



O PROBLEMA DE RECOBRIMENTO MÍNIMO DE UM CORPO EM TRÊS
DIMENSÕES POR ESFERAS DE DIFERENTES RAIOS

Marilis Bahr Karam Venceslau

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia de Sistemas e Computação, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia de Sistemas e Computação.

Orientador: Nelson Maculan Filho

Rio de Janeiro

Maio de 2015

O PROBLEMA DE RECOBRIMENTO MÍNIMO DE UM CORPO EM TRÊS
DIMENSÕES POR ESFERAS DE DIFERENTES RAIOS

Marilis Bahr Karam Venceslau

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ
COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE)
DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS
REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR
EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

Examinada por:

Prof. Nelson Maculan Filho, D.Sc.

Prof. Adilson Elias Xavier, D.Sc.

Prof^ª. Marcia Helena Costa Fampa, D.Sc.

Prof. Yuri Abitbol de Menezes Frota, D.Sc.

Prof. Thibaut Victor Gaston Vidal, D.Sc.

RIO DE JANEIRO, RJ – BRASIL
MAIO DE 2015

Venceslau, Marilis Bahr Karam

O Problema de recobrimento mínimo de um corpo em três dimensões por esferas de diferentes raios/Marilis Bahr Karam Venceslau. – Rio de Janeiro: UFRJ/COPPE, 2015. XIII, 136 p.: il.; 29,7cm.

Orientador: Nelson Maculan Filho

Tese (doutorado) – UFRJ/COPPE/Programa de Engenharia de Sistemas e Computação, 2015.

Referências Bibliográficas: p. 96 – 104.

1. Problema de Recobrimento. 2. Otimização Global. 3. Gamma Knife. I. Maculan Filho, Nelson. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Sistemas e Computação. III. Título.

À minha família.

Agradecimentos

Agradeço à UFRJ pela oportunidade de me permitir ampliar meus horizontes de estudo.

Aos professores do PESC/COPPE pelo incentivo e experiências compartilhadas. Em particular, aos professores Abílio Lucena e Adilson Xavier. As conversas com professor Adilson foram sempre proveitosas. Sou muito grata por seu apoio.

Aos meus colegas de turma e do Laboratório de Otimização - LabOtim: Renan, Rogério, Vinícius Layter e Daniela pela ajuda e companheirismo durante o tempo em que estivemos juntos.

À Fátima, secretária de OT, Solange, Sonia e Gutierrez da secretaria do PESC e Carol, secretária do LAND, pelo suporte técnico. A ajuda de vocês foi inestimável.

Ao Colégio Pedro II e à Chefe do Departamento de Matemática, por me terem concedido afastamento para estudos.

Aos colegas de trabalho do Campus São Cristóvão III pela ajuda e pelo estímulo para a conclusão deste trabalho, especialmente à Lúcia, Maria Helena e ao Marcos: serei eternamente grata por tudo o que vocês, meus amigos, fizeram por mim.

Aos familiares e amigos que estiveram ao meu lado, apoiando-me para eu prosseguir em minha pesquisa. Agradeço, de modo especial, à minha mãe, Marilene, e aos meus irmãos – Roselis, Daniëlle, Júnior e Guilherme.

Agradeço aos dois homens mais importantes da minha vida. Ao meu marido, Helder, que sempre me apoiou e cuja contribuição foi fundamental para a realização deste trabalho. Ao meu filho, Daniel, a quem, desde muito pequeno, incentivei a estudar e a nunca desistir e que, agora, homem feito, passou a ser meu incentivador.

Ao meu orientador, Professor Nelson Maculan, um pesquisador notável, uma pessoa extraordinária - pela generosidade, pela disposição em ajudar o próximo. Por tudo isso, Professor Maculan, o senhor se distingue tanto pelas suas realizações acadêmicas quanto pela sua forma de interagir no mundo. Obrigada por toda a disposição e paciência em me orientar. Foi um prazer ter sido sua aluna e orientanda.

Enfim, gostaria de agradecer a todos que me auxiliaram, direta ou indiretamente, para a conclusão deste trabalho.

Agradeço, sobretudo, a Deus, por sua graça de me conceder força na condução de mais esta etapa da minha vida.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

O PROBLEMA DE RECOBRIMENTO MÍNIMO DE UM CORPO EM TRÊS DIMENSÕES POR ESFERAS DE DIFERENTES RAIOS

Marilis Bahr Karam Venceslau

Maio/2015

Orientador: Nelson Maculan Filho

Programa: Engenharia de Sistemas e Computação

A radiocirurgia *Gamma Knife* representa um dos mais avançados meios disponíveis para o tratamento de tumores cerebrais, malformações vasculares e desordens funcionais. O equipamento *Gamma Knife* emite, com precisão, uma dose adequada de radiação ionizante (*shot*) à região do tumor alvo. Em boas aproximações, os shots podem ser modelados como esferas de diferentes tamanhos. Múltiplos shots podem ser usados para cobrir o tumor enquanto evita-se irradiar uma dose excessiva nos tecidos saudáveis ao seu redor. As características intrínsecas a esse problema possibilitam sua formulação como um problema de recobrimento no espaço euclidiano tridimensional \mathbb{R}^3 .

Neste trabalho consideraremos o problema de recobrimento de um corpo tridimensional por esferas de diferentes raios visando sua aplicação no planejamento de tratamentos de radiocirurgias *Gamma Knife*. Abordaremos o problema apenas do ponto de vista de uma cobertura geométrica, ou seja, dados um conjunto de esferas e um corpo, o objetivo é cobrir completamente o corpo utilizando o menor número possível de esferas, independentemente do problema da dosagem. A fim de resolver esse problema de programação matemática, adotaremos uma abordagem baseada na aplicação de técnicas de penalização e heurísticas estocásticas de busca.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

THE MINIMUM COVERING PROBLEM OF A THREE DIMENSIONAL BODY
USING DIFFERENT RADIUS SPHERES

Marilis Bahr Karam Venceslau

May/2015

Advisor: Nelson Maculan Filho

Department: Systems Engineering and Computer Science

Gamma Knife radiosurgery represents one of the most advanced means available for treating brain tumors, vascular malformations and functional disorders. Gamma Knife unit delivers accurately a suitable dose of ionizing radiation, called shot, to the target tumor region. In good approximation, these shots can be modeled as spheres of different sizes. Multiple shots can be used to cover the entire tumor, while avoiding an excessive dose to the surrounding healthy tissue. The intrinsic characteristics of this problem enable its formulation as a covering problem in the three-dimensional Euclidian space \mathbb{R}^3 .

In this work, we focus on the covering problem of a three dimensional body using different radius spheres, aiming its use on the treatment planning of Gamma Knife radiosurgery. We approach the problem just by the geometric covering point of view, that is, given a set of spheres and a body, the objective is to fully cover the body using the smallest possible number of spheres, regardless of the dosage issue. In order to solve this mathematical programming problem, we consider an approach based on the application of penalty and stochastic heuristic search techniques.

Sumário

Lista de Figuras	x
Lista de Tabelas	xiii
1 Introdução	1
2 O Problema de Recobrimento	5
2.1 A Radiocirurgia <i>Gamma Knife</i>	5
2.2 Definição do Problema de Recobrimento	9
3 Três problemas clássicos da Literatura	11
3.1 Algumas definições e resultados importantes	11
3.1.1 Algumas definições e resultados da Análise Convexa.	11
3.1.2 Reticulados	13
3.1.3 Reticulado cúbico de face centrada	19
3.1.4 Reticulado cúbico de corpo centrado	20
3.2 Um breve apanhado histórico sobre a Conjectura de Kepler	21
3.3 O Problema de Empacotamento de Esferas	23
3.4 O Problema do Kissing Number	26
3.5 O Problema de Recobrimento do espaço euclidiano n-dimensional por hiperesferas	27
4 Modelo proposto e discussão do método de resolução	31
4.1 O Modelo	32
4.2 Abordagem proposta para resolver (P1)	35
4.3 Reformulação do problema	37
4.3.1 Reformulação exata de (P1)	37
4.3.2 Reformulação de (P1) com penalização	38
4.3.3 Aprimoramento da função p_μ	39
4.4 Valores de α_{ij}	49
4.4.1 Esferas S_i e S_j possuem o mesmo raio	49
4.4.2 Esferas S_i e S_j possuem raios diferentes	52

4.5	Região de segurança	57
4.5.1	Corpo paralelo ao sólido T	59
4.5.2	Inclusão de esferas em T_ϵ	65
4.5.3	Suavização da função q	69
4.6	Árvore de decisão e índices de cobertura	73
4.6.1	Índices de cobertura	74
4.6.2	Árvores de decisão	79
5	Resultados computacionais	84
5.1	A implementação USCS	84
5.2	Instâncias de teste	86
5.3	Resultados	89
6	Conclusões e trabalhos futuros	93
6.1	Conclusões	93
6.2	Sugestões para trabalhos futuros	94
	Referências Bibliográficas	96
A	Curvas e superfícies paralelas	105
B	Sobre a distância de um ponto a uma elipse e a um elipsóide.	109
B.1	O envelope das normais a uma elipse	109
B.2	Determinação da distância de um ponto a uma elipse	111
B.3	Aplicação do método de Newton na resolução do problema de deter- minar a distância de um ponto a uma elipse	118
B.4	Distância aproximada à elipse	122
B.5	Distância de um ponto a um elipsóide	125
C	Índices de cobertura e o <i>Gamma Knife</i>	127
D	As técnicas de integração de Monte Carlo e quasi-Monte Carlo	129
D.1	Sequências de Halton	130
E	Exemplos extraídos do USCS	132
E.1	Tela de interfaceamento com o USCS	132
E.2	Sumário das estatísticas e índices	133
E.3	Histórico cronológico dos eventos de processamento	134

Lista de Figuras

3.1	(a) Reticulado bidimensional e (b) paralelepípedo fundamental determinado pelos vetores v_1 e v_2	14
3.2	(a) Reticulado cúbico, (b) reticulado cúbico de corpo centrado e (c) reticulado cúbico de face centrada.	14
3.3	As regiões fundamentais dos reticulados cúbico, cúbico de corpo centrado e cúbico de face centrada são, respectivamente, o cubo (a), o octaedro truncado (b) e o dodecaedro rômico (c).	15
3.4	(a) Reticulado bidimensional hexagonal e (b) os hexágonos regulares são as regiões de Voronoi para os pontos do reticulado em (a).	18
3.5	Para os pontos de \mathcal{A} indicados pelos círculos pequenos, as regiões de Delaunay são quadrados (A) e octógonos (B); as regiões de Voronoi são triângulos (como XYZ).	19
3.6	Região de Voronoi do reticulado cúbico de face centrada. Os pontos representam os centros do reticulado.	19
3.7	Dodecaedro rômico.	19
3.8	(a) Região de Voronoi (b) Partição do \mathbb{R}^3 por dodecaedros rômicos.	20
3.9	Região de Voronoi do reticulado cúbico de corpo centrado.	20
3.10	Octaedro truncado.	21
3.11	(a) Regiões de Voronoi e (b) Partição do \mathbb{R}^3 por octaedros truncados.	21
3.12	Empacotamento cúbico de face centrada.	22
3.13	Reticulado hexagonal com regiões de Voronoi transladadas e empacotamento de círculos associado.	25
3.14	Empacotamento de esferas. Os pontos rotulados com a são pontos do reticulado.	26
3.15	Empacotamento compacto hexagonal.	26
3.16	À esquerda, o reticulado compacto hexagonal e à direita, o reticulado cúbico de face centrada.	27
3.17	(a) Kissing Number em \mathbb{R} , (b) Kissing Number em \mathbb{R}^2 e (c) Kissing Number em \mathbb{R}^3	28

3.18	Recobrimento do plano por círculos. Os centros pertencem ao reticulado \mathbb{Z}^2	28
3.19	Recobrimento do plano por círculos. Os centros pertencem ao reticulado hexagonal.	29
3.20	(a) Ladrilhamento do espaço por octaedros truncados, (b) O recobrimento reticulado de corpo centrado no \mathbb{R}^3 é obtido colocando-se esferas circunscritas aos octaedros truncados. Observe que cada esfera possui 14 outras esferas adjacentes a ela.	29
4.1	Representação dos parâmetros α_{ij} , representando a máxima sobreposição das esferas.	33
4.2	Gráfico da função $p_\mu(t)$	38
4.3	Gráfico da função $p_{\mu,\lambda}(t)$	40
4.4	Gráfico de uma versão da função $p_w(t)$	41
4.5	Gráfico da função $p_w(t)$ para $n = 1$ e $n = 4$	43
4.6	Gráfico da função $p_w(t)$ para valores distintos de E , com $n = 4$	43
4.7	Gráfico da função $p_w(t)$ com $n = 4$ para $k = 2$ e $k = 4$	44
4.8	Gráfico da função $p_v(t)$ com $n = 4$ para $k = 2$ e $k = 4$	46
4.9	Gráfico da função $p_u(t)$, comparada com a função $p_v(t)$ para $k = 4$, ambas com $n = 4$	48
4.10	Octaedro truncado.	50
4.11	Seção plana do octaedro truncado de lado l inscrito na esfera de raio R	51
4.12	A face hexagonal do octaedro truncado de lado l é uma seção plana de um cubo de lado $l\sqrt{2}$	51
4.13	Disposição hexagonal de sete círculos recobrimdo uma região do \mathbb{R}^2	53
4.14	A mesma disposição hexagonal da anterior, após aplicação da transformação conforme no \mathbb{R}^2	53
4.15	Representação da região de segurança para o sólido T , onde ε é o parâmetro de tolerância.	58
4.16	Exemplo onde $x \in T$ pertence a três superfícies paralelas distintas.	60
4.17	Exemplos onde T é convexo e $D(x)$ não é diferenciável em sub-regiões de T	64
4.18	Representação dos sólidos T_ε , T e $T_{\varepsilon-r_i}$ e uma esfera S_i posicionada de forma a tangenciar ∂T_ε	67
4.19	Exemplo de cones tangente CT_p e normal CN_p para o caso de T elipsóide.	72
4.20	Comportamento aproximado dos principais índices de cobertura.	77
4.21	Representação do conjunto S de esferas, para $M = 4$ e $O = 6$	80

4.22	Representação gráfica de uma árvore de esferas, para $M = 3$ e $O = 3$, organizada segundo a assinatura de cada nó, aqui representada por uma sequência das letras a , b e c	82
4.23	Busca p no conjunto S de esferas, para $M = 4$ e $O = 6$	83
5.1	Resultado do recobrimento do sólido com o formato de um cubo.	90
5.2	Resultado do recobrimento do Elipsóide prolato.	91
5.3	Resultado do recobrimento do Elipsóide oblato.	92
A.1	Um retângulo e uma curva paralela.	106
A.2	Uma elipse com $a = 3$ e $b = 2$ e algumas de suas curvas paralelas.	106
A.3	Um elipsóide com $a = 8$, $b = 6$ e $c = 4$ e três de suas superfícies paralelas internas.	108
A.4	Superfícies paralelas degeneradas a uma distância de deslocamento d do elipsóide, com $a = 8$, $b = 6$ e $c = 4$	108
B.1	Uma elipse acompanhada por uma reta normal passando pelo ponto N e por sua evoluta, a astróide.	111
B.2	Ponto P exterior à astróide e as duas normais à elipse passando pelos pontos P_1 e P_2	115
B.3	Ponto P sobre a astróide e as três normais à elipse passando pelos pontos P_1 , P_2 e P_3	116
B.4	Ponto P interior à astróide e as quatro normais à elipse passando pelos pontos P_1 , P_2 , P_3 e P_4	116
B.5	Retas tangentes à astróide e à parábola passando pelo ponto P . As interseções com a elipse são os pontos N e M , respectivamente.	120
B.6	Uma elipse, a aproximação parabólica da astróide no quarto quadrante, e as curvas exata (traço fino) e aproximada (traço grosso) dos pontos com distância à elipse igual a 1 unidade.	124
B.7	Elipsóides de revolução.	125
E.1	Tela de interface com o USCS.	132

Lista de Tabelas

5.1	Tabela com os tipos de esferas disponíveis e suas características . . .	87
5.2	Tabela com as dimensões dos sólidos utilizados	87
5.3	Tabela com os resultados dos recobrimentos dos sólidos	89
B.1	Tabela de comparação das distâncias exatas e aproximadas de um ponto P à elipse.	123
D.1	Os primeiros 10 termos de uma sequência de Halton tridimensional utilizando as bases 2, 3 e 5.	131

Capítulo 1

Introdução

Problemas de empacotamento e recobrimento de regiões do espaço euclidiano por figuras regulares são clássicos na literatura e seus aspectos teóricos são estudados tanto nas áreas da geometria discreta e combinatória como na área da otimização matemática.

A teoria de empacotamento e recobrimento constitui o mais antigo tópico da geometria discreta e computacional. Teve origem na teoria dos números e cristalografia, na virada do século XX, e tornou-se uma teoria independente na década de 1950. Foi, então, que o termo geometria discreta foi introduzido como sinônimo para problemas concernentes a empacotamento, recobrimento e seu caso especial: partições (*tesselation*). A partir do final da década de 1980, passou-se a usar o termo geometria discreta e computacional a qual inclui, além da geometria computacional, não somente problemas concernentes a empacotamentos, recobrimentos e partições, mas também a teoria dos politopos, bem como geometria combinatória ambas no sentido de Hadwiger e Debrunner, e Erdős. A nova terminologia reflete novas tendências: além da teoria clássica de densidades de empacotamento e recobrimento, houve um aumento em pesquisas sobre empacotamento e recobrimento influenciadas pela combinatória e ciência da computação (FEJES TÓTH, 1999).

Os empacotamentos densos de objetos sólidos onde não há sobreposição são muito comuns em situações que ocorrem na natureza ou criadas pelo homem. Os problemas de empacotamento surgem em contextos tecnológicos, tais como: indústrias de embalagens; agricultura, por exemplo, armazenamento de grãos em silos; propelentes sólidos de foguetes; terapia de radiação estereotáxica (SOUTOU e DAI, 2002; WANG, 1999) e dão suporte a estrutura de inúmeros sistemas biológicos, por exemplo, estrutura de tecidos, membranas celulares e filotaxia¹ (TORQUATO e JIAO, 2010). Já os problemas de recobrimento têm aplicações práticas como, por exemplo, em localização e dimensionamento de centros de telecomunicação (XA-

¹Disposição das folhas nos eixos caulinares.

VIER *et al.*, 1989); na topologia de rede de comunicação sem fio (ADICKES *et al.*, 2002); buscas em espaços paramétricos de altas dimensões, por exemplo, buscas por ondas gravitacionais baseadas em padrões (PRIX, 2007) e terapia de radiação estereotáxica (LIBERTI *et al.*, 2009; LUBKE, 2014; NASCIMENTO *et al.*, 2010; PINTO, 2015).

Por outro lado, um grande número de modelos matemáticos em diversas áreas das ciências e engenharias tem levado à formulação de problemas de otimização onde a melhor solução (ótimo global) é necessária. Devido à natureza interdisciplinar da otimização global, surpreendentes progressos foram realizados neste campo nas últimas décadas. Muitos algoritmos computacionais poderosos e novos desenvolvimentos teóricos foram apresentados para resolver um espectro de problemas difíceis em muitas disciplinas.

O estudo da matemática discreta e otimização com aplicações médicas emergiu como uma nova importante área de pesquisa (DU *et al.*, 2000). Nos últimos anos, inúmeras aplicações médicas receberam o auxílio de técnicas de otimização. Dentre essas, a *terapia por radiação* é indiscutivelmente a área mais popular em planejamento de tratamentos onde técnicas de otimização são utilizadas (PARDALOS e EDWIN, 2009).

Terapia por radiação é um método usado para tratar câncer com radiação de alta energia que causa danos ao DNA das células na área a ser tratada, interferindo com suas habilidades de se reproduzirem e crescerem. As células cancerígenas são incapazes de reparar esses danos tão rapidamente, assim o crescimento dessas células é interrompido e o tumor diminui de tamanho. Dentre os tipos de *terapia por radiação*, temos a radiocirurgia estereotáxica cuja modalidade mais efetiva é a radiocirurgia conhecida como *Gamma Knife* (GANZ, 1997). Nesse procedimento, tumores ou malformações vasculares em regiões profundas do cérebro são tratados mediante a aplicação de feixes de radiação. Mais especificamente, o equipamento *Gamma Knife* emite, com precisão, uma dose adequada de radiação ionizante (*shot*) à região do tumor alvo. Os *shots* podem ser modelados aproximadamente como esferas de diferentes tamanhos. Dependendo do tamanho do tumor, múltiplos *shots* podem ser usados para cobrir a região. O objetivo é cobrir o tumor o máximo possível e, simultaneamente, minimizar ou evitar o volume de tecidos saudáveis atingidos pelos raios.

Vários autores tratam do problema de planejamento do tratamento da radiocirurgia *Gamma Knife* sob o ponto de vista de um problema de empacotamento, não permitindo que as esferas se sobreponham. Dessa forma, o problema é descrito como um modelo matemático de empacotamento de esferas de diâmetros diferentes em regiões tridimensionais, almejando uma densidade de empacotamento superior a 0,9.

A maior parte da literatura existente sobre o uso de técnicas de otimização e modelos matemáticos para automatizar o planejamento de tratamento de radiocirurgias *via Gamma Knife* visam fornecer uma solução completa para o problema de planejamento do tratamento, ou seja, trata-se das dosagens tanto quanto das posições dos *shots*. JITPRAPAIKULSARN (2005) contém um resumo detalhado de algumas abordagens existentes.

Mais recentemente, alguns autores abordaram o problema de recobrimento de corpos tridimensionais por esferas de diferentes raios visando a aplicação no planejamento de tratamentos de radiocirurgias *Gamma Knife*. Dados um conjunto de esferas e um corpo, o objetivo é cobrir completamente a região utilizando o menor número possível de esferas independentemente do problema da dosagem.

LIBERTI *et al.* (2009) tratam do problema sob o ponto de vista de recobrir um elipsóide (o tumor) por esferas (*shots*). Eles propuseram um modelo MINLP de programação matemática, cuja relaxação contínua é não-convexa. A partir dessa formulação, foi feita uma discretização do sólido a ser recoberto e, com o emprego de técnicas de linearização de restrições que utilizam variáveis 0-1, conseguiram aproximar o modelo por uma formulação de programação linear 0-1. Soluções interessantes foram obtidas com o uso do CPLEX. No entanto, esse enfoque pode se tornar inviável em discretizações mais finas, devido à enorme quantidade de variáveis.

Seguindo esse caminho, uma outra abordagem para tratar do problema de recobrir um sólido por esferas de diferentes diâmetros foi proposta por PINTO (2015). O problema foi formulado como um problema de otimização não-convexa, com restrições quadráticas e função objetivo linear. Utilizou-se técnicas de discretização afim de trabalhar com um modelo linear. Foi proposta, ainda, uma heurística e uma reformulação baseada em uma estrutura de grafos, onde a clique de peso máximo é a solução ótima do modelo original, com a finalidade de encontrar boas soluções em tempos razoáveis.

A nossa proposta, neste trabalho, é tratar do problema de recobrimento de um corpo tridimensional por esferas de diferentes raios utilizando o mesmo modelo MINLP não-convexo de programação matemática proposto por PINTO (2015). Entretanto, para resolver o problema, adotaremos uma abordagem alternativa, muito empregada na resolução de problemas de otimização global: aplicação de técnicas heurísticas, misturando técnicas estocásticas de busca com técnicas de penalização. Desta forma, não estaremos, a princípio, garantindo a otimalidade de nossas soluções. No entanto, os métodos desenvolvidos irão permitir pelo menos a seleção da melhor candidata à solução do problema original dentre um conjunto de possíveis soluções produzidas mediante a aplicação de heurísticas especialmente projetadas.

Iniciaremos, o Capítulo 2, fazendo uma breve descrição da modalidade de terapia por radiação *Gamma Knife*. O conhecimento dos requerimentos de protocolo

dessa aplicação irão ajudar-nos a compreender melhor as características do modelo proposto para o problema de recobrimento que será definido logo em seguida. No Capítulo 3, estudaremos um pouco os problemas de empacotamento, os de recobrimento dos espaços euclidianos n -dimensionais e o Kissing Number quando os centros das esferas idênticas participantes do recobrimento (ou empacotamento) formam um reticulado. Tal estudo se faz importante devido à necessidade de conhecermos os valores de alguns parâmetros presentes na formulação do problema e que são possíveis de serem determinados exatamente nesses reticulados. Apresentaremos, no Capítulo 4, o modelo proposto, as reformulações do problema, falaremos ainda sobre região de segurança, árvore de decisão e índices de cobertura, conceitos que serão empregados na resolução do problema. No Capítulo 5, comentaremos sobre os experimentos computacionais realizados e os resultados obtidos. Finalmente, no Capítulo 6, apresentaremos as conclusões e as propostas de trabalhos futuros. Alguns cálculos mais detalhados e noções básicas de tópicos necessários para o desenvolvimento deste trabalho serão apresentados nos apêndices.

Capítulo 2

O Problema de Recobrimento

Neste trabalho, abordaremos o problema de recobrimento de sólidos tridimensionais por esferas de diferentes raios tendo como uma aplicação o planejamento de tratamentos de radiocirurgias *Gamma Knife*. Dados um conjunto de esferas e um sólido, o objetivo é cobrir completamente a região utilizando o menor número possível de esferas sem considerar o problema da dosagem. A fim de obtermos uma compreensão melhor do problema e de sua modelagem, iniciaremos por fazer uma breve descrição desta modalidade de tratamento.

2.1 A Radiocirurgia *Gamma Knife*

A radiocirurgia *Gamma Knife* é uma ferramenta comprovadamente eficaz utilizada para tratar doenças neurológicas. É considerada superior à cirurgia invasiva, neurocirurgia e terapia de radiação, formas mais tradicionais de tratamento do câncer. Atualmente, mais de 500.000 pacientes foram tratados no mundo inteiro em 250 Centros *Gamma Knife* e cerca de 50.000 pacientes são tratados a cada ano (GANZ, 2011).

O trabalho desenvolvido pelo neurocirurgião Lars Leksell do Instituto de Karolinska, em Estocolmo, Suécia, no começo dos anos da década de 1950, em conjunto com o biofísico Borje Larsson, da Universidade Upsalla, deu origem ao aparelho de tratamento radiocirúrgico *Gamma Knife*. Enquanto pesquisavam por uma maneira não invasiva para tratar distúrbios funcionais, observaram que uma única dose de radiação podia ser usada para destruir quase completamente regiões de pequenas dimensões dentro do cérebro, como se fosse um bisturi, no entanto, sem a necessidade de abrir a caixa craniana (GANZ, 1997). A essa técnica chamaram de radiocirurgia estereotáxica (LEKSELL, 1951) .

