^2(1 + \frac{\langle a \rangle^2}{\sigma^2})\right]^2. \tag{3.36}$$

A função de partição (3.28) neste caso é então

$$\hat{Z}_m = \frac{(2e\pi)^{\frac{N}{2}}}{4N\sigma\pi}\sqrt{\frac{1}{N\beta}}\left[1 - m^2(1 + \frac{\langle a\rangle^2}{\sigma^2})\right]^{-2} e^{\frac{N}{2}\ln[1-m^2(1+\frac{\langle a\rangle^2}{\sigma^2})]} \tag{3.37}$$

Note que se lançarmos mão da propriedade de automediância no limite N$\rightarrow$ ∞ para calcular os momentos da variável aleatória **a**, isto é,

$$\lim_{N\to\infty}\frac{1}{N}\sum_i a_i^\alpha = \int_0^1 daa^\alpha = \frac{1}{\alpha+1}. \tag{3.38}$$

então $\frac{\langle a\rangle^2}{\sigma^2} = 3$ e, portanto, a equação (3.37) só é válida para $m < \frac{1}{2}$.

<u>Caso II:</u> $\frac{x_{ps}}{\beta} \nrightarrow 0$.

Neste caso não podemos desprezar o termo $\frac{1}{\beta}$, então as equações escritas como

$$F(x_{ps}) = x_{ps}(1-m^2) - \frac{Nm^2\langle a\rangle^2}{\frac{1}{\beta}+\frac{N\sigma^2}{x_{ps}}} - \frac{1}{2}\ln(x_{ps}), \tag{3.39}$$

$$g(x_{ps}) = \sqrt{\frac{4x_{ps}}{\frac{1}{\beta}+\frac{N\sigma^2}{x_{ps}}}}, \tag{3.40}$$

e o ponto de sela x_{ps} é encontrado resolvendo-se a equação $F'(x)|_{x=x_{ps}}$. O resultado é

$$x_{ps} = N\beta\left[\frac{\sigma m\langle a\rangle}{\sqrt{1-m^2}} - \sigma^2\right]. \tag{3.41}$$

Novamente, para obter $\hat{Z}_m$ basta calcular $F(x_{ps})$, $g(x_{ps})$ e $F''(x_{ps})$. O resultado final é o seguinte

$$F(x_{ps}) = -N\beta\left[m\langle a\rangle - \sigma\sqrt{1-m^2}\right]^2 - \frac{1}{2}\ln\left[N\beta\left(\frac{\sigma m\langle a\rangle}{\sqrt{1-m^2}} - \sigma^2\right)\right], \tag{3.42}$$

$$g(x_{ps}) = \frac{2N\beta}{\sqrt{N\sigma m\langle a\rangle}}\left[\frac{\sigma m\langle a\rangle}{\sqrt{1-m^2}} - \sigma^2\right], \tag{3.43}$$

$$F''(x_{ps}) = \frac{2\sqrt{1-m^2}}{\sigma m \langle a \rangle}. \tag{3.44}$$

Daí, a função de partição (3.28) é finalmente escrita da seguinte forma:

$$\hat{Z}_m = \frac{\pi^{\frac{N}{2}}}{N\pi\sigma}\sqrt{\frac{\beta}{N}}\left[\frac{\sigma m \langle a \rangle}{\sqrt{1-m^2}} - \sigma^2\right] e^{NF(x_{ps})}, \tag{3.45}$$

válida para $m \geq \frac{1}{2}$

3.4 Energia média

A energia média é calculada facilmente através da definição

$$E_m = -\lim_{\beta\to\infty} \frac{\partial}{\partial\beta} \ln \hat{Z}_m. \tag{3.46}$$

No intervalo em que $m < \frac{1}{2}$ substituímos $\hat{Z}_m$ por (3.37). Como $\hat{Z}_m \propto \beta^{-1/2}$, podemos desprezar todas as constantes e efetuar um cálculo simples

$$E_m = -\lim_{\beta\to\infty} \frac{\partial}{\partial\beta}\left[\frac{1}{2}\ln\beta\right] = 0. \tag{3.47}$$

Por outro lado, quando $x_{ps} \to \infty$, usamos (3.45) para obter a energia correspondente ao intervalo $m > \frac{1}{2}$. Observe que somente $F(x_{ps})$ depende de β o que resultará numa energia média não nula

$$\begin{aligned} E_m &= -\lim_{\beta\to\infty} \frac{\partial}{\partial\beta}\left[NF(x_{ps})\right] \\ &= N^2(m\langle a \rangle - \sigma\sqrt{1-m^2})^2. \end{aligned} \tag{3.48}$$

A energia calculada nesta seção, bem como aquela usada para derivar a dinâmica da relaxação dada pela equação (3.14) foram definidas como sendo o quadrado da energia exata para a versão inteira (equação (2.66)) derivada no capítulo anterior. Obviamente, para podermos compará-las com a expressão antiga não podemos esquecer de tirar a raiz quadrada, portanto redefinimos a energia relaxada como $E = \sqrt{E_m}$. Fazendo isto, verificamos que a energia média no problema relaxado, expressas

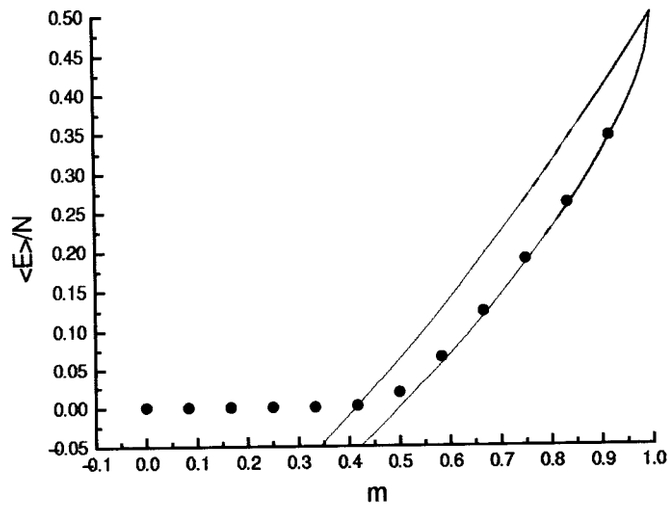


Figura 3.1: A linha vermelha é o resultado do cálculo exato do problema inteiro derivado no capítulo anterior. A linha verde é o resultado teórico do problema relaxado. Os pontos foram obtidos da simulação utilizando o algoritmo da seção 1.

pelas equações (3.47) e (3.48), concordam plenamente com a simulação conforme ilustra a figura 3.1. Alternativamente, podemos dizer que o algoritmo de relaxação funciona muito bem. A curva em vermelho é o resultado analítico exato para a configuração de Ising, enquanto a curva em verde é o resultado analítico obtido nesta seção para o problema relaxado. Os pontos foram obtidos fazendo-se 2000 experimentos com $N = 24$. Podemos verificar que a versão do problema de otimização contínuo fornece um limitante inferior para o problema de otimização inteiro, conforme o esperado.

3.5 Distribuição de probabilidade das cardinalidades

Para uma mesma realização dos números reais $\{a_i\}$, a dinâmica relaxada sem vínculo de cardinalidade (3.15) produz mínimos globais distintos que possuem diferentes valores de M (diferenças de cardinalidades). Nesta seção, vamos calcular a distribuição de probabilidade ou a freqüência dos mínimos globais com um dado valor de M.

O que desejamos encontrar é a função densidade de probabilidade de uma variável aleatória $Y = h(\mathbf{s})$, a qual, por sua vez, é uma função de um conjunto de variáveis aleatórias $\mathbf{s}$, cuja distribuição de probabilidade conjunta $f(\mathbf{s})$ é conhecida. A densidade de probabilidade de Y, denotada por $p(Y)$, é definida por [45]

$$p(Y) = \int_{-\infty}^{\infty} \left[\prod_i ds_i\right] \delta(Y - h(\mathrm{s}))f(\mathrm{s}). \tag{3.49}$$

No problema que estamos tratando, temos $y = |m|$; porém, por simplicidade, vamos calcular primeiro a distribuição para m. A função h carrega a informação denotada por m, ou seja,

$$h(\mathrm{s}) = \frac{1}{N}\sum_i s_i. \tag{3.50}$$

Já a distribuição conjunta dos s_i é a distribuição de Gibbs

$$f(\mathrm{s}) = \frac{e^{-\beta(\sum s_i a_i)^2}}{\hat{Z}_u}, \tag{3.51}$$

onde $\hat{Z}_u$ é dada pela equação (3.20). Assim, no caso relaxado, onde $s_i \in (-\infty, \infty)$, podemos escrever $p(m)$ (já normalizado) como

$$p(m) = \lim_{T\to 0} N \left\langle \left\langle \delta(Nm - \sum_j s_j) \right\rangle_T \right\rangle_a . \tag{3.52}$$

Usando a definição de média térmica $\langle ... \rangle_T$, ou seja, aquela que usa a distribuição de Gibbs, esta equação é reescrita da seguinte forma

$$p(m) = \lim_{\beta\to\infty} N \left\langle \frac{\int_{-\infty}^{\infty} (\prod_j ds_j)\delta(N - \sum s_j^2)\, e^{-\beta(\sum s_j a_j)^2} \delta(Nm - \sum s_j)}{\int_{-\infty}^{\infty} (\prod_j ds_j)\delta(N - \sum s_j^2)\, e^{-\beta(\sum s_j a_j)^2}} \right\rangle_a , \tag{3.53}$$

onde podemos observar que o numerador é $\hat{Z}_m$ enquanto o denominador é justamente $\hat{Z}_u$. Daí, utilizando os resultados obtidos anteriormente escrevemos

$$p(m) = \lim_{\beta\to\infty} N \left\langle \frac{\hat{Z}_m}{\hat{Z}_u} \right\rangle . \tag{3.54}$$

A rigor, para $m < \frac{1}{2}$, substituímos (3.20) e (3.37) na equação acima, obtendo

$$p(m) = \lim_{\beta\to\infty} N \sqrt{\frac{N < a^2 >}{}}\, e^{-\frac{N < a^2 >}{2\sigma^2} m^2}; \tag{3.55}$$

enquanto que para $m > \frac{1}{2}$ substituímos (3.20) e (3.45), encontrando

$$p(m) = 0,$$

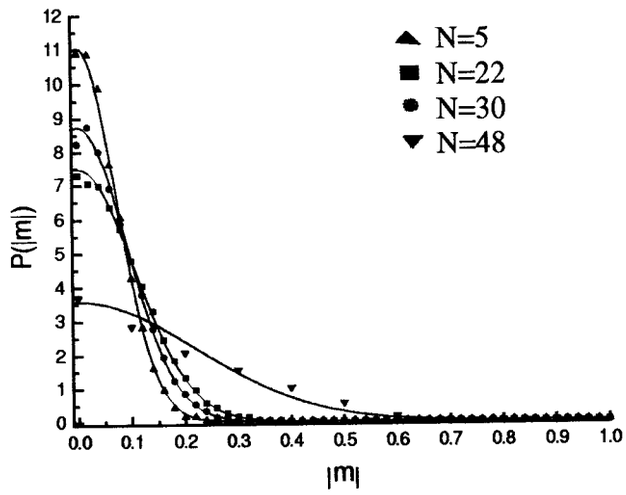


Figura 3.2: Distribuição do mínimos globais como função do vínculo de cardinalidade m para N fixo. Os pontos foram obtidos através de simulação.

pois $\hat{Z}_m = 0$ neste limite. A fim de eliminarmos a dependência em N vamos fazer uma mudança de variável passando de m, que assume valores entre -1 e 1, para $\tilde{m}$ definido como

$$\tilde{m} = \sqrt{N}m, \tag{3.56}$$

assumindo então valores entre -∞ e ∞. A maneira correta de efetuarmos esta mudança de variável é através da definição

$$p(\tilde{m}) = \int_{-\infty}^{\infty} dm\ p(m)\delta(\tilde{m} - \sqrt{N}m) = \sqrt{\frac{< a^2 >}{2\pi\sigma^2}}\, e^{-\frac{< a^2 >}{2\sigma^2}\tilde{m}^2}, \tag{3.57}$$

que está corretamente normalizado. Podemos agora derivar facilmente a distribuição de $y = \mid \tilde{m} \mid$:

$$\tilde{p}(y) = \int_{-\infty}^{\infty} p(\tilde{m})\delta(y - \mid \tilde{m} \mid)d\tilde{m} = 2\int_{0}^{\infty} p(\tilde{m})\delta(y - \tilde{m})d\tilde{m} = \sqrt{\frac{8}{\pi}}\, e^{-2y^2} \tag{3.58}$$

ou

$$\tilde{p}(\mid \tilde{m} \mid) = \sqrt{\frac{8}{\pi}}\, e^{-2|\tilde{m}|^2}. \tag{3.59}$$

Realizamos simulações para averigüar esse resultado, conforme ilustrado pela figura (3.2). Novamente a concordância entre a previsão teórica (linhas cheias) e os resultados numéricos (símbolos) é excelente. A probabilidade de encontrarmos partições ótimas cai drasticamente com m e vai a zero para $m > \frac{1}{2}$. Para entendermos melhor as implicações deste resultado vamos calcular os momentos da variável aleatória m e, em seguida, estudar a correlação entre os estados com energia ótima.

3.5.1 Momentos

Agora podemos calcular facilmente os momentos da distribuição (3.59). Este cálculo não oferece nenhuma dificuldade, veja:

$$<\mid m \mid^k>= 2\int_0^\infty d\tilde{m} \mid \tilde{m} \mid^k \sqrt{\frac{8}{\pi}}\, e^{-2|\tilde{m}|^2} = \sqrt{\frac{2}{\pi}}\frac{\Gamma[(\frac{k+1}{2})]}{(2)^{\frac{k+1}{2}}}. \tag{3.60}$$

Daí, a média e a variância são dadas, respectivamente, por

$$\langle \mid m \mid \rangle = \frac{1}{\sqrt{2N\pi}}, \qquad \sigma^2_{|m|} = \frac{1}{2N}(\frac{1}{2} - \frac{1}{\pi}). \tag{3.61}$$

Uma grandeza importante é a razão $r_m = \sqrt{\sigma^2_{|m|}} / \langle \mid m \mid \rangle$, pois, como já vimos, ela determina a automediância da variável aleatória $\mid m \mid$. Nas figuras 3.3a e 3.3b apresentamos o resutado de experimentos numéricos para estimar a dependência com N de $\langle \mid m \mid \rangle$ e r_m, respectivamente, para três tipos de configurações: o mínimo global do problema original obtido através de busca exaustiva no espaço de configuração de Ising para N$\leq$ 26 (símbolos verdes), o mínimo global do problema relaxado obtido com a dinâmica (3.15) (símbolos vermelhos) e, finalmente, a configuração obtida usando-se o método da diferença de Karmarkar e Karp (símbolos azuis), discutido no apêndice C. Observe que há um acordo muito bom entre as estimativas numéricas e a predições analíticas. Note ainda que $\mid m \mid$ não é automediante mesmo no limite de N grande, pois $r_m \rightarrow 0.5$, significando que os valores de $\mid m \mid$ associados às configurações em estudo dependem da realização específica das

variáveis aleatórias $\{a_i\}$.

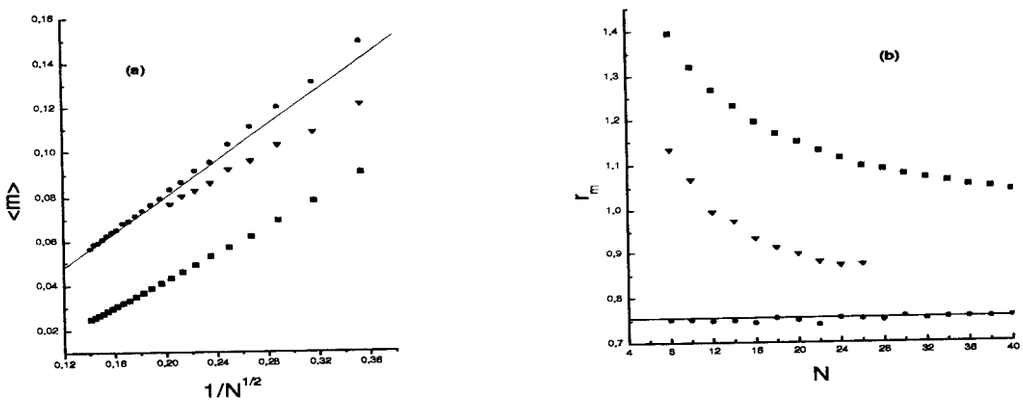


Figura 3.3: (a)A média do módulo da cardinalidade m dos mínimos globais em função de N. (b) A razão entre o desvio padrão e a média da grandeza m. As linhas contínuas são estimativas teóricas. Os pontos em vermelho são o resultado dos experimentos computacionais via algoritmo da relaxação. Os pontos azuis foram obtidos aplicando-se o algoritmo da diferença de Karmakar e Karp (o algoritmo apresentado no apêndice C). Os pontos verdes são o resultado da busca exaustiva para o problema inteiro.

3.6 Correlação

Uma propriedade interessante a ser investigada é a correlação entre os mínimos globais em função da diferença de cardinalidade $m = \frac{M}{N}$. Matematicamente a correlação entre duas configurações de spins s_{1j} e s_{2j} cada qual com diferença de cardinalidade m é definida como

$$Cor(m) = \left\langle \left\langle \frac{1}{N} \sum_j s_{1j} s_{2j} \right\rangle_T \right\rangle_a . \tag{3.62}$$

Para efetuar esse cálculo primeiro definimos uma energia auxiliar

$$E_{cor} = (\sum_j s_{1j} a_j)^2 - (\sum_j s_{2j} a_j)^2 - h \sum_j s_{1j} s_{2j} \tag{3.63}$$

e a função de partição associada,

$$Z_{cor} = \int \prod_{\alpha=1}^{2} \left[\prod_{j=1}^{N} ds_{\alpha j} \right] \delta(N - \sum_j s_{\alpha j}^2) \delta(M - \sum_j s_{\alpha j})\, e^{-\beta E_{cor}}. \tag{3.64}$$

Como estamos interessados em calcular (3.62) basta efetuar

$$Cor(m) = -\frac{1}{N} \lim_{\beta \to \infty} \frac{1}{\beta} \frac{\partial}{\partial h} \ln(Z_{cor}) \,|_{h=0} \,. \tag{3.65}$$

Devemos simplificar o expoente contendo o termo quadrático $(\sum_j s_{\alpha j} a_{\alpha j})^2$ introduzimos novas deltas de Dirac e, em seguida, escrever as deltas na sua representação integral. Daí,

$$\begin{aligned} Z_{cor} &= \int \prod_{\alpha=1}^{2} \left[\frac{d\tilde{Q}_\alpha d\tilde{P}_\alpha dW_\alpha d\widetilde{W}_\alpha}{(2\pi)^3} e^{iN\tilde{Q}_\alpha + iM\tilde{P}_\alpha + iW_\alpha \widetilde{W}_\alpha - \beta W_\alpha^2} \right] \\ &\times \int_{-\infty}^{\infty} \left(\prod_j ds_{2j} \right) e^{-i\tilde{Q}_2 \sum_j s_{2j}^2 - i\tilde{P}_2 \sum_j s_{2j} - i\tilde{W}_2 \sum_j s_{2j} a_j + \sum_j \frac{(\tilde{P}_1 + \tilde{W}_1 a_j)}{2\tilde{Q}_1} \beta h s_{j2}} \\ &\times \int_{-\infty}^{\infty} \left(\prod_j ds_{1j} \right) e^{-i\tilde{Q}_1 \sum_j s_{1j}^2 - i\tilde{P}_1 \sum_j s_{1j} - i\tilde{W}_1 \sum_j s_{1j} a_j - \beta h \sum_j s_{1j} s_{2j}} \end{aligned} \tag{3.66}$$

Primeiro efetuamos a integração sobre s_{1j}

$$\begin{aligned} \Lambda_1 &= \int_{-\infty}^{\infty} \left(\prod_j ds_{1j} \right) e^{-i\tilde{Q}_1 \sum_j s_{1j}^2 - i\tilde{P}_1 \sum_j s_{1j} - i\tilde{W}_1 \sum_j s_{1j} a_j - \beta h \sum_j s_{1j} s_{2j}} \\ &= \prod_j \{ \int_{-\infty}^{\infty} ds_{1j}\, e^{-i\tilde{Q}_1 (s_{1j} + \frac{\tilde{P}_1 + \tilde{W}_1 a_j - i\beta h s_{2j}}{2\tilde{Q}_1})^2 + i \frac{(\tilde{P}_1 + \tilde{W}_1 a_j - i\beta h s_{2j})^2}{4\tilde{Q}_1}} \} \\ &= \prod_j \sqrt{\frac{\pi}{i\tilde{Q}_1}}\, e^{i \frac{(\tilde{P}_1 + \tilde{W}_1 a_j - i\beta h s_{2j})^2}{4\tilde{Q}_1}}. \end{aligned} \tag{3.67}$$

No que segue vamos desprezar os termos que dependem de h^2. Devemos usar o resultado anterior para podermos integrar em s_{2j} devido ao acoplamento dos termos s_1 e s_2:

$$\Lambda_2 = \int_{-\infty}^{\infty} \left(\prod_j ds_{2j} \right) e^{-i\tilde{Q}_2 \sum_j s_{2j}^2 - i\tilde{P}_2 \sum_j s_{2j} - i\tilde{W}_2 \sum_j s_{2j} a_j + \sum_j \frac{(\tilde{P}_1 + \tilde{W}_1 a_j)}{2\tilde{Q}_1} \beta h s_{j2}}$$

$$= \prod_j \int_{-\infty}^{\infty} ds_{2j}\, e^{-i\tilde{Q}_2 s_{2j}^2 - i(\tilde{P}_2+\tilde{W}_2 a_j + i\frac{(\tilde{P}_1+\tilde{W}_1 a_j)}{2\tilde{Q}_1}\beta h)s_{j2}}$$

$$= \prod_j \sqrt{\frac{\pi}{i\tilde{Q}_2}}\, e^{-\frac{1}{4i\tilde{Q}_2}(\tilde{P}_2+\tilde{W}_2 a_j + i\frac{(\tilde{P}_1+\tilde{W}_1 a_j)}{2\tilde{Q}_1}\beta h)^2} \tag{3.68}$$

$$= \prod_j \sqrt{\frac{\pi}{i\tilde{Q}_2}}\, e^{-\frac{1}{4i\tilde{Q}_2}(\tilde{P}_2+\tilde{W}_2 a_j)^2 - \frac{1}{4\tilde{Q}_1\tilde{Q}_2}(\tilde{P}_1+\tilde{W}_1 a_j)(\tilde{P}_2+\tilde{W}_2 a_j)\beta h}. \tag{3.69}$$

A integral em W_α ($\alpha = 1, 2$) também pode ser calculada facilmente,

$$\int_{-\infty}^{\infty} dW_\alpha\, e^{-\beta W^2 - iW_\alpha \tilde{W}_\alpha} = \sqrt{\frac{\pi}{\beta}}\, e^{-\frac{\tilde{W}_\alpha^2}{4\beta}}.$$

Com esses resultados, a função de partição Z_{cor} fica escrita como

$$Z_{cor} = \frac{\pi}{\beta}\int_{-\infty}^{\infty} \frac{d\tilde{Q}_2 d\tilde{P}_2 d\tilde{W}_2}{(2\pi)^3}\, e^{iN\tilde{Q}_2 + iM\tilde{P}_2 - \frac{\tilde{W}_2^2}{4\beta}} \int_{-\infty}^{\infty} \frac{d\tilde{Q}_1 d\tilde{P}_1 d\tilde{W}_1}{(2\pi)^3}\, e^{iN\tilde{Q}_1 + iM\tilde{P}_1 - \frac{\tilde{W}_1^2}{4\beta}}$$
$$\times \prod_j \left[\sqrt{\frac{\pi^2}{i^2\tilde{Q}_1\tilde{Q}_2}}\, e^{-\frac{(\tilde{P}_1+\tilde{W}_1 a_j)^2}{4i\tilde{Q}_1} - \frac{1}{4\tilde{Q}_1\tilde{Q}_2}(\tilde{P}_1+\tilde{W}_1 a_j)(\tilde{P}_2+\tilde{W}_2 a_j)\beta h - \frac{1}{4i\tilde{Q}_2}(\tilde{P}_2+\tilde{W}_2 a_j)^2}\right]. \tag{3.70}$$

Voltando em (3.65) e usando (3.70), obtemos a seguinte expressão para a correlação,

$$Cor(m) = \frac{1}{N}\lim_{\beta\to\infty}\frac{1}{4\tilde{Q}_1\tilde{Q}_2}\sum_{j=1}^{N}(\tilde{P}_1 + \tilde{W}_1 a_j)(\tilde{P}_2 + \tilde{W}_2 a_j) \tag{3.71}$$

onde $\tilde{P}_1$, $\tilde{P}_2$, $\tilde{Q}_1$, $\tilde{Q}_2$, $\tilde{W}_1$ e $\tilde{W}_2$ são os pontos de sela de Z_{cor} calculados para $h = 0$, que denotamos por Z_{cor}^0

$$Z_{cor}^0 = \frac{\pi}{\beta}\int_{-\infty}^{\infty}\prod_{\alpha=1}^{2}\frac{d\tilde{Q}_\alpha d\tilde{P}_\alpha d\tilde{W}_\alpha}{(2\pi)^3}\, e^{iN\tilde{Q}_\alpha + iM\tilde{P}_\alpha - \frac{\tilde{W}_\alpha^2}{4\beta}}\prod_j\left[\sqrt{\frac{\pi}{i\tilde{Q}_\alpha}}\, e^{i\frac{(\tilde{P}_\alpha+\tilde{W}_\alpha a_j)^2}{4\tilde{Q}_\alpha}}\right]$$
$$= \frac{\pi}{\beta}\left\{\int_{-\infty}^{\infty}\frac{d\tilde{Q}d\tilde{P}d\tilde{W}}{(2\pi)^3}\, e^{iN\tilde{Q} + iM\tilde{P} - \frac{\tilde{W}^2}{4\beta}}\prod_j\left[\sqrt{\frac{\pi}{i\tilde{Q}}}\, e^{-\frac{1}{4i\tilde{Q}}(\tilde{P}+\tilde{W}a_j)^2}\right]\right\}^2. \tag{3.72}$$

Os pontos de sela são obtidos maximizando-se o argumento Ξ das exponenciais em (3.72), a saber,

$$\begin{aligned} \mathcal{S} &= i\tilde{Q} + im\tilde{P} - \frac{\tilde{W}^2}{4N\beta} - \sum_{j=1}^{N} \frac{(\tilde{P} + \tilde{W}a_j)^2}{4iN\tilde{Q}} - \frac{1}{2}\ln(i\tilde{Q}) \\ &= x + my + \frac{z^2}{4N\beta} + \frac{1}{4x}\left[y^2 + z^2\left\langle a^2\right\rangle + 2yz\left\langle a\right\rangle\right] - \frac{1}{2}\ln(x) \end{aligned} \tag{3.73}$$

onde fizemos a mudança de variáveis $x = i\tilde{Q}$, $y = i\tilde{P}$, $z = i\tilde{W}$, enquanto $\langle a\rangle$ e $\langle a^2\rangle$ são os momentos da variável aleatória **a**. A função de correlação em termos destas novas variáveis fica

$$Cor(m) = \lim_{\beta\to\infty} \frac{1}{4x^2}\left[y^2 + z^2\left\langle a^2\right\rangle + 2yz\left\langle a\right\rangle\right]. \tag{3.74}$$

Os ponto de sela são obtidos resolvendo-se o sistema de 3 equações obtidas derivando-se $\mathcal{S}$ em relação a x, y e z, e igualando a zero. Daí,

$$1 - \frac{1}{4x^2}\left[y^2 + z^2\left\langle a^2\right\rangle + 2yz\left\langle a\right\rangle\right] - \frac{1}{2x} = 0; \tag{3.75}$$

$$-\frac{z}{2N\beta} - \frac{z\left\langle a^2\right\rangle + y\left\langle a\right\rangle}{2x} = 0; \tag{3.76}$$

$$m - \frac{y + z\left\langle a\right\rangle}{2x} = 0. \tag{3.77}$$

Resolvendo (3.76) e (3.77) em termos de x encontramos

$$y = 2xm + \frac{2xm\left\langle a\right\rangle^2}{\sigma^2 + \frac{x}{\beta}}, \tag{3.78}$$

$$z = -\frac{2xm\left\langle a\right\rangle}{\sigma^2 + \frac{x}{\beta}}. \tag{3.79}$$

De (3.75) podemos concluir então que a correlação é dada simplesmente por

$$Cor(m) = \lim_{\beta \to \infty} 1 - \frac{1}{2x}. \tag{3.80}$$

Novamente há dois casos a serem considerados para que possamos tomar o limite $\beta \to \infty$ corretamente. Vamos primeiramente considerar o caso em que x é finito. No limite $\beta \to \infty$, as expressões (3.78) e (3.79) reduzem-se a

$$y = 2xm + \frac{2xm\langle a\rangle^2}{\sigma^2}, \tag{3.81}$$

$$z = -\frac{2xm\langle a\rangle}{\sigma^2}. \tag{3.82}$$

Substituindo as equações acima em (3.75) encontramos

$$1 - m^2\left[(1+\frac{\langle a\rangle^2}{\sigma^2})^2 - \frac{\langle a\rangle^2}{\sigma^4}\langle a^2\rangle - 2\frac{\langle a\rangle^2}{\sigma^2}(1+\frac{\langle a\rangle^2}{\sigma^2})\right] - \frac{1}{2x} = 0. \tag{3.83}$$

Usando a propriedade de automediância, ou seja, supondo

$$\langle a\rangle = \frac{1}{2}, \qquad \langle a^2\rangle = \frac{1}{3} \;\; e \;\; \sigma^2 = \frac{1}{12}, \tag{3.84}$$

a equação (3.83) para a variável x fica

$$1 - 4m^2 - \frac{1}{2x} = 0, \tag{3.85}$$

onde tiramos que $m < \frac{1}{2}$, já que x deve ser positivo e finito (veja a equação (3.73)). De (3.80) e (3.85) concluímos que, neste regime, a correlação é dada por

$$Cor(m) = \lim_{\beta \to \infty} 1 - \frac{1}{2x} = 4m^2. \tag{3.86}$$

Note que a correlação alcança seu valor máximo para $m = \frac{1}{2}$, indicando que o mínimo global é único a partir de $m = \frac{1}{2}$.

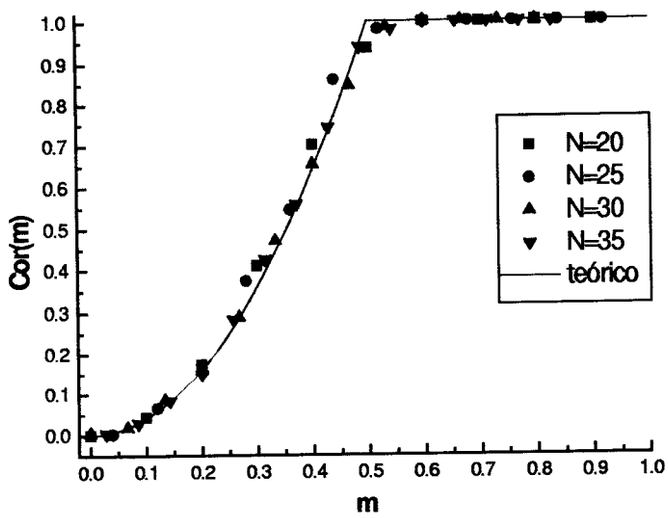


Figura 3.4: A curva continua mede a correlação entre dois estados s_i e s_j com energia mínima (ótimo global) com a mesma cardinalidade (M fixo) obtida analiticamente. Os pontos resultam da aplicação do algoritmo da relaxação para vários valores de N.

$$Cor(m) = \lim_{\beta \to \infty} 1 - \frac{1}{2x} = 1. \tag{3.87}$$

A figura 3.4 mostra as estimativas numéricas feitas para a correlação representadas pelos símbolos para diferentes valores de N. A linha sólida é a predição teórica derivada acima e o que observamos foi que ela ajustou perfeitamente os dados obtidos nas simulações.

3.7 Cálculo da energia de integralização

Nesta seção vamos usar técnicas de Mecânica Estatística para avaliar o desempenho médio de uma heurística simples, aqui denominada Heurística de Integralização. A energia de integralização é calculada por meio do seguinte critério: quando s_j for negativo, trocamos o respectivo valor por -1 e quando s_j for positivo, trocamos por 1. Denotando esta operação por $sign(s_j)$, a energia procurada fica dada por

$$\langle E_I \rangle = \lim_{T \to 0} \left\langle \left\langle | \sum a_j sign(s_j) | \right\rangle_T \right\rangle_a, \tag{3.88}$$

onde a média térmica $\langle ... \rangle_T$ é feita com a distribuição de probabilidade de Gibbs, isto é,

$$P(s) = \frac{e^{-\beta E(s)}}{Z_u}, \tag{3.89}$$

com $E(s)$ dada em (3.16) . Alternativamente, defindo a energia auxiliar

$$E_{aux}(s) = (\sum s_j a_j)^2 - h\beta \mid \sum a_j sign(s_j) \mid \tag{3.90}$$

e a função de partição associada

$$Z_I = \int_{-\infty}^{\infty} \left(\prod_j ds_j \right) \delta(N - \sum s_j^2)\, e^{-\beta E_{aux}}, \tag{3.91}$$

a energia de integralização pode ser recuperada através do seguinte procedimento

$$\langle E_I \rangle = - \lim_{\beta \to \infty} \frac{1}{\beta} \frac{\partial}{\partial h} \langle \ln Z_I \rangle \quad |_{h=0} \,. \tag{3.92}$$

Para obtermos essa energia devemos calcular explicitamente a função de partição. Com o auxílio de mais uma delta de Dirac reescrevemos Z_I como

$$\begin{aligned} Z_I &= \int_{-\infty}^{\infty} \left[\prod_j ds_j \right] \delta(N - \sum_j s_j^2) \int_{-\infty}^{\infty} dW dy\, \delta(W - \sum_j s_j a_j) \delta\left(y - \sum_j a_j sign(s_j) \right) e^{-\beta W^2 - \beta h |y} \\ &= \int_{-\infty}^{\infty} \frac{d\tilde{Q} d\tilde{R} dW d\tilde{y} dy}{(2\pi)^3} \int_{-\infty}^{\infty} \prod_j ds_j\, e^{-i\tilde{Q}\sum s_j^2 - i\tilde{R}\sum s_j a_j - i\tilde{y}\sum_j sign(s_j) a_j}. \end{aligned} \tag{3.93}$$

Para efetuar a integral em s_j temos que lembrar de dividir a integral em duas partes: uma referente a parte positiva $s_j > 0$ e $sign(s_j) = 1$ e a outra referente a parte negativa $s_j < 0$ e $sign(s_j) = -1$. Temos então

$$\int_{-\infty}^{\infty} \prod_j ds_j\, e^{-i\tilde{Q}\sum s_j^2 - i\tilde{R}\sum s_j a_j - i\tilde{y}\sum sign(s_j) a_j} = \prod_j \sqrt{\frac{\pi}{\tilde{Q}}} \cos(\tilde{y} a_j)\, e^{\frac{i\tilde{R}^2 a_j^2}{4\tilde{Q}}}. \tag{3.94}$$

Assim podemos reescrever a função de partição como

$$Z_I = \int_{-\infty}^{\infty} \frac{d\widetilde{Q}}{(2\pi)^3} e^{iN\widetilde{Q}} (\frac{\pi}{\widetilde{Q}})^{\frac{N}{2}} \int_{-\infty}^{\infty} d\widetilde{R}dW\, e^{-\beta W^2 + iW\widetilde{R} + \frac{i\widetilde{R}^2 \sum a_j^2}{4\widetilde{Q}}} \int_{-\infty}^{\infty} d\tilde{y}dy\, e^{-\beta h|y| + iy\tilde{y}} \prod_j \cos(\tilde{y}a_j). \quad (3.95)$$

Este é um bom momento para calcularmos E_I segundo a fórmula (3.92), pois as integrais em y e $\tilde{y}$ desacoplaram das integrais em W, $\widetilde{Q}$ e $\widetilde{R}$, levando a

$$E_I = \frac{1}{2\pi} \int_{-\infty}^{\infty} dy \mid y \mid \int_{-\infty}^{\infty} d\tilde{y}\, e^{iy\tilde{y}} \prod_j \cos(\tilde{y}a_j). \quad (3.96)$$

Efetuando a média sobre a_i obtemos

$$\begin{aligned} \langle E_I \rangle &= \frac{1}{2\pi} \int_{-\infty}^{\infty} dy \mid y \mid \int_{-\infty}^{\infty} d\tilde{y}\, e^{iy\tilde{y}} \left(\frac{\sin \tilde{y}}{\tilde{y}}\right)^N \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} dy \mid y \mid \int_{-\infty}^{\infty} d\tilde{y}\, e^{iy\tilde{y} + N \ln(\frac{\sin \tilde{y}}{\tilde{y}})}. \end{aligned} \quad (3.97)$$

A integral sobre $\tilde{y}$ é calculada utilizando-se a expansão

$$\ln(\frac{\sin \tilde{y}}{\tilde{y}}) \simeq -\frac{\tilde{y}^2}{6} - \frac{\tilde{y}^4}{180} + O(\tilde{y}^6), \quad (3.98)$$

Na verdade estamos usando a mesma idéia da aproximação de ponto de sela. A função na variável $\tilde{y}$ fica facilmente integrável,

$$\int_{-\infty}^{\infty} d\tilde{y}\, e^{iy\tilde{y} + N \ln(\frac{\sin \tilde{y}}{\tilde{y}})} = \int_{-\infty}^{\infty} d\tilde{y}\, e^{iy\tilde{y} - \frac{N\tilde{y}^2}{6}} = \sqrt{\frac{6\pi}{N}}\, e^{-\frac{3\tilde{y}^2}{2N}}. \quad (3.99)$$

Chegamos no resultado final ao integrarmos em y

$$\langle E_I \rangle = \frac{1}{\pi} \sqrt{\frac{6\pi}{N}} \int_{-\infty}^{\infty} dy \mid y \mid\, e^{-\frac{3\tilde{y}^2}{2N}} = \sqrt{\frac{2N}{3\pi}}. \quad (3.100)$$

A expressão (3.100) é a mesma encontrada para energia média aleatória [50], isto demonstra o completo fracasso dessa heurística. A figura 3.5 mostra os pontos obtidos a partir de 10000 experimentos

IFSC-USP SERVIÇO DE BIBLIOTECA INFORMAÇÃO

utilizando o algoritmo de relaxação para determinar as configurações s_j que minimizam a energia (3.1) e aplicando em seguida a heurística da intergralização sobre as configurações $(sign(s_j))$. Os pontos exibidos pela figura 3.5 estão em pleno acordo com a previsão teórica (linha cheia).

A idéia de integralização é muito utilizada para melhorar a solução em algoritmos mais complexos como o *plano de cortes* [42]. Talvez em outros contextos ela possa ser uma boa estratégia, mas a aplicação desta heurística no problema da partição númerica é tão ruim quanto ficar escolhendo soluções factíveis aleatoriamente. E aqui enfatizamos a importância de podermos fazer uma análise probabilística de problemas de otimização analiticamente. Quanto mais difícil for o algoritmo, mais difícil será a análise de seu desempenho, contudo, expressões analíticas são gerais e compactas tornando a comparação eficiente e a validação do algoritmo mais confiável.

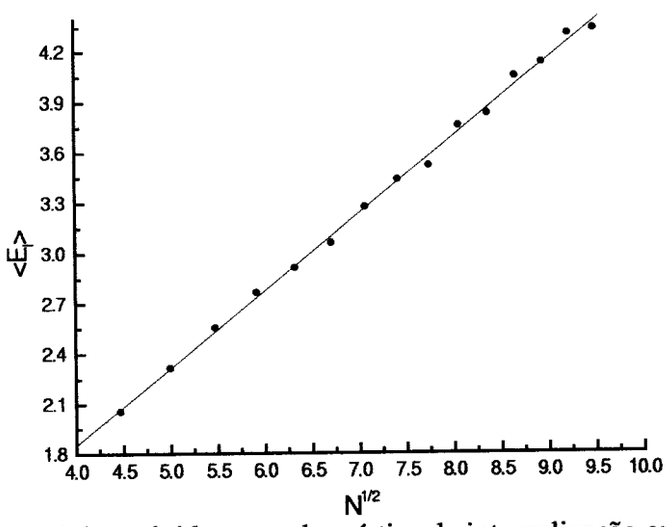


Figura 3.5: Energia minima obtida com a heurística da integralização em função de N. A linha cheia é o resultado teórico e os pontos foram obtidos via simulações.

Capítulo 4

O Espaço das Realizações do Problema da Partição Numérica

A maior parte das análises de problemas de otimização combinatorial via Mecânica Estatística tem se concentrado na caracterização de propriedades médias dos mínimos de uma dada função custo [1,25]. Usualmente a função custo depende de um grande conjunto de parâmetros fixos denominados instância ou realização de um problema de otimização. Como por exemplo, a distância entre n cidades no famoso problema do caixeiro viajante (apêndice A). No esquema usual da Mecânica Estatística esses parâmetros são tratados como variáveis lentas distribuídas de acordo com alguma função de probabilidade conhecida. Ainda no exemplo do caixeiro viajante, sabemos que o espaço de configuração tem dimensão $(n-1)!/2$ e para calcular a média sobre a função custo definiu-se uma distribuição de probabilidade como sendo a distribuição de Gibbs com temperatura $T = \frac{1}{\beta}$. O limite de temperatura zero seleciona as configuração que minimizam a função custo (estado fundamental). Claramente, nesta formulação as configurações são tratadas como variáveis rápidas. Este capítulo explora um ponto de vista oposto, a saber, para um dado conjunto de configurações queremos caracterizar o subespaço de instâncias para que estas configurações tenham uma função custo com valor bem determinado. Formalmente a proposta apresentada aqui tem por objetivo caracterizar o subespaço das realizações $\{a\}$ que torna um conjunto fixo de estados $\{s_j^\mu\}$ $\mu = 1, ..., P$, partições perfeitas, ou seja, $E(s_j^\mu) = 0$

$\forall\mu$. Para atingir este objetivo definimos uma energia no espaço de realização:

$$\mathcal{H} = \sum_{\mu=1}^{P}\left(\frac{1}{\sqrt{N}}\sum_{j=1}^{N} a_j s_j^{\mu}\right)^2, \tag{4.1}$$

de modo que as partições serão perfeitas somente se $\mathcal{H} = 0$. Vamos supor que P cresce linearmente com N, ou seja, $P = \alpha N$. Além disso assumimos também que as componentes s_j^{μ} sejam variáveis aleatórias estatisticamente independente escolhida segundo a distribuição de probabilidade:

$$\tilde{P}(s_j^{\mu}) = \frac{1}{2}\left(1 + \frac{m}{\sqrt{N}}\right)\delta(s_j^{\mu} - 1) + \frac{1}{2}\left(1 - \frac{m}{\sqrt{N}}\right)\delta(s_j^{\mu} + 1). \tag{4.2}$$

Os pesos da função delta foram escolhidos para que $\langle s_j^{\mu}\rangle = \frac{m}{\sqrt{N}}$. Esta escolha foi motivada pelo resultados do capítulo anterior (3.61), onde vimos que a variável m (diferença de cardinalidade média) escala como $\frac{1}{\sqrt{N}}$ e portanto usamos este resultado em (4.2) de forma a tornar o parâmetro m independente de N. É claro que para qualquer instância realística a, as diferentes componentes s_i de uma partição provavelmente não sejam estatisticamente independentes, ou melhor dizendo, a partição perfeita s não é aleatória. Restringir a análise às partições perfeitas aleatórias é a concessão que devemos fazer para tornar o problema tratável analiticamente.

Aplicaremos as técnicas da Mecânica Estatística para estudar o estado fundamental da Hamiltoniana $\mathcal{H}$ e enfatizaremos a análise de instâncias de energia zero. Este capítulo está organizado da seguinte forma: Na seção 1 usaremos o método das réplicas para calcular a densidade de energia livre no limite termodinâmico e derivamos um parâmetro de ordem na simetria de réplicas que descreve as propriedades do espaço de realizações. Em particular, mostramos que há um valor crítico $\alpha_c N$ que limita o número de partições perfeitas aleatórias. Em seguida, na seção 2, calculamos a densidade de probabilidade para uma dada entrada, digamos a_k, de ter um valor a. Isto é feito integrando a distribuição de probabilidade conjunta (a distribuição de Gibbs) sobre todas as entradas exceto a_k. No apêndice D estudamos as estabilidades da solução na hipótese de simetria com respeito à quebra de simetria e mostramos que as instâncias de energia zero podem ser confiavelmente descritas pelos nossos parâmetros de ordem.

4.1 Energia Livre

Seguindo a formulação de Gardner [46] para a análise do espaço de acoplamentos de redes neurais, vamos considerar os a_j como sendo as variáveis rápidas e os s_j^μ como sendo as variáveis lentas (fixas). A função de partição para este problema é escrita como

$$Z = \int_0^\infty \left[\prod_j da_j\right] \delta\left(Q - \frac{1}{N}\sum_j a_j^2\right) e^{-\beta\mathcal{H}}, \tag{4.3}$$

onde o vínculo esférico é necessário para garantir que nem todos os a_j anulem-se simultaneamente, o que claramente levaria a $\mathcal{H} = 0$ para qualquer conjunto de partições $\{s_j^\mu\}$. Naturalmente, esperamos que o parâmetro de normalização Q não influencie os resultados dos cálculos. Ao tomar o limite $T \to 0$ na equação (4.3) garantiremos que somente as instâncias que minimizam $\mathcal{H}$ contribuirão em Z.

A densidade de energia livre é definida por

$$-\beta f = \lim_{N\to\infty} \frac{1}{N} \langle \ln Z \rangle, \tag{4.4}$$

onde $\langle ... \rangle$ denota a média sobre as variáveis aleatórias s_j^μ distribuídas de acordo com a equação (4.2). A razão para se escrever $\langle \ln Z \rangle$ e não $\ln \langle Z \rangle$ é que $\ln Z$ é extensiva (isto é, escala linearmente com N) e portanto automediante (as flutuações de $\ln Z$ em torno da média $\langle \ln Z \rangle$ escalam com $\sqrt{N}$, de modo que o valor médio passa a ser significativo para descrever o comportamento das variáveis aleatórias). Para calcular a média em $\langle \ln Z \rangle$ vamos utilizar o método das réplicas: Este método consiste em primeiro cálcular a média de n funções de partição acopladas $\langle Z^n \rangle$ para n inteiro, para em seguida tomar a continuação analítica de $n \to 0$, em outras palavras, recorremos a identidade

$$\langle \ln Z \rangle = \lim_{n\to 0} \frac{1}{n} \langle Z^n - 1 \rangle, \tag{4.5}$$

onde $Z^n = \prod_\rho z^\rho$ é dado por

$$Z^n = \int_0^\infty \left[\prod_{j\rho} da_{j\rho}\right] \delta\left(Q - \frac{1}{N}\sum_j a_{j\rho}^2\right) e^{-\beta\mathcal{H}(a_{j\rho}, s_j^\mu)} \tag{4.6}$$

substituindo a energia dada por (4.1), e explicitando as funções delta de Dirac temos

$$\langle Z^n \rangle = \int_{-\infty}^{\infty} \left[\prod_{\rho} \frac{N d\tilde{Q}_\rho}{2\pi} e^{iN\tilde{Q}_\rho Q} \right] \int_0^{\infty} [\prod_{j\rho} da_{j\rho}\, e^{-i\tilde{Q}_\rho a_{j\rho}^2}]$$

$$\int_{-\infty}^{\infty} \left[\prod_{\mu\rho} \frac{dx_{\mu\rho} d\tilde{x}_{\mu\rho}}{2\pi} e^{i\tilde{x}_{\mu\rho} x_{\mu\rho} - \beta x_{\mu\rho}^2} \right] \left\langle e^{-\frac{i}{\sqrt{N}} \sum_{\mu\rho} \tilde{x}_{\mu\rho} \sum_j a_{j\rho} s_j^\mu} \right\rangle . \tag{4.7}$$

O primeiro passo será efetuarmos a média sobre s_j^μ. Recorrendo à distribuição de probabilidade (4.2)

$$\left\langle e^{-\frac{i}{\sqrt{N}} \sum_{\mu\rho} \tilde{x}_{\mu\rho} \sum_j a_{j\rho} s_j^\mu} \right\rangle = \int_{-\infty}^{\infty} \prod_{j\mu} ds_j^\mu\, \tilde{P}(s_j^\mu)\, e^{-\frac{i}{\sqrt{N}} \sum_\rho \tilde{x}_{\mu\rho} a_{j\rho} s_j^\mu}$$

$$= \prod_{j\mu} \left[\cos(\frac{1}{\sqrt{N}} \sum_\rho \tilde{x}_{\mu\rho} a_{j\rho}) + i \frac{m}{\sqrt{N}} \sin(\frac{1}{\sqrt{N}} \sum_\rho \tilde{x}_{\mu\rho} a_{j\rho}) \right] .$$

Como o argumento dessas funções trigonométricas é pequeno para N grande, podemos simplificar o resultado acima simplesmente expandindo em potências de $\frac{1}{N}$ e retendo o termo de correção de ordem mais baixa, obtendo assim a seguinte expressão para a média

$$\langle ... \rangle = \prod_{j\mu} \left[1 - \frac{1}{2N} (\sum_\rho \tilde{x}_{\mu\rho} a_{j\rho})^2 + \frac{im}{N} \sum_\rho \tilde{x}_{\mu\rho} a_{j\rho} \right]$$

$$= \prod_{j\mu} e^{\ln\left[1 - \frac{1}{2N} (\sum_\rho \tilde{x}_{\mu\rho} a_{j\rho})^2 + \frac{im}{N} \sum_\rho \tilde{x}_{\mu\rho} a_{j\rho}\right]}$$

$$= \prod_{j\mu} e^{-\frac{1}{2N} (\sum_\rho \tilde{x}_{\mu\rho} a_{j\rho})^2 + \frac{im}{N} \sum_\rho \tilde{x}_{\mu\rho} a_{j\rho}} . \tag{4.8}$$

Observando que o quadrado da somatória pode ser escrito como

$$(\sum_\rho \tilde{x}_{\mu\rho} a_{j\rho})^2 = \sum_\rho \tilde{x}_{\mu\rho}^2 a_{j\rho}^2 + 2 \sum_{\rho<\delta} \tilde{x}_{\mu\rho} \tilde{x}_{\mu\delta} a_{j\rho} a_{j\delta}, \tag{4.9}$$

reescrevemos a equação (4.8)

$$\langle ... \rangle = \prod_{j\mu} e^{-\frac{1}{2N} \sum_\rho \tilde{x}_{\mu\rho}^2 a_{j\rho}^2 - \frac{1}{N} \sum_{\rho<\delta} \tilde{x}_{\mu\rho} \tilde{x}_{\mu\delta} a_{j\rho} a_{j\delta} + \frac{im}{N} \sum_\rho \tilde{x}_{\mu\rho} a_{j\rho}} . \tag{4.10}$$

Nesta etapa, podemos definir os seguintes parâmetros de ordem:

$$q_{\rho\delta} = \frac{1}{N}\sum_j a_{j\rho}a_{j\delta} \tag{4.11}$$

$$R_\rho = \frac{1}{N}\sum_j a_{j\rho}. \tag{4.12}$$

Com o auxílio das funções delta de Dirac podemos inserir estes parâmetros na função de partição,

$$\int_{-\infty}^{\infty} N dq_{\rho\delta}\, \delta(Nq_{\rho\delta} - \sum_j a_{j\rho}a_{j\delta}) \tag{4.13}$$

$$\int_{-\infty}^{\infty} N dR_\rho\, \delta(NR_\rho - \sum_j a_{j\rho}) \tag{4.14}$$

de modo que a função de partição replicada média fica dada por

$$\begin{aligned}\langle Z^n\rangle &= \int_{-\infty}^{\infty}\left[\prod_\rho \frac{N d\tilde{Q}_\rho}{2\pi i}\, e^{N\tilde{Q}_\rho Q}\right]\int_{-\infty}^{\infty}\left[\prod_{\rho<\delta}\frac{N dq_{\rho\delta}d\tilde{q}_{\rho\delta}}{2\pi i}\, e^{Nq_{\rho\delta}\tilde{q}_{\rho\delta}}\right]\\ &\times\int_{-\infty}^{\infty}\left[\prod_\rho \frac{N d\tilde{R}_\rho dR_\rho}{2\pi i}\, e^{NR_\rho\tilde{R}_\rho}\right] e^{\alpha N\ln G_1(q,R)+N\ln G_2(\tilde{Q},\tilde{q},\tilde{R})}\end{aligned} \tag{4.15}$$

onde

$$G_1 = \int_{-\infty}^{\infty}\left[\prod_\rho \frac{dx_\rho d\tilde{x}_\rho}{2\pi}\right] e^{i\sum_\rho \tilde{x}_\rho x_\rho - \beta\sum_\rho x_\rho^2 - \frac{1}{2}Q\sum_\rho \tilde{x}_\rho^2 - \sum_{\rho<\delta}\tilde{x}_\rho\tilde{x}_\delta q_{\rho\delta} + im\sum_\rho \tilde{x}_\rho R_\rho} \tag{4.16}$$

e

$$G_2 = \int_0^{\infty}\left[\prod_\rho da_\rho\right] e^{-\sum_\rho \tilde{Q}_\rho a_\rho^2 - \sum_{\rho<\delta}\tilde{q}_{\rho\delta}a_\rho a_\delta - \sum_\rho \tilde{R}_\rho a_\rho}. \tag{4.17}$$

As integrais sobre os parâmetros de ordem $q_{\rho\delta}$ e R_ρ , bem como sobre os multiplicadores de Lagrange $\tilde{q}_{\rho\delta}$, $\tilde{R}_\rho$ e $\tilde{Q}_\rho$ podem ser calculadas pelo método de ponto de sela no limite $N \to \infty$.

Ansatz de Simetria de Réplicas

O próximo passo é calcular as funções G_1 e G_2 na condição de simetria de réplicas, isto é, supondo

$$\begin{aligned} \tilde{Q}_\rho &= \tilde{Q}, \qquad \tilde{R}_\rho = \tilde{R}, \qquad R_\rho = R \qquad \forall\rho \\ \tilde{q}_{\rho\delta} &= \tilde{q}, \qquad q_{\rho\delta} = q \qquad\qquad \rho < \delta. \end{aligned} \tag{4.18}$$

Esta hipótese irá facilitar enormemente o cálculo analítico de G_1 e G_2

Cálculo de G_1

Ao aplicarmos o ansatz acima na expressão de G_1 dada na equação (4.16), ficamos com

$$G_1 = \int_{-\infty}^{\infty} \left[\prod_\rho \frac{dx_\rho d\tilde{x}_\rho}{2\pi} \right] e^{i\sum_\rho \tilde{x}_\rho x_\rho - \beta \sum_\rho x_\rho^2 - \frac{1}{2}Q\sum_\rho \tilde{x}_\rho^2 - q\sum_{\rho<\delta} \tilde{x}_\rho \tilde{x}_\delta + imR\sum_\rho \tilde{x}_\rho}. \tag{4.19}$$

Utilizando a identidade (4.9) e a transforamação gaussiana

$$\int_{-\infty}^{\infty} Dz\, e^{i\sqrt{q}\sum_\rho \tilde{x}_\rho z} = e^{-\frac{q}{2}(\sum_\rho \tilde{x}_\rho)^2} \tag{4.20}$$

onde

$$Dz = \frac{dz}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}} \tag{4.21}$$

é a medida gaussiana, podemos desacoplar as diferentes réplicas que aparecem no termo cruzado $(\tilde{x}_\rho \tilde{x}_\delta)$ e ficar com uma expressão mais simples. A seguir, integramos em x e $\tilde{x}$,

$$\begin{aligned} G_1 &= \int_{-\infty}^{\infty} Dz \prod_\rho \int_{-\infty}^{\infty} \frac{dx d\tilde{x}}{2\pi}\, e^{i\tilde{x}x - \beta x^2 - \frac{1}{2}Q\tilde{x}^2 + \frac{q}{2}\tilde{x}^2 + imR\tilde{x} + i\sqrt{(q)}z\tilde{x}} \\ &= \int_{-\infty}^{\infty} Dz \left[\sqrt{\frac{1}{1+2\beta(Q-q)}}\, e^{-\frac{\beta(mR+\sqrt{(q)}z)^2}{1+2(Q-q)\beta}} \right]^n . \end{aligned} \tag{4.22}$$

Neste estágio vamos tomar o limite $n \to 0$. Para isto recorremos à seguinte aproximação

$$G_1 = \int_{-\infty}^{\infty} Dz\, (\varphi)^n \approx \int_{-\infty}^{\infty} Dz\, (1 + n \ln \varphi) \approx 1 + n \int_{-\infty}^{\infty} Dz \ln \varphi \tag{4.23}$$

onde φ é o termo entre colchetes na equação (4.22). Daí,

$$\begin{aligned} \frac{\ln G_1}{n} &= -\int_{-\infty}^{\infty} Dz \left\{ \ln\left[1 + 2\beta(Q-q)\right]^{\frac{1}{2}} + \frac{\beta[mR + \sqrt{(q)}z]^2}{1+2\beta(Q-q)} \right\} \\ &= -\frac{1}{2}\ln\left[1 + 2\beta(Q-q)\right] - \beta \frac{m^2R^2 + q}{1+2\beta(Q-q)} . \end{aligned} \tag{4.24}$$

Cálculo de G_2

Retomando a expressão (4.17) e seguindo o mesmo procedimento utilizado no cálculo de G_1, temos

$$G_2 = \int_{-\infty}^{\infty} Dz \int_0^{\infty} \left[\prod_\rho da_\rho \right] e^{-\tilde{Q}\sum_\rho a_\rho^2 - \frac{\tilde{q}}{2}\left[\left(\sum_\rho a_\rho\right)^2 - \sum_\rho a_\rho^2\right] - \tilde{R}\sum_\rho a_\rho}$$

$$= \int_{-\infty}^{\infty} Dz \left[\frac{1}{2}\sqrt{\frac{\pi}{(\tilde{Q}-\frac{\tilde{q}}{2})}}\, e^{\frac{\left(\tilde{R}-i\sqrt{(\tilde{q})}z\right)^2}{4(\tilde{Q}-\frac{\tilde{q}}{2})}} \operatorname{erfc}\left[\frac{(\tilde{R}-i\sqrt{(\tilde{q})}z)}{2(\tilde{Q}-\frac{\tilde{q}}{2})^{\frac{1}{2}}}\right]\right]^n . \tag{4.25}$$

Tomando o limite $n \to 0$ obtemos

$$\frac{\ln G_2}{n} \approx \frac{1}{2}\ln\pi - \frac{1}{2}\ln[(\tilde{Q}+\frac{\tilde{q}}{2})] + \frac{\tilde{R}^2+\tilde{q}}{4(\tilde{Q}+\frac{\tilde{q}}{2})} + \int_{-\infty}^{\infty} Dz \ln\frac{1}{2}\operatorname{erfc}\left[\frac{(\tilde{R}+\sqrt{(\tilde{q})}z)}{2(\tilde{Q}+\frac{\tilde{q}}{2})^{\frac{1}{2}}}\right], \tag{4.26}$$

onde efetuamos a mudança de variável $-\tilde{q} \to \tilde{q}$.