Em 1968, a primeira unidade *Gamma Knife* foi construída (LEKSELL, 1968) sendo designada especificamente para tratar de distúrbios funcionais. A partir de 1970, com o avanço de tecnologias, aplicações a outras doenças começaram a surgir.

Desde então, tem sido usado de forma gradual no tratamento de tumores cerebrais (especialmente aqueles de tamanho pequeno), malformações vasculares e distúrbios de dores na cabeça (GANZ, 1997). Atualmente, o *Gamma Knife* é utilizado, com êxito, nos seguintes tratamentos :

- tumores cerebrais malignos e benignos: metástases (originárias de outros locais do organismo com disseminação cerebral), neurinoma de nervo acústico (normalmente benigno e que pode gerar déficit auditivo), meningioma (na meninge, membrana que envolve o cérebro), hipófise (benigno, mas pode acarretar distúrbios hormonais e visuais) e gliomas (são tumores primários do cérebro);
- distúrbios vasculares: má-formação arteriovenosa cerebral (doença que pode causar sangramento e epilepsia), cavernomas e fistulas durais;
- distúrbios funcionais: neuralgia do trigêmeo (dor facial), epilepsia, Mal de Parkinson.

Além disso, estudos experimentais estão sendo realizados para avaliar a eficácia do emprego da técnica no tratamento do Transtorno Obsessivo Compulsivo (TOC) em estado grave.

O equipamento *Gamma Knife* irradia uma dose de radiação ionizante emitida de 201 fontes de Cobalto-60. Os 201 feixes irradiados são focados em um único ponto no espaço (denominados de *shot* de radiação), afetando uma região aproximadamente esférica com uma elevada dose de radiação. Isso ocorre devido à dose de tratamento prescrita, nas aplicações clínicas, ser de 50% *isodose line* (IDL), curva que envolve todos os pontos que recebem no mínimo 50% da dose máxima que é irradiada, e a distribuição da dose de um *shot* a 50% IDL ou mais ser aproximadamente uma esfera sólida no espaço tridimensional (WU, 1992).

Consideremos o tumor como uma região tridimensional. Sejam T o tumor, V_T o volume do tumor e V_{IDL} o volume do tumor coberto por 50% IDL. Então, V_{IDL} é chamado de *volume de empacotamento* e a razão $\frac{V_{IDL}}{V_T}$ é chamada de *densidade de empacotamento*. Em aplicações clínicas, a *densidade de empacotamento* requerida é maior do que 0,9 (SHAW, 1993; WU, 1996).

Um tratamento típico consiste em aplicar *shots*, possivelmente de tamanhos diferentes, centrados em diferentes pontos do tumor, tendo por objetivo irradiar uma determinada dose de radiação ao volume de tratamento a fim de obter a densidade requerida, enquanto o efeito nos tecidos saudáveis ao redor é minimizado. Sub-regiões do tumor não cobertas recebem o somatório das doses mais baixas (<50%) dos múltiplos *shots* que as circundam. Uma densidade de empacotamento alta é necessária para garantir que o somatório das doses recebidas por essas sub-regiões não cobertas seja suficientemente forte (WANG, 1999).

A radiocirurgia *Gamma Knife* inicia-se pela determinação da localização e do tamanho do tumor. Para isso, após a administração de anestesia local, um quadro

rígido de referência chamado dispositivo estereotáxico é fixado no crânio do paciente mediante quatro apoios ajustáveis e parafusos de fixação. A fixação de tal quadro é absolutamente imprescindível. Além deste dispositivo fornecer um sistema de coordenadas geométricas para cada estrutura do cérebro, através do qual a localização do tumor será precisamente determinada, também serve para imobilizar a cabeça do paciente dentro de um capacete acoplado a ele durante o tratamento. Este capacete é responsável pela irradiação dos *shots* e serve como um colimador que reduz o tamanho de cada feixe. Há, no total, quatro capacetes que se diferenciam entre si pelo tamanho dos *shots* que irradiam. Assim, como cada *shot* pode ser modelado como uma esfera, estão disponíveis quatro tamanhos de *shots* cujos raios medem 2 mm, 4 mm, 7 mm e 9 mm. A seguir, são feitas imagens de ressonância magnética ou tomografia computadorizada para determinar a posição do volume de tratamento em relação às coordenadas obtidas pelo dispositivo. Na etapa seguinte, um neurocirurgião, um oncologista e um físico trabalham juntos a fim de elaborar um bom plano de tratamento. Esse processo pode ser muito tedioso e demandar muito tempo.

A dificuldade em elaborar planos de tratamento varia substancialmente de um caso para outro. Em alguns casos, os tumores são pequenos o suficiente para serem tratados apenas com um *shot* de radiação e o planejamento estará pronto em poucos minutos. Por outro lado, quando o tumor é grande, ou tem um formato irregular ou está muito próximo a estruturas sensíveis, muitos *shots* de diferentes tamanhos podem ser necessários para se alcançar uma alta dose de radiação no volume de tratamento enquanto poupa-se os tecidos sadios ao redor. Nestes casos, o planejamento pode se tornar muito complexo.

Um plano de tratamento pode ser realizado combinando-se múltiplos *shots* de radiação a fim de tratar lesões de diferentes formas e tamanhos, sendo que o tamanho das lesões deve ser no máximo 3cm. Porém, esse procedimento geralmente resulta em um recobrimento não homogêneo, devido à sobreposição dos diferentes *shots*. Na prática, muitos casos são tratados com o uso de 1 a 15 *shots* de radiação (JITPRAPAIKULSARN, 2005; LIM, 2002) embora um número maior que 10 seja considerado impraticável pelas rotinas clínicas de tratamento. A equipe responsável pela elaboração do plano de tratamento deve determinar o número de *shots* de radiação que são necessários, os tamanhos dos *shots*, a localização e o peso que deve ser atribuído a cada um deles.

Durante o procedimento, a cada vez que há uma mudança no tamanho do *shot*, deve-se substituir o capacete em uso pelo capacete com o tamanho adequado. Como os capacetes são muito pesados, é necessário a utilização de máquinas para movê-los o que torna esta etapa bastante demorada. Além disso, ajustes na conexão entre o dispositivo estereotáxico fixado na cabeça do paciente e o capacete devem ser feitos após cada *shot*, pois a localização do centro do *shot* é determinada por esta conexão.

Na hora da irradiação dos raios gamas, a porta da unidade de tratamento se abre e a cama do paciente avança para dentro da abóboda blindada.

Devido a todo esse processo tornar-se bastante tedioso e demorado para muitos pacientes e a qualidade do plano de tratamento elaborado depender extremamente da experiência e paciência dos profissionais responsáveis, há o desejo de automatizar o processo de elaboração do plano de tratamento via *Gamma Knife*. Tendo por objetivo, determinar um bom recobrimento do volume do tumor e acelerar o processo de planejamento do tratamento.

Na elaboração de um plano de tratamento via *Gamma Knife*, além do requerimento da densidade, no mínimo 90% do tumor deve ser coberto, há vários outros requerimentos a serem atingidos: a radiação recebida pelos tecidos saudáveis ao redor do tumor deve ser a mínima possível; não deve haver sobreposição de *shots*, pois os níveis de radiação através do tumor devem ser uniformes tanto quanto possível; e deve-se usar o menor número possível de *shots*. Contudo, esses últimos objetivos de otimização são competitivos, então, um certo compromisso é necessário. Geralmente, o objetivo de minimizar o número de *shots* é considerado como de maior prioridade (WANG, 1999).

Muitos autores atacam o problema de planejamento do tratamento da radiocirurgia *Gamma Knife* sob o ponto de vista de um problema de empacotamento, não permitindo que as esferas se sobreponham. Dessa forma, o problema é descrito como um modelo matemático de empacotamento de esferas de diâmetros diferentes em regiões tridimensionais, almejando uma densidade de empacotamento superior a 0,9. O empacotamento de esferas idênticas em um espaço euclidiano tridimensional ocupa no máximo 74% desse espaço. Como o tratamento em questão disponibiliza quatro diferentes tamanhos de esferas, uma densidade mais alta é possível de ser obtida, de acordo com os requerimentos do tratamento. Por exemplo, SOUTOU e DAI (2002) tratam do problema de empacotar esferas (*shots*) de diferentes diâmetros em um polítopo tridimensional (tumor) de modo a agrupá-las sem que haja sobreposição das mesmas e o volume de empacotamento (soma dos volumes das esferas empacotadas) seja maximizado. O problema é formulado como um problema de otimização não-convexa com restrições quadráticas e função objetivo linear. A fim de resolver o problema, métodos de resolução baseados em técnica de *branch and bound* são apresentados.

A literatura existente sobre o uso de técnicas de otimização e modelos matemáticos para automatizar o planejamento de tratamento de radiocirurgias via *Gamma Knife* é mais abrangente do que o problema de recobrimento, objeto de nosso estudo, e visam fornecer uma solução completa para o problema de planejamento do tratamento, ou seja, trata-se das dosagens tanto quanto das posições dos *shots*. JIT-PRAPAIKULSARN (2005) contém um resumo detalhado de algumas abordagens

existentes.

Entretanto, nesse trabalho, desejamos não apenas um recobrimento parcial, mas uma cobertura total da região. Para esse feito, a restrição de não-sobreposição das esferas é relaxada, sendo permitido que esferas distintas possuam um volume de interseção. Porém, essa situação deve ser evitada, já que a sobreposição das esferas indica uma aplicação dobrada de radiação na região compreendida pela interseção, enquanto que, idealmente, o tumor deva ser recoberto de forma homogênea. Temos, então, um problema que recai na classe de problemas de recobrimento, com o objetivo de utilizar o menor número possível de esferas.

2.2 Definição do Problema de Recobrimento

O problema de recobrimento de interesse no tratamento por *Gamma Knife* pode ser descrito matematicamente como:

(*P*): Seja $B(x, \rho)$ a esfera de centro em x e raio ρ . Dados um conjunto compacto $T \subset \mathbb{R}^3$, um conjunto finito $R \subset \mathbb{R}_+$ de raios, um conjunto N indexando as esferas e uma função $\rho : N \rightarrow R$, encontrar um conjunto de esferas $\{B(x(i), \rho(i)) \mid i \in N\}$ de cardinalidade mínima recobrendo cada ponto de T .

O problema (*P*), como definido acima, pode ser formulado como um problema de programação matemática, com o objetivo de encontrar os centros de cada esfera e associar a cada uma delas um respectivo raio, através da função ρ , de forma que a união dessas esferas forme uma cobertura do conjunto T .

LIBERTI *et al.* (2009) propuseram um modelo MINLP de programação matemática para abordar o problema sob o ponto de vista de recobrir um elipsóide por esferas, sendo a relaxação contínua desse modelo MINLP não-convexa. Partindo dessa formulação, fez-se uma discretização do sólido a ser recoberto e, com o emprego de técnicas de linearização de restrições que utilizam variáveis 0-1, conseguiu-se aproximar o modelo por uma formulação de programação linear 0-1. Com o uso do CPLEX, foram obtidas soluções interessantes. No entanto, em discretizações mais finas, esse enfoque pode se tornar inviável, devido à quantidade de variáveis aumentar enormemente.

Uma outra abordagem encontrada na literatura é a formulação como um problema de programação geométrica signomial. Utilizando essa técnica, NASCIMENTO *et al.* (2010) apresentam uma metodologia de solução do problema de recobrimento de um elipsóide que não se baseia em técnicas de otimização discreta, tratando até mesmo os raios das esferas de forma contínua.

LUBKE (2014) trata do problema de recobrir corpos tridimensionais utilizando a Suavização Hiperbólica, técnica esta desenvolvida pelo professor Adilson Elias

Xavier (XAVIER *et al.*, 1989). Nesta abordagem, a modelagem matemática deste problema apresenta uma formulação *min-max-min* com natureza multi-nível intrínseca e é não diferenciável. A estratégia de suavização é desenvolvida utilizando uma classe especial C^∞ de função suavizada. A solução final é obtida através da resolução de uma sequência de subproblemas que gradualmente se aproximam do problema original. A técnica utilizada permite que as principais dificuldades apresentadas no problema original sejam superadas e gradualmente o problema se aproxime do original.

A abordagem mais recente para tratar desse problema como um problema de recobrimento foi proposta por PINTO (2015). O problema foi formulado como um problema de otimização não-convexa, com restrições quadráticas e função objetivo linear. Utilizou-se técnicas de discretização a fim de trabalhar com um modelo linear. Foi proposta, ainda, uma heurística e uma reformulação baseada em uma estrutura de grafos, onde a clique de peso máximo é a solução ótima do modelo original, com a finalidade de encontrar boas soluções em tempos razoáveis.

Capítulo 3

Três problemas clássicos da Literatura

Alguns problemas estão direta ou indiretamente relacionados com o problema de recobrimento por esferas. É o caso do problema de empacotamento de esferas em um container, ou no espaço euclidiano tridimensional, sem que haja sobreposição e cuja densidade seja máxima. No caso do empacotamento de esferas idênticas no \mathbb{R}^3 , seis esferas ficam dispostas em torno de uma esfera central. Reçamos, então, no Problema do Kissing Number. Devido ao relacionamento entre esses problemas e a natureza do problema tratado, faremos uma breve exposição desses assuntos. Assim, vamos iniciar esta seção enunciando alguns conceitos e resultados importantes que serão utilizados ao longo deste trabalho. Logo após falaremos um pouco sobre a famosa Conjectura de Kepler para, em seguida, tratarmos do problema de empacotamento em reticulados e do Problema do Kissing Number. Finalmente, trataremos do Problema de Recobrimento em reticulados.

3.1 Algumas definições e resultados importantes

Nesta seção, enunciaremos algumas definições da Análise Convexa e faremos uma breve apresentação sobre reticulados e suas regiões de Voronoi. Citaremos apenas os conceitos, notações e resultados necessários que serão utilizados nos capítulos seguintes.

3.1.1 Algumas definições e resultados da Análise Convexa.

Seja n um número natural. Um ponto no \mathbb{R}^n é uma n -upla de números reais $x = (x_1, x_2, x_3, \dots, x_n)$.

O *espaço euclidiano n -dimensional* \mathbb{R}^n é um *espaço vetorial real*, munido do

produto interno $\langle \cdot, \cdot \rangle : \mathbb{R}^n \rightarrow \mathbb{R}$ definido por

$$\langle x, y \rangle = \sum_{i=1}^n x_i y_i, \quad x, y \in \mathbb{R}^n. \quad (3.1)$$

O \mathbb{R}^n é um *espaço normado* cuja norma usual é a *norma euclidiana*

$$\|x\| = \sqrt{\sum_{i=1}^n x_i^2}, \quad x \in \mathbb{R}^n. \quad (3.2)$$

Esta norma dá origem à noção de *distância* $d(x, y)$ entre dois pontos x e $y \in \mathbb{R}^n$ definida por

$$d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}. \quad (3.3)$$

Dados o ponto $c \in \mathbb{R}^n$ e o número real $r > 0$, denotamos a *bola fechada n -dimensional* de centro c e raio r por:

$$B^n(c, r) = \{x \in \mathbb{R}^n : \|x - c\| \leq r\} \quad (3.4)$$

e a bola unitária *n -dimensional* de centro na origem será denotada por:

$$B^n = \{x \in \mathbb{R}^n : \|x\| \leq 1\}. \quad (3.5)$$

Chamaremos, daqui por diante, como é usual na literatura sobre empacotamentos e recobrimentos, as bolas fechadas (bolas sólidas) de *esferas*.

O *volume* de uma *hiperesfera n -dimensional* de raio r é dado pela fórmula:

$$V_n = \frac{\sqrt{\pi^n}}{(n/2)!} \cdot r^n, \quad \text{se } n \text{ for par,} \quad (3.6)$$

ou

$$V_n = \frac{2^n \sqrt{\pi^{(n-1)/2}} ((n-1)/2)!}{n!} \cdot r^n, \quad \text{se } n \text{ for ímpar.}$$

Um conjunto $X \subset \mathbb{R}^n$ é um *conjunto convexo* se, e somente se, dados x e $y \in \mathbb{R}^n$ e $\lambda \in \mathbb{R}$, $0 \leq \lambda \leq 1$, temos $(1 - \lambda)x + \lambda y \in X$. Um conjunto $C \subset \mathbb{R}^n$ é um *corpo convexo* se é convexo, compacto e o interior é não vazio. Dizemos que *não há sobreposição* entre dois corpos se eles não tem pontos interiores em comum. E, se não há sobreposição entre eles, mas há pontos comuns na fronteira, dizemos que eles se *tocam* (BRASS *et al.*, 2005).

Seja $C \subset \mathbb{R}^n$ um conjunto fechado e não-vazio. A *função distância a um corpo*

convexo é definida, para todo $x \in \mathbb{R}^n$, por:

$$d_C(x) := d(x, C) := \inf \{|x-y| \mid y \in C\}.$$

É fácil ver que o ínfimo é um mínimo, ou seja, existe ao menos um ponto $y \in C$ de menor distância de x .

Enunciaremos agora um teorema que nos será muito útil:

Teorema 3.1. (Teorema de Motzkin). *Seja $C \subset \mathbb{R}^n$ um conjunto fechado e não-vazio. As seguintes afirmações são equivalentes:*

- (i) C é convexo;
- (ii) para todo ponto $x \in C$, existe um único ponto mais próximo a x em C ;
- (iii) a função $d(x, C)$ é diferenciável para todo $x \in \mathbb{R}^n \setminus C$.

3.1.2 Reticulados

Os *reticulados* ou *arranjos regulares* têm muitas aplicações importantes em matemática pura, particularmente em conexão com álgebras de Lie, teoria dos números e teoria dos grupos. Aparecem também na matemática aplicada em conexão com a teoria de códigos, em criptografia e são usados de várias maneiras nas ciências. Por exemplo, em ciências dos materiais e física do estado sólido, um reticulado é um sinônimo para o *framework* de uma estrutura cristalina, um arranjo tridimensional de pontos regularmente espaçados coincidindo com as posições do átomo ou da molécula em um cristal.

Diz-se que um subconjunto Λ do espaço euclidiano \mathbb{R}^n é um *reticulado n -dimensional* em \mathbb{R}^n se existem vetores linearmente independentes v_1, v_2, \dots, v_n tal que

$$\Lambda = \left\{ u : u = \sum_{i=1}^n \alpha_i v_i, \alpha_i \in \mathbb{Z} \right\}.$$

O conjunto de vetores $\{v_1, v_2, \dots, v_n\}$ é chamado de *base do reticulado* Λ .

Observe que um *reticulado* Λ no \mathbb{R}^n é um subgrupo aditivo discreto cuja base gera o espaço vetorial \mathbb{R}^n . Em particular, a origem $O = (0, 0, \dots, 0)$ pertence ao reticulado e a soma e a diferença de quaisquer dois vetores do reticulado são também vetores do reticulado (Figura 3.1a).

O paralelepípedo

$$\mathcal{P} = \left\{ w : w = \sum_{i=1}^n \theta_i v_i : 0 \leq \theta_i < 1 \right\}$$

é chamado *paralelepípedo fundamental* que é um exemplo de uma *região fundamental* para o reticulado (Figura 3.1b), isto é, a justaposição de várias cópias dessa região

preenche o espaço inteiro com um único ponto do *reticulado* em cada cópia. Em geometria, o preenchimento do \mathbb{R}^n com cópias de um ou mais polítopos n -dimensionais sem que haja sobreposições ou espaços entre eles é chamado de *ladrilhamento* ou uma *tesselação* (*tiling* ou *tessellation*). CONWAY *et al.* (2011) definem o termo ladrilhamento ou tesselação de forma mais abrangente que inclui os arranjos não periódicos:

Definição 3.2. Um *ladrilhamento* ou uma *tesselação* é uma partição do espaço euclidiano \mathbb{R}^n em regiões fechadas cujos interiores são disjuntos.

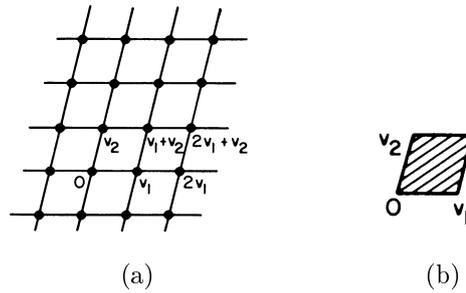


Figura 3.1: (a) Reticulado bidimensional e (b) paralelogramo fundamental determinado pelos vetores v_1 e v_2 .

No \mathbb{R}^3 , por exemplo, temos o reticulado cúbico (os pontos do reticulado estão localizados nos vértices de um cubo), o reticulado cúbico de corpo centrado (os pontos do reticulado estão localizados nos vértices e no centro de um cubo), e o reticulado cúbico de face centrada (os pontos do reticulado estão localizados nos vértices e no centro das faces de um cubo), conforme Figura 3.2.

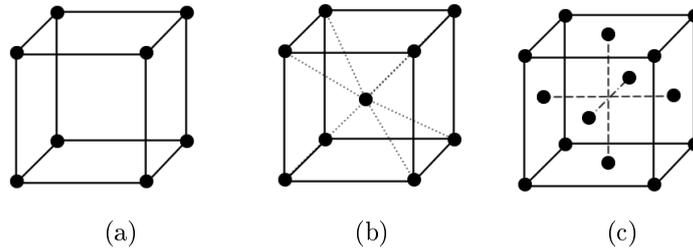


Figura 3.2: (a) Reticulado cúbico, (b) reticulado cúbico de corpo centrado e (c) reticulado cúbico de face centrada.

A Figura 3.3, mostra o cubo, o octaedro truncado e o dodecaedro rômbo - regiões fundamentais do reticulado cúbico, do reticulado cúbico de corpo centrado e do reticulado cúbico de face centrada, respectivamente.

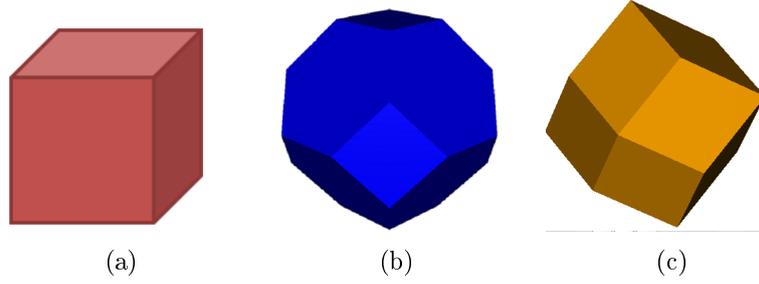


Figura 3.3: As regiões fundamentais dos reticulados cúbico, cúbico de corpo centrado e cúbico de face centrada são, respectivamente, o cubo (a), o octaedro truncado (b) e o dodecaedro rômico (c).

Embora haja muitas maneiras de escolher uma base e uma região fundamental para um *reticulado* Λ , o volume da região fundamental é unicamente determinado por Λ e o quadrado desse volume é chamado de *determinante* ou *discriminante* do reticulado. De fato, seja $\{v_1, v_2, \dots, v_n\}$ uma base de Λ . Se $v_i = (v_{i1}, v_{i2}, \dots, v_{im})$, para $i = 1, \dots, n$ onde $m \geq n$ (às vezes, é conveniente usar $m > n$ coordenadas para descrever um reticulado n -dimensional). A matriz é a *matriz geradora* do reticulado, e os vetores do reticulado consistem de todos os vetores

$$\xi M, \tag{3.7}$$

onde $\xi = (\xi_1, \xi_2, \xi_3, \dots, \xi_n)$ é um vetor arbitrário com componentes inteiras ξ_i . Assim, o reticulado Λ pode ser descrito como

$$\Lambda = \{\xi M : \xi \in \mathbb{Z}^n\}.$$

A matriz

$$A = MM^t, \tag{3.8}$$

onde M^t denota a matriz transposta da matriz M , é chamada *matriz de Gram* do reticulado Λ .

Observe que a matriz A pode ser escrita como

$$A = (\langle v_i, v_j \rangle)_{1 \leq i, j \leq n}. \tag{3.9}$$

Esta matriz é definida positiva. Então, o determinante do reticulado Λ é o determinante da *matriz de Gram* A :

$$\det \Lambda = \det A. \tag{3.10}$$

Se M é uma matriz quadrada

$$\det \Lambda = (\det M)^2. \tag{3.11}$$

Conseqüentemente, o volume do paralelepípedo fundamental \mathcal{P} é $\sqrt{\det\Lambda}$.

Como exemplo, considere o reticulado hexagonal bidimensional (Figura 3.4a). Uma matriz geradora óbvia é

$$M = \begin{pmatrix} 1 & 0 \\ \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix} \quad (3.12)$$

e a matriz de Gram para este reticulado é

$$A = MM^t = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}, \quad (3.13)$$

com $\det\Lambda = \det A = \frac{3}{4}$. Porém, para muitas finalidades é melhor usar a matriz geradora

$$M' = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix}, \quad (3.14)$$

tendo por matriz de Gram associada

$$A' = M'M'^t = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}, \quad (3.15)$$

e agora $\det\Lambda = \det A = 3$.

A matriz em (3.14) usa três coordenadas para definir um reticulado bidimensional, contudo todos os pontos estão no plano $x + y + z = 0$. As matrizes obtidas em (3.12) e (3.14) descrevem o reticulado hexagonal da Figura 3.4a, mas em diferentes coordenadas e em uma escala diferente. As vantagens da matriz em (3.14) sobre (3.12) é que o número irracional $\sqrt{3}$ desaparece das coordenadas e as simetrias do reticulado são mais fáceis de serem vistas. Da igualdade (3.14) temos que as três coordenadas podem ser permutadas sem mudar o reticulado. Formalmente, os reticulados definidos pelas matrizes em (3.12) e (3.14) são chamados de *reticulados equivalentes*.

Dois reticulados são *equivalentes* ou *similares* se um dos reticulados puder ser obtido do outro por rotação, reflexão e/ou mudança de escala, e denotaremos esta relação por \cong . Duas matrizes geradoras, M e M' , definem reticulados equivalentes se, e somente se, são relacionadas por $M' = cUMB$, onde c é uma constante não nula, U é uma matriz com elementos inteiros e determinante ± 1 , e B é uma matriz ortogonal real (com $B \cdot B^t = I$). As matrizes de Gram correspondentes se relacionam por $A' = c^2UAU^t$.