4.1.1 Energia livre com simetria de réplicas

No limite $N \to \infty$ as integrais (4.15) podem ser calculadas pelo método de ponto de sela. A expressão para a densidade de energia livre (4.4) é simplesmente

$$-\beta f = \frac{1}{n} extr\{Q\sum_{\rho}\tilde{Q}_\rho + \sum_{\rho<\delta} q_{\rho\delta}\tilde{q}_{\rho\delta} + \sum_{\rho}\tilde{R}_\rho R_\rho + \alpha\ln G_1 + \ln G_2\} \tag{4.27}$$

onde o extremo é tomado sobre os parâmetros de ponto de sela $q_{\rho\delta}$, R_ρ, $\tilde{q}_{\rho\delta}$, $\tilde{R}_\rho$ e $\tilde{Q}_\rho$. Agora podemos obter a energia livre nesse ansatz. Para isto notamos que

$$\frac{Q}{n}\sum_{\rho}\tilde{Q}_\rho = Q\tilde{Q}, \tag{4.28}$$

$$\frac{1}{n}\sum_{\rho}\tilde{R}_\rho R_\rho = R\tilde{R}, \tag{4.29}$$

$$\frac{1}{n}\sum_{\rho<\delta}\tilde{q}_{\rho\delta}q_{\rho\delta} = \frac{n(n-1)}{2n}\tilde{q}q = -\frac{\tilde{q}q}{2}, \tag{4.30}$$

de forma que a equação (4.27) fica escrita como

$$\begin{aligned} -\beta f_{rs} &= Q\tilde{Q} - \frac{\tilde{q}q}{2} + \frac{R\tilde{R}}{N} - \frac{\alpha}{2}\ln\left[1+2\beta(Q-q)\right] - \beta\alpha\frac{m^2R^2+q}{1+2\beta(Q-q)} + \frac{1}{2}\ln\pi \\ &\quad -\frac{1}{2}\ln[(\tilde{Q}+\frac{\tilde{q}}{2})] + \frac{\tilde{R}^2+\tilde{q}}{4(\tilde{Q}+\frac{\tilde{q}}{2})} + \int_{-\infty}^{\infty} Dz\ln\frac{1}{2}\operatorname{erfc}\left[\frac{(\tilde{R}+\sqrt{(\tilde{q})}z)}{2(\tilde{Q}+\frac{\tilde{q}}{2})^{\frac{1}{2}}}\right]. \end{aligned} \tag{4.31}$$

A densidade de energia ε é calculada segundo a sua fórmula clássica,

$$\varepsilon = -\frac{\partial}{\partial\beta}\frac{1}{N}\ln Z = -\frac{\partial}{\partial\beta}(-\beta f_{rs}), \tag{4.32}$$

que resulta em

$$\varepsilon = \frac{\alpha}{2}\frac{2(Q-q)}{1+2\beta(Q-q)} + \alpha\frac{q+m^2R^2}{1+2\beta(Q-q)} - \alpha\frac{q+m^2R^2}{[1+2\beta(Q-q)]^2}2\beta(Q-q). \tag{4.33}$$

Cálculo dos parâmetros de ponto de sela

Resta agora determinarmos os parâmetros q, $\tilde{q}$, R, $\tilde{R}$ e $\tilde{Q}$. Eles são obtidos extremizando f_{rs}, isto é, resolvendo-se as chamadas equações de ponto de sela que apresentamos a seguir:

a) $\frac{\partial}{\partial q}(-\beta f\ ) = 0$:

$$\tilde{q} - \frac{4\alpha\beta^2(q+m^2R^2)}{[1+2\beta(Q-q)]^2} = 0; \tag{4.34}$$

b) $\frac{\partial}{\partial R}(-\beta f\ ) = 0$:

$$\tilde{R} - \frac{2Rm^2\alpha\beta}{1+2\beta(Q-q)} = 0; \tag{4.35}$$

c) $\frac{\partial}{\partial \tilde{R}}(-\beta f\ ) = 0$:

$$R + \frac{\tilde{R}}{2(\tilde{Q}+\frac{\tilde{q}}{2})} - \frac{1}{\sqrt{\pi}(\tilde{Q}+\frac{\tilde{q}}{2})^{\frac{1}{2}}} \int_{-\infty}^{\infty} Dz \frac{e^{-\frac{(\tilde{R}+\sqrt{(\tilde{q})}z)^2}{4(\tilde{Q}+\frac{\tilde{q}}{2})}}}{\mathrm{erfc}\left[\frac{(\tilde{R}+\tilde{q}^{\frac{1}{2}}z)}{2(\tilde{Q}+\frac{\tilde{q}}{2})^{\frac{1}{2}}}\right]} = 0; \tag{4.36}$$

d) $\frac{\partial}{\partial \tilde{Q}}(-\beta f) = 0$:

$$Q - \frac{1}{2(\tilde{Q}+\frac{\tilde{q}}{2})} - \frac{\tilde{q}+\tilde{R}^2}{4(\tilde{Q}+\frac{\tilde{q}}{2})^2} + \frac{1}{\sqrt{\pi}2(\tilde{Q}+\frac{\tilde{q}}{2})^{\frac{3}{2}}} \int_{-\infty}^{\infty} Dz \frac{e^{-\frac{(\tilde{R}+\sqrt{(\tilde{q})}z)^2}{4(\tilde{Q}+\frac{\tilde{q}}{2})}}}{\mathrm{erfc}\left[\frac{(\tilde{R}+\sqrt{(\tilde{q})}z)}{2(\tilde{Q}+\frac{\tilde{q}}{2})^{\frac{1}{2}}}\right]} (\tilde{R}+\tilde{q}^{\frac{1}{2}}z) = 0; \tag{4.37}$$

e) $\frac{\partial}{\partial \tilde{q}}(-\beta f) = 0$:

$$q - \frac{\tilde{q}+\tilde{R}^2}{4(\tilde{Q}+\frac{\tilde{q}}{2})^2} + \frac{1}{\sqrt{\pi}(\tilde{Q}+\frac{\tilde{q}}{2})^{\frac{1}{2}}} \int_{-\infty}^{\infty} Dz \frac{e^{-\frac{(\tilde{R}+\sqrt{(\tilde{q})}z)^2}{4(\tilde{Q}+\frac{\tilde{q}}{2})}}}{\mathrm{erfc}\left[\frac{(\tilde{R}+\sqrt{(\tilde{q})}z)}{2(\tilde{Q}+\frac{\tilde{q}}{2})^{\frac{1}{2}}}\right]} \left(\frac{\tilde{R}+\sqrt{(\tilde{q})}z}{2(\tilde{Q}+\frac{\tilde{q}}{2})} - \tilde{q}^{\frac{-1}{2}}z\right) = 0. \tag{4.38}$$

Observemos que a diferença entre (4.37) e (4.38) leva a uma expressão mais simples,

$$Q - q - \frac{1}{2(\tilde{Q}+\frac{\tilde{q}}{2})} + \frac{1}{\sqrt{\pi(\tilde{Q}+\frac{\tilde{q}}{2})\tilde{q}}} \int_{-\infty}^{\infty} Dz \frac{z\, e^{-\frac{(\tilde{R}+\sqrt{(\tilde{q})}z)^2}{4(\tilde{Q}+\frac{\tilde{q}}{2})}}}{\mathrm{erfc}\left[\frac{(\tilde{R}+\sqrt{(\tilde{q})}z)}{2(\tilde{Q}+\frac{\tilde{q}}{2})^{\frac{1}{2}}}\right]} = 0. \tag{4.39}$$

As equações (4.36) e (4.39) permitem-nos reescrever a equação (4.37) como

$$2(\tilde{Q}+\frac{\tilde{q}}{2})Q = 1 - \tilde{R}R + \tilde{q}(Q-q). \tag{4.40}$$

Estas equações são gerais, válidas para quaisquer valores dos parâmetros de controle Q, α, m e β. Contudo, estamos interessados apenas no limite de temperatura zero ($\beta \to \infty$), pois neste limite podemos garantir que as únicas realizações $\{a_i\}$ que contribuem para a função de partição (4.3) são as que minimizam a energia $\mathcal{H}$. Entretanto, o limite $\beta \to \infty$ deve ser tomado com muito cuidado e, neste sentido, separamos a análise a seguir em duas partes.

Caso I: $\lim_{\beta\to\infty} \beta(Q-q) < \infty$.

A primeira maneira de efetuarmos o limite é definir $x = 2\beta(Q-q)$ e impor que ele seja finito no limite $\beta \to \infty$. Destarte, podemos fazer algumas simplificações nas equações de ponto de sela. A equação (4.34) mostra que $\tilde{q} \approx \beta^2$; explicitamente

$$\tilde{q} = \frac{4\alpha(q+m^2R^2)}{[1+x]^2}\beta^2, \tag{4.41}$$

enquanto a equação (4.35) mostra que $\tilde{R} \approx \beta$,

$$\tilde{R} = \frac{2Rm^2\alpha}{1+x}\beta. \tag{4.42}$$

Substituindo (4.41) e (4.42) em (4.40) obtemos

$$2\Delta Q = 1 - \frac{2m^2R^2}{1+x}\alpha\beta + \frac{2(Q+m^2R^2)}{[1+x]^2}\alpha\beta x \tag{4.43}$$

onde $\Delta \equiv \tilde{Q}+\frac{\tilde{q}}{2} \approx \beta$. No limite $\beta \to \infty$ a função *erro complementar* pode ser simplificada reduzindo a equação (4.36) à seguinte forma

$$\begin{aligned} R &= -\frac{\tilde{R}}{2\Delta} + \frac{1}{2\sqrt{\Delta}}\int_{-z_0}^{\infty} Dz\frac{(\tilde{R}+\sqrt{(\tilde{q})}z)}{\sqrt{\Delta}} \\ &= \frac{\sqrt{(\tilde{q})}}{2\Delta}\left[-\frac{z_0}{2}\mathrm{erfc}(\frac{z_0}{\sqrt{2}}) + \frac{e^{(-\frac{z_0^2}{2})}}{\sqrt{2\pi}}\right]. \end{aligned} \tag{4.44}$$

Já a equação (4.37) fica

$$Q = \frac{1}{\Delta} + \frac{\tilde{q}+\tilde{R}^2}{4\Delta^2} - \frac{\tilde{q}}{4\Delta^2}\int_{-z_0}^{\infty} Dz(z_0+z)^2$$

$$= \frac{\tilde{q}}{4\Delta^2}\left[\frac{1+z_0^2}{2}\mathrm{erfc}(\frac{z_0}{\sqrt{2}}) - z_0\frac{e^{(-\frac{z_0^2}{2})}}{\sqrt{2\pi}}\right]. \tag{4.45}$$

Por fim, a equação (4.39) reduz-se a

$$\begin{aligned} Q-q &= \frac{1}{2\Delta} - \frac{1}{2\Delta}\int_{-z_0}^{\infty} Dz(z_0+z)z \\ &= \frac{1}{4\Delta}\mathrm{erfc}(\frac{z_0}{\sqrt{2}}). \end{aligned} \tag{4.46}$$

Manipulando estas equações, podemos facilmente encontrar uma equação para determinar x, a saber,

$$x = \frac{A(z_0)\frac{z_0}{\alpha m^2}}{1-A(z_0)\frac{z_0}{\alpha m^2}} \tag{4.47}$$

com

$$A(z_0) = \frac{\mathrm{erfc}(\frac{z_0}{\sqrt{2}})}{-z_0\mathrm{erfc}(\frac{z_0}{\sqrt{2}}) + 2\frac{e^{(-\frac{z_0^2}{2})}}{\sqrt{2\pi}}} \tag{4.48}$$

onde z_0 é solução da equação

$$\frac{z_0^2}{\alpha m^4 - m^2 z_0^2} = \frac{\left[\frac{e^{-\frac{z_0^2}{2}}}{\sqrt{2\pi}} - \frac{1}{2}z_0\mathrm{erfc}\left(\frac{z_0}{\sqrt{2}}\right)\right]^2}{\frac{1+z_0^2}{2}\mathrm{erfc}(\frac{z_0}{\sqrt{2}}) - \frac{z_0\, e^{-\frac{z_0^2}{2}}}{\sqrt{2\pi}}}. \tag{4.49}$$

Naturalmente, as equações (4.47) - (4.49) são válidas apenas para $m \neq 0$. Para partições balanceadas ($m = 0$) temos $\tilde{R} = 0$ [veja equação (4.42)] e, portanto, $z_0 = 0$. Usando esse resultado no conjunto de equações (4.41)-(4.46) obtém-se

$$x = \frac{1}{\sqrt{2\alpha}-1}. \tag{4.50}$$

Alternativamente, esse resultado poderia ser obtido expandindo-se o conjunto de equações (4.47) - (4.49) para $z_0 \ll 1$.

Observe que tanto em (4.47) como em (4.50) existe um valor de α para o qual x diverge. A fim de

interpretar essa divergência, devemos calcular o limite de temperatura zero da energia ou função custo média ε dada para um β geral pela equação (4.33). O resultado final é

$$\frac{\varepsilon}{Q} = \frac{\alpha^2 m^2}{(1+x)^2(\alpha m^2 - z_0^2)}, \tag{4.51}$$

onde z_0 é dado pela solução de (4.49). Novamente, o caso $m = 0$ leva a uma indeterminação em (4.51) e deve ser tratado separadamente. De fato fazendo $m = 0$ diretamente em (4.33) e usando o valor de x dado em (4.50) obtemos

$$\frac{\varepsilon}{Q} = \frac{\alpha}{(1+x)^2} = \frac{1}{2}(\sqrt{2\alpha} - 1)^2. \tag{4.52}$$

Há dois resultados importantes derivados dessas equações. Primeiro, a norma Q apenas re-escala a energia; segundo, a divergência de x determina o valor de α em que a energia média anula-se, ou seja, torna-se possível encontrar realizações $\{a_i\}$ que resultem em partições perfeitas. Naturalmente, para determinar o valor máximo $\alpha_c N$ para o número de partições perfeitas em função do vínculo de cardinalidade m devemos substituir em (4.40) as equações (4.41) e (4.42) e em seguida usar (4.46). O resultado é

$$\alpha_c = \frac{1}{2}\mathrm{erfc}\left(\frac{z_0^*}{\sqrt{2}}\right), \tag{4.53}$$

onde z_0^* é a solucão da equação

$$A(z_0^*)\frac{z_0^*}{\alpha_c m^2} = 1. \tag{4.54}$$

Para $m = 0$, o resultado é simplesmente $\alpha_c(0) = \frac{1}{2}$. A figura 4.1 mostra α_c como função de m, obtido de acordo com o procedimento acima.

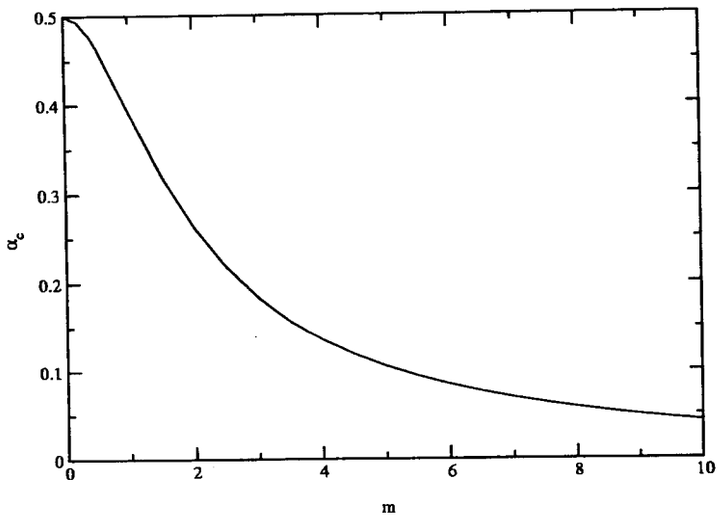


Figura 4.1: Limitante superior para o número de partições perfeitas aleatórias α_c como função do parâmetro m. Para partições balanceada ($m = 0$) encontramos $\alpha_c = \frac{1}{2}$.

Caso II: $\lim_{\beta\to\infty} 2\beta(Q-q) \to \infty$

Nesta segunda possibilidade devemos supor que $x \to \infty$ e, neste caso basta tomar o limite $\beta \to \infty$ nas equações (4.34) a (4.38). A equação (4.34) fica

$$\tilde{q} = \alpha\frac{(q + m^2R^2)}{(Q-q)^2}, \tag{4.55}$$

e a equação (4.35),

$$\tilde{R} = \alpha\frac{m^2R}{Q-q}. \tag{4.56}$$

As demais equações são escritas como

$$R = -\frac{\tilde{R}}{2\Delta} + \frac{1}{\sqrt{\pi\Delta}}\int_{-\infty}^{\infty} Dz\frac{e^{-\frac{(\tilde{R}+\sqrt{(\tilde{q})}z)^2}{4\Delta}}}{\mathrm{erfc}[\frac{(\tilde{R}+\tilde{q}^{\frac{1}{2}}z)}{2\sqrt{\Delta}}]}, \tag{4.57}$$

$$Q = \frac{1}{2\Delta} + \frac{\tilde{q} + \tilde{R}^2}{4\Delta^2} - \frac{1}{2\Delta\sqrt{\pi\Delta}} \int_{-\infty}^{\infty} Dz \frac{e^{-\frac{(\tilde{R}+\sqrt{(\tilde{q})}z)^2}{4\Delta}}}{\mathrm{erfc}[\frac{(\tilde{R}+\sqrt{(\tilde{q})}z)}{2\sqrt{\Delta}}]} (\tilde{R} + \sqrt{(\tilde{q})}z), \tag{4.58}$$

$$Q - q = \frac{1}{2\Delta} - \frac{1}{\sqrt{\pi\Delta\tilde{q}}} \int_{-\infty}^{\infty} Dz \frac{z\, e^{-\frac{(\tilde{R}+\sqrt{(\tilde{q})}z)^2}{4\Delta}}}{\mathrm{erfc}[\frac{(\tilde{R}+\tilde{q}^{\frac{1}{2}}z)}{2\sqrt{\Delta}}]}, \tag{4.59}$$

$$2\Delta Q = 1 - \tilde{R}R + \tilde{q}(Q - q). \tag{4.60}$$

Vamos definir novas variáveis,

$$\Gamma = \frac{\tilde{R}}{\sqrt{\Delta}}, \qquad \xi = \frac{\tilde{q}}{\Delta}, \tag{4.61}$$

e escrever novamente as equações de ponto sela:

$$\xi = \frac{\alpha}{\Delta} \frac{(1 + m^2R^2)}{(1 - q)^2}, \tag{4.62}$$

$$\Gamma = \frac{\alpha}{\sqrt{\Delta}} \frac{m^2R}{(1 - q)}, \tag{4.63}$$

$$R = -\frac{\Gamma}{2\sqrt{\Delta}} + \frac{1}{\sqrt{\pi\Delta}} \int_{-\infty}^{\infty} Dz \frac{e^{-\frac{\left(\Gamma+\xi^{\frac{1}{2}}z\right)^2}{4}}}{\mathrm{erfc}\left(\frac{\Gamma+\xi^{\frac{1}{2}}z}{2}\right)}, \tag{4.64}$$

$$q = 1 - \frac{1}{2\Delta} + \frac{1}{\Delta\sqrt{\pi\xi}} \int_{-\infty}^{\infty} Dz \frac{z\, e^{-\frac{\left(\Gamma+\xi^{\frac{1}{2}}z\right)^2}{4}}}{\mathrm{erfc}\left(\frac{\Gamma+\xi^{\frac{1}{2}}z}{2}\right)}, \tag{4.65}$$

$$\Delta = \frac{1}{2} - \frac{\Gamma}{2}R\sqrt{\Delta} + \frac{\xi}{2}\Delta(1 - q). \tag{4.66}$$

A variável Q foi eliminada simplesmente redefinindo

$$\Delta Q \to \Delta, \qquad \frac{R}{\sqrt{Q}} \to R, \qquad \frac{q}{Q} \to q. \tag{4.67}$$

Nunca é demais chamar a atenção para o fato de que essas equações independem da escolha da norma Q das realizações. A solução dessas equações deve ser obtida numericamente. Em particular, podemos determinar o valor de α_c simplesmente fazendo q$\rightarrow$ 1. Exibimos abaixo a figura 4.3 mostrando como q varia com α.

Podemos observar que, para um valor de α fixo, à medida que o valor de m aumenta, aumenta também a correlação entre as realizações que resultam em partições ótimas. É interessante mencionar que para $\alpha = 0$ temos $q = R^2 = \frac{2}{\pi}$ independentes do valor de m. Para concluir, observemos que tomando o limite $\beta \rightarrow \infty$ com q$\rightarrow$ Q na equação (4.51) obtemos $\frac{\varepsilon}{Q} = 0$ conforme o esperado.

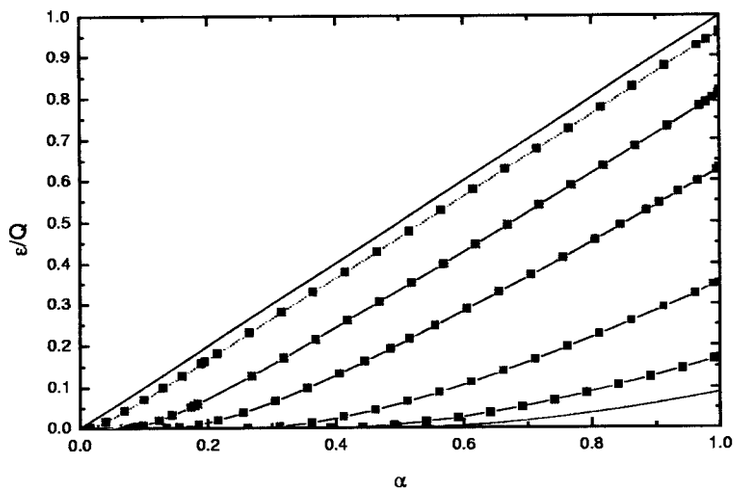


Figura 4.2: A densidade de energia ε normalizada por Q em função de α. A curva em verde é o resultado para $m = 0$ equanto a curva preta é um resultado assintótico (obtida para grandes valores de m). As curvas intermediárias correspondem aos valores: $m = 2, 4, 7, 10, 20$, de baixo para cima.

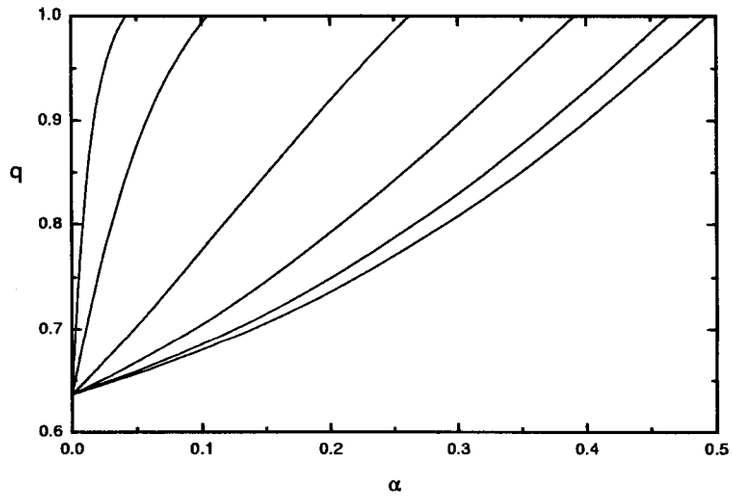


Figura 4.3: O parâmetro q contra α obtido resolvendo-se numericamente as equações de ponto de sela. Da direita para a esquerda as linhas correspondem aos seguintes valores de m: 0, 0.5, 1, 2, 5, 10

4.2 Distribuição de probabilidade das entradas

A abordagem probabilística tradicional para estudar problemas de otimização introduz uma distribuição de probabilidade sobre o espaço das realizações. O motivo para se fazer isto é que raramente se conhece uma distribuição de probabilidade realística. No **PPN**, por exemplo, usualmente considera-se as entradas a_k como sendo variáveis aleatórias estatisticamente independentes que seguem uma distribuição de probabilidade uniforme.

Nesta seção vamos determinar a distribuição de probabilidade de uma componente arbitrária, digamos a_k, da realização $\{a_j\}$ que minimiza a energia $\mathcal{H}$. Essa distribuição é definida por

$$\begin{aligned} \mathcal{P}_k(y) &= \lim_{\beta\to\infty} \langle \delta(y-a_k)\rangle_T \\ &= \lim_{\beta\to\infty} \left\langle \frac{1}{Z}\int_0^\infty (\prod_j da_j)\delta(Q-\frac{1}{N}\sum_j a_j^2)\delta(y-a_k)\, e^{-\beta\mathcal{H}(y)} \right\rangle \end{aligned} \tag{4.68}$$

onde Z e $\mathcal{H}$ são dadas pela equações (4.1) e (4.3) respectivamente. Como todas as entradas são equivalentes, podemos escrever $\mathcal{P}_k(y) = \mathcal{P}(y)\ \forall k$. Conseqüentemente, para avaliar a equação (4.68) podemos introduzir uma energia auxiliar

$$\mathcal{H}_{aux} = \sum_{\mu} \left(\frac{1}{\sqrt{N}} \sum_{j} a_j s_j^{\mu} \right)^2 + h \sum_{j} \delta(y - a_j) \tag{4.69}$$

de modo que

$$\mathcal{P}(y) = - \lim_{\beta \to \infty} \frac{1}{N\beta} \frac{\partial}{\partial h} \langle \ln Z_{aux} \rangle |_{h=0} \tag{4.70}$$

onde

$$Z_{aux} = \int_0^{\infty} \left[\prod_j da_j \right] \delta(Q - \frac{1}{N} \sum_j a_j^2)\, e^{-h \sum_j \delta(y-a_j)} \int_{-\infty}^{\infty} \prod_{\mu} dx_{\mu} \delta(x_{\mu} - \frac{1}{\sqrt{N}} \sum_j a_j s_j^{\mu})\, e^{-\beta x_{\mu}^2}. \tag{4.71}$$

Obviamente, as entradas não são estatisticamente independentes, sendo a sua distribuição de probabilidade conjunta dada exatamente pela distribuição de Gibbs.

Dado que precisamos calcular a média do logaritmo da função de partição, utilizaremos novamente o método das réplicas. Seguindo os passos da seção 1, escrevemos a função de partição replicada como

$$\begin{aligned} \langle Z_{aux}^n \rangle &= \int_{-\infty}^{\infty} \left[\prod_{\rho} \frac{N d\tilde{Q}_{\rho}}{2\pi}\, e^{iN\tilde{Q}_{\rho} Q} \right] \int_0^{\infty} [\prod_{j\rho} da_{j\rho}\, e^{-i\tilde{Q}_{\rho} a_{j\rho}^2 - h\delta(y-a_j)}] \\ & \int_{-\infty}^{\infty} \left[\prod_{\mu\rho} \frac{dx_{\mu\rho} d\tilde{x}_{\mu\rho}}{2\pi}\, e^{i\tilde{x}_{\mu\rho} x_{\mu\rho} - \beta x_{\mu\rho}^2} \right] \left\langle e^{-\frac{i}{\sqrt{N}} \sum_{\mu\rho} \tilde{x}_{\mu\rho} \sum_j a_{j\rho} s_j^{\mu}} \right\rangle . \end{aligned} \tag{4.72}$$

Efetuando os cálculos como anteriormente, encontramos

$$\begin{aligned} \langle Z_{aux}^n \rangle &= \int_{-\infty}^{\infty} \left[\prod_{\rho} \frac{N d\tilde{Q}_{\rho}}{2\pi i}\, e^{N\tilde{Q}_{\rho} Q} \right] \int_{-\infty}^{\infty} \left[\prod_{\rho<\delta} \frac{N dq_{\rho\delta} d\tilde{q}_{\rho\delta}}{2\pi i}\, e^{N q_{\rho\delta} \tilde{q}_{\rho\delta}} \right] \\ & \times \int_{-\infty}^{\infty} \left[\prod_{\rho} \frac{N d\tilde{R}_{\rho} dR_{\rho}}{2\pi i}\, e^{N R_{\rho} \tilde{R}_{\rho}} \right] e^{\alpha N \ln G_1(q,R) + N \ln G_2(\tilde{Q},\tilde{q},\tilde{R},h)} \end{aligned} \tag{4.73}$$

onde

$$G_1(q,R) = \int_{-\infty}^{\infty} \left[\prod_\rho \frac{dx_\rho d\tilde{x}_\rho}{2\pi}\right] e^{i\sum_\rho \tilde{x}_\rho x_\rho - \beta\sum_\rho x_\rho^2 - \frac{1}{2}Q\sum_\rho \tilde{x}_\rho^2 - \sum_{\rho<\delta}\tilde{x}_\rho\tilde{x}_\delta q_{\rho\delta} + im\sum_\rho \tilde{x}_\rho R_\rho} , \tag{4.74}$$

$$G_2(\tilde{Q},\tilde{q},\tilde{R},h) = \int_0^{\infty} \left[\prod_\rho da_\rho\right] e^{-\sum_\rho \tilde{Q}_\rho a_\rho^2 - \sum_{\rho<\delta}\tilde{q}_{\rho\delta}a_\rho a_\delta - \sum_\rho \tilde{R}_\rho a_\rho - h\delta(y-a_j)} . \tag{4.75}$$

Segundo o método das réplicas, a distribuição que foi definida em (4.70) relaciona-se com $\langle Z^n\rangle$ da seguinte forma

$$\mathcal{P}(y) = -\lim_{\beta\to\infty}\frac{1}{N\beta}\frac{\partial}{\partial h}\langle \ln Z_a\rangle |_{h=0} = -\lim_{\beta\to\infty}\frac{1}{N\beta}\frac{\partial}{\partial h}\lim_{n\to 0}\frac{1}{n}\ln\langle Z_a^n\rangle |_{h=0} , \tag{4.76}$$

o que resulta em

$$\mathcal{P}(y) = -\frac{\partial}{\partial h}\frac{1}{n} extr\{Q\sum_\rho \tilde{Q}_\rho + \sum_{\rho<\delta} q_{\rho\delta}\tilde{q}_{\rho\delta} + \sum_\rho \tilde{R}_\rho R_\rho + \alpha\ln G_1 + \ln G_2\} |_{h=0} . \tag{4.77}$$

Somente G_2 depende de h , conseqüentemente

$$\mathcal{P}(y) = -\frac{1}{n}\frac{\partial}{\partial h}\ln G_2 |_{h=0} . \tag{4.78}$$

Quando a simetria de réplica é imposta obtemos

$$\begin{aligned}
\mathcal{P}(y) &= -\frac{1}{n}\frac{\partial}{\partial h}\ln\int_{-\infty}^{\infty} Dz \left[\int_0^{\infty} da\, e^{-\tilde{Q}a^2 - \tilde{R}a + \frac{\tilde{q}}{2}a^2 + i\sqrt{(\tilde{q})}az - h\delta(y-a)}\right]^n \\
&= -\frac{1}{n}\frac{\partial}{\partial h}\int_{-\infty}^{\infty} Dz \ln\left(\int_0^{\infty} da\, e^{-\tilde{Q}a^2 - \tilde{R}a + \frac{\tilde{q}}{2}a^2 + i\sqrt{(\tilde{q})}az - h\delta(y-a)}\right)^n \\
&= \int_{-\infty}^{\infty} Dz \frac{\delta(y-a)\, e^{-\tilde{Q}a^2 - \tilde{R}a + \frac{\tilde{q}}{2}a^2 + i\sqrt{(\tilde{q})}az}}{\int_0^{\infty} da\, e^{-\tilde{Q}a^2 - \tilde{R}a + \frac{\tilde{q}}{2}a^2 + i\tilde{q}^{\frac{1}{2}}az}} \\
&= \int_{-\infty}^{\infty} Dz \frac{e^{-\left(\tilde{Q}+\frac{\tilde{q}}{2}\right)y^2 - \left(\tilde{R}+\sqrt{\tilde{q}}z\right)y}}{\frac{1}{2}\sqrt{\frac{\pi}{\tilde{Q}+\frac{\tilde{q}}{2}}}\, e^{\frac{(\tilde{R}+\sqrt{\tilde{q}}z)^2}{4(\tilde{Q}+\frac{\tilde{q}}{2})}}\, \mathrm{erfc}\left(\frac{(\tilde{R}+\sqrt{\tilde{q}}z)}{2\sqrt{(\tilde{Q}+\frac{\tilde{q}}{2})}}\right)} .
\end{aligned} \tag{4.79}$$

A seguir vamos considerar apenas o regime onde existem partições perfeitas, isto é, $\alpha < \alpha_c(m)$. Para facilitar a análise vamos agora calcular a função de probabilidade acumulada $C(y)$

$$\begin{aligned} C(y) &= \int_0^y dy' \mathcal{P}(y') \\ &= \int_{-\infty}^{\infty} Dz \frac{\left[\mathrm{erf}\left(\sqrt{\Delta}y + \frac{\Gamma+\sqrt{\xi}z}{2}\right) - \mathrm{erf}\left(\frac{\Gamma+\sqrt{\xi}z}{2}\right)\right]}{\mathrm{erfc}\left(\frac{\Gamma+\sqrt{\xi}z}{2}\right)} \\ &= \int_{-\infty}^{\infty} Dz \left[1 - \frac{\mathrm{erfc}\left(\sqrt{\Delta}y + \frac{\Gamma+\sqrt{\xi}z}{2}\right)}{\mathrm{erfc}\left(\frac{\Gamma+\sqrt{\xi}z}{2}\right)}\right], \end{aligned} \tag{4.80}$$

onde Γ, Δ e ξ são obtidos resolvendo-se a equações (4.62) - (4.66). No limite $q \to 1$ é possível calcular esta integral pois $\Gamma \to \infty$ e $\xi \to \infty$. Temos que determinar primeiro o limite de integração que contribui na integral. Note que para $\sqrt{\Delta}y + \frac{\Gamma+\sqrt{\xi}z}{2} < 0$ o integrando de (4.80) se anula. A função erro complementar pode ser desprezada, portanto encontramos a seguinte expressão para a distribuição acumulada,

$$\begin{aligned} C(y) &= \int_{-2\sqrt{\frac{\Delta}{\xi}}y - \frac{\Gamma}{\sqrt{\xi}}}^{\infty} Dz = 1 - \int_{-\infty}^{-2\sqrt{\frac{\Delta}{\xi}}y - \frac{\Gamma}{\sqrt{\xi}}} Dz \\ &= 1 - \frac{1}{2}\mathrm{erfc}\left(\frac{2\sqrt{\frac{\Delta}{\xi}}y + \frac{\Gamma}{\sqrt{\xi}}}{\sqrt{2}}\right). \end{aligned} \tag{4.81}$$

É possível expressar os expoentes $\sqrt{\frac{\Delta}{\xi}}$ e $\frac{\Gamma}{\sqrt{\xi}}$ em termos de parâmetros conhecidos. Assim, usando $z_0 = \frac{\Gamma}{\sqrt{\xi}}$ e manipulando as relações (4.46) e (4.62) encontramos

$$\tau = \sqrt{\frac{\Delta}{\xi}} = \frac{\sqrt{m^2\alpha_c - z_0^2}}{2m}, \tag{4.82}$$

de modo que (4.81) fica escrita como

$$C(y) = 1 - \frac{1}{2}\mathrm{erfc}\left(\frac{2\tau y + z_0}{\sqrt{2}}\right). \tag{4.83}$$

Para $m = 0$, sabemos que $z_0 = 0$ e τ é simplesmente

$$\tau = \sqrt{\frac{\Delta}{\xi}} = \frac{\Delta(1-q)}{\sqrt{\alpha_c}} = \frac{\sqrt{\alpha_c}}{2} = \frac{1}{2\sqrt{2}},$$

resultando em

$$C(y) = 1 - \frac{1}{2}\text{erfc}\left(\frac{y}{2}\right). \tag{4.84}$$

O que é notável nas distribuições (4.83) e (4.84) é que $C(0) \neq 0$ indicando que $\mathcal{P}(y)$ apresenta uma função delta de Dirac na origem. Uma vez conhecida a função acumulada, podemos derivar imediatamente a função densidade de probabilidade. Neste caso, a relação entre a densidade de probabilidade e a função acumulada é

$$\mathcal{P}(y) = C(0)\delta(y) + \frac{dC(y)}{dy}. \tag{4.85}$$

Daí, efetuando a derivada, chegamos à expressão da densidade de probabilidade para $\alpha = \alpha_c$:

$$\mathcal{P}(y) = \delta(y)\left[1 - \frac{1}{2}\text{erfc}\left(\frac{z_0}{\sqrt{2}}\right)\right] + \sqrt{\frac{2}{\pi}}\tau\, e^{-\frac{(z_0+2\tau y)^2}{2}}. \tag{4.86}$$

Entretanto, abaixo de α_c não há nada que se possa fazer para simplificar a expressão da distribuição acumulada dada na equação (4.80). A figura 4.4 mostra a distribuição acumulada para partições balanceadas $m = 0$. Note que C(0)$\neq$ 0 apenas para $\alpha = \alpha_c$.