Se $\det(U) = \pm 1$ e $c = 1$, então M e M' são *reticulados congruentes*. Por exemplo, o reticulado cúbico de face centrada ocorre em ambas as sequências A_n e D_n , já que

$A_3 \cong D_3$.

Alguns tipos importantes de reticulados:

$$\mathbb{Z}^n = \{(x_1, \dots, x_n) : x_i \in \mathbb{Z}\}$$

$$A_n = \{(x_0, x_1, \dots, x_n) \in \mathbb{Z}^{n+1} : x_0 + x_1 + \dots + x_n = 0\}$$

$$D_n = \{(x_1, \dots, x_n) \in \mathbb{Z}^n : x_1 + \dots + x_n \text{ é par}\}$$

$$E_8 = \left\{ (x_1, \dots, x_8) : \forall x_i \in \mathbb{Z} \text{ ou } \forall x_i \in \frac{1}{2} + \mathbb{Z}, x_1 + \dots + x_8 \text{ é par} \right\}$$

$$E_7 = \{(x_1, x_2, \dots, x_8) \in E^8 : x_1 + \dots + x_8 = 0\}$$

$$E_6 = \{(x_1, x_2, \dots, x_8) \in E^8 : x_1 + x_8 = x_2 + \dots + x_7 = 0\}$$

Todo reticulado $\Lambda \subset \mathbb{R}^n$ tem um *reticulado dual* Λ^* , definido por :

$$\Lambda^* = \{x \in \mathbb{R}^n : \langle x, y \rangle \in \mathbb{Z}, \forall y \in \Lambda\}.$$

Seja $\mathcal{A} = \{p_1, p_2, \dots\}$ um conjunto discreto de pontos no \mathbb{R}^n . Chamamos de *raio de recobrimento* de \mathcal{A} a menor cota superior para a distância de qualquer ponto pertencente ao \mathbb{R}^n ao ponto mais próximo p_i , $p_i \in \mathcal{A}$, geralmente denotada por R . Assim:

$$R = \sup_{x \in \mathbb{R}^n} \inf_{p \in \mathcal{A}} d(x, p). \quad (3.16)$$

Se a cota superior não existe, dizemos que $R = \infty$. Então, as esferas de raio R centradas nos pontos de \mathcal{A} recobrirão o \mathbb{R}^n , e nenhum outro raio menor do que ele o fará.

A *região de Voronoi* de p_i , $p_i \in \mathcal{A}$, é o conjunto de pontos $x \in \mathbb{R}^n$ que são mais próximos de p_i do que de qualquer outro p_j . Assim,

$$\mathcal{V}(p_i) = \{x \in \mathbb{R}^n : d(x, p_i) \leq d(x, p_j), \forall j \neq i\}.$$

Por exemplo, as regiões de Voronoi do reticulado bidimensional hexagonal são hexágonos regulares (Figura 3.4).

Os interiores das regiões de Voronoi são disjuntos, embora tenham faces em comum. Cada face está contida em um hiperplano que é equidistante de dois pontos

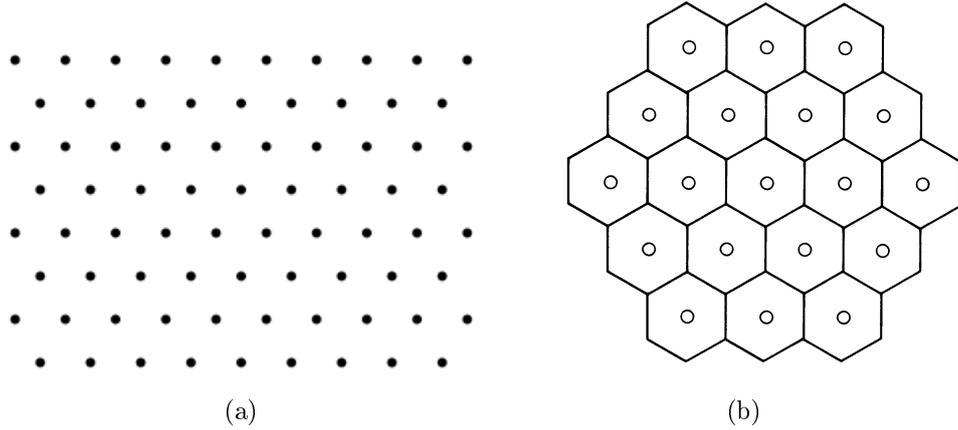


Figura 3.4: (a) Reticulado bidimensional hexagonal e (b) os hexágonos regulares são as regiões de Voronoi para os pontos do reticulado em (a).

vizinhos p_i e p_j , chamado *plano bissetor* de p_i e p_j e representado por

$$\mathcal{B}(p_i, p_j) = \{x \in \mathbb{R}^n : d(x, p_i) = d(x, p_j)\}, \forall i \neq j.$$

$\mathcal{B}(p_i, p_j)$ é perpendicular ao segmento de extremos p_i e p_j e passa pelo seu ponto médio. As regiões de Voronoi são politopos convexos cuja união é todo o \mathbb{R}^n . Se \mathcal{A} é um *reticulado* Λ , então, as regiões de Voronoi são congruentes e tem volume igual a $(\det \Lambda)^{1/2}$.

Os vértices das regiões de Voronoi são especialmente interessantes. Eles incluem os pontos do \mathbb{R}^n cuja distância de \mathcal{A} é um máximo local. Se existir um ponto cuja distância de \mathcal{A} é um máximo absoluto, então este ponto é chamado máximo profundo.

Para cada ponto que é um vértice de uma região de Voronoi existe uma região de Delaunay, a qual é o politopo determinado pelo fecho convexo dos pontos de \mathcal{A} mais próximos daquele ponto.

As regiões de Delaunay formam uma partição do \mathbb{R}^n em regiões convexas que é um tipo de dual para a partição do \mathbb{R}^n em regiões de Voronoi. Como exemplo, considere o conjunto de pontos \mathcal{A} indicados pelos pequeninos círculos na Figura 3.5. As regiões de Voronoi são triângulos tais como o triângulo XYZ , e existem dois tipos de regiões de Delaunay: quadrados como o quadrado A e octógonos como o octógono B .

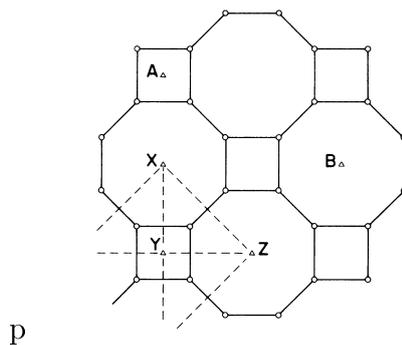


Figura 3.5: Para os pontos de \mathcal{A} indicados pelos círculos pequenos, as regiões de Delaunay são quadrados (A) e octógonos (B); as regiões de Voronoi são triângulos (como XYZ).

A seguir estudaremos as regiões de Voronoi de dois reticulados importantes.

3.1.3 Reticulado cúbico de face centrada

No reticulado cúbico de face centrada (Figura 3.2c) a região de Voronoi (Figura 3.6) é um dodecaedro rômbo, um dos poliedros de Catalan.

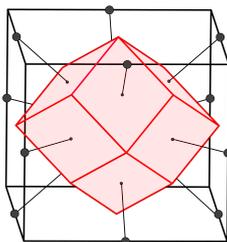


Figura 3.6: Região de Voronoi do reticulado cúbico de face centrada. Os pontos representam os centros do reticulado.

Um dodecaedro rômbo centrado na origem tem as coordenadas de 6 vértices permutando as coordenadas de $(\pm 1, 0, 0)$ e, de 8 vértices, as de $(\pm \frac{1}{2}, \pm \frac{1}{2}, \pm \frac{1}{2})$ (Figura 3.7). Os dodecaedros rômbo formam uma partição do \mathbb{R}^3 (Figura 3.8) e as regiões de Delaunay formam uma tesselação para o \mathbb{R}^3 alternando tetraedros e octaedros.

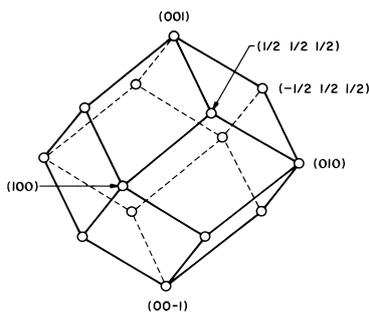


Figura 3.7: Dodecaedro rômbo.

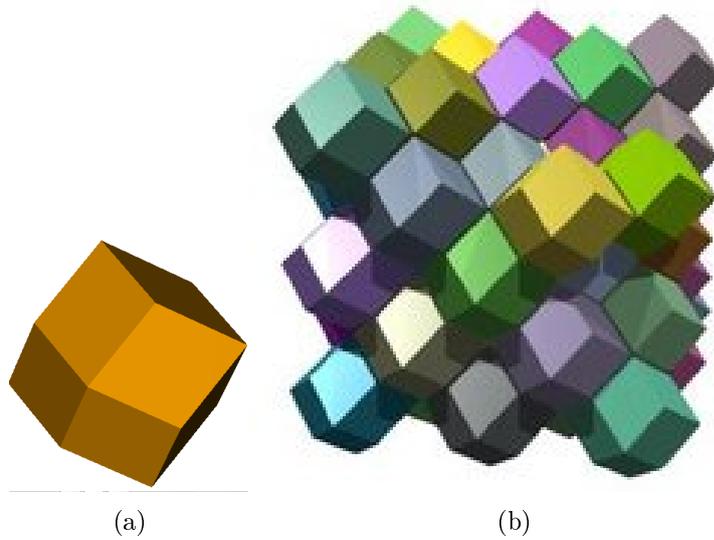


Figura 3.8: (a) Região de Voronoi (b) Partição do \mathbb{R}^3 por dodecaedros rômnicos.

3.1.4 Reticulado cúbico de corpo centrado

No reticulado cúbico de corpo centrado (Figura 3.2b) a região de Voronoi (Figura 3.9) é um octaedro truncado, um dos poliedros de Arquimedes.

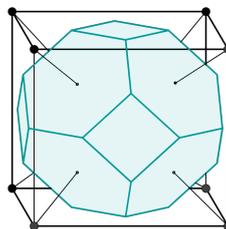


Figura 3.9: Região de Voronoi do reticulado cúbico de corpo centrado.

Um octaedro truncado centrado na origem tem as coordenadas de 24 vértices permutando as coordenadas de $(\pm 1, \pm \frac{1}{2}, 0)$, por exemplo, $a = (0, \frac{1}{2}, 1)$, $b = (0, 1, \frac{1}{2})$, $c = (\frac{1}{2}, 1, 0)$, $d = (1, \frac{1}{2}, 0)$, $e = (1, 0, \frac{1}{2})$, $f = (\frac{1}{2}, 0, 1)$, $g = (0, -\frac{1}{2}, 1)$, $h = (-\frac{1}{2}, 0, 1)$, como ilustrado na Figura (3.10).

Os octaedros truncados formam uma partição do \mathbb{R}^3 (Figura 3.11b). As regiões de Delaunay para este reticulado são todas tetraedros semirregulares.

Observe que as interseções entre duas regiões de Voronoi adjacentes são as faces do octaedro truncado (quadrados ou hexágonos regulares). As interseções de três regiões adjacentes são as arestas, e as de quatro regiões adjacentes, os vértices (Figura 3.11a)

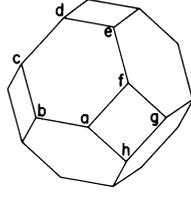


Figura 3.10: Octaedro truncado.

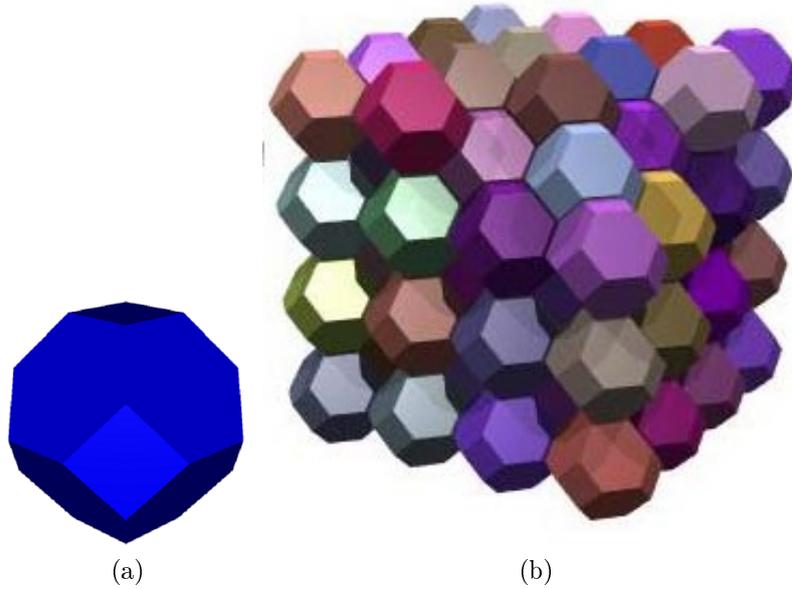


Figura 3.11: (a) Regiões de Voronoi e (b) Partição do \mathbb{R}^3 por octaedros truncados.

3.2 Um breve apanhado histórico sobre a Conjectura de Kepler

O empacotamento de esferas é um dos mais fascinantes e desafiadores assuntos da matemática (ZONG e TALBOT, 1999). É um dos problemas geométricos mais famosos, e o mais antigo da geometria discreta (HALES, 2000). Há quatro séculos atrás, KEPLER (1611), influenciado por Thomas Harriot, realizou um estudo sobre a densidade¹ de empacotamento de esferas e publicou um ensaio no qual descrevia o empacotamento cúbico de face centrada (Figura 3.12) e afirmava que:

“Nenhum empacotamento de esferas de mesmo raio em três dimensões tem densidade maior do que o empacotamento cúbico de face centrada.”

Tal afirmação ficou conhecida como *Conjectura de Kepler*.

Décadas depois, em 1694, David Gregory e Issac Newton discutiram a questão sobre quantas esferas de raios iguais podem ser arranjadas de modo a tocar uma

¹Proporção do espaço ocupada pelas esferas.

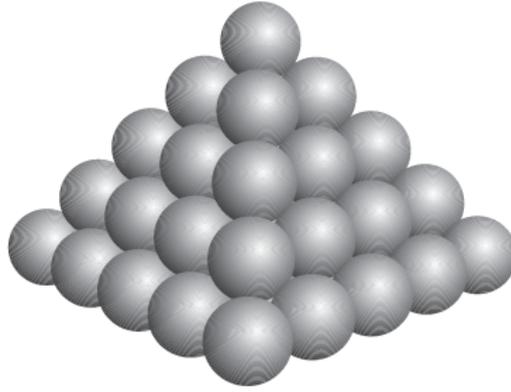


Figura 3.12: Empacotamento cúbico de face centrada.

esfera dada. Este é o *Problema do Kissing Number* de esferas em n dimensões. E, assim, propuseram o *Problema de Gregory Newton*:

“Pode uma esfera tocar 13 esferas de mesmo tamanho?”

Na literatura, às vezes, esse problema é mencionado como o problema das treze esferas. Newton acreditava que no máximo eram 12 esferas, enquanto que Gregory alegava que era possível serem 13 esferas. Porém, somente 250 anos depois, LEECH (1956) provou que Newton tinha razão: o Kissing Number máximo em 3 dimensões é 12. Analogamente, em duas dimensões, exatamente seis círculos, não mais, podem tocar um círculo central.

Desde então, esses problemas e problemas correlatos, como, por exemplo, o problema de recobrimento por esferas (assunto principal de nosso trabalho) o qual é uma espécie de problema dual do problema de empacotamento (CONWAY e SLOANE, 1999), têm atraído a atenção de muitos proeminentes matemáticos.

No século XIX, Carl Friedrich GAUSS (1831) provou que o empacotamento cúbico de face centrada é o mais denso arranjo no qual os centros das esferas formam um arranjo regular, mas deixou em aberto se um empilhamento irregular de esferas pode ser ainda mais denso.

A versão bidimensional da Conjectura de Kepler afirma que o empacotamento regular hexagonal é o mais denso em duas dimensões. Essa versão foi primeiramente provada por THUE (1892). Havia um erro na prova e Thue a corrigiu mais tarde em 1910 (BRASS *et al.*, 2005). Entretanto, muitos acreditam que a primeira prova completa sem falhas foi apresentada por László Fejes Tóth em 1940 (CONWAY e SLOANE, 1999; FUKSHANSKY, 2011; ROGERS, 1964).

No início do século XX, HILBERT (1901) incluiu a conjectura na sua lista dos 23 problemas insolúveis da matemática sendo uma parte importante do seu 18º problema. Todavia, o primeiro avanço significativo para provar a Conjectura de Kepler foi feito por László Fejes Tóth em 1953 (HALES, 2006). Embora a conexão entre o

problema de Gregory Newton e a Conjectura de Kepler não seja óbvia, FEJES TÓTH (1953), no primeiro trabalho descrevendo uma estratégia para provar a Conjectura de Kepler, fez uma versão quantitativa do problema de Gregory-Newton como uma primeira etapa da estratégia. E, mostrou que o problema de determinar a densidade máxima de todos os arranjos (regulares e irregulares) podia ser reduzido a um número finito, porém muito grande de cálculos. Mais tarde, sugeriu que computadores poderiam ajudar a resolver o problema (FEJES TÓTH, 1964). ROGERS (1958) estabeleceu um *bound* superior de $0,7796\dots$ para o problema. Por muitos anos, esse foi o melhor *bound* conhecido até que LINDSEY (1986) melhorou-o obtendo um valor de $0,7784\dots$. Porém, MUDER (1993) estabeleceu o melhor *bound* superior conhecido sobre densidade de empacotamento tridimensional: $0,773055\dots$. A Conjectura de Kepler, um dos principais problemas matemáticos de nossa época (SZPIRO, 2003), permaneceu aberta por quase quatro séculos até ter sido confirmada por HALES (2005) que deu uma prova auxiliada por computadores.

3.3 O Problema de Empacotamento de Esferas

Os problemas de empacotamento consistem em dispor um conjunto de n objetos em um dado container com o objetivo de se obter a maior densidade de empacotamento possível sem que haja sobreposição desses objetos. Em particular, problemas de empacotamento de objetos circulares em duas ou três dimensões tem um vasto leque de aplicações indo das ciências naturais à engenharia, passando pelas atividades comuns do dia a dia.

O problema de empacotamento de círculos tem uma longa história e uma variedade extensa de aplicações tais como: corte circular, carregamento de containeres, empacotamento de cilindros, dispersão de facilidades e rede de comunicações (CASTILLO *et al.*, 2008). Uma introdução histórica ao problema encontra-se em (SZABÓ *et al.*, 2007).

Em termos de otimização, há duas maneiras diferentes (mas equivalentes) de abordar o problema:

- maximizar o raio associado com os n círculos quando as dimensões (área) do container são fixas ou
- minimizar o tamanho (área) do container para acomodar n círculos de raios fixos.

Empacotar objetos circulares é um desafio à geometria discreta e computacional (SZABÓ *et al.*, 2007). Se houver um número grande de objetos circulares para empacotar, a solução ótima será muito difícil de se obter. Uma solução ótima pode

sofrer rotação, reflexão ou uma reordenação dos objetos circulares. Assim, o número de soluções ótimas equivalentes explode quando o número de objetos circulares aumenta (HIFI e M'HALLAH, 2009a).

Os Problemas de Empacotamento de objetos circulares resultam em problemas de otimização onde as posições dos objetos circulares são contínuas enquanto a estrutura de um padrão ótimo tem uma natureza discreta (OSTERGARD, 2008). As tendências mais recentes são desenvolver técnicas que abordem estes dois aspectos simultaneamente (HIFI e M'HALLAH, 2009b).

Muitas versões de empacotamento de objetos circulares no plano foram formuladas como problemas de otimização não-convexa, com a esperança que essas formulações quando resolvidas através de *solvers* de programação não-linear disponíveis produzam boas soluções aproximadas (MLADNOVIC *et al.*, 2005). Contudo, a maioria dos *solvers* de programação não-linear falham em identificar ótimos globais. Na verdade, problemas com objetos circulares são problemas de otimização muito difíceis, pois estes possuem um número muito grande de variáveis e mínimos locais. Assim, é necessário atacá-los com algoritmos que combinam busca local com heurísticas a fim de explorar soluções de uma forma ampla no espaço de busca (LOCATELLI e SCHOEN, 2003). As tendências mais recentes têm adotado métodos de aproximação (heurística) que combinam busca local (heurística) e método de busca local (exato) exaustiva de mínimos locais ou suas aproximações (STOYAN e YAS'KOV, 2004).

O problema geral de empacotamento de esferas investiga o empacotamento mais denso de esferas congruentes no espaço n -dimensional. Este problema tem importância prática em dimensões maiores do que três.

Podemos descrever um empacotamento de esferas no \mathbb{R}^n somente especificando seus centros e o raio.

Em um empacotamento de esferas, se o conjunto dos centros das esferas forma um reticulado no \mathbb{R}^n , chamamos este empacotamento de *empacotamento reticulado*. Assim, o conjunto dos centros das esferas forma um grupo aditivo. Em particular, há uma esfera com centro na origem e se u e v são centros de duas esferas, então também há esferas com centros em $u+v$ e $u-v$. Em cristalografia, esses reticulados são chamados de Reticulados de Bravais.

A forma mais frequente para medir a eficiência de um empacotamento é determinar a sua densidade. A densidade Δ de um empacotamento reticulado é a proporção do espaço ocupado pelas esferas. Mais precisamente,

$$\Delta = \frac{\text{volume de uma esfera}}{\text{volume de uma região fundamental}} = \frac{V_n}{\sqrt{\det\Lambda}},$$

onde V_n é o volume de uma esfera n -dimensional de raio r , dado pela fórmula 3.6.

Se o reticulado hexagonal é definido por (3.12) podemos tomar círculos de raio $\frac{1}{2}$, obtendo o empacotamento mostrado na Figura 3.13, com densidade:

$$\Delta = \frac{\pi}{\sqrt{12}} \cong 0,9069.$$

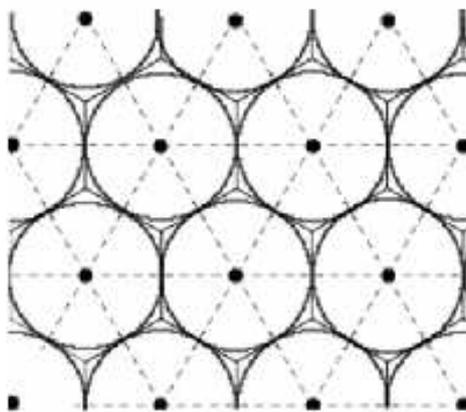


Figura 3.13: Reticulado hexagonal com regiões de Voronoi transladadas e empacotamento de círculos associado.

O empacotamento cúbico de face centrada (Figura 3.12) tem densidade:

$$\Delta = \frac{\pi}{\sqrt{18}} \cong 0,7405.$$

Para um empacotamento associado ao reticulado Λ , com regiões de Voronoi congruentes ao politopo V , dizemos que o raio de empacotamento ρ (*inradius* de V) é o raio da maior esfera inscrita no politopo V .

Para um bom empacotamento reticulado tentamos maximizar ρ , isto é, almejamos escolher os centros das esferas de maneira que o raio das esferas circunscritas às regiões de Voronoi seja o maior possível. No caso do empacotamento não ser regular, consideramos o menor raio das esferas circunscritas entre todas as regiões de Voronoi.

Nem todos os empacotamentos são arranjos regulares. Em três dimensões, por exemplo, há *empacotamentos não regulares* que são tão densos quanto o empacotamento cúbico de face centrada. Isto ocorre porque o empacotamento cúbico de face centrada pode ser construído por camadas, começando por uma camada de esferas colocadas no arranjo regular hexagonal (Figura 3.14) com os centros nos pontos rotulados com a . Há, então, duas maneiras (equivalentes) de colocar a segunda camada: as esferas podem ser colocadas acima nas posições rotuladas com b ou naquelas com c . Suponha que as colocamos em b . Então, a terceira camada pode ser formada colocando as esferas na posição rotulada com a , diretamente acima da primeira camada, ou na posição marcada com c . Analogamente, existem duas escolhas

para cada camada. Ordenar as camadas nas sequências

$$\dots abcabcabc \dots \text{ ou } \dots acbacbacb \dots$$

resulta no *empacotamento cúbico de face centrada* (Figuras 3.12 e 3.16). Mas existem muitas outras possibilidades (não regulares), tais como $\dots acbabacbca \dots$, todas com a mesma densidade.

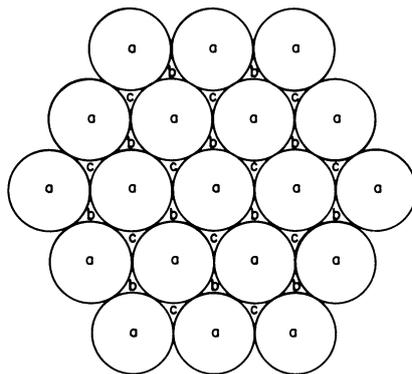


Figura 3.14: Empacotamento de esferas. Os pontos rotulados com a são pontos do reticulado.

Ordenar as camadas na sequência $\dots ababab \dots$ resulta no *empacotamento compacto hexagonal* (Figuras 3.15 e 3.16).

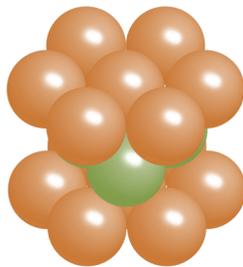


Figura 3.15: Empacotamento compacto hexagonal.

Este é um empacotamento periódico, obtido por colocar uma configuração fixa de esferas, digamos s esferas, em cada região fundamental de um reticulado Δ .

Os melhores empacotamentos conhecidos, segundo registram CONWAY e SLOANE (1999), são: Z , A_2 , A_3 , D_4 , D_5 , E_6 , E_7 , E_8 , K_{12} , Λ_{16} e Λ_{24} . Desses, somente os três primeiros são os mais densos possíveis nas respectivas dimensões.

3.4 O Problema do Kissing Number

O Problema do Kissing Number é um problema geométrico básico e recebeu esse nome oriundo do bilhar: duas bolas se “beijam” (*kiss*) se elas se tocam. O Problema

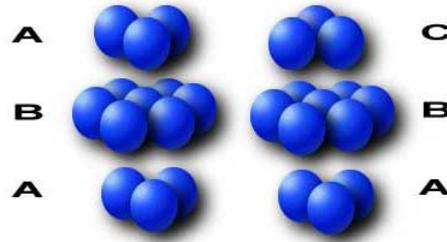


Figura 3.16: À esquerda, o reticulado compacto hexagonal e à direita, o reticulado cúbico de face centrada.

do Kissing Number é, então, determinar quantas bolas podem tocar uma bola central ao mesmo tempo se são todas de mesmo tamanho. Em termos matemáticos, Kissing Number em D -dimensões é o número de hiperesferas de raio R que podem ser dispostas em volta de uma hiperesfera central de raio R tal que cada uma das esferas periféricas tocam a esfera central sem que haja sobreposição das mesmas. Determinar o Kissing Number máximo em várias dimensões é um problema bem conhecido em Geometria Combinatorial.

Denotamos o Problema do Kissing Number em D -dimensões por $KNP(D)$. Em uma e duas dimensões, os resultados são triviais: $KNP(1) = 2$ (Figura 3.17a) e $KNP(2) = 6$ (Figura 3.17b). Porém, este fato não é nada natural no \mathbb{R}^3 . É fácil arranjar 12 esferas de modo que cada uma toque uma esfera central, mas sobra bastante espaço, e não é óbvio que não haja nenhuma maneira de empacotar uma 13ª esfera. De fato, se as 12 bolas são posicionadas nos pontos correspondentes aos vértices de um icosaedro regular concêntrico com a bola central, há tanto espaço extra que, quaisquer duas das 12 esferas de fora podem trocar de lugar através de um movimento contínuo sem nenhuma das esferas de fora perder contato com a esfera central. O arranjo icosaedral de 12 átomos em volta de um átomo central ocorre em certas ligas metálicas, outros compostos e em estruturas não-peródicas conhecidas como *quasicristais*. Como já mencionamos anteriormente, o problema ganhou fama devido a um debate entre Newton e Gregory e, dois séculos depois, LEECH (1956) provou que o $KNP(3)$ é 12 (Figura 3.17c). Em 2003, Oleg Musin provou que $KNP(4)$ é 24. Para dimensões maiores que 4 só são conhecidos $KNP(8) = 240$ e $KNP(24) = 196.560$.

3.5 O Problema de Recobrimento do espaço euclidiano n -dimensional por hiperesferas

Qual é a maneira mais econômica de recobrir o espaço euclidiano n -dimensional com esferas iguais que se sobrepõem?

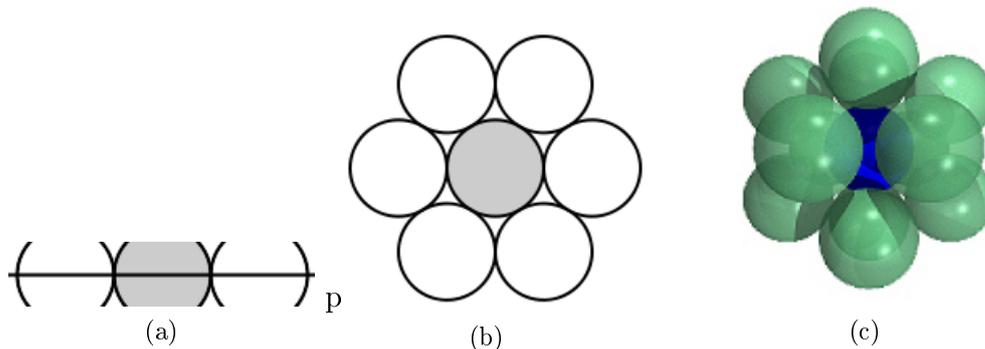


Figura 3.17: (a) Kissing Number em \mathbb{R} , (b) Kissing Number em \mathbb{R}^2 e (c) Kissing Number em \mathbb{R}^3 .

As figuras abaixo mostram duas maneiras de recobrir o plano com círculos que se sobrepõem. Na Figura 3.18 os centros dos círculos pertencem ao reticulado quadrado \mathbb{Z}^2 , e na Figura 3.19 ao reticulado hexagonal. Claramente, o reticulado hexagonal é um recobrimento mais eficiente do que o reticulado quadrado \mathbb{Z}^2 , pois a sobreposição entre as esferas é menor.

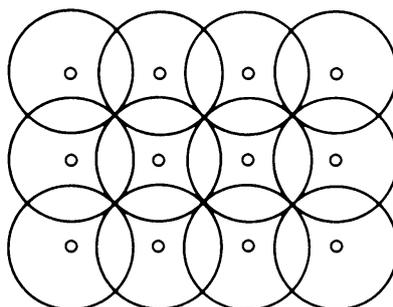


Figura 3.18: Recobrimento do plano por círculos. Os centros pertencem ao reticulado \mathbb{Z}^2 .

Mais precisamente, definiremos a *densidade* Θ de um recobrimento da mesma maneira que a densidade Δ de um empacotamento. Suponhamos que um arranjo de hiperesferas de raio R recobre o \mathbb{R}^n . Se os centros formam um reticulado Λ , então a densidade é definida como o número médio de esferas que contém um ponto do espaço ou:

$$\Theta = \frac{\text{volume de uma esfera}}{\text{volume de uma região fundamental}} = \frac{V_n}{\sqrt{\det\Lambda}}.$$

Θ é também chamado de espessura ou espessividade de recobrimento. Observe que $\Delta \leq 1 \leq \Theta$.

Como no caso de empacotamentos de esferas no \mathbb{R}^n , podemos também descrever um recobrimento do \mathbb{R}^n por esferas somente especificando seus centros e o raio.

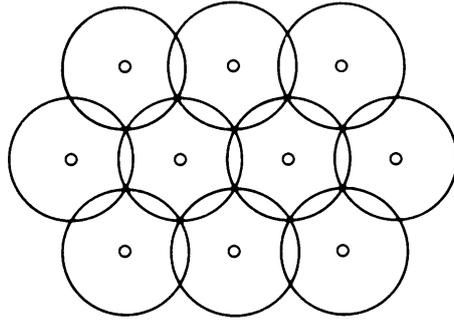


Figura 3.19: Recobrimento do plano por círculos. Os centros pertencem ao reticulado hexagonal.

Dados os recobrimentos da Figura 3.19, temos que $\Theta = \frac{\pi}{2} \cong 1,5708$ para o reticulado quadrado e $\Theta = \frac{2\pi}{3\sqrt{3}} \cong 1,2092$ para o reticulado hexagonal.

O *Problema de Recobrimento* investiga qual é o recobrimento mais eficiente ou econômico do espaço euclidiano n -dimensional por esferas, ou seja, pelo recobrimento com a menor densidade.

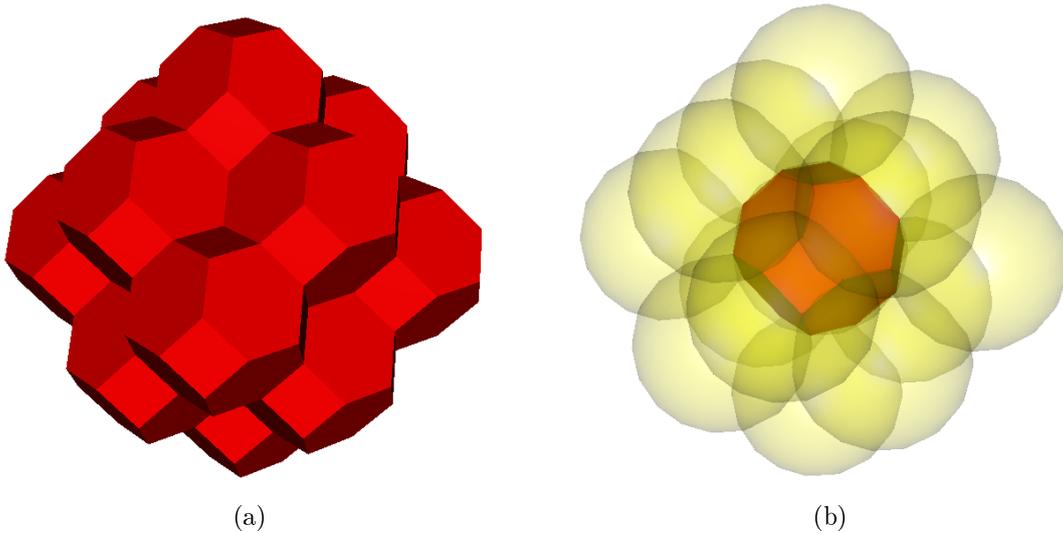


Figura 3.20: (a) Ladrilhamento do espaço por octaedros truncados, (b) O recobrimento reticulado de corpo centrado no \mathbb{R}^3 é obtido colocando-se esferas circunscritas aos octaedros truncados. Observe que cada esfera possui 14 outras esferas adjacentes a ela.

KERSHNER (1939) mostrou que nenhum arranjo de círculos pode recobrir o plano mais eficientemente do que o reticulado hexagonal bidimensional (Figura 3.19). Uma prova mais concisa é apresentada em (FEJES TÓTH, 1972). Em três dimensões, BAMBAH (1954) mostrou que o melhor recobrimento conhecido é o recobrimento cúbico de corpo centrado (Figura 3.20b, para maiores detalhes ver Seção 3.1.4), onde os centros das esferas formam o reticulado cúbico de corpo centrado. Este recobrimento é ótimo entre os recobrimentos reticulados. Em dimensões 4 e

5, os melhores recobrimentos reticulados conhecidos são, respectivamente, A_4^* e A_5^* . Porém, como ocorre com empacotamentos, recobrimentos ótimos não são conhecidos em dimensões mais altas.

Para um recobrimento reticulado Λ , com regiões de Voronoi congruentes ao polítopo V , dizemos que o raio de recobrimento R é o raio da menor esfera circunscrita ao polítopo V (*circumradius* of V).

Para um bom recobrimento reticulado tentamos minimizar R , isto é, almejamos escolher os centros das esferas de maneira que o raio das esferas circunscritas às regiões de Voronoi seja o menor possível. No caso do recobrimento não ser regular, consideramos o maior raio entre todas as regiões de Voronoi.

Na literatura, há poucos trabalhos de otimização que tratam do problema de recobrimento finito e alguns deles já foram citados nos capítulos anteriores.

Capítulo 4

Modelo proposto e discussão do método de resolução

Alguns autores abordam o problema de planejamento do tratamento da radiocirurgia *Gamma Knife* sob o ponto de vista de um problema de empacotamento, não permitindo que as esferas se sobreponham. Dessa forma, o problema é descrito como um modelo matemático de empacotamento de esferas de diâmetros diferentes em regiões tridimensionais, almejando uma densidade de empacotamento superior a 0,9. O empacotamento de esferas idênticas em um espaço euclidiano tridimensional ocupa no máximo 74% desse espaço. Como o tratamento em questão disponibiliza quatro diferentes tamanhos de esferas, uma densidade mais alta é possível de ser obtida, de acordo com os requerimentos do tratamento.

Todavia, desejamos, nesse trabalho, não apenas um recobrimento parcial, mas uma cobertura total da região. Para esse feito, a restrição de não-sobreposição das esferas é relaxada, sendo permitido que esferas distintas possuam um volume de interseção. Porém, essa situação deve ser evitada, já que a sobreposição das esferas indica uma aplicação dobrada de radiação na região compreendida pela interseção, enquanto que, idealmente, o tumor deva ser recoberto de forma homogênea. Temos, então, um problema que recai na classe de problemas de recobrimento, com o objetivo de utilizar o menor número possível de esferas.

Além disso, devido ao requerimento da radiação recebida pelos tecidos saudáveis ao redor do tumor ser a mínima possível, necessitaremos definir uma *região de segurança* para o tumor. As esferas participantes da cobertura, não deverão ocupar a região externa a essa *região de segurança*. Se considerarmos a união do tumor e da *região de segurança*, então poderemos também interpretar o problema como sendo um empacotamento relaxado de esferas, permitindo que as esferas possam se sobrepor.

Neste capítulo, propomos uma nova abordagem para resolver o modelo MINLP de programação matemática apresentado por PINTO (2011).

4.1 O Modelo

Seja S um conjunto de esferas. O **problema clássico de recobrimento ótimo** consiste em cobrir um sólido $T \subseteq \mathbb{R}^3$ com as esferas pertencentes a S , utilizando o menor número possível dessas esferas. De outra forma, devemos posicionar as esferas de modo que cada ponto $p \in T$, pertença também a alguma das esferas. Propomos, nesta seção, uma descrição formal do problema de recobrimento de um sólido tridimensional por esferas de diferentes diâmetros, em termos de uma formulação como um problema de programação matemática.

Seja S um conjunto formado por n esferas, $S = \{S_1, \dots, S_n\}$. Assim, para cada $i \in N$, $N = \{1, \dots, n\}$, seja S_i uma esfera centrada em $x^i \in \mathbb{R}^3$ e com raio $r_i > 0$. Assumiremos que um recobrimento de T por esferas de S existe. Uma solução para o problema clássico de recobrimento ótimo pode ser definida por variáveis x e y , onde:

- $x \in T^n$,
- $y \in \{0, 1\}^n$, sendo $y_i = \begin{cases} 1, & \text{se } \mu(T \setminus \bigcup_{j \neq i} S_j) > 0 \\ 0, & \text{caso contrário} \end{cases}$, $\forall i \in \{1, \dots, n\}$,

onde μ é a medida de Lebesgue em \mathbb{R}^3 . A variável $x = \{x^1, \dots, x^n\}$ representa um vetor contendo o centro de cada esfera e a variável binária $y = \{y_1, \dots, y_n\}$ determina quais esferas do conjunto S estão presentes na solução. O termo

$$V_{S_i} = T \setminus \bigcup_{j \neq i} S_j$$

representa o que sobra do volume T , após subtraídas as regiões pertencentes a cada uma das esferas, com exceção da esfera S_i . Logo, se $\mu(V_{S_i}) > 0$, então a esfera S_i tem contribuição positiva no recobrimento e, portanto, deve participar da solução, resultando em $y_i = 1$. Para desenvolvimentos futuros, definamos $I_k \subset N$, para $k \in N_K = \{0, 1, \dots, K\}$, que representará uma determinada escolha de esferas, dentre $K + 1$ possíveis alternativas, incluindo seleção nula de esferas, dada sempre por $I_0 = \emptyset$. Desta forma, dado $k \in N_K$, teremos $y_i = 1$ se $i \in I_k$, e $y_i = 0$ se $i \notin I_k$.

Para apresentar a formulação, primeiramente, consideremos os parâmetros

$$\alpha_{ij}, \quad 1 \leq i < j \leq n, \quad e$$

$$c_i, \quad 1 \leq i \leq n.$$

O parâmetro α_{ij} mede o quanto as esferas S_i e S_j podem se sobrepor. Se $\alpha_{ij} = 0$, então $|S_i \cap S_j| \leq 1$, o que significa que as esferas podem apenas se tangenciar.

Devemos ter

$$\alpha_{ij} < \min\{r_i, r_j\}, \quad \forall 1 \leq i < j \leq n.$$

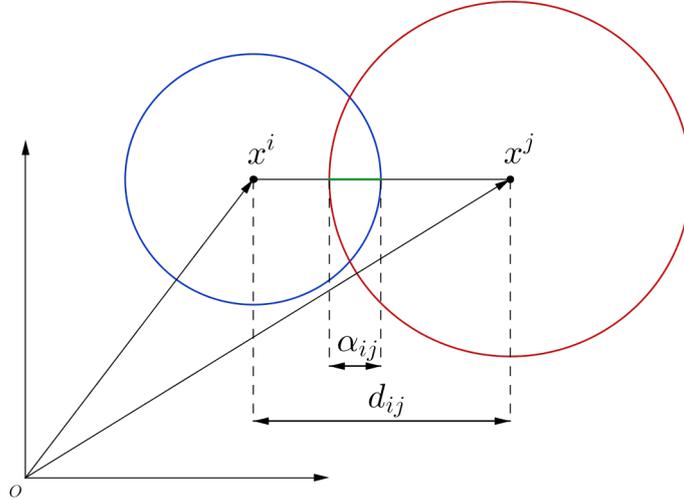


Figura 4.1: Representação dos parâmetros α_{ij} , representando a máxima sobreposição das esferas.

Mais ainda, podemos expressar α_{ij} como

$$\alpha_{ij} = \rho \min\{r_i, r_j\}, \quad 0 \leq \rho < 1.$$

Seja d_{ij} a distância entre os centros das esferas S_i e S_j . Então,

$$d_{ij} = r_i + r_j - \alpha_{ij}, \quad (4.1)$$

se as esferas S_i e S_j , possuem sobreposição. Caso contrário, d_{ij} assume o valor $r_i + r_j$.

O parâmetro c_i mede o quanto a esfera S_i contribui para o recobrimento.

Assim, a seguinte formulação é proposta:

$$(P1) : \max \quad \sum_{i=1}^n c_i y_i \quad (4.2)$$

$$\text{s. a } \|x^i - x^j\|_2^2 \geq d_{ij}^2 (y_i + y_j - 1), \quad \forall i < j \leq n \quad (4.3)$$

$$x \in T^n \quad (4.4)$$

$$y \in \{0, 1\}^n \quad (4.5)$$

a qual define o *problema de recobrimento α -ótimo* de T .

As variáveis de decisão são os $x^i \in \mathbb{R}^3$, centros das esferas, e os $y_i \in \{0, 1\}$,

que determinam se cada esfera está ou não presente na cobertura. Observe que, embora o objetivo seja o de minimizar a quantidade de esferas na solução, a direção de otimização é a maximização. Neste modelo, maximiza-se o volume das esferas no recobrimento, de forma a priorizar as maiores e, por consequência, diminuindo a quantidade de esferas na solução.

É importante observar que, em consequência de desejarmos que a solução contenha esferas posicionadas de forma a restringir a ocupação da região externa a T , o mínimo possível e das restrições quanto às interseções entre as esferas, definidas pelos parâmetros α_{ij} , talvez não seja possível recobrir completamente o conjunto T .

A partir desta formulação, temos o seguinte lema que define a existência de um conjunto de valores para os parâmetros α_{ij} e c_i , com os quais podemos retomar a solução do problema clássico.

Lema 4.1. *Existem $\{\alpha_{ij} \geq 0\}_{1 \leq i < j \leq n}$ e $\{c_i \geq 0\}_{1 \leq i \leq n}$ para os quais uma solução ótima de (4.2)-(4.5) é uma solução ótima para o problema clássico de recobrimento ótimo.*

Demonstração: Sejam (x^*, y^*) uma solução ótima do problema clássico de recobrimento ótimo e

$$dist_{ij} = \|(x^i)^* - (x^j)^*\|$$

a distância entre x^i e x^j , respectivamente, os centros das esferas S_i e S_j , com $1 \leq i < j \leq n$. Segue que :

$$\alpha_{ij} = \begin{cases} 0, & \text{se } dist_{ij} \geq r_i + r_j; \\ r_i + r_j - dist_{ij}, & \text{se } dist_{ij} < r_i + r_j. \end{cases}$$

Tome

$$c_i = y_i^*, \text{ para todo } 1 \leq i \leq n.$$

Afirmamos que (x^*, y^*) é uma solução ótima do problema de recobrimento α -ótimo com os parâmetros $\{\alpha_{ij}\}_{1 \leq i < j \leq n}$ e $\{c_i\}_{1 \leq i \leq n}$ como definidos acima.

De fato, utilizando os parâmetros escolhidos, a solução ótima do problema clássico também será viável em (4.2)-(4.5). Basta mostrar, então, que ela terá o melhor valor da função objetivo. Isto de fato ocorre, pois os y_i^* assumem valores 0 ou 1, para todo i , e como $c_i = y_i^*$, temos que

$$\sum_{i=1}^n c_i y_i = \sum_{i|y_i^*=1} y_i.$$

Esse somatório atingirá seu valor máximo se tomarmos $y_i = 1$ quando $y_i^* = 1$, ou

seja, quando dermos às variáveis y_i os valores y_i^* . Portanto, (x^*, y^*) é uma solução ótima de (4.2)-(4.5) com os parâmetros tais como definidos. ■

Tanto o problema clássico quanto os problemas de recobrimento α -ótimos não podem ser resolvidos explicitamente, a menos que a restrição (4.4) torne-se explícita. Em casos simples T pode ser descrito através de equações matemáticas, como por exemplo quando T é um paralelepípedo ou um elipsóide. Discutiremos este assunto detalhadamente mais adiante.

Como já mencionado, o problema $(P1)$ é classificado como um MINLP não-convexo. Observemos que os problemas MINLP são naturalmente não-convexos pois as variáveis inteiras tornam a região viável não-convexa. Devido às não linearidades, os problemas do tipo MINLP, mesmo após as restrições de integralidade serem relaxadas, são em geral não-convexos. Em nosso caso, mesmo se considerássemos uma relaxação das variáveis y_i , percebemos que as restrições (4.3) são não-convexas. Podemos, portanto, dividir as fontes de não-convexidade presentes em $(P1)$ em três categorias de restrições:

- NC_1 - **restrições de integralidade**: tratam-se das restrições (4.5) sobre as variáveis y_i , $i = 1, 2, 3, \dots, n$, que determinam a presença ou não da esfera S_i no recobrimento;
- NC_2 - **restrições de sobreposição**: tratam-se das restrições não-convexas (4.3) que restringem o quanto cada esfera pode se sobrepor a uma outra esfera;
- NC_3 - **restrições do sólido**: dependendo do sólido a ser recoberto, as restrições (4.4) poderão ser não-convexas.

A razão para a divisão apresentada acima está diretamente relacionada com a abordagem proposta para resolver o problema $(P1)$ conforme descrito a seguir.

4.2 Abordagem proposta para resolver $(P1)$

Como o problema $(P1)$ é do tipo MINLP não-convexo iremos adotar uma abordagem baseada na aplicação de técnicas de penalização e heurísticas estocásticas de busca. Desta forma, não estaremos a princípio garantindo a otimalidade de nossas soluções. No entanto, os métodos desenvolvidos irão permitir pelo menos a seleção da melhor candidata à solução do problema original dentre um conjunto de possíveis soluções produzidas mediante a aplicação de heurísticas especialmente projetadas.

Iremos agora analisar brevemente cada fonte de não-convexidade e apresentar a maneira escolhida para tratá-las.

As restrições de integralidade (4.5), onde as variáveis y_i , $i = 1, 2, 3, \dots, n$, são variáveis binárias, são a fonte de não-convexidade NC_1 . Estas variáveis representam o processo de escolha das esferas que irão participar do recobrimento do sólido T .

Existem dois métodos tradicionalmente utilizados para a resolução exata de problemas com variáveis inteiras: o método dos planos de corte e o método *Branch and Bound*. De uma forma geral, estes métodos organizam a busca por soluções ótimas dentro da região viável, dividindo-a em sub-regiões onde a busca deve ser priorizada, ao mesmo tempo em que eliminam as sub-regiões que não irão fornecer candidatas adequadas, tais como sub-regiões inviáveis ou com melhor sub-região possível.

Em nosso caso, procurou-se aplicar a ideia de árvores de decisão (subjacente ao método *Branch and Bound*) utilizando índices de cobertura, os quais serão definidos a posteriori, como as variáveis de decisão em cada nó. Devemos notar que cada nó dessa árvore irá representar uma determinada escolha de esferas utilizadas para o recobrimento. Dadas as características e interação entre os índices de cobertura adotados, nossa árvore de decisão será explorada de forma linear com *backtracking*.

Na Seção 4.6 serão apresentadas a árvore de decisão, sua topologia, os índices de cobertura utilizados e sua aplicação na árvore de decisão. Uma discussão geral sobre os índices de cobertura aplicados à técnica do *Gamma Knife* pode ser encontrada no Apêndice C. Estes índices utilizam o cálculo de volumes em sua definição. Para a realização da computação destes volumes adotou-se a técnica de quasi-Monte Carlo, descrita no Apêndice D.

As restrições (4.3) de sobreposição de esferas, fonte de não-convexidade NC_2 , são bastante complexas e não serão tratadas de forma exata mas sim mediante a aplicação de técnicas de penalização. O tratamento de NC_2 será na verdade a primeira providência a ser tomada em uma reformulação do problema original ($P1$) e será apresentada na Seção 4.3.

As restrições (4.4) quanto à posição das esferas em relação ao sólido T também serão tratadas mediante a aplicação de técnicas de penalização. Se estas restrições forem não-convexas, o que em última análise resultam da não-convexidade do sólido T , elas se constituem, então, na fonte de não-convexidade NC_3 . Neste trabalho trataremos somente de sólidos convexos e, portanto, não iremos lidar explicitamente com a fonte de não-convexidade NC_3 , apesar de que as técnicas básicas utilizadas também são válidas para sólidos não-convexos. O problema mais sério no tratamento das restrições (4.4), no entanto, se manifesta mesmo para sólidos convexos e está relacionado com a não regularidade das superfícies paralelas internas ao sólido T . Estes assuntos serão tratados na Seção 4.5.

4.3 Reformulação do problema

4.3.1 Reformulação exata de (P1)

Conforme definido anteriormente, consideremos uma determinada seleção fixa de esferas do conjunto S , com índices agrupados em $I_k \subset N$, para $k \in N_K = \{0, \dots, K\}$, onde N_K é um conjunto indexando todas as possíveis seleções de esferas consideradas. Se todas as esferas de uma seleção $k \in N_K$ participam do recobrimento, então $y_i = 1$ para todo $i \in I_k$, enquanto $y_i = 0$ para todo $i \notin I_k$. Daqui por diante, chamaremos esta k -ésima seleção de *nó*, que também designará informalmente o próprio conjunto I_k . Posteriormente discutiremos a formação do conjunto N_K .

Para um nó k genérico, o problema (P1) pode ser reescrito como:

$$(PN1) : \max \sum_{i \in I_k} c_i \quad (4.6)$$

$$\text{s. a.} \quad \|x^i - x^j\|^2 \geq d_{ij}^2, \quad i < j, i, j \in I_k \quad (4.7)$$

$$x \in T^n \quad (4.8)$$

A função objetivo do problema (PN1) é constante e, portanto, podemos considerá-lo como um Problema de Satisfação de Restrições (*Constraint Satisfaction Problem – CSP*). Por outro lado, as restrições (4.7) são não-convexas e sendo a variável x contínua, temos portanto que o problema (PN1) é um problema de otimização global.