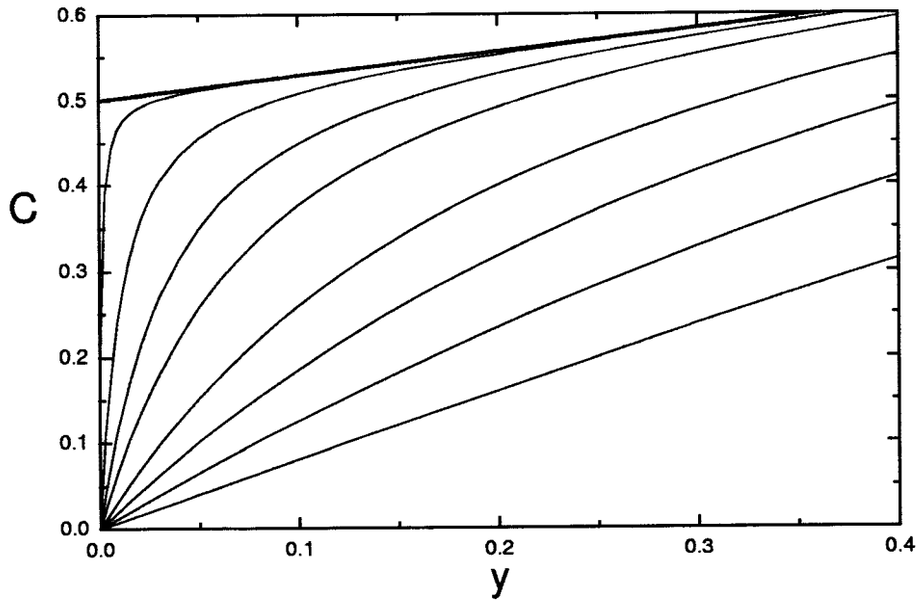


Figura 4.4: Distribuição de probabilidade acumulada de entradas de realizações com energia zero para partições balanceadas para $\alpha = 0.01$, 0.25, 0.36, 0.42, 0.48, 0.49 *e* 0.499, vistos de baixo para cima. A curva em vermelho corresponde à distribuição acumulada quando $\alpha = \alpha_c$.

Capítulo 5

Conclusão

A motivação principal desta tese foi explorar a utilização das idéias e ferramentas desenvolvidas na Mecânica Estatística de sistemas desordenados para estudar o clássico problema da partição numérica **PPN**. Isso é possível neste caso porque podemos facilmente formular esse problema de otimização combinatorial como um problema de spins de Ising com interações aleatórias.

Existem vários motivos que tornam o **PPN** um problema de otimização combinatorial ideal para ser investigado tanto analítica como numericamente. Primeiro, ele é um dos principais representantes da classe de problemas $\mathcal{NP}$-completos, ou seja, é um problema que pode ser reduzido polinomialmente a todos os problemas daquela classe. Segundo, apesar de apresentar uma formulação matemática muito simples, o espaço de soluções do **PPN** é dotado de uma estrutura extremamente rica, levando a diversas propriedades não triviais das soluções ótimas. Terceiro, essa simplicidade facilita muito o estudo teórico-analítico, bem como a elaboração de novos algoritmos.

Podemos neste momento perguntar o porquê do uso da mecânica estatística neste tipo de estudo. O principal objetivo dos pesquisadores na área específica de otimização combinatorial é a procura de evidências, propriedades e 'insights' que ajudem a provar as conjecturas $\mathcal{P} = \mathcal{NP}$ ou $\mathcal{P} \neq \mathcal{NP}$. É quase consenso entre aqueles pesquisadores que um novo conjunto de ferramentas matemáticas deverá ser desenvolvido para que possamos de fato resolver essa questão que, sem dúvida, é uma das mais desafiadoras da atualidade na ciência da computação. Deste ponto de vista, a mecânica estatística representa uma alternativa possível, merecedora de atenção, para o desenvolvimento de tais técnicas matemáticas mais eficientes para o estudo de problemas de otimização.

Em geral é muito difícil usar a teoria formal de probabilidade para derivar expressões analíticas para a função custo ótima ou então caracterizar as soluções ótimas de um dado problema de otimização. Em 1997, quando iniciamos os trabalhos que resultaram nesta tese, havia poucos resultados analíticos disponíveis na literatura sobre o problema da partição numérica. Em particular, Karmarkar e colaboradores [40], utilizando métodos rigorosos de teoria de probabilidade, conseguiram prever somente a ordem de grandeza da energia ou função custo ótima esperada para o **PPN**, a saber, $\mathcal{O}(\sqrt{N}2^{-N})$ para o caso de partições sem vínculo de cardinalidade e $\mathcal{O}(N2^{-N})$ para partições balanceadas. É importante mencionar também o trabalho pioneiro de Gent e Walsh [66] que, através de simulação computacional e aproximações do tipo 'annealed', sinalizou a existência de uma transição de fase no **PPN**.

A análise de mecânica estatística do problema da partição numérica apresentada nesta tese pode ser dividida em duas partes, a saber, uma abordagem do problema direto e uma abordagem do problema inverso. A seguir vamos resumir os principais resultados e conclusões dessas duas abordagens.

Na abordagem direta formulamos o **PPN** como um modelo de spins de Ising $s_i = \pm 1; i = 1, \ldots, N$ com interações J_{ij} antiferromagnéticas de intensidades aleatórias. Mais especificamente, as configurações de spins de Ising representam as partições ($s_i = 1$ caso o item i pertença ao conjunto A e $s_i = -1$ caso contrário) e os acoplamentos são dados por produtos dos valores dos itens, $J_{ij} = a_i a_j$. Em particular, consideramos os itens $a_i; i = 1, \ldots, N$ como variáveis aleatórias estatisticamente independentes geradas segundo uma distribuição de probabilidade uniforme no intervalo unitário. Dessa forma, a energia ou função custo passa a ser uma variável aleatória. No capítulo 2 derivamos um limitante inferior para a energia ótima média na aproximação 'annealed' (recozida) tanto para o caso com vínculo quanto para o caso sem vínculo de cardinalidade. Ainda, através de simulações numéricas mostramos que a largura relativa da distribuição das energias ótimas E^*, denotada por r (ver equação (1.23)) tende para o valor 1 no limite $N \to \infty$ tanto para a versão sem vínculo quanto para a versão balanceada. Isso significa que as flutuações típicas de E^* são da mesma ordem de grandeza de seu valor esperado e, portanto, a energia ótima não é uma grandeza automediante. Particularmente relevante é o resultado de que quando o valor do parâmetro m, que mede a diferença

(apropriadamente escalada) entre as cardinalidades dos dois conjuntos, está acima de um valor crítico dado por $m_c = \sqrt{2} - 1$ a energia ótima torna-se extensiva, permitindo então a realização do cálculo exato das propriedades termodinâmicas das soluções ótimas do **PPN**.

Ainda no capítulo 2 e também ilustrando o poder das ferramentas padrões da mecânica estatística de sistemas desordenados, calculamos o número médio de mínimos locais, ou seja, estados que poderiam ser 'armadilhas' para heurísticas de busca que minimizam localmente a função custo. Para partições sem vínculo de cardinalidade encontramos que essa grandeza média escala com $2.764\frac{2^N}{N^{3/2}}$ para N grande, e que a energia típica desses mínimos locais é aproximadamente $2/N$ (lembremos que a energia de um mínimo global é da ordem de $\sqrt{N}2^{-N}$). É interessante comparar esses resultados com os do modelo de Sherrington-Kirkpatrick (**SK**) para o qual o limitante inferior da energia do estado fundamental é $-0.833N$ [9] e o número médio de mínimos locais escala com $e^{0.199N}$ [48]. Portanto, no **PPN** existem muito mais mínimos locais e os ótimos globais são muito mais profundos que os encontrados no modelo **SK**. Esses resultados ajudam a explicar o fracasso dos algoritmos de busca local, tal como o popular 'simulated annealing', em produzir boas soluções para o PPN [16].

Uma técnica tradicional de se obter limitantes inferiores rigorosos para a função custo de problemas de programação inteira, como é o caso do **PPN**, é através da relaxação dos vínculos. O capítulo 3 foi dedicado ao estudo da versão contínua ou esférica do problema da partição numérica em que os spins de Ising passam agora a spins esféricos, ou seja, $-\infty < s_i < \infty$ mas $\sum_i s_i^2 = N$. Neste caso, dizemos que o vínculo da integralidade foi relaxado. Como era esperado, foi possível resolver o problema relaxado analiticamente e obter expressões bastante simples para as diversas grandezas termodinâmicas de interesse. Infelizmente, o valor esperado da energia ótima (caso sem vínculo de cardinalidade) obtido foi zero, resultando em um limitante inferior sem nenhuma utilidade. Entretanto, algumas propriedades estatísticas das soluções ótimas mostraram-se muito interessantes. Por exemplo, mostramos que a maior concentração de soluções ótimas possui 'partições' balanceadas $m = 0$, conforme refletido no pico da distribuição das diferenças de cardinalidade das soluções ótimas $P(m)$ que aparece em $m = 0$. Com o auxílio dessa distribuição mostramos também que $|m|$ não é uma grandeza automediante. Ainda, através do cálculo da função de correlação que mede a sobreposição de pares de

mínimos globais mostramos que para $m > 0.5$ há somente um estado com energia mínima (mínimo global). Finalizamos o capítulo 3 e a primeira parte da tese com um resultado de interesse prático, a saber, a análise do desempenho de uma heurística que denominamos heurística de integralização. Nessa heurística geramos as partições, isto é, as configurações de Ising, simplesmente tomando o sinal dos spins esféricos da solução ótima relaxada. Mostramos analiticamente que tal heurística é tão ineficiente quanto uma heurística que escolhe os estados no espaço de configurações de Ising de forma aleatória. Todos os resultados discutidos acima representaram contribuições originais para o entendimento do problema da partição numérica.

Para motivar a formulação do problema inverso, que é o objeto da segunda parte da tese (capítulo 4), observamos primeiramente que a abordagem tradicional da ciência da computação para a validação de algoritmos concentra-se quase que exclusivamente no espaço de realizações (ou instâncias), buscando por instâncias em que o algoritmo apresente o mais pobre desempenho (análise do pior caso). Por sua vez, as análises de mecânica estatística, conforme ilustrado na primeira parte desta tese, têm se concentrado principalmente no espaço de configurações, sendo as instâncias amostradas segundo uma distribuição de probabilidade dada *a priori* [25]. Inspirados pelo trabalho de Elizabeth Gardner em redes neurais [46,47], investigamos as propriedades estatísticas do espaço de realizações do problema da partição numérica, dado um conjunto fixo de partições. Especificamente, procuramos nesse espaço pelas instâncias mais fáceis a fim de mostrar que existe um número máximo de partições perfeitas aleatórias $\alpha_c(m)N$. Em particular, para partições balanceadas ($m = 0$) encontramos $\alpha_c(0) = \frac{1}{2}$. Mais ainda, mostramos que partições perfeitas aleatórias desbalanceadas ($m \neq 0$) existem somente se a diferença das cardinalidades dos dois conjuntos $\sum_i s_i$ escalar com $mN^{1/2}$. Esses resultados podem ter relevância para a criptografia [77], pois fornecem limitantes superiores para o número de mensagens binárias descorrelacionadas (partições perfeitas aleatórias) que podem ser codificadas em uma instância arbitrária. Outro resultado obtido no capítulo 4 que também pode ser de interesse em criptografia é a distribuição de probabilidade dos itens a_k de uma realização que suporta partições perfeitas, desde que sua forma pode dar indícios quanto ao número de padrões codificados (αN) na instância.

É importante tentar descobrir se o limitante superior para o número de partições perfeitas *aleatórias* ($\alpha_c(m)$) que derivamos nesta tese também dá alguma informação sobre o número de partições perfeitas arbitrárias (correlacionadas). Primeiramente, observamos que é mais difícil encontrar soluções (instâncias $\mathbf{a}$ com energia zero) para o conjunto de equações lineares $\mathbf{a} \cdot \mathbf{s}^l = 0; l = 1, \ldots, \alpha N$ no caso de partições $\mathbf{s}^l$ aleatórias do que no caso correlacionado. Conseqüentemente, se para $\alpha \leq \alpha_c(m)$ podemos garantir a existência daquelas soluções para partições aleatórias, também asseguramos a existência de soluções para partições correlacionadas. Isso leva a conclusão de que $\alpha_c(m)N$ fornece um limitante inferior para o número máximo de partições perfeitas que podem ser encontradas em uma instância arbitrária. É claro que esse limitante torna-se exato para o caso de partições aleatórias. Portanto, podemos considerar o limitante $\alpha_c(m)N$ como uma característica geral e robusta do problema da partição numérica, no sentido de que não depende de nenhuma hipótese arbitrária sobre as instâncias e/ou partições. Novamente, devemos enfatizar que os resultados do estudo do problema inverso apresentados nesta tese são originais e acreditamos que esse tipo de análise possa ser extendida com o mesmo êxito para outros problemas de otimização.

Concluindo, desejamos mencionar que a perspectiva de continuação desse trabalho concentra-se na generalização do problema da partição numérica para mais de dois conjuntos, que possivelmente pode ser mapeado em um modelo de spins de Potts. A perspectiva futura é então reproduzir o que for possível da análise realizada nesta tese para o problema generalizado [98].

Apêndice A

Complexidade Computacional

A.0.1 Motivação

O ramo da ciência da computação teórica conhecido como complexidade computacional concentra-se na classificação dos problemas de acordo com os recursos computacionais (tempo de CPU, memória) requeridos para resolvê-los. Isto levou a uma noção precisamente definida de problemas *tratáveis* e *intratáveis*. Esta noção poderia ser transferida para problemas de *solubilidade* de modelos em Mecância Estatística, onde não entendemos porque o modelo de Ising é exatamente solúvel em duas dimensões, mas não foi ainda resolvido em três.

O sucesso da teoria de complexidade computacional é baseado em atitudes pessimistas: a intratabilidade de um problema é definida com base no pior caso (ou pior exemplo) possível. Freqüentemente, no caso de realizações aleatórias do problema, o cenário da pior instância difere consideravelmente da instância típica, ou seja, aquela escolhida aleatóriamente no conjunto de realizações. Uma observação comum é que problemas difíceis são tipicamente fáceis de resolver. Para conseguir exemplos reais difíceis, os parâmetros do conjunto fundamental que determina a distribuição de probabilidade das realizações devem ser cuidadosamente ajustados para certos valores críticos. Variando os parâmetros através da região crítica observa-se uma mudança abrupta na complexidade típica – muito similar às mudanças abruptas associadas às transições de fase em sistemas físicos. Transições de fase em sistemas físicos e em sistemas computacionais são bem estudados dentro do esquema da Mecânica Estatística. Além das transições de fase, a Mecânica Estatística oferece meios para análises probabilísticas gerais de problemas computacionais. O procedimento usual é formular um problema

de otimização como um vidro de spin e analisar as propriedade em baixa temperatura. Esta abordagem física freqüentemente fornece resultados que vão além dos obtidos pelos métodos tradicionais.

Uma outra estimulante conexão entre Física e Ciência da Computação é fornecida pela *computação quântica.* Há algumas evidências teóricas de que computadores projetados para explorar as propriedades de sistemas quânticos como artifícios computacionais são mais poderosos do que computadores baseados em conceitos clássicos. A esperança é que problemas considerados intratáveis em computadores clássicos se tornem tratáveis quando colocados em um computador quântico. Em última instância, algum conhecimento de complexidade computacional nos ajudará a entender as promessas e as limitações dos computadores quânticos.

A.0.2 A medida da complexidade

Algoritmos

A complexidade computacional de um problema é uma medida dos recursos computacionais, tipicamente tempo, requeridos para resolver o problema. O que pode ser medido (ou computado) é o tempo que um particular *algoritmo* usa para resolver o problema. Este tempo, por sua vez, depende da implementação do algoritmo bem como do computador que o executa. A teoria da complexidade computacional fornece uma noção de complexidade que é totalmente independente dos detalhes de implementação e do computador à disposição. Sua definição precisa requer um considerável formalismo. Isto não é uma surpresa, desde que está relacionado a uma questão altamente não trivial que toca nos fundamentos da matemática: *O que entendemos ao dizermos que um problema é solúvel?* Pensando nesta questão chegamos ao teorema da incompleteza de Gödel [97], maquina de Turing e a Tese de Church-Turing em funções computáveis (veja [79] para uma introdução nestes tópicos). Neste apêndice adotaremos um ponto de visto mais informal e pragmático. Um problema é solúvel se ele pode ser resolvido por um computador escrito em sua linguagem de programação favorita. O tempo de execução ou a *complexidade temporal* de seu programa deve ser então definida com algum cuidado para servir como uma medida significativa para a complexidade do problema.

Complexidade temporal

Em geral o tempo de execução depende do tamanho do problema e de uma entrada específica de dados (a realização). Por exemplo, ordenar 1000 números leva mais tempo do que ordenar 10 números e alguns algoritmos executam mais rapidamente se os dados estiverem parcialmente ordenados. Para minimizar a dependência em uma realização específica considere a *complexidade temporal do pior caso* $T(n)$,

$$T(n) = \max_{|x|=n} t(x) \tag{A.1}$$

onde $t(x)$ é o tempo de execução do algoritmo para entrada de dados x (em unidades arbitrárias) e o máximo é tomado sobre todas as realizações de tamanho n. O pior tempo é um limitante superior para o tempo de execução observável.

Uma medida do tempo de complexidade deveria ser baseada em unidades de tempo independentes do **clock** de uma CPU específica. Tal unidade é fornecida pelo tempo que a CPU gasta para desenvolver uma operação elementar, como a adição de dois números inteiros. Medir o tempo nestas unidades significa contar o número de operações elementares executados pelo seu algoritmo. Este número, por sua vez, depende fortemente de detalhes de implementação do algoritmo – programadores experientes e compiladores otimizados tentarão reduzí-lo. Portanto, não consideraremos o número preciso $T(n)$ de operações elementares, mas somente o comportamento assintótico de $T(n)$ para grandes valores de n denotado pelo símbolo $\mathcal{O}$. Essa notação é usada para expressar a complexidade do tempo de execução de um algoritmo, em função de um valor n que é correspondente ao tamanho da entrada do problema. Assim , $\mathcal{O}(f(n))$ indica que o tempo necessário para obter um resultado do algoritmo é proporcional a $f(n)$. Essa medida tem por objetivo avaliar a eficiência do algoritmo em termos de sua complexidade, normalmente considerada no pior caso. O valor de $f(n)$ não tem uma relação exata com o tempo e sim com sua ordem de grandeza, o que serve de parâmetro de comparação entre diferentes algoritmos. Por exemplo, para valores de n suficientemente grandes as funções $f_1(n) = n^2$, $f_2(n) = 2n^2 + 3$ e $f_3 = 3n^2$, têm a mesma ordem de grandeza, a qual é proporcional a n^2, ou seja $\mathcal{O}(n^2)$. Por definição, se existir uma função $g(n)$ tal que $f(n) \leq cg(n)$ para uma constante $c > 0$, diremos que $\mathcal{O}(f(n)) = \mathcal{O}(g(n))$. Normalmente a função g representa uma família de funções

Complexidade	$n = 10$	$n = 50$	$n = 100$	$n = 1000$	$n = 10000$
$\mathcal{O}(\log_2 n)$	0.0000033	0.0000056	0.0000066	0.0000099	0.000013
$\mathcal{O}(n)$	0.000010	0.000050	0.000100	0.000100	0.010000
$\mathcal{O}(n \log_2 n)$	0.000033	0.000283	0.000684	0.009966	0.132877
$\mathcal{O}(n^2)$	0.000100	0.002500	0.01000	1 s	1min 40 s
$\mathcal{O}(n^3)$	0.001000	0.125000	1 s	16 min 40 s	11 dias 14h
$\mathcal{O}(n^5)$	0.100000	5 min 13s	2h 47 min	32 anos	∞
$\mathcal{O}(2^n)$	0.001024	35 anos	$4\ 10^{16}$ anos	∞	∞
$\mathcal{O}(3^n)$	0.059049	$2\ 10^{10}$ anos	∞	∞	∞

Tabela A.1: A ordem de complexidade versus o tempo de processamento. A unidade, quando não especificada está em segundos.

conhecida como: $\mathcal{O}(n)$, $\mathcal{O}(n^2)$, $\mathcal{O}(n^3)$, $\mathcal{O}(e^n)$, $\mathcal{O}(n!)$, $\mathcal{O}(\log_2 n)$, $\mathcal{O}(n \log_2 n)$, etc. A notação $\mathcal{O}(1)$ é usada para indicar uma complexidade constante, ou seja , independente de n; no entanto isto é bastante incomum. É instrutivo inspecionar a tabela 1 para termos uma idéia da grandeza e fazermos uma comparação entre funções comumente usadas na medição da complexidade de algoritmo para alguns valores de n. Para melhor visualização, os valores indicados representam o tempo médio esperado de execução, expressos em segundos quando não indicados. Dessa tabela, podemos concluir que para valores de n não superiores a 1000 a complexidade cúbica $\mathcal{O}(n^3)$ seria a máxima admitida. Para grandes conjuntos de dados ($n \geq 10000$) o ideal seria a complexidade não superior a $\mathcal{O}(n^2)$. Em geral, $\mathcal{O}(n^m)$ para $m > 3$ ou complexidade exponencial são indesejáveis. Outro fator a ser considerado, por curiosidade, seria o impacto do aumento da velocidade do computador sobre o tempo de execução do algoritmo. Se em um tempo fixo t executamos um problema de tamanho n e complexidade $\mathcal{O}(n^3)$, então um aumento em 1000 vezes na velocidade de um computador nos permite resolver um problema 10 vezes maior no mesmo intervalo de tempo. Por outro lado, se a complexidade fosse $\mathcal{O}(2^n)$, o aumento no tamanho do problema seria de apenas 10 unidades (valor constante).

Multiplicando duas matrizes $n \times n$ requer n^3 multiplicações, de acordo com a fórmula do livro texto, isto significa que o problema de multiplicar duas matrizes $n \times n$ tem complexidade $\mathcal{O}(n^3)$? Não. A fórmula do livro texto é um particular algoritmo, e o tempo de complexidade de um algoritmo é somente um *limitante superior* para a complexidade inerente ao problema. De fato, algoritmos mais

rápidos para multiplicação de matrizes com complexidade $\mathcal{O}(n^\alpha)$ e $\alpha < 3$ foram encontrados nas últimas décadas, e o recorde atual está sendo $\alpha = 2.376$ [80]. Desde que o produto de matrizes tem n^2 entradas, α não pode ser menor do que 2, e é uma questão aberta se este limite inferior poderia ser atingido por algum algoritmo. Outro problema onde o limitante superior para a complexidade de um algoritmo encontra um limitante inferior inerente é o ordenamento de n ítens. Sob hipóteses gerais, as comparações entre pares de ítens são as únicas fontes de informação e pode ser mostrado que $\mathcal{O}(n \log n)$ é o limitante inferior para o número de comparações necessária para se ordenar n ítens no pior caso [78, chap. 9]. Este limite é encontrado pelos algoritmos heapsort or mergesort.

Tamanho do problema

Nossa medida do tempo de complexidade depende ainda de uma noção um pouco ambígua do tamanho do problema. No exemplo da multiplicação de matrizes, tacitamente tomamos o número n de linhas de entrada de uma matriz como uma medida 'natural' do tamanho. Usando o número de elementos $m = n^2$ 'aceleramos' o algoritmo $\mathcal{O}(n^3)$ para $\mathcal{O}(m^{3/2})$ sem mudar uma simples linha do código do programa. Veja que uma definição não ambígüa do tamanho do problema exige comparar algoritmos. Em complexidade computacional, todos os problemas que podem ser resolvidos por um algoritmo polinomial, isto é, um algoritmo com tempo de complexidade $\mathcal{O}(n^k)$ para algum k, são colocados juntos e chamados de *tratáveis* . Por outro lado, problemas que podem somente ser resolvidos por algoritmos com tempo de execução não polinomial como $\mathcal{O}(2^n)$ ou $\mathcal{O}(n!)$ são também agrupados em outra família e recebem denominação de *intratáveis*. Existem motivos práticos e teóricos para essa classificação um tanto grosseira. Uma das vantagens é que ela não distingüe entre $\mathcal{O}(n^3)$ e $\mathcal{O}(n^{3/2})$. Conseqüentemente podemos nos dar ao luxo de ser superficiais e assim evitar a definição da medida 'natural' do tamanho do problema.

A.0.3 Problemas tratáveis e intratáveis

Crescimento polinomial vs. exponencial

Do ponto de vista prático, um algoritmo exponencial significa um limite extremamente restritivo para o tamanho acessível do problema. Suponha que com um dado equipamento possamos resolver

um problema de tamanho n dentro de um determinado prazo. Se o algoritmo tem complexidade $\mathcal{O}(2^n)$, um problema de tamanho $n+1$ precisará duas vezes mais tempo, trazendo-o definitivamente para fora do prazo. O aumento causado por um algoritmo $\mathcal{O}(n)$ ou $\mathcal{O}(n^2)$ por outro lado é menos dramático e pode ser facilmente compensado por uma atualização de hardware. Pode-se alegar que um algoritmo $\mathcal{O}(n^{100})$ tem um desempenho melhor do que um algoritmo $\mathcal{O}(2^n)$ somente para tamanhos de problemas que nunca ocorrerão em suas aplicações. Um algoritmo polinomial para um problema é tipicamente $\mathcal{O}(n^k)$, $k = 1, 2$ ou 3. Algoritmos polinomiais com $k > 10$ são raros e surgem em problemas bastante exóticos.

Árvores

Como exemplo, considere o seguinte problema de projeto de redes. Você tem uma empresa com muitos escritórios e você quer arrendar linhas telefônicas para conectá-las. A companhia telefônica cobra diferentes preços para conectar pares de cidades, e sua tarefa é selecionar um conjunto de linhas que conecte todos os seus escritórios a um custo total mínimo.

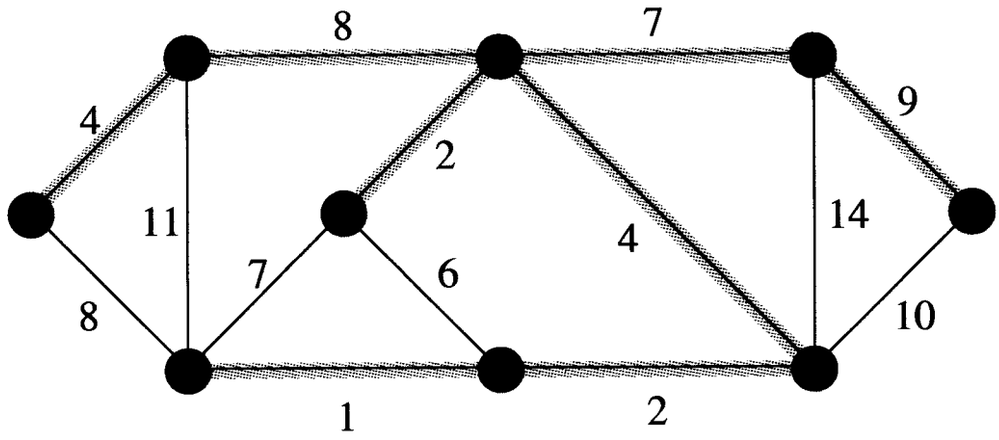


Figura A.1: Um grafo valorado e sua árvore geradora mínima (arestas coloridas).

Em termos matemáticos, as cidades e as linhas entre elas formam os vértices V e as arestas E, respectivamente, de um grafo valorado $G = (V, E)$, onde o valor de uma aresta é o custo da correspondente linha a ser arrendada. Sua tarefa é encontrar um subgrafo que conecte todos os vértices do grafo, isto é, um subgrafo gerador, cujas arestas tenham um peso total mínimo. Seu subgrafo não deve conter ciclos, já que você pode remover uma aresta do ciclo deixando todos os nós conectados e reduzindo

o custo. Um grafo sem ciclos é uma árvore, então o que você está procurando é uma *árvore geradora mínima* em um grafo valorado (Fig. A.1).

> ÁRVORE GERADORA MÍNIMA (AGM): Dado um grafo valorado $G = (V, E)$ com pesos não negativos. Encontre um árvore geradora $T \subseteq G$ com peso total mínimo.

Como encontrar uma árvore geradora ? Uma abordagem ingênua é gerar todas as árvores possíveis com n vértices e permanecer com aquela que tem menor peso. A enumeração de todas as árvores pode ser feita usando os códigos de Prüfer [81], mas a fórmula de Cayley nos diz que existem n^{n-2} diferentes árvores com n vértices. Já para $n = 100$ existem mais árvores do que átomos no universo! Conseqüentemente uma busca exaustiva é inviável mesmo para problemas pequenos. A idéia matemática que torna o AGM um problema tratável é este:

> **Lema**: Seja $U \subset V$ qualquer subconjunto de vértices de $G = (V, E)$, e seja e a aresta com menor peso de todas conectando U e $V - U$. Então e é parte de uma árvore geradora mínima.

Prova: Por contradição. Suponha que T seja uma árvore geradora mínima que não contenha e. Seja $e = (u, v)$ com $u \in U$ e $v \in V - U$. Então, devido a T ser uma árvore geradora, ela contém um único caminho de u para v, que juntos com e formam um ciclo em G. Este caminho tem que incluir uma outra aresta f conectando U e $V - U$. Agora $T + e - f$ é uma outra árvore geradora com peso total menor do que T. Então T não era uma árvore geradora mínima.

O lema permite crescer uma árvore geradora aresta por aresta usando o algoritmo de Prim. Por exemplo:

PRIM(G)
Input: grafo valorado $G(V, E)$
Output: árvore geradora mínima $T \subseteq G$
begin
  Seja T um vértice simples v de G
  while T tem menos do que n vértices
    encontre a aresta mínima conectando T a $G - T$
    adicione ela a T
  end
end

O tempo de complexidade do algoritmo de Prim depende da estrutura de dados usada para organizar as arestas, mas em qualquer caso $\mathcal{O}(n^2 \log n)$ é um limite superior. (veja [82] para algoritmos mais rápidos). Equipados com tal algoritmo polinomial podemos encontrar a árvore geradora mínima com milhares de nós dentro de segundos em um microcomputador. Compare isto com uma busca exaustiva! De acordo com nossa definição AGM é um problema tratável.

Itinerários intratáveis

Encorajados pelo algoritmo eficiente para o AGM iremos agora investigar um problema similar. Sua tarefa é planejar um itinerário para o caixeiro viajante que deve visitar n cidades. Você tem um mapa com todas as cidades e distância entre elas. Em qual ordem o viajante deveria visitar as cidades para minimizar a distância total da viagem? Numere as cidades arbitrariamente e anote a distância em uma matriz (d_{ij}), onde d_{ij} denota a distância entre a cidade i e a cidade j. Um percurso é dado por uma *permutação cíclica* $\pi : [1 \ldots n] \mapsto [1 \ldots n]$, onde $\pi(i)$ denota a sucessora da cidade i, e seu problema pode ser definido como:

PROBLEMA DO CAIXEIRO VIAJANTE (PCV): Dada uma matriz distância $n \times n$ com elementos $d_{ij} \geq 0$. Encontre uma permutação cíclica $\pi : [1 \ldots n] \mapsto [1 \ldots n]$ que minimiza

$$c(\pi) = \sum_{i=1}^{n} d_{i\pi(i)}. \tag{A.2}$$

O PCV provavelmente é o mais famoso problema de otimização e existe uma vasta literatura especialmente devotada a ele, veja [74, 75] e suas referências. Não é muito difícil encontrar boas soluções, mesmo para problemas grandes, mas como obter a *melhor* solução para uma realização qualquer? Existem $(n-1)!$ permutações cíclicas, e o cálculo do comprimento de um percurso pode ser feito em tempo $\mathcal{O}(n)$, conseqüentemente uma busca exaustiva tem complexidade $\mathcal{O}(n!)$. Novamente esta abordagem é limitada mesmo para pequenas amostras. Será que existe uma idéia matemática que nos forneça um atalho para a solução ótima, como no AGM? Ninguém sabe! Apesar dos esforços de muitas pessoas brilhantes, nenhum algoritmo polinomial para o PCV foi encontrado. Existem alguns algoritmos engenhosos e eficientes (isto é, polinomial) que encontra boas soluções, mas que não garantem a otimalidade [75]. De acordo com nossa definição, o o problema PCV é intratável.

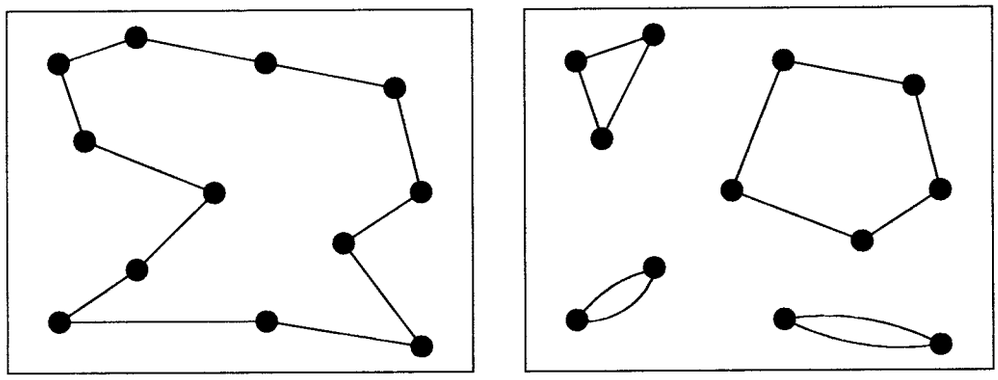


Figura A.2: Mesma instância, diferentes problemas: Uma configuração válida do PROBLEMA DO CAIXEIRO VIAJANTE (esquerda) e o problema DESIGNAÇÃO (direita). Enquanto DESIGNAÇÃO pode ser resolvido em tempo polinomial, o PCV é intratável.

Por que o PCV é intratável? Novamente ninguém sabe. Não há nenhuma *prova* que exclua a existência de um algoritmo polinomial para o PCV, então pode ser que algum dia alguém venha como um algoritmo polinomial e o correspondente 'insight' matemático, embora isso seja muito pouco provável. A intratabilidade do PCV causa espanto a todos que se deparam com um problema quase idêntico :

DESIGNAÇÃO: Dada uma matriz custo $n \times n$ com elementos $d_{ij} \geq 0$. Encontre uma permutação $\pi : [1 \ldots n] \mapsto [1 \ldots n]$ que minimize

$$c(\pi) = \sum_{i=1}^{n} d_{i\pi(i)}. \tag{A.3}$$

A única diferença entre o PCV e o problema da DESIGNAÇÃO é que o último permite todas as permutações de n ítens, e não somente as cíclicas. Se d_{ij} denota a distância entre cidades, o problema da DESIGNAÇÃO corresponde a minimização do percurso total para um número variável de viajantes, onde cada um tem o seu próprio *sub-percurso* (Fig. A.2).

A aplicação clássica do problema DESIGNAÇÃO é a atribuição de n tarefas para n trabalhadores, sujeitos ao vínculo de que cada trabalhador receba exatamente uma tarefa. Seja d_{ij} o custo da tarefa j quando realizada pelo operário i, e $\pi(i)$ denota a tarefa atribuída ao operário i, a DESIGNAÇÃO é o problema de minimizar o custo total. Existem $n!$ possíveis atribuições de n tarefas a n trabalhadores, conseqüentemente a enumeração exaustiva está proibida. Em contraste ao problema PCV, o problema DESIGNAÇÃO pode ser resolvido em tempo polinomial, por exemplo usando o algoritmo húngaro $\mathcal{O}(n^3)$ [42].

A.1 Classes de Complexidade

A.1.1 Problemas de Decisão

Até aqui temos discutido problemas de otimização: resolver ÁRVORE GERADORA MÍNIMA, PCV ou DESIGNAÇÃO implica que nós comparamos um número exponencial de soluções factíveis umas com as outras e retemos a solução ótima (a busca exaustiva faz isto explicitamente). Talvez possamos aprender mais sobre a barreira que separa os problemas *tratáveis* dos problemas *intratáveis* se investigássemos versões mais simples, ou seja, problemas para os quais as soluções são reconhecidas sem comparações explícitas ou implícitas de todas as soluções factíveis. Então vamos considerar *problemas de decisão* cuja solução é ou 'sim' ou 'não'.

Qualquer problema de otimização pode ser convertido em um problema de decisão pela adição de um limite B para a realização. Exemplos:

AGM (DECISÃO): Dado um grafo valorado $G = (V, E)$ com pesos não negativos e um número $B \geq 0$. G contém uma árvore geradora T com peso total $\leq B$?

PCV (DECISÃO): Dada uma matriz distância $n \times n$ com elementos $d_{ij} \geq 0$ e um número

$B \geq 0$. Existe um percurso π com comprimento $\leq B$?

Em um problema de decisão, as soluções factíveis não são avaliadas umas em relação às outras, mas através de um critério 'absoluto': Um percurso no PCV ou tem comprimento $\leq B$ ou não. A AGM(D) pode ser resolvido em tempo polinomial já que basta simplesmente resolver a variante otimização AGM e comparar o resultado ao parâmetro B. Para o PCV(D) esta abordagem não ajuda. De fato veremos na seção A.1.3 que existe um algoritmo polinomial para PCV(D) se e somente se existe um algoritmo polinomial para PCV. Vamos considerar outros problemas de decisão que não derivam de um problema de otimização e tentar aprender mais sobre a barreira que separa os problemas intratáveis daqueles ditos tratáveis.

Circuito euleriano

Nosso primeiro e 'genuino' problema de decisão data do século XVIII quando a cidade de Königsberg (agora Kaliningrado) tinha sete pontes que atravessavam o rio Pregel (Fig. A.3). Uma charada popular da época perguntava se era possível andar através da cidade atravessando cada uma das pontes exatamente uma vez e retornar à casa.

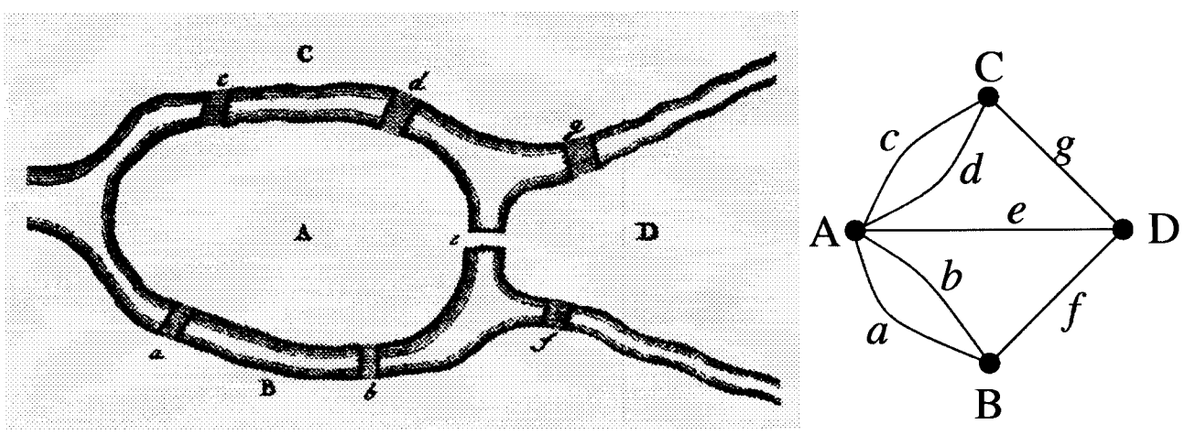


Figura A.3: As sete pontes de of Königsberg, desenhada como no artigo de Euler de 1736 [88] (esquerda) e representado como um grafo (direita). No grafo, a terra e as ilhas são condensadas em pontos (vértices) e cada ponte é desenhada como uma linha (arestas).

Foi Leonhard Euler quem resolveu este enigma em 1736 [88]. Primeiramente Euler reconheceu que para a solução do problema importava somente o padrão de interconexão dos aterros e ilhas – um grafo

$G = (V, E)$ em termos modernos. O grafo correspondente ao enigma das pontes de Königsberg tinha 4 vértices para representar dois aterros e duas ilhas e 7 arestas para representar as pontes (Fig. A.3). O artigo de Euler que tratava do problema da ponte de Königsberg pode ser considerado como a origem da teoria dos grafos.

Para generalizar o problema da ponte de Königsberg precisamos de alguma terminologia da teoria dos grafos [81]. Um *caminho* em um grafo $G = (V, E)$ é uma seqüência alternada de vértices $v \in V$ e arestas $(v, v') \in E$,

$$v_1, (v_1, v_2), v_2, (v_2, v_3), \ldots, (v_{l-1}, v_l), v_l.$$

Note que a seqüência começa e termina com um vértice, e cada aresta é incidente com cada vértice imediatamente anterior e posterior a ele. Um caminho é dito *fechado* se $v_l = v_1$; caso contrário é *aberto*. Um caminho é chamado *percurso* se todas as suas arestas são distintas, e um percurso fechado é chamado de *circuito*. O que os andarilhos em Königsberg tentaram encontrar foi um circuito contendo todas os vértices. Em homenagem a Leonhard Euler tal circuito é chamado de *circuito euleriano*. Equipado com toda essa terminologia estamos prontos para generalizar o problema da ponte de Königsberg :

> CIRCUITO EULERIANO: Dado um grafo $G = (V, E)$. Perguntamos: G contém um circuito euleriano ?

Obviamente este é um problema de decisão: A resposta é 'sim' ou 'não', e qualquer circuito pode ser verificado ser Euleriano ou não sem a necessidade de inspecionar todos os circuitos possíveis.

Mais uma vez a busca exaustiva resolveria este problema, mas a intratabilidade deste problema fôra já observada por Euler. Mais de 200 anos antes do advento do computador ele escreveu "O problema particular das sete pontes de Königsberg poderia ser resolvido pela enumeração cuidadosa de todos os caminhos, desse modo determina-se por inspeção qual deles, se houver, satisfaz as exigências. Este método de solução, contudo, é muito tedioso e também difícil devido ao grande número de combinações possíveis, e em outros problemas onde muitas pontes podem estar envolvidas ele não se aplicaria."(citado em [87]).

Euler resolveu o problema da ponte de Königsberg não pela listagem de todos os percursos possíveis

e sim através de uma intuição matemática. Ele notou que em um circuito você deve partir de um vértice via uma aresta diferente daquela que leva você de volta a ele. Em outras palavras, o grau do vértice (que é o número de arestas adjacentes ao vértice) deve ser par. Esta é obviamente uma condição necessária, mas Euler provou que ela é também suficiente:

> *Teorema:* Um grafo conectado $G = (V, E)$ contém um circuito euleriano se e somente se o grau de todos os vértices $v \in V$ é par.

O teorema de Euler nos permite imaginar um algoritmo polinomial para CIRCUITO EULERIANO : Inspecione o grau de todos os vértices no grafo. Se um vértice tiver grau ímpar, retorne 'não'. Se todos os vértices chegados tiverem grau par, retorne 'sim'. O tempo de execução deste algoritmo depende do código do grafo. Se $G = (V, E)$ é codificado como uma matriz adjacente $|V| \times |V|$ com entradas a_{ij} = número de arestas conectando v_i and v_j, o tempo de processamento será de $\mathcal{O}(|V|^2)$. Graças a Euler, CIRCUITO EULERIANO é um problema tratável. Os moradores de Königsberg, por outro lado, aprenderam de Euler que eles nunca encontrariam um caminho para suas casas atravessando cada uma das sete pontes exatamente uma vez.

Ciclos hamiltonianos

Um outro problema de decisão está associado com o matemático e astrônomo Sir William Rowan Hamilton. No ano de 1859, Hamilton propos um novo enigma chamado de Jogo do Icosaedro (Fig. A.4).

A generalização do Jogo do Icosaedro evoca mais algumas definições da teoria dos grafos: um caminho fechado num grafo é chamado de *ciclo* se todos os seus vértices (exceto o primeiro e o último) forem distintos. Um ciclo Hamiltoniano é um ciclo que contém todos os vértices do grafo. A generalização do jogo do icosaedro fica

> CICLO HAMILTONIANO: Dado um grafo $G = (V, E)$. Será que G contém um ciclo hamiltoniano ?

Existe uma certa semelhança entre CIRCUITO EULERIANO e CICLO HAMILTONIANO. No primeiro podemos passar pelas arestas uma única vez; no último, visitamos cada vértice uma única vez. Ape-

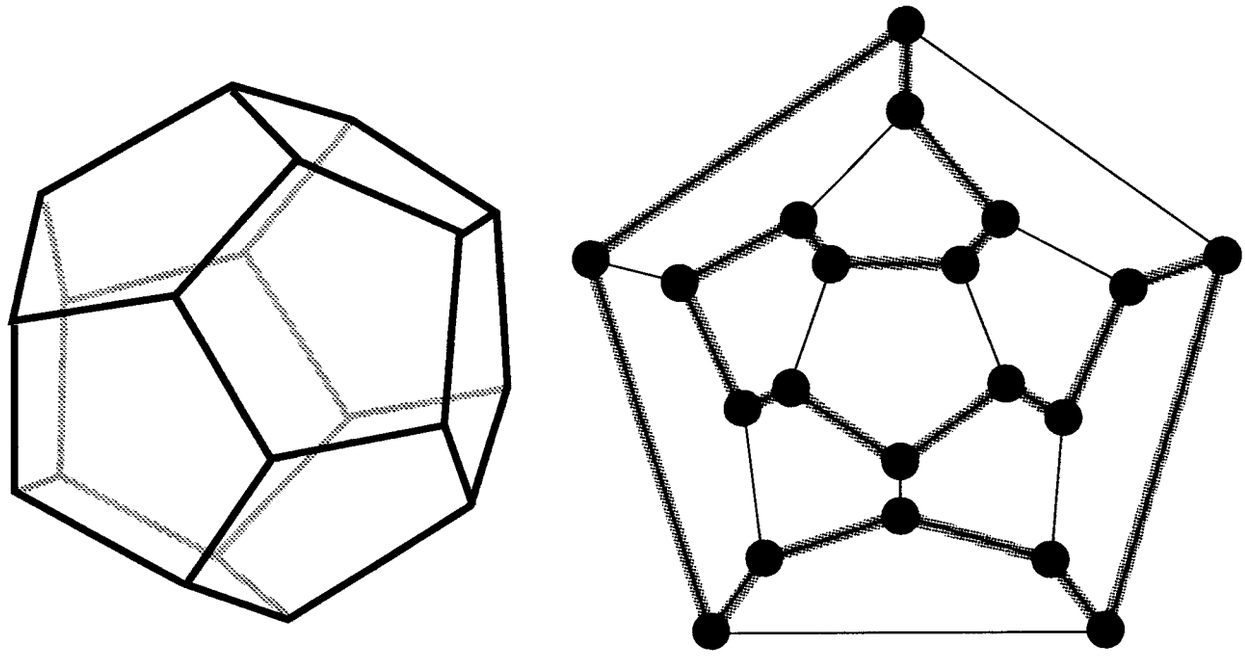


Figura A.4: O Jogo do Icosaedro de Hamilton: Encontre uma rota ao longo das arestas de um dodecaedro (esquerda), passando em cada vértice exatamente uma vez e retornando ao ponto inicial. Uma solução está indicada no grafo planar (arestas sombreadas) que é isomorfo ao dodecaedro (direita).

sar da semelhança, os dois problemas apresentam graus de dificuldades completamente diferentes. A disponibilidade de uma intuição matemática no CICLO HAMILTONIANO não fornece nem um algoritmo polinomial nem uma prova de que ele seja impossível. CICLO HAMILTONIANO é intratável e ninguém sabe o porquê.

Coloração

Imagine que desejemos arranjar os palestrantes em um congresso de forma que nenhum participante seja forçado a perder uma palestra que ele gostaria de ouvir. Supondo uma boa oferta de salas de seminários, assegurando tantas palestras paralelas quanto desejarmos, podemos terminar o evento dentro de k seções não paralelas? Este problema pode ser formulado em termos de grafos: Seja G o grafo cujos vértices representem os palestrantes e que dois palestrantes sejam adjacentes se e somente se houver um participante desejando assistir a ambos. Sua tarefa é associar uma das k seções a cada vértice de modo que vértices adjacentes tenham diferentes blocos ou seções. Uma formulação comum deste problema usa cores ao invés de blocos.

k-COLORAÇÃO: Dado um grafo $G = (V, E)$. Existe uma coloração de vértices de G usando no máximo k diferentes cores tal que nenhum vértice adjacente tenha a mesma cor ?

Quando $k = 2$ este problema é tratável e a construção de um algoritmo polinomial é um exercício relativamente fácil. Para $k = 3$ a situação muda consideravelmente: 3-COLORAÇÃO é intratável. Note que para k maiores o problema fica mais fácil novamente: todo *grafo planar* sem laço (são arestas cujos extremos coincidem) é sempre colorível com 4 cores! Este é o famoso teorema das 4 cores, que ilumina o caminho para provar se um grafo é k-colorável, basta verificar se ele tem laço ou não. O 3-COLORAÇÃO permanece intratável mesmo quando restringimos a grafos planares.

Satisfabilidade

Fechamos esta seção com um problema de decisão que não é da teoria de grafos mas da lógica Booleana. Uma variável booleana x pode assumir um dentre dois valores 0 (falso) ou 1 (verdadeiro). Variáveis booleanas podem ser combinadas em *cláusulas* usando os operadores booleanos

- NOT $\bar{\cdot}$ (negação): A cláusula $\overline{x}$ é verdadeira ($\overline{x} = 1$) se e somente se x é falso ($x = 0$).
- AND $\wedge$ (conjunção): A cláusula $x_1 \wedge x_2$ é verdadeira ($x_1 \wedge x_2 = 1$) se e somente se ambas as variáveis são verdadeiras: $x_1 = 1$ e $x_2 = 1$
- OR $\vee$ (disjunção): A cláusula $x_1 \vee x_2$ é verdadeira ($x_1 \vee x_2 = 1$) se e somente se pelo menos uma das variáveis for verdadeira: $x_1 = 1$ ou $x_2 = 1$.

Uma variável x ou a negação de $\overline{x}$ é chamada uma *literal*. Cláusulas diferentes podem ser combinadas para fornecerem fórmulas booleanas complexas, isto é,

$$F_1(x_1, x_2, x_3) = (x_1 \vee \overline{x_2} \vee x_3) \wedge (x_2 \vee \overline{x_3}) \wedge (\overline{x_1} \vee x_2) \wedge (\overline{x_1} \vee \overline{x_3}). \tag{A.4}$$

Uma fórmula booleana retorna 1 ou 0, dependendo das associações das variáveis booleanas. No exemplo acima, $F_1 = 1$ para $x_1 = 1, x_2 = 1, x_3 = 0$ e $F_1 = 0$ para $x_1 = x_2 = x_3 = 1$. Uma fórmula

F é chamada *satisfatível*, se existir pelo menos uma atribuição das variáveis tais que a fórmula seja verdadeira. Daí F_1 é satisfatível, enquanto que

$$F_2(x_1, x_2) = (\overline{x_1} \vee x_2) \wedge \overline{x_2} \wedge x_1 \tag{A.5}$$

não é satisfatível. Toda fórmula booleana pode ser escrita na *Forma Normal Conjuntiva* (FNC), isto é, como um conjunto de cláusulas C_k combinadas exclusivamente pelo operador AND:

$$F = C_1 \wedge C_2 \wedge \cdots \wedge C_m \tag{A.6}$$

e as literais em cada cláusula são combinadas com o operador OR. Os exemplos F_1 e F_2 estão ambas escritos na FNC. Cada cláusula pode ser considerada como um vínculo nas variáveis e, portanto satisfazer uma fórmula significa satisfazer um conjunto de vínculos simultaneamente (possivelmente conflitantes). Conseqüentemente

> SATISFABILIDADE(SAT): Dadas cláusulas disjuntas $C_1, C_2, \ldots, C_m$ de literais, onde uma literal é um conjunto de variáveis ou variáveis negadas de um conjunto $\{x_1, x_2, \ldots, x_n\}$. Existe alguma atribuição de variáveis que satisfaça todas as cláusulas simultaneamente?

Pode ser considerado como um protótipo de um problema de satisfabilidade com vínculo [89]. Fixando o número de literais em cada cláusula ficamos com a seguinte formulação

> k–SAT: Dadas cláusulas disjuntas $C_1, C_2, \ldots, C_m$ de k literals cada, onde um literal é uma variável ou uma variável negada de um conjunto $\{x_1, x_2, \ldots, x_n\}$. Existe alguma atribuição que satisfaça todas as cláusulas simultaneamente?

Algoritmos polinomiais são conhecidos para 1-SAT e 2-SAT [90], mas nenhum algoritmo polinomial é conhecido para o problema geral SAT e k-SAT se $k > 2$.

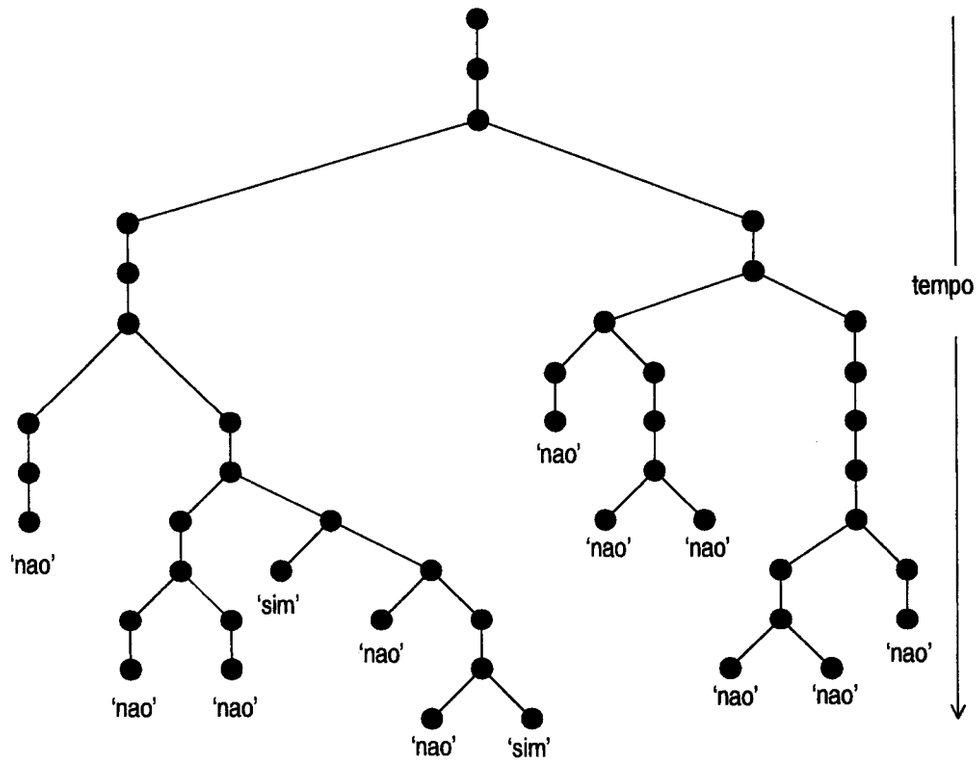


Figura A.5: Exemplo da execução da história de um algoritmo não determinístico.

A.1.2 As classes $\mathcal{P}$ e $\mathcal{NP}$

Problemas Tratáveis

Já vimos bastante exemplos e podemos agora introduzir uma importante classe de complexidade para problemas de decisão. A classe de problemas de decisão tratáveis é fácil de definir: ela consiste daqueles problemas para o qual é conhecido um algoritmo polinomial. A correspondente classe é chamada de 'polinomial' $\mathcal{P}$:

> **Definição:** Um problema de decisão P é um elemento da classe $\mathcal{P}$ se e somente se ele pode ser resolvido por um algoritmo em tempo polinomial.

CIRCUITO EULERIANO, 2-COLORAÇÃO, AGM(D) etc., estão in $\mathcal{P}$.

Algoritmo não determinístico

A definição da segunda classe de complexidade envolve o conceito de *algoritmo não determinístico*. Um algoritmo não determinístico é como um algoritmo comum, exceto que ele pode usar uma

instrução adicional muito poderosa [91]:

goto both rótulo 1, rótulo 2

Esta instrução divide a computação em dois processos paralelos, que continua de cada uma das instruções indicadas pelos 'rótulo 1' e 'rótulo 2'. Ao encontrar mais e mais instruções como essas a computação vai se ramificando, como uma árvore em um número de computações paralelas que potencialmente pode crescer como uma função exponencial do tempo transcorrido (veja Fig. A.5). Um algoritmo não determinístico pode desenvolver um número exponencial de computações em tempo polinomial! No mundo dos computadores convencionais, algoritmo não deterministico é um *conceito teórico* somente, mas em computação quântica isto pode mudar. Precisamos desse conceito para definir a classe $\mathcal{NP}$ de problemas 'Não determinístico Polinomial':

Definição: Um problema de decisão P está na classe $\mathcal{NP}$, se e somente se ele pode ser resolvido em tempo *polinomial* por um *algoritmo não determinístico*.