Se considerarmos a função $p : \mathbb{R}_+ \rightarrow \overline{\mathbb{R}}$ definida por:

$$p(t) = \begin{cases} +\infty, & \text{se } 0 \leq t < 1 \\ 0, & \text{se } t \geq 1 \end{cases},$$

podemos reformular (PN1) da seguinte maneira:

$$(PN2) : \max \sum_{i \in I_k} c_i - \sum_{\substack{i, j \in I_k \\ i < j}} p\left(\frac{\|x^i - x^j\|}{d_{ij}}\right) \quad (4.9)$$

$$\text{s.a.} \quad x \in T^n \quad (4.10)$$

Devemos observar que o argumento da função p dispensa a elevação ao quadrado presente na restrição (4.7). A fim de simplificar a expressão da função objetivo, usaremos $C_k = \sum_{i \in I_k} c_i$ nas formulações seguintes.

4.3.2 Reformulação de (P1) com penalização

Apesar de equivalente, a formulação (PN2) não pode ser resolvida por nenhum *solver* conhecido, já que a função p não é contínua para $t = 1$, além de assumir o valor $+\infty$, o que somente é válido para o conjunto dos reais estendidos $\overline{\mathbb{R}}$. A formulação (PN2) pode, no entanto, ser interpretada como a versão limite de uma penalização do problema (PN1). Podemos, então, considerar a substituição de p por uma versão suavizada e parametrizada p_μ de tal forma que p_μ tenda para p quando $\mu \rightarrow \infty$.

Com base na teoria dos métodos de penalização e barreiras, amplamente descrita na literatura (ver, por exemplo, LUENBERGER e YE (2008)), consideremos primeiramente a função p_μ definida por:

$$p_\mu(t) = \begin{cases} \mu(t-1)^2 & , \text{ se } 0 \leq t < 1 \\ 0 & , \text{ se } t \geq 1, \end{cases}$$

onde $\mu > 0$ é uma constante que especifica o nível de penalização aplicado.

É importante observarmos que p_μ é diferenciável no intervalo $(0, +\infty)$. Além disso, temos que $p_\mu \rightarrow p$ quando $\mu \rightarrow +\infty$. A Figura 4.2 mostra o gráfico de p_μ para alguns valores arbitrários de μ .

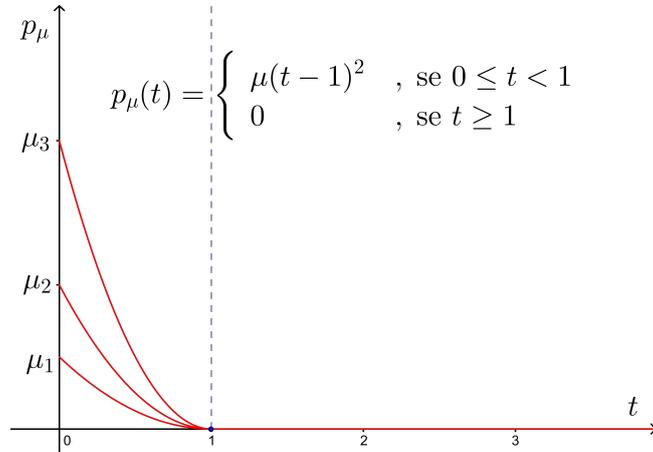


Figura 4.2: Gráfico da função $p_\mu(t)$

Podemos agora considerar a versão penalizada do problema (PN1), simplesmente pela substituição da função $p(t)$ presente no problema (PN2) pela função $p_\mu(t)$:

$$(PN3) : \max C_k - \sum_{\substack{i,j \in I_k \\ i < j}} p_\mu \left(\frac{\|x^i - x^j\|}{d_{ij}} \right) \quad (4.11)$$

$$\text{s.a.} \quad x \in T^n \quad (4.12)$$

Dado um valor para o parâmetro μ e um valor inicial x_0 , o problema (PN3) pode ser facilmente resolvido por um solver não linear convencional, se estivermos considerando somente a busca de um ótimo local. A solução (local) ótima é aproximada fazendo-se $\mu \rightarrow +\infty$.

No entanto, mesmo que seja possível exprimir a restrição (4.12) como uma restrição convexa, o problema (PN3) continua sendo um problema de otimização global e providências (heurísticas) devem ser tomadas para garantir que uma solução do problema penalizado seja suficientemente boa, a ponto de ser considerada uma candidata aceitável como solução do nó em questão. Seguindo esta diretriz, discutiremos aprimoramentos para a função p_μ na próxima seção.

Torna-se oportuno mencionar, neste momento, que o valor inicial x_0 , utilizado na busca da solução ótima (local) de cada nó, será obtido a partir da melhor configuração obtida no nó anterior acrescentando-se uma nova esfera em uma posição randômica convenientemente escolhida. Desta forma, estaremos progredindo no sentido de buscarmos boas soluções para cada nó.

4.3.3 Aprimoramento da função p_μ

A função p_μ apresentada acima corresponde à aplicação tradicional da teoria dos métodos de penalização e barreiras. Tratando-se do nosso problema, mais especificamente, seria conveniente que algumas características adicionais pudessem estar presentes na função p_μ . Porém, a inclusão de novas características à função p_μ irá pouco a pouco nos distanciar da abordagem inicial, voltada para o uso tradicional de penalizações, e irá nos conduzir a uma abordagem mais semelhante aos métodos que empregam funções que descrevem potenciais, onde se busca a minimização da energia total. Desta forma, o parâmetro μ (e/ou outros parâmetros acrescentados posteriormente) irá estar mais ligado à conformação geral da curva de penalização do que à ideia de uma função limite quando os parâmetros tendem a valores pré-estabelecidos, como é o caso de p_μ quando $\mu \rightarrow +\infty$. Assim, veremos que os parâmetros serão fixados em valores finitos escolhidos para se conseguir as melhores características para a função de penalização.

Iremos, primeiramente, embutir em p_μ um nível de penalização maior para as esferas distantes, de tal forma que haja uma tendência natural para a aglomeração das mesmas, algo esperado para a solução ótima do problema (PN1). Isto pode ser obtido pela modificação do nível constante e igual a zero quando $t \geq 1$, originalmente presente em p_μ , para uma rampa com inclinação dada pelo parâmetro $\lambda > 0$. Desta forma, podemos redefinir p_μ como uma nova função a dois parâmetros dada por:

$$p_{\mu,\lambda}(t) = \begin{cases} \mu(t-1)^2 & , \text{ se } 0 \leq t < 1 + \frac{\lambda}{2\mu}, \\ \lambda(t-1) - \frac{\lambda^2}{4\mu} & , \text{ se } t \geq 1 + \frac{\lambda}{2\mu}. \end{cases}$$

Esta função pode ser facilmente deduzida se considerarmos uma reta tangenciando uma parábola. A função $p_{\mu,\lambda}$ foi construída para ser diferenciável (e portanto, contínua) no intervalo $(0, +\infty)$, à semelhança de p_μ , e satisfazendo $p_{\mu,\lambda}(1) = 0$. O ajuste do parâmetro λ permite graduar a tendência de aglomeração das esferas pois o nível de penalização pode se tornar arbitrariamente elevado com o aumento de λ , e obviamente com a distância entre as esferas.

A Figura 4.3 mostra o gráfico de $p_{\mu,\lambda}$ para alguns valores arbitrários de λ .

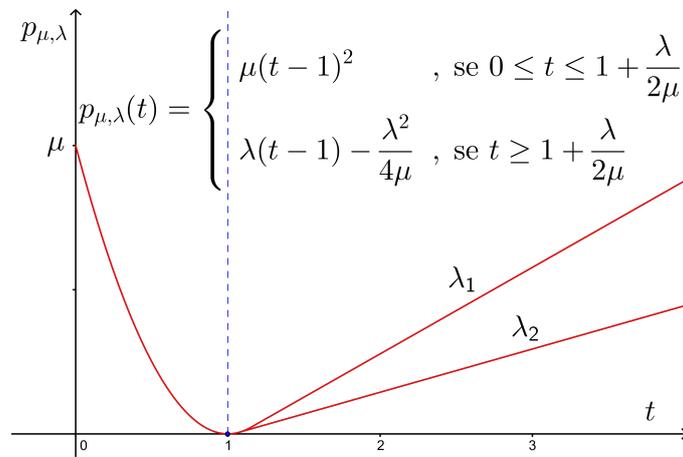


Figura 4.3: Gráfico da função $p_{\mu,\lambda}(t)$

Todavia, $p_{\mu,\lambda}$ ainda pode ser aperfeiçoada. Suas principais deficiências são:

- desejamos que, para quaisquer duas esferas distantes, a penalização assuma um nível aproximadamente estável (constante), restringindo o efeito de aglomeração das esferas às situações em que elas já estão relativamente próximas;
- a faixa de valores de t na vizinhança do ponto $t = 1$ não é especialmente privilegiada com baixa penalização, o que seria desejável quando as esferas estiverem próximas, visando eliminar possíveis lacunas no recobrimento de T .
- é uma função definida por partes, e isto pode dificultar sua implementação em um *solver*;

Teremos, portanto, que avaliar uma nova função onde as deficiências citadas acima não estejam presentes. As funções p_μ e $p_{\mu,\lambda}$, desenvolvidas até este ponto partindo da função p , não podem ser ajustadas por serem naturalmente definidas por partes e, portanto, decidimos abandoná-las em busca de uma função mais adequada.

Uma função, que apresenta todas as características desejadas, é a função utilizada em Bioquímica para descrever a interação de Van der Waals entre átomos e

moléculas (JONES, 1924). Abaixo apresentamos uma expressão bastante geral para esta função

$$p_w(t) = E + \frac{C_m}{t^m} - \frac{C_n}{t^n}, \quad m, n \in \mathbb{N}, m > n. \quad (4.13)$$

A Figura 4.4 apresenta o gráfico da função p_w , onde a conformação desejada é apresentada. A função p_w é racional e definida por uma expressão única para $t > 0$, eliminando o problema da “definição por partes” presente nas versões anteriores. Também percebemos uma alta penalização para $t < 1$, assim como uma estabilização do nível de penalização quando $t \rightarrow \infty$. É notório que a vizinhança do ponto $t = 1$ é altamente privilegiada por níveis de penalização próximos de zero.

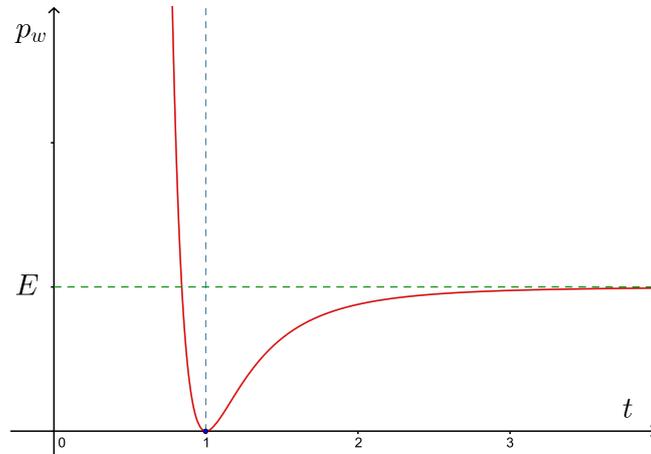


Figura 4.4: Gráfico de uma versão da função $p_w(t)$

A princípio, os valores dos parâmetros E , m , n , C_m e C_n são livres, mas a adequação à conformação gráfica, apresentada na Figura 4.4, irá trazer restrições sobre alguns destes parâmetros.

Iremos, primeiramente, determinar os valores de C_n e C_m . Diferenciando a função (4.13) obtemos:

$$p'_w(t) = -\frac{mC_m}{t^{m+1}} + \frac{nC_n}{t^{n+1}}. \quad (4.14)$$

Observemos que, em $t = 1$, a função (4.13) atinge o seu valor mínimo e, portanto, $p'_w(1) = 0$. Deste fato, temos que:

$$-mC_m + nC_n = 0 \quad (4.15)$$

e

$$C_m = \frac{n}{m}C_n. \quad (4.16)$$

Substituindo (4.16) em (4.13) obtemos:

$$p_w(t) = E + \frac{nC_n}{mt^m} - \frac{C_n}{t^n} \quad (4.17)$$

que pode ser reescrita como:

$$p_w(t) = E + \frac{C_n}{mt^m} [n - mt^{m-n}], \quad m > n. \quad (4.18)$$

Por outro lado, $t = 1$ também é zero da função, o que implica em

$$p_w(1) = 0.$$

Por conseguinte,

$$E + \frac{n-m}{m}C_n = 0,$$

implica em

$$E = \frac{m-n}{m}C_n$$

e

$$C_n = \frac{m}{m-n}E. \quad (4.19)$$

Substituindo (4.19) em (4.18) obtemos:

$$p_w(t) = E \left[1 - \frac{m \left(\frac{1}{t}\right)^n - n \left(\frac{1}{t}\right)^m}{m-n} \right], \quad m > n. \quad (4.20)$$

A função de Van de Walls é comumente utilizada na configuração Lennard-Jones com $n = 6$ e $m = 12$ (JONES, 1924), ou, mais geralmente, com $m = 2n$. Se, portanto, fizermos $m = 2n$ a função p_w pode ser expressa por

$$p_w(t) = E \left(1 - \frac{1}{t^n} \right)^2. \quad (4.21)$$

A Figura 4.5 mostra o gráfico de p_w para alguns valores arbitrários de n , enquanto a Figura 4.6 mostra o gráfico de p_w para alguns valores arbitrários de E .

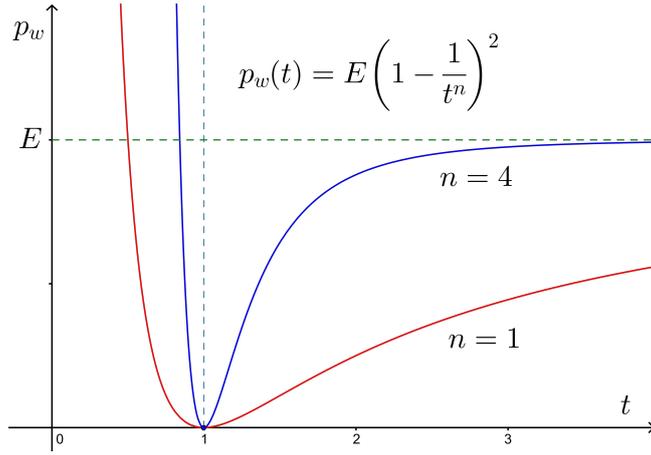


Figura 4.5: Gráfico da função $p_w(t)$ para $n = 1$ e $n = 4$

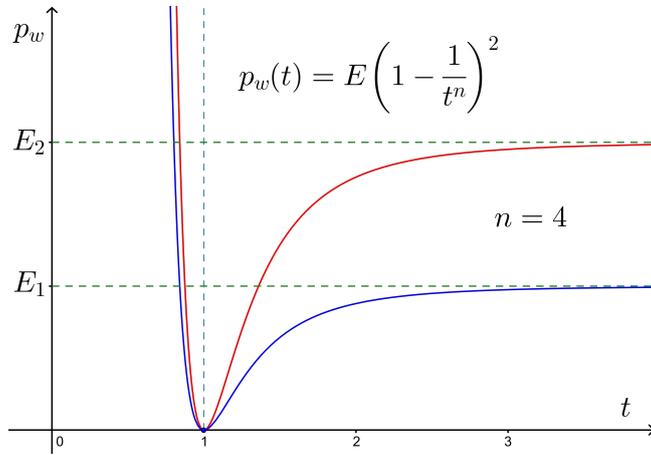


Figura 4.6: Gráfico da função $p_w(t)$ para valores distintos de E , com $n = 4$

A função p_w também pode ser expressa por

$$p_w(t) = E \left(\frac{1}{t^n} - 1 \right)^k, \quad (4.22)$$

onde $k = 2$, mas de uma forma geral iremos considerar k par e possivelmente maior que dois. Se definirmos $f(\xi) = E(\xi - 1)^k$ e $g(\xi) = 1/\xi^n$, temos $p_w(t) = (f \circ g)(t)$. A função $f(\xi)$ para $k = 2$ é uma parábola com vértice no ponto $(1, 0)$ e interseção com o eixo vertical no ponto $(0, E)$. Podemos, portanto, obter p_w como a composição de uma parábola com uma função que associa o intervalo $(0, 1]$ ao intervalo $[1, +\infty)$ e o intervalo $[1, +\infty)$, no intervalo $(0, 1]$. Desta forma, pode-se compreender a razão para o formato do gráfico de p_w .

As derivadas primeira e segunda de p_w podem ser facilmente calculadas:

$$p'_w(t) = -E \frac{kn}{t^{n+1}} \left(\frac{1}{t^n} - 1 \right)^{k-1} \quad (4.23)$$

$$p''_w(t) = E \frac{kn(n+1)}{t^{n+2}} \left(\frac{1}{t^n} - 1 \right)^{k-2} \left(\frac{\frac{kn+1}{n+1}}{t^n} - 1 \right) \quad (4.24)$$

Pretendemos, com esses comentários, introduzir um aspecto ainda não perfeitamente contemplado em p_w , e que será consequência dos desenvolvimentos apresentados na Seção 4.4. Na vizinhança do ponto $t = 1$ o gráfico de p_w se comporta localmente como uma parábola, não podendo assim assegurar uma região suficientemente ampla onde o nível de penalização possa ser considerado desprezível. Este fato pode ser facilmente verificado mediante o cálculo de p''_w para $k = 2$ no ponto $t = 1$, ou seja, $p''_w(1) = 2n^2E$. Como $p''_w(1) > 0$ e $p'_w(1) = 0$, não temos uma vizinhança de $t = 1$ onde o gráfico de p_w possua um formato próximo de uma reta horizontal e suficientemente plano. Seria necessário que, pelo menos, $p''_w(1) = 0$ para que este objetivo fosse alcançado.

Retornando à decomposição de p_w apresentada acima, percebemos que se substituirmos a parábola por uma curva de quarto grau talvez tenhamos sucesso na tentativa de anularmos a segunda derivada. Esta alteração pode ser implementada simplesmente fazendo-se $k = 4$ na Definição 4.22 de $p_w(t)$. Para $k = 4$ podemos verificar que $p''_w(1) = 0$, mas o formato geral do gráfico não é substancialmente alterado. A Figura 4.7 apresenta os gráfico de p_w para $n = 4$, considerando-se $k = 2$ e $k = 4$, a título de comparação. Observe que todas as características discutidas anteriormente continuam válidas.

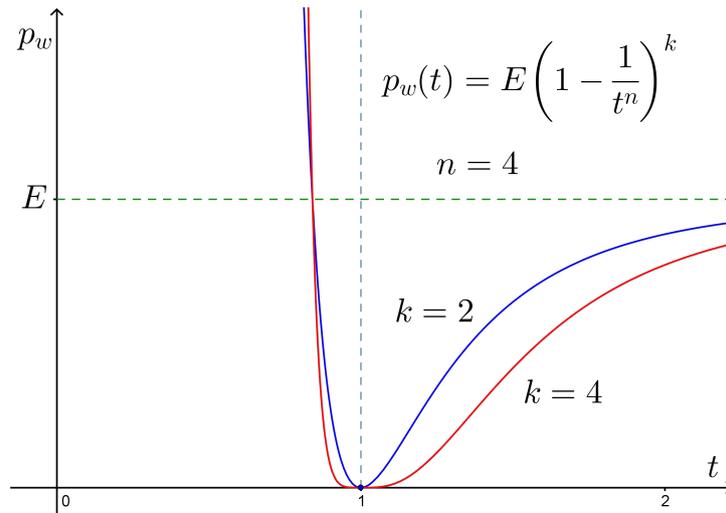


Figura 4.7: Gráfico da função $p_w(t)$ com $n = 4$ para $k = 2$ e $k = 4$

Em teoria, o desenvolvimento apresentado acima seria satisfatório. No entanto,

a versão da função $p_w(t)$ dada por (4.22), pode gerar problemas de *overflow* quando utilizada em uma implementação computacional, pois

$$\lim_{t \rightarrow 0} p_w(t) = \infty.$$

Desta forma, para evitar *overflow*, faremos uma transformação afim na variável independente da função $p_w(t)$, a qual denotaremos abaixo por $p_v(t)$. Assim, redefiniremos a função $p_w(t)$ como:

$$p_v(t) = p_w(a_k t + b_k) = E \left(1 - \frac{1}{(a_k t + b_k)^n} \right)^k,$$

onde a_k e b_k são constantes a serem determinadas de maneira a satisfazer as seguintes condições:

$$\begin{cases} p_v(0) = L & , \text{ com } L \gg E, \\ p_v(1) = 0. \end{cases}$$

Como

$$p_v(0) = E \left(1 - \frac{1}{b_k^n} \right)^k = L,$$

então

$$b_k = \frac{1}{\sqrt[n]{\sqrt[k]{L/E} + 1}}. \quad (4.25)$$

De

$$p_v(1) = E \left(1 - \frac{1}{(a_k + b_k)^n} \right)^k = 0$$

segue que

$$a_k + b_k = 1 \Rightarrow a_k = 1 - b_k. \quad (4.26)$$

Por conseguinte, temos que

$$p_v(t) = E \left(1 - \frac{1}{((1 - b_k)t + b_k)^n} \right)^k,$$

onde b_k é dado pela expressão (4.25) acima.

A Figura 4.8 apresenta o gráfico de p_v para $n = 4$ considerando-se $k = 2$ e $k = 4$.

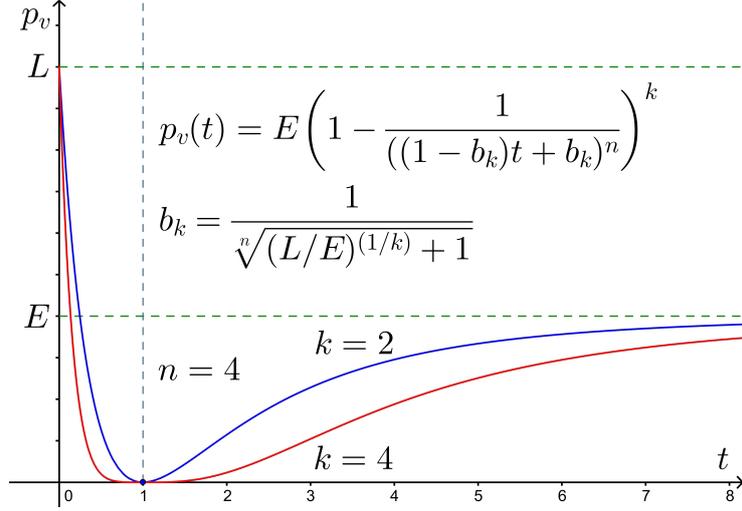


Figura 4.8: Gráfico da função $p_v(t)$ com $n = 4$ para $k = 2$ e $k = 4$

Os desenvolvimentos da próxima seção irão demonstrar a necessidade da existência de uma vizinhança de $t = 1$ com penalidade desprezível. Apesar das boas características apresentadas por esta última versão de p_v , notamos que, para t próximo de 1, esta função assume valores muito próximos de zero, mas nunca iguais a zero. Isto poderia trazer dificuldades quanto à definição da vizinhança de $t = 1$ com penalização desprezível, pois o valor ideal para uma penalização desprezível é zero, se isto for possível.

Podemos atingir este objetivo por intermédio das seguintes transformações da função p_v com $k = 2$:

1. uma translação vertical negativa de módulo δ ,
2. eliminação (“corte”) da parte negativa,
3. elevação ao quadrado para assegurar a diferenciabilidade da função resultante.

Os passos 2 e 3 são procedimentos padrões na criação de funções que implementam barreiras e penalidades (LUENBERGER e YE, 2008). Desta forma, iremos definir uma nova função, que denominaremos p_u , pela expressão

$$p_u(t) = [\max(0, p_v(t) - \delta)]^2, \quad (4.27)$$

onde δ é uma constante a ser determinada em função da largura almejada para a região com penalização zero. Conforme mencionado acima, utilizamos a função p_v , na definição de p_u , fazendo $k = 2$.

Para evitar confusões com os parâmetros da função p_v que dá origem à função

p_u , denominaremos

$$\begin{aligned} L_2 &= p_u(0), \\ E_2 &= \lim_{t \rightarrow \infty} p_u(t), \end{aligned}$$

os quais devem guardar um relacionamento direto com os níveis E e L da função p_v com $k = 2$ utilizada na criação de p_u . Observemos que

$$L_2 = (L - \delta)^2 \Rightarrow L = \sqrt{L_2} + \delta, \quad (4.28)$$

e

$$E_2 = (E - \delta)^2 \Rightarrow E = \sqrt{E_2} + \delta. \quad (4.29)$$

Estes valores de E e L são utilizados para construir uma função p_v que, mediante Definição 4.27, é empregada na construção da função p_u com níveis E_2 e L_2 . É interessante observar que a elevação ao quadrado presente na definição de p_u faz com que esta, para valores pequenos de δ , tenha um comportamento muito próximo ao da função p_v com $k = 4$, $E = E_2$ e $L = L_2$. De fato, se tomarmos $\delta = 0$ e construirmos uma versão de p_v com $k = 2$, $E = \sqrt{E_2}$ e $L = \sqrt{L_2}$, teremos $p_u(t) = [p_v(t)]^2$.

Se agora considerarmos os pontos $0 < t_1 < 1$ e $t_2 > 1$ como extremos do intervalo onde a penalidade é zero, podemos facilmente obtê-los resolvendo $p_v(t) = \delta$, resultando em

$$t_1 = \frac{\frac{1}{\sqrt[n]{1+\sqrt{\delta/E}}} - b_2}{1 - b_2}$$

e

$$t_2 = \frac{\frac{1}{\sqrt[n]{1-\sqrt{\delta/E}}} - b_2}{1 - b_2},$$

onde b_2 é calculado com base nas definições de E e L dadas por (4.28) e (4.29). Fazendo uma simples manipulação algébrica, considerando-se a expressão completa de b_2 , obtemos

$$\frac{t_2}{t_1} = \frac{\sqrt[n]{\frac{1+\sqrt{L/E}}{1-\sqrt{\delta/E}}} - 1}{\sqrt[n]{\frac{1+\sqrt{L/E}}{1+\sqrt{\delta/E}}} - 1}.$$

Para $\delta = 0$ temos a igualdade $t_1 = t_2 = 1$, em consequência, o intervalo fica reduzido a um ponto, como esperado.

A Figura 4.9 apresenta o gráfico de p_u , para $n = 4$. Os pontos t_1 e t_2 estão destacados no eixo t .