Solubilidade por um algoritmo não determinístico significa o seguinte: Todos os ramos da computação irão parar, retornando 'sim' ou 'não'. Dizemos que o algoritmo global retorna 'sim' se e somente se *qualquer* um de seus ramos relate 'sim'. A resposta é 'não', se nenhum dos ramos relata 'sim'. Dizemos ainda que um algoritmo não determinístico resolve um problema de decisão em tempo polinomial, se o número de passos usados para encontrar o ramo que relata 'sim' é limitado por um polinômio no tamanho do problema. Requeremos solubilidade polinomial somente para 'sim'. Note que os algoritmos convencionais (determinísticos) são casos especiais de algoritmos não determinísticos que não usam a instrução **goto both**. Segue imediatamente que $\mathcal{P} \subseteq \mathcal{NP}$. Todos os problemas de decisão discutidos na seção precedente são membros de $\mathcal{NP}$. Aqui está um algoritmo não determinístico para o SAT:

```
SATISFABILIDADE(F)
Input: Fórmula Booleana F(x_1, ..., x_n)
Output: 'sim' se F é satisfatível, caso contrário, 'não'
begin
  for i = 1 to n
    goto both rótulo 1, rótulo 2
    rótulo 1: x_i = verdadeiro; continue
    rótulo 2: x_i = falso; continue
  end
  if F(x_1, ..., x_n) = verdadeiro then return 'sim'
                                   else return 'não'
end
```

O laço **for** ramifica para cada iteração: em um ramo x_i = verdadeiro e no outro ramo x_i = falso (a instrução **continue**) começa a próxima iteração para o laço). Depois de executar o laço **for** temos 2^n ramos de computação, um para cada uma das possíveis atribuições das n variáveis Boolenas. O poder dos algoritmos não determinísticos é que eles permitem a enumeração exaustiva de um número exponencialmente grande de candidatos a solução em tempo polinomial. Se a avaliação de cada candidato a solução (calculando $F(x_1, \ldots, x_n)$ no exemplo acima) por sua vez pode ser feito em tempo polinomial, o algoritmo não determinístico total é polinomial. Para um problema da classe $\mathcal{NP}$, a única fonte de intratabilidade é o tamanho exponencial do espaço de busca.

Um primeiro mapa para a complexidade

A Fig. A.6 resume o que temos feito até aqui. A classe $\mathcal{NP}$ consiste de todos os problemas de decisão cuja única fonte de dificuldade é o tamanho do espaço de busca que cresce exponecialmente com o tamanho do problema. Esses problemas são intratáveis a menos que algum 'insight' matemático nos forneça um atalho polinomial para evitar a busca exaustiva. Tal 'insight' promove um problema para a classe $\mathcal{P}$ de problemas solúveis polinomialmente.

A classe $\mathcal{NP}$ não somente contém um grande número de problemas com importantes aplicações, mas também apresenta um desafio real: Todos os problemas em $\mathcal{NP}$ ainda tem uma chance de estar em $\mathcal{P}$.

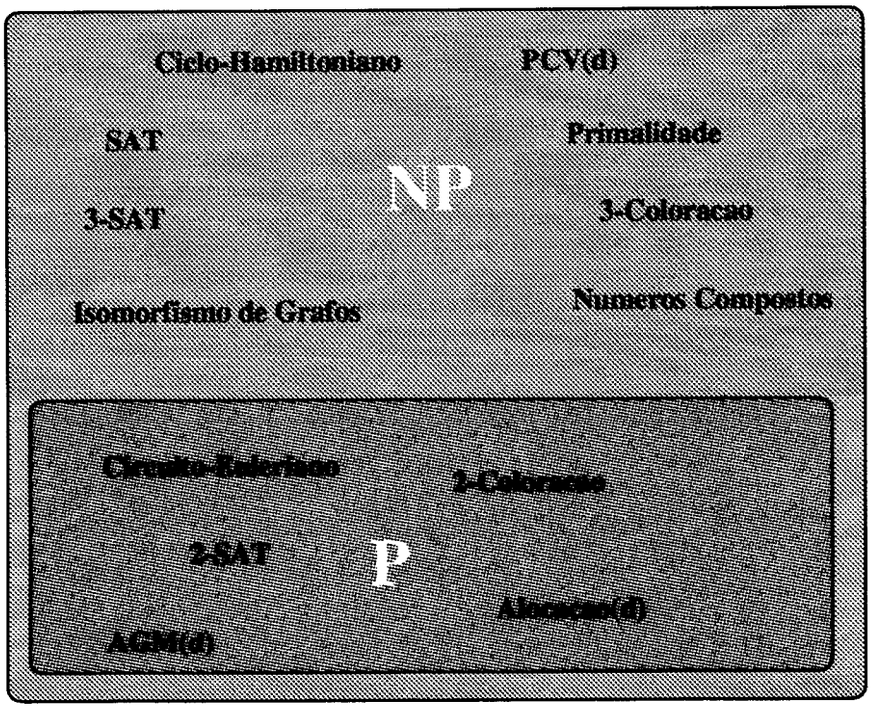


Figura A.6: Um primeiro mapa para a complexidade. Todos problems indicados estão definidos dentro do texto. Problemas com um (D) são as versões decisão dos problemas de otimização.

A prova da não existência de um algoritmo polinomial para um simples problema $\mathcal{NP}$ estabeleceria que $\mathcal{P} \neq \mathcal{NP}$. Já que esta prova não foi encontrada,

$$\mathcal{P} \stackrel{?}{=} \mathcal{NP} \tag{A.7}$$

representa a mais famosa conjectura aberta em ciência da computação teórica.

A.1.3 $\mathcal{NP}$-completeza

Reduções polinomiais

As complexidades computacionais de dois problemas P_1 e P_2 podem ser relacionadas pela construção de um algoritmo para P_1 que use um algoritmo para P_2 como uma 'sub-rotina'. Considere o seguinte algoritmo que relaciona o CICLO HAMILTONIO ao PCV(D):

CICLO HAMILTONIANO(G)
Input: Grafo $G = (V, E)$
Output: 'sim' se G contém um ciclo Hamiltoniano, caso contrário 'não'

```
(1)  begin
(2)     n := |V|
(3)     for i = 1 to n
(4)        for j = 1 to n
(5)           if (v_i, v_j) ∈ E then d_ij := 1
(6)                             else d_ij := 2
(7)     if PCV-decisão (d, B := n) = 'sim' then return 'sim'
(8)                                        else return 'não'
(9)  end
```

Este algoritmo resolve o CICLO HAMILTONIANO resolvendo uma realização apropriada do PCV(D). No laço **for** (linhas 3-5) a matriz distância d é formada com entradas $d_{ij} = 1$ se existir uma aresta (v_i, v_j) em G e caso contrário $d_{ij} = 2$. Agora um ciclo hamiltoniano em G é um percurso válido no correspondente PCV com todas as distâncias entre cidades tendo comprimento 1, isto é, com comprimento total n. Reciprocamente, suponha que o PCV tenha um itinerário de comprimento n. Desde que as distâncias entre cidades são 1 ou 2 e a soma no percurso é n, o comprimento total n implica que cada par de cidades visitadas sucessivamente deve ter distância 1, isto é as aresta do trajeto estão em G e correspondem a um ciclo hamiltoniano. Conseqüentemente a sub-rotina que resolve PCV(D) (linha 7) fornece a solução para o CICLO HAMILTONIANO. A pergunta que poderiamos fazer seria: de que modo este algoritmo relaciona a complexidade computacional do CICLO HAMILTONIANO com o PCV(D)? Note que este é um algoritmo polinomial se as chamadas para a sub-rotina PCV(D) forem considerados operações elementares. Se alguém apresentar um algoritmo polinomial para o PCV(D), imediatamente teremos um algoritmo polinomial para o CICLO HAMILTONIANO! Dizemos que CICLO HAMILTONIANO é *redutível polinomialmente* ao PCV(D) e escrevemos

$$\text{CICLO HAMILTONIANO} \leq \text{PCV(D)}. \tag{A.8}$$

Em muitos livros, redutibilidade polinomial é denotado por '$\propto$' ao invés de '$\leq$'. Seguimos [93] e

usamos '$\leq$' porque esta notação enfatiza uma importante conseqüência da redutibilidade polinomial: a existência de uma redução de P_1 para P_2 exclui a possibilidade de que P_2 possa ser resolvido em tempo polinomial, mas não a de P_1. Conseqüentemente $P_1 \leq P_2$ pode ser lido como P_1 *não pode ser mais difícil do que* P_2. Aqui está a definição (informal):

Definição: Dizemos que um problema P_1 é *redutível polinomialmente* para um problema P_2 e escrevemos $P_1 \leq P_2$ se existe um algoritmo polinomial para P_1 dado que existe um algoritmo polinomial para P_2.

Problemas $\mathcal{NP}$-completos

Outras reduções polinomiais similares a Eq.(A.8) são:

$$\begin{aligned} \text{SAT} &\geq \text{3-SAT} \\ \text{3-SAT} &\geq \text{3-COLORAÇÃO} \\ \text{3-COLORAÇÃO} &\geq \text{CICLO HAMILTONIANO} \end{aligned} \tag{A.9}$$

Veja [1,24] para os algoritmos de redução correspondentes. A redutibilidade polinomial é transitiva: $P_1 \leq P_2$ e $P_2 \leq P_3$ implica $P_1 \leq P_3$. Da transitividade e Eqs. (A.8) e (A.9) segue que cada um dos SAT, 3-SAT, 3-COLORAÇÃO e CICLO HAMILTONIANO reduzem-se ao PCV(D), isto é, um algoritmo polinomial PCV(D) implica em um algoritmo polinomial para todos esses problemas! Isto é incrível. A redutibilidade polinomial foi revelada por Stephen Cook em 1971 [23] através do seguinte e notável teorema:

Teorema: (Cook, 1971) Todo os problemas em $\mathcal{NP}$ são redutíveis polinomiamente ao SAT,

$$\forall P \in \mathcal{NP} : P \leq \text{SAT} \tag{A.10}$$

Este teorema significa que

1. nenhum problema em $\mathcal{NP}$ é mais difícil do que SAT ou que o SAT está entre os mais difíceis em $\mathcal{NP}$.

2. um algoritmo polinomial para SAT implicaria um algoritmo polinomial para *todo* problema em $\mathcal{NP}$, i.e. implicaria $\mathcal{P} = \mathcal{NP}$.

Parece que o SAT é muito especial, mas de acordo com a transitividade e Eqs. (A.9) e (A.8) ele pode ser trocado pelo 3-SAT, 3-COLORAÇÃO, CICLO HAMILTONIANO ou PCV(D). Estes problemas formam uma nova classe de complexidade:

Definição: Um problema P é chamado $\mathcal{NP}$-completo se $P \in \mathcal{NP}$ e $Q \leq P$ para todo $Q \in \mathcal{NP}$

A classe dos problemas $\mathcal{NP}$-completo coleta os problemas mais difíceis em $\mathcal{NP}$. Se qualquer um deles tiver um algoritmo eficiente, então *todos* os problemas em $\mathcal{NP}$ poderão ser resolvidos eficientemente, isto é, $\mathcal{P} = \mathcal{NP}$. Isto é extremamente improvável, ainda mais se levarmos em conta os esforços infrutíferos de muitas pessoas brilhantes para encontrar um algoritmo polinomial para problemas como CICLO HAMILTONIANO ou PCV(D).

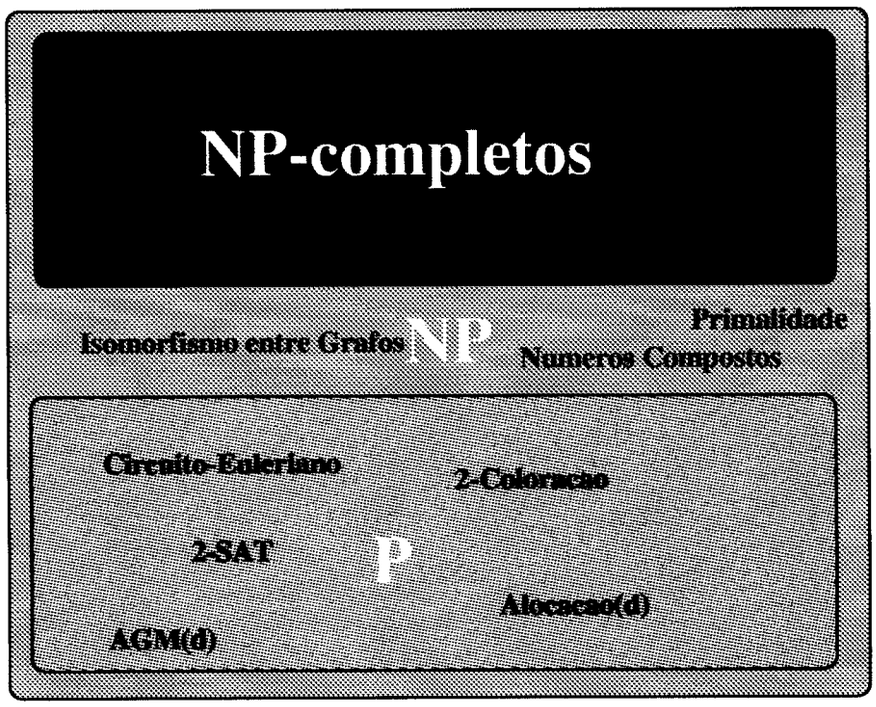


Figura A.7: O mapa da complexidade revisado.

O mapa de $\mathcal{NP}$

Desde o teorema de Cook muitos problemas têm sido mostrados pertencer a classe $\mathcal{NP}$-completo. Uma lista compreensiva e atualizada de centenas de problemas $\mathcal{NP}$-completo pode ser encontrada na rede [94]. Nosso mapa de $\mathcal{NP}$ precisa ser remodelado (Fig. A.7). Todos os problemas intratáveis que discutimos até aqui são $\mathcal{NP}$-completos, exceto NÚMEROS COMPOSTOS e PRIMALIDADE definidos respectivamente como

NÚMEROS COMPOSTOS: Dado um número inteiro positivo N. Existem números inteiros $p > 1$ e $q > 1$ tais que $N = pq$?

PRIMALIDADE: Dado um inteiro positivo N. N é primo?

PRIMALIDADE é a negação ou o complemento do NÚMEROS COMPOSTOS. O primeiro problema consiste em fatorar um número inteiro como o produto de dois outros números inteiros. O segundo é a negação ou o complemento do primeiro. Eles são de crucial importância em criptografia; em particular, a vantagem do algoritmo **RSA** [41] está em explorar a propriedade deles serem $\mathcal{NP}$. Para ambos os problemas nenhum algoritmo polinomial é conhecido bem como não há nenhuma redução que os classifiquem como $\mathcal{NP}$-completos. Um outro problema $\mathcal{NP}$ que resiste a classificação em $\mathcal{P}$ e em $\mathcal{NP}$ é o seguinte:

ISOMORFISMO DE GRAFOS: Dado dois grafos $G = (V, E)$ e $G'(V, E')$ no mesmo conjunto de nós. G e G' são isomórficos, isto é existe uma permutação π de V tal que $G' = \pi(G)$, onde por $\pi(G)$ denotamos o grafo $(V, \{[\pi(u), \pi(v)] : [u, v] \in E\})$?

Existem mais problemas em $\mathcal{NP}$ que resistem a classificação em $\mathcal{P}$ ou $\mathcal{NP}$, mas nenhum deles foi provado não pertencer a $\mathcal{P}$ ou $\mathcal{NP}$. No momento há três alternativas de mapas para $\mathcal{NP}$ bem ilustradas pela Fig. A.8.

redução polinomial que estabelece

$$P(D) \leq P(L) \leq P(O). \tag{A.11}$$

Se a variante de decisão de um problema de otimização é $\mathcal{NP}$-completo, não existe nenhum algoritmo eficiente para o problema de otimização a menos que $\mathcal{P} = \mathcal{NP}$. Um problema de otimização como

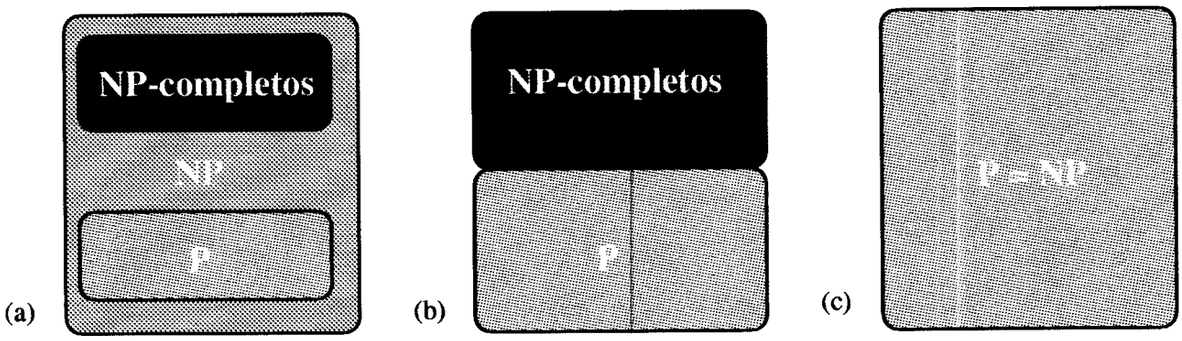


Figura A.8: Três tentativas de mapas $\mathcal{NP}$. É muito provável (mas não com certeza) que (a) seja o mapa correto.

A.1.4 Além de $\mathcal{NP}$

Problemas de otimização

Como a classificação dos problemas de decisão relaciona-se com a dos problemas de otimização? Uma realização geral de um problema de otimização é um par (F, c), onde F é o conjunto das soluções factíveis e c é a função custo $c : F \to \mathbb{R}$. Consideraremos somente *otimização combinatorial* onde o conjunto F é contável. Um problema algorítmico P pode ser classificado nas seguintes categorias:

1. O problema de otimização $P(O)$: Encontre a *solução factível* $f^* \in F$ que minimiza a função custo.

2. O problema da localização $P(L)$: Encontre o *custo* $c^* = c(f^*)$ de solução mínima.

3. O problema de decisão $P(D)$: Dado um limitante $B \in \mathbb{R}$, existe uma solução factível $f \in F$ tal que $c(f) \leq B$?

Sob a hipótese de que a função custo c pode ser avaliada em tempo polinomial, é direto escrever a redução polinomial que estabelece

$$P(D) \leq P(L) \leq P(O). \tag{A.11}$$

Se a variante de decisão de um problema de otimização é $\mathcal{NP}$-completo, não existe nenhum algoritmo eficiente para o problema de otimização a menos que $\mathcal{P} = \mathcal{NP}$. Um problema de otimização como

PCV, cuja variante de decisão é $\mathcal{NP}$-completo é denotado $\mathcal{NP}$-árduo. Na verdade, os problemas de otimização são $\mathcal{NP}$-árduos, pois sendo compostos de vários problemas de decisão, eles seriam tão difíceis quantos os problemas de decisão. Por exemplo, se o problema de decisão for $\mathcal{O}(f(n))$, então a variante de otimização seria da ordem de $\mathcal{O}(m \cdot f(n))$. De forma geral,

$$\mathcal{NP}\text{-árduo} \geq \mathcal{NP}\text{-completo} \geq \mathcal{NP}.$$

Um algoritmo para um problema de decisão implica em um algoritmo polinomial para a variante problema de otimização ou de localização? Sim, para responder a esta questão basta provar o inverso da Eq. (A.11),

$$P(O) \leq P(L) \leq P(D). \tag{A.12}$$

Não apresentaremos a prova, mas enfatizamos que o estudo dos problemas de decisão, tidos como versões mais fáceis, é suficiente para entendermos os problemas intratáveis, e que todos os tipos de formulação dos problemas algorítmicos estão extremamente articulados entre si.

Problemas de contagem

Até aqui estudadamos dois estilos relacionados de problemas: Problemas de decisão que perguntam se uma solução desejada existe e o problema de otimização exige que a solução seja produzida. Um terceiro importante e fundamentalmente diferente tipo de problema pergunta quantas soluções existem. A variante *contagem* do SAT é enunciada a seguir

> #SAT: Dada uma expressão Booleana, compute o número de diferentes atribuições verdadeiras que o torna satisfatível

Similarmente, #CICLO HAMILTONIANO pergunta pelo número de diferentes ciclos hamiltonianos em um dado grafo, #PCV pelo número de diferentes rotas com comprimento $\leq B$ e assim por diante.

> **Definição:** Um problema de contagem $\#P$ é um par (F, d), onde F é o conjunto de todas as soluções factíveis e d é uma função de decisão $d : F \mapsto \{0, 1\}$. A saída do $\#P$ é o número de $f \in F$ with $d(f) = 1$. A classe $\#\mathcal{P}$ (pronunciada "número P") consiste de

todos os problemas de contagem com uma função decisão d que pode ser avaliada em tempo polinomial.

Como na classe $\mathcal{NP}$, $\#\mathcal{P}$ coleta todos os problemas cuja única fonte de intratabilidade é o número de soluções factíveis. Um algoritmo polinomial para um problema de contagem $\#P$ implica em um algoritmo polinomial para o problema de decisão associado P: $P \leq \#P$. Conseqüentemente é muito improvável que o #SAT possa ser resolvido eficientemente. De fato podemos definir redutibilidade polinomial para problemas de contagem e provar que todo problema em $\#\mathcal{P}$ reduz-se polinomialmente a #SAT [43]:

Teorema: #SAT é $\#\mathcal{P}$-completo.

Como você deve ter imaginado , #CICLO HAMILTONIANO e #PCV também são $\#\mathcal{P}$-completos. Apesar da similaridade entre $\mathcal{NP}$ e $\#\mathcal{P}$, os problemas de contagem são inerentemente mais difíceis do que os problemas de decisão. Isto está documentado por aqueles problemas $\#\mathcal{P}$-completos, para os quais os correspondentes problemas de decisão podem ser resolvidos em tempo polinomial, o exemplo clássico é o problema de calcular o *permanente* de uma matriz [96].

Apêndice B

O cálculo exato

A solução exata da termodinâmica do problema da partição numérica é devida a Mertens [37]. Vamos considerar apenas o caso mais simples, ou seja, do **PPN** sem vínculo de cardinalidade. Sabemos que a função de partição é dada por

$$Z = \sum_{s_j} e^{\frac{-1}{T}|\sum_j a_j s_j|}. \tag{B.1}$$

Podemos omitir algumas passagens, pois o cálculo é idêntico àquele desenvolvido na aproximação 'annealed' até a expressão (2.21), daí função Z fica

$$Z = 2^N \int_{-\infty}^{\infty} \frac{d\hat{x}}{2\pi} \prod_{j=1}^{N} \cos\left(\frac{a_j \hat{x}}{T}\right) \frac{2}{1+\hat{x}^2}, \tag{B.2}$$

fazendo a seguinte mudança de variável: $y = \arctan \hat{x}$, a equação (B.2) é reescrita de uma forma muito elegante

$$Z = 2^N \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{dy}{\pi} e^{NG(y)} \tag{B.3}$$

onde

$$G(y) = \frac{1}{N} \sum_{j=1}^{N} \ln \cos(\frac{a_j}{T} \tan y). \tag{B.4}$$

Para grandes valores de N, a independência estatística dos a_j nos permite aplicar a *lei dos grandes números* (ou a propriedade de automediância), isto é, trocar a soma pela média sobre a:

$$G(y) \approx \left\langle \ln \cos(\frac{a}{T} \tan(y)) \right\rangle. \tag{B.5}$$

Esta troca é a principal razão pela qual vidros de spin com acoplamento que fatorizam-se, $J_{ij} = -a_i a_j$, são comparativamente fáceis de se resolver. Isto nos alivia de calcular a média sobre $\ln Z$, que pode ser muito difícil em outros modelos de vidros de spin. A integral na equação (B.3) pode ser efetuada assintoticamente para grandes valores de N usando o *método de ponto de sela.* Para encontrar o máximo de $G(y)$, suporemos que **a** pode somente tomar valores que são múltiplos inteiros de um número fixo Δa. Para uma distribuição inteira $\Delta a = 1$, e para uma distribuição contínua (ponto flutuante) Δa é o menor número que pode ser representado com o número de bits disponíveis. Com esta hipótese, as soluções de

$$G'(y) = \left\langle \frac{a}{T} \tan(\frac{a}{T} \tan y) \right\rangle \cdot (1 + \tan^2 y) = 0 \tag{B.6}$$

são dadas por

$$y_k = \arctan(\frac{\pi T}{\Delta a} k) \qquad k = 0, \pm 1, \pm 2, ... \tag{B.7}$$

Note que $\tan\left(\frac{a}{T} \tan y_k\right) = 0$ para todos os valores $a = n\Delta a$. É claro que temos que considerar todos pontos de sela quando efetuamos a integral (B.3):

$$Z \approx 2^N \sum_k \int_{-\infty}^{\infty} \frac{dy}{\pi} e^{-\frac{N}{2} G''(y_y) y^2} = 2^N \frac{\sqrt{2}}{\sqrt{\pi N}} \sum_k \frac{1}{\sqrt{G''(y_k)}}, \tag{B.8}$$

com

$$G''(y_k) = \frac{\langle a^2 \rangle}{T^2} \left[1 + \left(\frac{\pi T}{\Delta a} \right)^2 k^2 \right]^2. \tag{B.9}$$

usando a seguinte identidade

$$\sum_{k=0\pm1,...} \frac{1}{1 + (xk)^2} = \frac{\pi}{x} \cdot \coth \frac{\pi}{x} \tag{B.10}$$

finalmente obtemos o valor exato da função de partição

$$Z = 2^N \cdot \frac{\Delta a}{\sqrt{\frac{\pi}{2} N \langle a^2 \rangle}} \cdot \coth \frac{\Delta a}{T}. \tag{B.11}$$

No cáculo aproximado só consideramos um ponto de sela, e a grande sutileza desse cálculo foi escrever os a_j como múltiplos de Δa. A função de partição Z imediatamente fornece a energia livre

$$F(T) = -TN\ln 2 + \frac{T}{2}\ln\frac{\pi N\langle a^2\rangle}{2\Delta a^2} - T\ln\coth\frac{\Delta a}{T} \tag{B.12}$$

e a energia média

$$\langle E\rangle = \frac{\Delta a}{\sinh\frac{\Delta a}{T}\cosh\frac{\Delta a}{T}}. \tag{B.13}$$

Fixando $\Delta a > 0$, então o $\lim_{T\to 0}\langle E\rangle_T = 0$, isto é, o estado fundamental são partições perfeitas. Quantas partições perfeitas podemos esperar? A resposta é dada pela entropia S, que pode ser calculada a partir da energia livre

$$S = \lim_{T\to 0}\frac{\partial F(T)}{\partial T}. \tag{B.14}$$

Efetuando esse cálculo encontramos

$$S = N(\kappa_c - \kappa)\ln 2 + \tilde{S}(\frac{\Delta a}{2T}), \tag{B.15}$$

com

$$\kappa_c = 1 - \frac{\ln(\frac{\pi}{6}N}{N2\ln 2} \tag{B.16}$$

$$\kappa = \frac{\ln(\frac{3}{\Delta a^2}\langle a^2\rangle}{N2\ln 2}, \tag{B.17}$$

e a contribuição térmica da entropia é

$$\tilde{S}(\frac{\Delta a}{T}) = \ln\coth\frac{\Delta a}{T} + \frac{\Delta a}{T}\frac{\coth^2\frac{\Delta a}{T} - 1}{\coth\frac{\Delta a}{T}}. \tag{B.18}$$

Para Δa finito, $\widetilde{S}$ anula quando a temperatura vai a zero e aumenta monotonicamente com T. Neste caso, a entropia a temperatura zero é dado por $N(\kappa_c - \kappa)\ln 2$. Se $\kappa < \kappa_c$, temos uma entropia extensiva e conseqüentemente um número exponencial de partições perfeitas. O produto $N.\kappa$ é uma medida para o número de bits necessários para codificar os a_j . Suponha que a_j seja um número inteiro com b-bits independentes, então $\Delta a = 1$ e

IFSC-USP SERVIÇO DE BIBLIOTECA INFORMAÇÃO

$$\kappa = \frac{b}{N} + \frac{1}{2N}\log_2(1 - \frac{3}{2}2^{-b} + \frac{1}{2}2^{-2b}) = \frac{b}{N} + \frac{1}{N}O(2^{-b}). \tag{B.19}$$

Neste caso a condição $\kappa < \kappa_c$, pode ser traduzida em

$$b < N - \frac{1}{2}\log_2(\frac{\pi}{6}N). \tag{B.20}$$

Aqui simplesmente usamos (B.16) que é um limintante superior para κ. Esta desigualdade deve ser satisfeita para que exista partições perfeitas.

O que acontece se $N < b$ ou, equivalentemente, $\kappa > \kappa_c$? De acordo com as considerações acima, esperamos uma energia ótima que seja exponencialmente pequena, $O(N)$, porém maior do que zero. O problema é que a entropia a temperatura zero fica negativa. Isto está errado, porque a entropia a temperatura zero é por definição o logaritmo do número de estados no estado fundamental, o que é neste caso pelo menos igual a $\ln 2$. Devemos ter mais cuidado com o limite $T \to 0$, para obter a entropia nesta temperatura. Como $S \geq \ln 2$, em termos de Δa a condição $\kappa > \kappa_c$ significa

$$2^{-N} > \Delta a\sqrt{\frac{2}{\pi N\langle a^2\rangle}} \tag{B.21}$$

isto é, essencialmente $\Delta a = O(2^{-N})$. Neste regime as contribuições de $\widetilde{S}$ são de ordem $O(N)$ para qualquer temperatura finita T,

$$\widetilde{S}(\frac{\Delta a}{T}) = \ln\left(\frac{T}{\Delta a}\right) + 1 + O(\frac{\Delta a^2}{T^2}), \tag{B.22}$$

e conseqüentemente não podem ser desprezadas. Tecnicamente vamos introduzir uma temperatura efetiva T_0 abaixo da qual o sistema não pode ser "congelado ". T_0 garante que as contribuições de $\widetilde{S}$ permaneçam $O(N)$. Este valor pode ser calculado a partir do limitante inferior de S :

$$\ln 2 = N(\kappa_c - \kappa)\ln 2 + \tilde{S}(\frac{\Delta a}{T}) \approx N(\kappa_c - \kappa)\ln 2 + \ln\left(\frac{T_0}{\Delta a}\right). \tag{B.23}$$

Daqui tiramos que

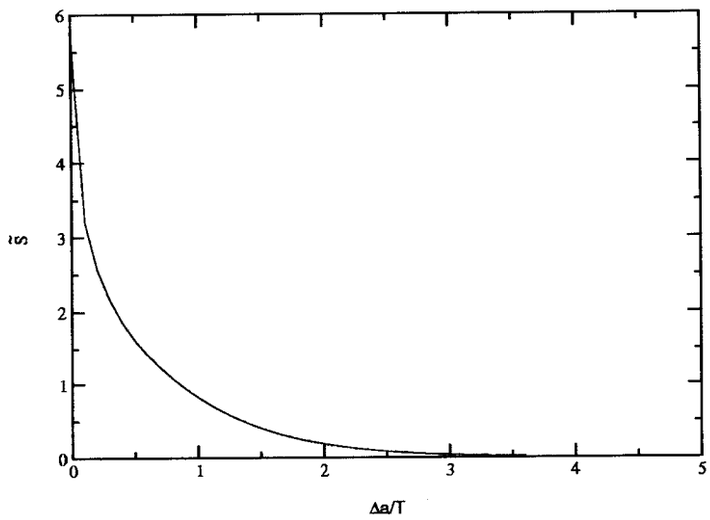


Figura B.1: Contribuição térmica da entropia $\tilde{S}$ contra $\frac{\Delta a}{T}$

$$T_0 = 2\Delta a 2^{N(\kappa-\kappa_c)} = \sqrt{2\pi N \langle a^2 \rangle} 2^{-N}. \tag{B.24}$$

Neste regime a energia do estado fundamental média $< E >$ não mais é zero e sim

$$\langle E \rangle = T_0 = \sqrt{2\pi N \langle a^2 \rangle} 2^{-N} \tag{B.25}$$

Esta equação está de acordo com as estimativas previamente conhecidas [40] para o valor médio da função custo (energia) E, que era $O(\sqrt{N}2^{-N})$, para conjuntos de números reais. A novidade aqui é o prefator $\sqrt{2\pi \langle a^2 \rangle}$.

Para verificar (B.25) vamos considerar a versão contínua do problema da partição numérica, onde a_i são números reais uniformemente distribuídos no intervalo [0,1). Neste formalismo isto significa que $\Delta a \to 0$ e $\sum_j a_j^2 = N/3$. Estamos no regime $\kappa > \kappa_c$ e a equação (B.25) se torna

$$\langle E \rangle = \sqrt{\frac{2}{3}\pi N} 2^{-N} = 1.447\sqrt{N} 2^{-N} \tag{B.26}$$

O acordo da equação (B.26) é quando comparada a dados numéricos é convincente, comparados com nosso resultado exibidos na figura (2.1), onde a linha contínua foi obtida através de uma regressão linear sobre os dados da simulação.

Note que $\kappa_c(N \to \infty) = 1$. A estimativa assintóé devido a influência de $O(\frac{1}{N}\ln N)$ na equação (B.16) que não pode ser desprezado para sistemas pequenos acessíveis para simulações ($N \leq 30$). Resumindo o que encontramos até agora: a análise estatística do PPN revelou duas diferentes fases distingüidas pelo parâmetro κ (B.17), que corresponde ao número de bits significantes para codificar as entradas a_j divido por N. Para $\kappa < \kappa_c$ temos um número exponencial de partições perfeitas e conseqüentemente um número exponencial de soluções do PPN. Para $\kappa > \kappa_c$ temos somente duas soluções ótimas cuja função é dada por(B.26).

Apêndice C

Heurísticas para a Partição Numérica

Nosso objetivo neste apêndice é apresentar uma revisão das principais heurísticas disponíveis na literatura para resolver o problema da partição numérica (PPN). Em particular, destacamos a heurística da diferença de Karmarkar & Karp (KK), que fornece uma solução que escala com $N^{-\alpha \ln N}$. Através de simulações, estimamos o valor da constante α. Discutimos também um algoritmo de enumeração exato que fornece uma solução ótima para o PPN e utiliza a heurística KK como elemento básico de sua implementação.

C.0.5 A heurística do avarento

Uma heurística pode ser um procedimento extremamente simples, como por exemplo, a heurística do avarento que descreveremos a seguir e que talvez seja o procedimento mais óbvio para se tentar resolver o PPN. A receita é:

Algoritmo:

- Ordenar uma lista de números de forma decrescente
- Colocar o maior número em um subconjunto arbitrário
- Cada número remanescente é colocado no subconjunto com a menor soma total

Por exemplo, dada uma lista ordenada (8,7,6,5,4) essa heurística procede da seguinte forma:

(1) $A = 8 \qquad A^c = \emptyset$

(2) $A = 8 \qquad A^c = 7$

(3) $A = 8 \qquad A^c = 7, 6$

(4) $A = 8, 5 \qquad A^c = 7, 6$

(5) $A = 8, 5, 4 \qquad A^c = 7, 6$

Portanto, a partição encontrada tem valor 4. A heurística do avarento requer $\mathcal{O}(N \ln N)$ para ordenar uma lista de N números e $\mathcal{O}(N)$ para designá-los aos subconjuntos, resultando num tempo de complexidade $\mathcal{O}(N \ln N)$.

C.0.6 A heurística de Karmarkar & Karp

Assim como a heurística do avarento discutida acima, muitas outras heurísticas foram propostas para minimizar a função custo do problema da partição numérica (nesta tese mesmo propomos a heurística da integralização baseada no algoritmo de relaxação). A mais bem sucedida e por isso mesmo também a mais popular é a heurística de Karmarkar & Karp (KK) – um algoritmo polinomial que fornece soluções aproximadas. De forma objetiva, enunciamos a seguir o método da diferença dos conjuntos proposto por Karmarkar e Karp [56]:

Algoritmo: Dada uma seqüência $A = \{a_1, a_2, ..., a_N\}$, repita os passos seguintes $N - 1$ vezes.

- Passo 1. Percorra a seqüência A para determinar o maior elemento a_j e o segundo maior elemento a_k.
- Passo 2. Troque a_j por $a_j - a_k$ e troque a_k por zero na seqüência A. Observe que um item de valor zero pode ser ignorado (removido da lista) pois não contribui para a soma dos ítens no conjunto.

Vejamos um exemplo: dada a mesma lista ordenada anterior $(8, 7, 6, 5, 4)$, primeiramente trocamos os elementos 8 e 7 pela sua diferença e então inserimos o resultado novamente na lista obtendo $(6, 5, 4, 1)$. A seguir, o 6 e o 5 são trocados pela sua diferença fornecendo $(4, 1, 1)$. O 4 e o 1 são trocados pela sua diferença levando a $(3, 1)$ e, finalmente, a diferença dos dois últimos números no subconjunto final é igual a 2. A heurística KK também falha para encontrar a partição ótima, mas produz um resultado melhor do que a herística do avarento. De fato, KK é a melhor heurística polinomial já proposta para o PPN. Para entender o que esta heurística está fazendo vamos tentar construir a partição gerada por ela que resultou numa diferença igual a 2. Imagine que cada operação adicione uma aresta entre dois números, significando que eles estão em subconjuntos diferentes. O grafo resultante forma uma árvore geradora, sendo os nós formados pelos elementos da lista, que podem ser 2-colorível (jargão da teoria dos grafos, ver apêndice A), determinando assim a partição, de forma que os números com a mesma cor vão para o mesmo subconjunto. Em geral esta árvore possui $N - 1$ arestas geradas pela operação de diferença. A figura C.1 ilustra bem este mecanismo. Note apenas que nós adjacentes não recebem a mesma cor.

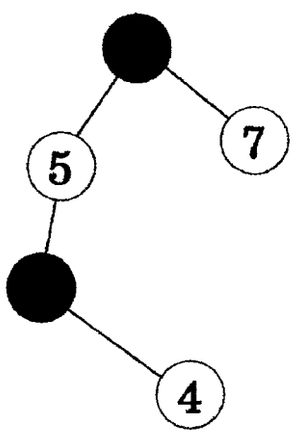


Figura C.1: Árvore representativa da construção da melhor partição dada pela heurística KK

Este algoritmo requer $\mathcal{O}(N \ln N)$ para ordenar uma lista de N números e $\mathcal{O}(N)$ para colorir o grafo, produzindo um tempo de complexidade de $\mathcal{O}(N \ln N)$. Outra característica importante é que quanto mais números colocamos na lista, mais operações de diferença são executadas e, conseqüentemente, menor será o resíduo. Mais especificamente, Karmarkar e Karp mostraram que o valor final encon-

trado é $\mathcal{O}(N^{-\alpha \ln N})$ para alguma constante α.

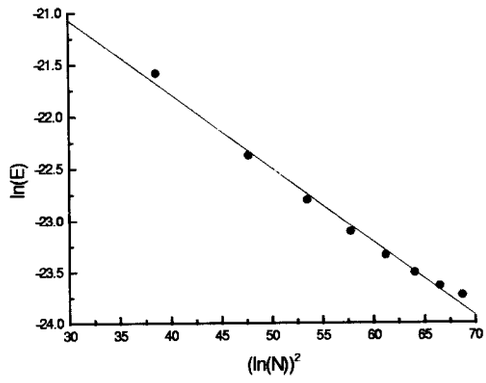


Figura C.2: Logaritmo da função custo do PPN encontrado pela heurística KK em função do quadrado do logaritmo do tamanho do problema N. A linha contínua foi obtida por regressão linear e a inclinação fornece o valor da constante $\alpha = 0.07$.

Através de simulações estimamos o valor dessa constante como $\alpha \approx 0.07$. Cada ponto na figura C.2 é o resultado da média sobre 500 experimentos (realizações das seqüências) independentes. Observe que KK é uma heurística determinística.

C.0.7 O algoritmo de Karmarkar & Karp Completo

Ambas heurísticas apresentadas anteriormente têm complexidade $\mathcal{O}(N \ln N)$, mas encontram soluções apenas aproximadas. Para obter a solução ótima o algoritmo óbvio é uma busca exaustiva em uma árvore binária cujo tempo de complexidade é $\mathcal{O}(2^N)$. O critério de parada de um tal algoritmo é encontrar uma partição perfeita, se ela existir; caso contrário, temos que ir até o fim guardando em cada passo o valor da menor diferença. Korf propôs um algoritmo de busca eficiente para obter a solução ótima do PPN inspirando-se na heurística KK; o algoritmo resultante foi denominado Karmarkar & Karp Completo (KKC) [54]. A idéia do KKC é realmente muito simples. A cada passo a heurística KK troca os dois maiores números de uma lista ordenada pela sua diferença. A outra possibilidade é trocar os dois maiores números pela sua soma, indicando assim que os dois números vão para o mesmo subconjunto. O KKC também é uma busca em árvore (ver figura C.3), onde no ramo da

esquerda trocamos os dois maiores números pela diferença enquanto no ramo da direita os trocamos pela soma. A diferença tem que ser colocada na ordem correta dentro da lista, mas a soma fica sempre no topo pois ela certamente fornece um número maior do que os remanescentes. A busca em cada ramo não precisa ir até o fim: existe um critério de corte ou parada. Um ramo termina quando o maior elemento for maior ou igual do que a soma de todos os elementos remanescentes, ou então quando encontramos uma partição ótima, isto é, de diferença zero. Note que a primeira solução que o KKC retorna é aquela fornecido pela heurística KK. A figura C.3 ilustra esse procedimento. Observe que, conforme esperado, o algoritmo KKC encontra a partição ótima, neste caso a partição perfeita $\{8, 7\}$ e $\{6, 5, 4\}$.

Algoritmo:

- Passo 1. Disponha os elementos da seqüência A na ordem decrescente.
- Passo 2. Aplique a heurística KK e guarde os elementos, sempre ordenados, correspondente a cada passo em um vetor B(n), ao qual associamos um nó. Neste ponto rotule o nó explorado por 1.
- Passo 3. Volte ao nó precedente e troque o maior elemento pela soma e o segundo maior elemento por zero, gerando um novo nó. Rotule o nó associado a essa operação por 2.
- Passo 4. Guarde o menor valor encontrado na variável AUX. Se AUX for igual a zero, termine o programa; caso contrário, procure um nó com rótulo igual a 1.
- Passo 5. Se o maior elemento for menor do que a soma dos remanescentes, termine o programa; caso contrário, procure outro nó com rótulo igual a 1 e aplique o passo 3 novamente.
- Passo 6. Termine o programa quando AUX for igual a zero ou quando todos os nós tiverem sido explorados.

É interessante comparar os desempenhos das heurísticas do avarento e KK com o desempenho ótimo obtido pelo algoritmo KKC. O resultado dessa comparação é mostrado na figura C.4 e ilustra de

forma inequívoca o desafio representado pelo problema da partição numérica mesmo para problemas relativamente pequenos: as funções custo obtidas com métodos distintos diferem por algumas ordens de magnitude! Já a figura C.5 apresenta o número de nós pesquisados pelo KKC para encontrar a solução ótima: quanto maior o número de nós, maior é a duração da busca e, portanto, mais difícil é o problema.

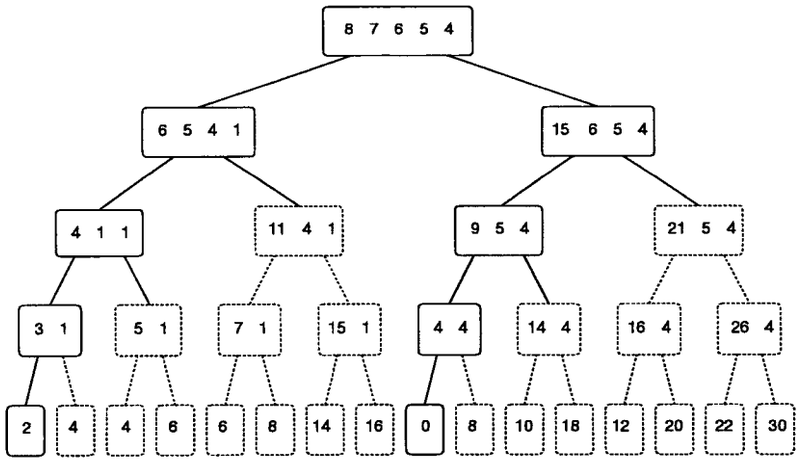
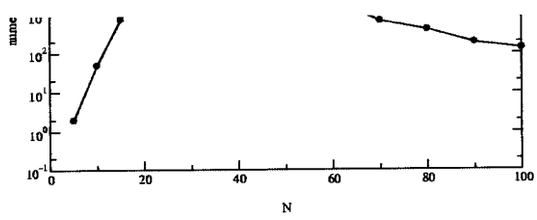


Figura C.3: Árvore gerada pelo algoritmo KKC quando aplicado a uma lista formada pelos números 8,7,6,5,4. Ramos esquerdos: troca-se os dois maiores elementos pela diferença entre eles. Ramos direitos: troca-se os dois maiores elementos pela sua soma. As linhas pontilhadas representam os nós cortados pelo algoritmo durante a busca.



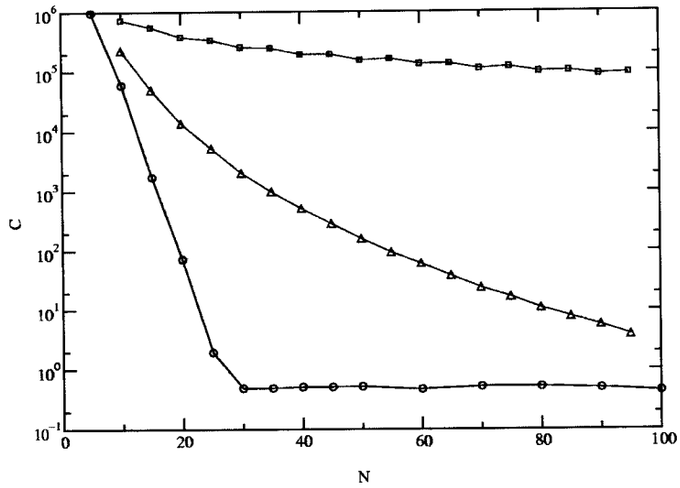


Figura C.4: C é a função custo (diferença) mínima e N é o tamanho do problema. Todos os pontos foram obtidos fazendo-se a média sobre 1000 experimentos independentes. Os valores inteiros a_i a serem particionados foram gerados aleatoriamente usando uma distribuição uniforme no intervalo $(0, 10^7]$. As linhas vermelha, verde e preta correspondem à heurística do avarento, à heurística KK e ao algoritmo KKC, respectivamente.

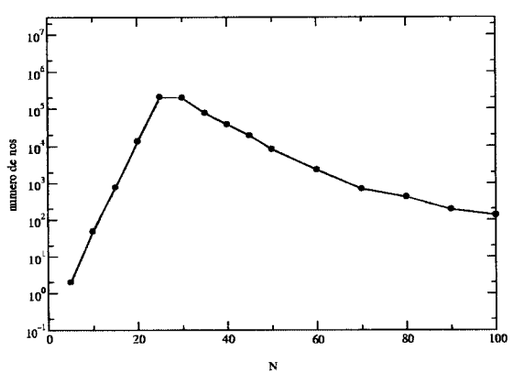


Figura C.5: O número médio de nós gerados durante a busca da solução ótima pelo KKC como função do tamanho do problema. Os parâmetros do experimento são os mesmos da figura anterior.

Apêndice D

Análise de Estabilidade

Para determinar a validade das soluções com simetria de réplicas consideradas no capítulo 4, devemos analisar os efeitos das flutuações dos parâmetro de ordem $q_{\alpha\beta}$ e $\tilde{q}_{\alpha\beta}$ em torno dos valores simétricos em relação as réplicas q e $\tilde{q}$, respectivamente. Para isto devemos calcular e diagonalizar a matriz de derivadas segundas da energia livre (4.27). Seguindo o procedimento padrão [46], [47], [49], a condição de estabilidade é obtida para

$$W = \alpha\lambda\lambda' < 1 \tag{D.1}$$

e λ e λ' são autovalores especiais a serem definidos a seguir.

O autovalor λ é definido por

$$\lambda = \overline{P} - 2\,\overline{Q} + \overline{R} \tag{D.2}$$

onde

$$\overline{P} = \frac{\partial^2 \ln G_1}{\partial q_{\nu\delta}\partial q_{\nu\delta}}, \tag{D.3}$$

$$\overline{Q} = \frac{\partial^2 \ln G_1}{\partial q_{\nu\delta}\partial q_{\nu\gamma}}, \tag{D.4}$$

$$\overline{R} = \frac{\partial^2 \ln G_1}{\partial q_{\nu\delta}\partial q_{\gamma\delta}}, \tag{D.5}$$

e

$$G_1 = \int_{-\infty}^{\infty} \left[\prod_{\rho} \frac{d\tilde{x}_\rho}{\sqrt{2\beta\pi}} \right] e^{-\sum_\rho \frac{\tilde{x}_\rho^2}{4\beta} - \frac{1}{2}Q\sum_\rho \tilde{x}_\rho^2 - \sum_{\rho<\delta} \tilde{x}_\rho \tilde{x}_\delta q_{\rho\delta} + im\sum_\rho \tilde{x}_\rho R_\rho}. \tag{D.6}$$

A barra sobre as grandezas definidas em (D.3) - (D.5) é apenas para diferenciá-las dos parâmetros de ordem introduzidos anteriormente. O cálulo deve ser efetuado derivando-se duas vezes $\ln G_1$ em relação a $q_{\alpha\beta}$ e logo a seguir aplicando-se o ansatz de simetria de réplica. Daí

$$\frac{\partial \ln G_1}{\partial q_{\nu\delta}} = \frac{\int_{-\infty}^{\infty} \left[\prod_\rho \frac{d\tilde{x}_\rho}{\sqrt{2\beta\pi}} \right] \tilde{x}_\nu \tilde{x}_\delta e^{-\sum_\rho \frac{\tilde{x}_\rho^2}{4\beta} - \frac{1}{2}Q\sum_\rho \tilde{x}_\rho^2 - \sum_{\rho<\delta} \tilde{x}_\rho \tilde{x}_\delta q_{\rho\delta} + im\sum_\rho \tilde{x}_\rho R_\rho}}{G_1} \tag{D.7}$$

e

$$\begin{aligned} \frac{\partial^2 \ln G_1}{\partial q_{\nu\delta}\partial q_{\nu\delta}} &= \frac{\int_{-\infty}^{\infty} \left(\prod_\rho \frac{d\tilde{x}_\rho}{\sqrt{2\beta\pi}} \right) (\tilde{x}_\nu \tilde{x}_\delta)^2 e^{-\sum_\rho \frac{\tilde{x}_\rho^2}{4\beta} - \frac{1}{2}Q\sum_\rho \tilde{x}_\rho^2 - \sum_{\rho<\delta} \tilde{x}_\rho \tilde{x}_\delta q_{\rho\delta} + im\sum_\rho \tilde{x}_\rho R_\rho}}{G_1} \\ &- \left[\frac{\int_{-\infty}^{\infty} \left(\prod_\rho \frac{d\tilde{x}_\rho}{\sqrt{2\beta\pi}} \right) \tilde{x}_\nu \tilde{x}_\delta e^{-\sum_\rho \frac{\tilde{x}_\rho^2}{4\beta} - \frac{1}{2}Q\sum_\rho \tilde{x}_\rho^2 - \sum_{\rho<\delta} \tilde{x}_\rho \tilde{x}_\delta q_{\rho\delta} + im\sum_\rho \tilde{x}_\rho R_\rho}}{G_1} \right]^2 . \end{aligned} \tag{D.8}$$

Lembrando que com a condição de simetria de réplicas podemos escrever

$$\sum_{\rho<\delta} \tilde{x}_\rho \tilde{x}_\delta q_{\rho\delta} = \frac{q}{2} \left[(\sum_\rho x_\rho)^2 - \sum_\rho x_\rho^2 \right] \tag{D.9}$$

e

$$e^{-\frac{q}{2}(\sum_\rho x_\rho)^2} = \int_{-\infty}^{\infty} Dz e^{i\sqrt{q}z\sum_\rho x_\rho} \tag{D.10}$$

obtemos

$$\frac{\partial^2 \ln G_1}{\partial q_{\nu\delta}\partial q_{\nu\delta}} = \frac{\int_{-\infty}^{\infty} Dz \int_{-\infty}^{\infty} \left(\prod_\rho \frac{d\tilde{x}_\rho}{\sqrt{2\beta\pi}} \right) (\tilde{x}_\nu \tilde{x}_\delta)^2 e^{-\sum_\rho \frac{\tilde{x}_\rho^2}{4\beta} - \frac{1}{2}Q\sum_\rho \tilde{x}_\rho^2 + \frac{q}{2}\sum_\rho x_\rho^2 + i\sqrt{q}z\sum_\rho x_\rho + imR\sum_\rho \tilde{x}_\rho}}{\int_{-\infty}^{\infty} Dz \int_{-\infty}^{\infty} \left(\prod_\rho \frac{d\tilde{x}_\rho}{\sqrt{2\beta\pi}} \right) e^{-\sum_\rho \frac{\tilde{x}_\rho^2}{4\beta} - \frac{1}{2}Q\sum_\rho \tilde{x}_\rho^2 + \frac{q}{2}\sum_\rho x_\rho^2 + i\sqrt{q}z\sum_\rho x_\rho + imR\sum_\rho \tilde{x}_\rho}} -$$

$$\left[\frac{\int_{-\infty}^{\infty}\left(\prod_\rho \frac{d\tilde{x}_\rho}{\sqrt{2\beta\pi}}\right)\tilde{x}_\nu\tilde{x}_\delta e^{-\sum_\rho \frac{\tilde{x}_\rho^2}{4\beta}-\frac{1}{2}Q\sum_\rho \tilde{x}_\rho^2+\frac{q}{2}\sum_\rho x_\rho^2+i\sqrt{q}z\sum_\rho x_\rho+imR\sum_\rho \tilde{x}_\rho}}{\int_{-\infty}^{\infty}Dz\int_{-\infty}^{\infty}\left(\prod_\rho \frac{d\tilde{x}_\rho}{\sqrt{2\beta\pi}}\right)e^{-\sum_\rho \frac{\tilde{x}_\rho^2}{4\beta}-\frac{1}{2}Q\sum_\rho \tilde{x}_\rho^2+\frac{q}{2}\sum_\rho x_\rho^2+i\sqrt{q}z\sum_\rho x_\rho+imR\sum_\rho \tilde{x}_\rho}}\right]^2 . \quad \text{(D.11)}$$

Agora não há mais necessidade de mantermos os índices e, assim sendo, escrevemos explicitamente

$$\frac{\partial^2 \ln G_1}{\partial q_{\nu\delta}\partial q_{\nu\delta}} = \frac{\int_{-\infty}^{\infty}Dz\left[\int_{-\infty}^{\infty}\frac{d\tilde{x}}{\sqrt{2\beta\pi}}\tilde{x}^2 e^{-\phi\tilde{x}^2+i\psi\tilde{x}}\right]^2\left[\int_{-\infty}^{\infty}\frac{d\tilde{x}}{\sqrt{2\beta\pi}}e^{-\phi\tilde{x}^2+i\psi\tilde{x}}\right]^{n-2}}{\int_{-\infty}^{\infty}Dz\left[\int_{-\infty}^{\infty}\frac{d\tilde{x}}{\sqrt{2\beta\pi}}e^{-\phi\tilde{x}^2+i\psi\tilde{x}}\right]^{n}} - \left[\frac{\int_{-\infty}^{\infty}Dz\left[\int_{-\infty}^{\infty}\frac{d\tilde{x}}{\sqrt{2\beta\pi}}\tilde{x}e^{-\phi\tilde{x}^2+i\psi\tilde{x}}\right]^2\left[\int_{-\infty}^{\infty}\frac{d\tilde{x}}{\sqrt{2\beta\pi}}e^{-\phi\tilde{x}^2+i\psi\tilde{x}}\right]^{n-2}}{\int_{-\infty}^{\infty}Dz\left[\int_{-\infty}^{\infty}\frac{d\tilde{x}}{\sqrt{2\beta\pi}}e^{-\phi\tilde{x}^2+i\psi\tilde{x}}\right]^{n}}\right]^2 \quad \text{(D.12)}$$

onde introduzimos a notação

$$\phi = \frac{1}{4\beta}+\frac{1}{2}(Q-q), \quad \text{(D.13)}$$

$$\psi = mR+\sqrt{q}z. \quad \text{(D.14)}$$

Finalmente, no limite $n \to 0$ obtemos

$$\frac{\partial^2 \ln G_1}{\partial q_{\nu\delta}\partial q_{\nu\delta}} = \int_{-\infty}^{\infty}Dz\frac{\left[\int_{-\infty}^{\infty}\frac{d\tilde{x}}{\sqrt{2\beta\pi}}\tilde{x}^2 e^{-\phi\tilde{x}^2+i\psi\tilde{x}}\right]^2}{\left[\int_{-\infty}^{\infty}\frac{d\tilde{x}}{\sqrt{2\beta\pi}}e^{-\phi\tilde{x}^2+i\psi\tilde{x}}\right]^2} - \left[\int_{-\infty}^{\infty}Dz\frac{\left[\int_{-\infty}^{\infty}\frac{d\tilde{x}}{\sqrt{2\beta\pi}}\tilde{x}e^{-\phi\tilde{x}^2+i\psi\tilde{x}}\right]^2}{\left[\int_{-\infty}^{\infty}\frac{d\tilde{x}}{\sqrt{2\beta\pi}}e^{-\phi\tilde{x}^2+i\psi\tilde{x}}\right]^2}\right]^2 . \quad \text{(D.15)}$$

Repetindo este procedimento, encontramos

$$\frac{\partial^2 \ln G_1}{\partial q_{\nu\delta}\partial q_{\nu\gamma}} = \int_{-\infty}^{\infty}Dz\frac{\left[\int_{-\infty}^{\infty}\frac{d\tilde{x}}{\sqrt{2\beta\pi}}\tilde{x}^2 e^{-\phi\tilde{x}^2+i\psi\tilde{x}}\right]\left[\int_{-\infty}^{\infty}\frac{d\tilde{x}}{\sqrt{2\beta\pi}}\tilde{x}e^{-\phi\tilde{x}^2+i\psi\tilde{x}}\right]^2}{\left[\int_{-\infty}^{\infty}\frac{d\tilde{x}}{\sqrt{2\beta\pi}}e^{-\phi\tilde{x}^2+i\psi\tilde{x}}\right]^3} - \left[\int_{-\infty}^{\infty}Dz\frac{\left[\int_{-\infty}^{\infty}\frac{d\tilde{x}}{\sqrt{2\beta\pi}}\tilde{x}e^{-\phi\tilde{x}^2+i\psi\tilde{x}}\right]^2}{\left[\int_{-\infty}^{\infty}\frac{d\tilde{x}}{\sqrt{2\beta\pi}}e^{-\phi\tilde{x}^2+i\psi\tilde{x}}\right]^2}\right]^2 , \quad \text{(D.16)}$$

$$\frac{\partial^2 \ln G_1}{\partial q_{\nu\delta}\partial q_{\gamma\delta}} = \int_{-\infty}^{\infty} Dz \frac{\left[\int_{-\infty}^{\infty} \frac{d\tilde{x}}{\sqrt{2\beta\pi}} \tilde{x} e^{-\phi\tilde{x}^2+i\psi\tilde{x}}\right]^4}{\left[\int_{-\infty}^{\infty} \frac{d\tilde{x}}{\sqrt{2\beta\pi}} e^{-\phi\tilde{x}^2+i\psi\tilde{x}}\right]^4} - \left[\int_{-\infty}^{\infty} Dz \frac{\left[\int_{-\infty}^{\infty} \frac{d\tilde{x}}{\sqrt{2\beta\pi}} \tilde{x} e^{-\phi\tilde{x}^2+i\psi\tilde{x}}\right]^2}{\left[\int_{-\infty}^{\infty} \frac{d\tilde{x}}{\sqrt{2\beta\pi}} e^{-\phi\tilde{x}^2+i\psi\tilde{x}}\right]^2}\right]^2 . \tag{D.17}$$

Daí, de acordo com (D.2) o autovalor procurado é

$$\begin{aligned} \lambda &= \int_{-\infty}^{\infty} Dz \frac{\left[\int_{-\infty}^{\infty} d\tilde{x}\tilde{x}^2 e^{-\phi\tilde{x}^2+i\psi\tilde{x}}\right]^2}{\left[\int_{-\infty}^{\infty} d\tilde{x} e^{-\phi\tilde{x}^2+i\psi\tilde{x}}\right]^2} - 2\frac{\left[\int_{-\infty}^{\infty} d\tilde{x}\tilde{x}^2 e^{-\phi\tilde{x}^2+i\psi\tilde{x}}\right]\left[\int_{-\infty}^{\infty} d\tilde{x}\tilde{x} e^{-\phi\tilde{x}^2+i\psi\tilde{x}}\right]^2}{\left[\int_{-\infty}^{\infty} d\tilde{x} e^{-\phi\tilde{x}^2+i\psi\tilde{x}}\right]^3} \\ &+ \frac{\left[\int_{-\infty}^{\infty} d\tilde{x}\tilde{x}\, e^{-\phi\tilde{x}^2+i\psi\tilde{x}}\right]^4}{\left[\int_{-\infty}^{\infty} d\tilde{x}\, e^{-\phi\tilde{x}^2+i\psi\tilde{x}}\right]^4}. \end{aligned} \tag{D.18}$$