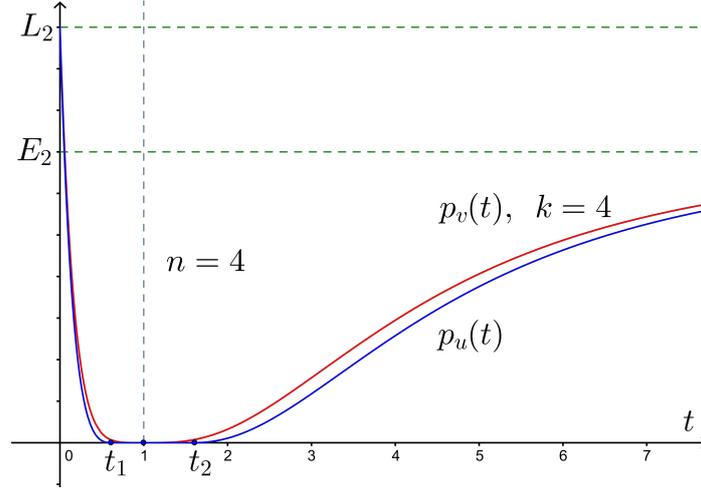


Figura 4.9: Gráfico da função $p_u(t)$, comparada com a função $p_v(t)$ para $k = 4$, ambas com $n = 4$

A função p_u será nossa escolha para a resolução dos problemas-nó que, portanto, passarão a ser expressos por

$$(PN4) : \max C_k - \sum_{\substack{i,j \in I_k \\ i < j}} p_u \left(\frac{\|x^i - x^j\|}{d_{ij}} \right) \quad (4.30)$$

$$\text{s.a.} \quad x \in T^n \quad (4.31)$$

Se permitirmos que a distância entre os centros x^i e x^j das esferas S_i e S_j , $1 < i < j < n$, possa variar com penalidade zero dentro de um intervalo dado por

$$m_{ij} \leq \|x^i - x^j\| \leq M_{ij}, \quad (4.32)$$

basta encontrarmos δ^* tal que

$$\frac{t_2}{t_1} = \frac{M_{ij}}{m_{ij}}. \quad (4.33)$$

De posse de δ^* , por proporcionalidade, temos que

$$\frac{d_{ij}}{m_{ij}} = \frac{1}{t_1} \Rightarrow d_{ij} = \frac{m_{ij}}{t_1},$$

e

$$\frac{M_{ij}}{d_{ij}} = \frac{t_2}{1} \Rightarrow d_{ij} = \frac{M_{ij}}{t_2}.$$

Assim podemos definir d_{ij} para ser utilizado na função objetivo, conforme (4.30).

4.4 Valores de α_{ij}

Na seção anterior desenvolvemos uma heurística, baseada nos métodos de penalização, que tem por objetivo posicionar duas esferas dentro de um intervalo de distâncias permitidas. A penalização é zero quando as distâncias estão dentro desta faixa e aumenta gradualmente quando as esferas começam a se afastar além dos extremos inferior ou superior deste intervalo. A determinação dos extremos deste intervalo é a questão central desta seção.

4.4.1 Esferas S_i e S_j possuem o mesmo raio

O Lema 4.1 garante a existência dos parâmetros $\{\alpha_{ij} \geq 0\}_{1 \leq i < j \leq n}$ e $\{c_i \geq 0\}_{1 \leq i \leq n}$ para os quais uma solução ótima de (4.2)-(4.5) é uma solução ótima para o problema clássico de recobrimento ótimo. Porém, esses valores não são conhecidos. Um dos objetivos deste trabalho é determinar bons valores para α_{ij} , o parâmetro que mede o quanto duas esferas se sobrepõem. No caso do recobrimento do \mathbb{R}^3 por esferas idênticas, o recobrimento reticulado (Figura 3.2b) mais eficiente e o único localmente ótimo entre os arranjos regulares (CONWAY e SLOANE, 1999) é o *reticulado cúbico de corpo centrado*. Embora, neste trabalho, tratemos de esferas de raios diferentes, os valores de α_{ij} para este recobrimento servirão como ponto de partida para a determinação de um intervalo de valores que serão utilizados na resolução do problema proposto.

A seguir demonstraremos que, nesse recobrimento, $\alpha_{ij} = 0,2111R$ ou $\alpha_{ij} = 0,4508R$ quando há sobreposição entre duas esferas. Sejam x^i o centro da esfera S_i , R o raio das esferas, d_{ij} a distância entre os centros das esferas S_i e S_j e definida por α_{ij} o parâmetro que mede o quanto as esferas se sobrepõem, conforme indicado na Figura 4.1.

Como os centros das esferas formam um reticulado cúbico de corpo centrado (ver Seção 3.1.4), cada centro x^i está exatamente no centro de um octaedro truncado inscrito na esfera S_i (Figura 3.20b). Para uma melhor compreensão, imaginemos o espaço \mathbb{R}^3 preenchido por octaedros truncados idênticos (Figuras 3.11b e 3.20a) e cada octaedro circunscrito por uma esfera (Figura 3.20b). Tomemos duas esferas adjacentes neste reticulado, S_i e S_j , que se sobrepõem. Para determinarmos o valor de α_{ij} , usando a definição, primeiro devemos calcular a distância d_{ij} . Mas d_{ij} também é a distância entre os centros dos dois octaedros truncados circunscritos às esferas S_i e S_j . Observe que, passando um plano pelos pontos comuns às superfícies dessas duas esferas obtemos uma seção plana circular. Esta seção está circunscrita à face comum aos dois octaedros truncados inscritos nestas duas esferas. Estes sólidos têm dois tipos de faces, quadrados e hexágonos regulares. Como os lados dos quadrados e dos hexágonos regulares são congruentes, a área dessa seção circular

terá dois valores distintos, dependendo do polígono inscrito ser um quadrado ou um hexágono regular. A seção circular tem área maior se a face de interseção entre os dois octaedros truncados for um hexágono regular. Isto implica que a sobreposição das esferas é maior neste caso.

Voltemos ao cálculo de d_{ij} . Para determinar a distância entre dois octaedros truncados adjacentes basta determinar o dobro da distância do centro de um deles ao centro da face comum a eles.

Seja l a medida das arestas do octaedro truncado. Suponhamos, primeiro que a interseção entre dois octaedros truncados adjacentes seja uma face quadrada.

O octaedro truncado de aresta l pode ser obtido pelo truncamento de um octaedro regular de lado $3l$ ou através do cubo circunscrito de lado $2\sqrt{2}l$, conforme mostrado na Figura 4.10.

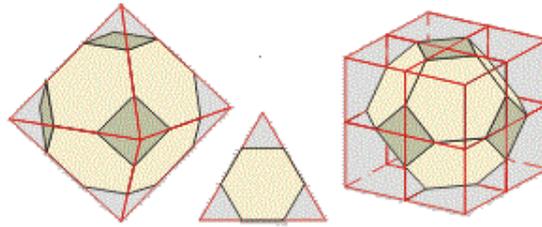


Figura 4.10: Octaedro truncado.

Consideremos o octaedro truncado construído a partir do cubo de lado $2\sqrt{2}l$. Tomemos uma seção plana do octaedro truncado, passando pelo seu centro e pelas diagonais das quatro faces quadradas (Figura 4.10). A seção plana obtida é um octógono de lados l e $l\sqrt{2}$. A Figura 4.11 mostra o octógono inscrito na circunferência de raio R e no quadrado de lado $2\sqrt{2}l$ (observe que o quadrado está dividido em quadrados menores de lado $\frac{l\sqrt{2}}{2}$).

A distância entre os centros destes octaedros é $2\sqrt{2}l$, e portanto, os centros das esferas também distam $d_{ij} = 2\sqrt{2}l$. Como $d_{ij} = 2R - \alpha_{ij}$, então $\alpha_{ij} = 2R - d_{ij}$. Logo,

$$\alpha_{ij} = 2R - 2\sqrt{2}l. \quad (4.34)$$

Aplicando o Teorema de Pitágoras ao triângulo retângulo de catetos $l\sqrt{2}$ e $\frac{l\sqrt{2}}{2}$, e hipotenusa R , temos que:

$$l = \frac{\sqrt{10}}{5}R \quad (4.35)$$

Substituindo 4.35 em 4.34 obtemos:

$$\alpha_{ij} = \frac{10 - 4\sqrt{5}}{5}R = 2 \left(1 - \frac{2\sqrt{5}}{5} \right) R$$

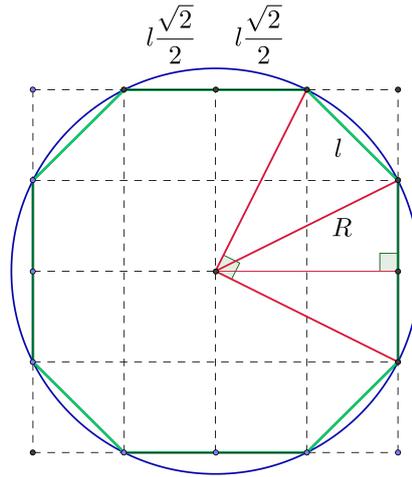


Figura 4.11: Seção plana do octaedro truncado de lado l inscrito na esfera de raio R .

Neste caso, chamemos α_{ij} de α_m , isto é, se a interseção entre dois octaedros truncados adjacentes for uma face quadrada. Logo,

$$\alpha_m = 2 \left(1 - \frac{2\sqrt{5}}{5} \right) R$$

ou

$$\alpha_m \cong 0,2111R.$$

Se $R = 1$, então $\alpha_m \cong 0,2111$.

Analogamente, suponhamos que a interseção entre dois octaedros truncados adjacentes seja uma face hexagonal. Dividimos o cubo da Figura 4.10 em oito cubos menores idênticos de lado $l\sqrt{2}$. Cada face hexagonal é uma seção plana de um cubo menor (4.12).

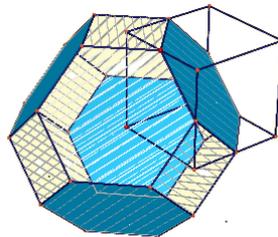


Figura 4.12: A face hexagonal do octaedro truncado de lado l é uma seção plana de um cubo de lado $l\sqrt{2}$.

Os vértices do hexágono regular são pontos médios das arestas do cubo menor. Eles são equidistantes do centro do octaedro truncado e do vértice do cubo oposto a este ponto. Então, os vértices do hexágono pertencem ao plano bissetor que é perpendicular à diagonal do cubo menor passando pelo centro deste. Portanto, o

centro do hexágono regular coincide com o ponto médio da diagonal do cubo menor. A diagonal do cubo menor mede $l\sqrt{2}\sqrt{3} = l\sqrt{6}$. Assim, a distância entre o centro do octaedro truncado e uma de suas faces hexagonais é $\frac{l\sqrt{6}}{2}$. A distância entre os centros dos dois octaedros truncados é $l\sqrt{6}$ e, conseqüentemente, os centros das esferas distam $d_{ij} = l\sqrt{6}$. Então, $\alpha_{ij} = 2R - d_{ij}$, e usando a equação (4.35) obtemos:

$$\alpha_{ij} = \alpha_M = \frac{10 - 2\sqrt{15}}{5}R = 2 \left(1 - \frac{\sqrt{15}}{5} \right) R$$

Neste caso, chamemos α_{ij} de α_M , isto é, se a interseção entre dois octaedros truncados adjacentes for uma face hexagonal regular. Logo,

$$\alpha_M = 2 \left(1 - \frac{\sqrt{15}}{5} \right) R$$

ou

$$\alpha_M \cong 0,4508R.$$

Se $R = 1$, então $\alpha_M \cong 0,4508$.

Assim, dependendo da interseção entre os octaedros truncados inscritos nas esferas serem faces quadradas ou hexagonais, temos dois valores distintos para α_{ij} , $\alpha_m = 0,2111R$ e $\alpha_M = 0,4508R$, respectivamente.

4.4.2 Esferas S_i e S_j possuem raios diferentes

Não existe nenhuma teoria geral que determine de forma exata qual deve ser a distância entre duas esferas de raios diferentes quando estas são utilizadas em um recobrimento. O problema original ($P1$) inclui uma restrição sobre a distância mínima em termos de α_{ij} , mas não especifica como estes parâmetros devem ser calculados.

Vimos na seção anterior que, no caso de esferas de mesmo raio, o recobrimento mais eficiente do \mathbb{R}^3 é conseguido com os centros das esferas formando o reticulado cúbico de corpo centrado, o que nos fornece dois valores distintos para α_{ij} . Para recobrir um sólido T utilizando esferas de raios diferentes devemos utilizar uma abordagem que seja compatível com a situação dos raios iguais.

Se, momentaneamente, considerarmos o problema de recobrir o \mathbb{R}^2 com círculos de mesmo raio, o melhor arranjo para os seus centros é o reticulado hexagonal (ver Seção 3.5). Consideremos agora uma transformação conforme aplicada sobre o plano \mathbb{R}^2 contendo estes círculos. Sabemos que uma transformação conforme é caracterizada pela preservação de ângulos entre curvas e, em particular, transforma círculos em círculos¹. Em geral, os centros dos círculos transformados não são, no

¹Uma reta pode ser interpretada como um círculo com centro no infinito

entanto, o mapeamento dos centros dos círculos originais.

A transformação conforme dada pela função complexa

$$f(z) = \frac{z}{cz + 1}$$

possui a conveniente propriedade de ser a função identidade para $c = 0$. Quando c aumenta, esta função passa a transformar círculos em proporções diferentes dependendo de suas posições no \mathbb{R}^2 , de tal forma que os círculos originais, de raios iguais, são transformados em círculos com raios diferentes. Os pontos de interseção dos círculos são transformados, mas a propriedade de interseção de três círculos é preservada e, portanto, os círculos transformados ainda são um recobrimento do \mathbb{R}^2 .

Para ilustrar a ação desta função, na Figura 4.13 apresentamos uma parte do arranjo hexagonal composta de sete círculos no \mathbb{R}^2 , enquanto na Figura 4.14 apresentamos a ação desta função sobre estes mesmos círculos.

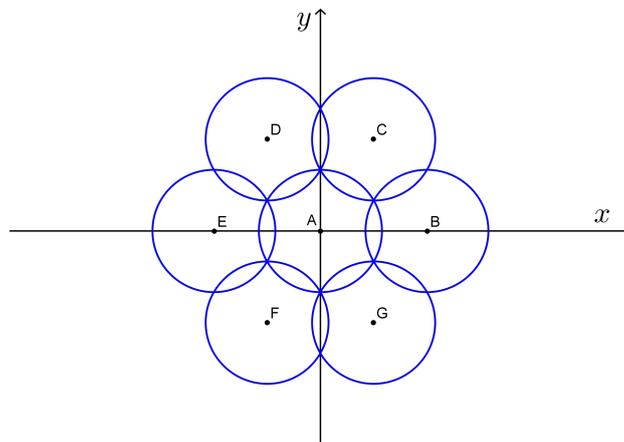


Figura 4.13: Disposição hexagonal de sete círculos recobrindo uma região do \mathbb{R}^2 .

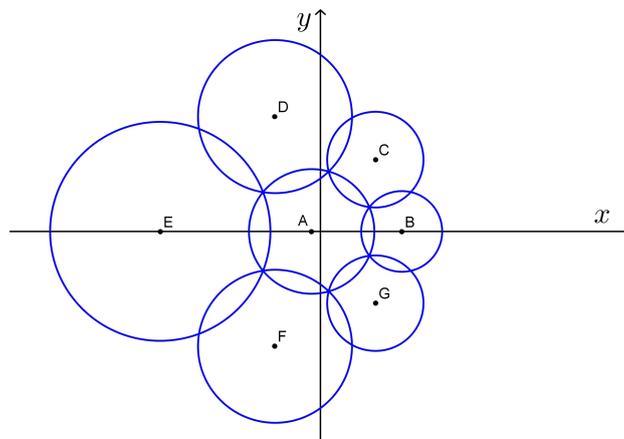


Figura 4.14: A mesma disposição hexagonal da anterior, após aplicação da transformação conforme no \mathbb{R}^2 .

Devemos notar que, apesar de estarmos utilizando uma função complexa para

representar nossa transformação conforme, nosso interesse está centrado na transformação em si e em suas propriedades gerais, o que nos permitirá estendermos nosso raciocínio para o espaço tridimensional.

Considere mais uma vez um arranjo de círculos de raio R onde seus centros formam um reticulado hexagonal. Sejam C_i e C_j dois círculos distintos sobrepostos. A transformação conforme preserva o ângulo $\theta_2 = 120^\circ$ formado entre um dos pontos de interseção de suas circunferências e seus centros. A distância entre os centros dos círculos transformados utilizando a lei dos cossenos, se seus raios forem conhecidos. Assim, para os círculos transformados C'_i e C'_j , com raios r_i e r_j , temos que a distância entre eles é dada por

$$d_{ij}^2 = r_i^2 + r_j^2 - 2r_i r_j \cos(120^\circ) = r_i^2 + r_j^2 + r_i r_j. \quad (4.36)$$

Se utilizarmos a definição de α_{ij} para estes círculos (sobrepostos), conforme (4.1), temos

$$d_{ij} = r_i + r_j - \alpha_{ij}. \quad (4.37)$$

Utilizando as equações acima, podemos, então, escrever

$$\alpha_{ij} = r_i + r_j - d_{ij} = r_i + r_j - \sqrt{r_i^2 + r_i r_j + r_j^2}.$$

É claro que, nesta situação, torna-se mais simples utilizar diretamente a expressão (4.36) para calcular d_{ij} , sem envolver diretamente α_{ij} .

Em um recobrimento bidimensional com círculos de raios diferentes, poderíamos, inspirados pela transformação conforme acima descrita, definir heurísticamente que a distância ideal entre dois círculos quaisquer C_i e C_j , com raios r_i e r_j , deveria ser a descrita por (4.36). Em particular, se os raios fossem todos iguais a R , teríamos que a distância ideal deveria ser dada por

$$d_{ij}^2 = R^2 + R^2 + RR \implies d_{ij} = \sqrt{3}R,$$

e os círculos sobrepostos seriam forçados a assumir esta distância, mediante o processo de otimização baseado nos desenvolvimentos da seção anterior.

Retornando ao espaço tridimensional, um raciocínio semelhante pode ser aplicado inicialmente a um arranjo de quatro esferas S_i , $i = 1, \dots, 4$, de mesmo raio R , dispostas nos vértices de um tetraedro regular de lado l , o qual representa a distância comum entre duas esferas quaisquer. O ponto interior mais distante de todos os vértices está localizado no centro deste tetraedro. Para que este ponto seja coberto por esferas iguais é necessário que o raio destas esferas seja pelo menos igual à distância deste ponto a um dos vértices. Podemos expressar o lado l deste

tetraedro em função do raio das esferas:

$$l = \frac{2\sqrt{6}}{3}R,$$

que também pode ser escrito como

$$l^2 = \frac{8R^2}{3} = R^2 + R^2 - 2R^2 \left(-\frac{1}{3}\right).$$

Comparando esta expressão com a lei dos cossenos, chegamos à conclusão que o ângulo θ_3 do centro do tetraedro à dois de seus vértices é

$$\theta_3 = \arccos \left(-\frac{1}{3}\right) \cong 109,47^\circ.$$

Nesta configuração bastante regular para as esferas, mais uma vez utilizando a expressão(4.1) temos

$$\alpha_{ij} = R + R - \frac{2\sqrt{6}}{3}R = 2 \left(1 - \frac{\sqrt{6}}{3}\right) R \cong 0,3670R.$$

Por intermédio de racionalizações, podemos facilmente comprovar que

$$2 \left(1 - \frac{2\sqrt{5}}{5}\right) R < 2 \left(1 - \frac{\sqrt{6}}{3}\right) R < 2 \left(1 - \frac{\sqrt{15}}{5}\right) R,$$

ou seja,

$$\alpha_m < \alpha_{ij} < \alpha_M. \quad (4.38)$$

Temos assim um valor para α_{ij} que está entre os limites α_m e α_M obtidos no desenvolvimento da seção anterior para o recobrimento reticulado de corpo centrado, considerado o mais eficiente e o único localmente ótimo entre os arranjos regulares. Utilizar o arranjo tetraedral como uma disposição intermediária para esferas iguais durante o processo de otimização é, portanto, totalmente compatível com as restrições oriundas do recobrimento reticulado de corpo centrado.

Seguindo os passos do desenvolvimento para o espaço bidimensional, considere-mos agora uma transformação conforme² deste arranjo de quatro esferas. O ângulo θ_3 entre o centro do tetraedro à dois vértices quaisquer permanece o mesmo. Utilizando a lei dos cossenos obtemos

$$d_{ij}^2 = r_i^2 + r_j^2 - 2r_i r_j \left(-\frac{1}{3}\right) = r_i^2 + r_j^2 + \frac{2}{3}r_i r_j. \quad (4.39)$$

²As transformações conformes no \mathbb{R}^n , para $n > 2$, são restritas às transformações de Möbius, segundo o Teorema de Liouville encontrado na geometria conforme.

Para esta configuração de esferas, temos que

$$\alpha_{ij} = r_i + r_j - d_{ij} = r_i + r_j - \sqrt{r_i^2 + r_j^2 + \frac{2}{3}r_i r_j}.$$

Em termos de implementação computacional, mais uma vez notamos ser mais conveniente utilizar diretamente a expressão (4.39) para calcular d_{ij} , sem envolver diretamente α_{ij} .

Em um recobrimento tridimensional com esferas de raios diferentes, iremos, com base nos resultados da transformação conforme no \mathbb{R}^3 acima descritos, definir heurísticamente que a distância ideal entre duas esferas quaisquer S_i e S_j , com raios r_i e r_j , deve ser a descrita por (4.39). Devemos lembrar que esta distância d_{ij} não será, na verdade, privilegiada, pois conforme (4.32), faremos com que todas as distâncias dentro do intervalo $[m_{ij}, M_{ij}]$ tenham penalização nula. As desigualdades (4.38) não só confirmam a existência deste intervalo $[m_{ij}, M_{ij}]$, como também que d_{ij} dado por (4.39) pertence a ele. Notemos que α_M está associado à determinação de m_{ij} e α_m está associado à determinação de M_{ij} , pois queremos $m_{ij} < M_{ij}$, mas as distâncias d_{ij} e os α_{ij} variam em sentidos opostos.

Devemos agora determinar os valores m_{ij} e M_{ij} utilizando a transformação conforme. Para tal, calculemos os cossenos dos ângulos θ_m e θ_M associados a α_M e α_m , respectivamente, utilizando inicialmente esferas de mesmo raio R . Temos

$$m_{ij} = 2R - \alpha_M = 2R \left(\frac{\sqrt{15}}{5} \right),$$

logo

$$4R^2 \frac{3}{5} = m_{ij}^2 = 2R^2 - 2R^2 \cos \theta_m \Rightarrow \cos \theta_m = -\frac{1}{5}.$$

Da mesma forma,

$$M_{ij} = 2R - \alpha_m = 2R \left(\frac{2\sqrt{5}}{5} \right),$$

logo

$$4R^2 \frac{4}{5} = M_{ij}^2 = 2R^2 - 2R^2 \cos \theta_M \Rightarrow \cos \theta_M = -\frac{3}{5}.$$

Se agora considerarmos esferas de raios diferentes dados por r_i e r_j , aplicamos a transformação conforme e, como os ângulos permanecem inalterados, podemos definir:

$$\begin{aligned} m_{ij}^2 &= r_i^2 + r_j^2 - 2r_i r_j \cos \theta_m = r_i^2 + r_j^2 + \frac{2}{5}r_i r_j \\ M_{ij}^2 &= r_i^2 + r_j^2 - 2r_i r_j \cos \theta_M = r_i^2 + r_j^2 + \frac{6}{5}r_i r_j \end{aligned}$$

Para utilizarmos a função p_u da seção anterior precisamos antes encontrar um δ^* tal que, conforme (4.33),

$$\frac{t_2}{t_1} = \frac{M_{ij}}{m_{ij}} = \sqrt{\frac{r_i^2 + r_j^2 + \frac{6}{5}r_i r_j}{r_i^2 + r_j^2 + \frac{2}{5}r_i r_j}} = \sqrt{1 + \frac{4r_i r_j}{5(r_i^2 + r_j^2) + 2r_i r_j}},$$

que pode ser simplificado para

$$\frac{t_2}{t_1} = \sqrt{1 + \frac{4}{5\left(\frac{r_i}{r_j} + \frac{r_j}{r_i}\right) + 2}}$$

Assumindo que os raios das esferas utilizadas são todos conhecidos, os quocientes acima podem ser pré-calculados, assim como os valores correspondentes de δ^* , t_2 e t_1 , o que finalmente permite o cálculo do d_{ij} que irá ser utilizado pela função p_u de cada nó, conforme formulação (PN4).

4.5 Região de segurança

Discutiremos agora mais detalhadamente a restrição (4.31). Esta restrição obriga o centro de cada esfera a permanecer dentro do sólido T a ser recoberto. Em um primeiro momento, com base nos desenvolvimentos da última seção, iremos definir, para $z \in \mathbb{R}^3$, a seguinte função:

$$q(z) = \begin{cases} 0 & , z \in T, \\ +\infty & , z \notin T. \end{cases} \quad (4.40)$$

Esta função é chamada *função característica* do conjunto T e comumente representada por χ_T . A restrição $x \in T^n$ pode, então, ser substituída pelas restrições $q(x^i) \leq 0$, $i = 1, \dots, n$.

Se o sólido T for convexo, $q(z)$ é uma função convexa e estas restrições são, portanto, todas convexas. No caso geral, em que o sólido T é não-convexo, este conjunto de restrições é não-convexo, constituindo-se, então, em uma fonte de não-convexidade que denominamos NC_3 .

Com relação à aplicação desta modelagem no tratamento por *Gamma Knife*, onde o sólido T representa um tumor cerebral, é de praxe utilizar técnicas que não permitam uma ocupação muito grande da parte externa de T pelas esferas que modelam as doses de radiação aplicada. Essa medida de segurança é tomada porque o tumor está envolvido por tecidos saudáveis e não é desejável que estes tecidos sejam atingidos pela radiação. Com o propósito de obter uma solução com as esferas posicionadas de forma a minimizar a irradiação de tecidos saudáveis, restringeremos

a ocupação da região externa a T através de um parâmetro de tolerância $\varepsilon > 0$ e denominaremos a região delimitada de *região de segurança*, conforme representado na Figura 4.15.

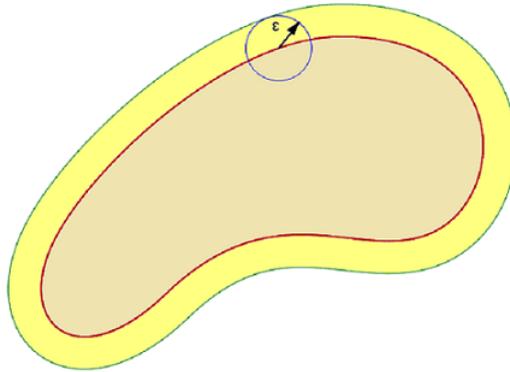


Figura 4.15: Representação da região de segurança para o sólido T , onde ε é o parâmetro de tolerância.

Passaremos, portanto, a exigir que somente os pontos situados a uma distância euclidiana de no máximo ε unidades da superfície de T possam sofrer os efeitos da radiação. Os pontos situados além desta fronteira não devem sofrer nenhum efeito da radiação.

Visando a construção de um modelo matemático para um tumor, consideremos agora uma região genérica $T \subset \mathbb{R}^3$ compacta³, conexa⁴, com dimensão uniforme igual a três⁵, e com fronteira conexa homeomorfa à esfera $S^2 = \{(u, v, w) \in \mathbb{R}^3 \mid u^2 + v^2 + w^2 = 1\}$. Estas condições tem por objetivo excluir modelos de tumores tais como:

1. tumores “ilimitados” ou sem fronteira definida (na prática eles não existem);
2. tumores formados por mais de uma “peça” (cada peça deve ser tratada separadamente);
3. tumores com partes “planas” ou filamentosas (não podem ser tratados por nossa abordagem);
4. tumores multi-toroidais (com uma ou mais “alças”) ou com cavidades em seu interior (não podem ser tratados diretamente por nossa abordagem).

³Região limitada e fechada.

⁴Mais precisamente, conexa por caminhos, ou seja, dados dois pontos P_1 e P_2 em T , existe uma função contínua $\varphi : [0, 1] \rightarrow T$, com $\varphi(0) = P_1$ e $\varphi(1) = P_2$, que conecta estes dois pontos.

⁵Todo ponto no interior de T possui uma vizinhança homeomorfa ao interior da bola B^3 .

Uma propriedade a princípio não requerida é a convexidade de T . É claro que o modelo matemático de um tumor pode resultar em uma região não-convexa. Com o propósito de simplificarmos análises posteriores, iremos trabalhar somente com sólidos convexos, embora algumas destas análises possam ser adaptadas para o caso de T não-convexo⁶. Assim sendo, não iremos lidar explicitamente com a fonte de não-convexidade NC_3 .

4.5.1 Corpo paralelo ao sólido T

Pela definição usual de distância a um conjunto, pontos no exterior de T estarão a uma distância positiva de T e pontos no interior de T estarão a uma distância nula de T , ou seja:

$$d(x, T) = \min_{z \in T} d(x, z),$$

onde utilizamos “min” pois T é compacto.

Denotaremos por ∂T a superfície que representa a fronteira de T , e, como já mencionado acima, assumiremos que a fronteira ∂T é conexa.

As superfícies paralelas a ∂T serão denotadas por $(\partial T)_d$, seguindo a notação apresentada no Apêndice A, onde d pode assumir qualquer valor real, as quais servirão para definir a fronteira da região de segurança, assim como outras superfícies associadas.

Devemos observar que ∂T (ou uma de suas superfícies paralelas $(\partial T)_d$) pode ser uma superfície regular (vide Definição A.1). Em análises posteriores, iremos separar os casos em que estas superfícies são regulares dos casos em que elas não são.

Para ∂T regular e T não necessariamente convexo, podemos definir

$$\begin{aligned} d_{\text{inf}} &= \arg \inf_d \{(\partial T)_d \text{ é regular}\}, \\ d_{\text{sup}} &= \arg \sup_d \{(\partial T)_d \text{ é regular}\}, \end{aligned}$$

onde em geral teremos $d_{\text{inf}} < 0$ e $d_{\text{sup}} > 0$. Para T não-convexo existirão valores de $d > 0$ para os quais a superfície $(\partial T)_d$ apresentará auto-interseções, deixando de ser regular, e portanto teremos $0 \leq d_{\text{sup}} < +\infty$. Caso T seja convexo teremos $(\partial T)_d$ regular para todo $d > 0$ e, portanto, $d_{\text{sup}} = +\infty$.

Para ∂T não necessariamente regular, mas retornando a considerar T convexo, temos que as superfícies $(\partial T)_d$ serão externas a T se $d > 0$, coincidirão com ∂T se $d = 0$, e serão internas, parcialmente internas ou até mesmo externas a T dependendo do valor de d quando $d < 0$. Notamos que as superfícies $(\partial T)_d$ podem ser bastante complexas quando $d < 0$.

⁶Por exemplo, quando o sólido pode ser dividido em sub-regiões convexas.

Por outro lado, para todo $d \geq 0$ e para $d < 0$ próximo de zero, as superfícies paralelas à distância d são de fácil compreensão e iremos definir T_d como sendo a região limitada que tem por fronteira $(\partial T)_d$. Podemos, então, escrever $\partial T_d = (\partial T)_d$, ou seja, a superfície fronteira de T_d é a superfície paralela à distância d de T . Chamaremos T_d de *corpo paralelo* a T à distância d .

Em especial, temos que T_ε , o corpo paralelo a T à distância $\varepsilon > 0$, pode ser interpretado como a união de T com a sua região de segurança, e identificamos $\partial T_\varepsilon = (\partial T)_\varepsilon$. Dada a convexidade de T , temos que não existe nenhuma restrição sobre o maior valor que ε possa assumir⁷.

Iremos agora estender a definição de corpo paralelo para qualquer $d < 0$. Primeiramente, para um ponto $x \in T$ arbitrário, temos

$$d(x, \partial T) = \min_{z \in \partial T} d(x, z),$$

onde utilizamos “min” pois como T é compacto, ∂T também é compacta.

Dado $d < 0$, temos que $d(x, \partial T) = |d|$ implica, por definição, que $x \in (\partial T)_d$. A implicação inversa, no entanto, não é válida, isto é, $x \in (\partial T)_d$ não implica em $d(x, \partial T) = |d|$. Isto porque, em geral, x irá pertencer a mais de uma superfície $(\partial T)_d$, para valores de d distintos. Na Figura 4.16 apresentamos a seção transversal de um elipsóide prolato T e um ponto x em seu interior, juntamente com os traços de três superfícies paralelas às quais x pertence.

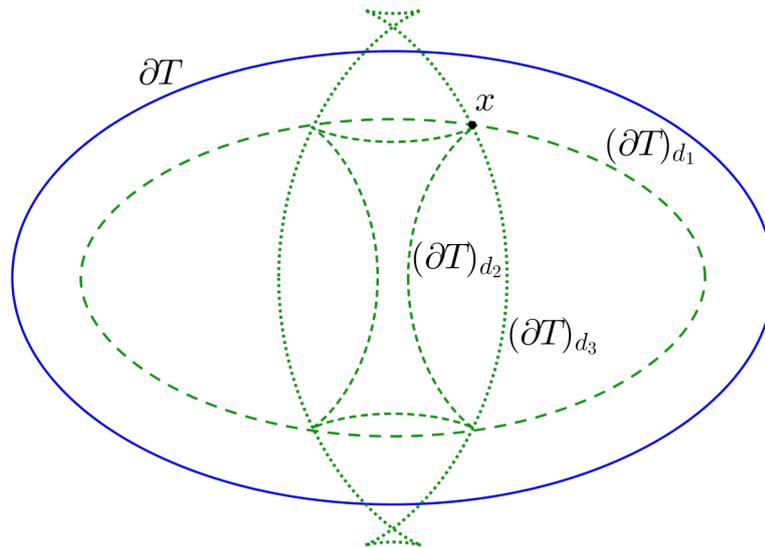


Figura 4.16: Exemplo onde $x \in T$ pertence a três superfícies paralelas distintas.

Partindo da definição preliminar de corpo paralelo apresentada acima, tomaremos as seções de superfícies $(\partial T)_d$ com os valores de $d < 0$ mais próximos de zero.

⁷Para T não-convexo existirão valores de $\varepsilon > 0$ para os quais ∂T_ε apresentará auto-interseções.

Podemos fazer isto se definirmos a função $D : T \rightarrow (-\infty, 0]$ como

$$D(x) = \arg \max_{x \in (\partial T)_d} d \quad (4.41)$$

Pelo exposto acima, temos claramente que

$$|D(x)| = \left| \arg \max_{x \in (\partial T)_d} d \right| = \arg \min_{x \in (\partial T)_d} |d| = d(x, \partial T),$$

pois uma das superfícies $(\partial T)_d$ será justamente aquela para a qual $|d|$ será igual a $d(x, \partial T)$. Como $x \in T$ implica em $D(x) \leq 0$, esta última expressão pode ser reescrita como

$$D(x) = -d(x, \partial T) = -\min_{z \in \partial T} d(x, z) = \max_{z \in \partial T} -d(x, z). \quad (4.42)$$

Observemos que a função $D(x)$ determina o valor de d correspondente à superfície $(\partial T)_d$ mais próxima a ∂T , eliminando assim as multiplicidades oriundas do fato de x pertencer a mais de uma superfície $(\partial T)_d$.

A função $D(x)$ que acabamos de definir, acima, goza das seguintes propriedades:

Propriedades da função $D(x)$

1. Se $x \in \partial T$, então $D(x) = 0$.

Óbvio, pois para $x \in \partial T$, $D(x) = -d(x, \partial T) = 0$.

2. A função $D(x)$ é contínua.

Por definição, $D(x) = \max_{z \in \partial T} -d(x, z)$. Como a função distância é contínua e o máximo de um conjunto qualquer de funções contínuas também é contínuo, temos que $D(x)$ é contínua.

3. A função $D(x)$ atinge os valores máximo e mínimo em T .

Como T é compacto e $D(x)$ é contínua, pelo Teorema do Valor Extremo⁸ temos que $D(x)$ atinge os valores mínimo e máximo em T . Como $D(x) \leq 0$ e $D(x) = 0$ para $x \in \partial T$, então $D(x)$ assume o valor máximo igual a zero na fronteira ∂T . Para qualquer outro x no interior de T , teremos $d(x, \partial T) > 0$ e, portanto, ∂T é exatamente o conjunto de todos os pontos onde $D(x)$ atinge o valor máximo. Quanto ao valor mínimo de d , denominá-lo-emos d^* . Portanto, a função $D(x)$ pode ser melhor representada como uma função $D : T \rightarrow [d^*, 0]$, onde d^* depende do sólido T .

Provaremos a seguir um importante resultado sobre a função $D(x)$.

⁸O Teorema do Valor Extremo da Análise Real afirma que toda função contínua definida em um compacto atinge seus valores máximo e mínimo.

Teorema 4.2. *Seja $T \subset \mathbb{R}^3$ convexo, compacto e com interior não vazio, e $F(x) = -D(x) = d(x, \partial T)$. Então, $F(x)$ é uma função côncava ($D(x)$ é uma função convexa).*

Demonstração. Sejam $x_1, x_2 \in T$, $d_1 = d(x_1, \partial T)$ e $d_2 = d(x_2, \partial T)$.

Para $0 \leq \lambda \leq 1$, sejam $x_\lambda = (1 - \lambda)x_1 + \lambda x_2$ e $d_\lambda = (1 - \lambda)d_1 + \lambda d_2$.

Temos que provar que

$$F((1 - \lambda)x_1 + \lambda x_2) \geq (1 - \lambda)F(x_1) + \lambda F(x_2),$$

ou

$$d((1 - \lambda)x_1 + \lambda x_2, \partial T) \geq (1 - \lambda)d(x_1, \partial T) + \lambda d(x_2, \partial T),$$

o que pode também ser escrito como

$$d(x_\lambda, \partial T) \geq d_\lambda.$$

Suponhamos, por absurdo, que $d(x_\lambda, \partial T) < d_\lambda$. Então, como ∂T é compacto, existe $y \in \partial T$ tal que $d(x_\lambda, y) = d(x_\lambda, \partial T)$, e, portanto, $d(x_\lambda, y) < d_\lambda$.

Definamos

$$y_1 = x_1 + (y - x_\lambda)d_1/d_\lambda$$

e

$$y_2 = x_2 + (y - x_\lambda)d_2/d_\lambda.$$

Temos

$$d(y_1, x_1) = d(y, x_\lambda)d_1/d_\lambda < d_\lambda d_1/d_\lambda = d_1.$$

Logo, y_1 está no interior da esfera $B(x_1, d_1)$ centrada em x_1 de raio d_1 e $B(x_1, d_1) \subset T$, pois d_1 é o maior valor d para o qual $B(x_1, d) \subset T$.

Da mesma forma,

$$d(y_2, x_2) = d(y, x_\lambda)d_2/d_\lambda < d_\lambda d_2/d_\lambda = d_2.$$

Logo, y_2 está no interior da esfera $B(x_2, d_2)$ e temos $B(x_2, d_2) \subset T$. Mas,

$$\begin{aligned} (1 - \lambda)y_1 + \lambda y_2 &= (1 - \lambda)x_1 + (y - x_\lambda)(1 - \lambda)d_1/d_\lambda + \lambda x_2 + (y - x_\lambda)\lambda d_2/d_\lambda \\ &= x_\lambda + (y - x_\lambda)d_\lambda/d_\lambda \\ &= y. \end{aligned}$$

Portanto, como y é uma combinação convexa de dois pontos que pertencem ao interior de T , temos que y também pertence ao interior de T . Mas isto é um absurdo,

pois supomos que $y \in \partial T$. Logo $d(x_\lambda, \partial T) \geq d_\lambda$, $F(x)$ é côncava e consequentemente $D(x)$ é convexa. \square

É possível estendermos o domínio de $D(x)$. Para qualquer $d > 0$ temos T_d bem definido e compacto. Podemos, então, estender $D(x)$ para $x \in T_d \setminus T$ definindo $D(x) = d(x, \partial T) > 0$. Este processo, sendo válido para qualquer $d > 0$, permite estender $D(x)$ para todo $x \in \mathbb{R}^3$. O mínimo desta função estendida permanece sendo d^* , pois os valores acrescentados à imagem de $D(x)$ são todos positivos.

Doravante iremos sempre utilizar esta última versão de $D(x)$, assim definida:

$$D(x) = \begin{cases} -d(x, \partial T) & , \text{ se } x \in T \\ d(x, \partial T) & , \text{ se } x \notin T \end{cases},$$

como uma função $D : \mathbb{R}^3 \rightarrow [d^*, +\infty)$. Podemos simplificar mais ainda a expressão de $D(x)$ se utilizarmos a função indicadora de um conjunto

$$I_T(x) = \begin{cases} 1 & , \text{ se } x \in T \\ 0 & , \text{ se } x \notin T \end{cases},$$

definindo $\varsigma_T(x) = 1 - 2I_T(x)$, então podemos escrever simplesmente que

$$D(x) = (1 - 2I_T(x))d(x, \partial T) = \varsigma_T(x)d(x, \partial T).$$

Utilizando a mesma argumentação apresentada para a versão original de $D(x)$, é imediato mostrarmos que $D(x)$ estendida permanece sendo uma função contínua.

Para $d > 0$ ou $d \leq 0$ próximo de zero, podemos afirmar, lembrando que T é convexo, que

$$(\partial T)_d = \{x \in \mathbb{R}^3 \mid D(x) = d\}.$$

Alguns parágrafos acima utilizamos a fronteira $(\partial T)_d$ para limitar uma região que definimos como sendo o corpo paralelo T_d . Utilizando diretamente $D(x)$, podemos agora redefinir T_d , para qualquer $d \in \mathbb{R}$, como sendo o conjunto de nível d da função $D(x)$, ou seja,

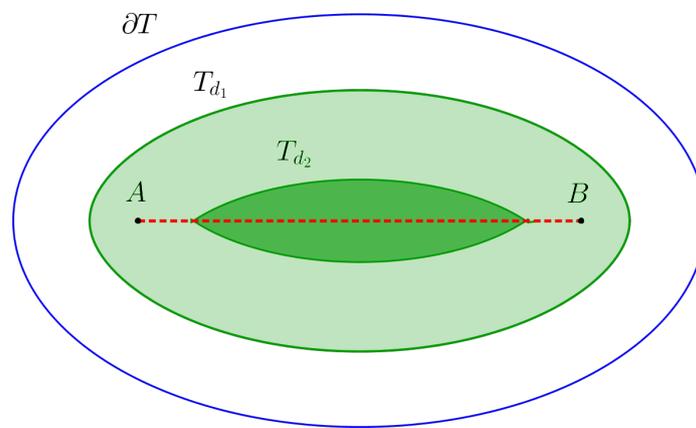
$$T_d = \{x \in \mathbb{R}^3 \mid D(x) \leq d\}.$$

Temos agora uma definição para o corpo paralelo que depende somente da função $D(x)$, e para o qual a igualdade $\partial T_d = (\partial T)_d$ somente é válida para $d > 0$ ou $d \leq 0$ próximo de zero. Doravante, estaremos constantemente nos referindo aos corpos paralelos T_d . É claro que $T_0 = T$, $T_d = \emptyset$ para $d < d^*$ e, mais uma vez, T_ε é a união de T com a sua região de segurança de largura $\varepsilon > 0$.

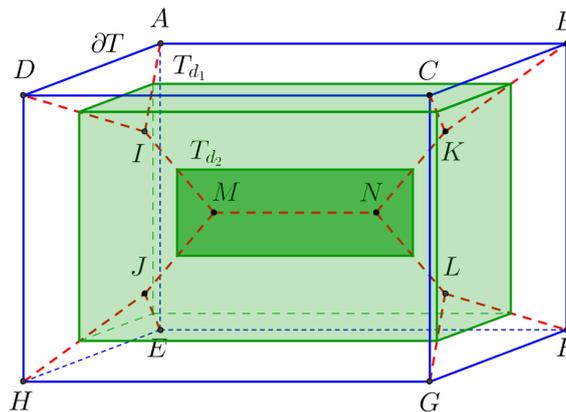
Demonstramos, acima, que $D(x)$ é contínua. Quanto à diferenciabilidade, é

possível que $D(x)$ seja não diferenciável em T mesmo com T convexo e ∂T regular. Um fato importante é dado pelo Teorema de Motzkin (Teorema 3.1), o qual afirma que, independentemente de ∂T ser regular ou não, T convexo implica em $D(x)$ ser sempre diferenciável em $\mathbb{R}^3 \setminus T$.

Na Figura 4.17a, apresentamos a seção transversal de um elipsóide prolato T e dois corpos paralelos às distâncias d_1 e d_2 . A fronteira ∂T é regular, mas a função $D(x)$ não é diferenciável no segmento \overline{AB} , pois neste segmento existe uma mudança brusca na região de ∂T utilizada para o cálculo das distâncias $d(x, \partial T)$. Notamos que a fronteira do corpo paralelo à distância d_1 é regular, enquanto que a fronteira do corpo paralelo à distância d_2 não é regular.



(a) Seção de um elipsóide prolato: ∂T é regular, mas $D(x)$ não é diferenciável no segmento \overline{AB} .



(b) Paralelepípedo: ∂T não é regular e $D(x)$ não é diferenciável nos segmentos \overline{MN} , \overline{AI} , \overline{DI} , \overline{IM} , \overline{HJ} , \overline{EJ} , \overline{JM} , \overline{CK} , \overline{BK} , \overline{KN} , \overline{GL} , \overline{FL} e \overline{LN} .

Figura 4.17: Exemplos onde T é convexo e $D(x)$ não é diferenciável em sub-regiões de T .

Quando ∂T não é regular a situação de diferenciabilidade normalmente se agrava. Na Figura 4.17b apresentamos um paralelepípedo T e sua extensa região de não diferenciabilidade, juntamente com dois corpos paralelos às distâncias d_1 e d_2 .

Aplicando mais uma vez o Teorema de Motzkin temos que, em contraste com $D(x)$, e independentemente de ∂T ser regular ou não, $d(x, T_d)$ será sempre diferenciável em $\mathbb{R}^3 \setminus T_d$, para T convexo. Esta propriedade será explorada em desenvolvimentos posteriores.

4.5.2 Inclusão de esferas em T_ε

Dizer que nenhum ponto de uma esfera S_i ultrapassa a fronteira da região de segurança equivale a dizer que

$$S_i \subset T_\varepsilon.$$

Para restringirmos a localização das esferas ao interior de T_ε , estaremos sempre trabalhando com restrições sobre seus centros. Considerando que $r_i \geq \varepsilon$, $i = 1, \dots, n$, (esta pré-condição será tratada mais à frente), provaremos que

$$S_i \subset T_\varepsilon \iff x^i \in T_{\varepsilon - r_i}, \quad i = 1, \dots, n.$$

onde $T_{\varepsilon - r_i}$ é o corpo paralelo a T à distância $\varepsilon - r_i \leq 0$.

Teorema 4.3. *Seja $T \subset \mathbb{R}^3$ convexo, compacto e com interior não vazio, S uma esfera com raio r e centro $x_c \in T$, $\varepsilon > 0$, tal que $\varepsilon - r \leq 0$ e $T_{\varepsilon - r}$ o corpo paralelo a T , à distância $\varepsilon - r$. Então:*

$$S \subset T_\varepsilon \iff x_c \in T_{\varepsilon - r}.$$

Demonstração. Suponhamos inicialmente que

$$x_c \in T_{\varepsilon - r}.$$

Pela definição de corpo paralelo, temos

$$D(x_c) \leq \varepsilon - r \implies -\min_{z \in \partial T} d(x_c, z) \leq \varepsilon - r \implies r - \min_{z \in \partial T} d(x_c, z) \leq \varepsilon.$$

Seja agora $x \in S$, o que implica em $d(x, x_c) \leq r$. Substituindo na desigualdade acima temos

$$d(x, x_c) - \min_{z \in \partial T} d(x, z) \leq \varepsilon \implies -\min_{z \in \partial T} (d(x_c, z) - d(x, x_c)) \leq \varepsilon.$$

Aplicando a desigualdade triangular $d(x_c, z) - d(x, x_c) \leq d(x, z)$ à desigualdade

acima obtemos

$$-\min_{z \in \partial T} d(x, z) \leq -\min_{z \in \partial T} (d(x_c, z) - d(x, x_c)) \leq \varepsilon.$$

Mas esta última desigualdade significa que:

$$-\min_{z \in \partial T} d(x, z) \leq \varepsilon \implies D(x) \leq \varepsilon \implies x \in T_\varepsilon.$$

Logo,

$$S \subset T_\varepsilon.$$

Suponhamos agora que

$$S \subset T_\varepsilon.$$

Para pontos $y \in S \setminus T$ isto significa que $D(y) = d(y, \partial T) \leq \varepsilon$.

Iremos também supor, por absurdo, que

$$x_c \notin T_{\varepsilon-r}.$$

Como $x_c \in T$, isto implica em

$$D(x_c) > \varepsilon - r \implies -d(x_c, \partial T) > \varepsilon - r \implies r - d(x_c, \partial T) > \varepsilon.$$

Como ∂T é compacto, seja $y_1 \in \partial T$ tal que $d(x_c, y_1) = d(x_c, \partial T)$. Logo, aplicando a desigualdade acima, obtemos

$$d(x_c, y_1) < r - \varepsilon < r.$$

Consideremos a semi-reta m com origem em x_c passando por y_1 . Como T é convexo, temos, para $y \in m$:

$$\begin{aligned} 0 \leq d(y, x_c) \leq d(y_1, x_c) &\implies y \in T, \\ d(y, x_c) > d(y_1, x_c) &\implies y \notin T. \end{aligned}$$

Seja agora

$$y_2 = x_c + \frac{r}{d(x_c, y_1)}(y_1 - x_c).$$

Notemos que $y_2 \in m$ e, como $d(x_c, y_1) < r$, temos que $y_2 \notin T$. Por outro lado,

$$d(y_2, x_c) = \|y_2 - x_c\| = \left\| x_c + \frac{r}{d(x_c, y_1)}(y_1 - x_c) - x_c \right\| = r,$$

ou seja, $y_2 \in S$. Temos portanto que $y_2 \in S \setminus T$.

Finalmente, como x_c, y_1 e y_2 pertencem a m , temos

$$d(y_2, \partial T) = d(y_2, y_1) = d(y_2, x_c) - d(y_1, x_c) = r - d(x_c, y_1) > \varepsilon.$$

Mas isto contradiz a hipótese de que $S \subset T_\varepsilon$. Temos portanto que $x_c \in T_{\varepsilon-r}$. \square

Para o caso $\partial T_{\varepsilon-r_i}$ regular, a Figura 4.18 apresenta um caso extremo em que uma esfera S_i tem seu centro sobre a superfície $\partial T_{\varepsilon-r_i}$ sendo, portanto, tangente à superfície ∂T_ε . Nesta, P representa o ponto de tangência de S_i com T_ε , m a reta normal a ∂T_ε que passa por P e pelo centro x^i da esfera S_i , o qual é representado pelo ponto A , e finalmente Q é o ponto de interseção de m com ∂T . Temos que $d(P, x^i) = \overline{AP} = r_i$. Como $P \in \partial T_\varepsilon$ e ∂T_ε é uma curva paralela à ∂T , a distância de P a ∂T é $\overline{QP} = \varepsilon$. E assim, a distância de x^i a ∂T é igual a $\overline{AQ} = |\varepsilon - r_i|$.

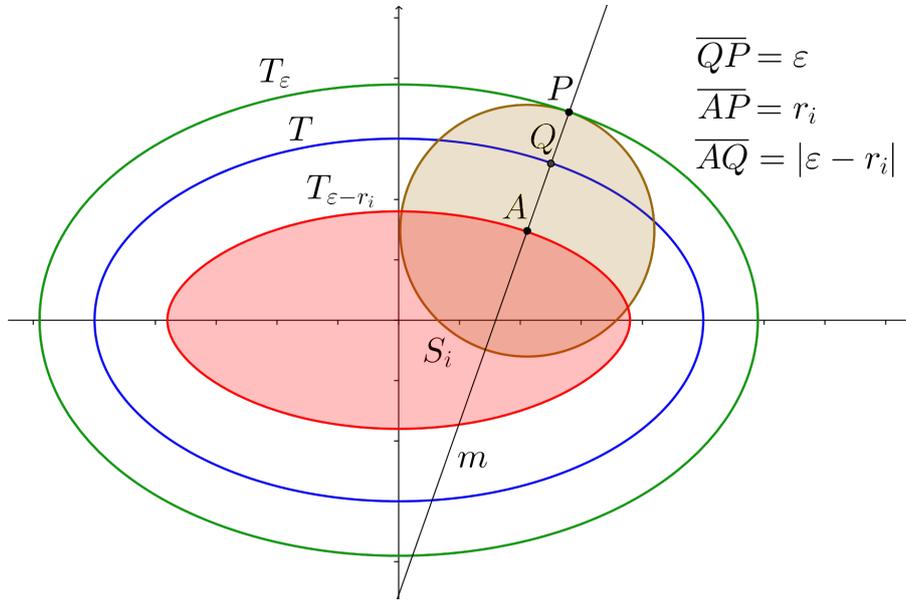


Figura 4.18: Representação dos sólidos T_ε, T e $T_{\varepsilon-r_i}$ e uma esfera S_i posicionada de forma a tangenciar ∂T_ε .