Podemos escrever λ em um formato mais compacto:

$$\lambda = \int_{-\infty}^{\infty} Dz \left[\langle \tilde{x}^2 \rangle - \langle \tilde{x} \rangle^2\right]^2 \tag{D.19}$$

onde

$$\langle f(\tilde{x}) \rangle = \frac{\int_{-\infty}^{\infty} d\tilde{x} f(\tilde{x}) e^{-\phi\tilde{x}^2+i\psi\tilde{x}}}{\int_{-\infty}^{\infty} d\tilde{x} e^{-\phi\tilde{x}^2+i\psi\tilde{x}}}. \tag{D.20}$$

Os momentos podem ser facilmente obtidos, usando-se os seguintes resultados

$$\int_{-\infty}^{\infty} d\tilde{x}\tilde{x}^2 e^{-\phi\tilde{x}^2+i\psi\tilde{x}} = \left(\frac{1}{2\phi} - \frac{\psi^2}{4\phi^2}\right)\sqrt{\frac{\pi}{\phi}} e^{-\frac{\psi^2}{4\phi}}, \tag{D.21}$$

$$\int_{-\infty}^{\infty} d\tilde{x}\tilde{x} e^{-\phi\tilde{x}^2+i\psi\tilde{x}} = \frac{i\psi}{2\phi}\sqrt{\frac{\pi}{\phi}} e^{-\frac{\psi^2}{4\phi}}, \tag{D.22}$$

$$\int_{-\infty}^{\infty} d\tilde{x} e^{-\phi\tilde{x}^2+i\psi\tilde{x}} = \sqrt{\frac{\pi}{\phi}} e^{-\frac{\psi^2}{4\phi}}. \tag{D.23}$$

Resultando em uma expressão extremamente simples para λ, a saber,

$$\lambda = \int_{-\infty}^{\infty} Dz \left[\left(\frac{1}{2\phi} - \frac{\psi^2}{4\phi^2}\right) + \frac{\psi^2}{4\phi^2}\right]^2 = \frac{1}{(Q-q)^2}. \tag{D.24}$$

Vamos agora definir e calcular o outro autovalor mencionado em (D.1),

$$\lambda' = \vec{P}' - 2\,\vec{Q}' + \vec{R}' \tag{D.25}$$

onde

$$\vec{P}' = \frac{\partial^2 \ln G_2}{\partial \tilde{q}_{\nu\delta}\partial \tilde{q}_{\nu\delta}}, \tag{D.26}$$

$$\vec{Q}' = \frac{\partial^2 \ln G_2}{\partial \tilde{q}_{\nu\delta}\partial \tilde{q}_{\nu\gamma}}, \tag{D.27}$$

$$\vec{R}' = \frac{\partial^2 \ln G_2}{\partial \tilde{q}_{\nu\delta}\partial \tilde{q}_{\gamma\delta}}, \tag{D.28}$$

e

$$G_2 = \int_0^\infty [\prod_\rho da_\rho] e^{-\sum_\rho \tilde{Q}_\rho a_\rho^2 - \sum_{\rho<\delta} \tilde{q}_{\rho\delta} a_\rho a_\delta - \sum_\rho \tilde{R}_\rho a_\rho}. \tag{D.29}$$

Primeiramente vamos efetuar as derivadas em (D.26) para encontrar $\vec{P}'$,

$$\frac{\partial^2 \ln G_2}{\partial \tilde{q}_{\nu\delta}\partial \tilde{q}_{\nu\delta}} = \frac{\int_0^\infty [\prod_\rho da_\rho]\,(a_\nu a_\delta)^2\,\Omega_\rho}{G_2} - \left[\frac{\int_0^\infty [\prod_\rho da_\rho] a_\nu a_\delta \Omega_\rho}{G_2}\right]^2 \tag{D.30}$$

onde $\Omega_\rho = e^{-\sum_\rho \tilde{Q}_\rho a_\rho^2 - \sum_{\rho<\delta} \tilde{q}_{\rho\delta} a_\rho a_\delta - \sum_\rho \tilde{R}_\rho a_\rho}$. Agora aplicamos o ansatz de simetria de réplica para escrever Ω_ρ como

$$\begin{aligned}
\Omega_\rho &= \int_{-\infty}^{\infty} Dz e^{-\tilde{Q}\sum_\rho a_\rho^2 + i\sqrt{\tilde{q}}z\sum_\rho a_\rho - \frac{\tilde{q}}{2}\sum_\rho a_\rho^2 - \tilde{R}\sum_\rho a_\rho} \\
&= \int_{-\infty}^{\infty} Dz \prod_{\rho=1}^{n} e^{-\tilde{Q}a_\rho^2 - \frac{\tilde{q}}{2}a_\rho^2 + i\sqrt{\tilde{q}}z a_\rho - \tilde{R}a_\rho} \\
&= \int_{-\infty}^{\infty} Dz \prod_{\rho=1}^{n} e^{-\Delta a_\rho^2 - \chi a_\rho}
\end{aligned} \tag{D.31}$$

onde utilizamos a notação

$$\Delta = \tilde{Q} + \frac{\tilde{q}}{2}, \tag{D.32}$$

$$\chi = \tilde{R} + \sqrt{\tilde{q}}z. \tag{D.33}$$

Assim reescrevemos $\bar{P}'$ como

$$\frac{\partial^2 \ln G_2}{\partial \tilde{q}_{\nu\delta}\partial \tilde{q}_{\nu\delta}} = \frac{\int_{-\infty}^{\infty} Dz \left[\int_0^{\infty} daa^2 e^{-\Delta a^2 - \chi a}\right]^2 \left[\int_0^{\infty} da_\rho e^{-\Delta a_\rho^2 - \chi a_\rho}\right]^{n-2}}{\int_{-\infty}^{\infty} Dz \left[\int_0^{\infty} da_\rho e^{-\Delta a_\rho^2 - \chi a_\rho}\right]^{n}} - \left[\frac{\int_{-\infty}^{\infty} Dz \left[\int_0^{\infty} daae^{-\Delta a^2 - \chi a}\right]^2 \left[\int_0^{\infty} da_\rho e^{-\Delta a_\rho^2 - \chi a_\rho}\right]^{n-2}}{\int_{-\infty}^{\infty} Dz \left[\int_0^{\infty} da_\rho e^{-\Delta a_\rho^2 - \chi a_\rho}\right]^{n}}\right]^2. \tag{D.34}$$

Tomando o limite n$\rightarrow$ 0 obtemos finalmente

$$\bar{P}' = \int_{-\infty}^{\infty} Dz \frac{\left[\int_0^{\infty} daa^2 e^{-\Delta a^2 - \chi a}\right]^2}{\left[\int_0^{\infty} da_\rho e^{-\Delta a_\rho^2 - \chi a_\rho}\right]^2} \left[\int_{-\infty}^{\infty} Dz \frac{\left[\int_0^{\infty} daae^{-\Delta a^2 - \chi a}\right]^2}{\left[\int_0^{\infty} da_\rho e^{-\Delta a_\rho^2 - \chi a_\rho}\right]^2}\right]^2. \tag{D.35}$$

Repetindo o mesmo procedimento para calcular $\bar{Q}'$ e $\bar{R}'$

$$\bar{Q}' = \int_{-\infty}^{\infty} Dz \frac{\left[\int_0^{\infty} daa^2 e^{-\Delta a^2 - \chi a}\right] \left[\int_0^{\infty} daae^{-\Delta a^2 - \chi a}\right]^2}{\left[\int_0^{\infty} da_\rho e^{-\Delta a_\rho^2 - \chi a_\rho}\right]^3} - \left[\int_{-\infty}^{\infty} Dz \frac{\left[\int_0^{\infty} daae^{-\Delta a^2 - \chi a}\right]^2}{\left[\int_0^{\infty} da_\rho e^{-\Delta a_\rho^2 - \chi a_\rho}\right]^2}\right]^2 \tag{D.36}$$

e

$$R' = \int_{-\infty}^{\infty} Dz \frac{\left[\int_0^{\infty} daa^2 e^{-\Delta a^2 - \chi a}\right]^4}{\left[\int_0^{\infty} da_\rho e^{-\Delta a_\rho^2 - \chi a_\rho}\right]^4} \left[\int_{-\infty}^{\infty} Dz \frac{\left[\int_0^{\infty} daae^{-\Delta a^2 - \chi a}\right]^2}{\left[\int_0^{\infty} da_\rho e^{-\Delta a_\rho^2 - \chi a_\rho}\right]^2}\right]^2. \tag{D.37}$$

Finalmente escrevemos λ' de forma compacta como

$$\lambda' = \int_{-\infty}^{\infty} Dz \left[\langle a^2 \rangle - \langle a \rangle^2\right]^2 \tag{D.38}$$

onde

$$\langle f(a)\rangle = \frac{\int_{-\infty}^{\infty} da f(a) e^{-\Delta a^2 - \chi a}}{\int_{-\infty}^{\infty} da e^{-\Delta a^2 - \chi a}}. \tag{D.39}$$

Os momentos podem ser facilmente obtidos usando-se os resultados

$$\int_0^{\infty} da a^2 e^{-\Delta a^2 - \chi a} = \frac{1}{2}\left(\frac{1}{2\Delta} + \frac{\chi^2}{4\Delta^2}\right)\sqrt{\frac{\pi}{\Delta}} e^{\frac{\chi^2}{4\Delta}} \mathrm{erfc}(\frac{\chi}{2\sqrt{\Delta}}) - \frac{\chi}{4\Delta^2}, \tag{D.40}$$

$$\int_0^{\infty} da a e^{-\Delta a^2 - \chi a} = \frac{1}{2\Delta} - \frac{1}{4}\frac{\chi}{\Delta}\sqrt{\frac{\pi}{\Delta}} e^{\frac{\chi^2}{4\Delta}} \mathrm{erfc}(\frac{\chi}{2\sqrt{\Delta}}), \tag{D.41}$$

$$\int_0^{\infty} da e^{-\Delta a^2 - \chi a} = \frac{1}{2}\sqrt{\frac{\pi}{\Delta}} e^{\frac{\chi^2}{4\Delta}} \mathrm{erfc}(\frac{\chi}{2\sqrt{\Delta}}), \tag{D.42}$$

de modo que

$$\langle a\rangle = \frac{e^{-\frac{\chi^2}{4\Delta}}}{\sqrt{\pi\Delta}\mathrm{erfc}(\frac{\chi}{2\sqrt{\Delta}})} - \frac{1}{2}\frac{\chi}{\Delta}, \tag{D.43}$$

$$\langle a^2\rangle = \left(\frac{1}{2\Delta} + \frac{\chi^2}{4\Delta^2}\right) - \frac{1}{2}\frac{\chi}{\Delta}\frac{e^{-\frac{\chi^2}{4\Delta}}}{\sqrt{\pi\Delta}\mathrm{erfc}(\frac{\chi}{2\sqrt{\Delta}})}. \tag{D.44}$$

Definindo

$$A = \frac{e^{-\frac{\chi^2}{4\Delta}}}{\sqrt{\pi\Delta}\mathrm{erfc}(\frac{\chi}{2\sqrt{\Delta}})} \tag{D.45}$$

o autovalor λ' reduz-se a

$$\lambda' = \int_{-\infty}^{\infty} Dz \left[\frac{1}{2\Delta} + \frac{\chi A}{2\Delta} - A^2\right]^2. \tag{D.46}$$

Conhecidos as expressões de λ e λ', podemos prontamente determinar a região de estabilidade, bastando usar a condição (D.1). O fato é que W só pode ser avaliado numericamente, exceto no limite $q \to 1$, onde mostramos que $W \to 1$. Isto significa que somente a região abaixo de α_c é estável. Neste

limite temos $\chi \to \infty$ e $\Delta \to \infty$ segundo as equações (4.62) e (4.63). Assim expandindo (D.45)

$$A \approx \frac{\chi}{2\Delta}(1 + \frac{2\Delta}{\chi^2} + ...) = \frac{\chi}{2\Delta} + \frac{1}{\chi} + ..., \qquad para \ z > -z_0^* \tag{D.47}$$

e,substituindo na equação (D.46) a integral se anula. Para $z < -z_0^*$, observamos que

$$A \approx \frac{e^{-\frac{\chi^2}{4\Delta}}}{2\sqrt{\pi\Delta}} \to 0. \tag{D.48}$$

Daí,

$$\lambda' = \int_{-\infty}^{-z_0^*} Dz \left[\frac{1}{2\Delta}\right]^2 = \left[\frac{1}{2\Delta}\right]^2 \frac{1}{2}\mathrm{erfc}(\frac{z_0^*}{\sqrt{2}}) \tag{D.49}$$

e, consequentemente, quando $\alpha = \alpha_c$

$$W = \alpha_c \left[\frac{1}{1-q}\right]^2 \left[\frac{1}{2\Delta}\right]^2 \alpha_c = 1, \tag{D.50}$$

onde temos usado as equações (4.46) e (4.53), que nos permite mostrar que

$$\alpha_c = 2\Delta(1-q). \tag{D.51}$$

Verificamos assim que a condição de estabilidade está sempre satisfeita para $\alpha \leq \alpha_c$.

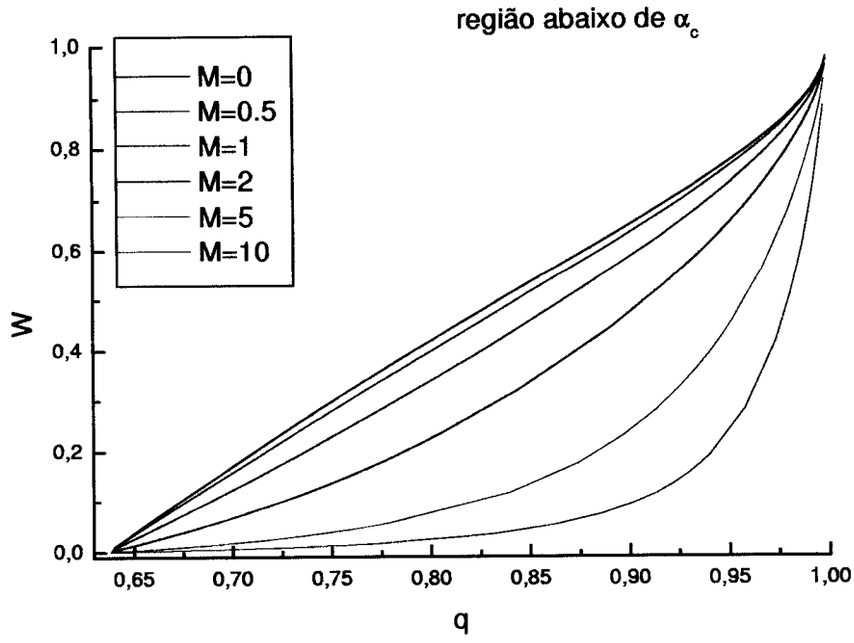


Figura D.1: O valor da grandeza W em função do parâmetro q na região abaixo de α_c, para vários valores da variável m.

Bibliografia

[1] M. R. Garey and D. S. Johnson. *Computers and Intractability. A Guide to the Theory of NP-Completeness.* W.H. Freeman, New York, 1997.

[2] V. Klee and G. J. Minty. How good is the simplex algorithm?. in O. Shisha(ed.), *Inequalities* **III**, Academic Press, New York, 159-175 (1972).

[3] S. F. Edwards and P. W. Anderson. *J. Phys.* **F5**:965, (1975).

[4] J. L. Van Hemmen and R. G. Palmer. *J. Phys. A: Math. Gen.* **12** 563 (1979).

[5] T. Y. Fu *Lectures in the Sciences of Complexity ed D Stein* (Reading, MA: Addison-Wesley; London: Longman) (1989).

[6] J. F. Fontanari. *J. Phys. A: Math. Gen.* **28** 4751 (1995).

[7] J. Inoue. *J. Phys. A: Math. Gen.* **30** 1047 (1997).

[8] W. Feller. *An Introduction to Probability Theory and its Applications*, **vol. II**. (New York: Wiley) (1957).

[9] G. Toulouse and J. Vannimenus. *Phys. Rep.* **67** 47 (1980).

[10] J. Vannimenus and M. Mézard. *J. Phys. Lett.* **45** L1145 (1984).

[11] M.J. Lighthill. *Introduction to Fourier Analysis and Generalised Functions.* Cambridge University Press (1959).

[12] P. W. Anderson. *Physica A* **140** 405 (1986).

[13] H. E. Daniels.*Ann. Math. Stat.* **25** 631 (1954).

[14] N.G. De Bruijn. *Asymptotic Methods in Analysis.* John Wiley & Sons, New York, (1961).

[15] J. Mathews and R.L. Walker. *Mathematical Methods of Physics.* Addison-Wesley Publishing Company, Inc., Redwood City, California, (1970).

[16] D. S. Johnson, C. R. Aragon, L. A. McGeoch and C. Schevon. *Operations Research* **39** 378 (1991).

[17] A. Erdélyi. *Tables of Integrals Transforms* (New York: McGraw-Hill),(1954).

[18] F. Tanaka and S. F. Edwards. *J. Phys. F: Metal Phys.* **13** 2769 (1980).

[19] A. Cobham, The intrinsic computational difficulty of function Proc. 1964 International Congress for Logic Methodology and Phylosophy of Science, North Holland, Amsterdam 24-30

[20] J. Edmonds. *Canad. J. Math.* **17** 449–467 (1965).

[21] B. G. Dantzig. *Econometrics* **28** 30-44 (1960).

[22] J. Edmonds. *Bull. Amer. Math. Soc.* **68** 494–499 (1962).

[23] S. Cook. The complexity of theorem proving procedures. In *Proceedings of the Third Annual ACM Symposium on Theory of Computing*, 151–158 (1971).

[24] R.M. Karp. Complexity of computer computations. In R.E. Miller and J.W. Thatcher, editors, *Reducibility Among Combinatorial Problems*, 85–103, New York, (1972). Plenum Press.

[25] M. Mézard, G. Parisi, and M.A. Virasoro. *Spin glass theory and beyond.* World Scientific, Singapore, (1987).

[26] M. Mézard and G. Parisi. *Europhys. Lett.***2(12)**:913–918, December (1986).

[27] W. Krauth and M. Mézard. *Europhys. Lett.*,**8(3)**:213–218 (1989).

[28] Y. Fu and P. W. Anderson. *J. Phys. A*, **19**:1605–1620 (1986).

[29] S. Kirkpatrick and B. Selman. *Science*, **264**:1297–1301, May (1994).

[30] R. Monasson and R. Zecchina. *Phys. Rev. E*, **56(2)**:1357–1370 (1997).

[31] M. Mézard and G. Parisi. *J. Physique Lett.* **46**:L771–L778 (1985).

[32] Giorgio Parisi. arXiv: *cond-mat/9801176*, (1998).

[33] V.S. Dotsenko. *J. Phys. A*, **33**:2015–2030 (2000).

[34] D. J. Aldous. http://www.stat.berkeley.edu/users/aldous, (2000).

[35] R.Monasson, R. Zecchina, S. Kirkpatrick, B. Selman, and L. Troyanksy. *Nature*, **400**:133–137, July (1999).

[36] M. Weigt and A.K. Hartmann. *Phys. Rev. Lett.*,**84**:6118 (2000).

[37] Stephan Mertens. *Phys. Rev. Lett.*, **81(20)**:4281–4284, November (1998).

[38] E.G. Coffman and G. S. Lueker. John Wiley & Sons, New York, (1991).

[39] Li-Hui Tsai. *SIAM J.Comput.*, **21(1)**:59–64 (1992).

[40] N. Karmarkar, R. M. Karp, G. S. Lueker, and A. M. Odlyzko. *J. Appl. Prob.*, **23**:626–645 (1986).

[41] R. L. Rivest, A. Shamir, L. Adelman, "On Digital Signatures and Public Key Cryptosystems. *MIT Laboratory for Computer Science Technical Memorandum* **82**, April (1977).

[42] C.H. Papadimitriou and K. Steiglitz. *Combinatorial Optimization.* Prentice-Hall, Englewood Cliffs, New Jersey, (1982).

[43] C. H. Papadimitriou. *Computational Complexity.* Addison-Wesley Publishing Company, Reading, Massachusetts, (1994).

[44] M. Plischke and B. Bergersen. *Equilibrium Statistical Physics.* World Scientific Publishing Co. Pte. Ltd., Singapore (1999).

[45] L. Reichl. *A modern course in Statistical Physics.* Austin: University of Texas Press (1980).

[46] E. Gardner and B. Derrida. *J.Phys.A: Math. Gen.* **21** 271 (1988).

[47] E. Gardner. *J.Phys.A: Math. Gen.* **21** 257 (1988).

[48] A. J. Bray and M. A. Moore. *J. Phys. C: Solid St. Phys.* **13** L469 (1980).

[49] J. R. L. Almeida and D. J. Thouless.*J. Phys. A* **11** 983 (1978).

[50] F. F. Ferreira and J. F. Fontanari. *J. Phys A:Math. Gen.* **31** 3417 (1998).

[51] F. F. Ferreira and J. F. Fontanari. *Physica A* **269** 58 (1999).

[52] F. F. Ferreira and J. F. Fontanari. *J Phys A:Math. Gen.* **33** 7265 (2000).

[53] R. E. Korf. In Chris S. Mellish, editor, *Proc. of IJCAI-95*, 266–272, San Mateo, CA, (1995). Morgan Kaufmann.

[54] Richard E. Korf. *Artificial Intelligence*, **106**:181–203 (1998).

[55] A. B. Diaz and P. Moscato. *Las Nuevas Tecnicas Heuristicas y las Redes Neuronales* ed B A Diaz (Madrid: Paraninfo), (1996).

[56] N. Karmakar and R. M. Karp. Report No. UCB/CSD 81/113, Computer Science Division, University of California, Berkeley, (1982).

[57] D. Sherrington and S. Kirkpatrick *Phys. Rev. Lett.* **35** (1975).

[58] F. Barahona. *J. Phys.A*, **15**:3241 (1982).

[59] N.G. De Bruijn. John Wiley & Sons, New York, (1961).

[60] P. Cheeseman, B. Kanefsky, and W. M. Taylor. Where the *really* hard problems are. In J. Mylopoulos and R. Rediter, editors, *Proc. of IJCAI-91*, 331–337, San Mateo, CA, (1991). Morgan Kaufmann.

[61] J. Galambos. *The Asymptotic Theory of Extreme Order Statistics*. Robert E. Krieger Publishing Co., Malabar, Florida, (1987).

[62] I. P. Gent and T. Walsh. Phase transitions from real computational problems. In *Proceedings of the 8th International Symposium on Artificial Intelligence*, pages 356–364, Monterrey, (1995). ITESM.

[63] I. P. Gent and T. Walsh. *Computational Intelligence*, **14(3)**:430–451 (1998).

[64] D.J. Gross and M. Mézard. The simplest spin glass. *Nucl. Phys. B*, **240**:431–452, 1984.

[65] W.D. Harvey and M.L. Ginsberg. Limited discrepancy search. In *Proceedings IJCAI-95*, 607–613, Montreal, Quebec, (1995).

[66] Ian P. Gent and Toby Walsh. Phase transitions and annealed theories: Number partitioning as a case study. In W. Wahlster, editor, *Proc. of ECAI-96*, pages 170–174, New York, 1996. John Wiley & Sons.

[67] S. Kirkpatrick, Jr. C.D. Gelatt, and M.P. Vecchi. *Science*, **220**(4598):671–680, May (1983).

[68] H. Rieger. Frustrated systems: Ground state properties via combinatorial optimization. In J. Kertesz and I. Kondor, editors, *Advances in Computer Simulation*, volume 501 of *Lecture Notes in Physics*, Berlin Heidelberg New York, 1998. Springer-Verlag.

[69] W. Ruml, J.T. Ngo, J. Marks, and S.M. Shieber. *Journal of Optimization Theory and Applications*, P**89(2)**:251–291, May (1996).

[70] Dorit Aharonov. Quantum computation. In Dietrich Stauffer, editor, *Annual Reviews of Computational Physics VI*. World Scientific, 1998. arXiv:quant-ph/9812037.

[71] R. Feynman. Simulating physics with computers. *Int. J. Theor. Phys.*, 21(6/7):467–488, 1982.

[72] M. Jerrum and A. Sinclair. Polynomial-time approximation algorithms for the Ising model. In *Proc. 17th ICALP*. EATCS, 1990.

[73] V.R. Pratt. *SIAM J. Comput.*, **4**:214–220, (1975).

[74] E.L. Lawler, J.K. Lenstra, A.H.G. Rinnoy Kan, and D.B. Shmoys, editors. *The Traveling Salesman Problem*. Wiley-Interscience series in discrete mathematics and optimization. John Wiley & Sons, New York, 1985.

[75] G. Reinelt. volume **840** of *Lecture Notes in Computer Science*. Springer-Verlag, Berlin Heidelberg New York, (1994).

[76] P.W. Shor. *SIAM J.Comp.*, **26(5)**:1484–1509, (1997).

[77] A. Shamir. Proc. 11th Ann. ACM Symp. on Theory of Computing, Association for Computing Machinery (New York) 118-129. (1979).

[78] T. H. Cormen, C. E. Leiserson, and R. L. Rivest. *Introduction to Algorithms*. The MIT Electrical Engineering and Computer Science Series. The MIT Press, Cambridge, Massachusetts, 1990.

[79] Roger Penrose. *The Emperor's new Mind.* Oxford University Press, (1989).

[80] D. Coppersmith and S. Winograd. *J. Symbolic Computation*, **9(3)**:251–280 (1990).

[81] B. Bollobás. *Modern Graph Theory*, volume 184 of *Graduate Texts in Mathematics*. Springer-Verlag, Berlin (1998).

[82] H.N. Gabow, Z. Galil, T.H. Spencer, and R.E. Tarjan. *Combinatorica*, **6**:109–122, (1986).

[83] Stephan Mertens. *Phys. Rev. Lett.*, 84(7):1347–1350, February 2000.

[84] Stephan Mertens. *Theor. Comp. Sci.*, to appear 2001.

[85] R.E. Ladner. *Journal of the ACM*, **22**:155–171 (1975).

[86] G. V. R. Viana. Meta-heurísticas e programação paralela em otimização combinatória. Fortaleza: UFC (1998).

[87] H. R. Lewis and C. H. Papadimitriou.. *Scientific American*, **1** 96–109 (1978).

[88] L. Euler. *Comm. Acad. Sci. Imper. Petropol.*, **8**:128–140, (1736).

[89] V. Kumar. *AI Magazine*, **13(1)**:32–44 (1992)

[90] B.Aspvall, M. Plass, and R. Endre Tarjan. *Information Processing Letters*, **8(3)**:121–123 (1979).

[91] D.S. Johnson and C.H. Papadimitriou. Computational complexity. In Lawler et al. [74], pages 37–85.

[92] D.J.A. Welsh. The computational complexity of some classical problems from statistical physics. In G.R. Grimmett and D.J.A. Welsh, editors, *Disorder in Physical Systems*, pages 307–321. Clarendon Press, Oxford, 1990.

[93] G. Ausiello, P. Crescenzi, G. Gambosi, V. Kann, A. Marchetti-Spaccamela, and M. Protassi. *Complexity and Approximation.* Springer-Verlag, Berlin Heidelberg New York, 1999.

[94] Pierluigi Crescenzi and Viggo Kann. A compendium of NP optimization problems. http://www.nada.kth.se/~viggo/wwwcompendium.

[95] B. Hayes. *American Scientist*, **85(2)**:108–112, March-April (1997).

[96] L.G. Valiant. *Theor.Comp. Sci.*, **8**:189–201, (1979).

[97] E. Nagel and J. R. Newman, Prova de Gödel, *coleção debates* , editora perspectiva (1973).

[98] A. R. Lima and M. A. F. de Menezes, Entropy-based analysis of the Number Partitioning Problem. Phys. Rev. E63, RC,020106-1 (2001).

```
//*********************************************/
//29/03/2001
//qualificacao
/

#include <stdio.h>
#include <stdlib.h>
#include <math.h>
#include <malloc.h>
#include <time.h>

#define NM 50  //dimensao da matriz (NM x NM)
#define L NM*NM       //dimensao do vetor
//#define TEMPO 30000
#define NMEDIAS 30
#define NV 2           //numero de features
#define P 2.           //P=NV.
#define P_d 0.         ///difusao


void vizinhos(int k,int V[4]);
void difusao();
int  sitios();
void atualizacao();
int poisson(double lambda);
void cond_inicial(double lambda);
void estrutura(int n,int a, int b, int c, int d, int e) ;
void contagem();
void funcprint_maiores(int nprint,int jamx);
void funcprint(int nprint);
void impressao();
void quick_struct(int left, int right);
int medida_cluster(int va[L]);

int *A;
struct copia {
  int feature[NV];
};
struct copia individuo[L],tipo[L];
struct ordem {
  int posicao;
  int quantidade;
};
struct ordem ordenacao[L],ordenacao_aux;
void main ()
{
  int j,m,stime,a,i,jmax,n,na,va[L];
  double lambda,tamanho,xux,maior_cl;
  long ltime;



  FILE *br;
  br=fopen("teste.dat","a");

  ltime = time(NULL);
  stime = (unsigned) ltime/2;
  srand(stime);

  A = malloc(L * sizeof (int *));

  for(n = 0; n <= 10;n+=2){
    lambda = n/10.;
    tamanho = 0.;
    for (m = 0; m < NMEDIAS; m++){
```

```
                                        ************************/
//29/03/2001
//qualificacao
/*

#include <stdio.h>
#include <stdlib.h>
#include <math.h>
#include <malloc.h>
#include <time.h>

#define NM 50  //dimensao da matriz (NM x NM)
#define L NM*NM       //dimensao do vetor
//#define TEMPO 30000
#define NMEDIAS 30
#define NV 2            //numero de features
#define P 2.            //P=NV.
#define P_d 0.          ///difusao


void vizinhos(int k,int V[4]);
void difusao();
int  sitios();
void atualizacao();
int poisson(double lambda);
void cond_inicial(double lambda);
void estrutura(int n,int a, int b, int c, int d, int e) ;
void contagem();
void funcprint_maiores(int nprint,int jamx);
void funcprint(int nprint);
void impressao();
void quick_struct(int left, int right);
int medida_cluster(int va[L]);

int *A;
struct copia {
  int feature[NV];
};
struct copia individuo[L],tipo[L];
struct ordem {
  int posicao;
  int quantidade;
};
struct ordem ordenacao[L],ordenacao_aux;
void main ()
{
  int j,m,stime,a,i,jmax,n,na,va[L];
  double lambda,tamanho,xux,maior_cl;
  long ltime;



  FILE *br;
  br=fopen("teste.dat","a");

  ltime = time(NULL);
  stime = (unsigned) ltime/2;
  srand(stime);

  A = malloc(L * sizeof (int *));

  for(n = 0; n <= 10;n+=2){
    lambda = n/10.;
    tamanho = 0.;
    for (m = 0; m < NMEDIAS; m++){
```