Dada a equivalência provada acima, iremos substituir as restrições originais da região de segurança

$$S_i \subset T_\varepsilon$$

pelas restrições sobre os centros x^i de cada esfera S_i

$$x^i \in T_{\varepsilon-r_i}, \quad i = 1, \dots, n,$$

que podem ser reescritas como

$$D(x^i) \leq \varepsilon - r_i, \quad i = 1, \dots, n. \quad (4.43)$$

Este conjunto de restrições também pode ser codificado com a ajuda de uma nova versão da função q , assim definida

$$q(d, x) = \begin{cases} 0 & , D(x) \leq d \text{ (ou seja, } x \in T_d) \\ +\infty & , D(x) > d \text{ (ou seja, } x \notin T_d) \end{cases}, \quad (4.44)$$

e expresso por

$$q(\varepsilon - r_i, x^i) \leq 0, \quad i = 1, \dots, n. \quad (4.45)$$

Permanece o fato de que q pode ser interpretada como uma função característica, agora do corpo paralelo T_d . Daqui em diante, quando nos referirmos à função q estaremos sempre nos referindo à esta última versão, que envolve o parâmetro d .

Em resumo, ao levarmos em consideração a região de segurança, a restrição $x \in T^n$ deve ser complementada pelas restrições $x^i \in T_{\varepsilon - r_i}$, $i = 1, \dots, n$, na formulação dos problemas-nó:

$$(PN5) : \max \quad C_k - \sum_{\substack{i,j \in I_k \\ i < j}} p_u \left(\frac{\|x^i - x^j\|}{d_{ij}} \right) \quad (4.46)$$

$$\text{s.a.} \quad x \in T^n, \quad (4.47)$$

$$x^i \in T_{\varepsilon - r_i}, \quad i \in I_k. \quad (4.48)$$

Esta formulação dos problemas-nó também pode ser reescrita utilizando as desigualdades (4.43), como apresentado abaixo:

$$(PN6) : \max \quad C_k - \sum_{\substack{i,j \in I_k \\ i < j}} p_u \left(\frac{\|x^i - x^j\|}{d_{ij}} \right) \quad (4.49)$$

$$\text{s.a.} \quad x \in T^n, \quad (4.50)$$

$$D(x^i) \leq \varepsilon - r_i, \quad i \in I_k, \quad (4.51)$$

ou então, utilizando (4.45), esta formulação também pode ser expressa por

$$(PN7) : \max \quad C_k - \sum_{\substack{i,j \in I_k \\ i < j}} p_u \left(\frac{\|x^i - x^j\|}{d_{ij}} \right) \quad (4.52)$$

$$\text{s.a.} \quad x \in T^n, \quad (4.53)$$

$$q(\varepsilon - r_i, x^i) \leq 0, \quad i \in I_k, \quad (4.54)$$

que também pode ser reescrita como

$$(PN8) : \max C_k - \sum_{\substack{i,j \in I_k \\ i < j}} p_u \left(\frac{\|x^i - x^j\|}{d_{ij}} \right) - \sum_{i \in I_k} q(\varepsilon - r_i, x^i) \quad (4.55)$$

$$\text{s.a.} \quad x \in T^n \quad (4.56)$$

A versão (PN6) utiliza restrições explícitas e pode ser implementada diretamente em um solver se contarmos com uma versão suavizada da função $D(x)$, ou seja, se trabalharmos com uma versão suavizada de (PN6). Esta não é uma boa opção pois a suavização direta de $D(x)$ pode causar distorções na fronteira dos corpos paralelos T_d .

As versões (PN7) e (PN8) não podem ser implementadas diretamente em um solver, dadas as características de não diferenciabilidade e valores infinitos assumidos pela função q . Uma versão suavizada q_μ de q , que se aproxime suficientemente de q quando $\mu \rightarrow \infty$, pode ser utilizada para implementar versões suavizadas de (PN7) e (PN8). Notamos que a suavização de q pode ser feita de forma a garantir a preservação das fronteiras dos corpos paralelos T_d . A versão suavizada de (PN7) seria uma versão muito próxima de (PN5), enquanto que a versão suavizada de (PN8), com a função q_μ fazendo parte da função objetivo, seria uma versão penalizada de (PN5).

É importante ressaltarmos que a regularidade de ∂T não é um pré-requisito para garantir a existência de funções q de fácil manipulação. É o caso dos paralelepípedos, que possuem descrições absolutamente triviais para seus corpos paralelos T_d . Por outro lado, mesmo com ∂T regular, temos ∂T_d não regular para $d < d_{\text{inf}}$, o que na prática pode dificultar a construção das funções q para estes valores de d .

Como já mencionado anteriormente, temos que $D(x)$, em geral, é não diferenciável em uma sub-região de T . Desta forma, sua utilização em problemas de otimização cria dificuldades computacionais, pois a grande maioria dos otimizadores não-lineares partem do princípio que as funções envolvidas são diferenciáveis, e utilizam o gradiente, de uma forma ou de outra, como guia na busca por mínimos locais, parando a busca quando seu módulo é suficientemente pequeno.

4.5.3 Suavização da função q

Por intermédio da função q , o problema (PN8) pode ser interpretado como uma versão limite de uma penalização do problema (PN5). Podemos, então, considerar a substituição de q por uma versão suavizada e parametrizada q_μ , de tal forma que q_μ tenda para q quando $\mu \rightarrow \infty$.

Com base na teoria dos métodos de penalização e barreiras, consideremos a função q_μ , definida por:

$$q_\mu(d, x) = \mu d(x, T_d)^2.$$

Obviamente, se $x \in T_d$, então $d(x, T_d) = 0$ e portanto $q_\mu(d, x) = 0$. Por outro lado, se $x \notin T_d$, então $d(x, T_d) > 0$ e $q_\mu(d, x) \rightarrow \infty$ quando $\mu \rightarrow \infty$. Assim sendo, q_μ se comporta aproximadamente como a função q quando $\mu \rightarrow +\infty$.

Como T_d é convexo, e mais uma vez fazendo menção ao Teorema de Motzkin, temos que, independentemente de ∂T ser regular ou não, $d(x, T_d)$ será sempre diferenciável em $\mathbb{R}^3 \setminus T_d$. Adicionalmente, apesar da função distância $d(x, T_d)$ não ser diferenciável na fronteira ∂T_d , a função $d(z, T_d)^2$ é diferenciável em todos os pontos desta região.

Temos mencionado o método de penalização para justificar a construção de q_μ , mas esta função também pode ser utilizada diretamente para gerar uma versão suavizada de (PN7), sem incorrer nos problemas de balanceamento de penalizações na função objetivo.

Antes de continuarmos, observemos que, quando consideramos a restrição $x^i \in T_{\varepsilon-r_i}$ em conjunto com a restrição $x^i \in T = T_0$, podemos substituí-las pela restrição única $x^i \in T_0 \cap T_{\varepsilon-r_i}$. Portanto, se definirmos $d^i = \min(0, \varepsilon - r_i)$, temos que $x^i \in T_0 \cap T_{\varepsilon-r_i}$ pode ser abreviadamente expressa por $x^i \in T_{d^i}$. Esta restrição pode portanto ser reescrita, utilizando q_μ , como

$$q_\mu(d^i, x^i) = \mu d(x^i, T_{d^i})^2 \leq 0,$$

onde $d^i = \min(0, \varepsilon - r_i)$.

Retornando ao desenvolvimento de q_μ , temos a princípio somente $D(x)$ disponível, e seria importante tentarmos exprimir q_μ com base em $D(x)$.

Se ∂T e ∂T_d são regulares para um certo $d < 0$, temos

$$d(x, T_d) = \begin{cases} d(x, \partial T) - d = D(x) - d, & \text{para } x \in \mathbb{R}^3 \setminus T \\ -d(x, \partial T) - d = D(x) - d, & \text{para } x \in T \setminus T_d \\ 0, & \text{para } x \in T_d \end{cases}.$$

Estas relações surgem diretamente do paralelismo entre as superfícies ∂T e ∂T_d . Podemos simplificar esta definição para

$$d(x, T_d) = \max(0, D(x) - d),$$

pois $D(x) - d \leq 0$ para $x \in T_d$, o que nos leva a

$$q_\mu(d, x) = \mu [\max(0, D(x) - d)]^2, \quad x \in T.$$

Para a situação ∂T e ∂T_{d^i} regulares teremos portanto que a restrição sobre o centro x^i da esfera S^i será dada por

$$q_\mu(d^i, x^i) = \mu[\max(0, D(x^i) - d^i)]^2 \leq 0,$$

onde $d^i = \min(0, \varepsilon - r_i)$.

A situação descrita acima para ∂T e ∂T_{d^i} regulares, a qual pode ser analiticamente descrita de forma simples, muda completamente quando ∂T_{d^i} não é regular. Isto sempre acontecerá quando ∂T não é regular, ou quando ∂T é regular mas $d^i < d_{\text{inf}}$. Nestes casos as superfícies ∂T_d paralelas a T e internas a T ($d < 0$) não são as mesmas superfícies $\partial(T_{d^i})_d$ paralelas a T_{d^i} e externas a T_{d^i} ($d > 0$). Não é simples, portanto, relacionar diretamente $d(x, T_d)$ com $D(x) = \sigma_T(x)d(x, \partial T)$.

Este é o caso com o paralelepípedo, onde as superfícies $\partial(T_{d^i})_d$ sempre serão regulares para $d > 0$, pois sendo T_{d^i} compacto e convexo resulta em $d(x, T_{d^i})$ diferenciável. No entanto, $d(x, \partial T)$ é não diferenciável em uma sub-região dentro de T .

Pelo exposto acima, a não regularidade de ∂T_d impede a computação direta de $d(x, T_d)$ com base em $D(x)$. No entanto, já sabemos que T convexo implica na convexidade dos corpos paralelos T_d . Por sua vez, a convexidade de T_d resulta na unicidade da projeção x_p de um ponto x externo a T_d . Consideremos $d < 0$.

Se a superfície ∂T_d é regular em uma vizinhança de x_p , temos bem definido o plano tangente a T_d em x_p . Denominaremos CT_p ao semi-espço definido pelo plano tangente a T_d em x_p e que contém totalmente o sólido T_d , e CN_p à semi-reta com origem em x_p passando por x . A convexidade de T_d nos garante que a semi-reta CN_p é perpendicular a CT_p . Como $d(x, CT_p) = d(x, x_p)$, $T_d \subset CT_p$ e $x_p \in T_d$, temos que $d(x, T_d) = d(x, x_p)$. Logo, se a projeção (única) x_p de x sobre T_d é um ponto cuja vizinhança em ∂T_d é regular, temos que $d(x, T_d) = d(x, x_p) = -d(x, \partial T) - d$. Este resultado não é novo e também surgiu quando consideramos que ∂T e ∂T_d eram regulares.

Se a superfície ∂T_d não é regular em x_p não podemos definir o plano tangente a T_d em x_p . No entanto, como T_d é convexo, podemos definir seus cones tangente CT_p e normal CN_p em x_p dados por:

$$\begin{aligned} CT_p &= cl(\{v \in \mathbb{R}^3 \mid \exists t > 0 \text{ tal que } x_p + tv \in T_d\}), \\ CN_p &= \{s \in \mathbb{R}^3 \mid \langle s, z - x_p \rangle \leq 0, \forall z \in T_d\}, \end{aligned}$$

onde $cl()$ significa o fecho de um conjunto e \langle, \rangle representa o produto escalar de dois vetores.

Podemos agora facilmente provar que estes cones generalizam a situação regular.

Suponhamos que x externo a T_d seja tal que, para um determinado x_p tenhamos $x \in CN_p$. Por definição, $T_d \subset CT_p$. Provaremos que $d(x, T_d) = d(x, x_p)$.

De fato, seja $y \in T_d$ e θ o ângulo formado pelos segmentos $\overline{x_p y}$ e $\overline{x_p x}$. Pela lei dos cossenos temos:

$$d(x, y)^2 = d(x, x_p)^2 + d(x_p, y)^2 - 2d(x, x_p)d(x_p, y) \cos \theta.$$

Como $x \in CN_p$ e $y \in T_d$, temos $\cos \theta \leq 0$. Assim,

$$d(x, y)^2 \geq d(x, x_p)^2 + d(x_p, y)^2 \geq d(x, x_p)^2 \implies d(x, y) \geq d(x, x_p).$$

Finalmente,

$$d(x, T_d) = \min_{y \in T_d} d(x, y) = d(x, x_p),$$

e portanto, se $x \in CN_p$ para algum x_p , temos que $d(x, T_d) = d(x, x_p)$. Esta propriedade será utilizada na implementação de $d(x, T_d)$ quando for possível identificar os cones normais de T_d , como será o caso para T elipsóide.

A Figura 4.19 apresenta uma seção de um elipsóide com os cones tangente e normal em um ponto onde ∂T_d não é regular.

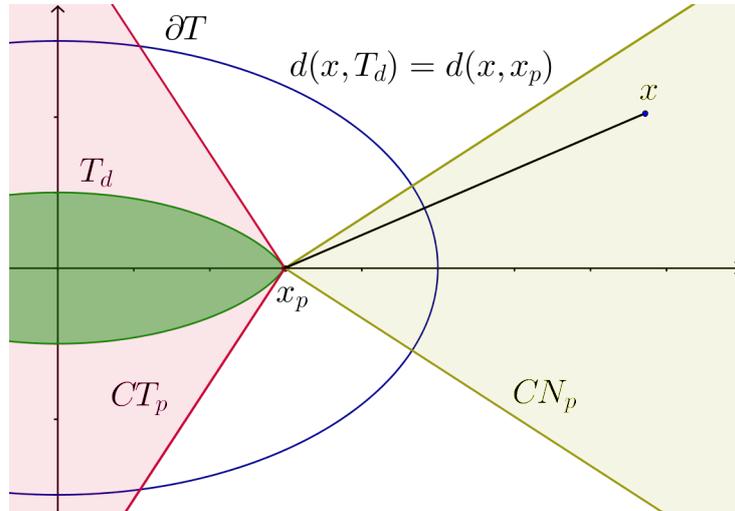


Figura 4.19: Exemplo de cones tangente CT_p e normal CN_p para o caso de T elipsóide.

Podemos agora incluir a função q_μ na formulação final dos problemas-nó. Optamos por uma versão suavizada de (PN8), como apresentado abaixo:

$$(PN) : \max C_k - \sum_{\substack{i,j \in I_k \\ i < j}} p_u \left(\frac{\|x^i - x^j\|}{d_{ij}} \right) - \sum_{i \in I_k} q_\mu(d^i, x^i) \quad (4.57)$$

$$\text{s.a.} \quad x \in T^n. \quad (4.58)$$

A restrição $x \in T^n$ foi mantida, apesar de já estar embutida na parcela de penalização dada por q_μ , como auxílio durante a depuração da implementação computacional. Ela é obviamente dispensável na implementação definitiva.

4.6 Árvore de decisão e índices de cobertura

Nesta seção iremos tratar das restrições de integralidade (4.5), as quais se constituem na fonte de não-convexidade NC_1 . Estas restrições forçam as variáveis y_i , $i = 1, 2, 3, \dots, n$, a se tornarem variáveis binárias, representando assim o processo de escolha das esferas que irão participar do recobrimento do sólido T .

Lembremos que chamamos de nó a uma determinada seleção fixa de esferas do conjunto S que irão participar do recobrimento, sendo $I_k \subset N$ o conjunto de índices dessas esferas, para cada nó $k \in N_K$, de tal forma que $y_i = 1$ para todo $i \in I_k$, enquanto $y_i = 0$ para todo $i \notin I_k$. O número K de nós considerados, assim como a constituição de cada I_k para $k \in N_K$, serão definidos abaixo.

Para cada nó estaremos resolvendo os problemas (PN) , conforme discutido na seção anterior. Estes problemas possuem funções objetivo constituídas por C_k , que é um termo relacionado com o número de esferas selecionadas, e um termo de penalização.

O termo de penalização implementa uma heurística que orienta a formação de um melhor arranjo para a sobreposição das esferas, uma vez que esta sobreposição é absolutamente necessária para a realização de um recobrimento. O processo de otimização naturalmente busca zerar este termo, independentemente do nó considerado. Em situações ideais, este termo seria zero para todos os nós considerados, não havendo, portanto, contribuição direta deste termo para a seleção do nó ótimo.

O termo C_k , considerando sua aplicação ao problema *Gamma Knife*, tem por motivação original a priorização das esferas de maior raio, por intermédio de parâmetros c_i com valores elevados, de tal forma que os nós com as maiores esferas apresentariam, a princípio, os valores mais elevados da função objetivo. Isto é desejável se quisermos reduzir o número total de esferas utilizado para o recobrimento, o que no contexto do *Gamma Knife* se traduz em um menor número de *shots*.

Percebemos, então, que nossa função objetivo não tem condições de medir diretamente a qualidade de um recobrimento, apesar dos esforços heurísticos neste sentido. Temos portanto que utilizar ferramentas pós-otimização para realizar a medição da qualidade de recobrimento de cada nó. Estas ferramentas serão construídas com base em um conjunto de índices de cobertura, os quais serão tratados na próxima subseção. Em seguida serão discutidas as árvores de decisão, as quais organizam o conjunto N_K de nós em estruturas de árvores e utilizam os índices de cobertura no processo de decisão dos ramos que serão percorridos.

4.6.1 Índices de cobertura

Dado um sólido T a ser recoberto pelo conjunto previamente selecionado I_k de esferas, iremos definir agora uma série de índices de cobertura cujo propósito será medir a qualidade do recobrimento de T . Estes índices de cobertura serão baseados na medição de volumes, e parte-se do princípio de que todo os volumes considerados podem ser medidos utilizando-se a medida de Lebesgue em \mathbb{R}^3 .

Os volumes básicos utilizados são:

- V_T : Volume do sólido T ;
- V_{SR} : Volume de T_ϵ , que é a união de T com sua região de segurança;
- V_{B_i} : Volume da esfera S_i , $i \in I_k$;
- V_B : Volume da união $B = \bigcup_{i \in I_k} S_i$;
- V_{TB} : Volume da interseção $T \cap B$;
- V_{SRB} : Volume da interseção $T_\epsilon \cap B$;
- V_{IB2} : Volume da interseção de duas ou mais esferas;
- V_{IB3} : Volume da interseção de três ou mais esferas;
- V_{IB4} : Volume da interseção de quatro ou mais esferas.

A computação destes volumes foi realizada mediante a adoção da integração de quasi-Monte Carlo, descrita no Apêndice D e que será discutida com maiores detalhes no Capítulo 5. Alguns volumes complementares foram definidos para auxiliar na implementação desta técnica:

- V_C : Volume de um cubo que contém a grade de integração;
- V_G : Volume da grade de integração.

Quanto aos índices definidos, temos:

Índice de Cobertura, I_{COV}

Definido por $I_{COV} = \frac{V_{TB}}{V_T}$.

Este é o principal índice, medindo a proporção de T recoberta pela união das esferas.

Valor ideal: 1 (100%).

Índice de Eficiência 1, I_{E1}

Definido por $I_{E1} = \frac{V_{TB}}{V_B}$.

Este índice mede a proporção da união das esferas que é efetivamente utilizada no recobrimento.

Valor ideal: 1 (100%), mas na prática, dado que normalmente existe alguma cobertura da região de segurança, um valor típico seria da ordem de 85%.

Índice de Eficiência 2, I_{E2}

Definido por $I_{E2} = \frac{V_{SRB}}{V_B}$.

Este índice mede a proporção da união das esferas que está contida em T_ϵ .

Valor ideal: 1 (100%).

Índice de Interseção 2, I_{B2}

Definido por $I_{B2} = \frac{V_{IB2}}{V_B}$.

Este índice mede a proporção do volume das interseções de duas ou mais esferas com relação ao volume da união das esferas.

Valor ideal: Não definido, mas deve ser o menor possível. Normalmente é necessária alguma sobreposição de esferas no recobrimento e, portanto, este índice costuma ser maior que zero.

Índice de Interseção 3, I_{B3}

Definido por $I_{B3} = \frac{V_{IB3}}{V_B}$.

Este índice mede a proporção do volume das interseções de três ou mais esferas com relação ao volume da união das esferas.

Valor ideal: Não definido, mas deve ser o menor possível.

Índice de Interseção 4, I_{B4}

Definido por $I_{B4} = \frac{V_{IB4}}{V_B}$.

Este índice mede a proporção do volume das interseções de quatro ou mais esferas com relação ao volume da união das esferas.

Valor ideal: 0.

O índice I_{B1} não é definido pois a lei de formação destes índices nos daria sempre

$$I_{B1} = \frac{V_B}{V_B} = 1.$$

Por outro lado, a definição destes índices implica em que sempre teremos $I_{B4} \leq I_{B3} \leq I_{B2}$. A proporção dos volumes somente das interseções duplas pode ser calculado por $I_{B2} - I_{B3}$ e a proporção dos volumes somente das interseções triplas pode ser calculado por $I_{B3} - I_{B4}$, mas estas informações não serão úteis no contexto de nossos desenvolvimentos.

Com base nestes índices primários, podemos definir os dois índices de conformidade mais mencionados na literatura e amplamente utilizado em centros de radio-cirurgia:

Índice de Shaw, I_S

Definido por $I_S = \frac{V_{TB}}{V_T} = I_{COV}$.

O índice de conformidade de Shaw (SHAW, 1993) é o mais conhecido e utilizado, antes do surgimento do índice de Paddick, medindo a proporção de T coberta pela união das esferas. É por definição igual a I_{COV} .

Valor ideal: 1 (100%).

Índice de Paddick, I_{PDK}

Definido por $I_{PDK} = \frac{V_{TB}^2}{V_T V_B} = \frac{V_{TB}}{V_T} \frac{V_{TB}}{V_B} = I_{COV} I_{E1}$.

Este é um índice de conformidade que é considerado superior ao índice de Shaw, pois considera a posição relativa entre os volumes V_T e V_B . PADDICK (2000) denomina o fator I_{COV} de "fator de cobertura do alvo (target coverage)" e o fator I_{E1} de "seletividade do plano (plan selectivity)".

Valor ideal: 1 (100%), mas na prática existe recobrimento da região de segurança, o que leva este índice, em casos de boa cobertura, a valores da ordem de 85%, mesmo que I_{COV} seja 100%.

Uma discussão geral sobre os índices de cobertura aplicados à técnica do *Gamma Knife* pode ser encontrada nos artigos de FEUVRET *et al.* (2006); PADDICK (2000).

Dentre os dois índices de conformidade apresentados, o índice de Paddick é o mais completo por também considerar a eficiência da união das esferas utilizadas na cobertura.

Para uma avaliação geral do comportamento destes índices com relação ao número de esferas utilizadas, iremos considerar, a título de exemplo, o caso particular de uma cobertura de um sólido T por esferas de mesmo tamanho. Na Figura 4.20, apresentamos os gráficos aproximados dos principais índices em função do número N_S de esferas iguais.

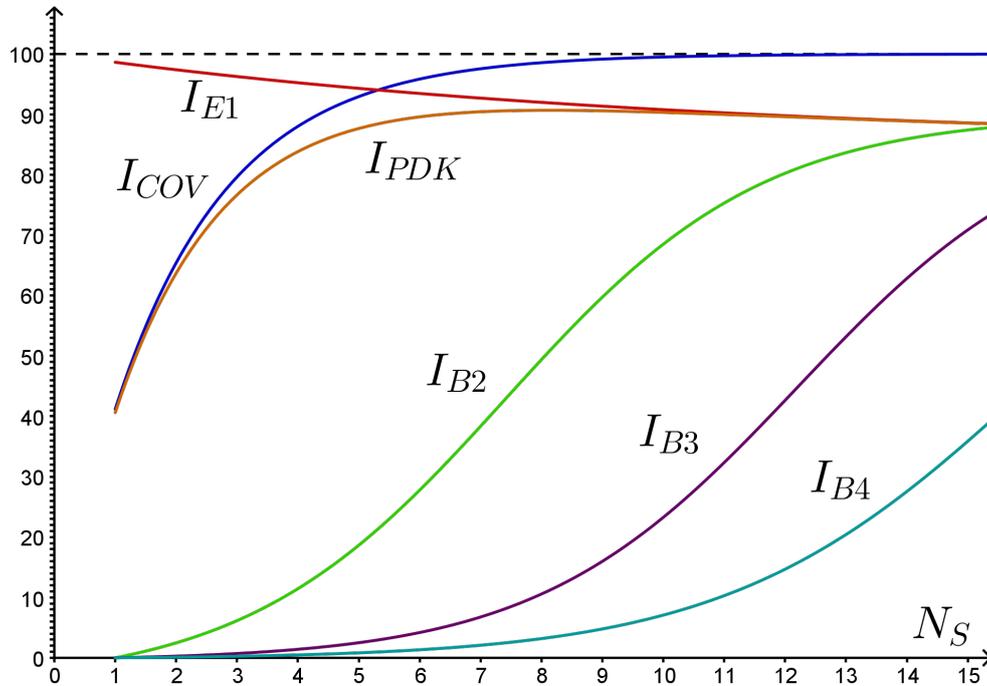


Figura 4.20: Comportamento aproximado dos principais índices de cobertura.

Temos o seguinte comportamento dos índices quando o número de esferas utilizado tende a aumentar:

- O índice de cobertura I_{COV} tende a aumentar, pois mais esferas estarão disponíveis para o recobrimento;
- O índice de eficiência I_{E1} tende a diminuir, pois uma porção maior das esferas tende a cobrir a região de segurança;
- O índice de Paddick tende inicialmente a aumentar, possuindo um comportamento muito similar ao do índice I_{COV} . À medida que o número de esferas continua a aumentar, no entanto, este índice passa a ser limitado pelo índice I_{E1} .
- Os índices I_{B2} , I_{B3} e I_{B4} , que obrigatoriamente seguem as desigualdades $I_{B4} \leq I_{B3} \leq I_{B2}$, começam com valores nulos para uma esfera e tendem a aumentar significativamente quando o número de esferas se torna grande.

Apesar deste comportamento ter sido apresentado para o caso especial de esferas de mesmo raio, no caso de raios diferentes o comportamento geral é muito similar.

Veremos agora como este conjunto de índices pode nos auxiliar na seleção da melhor cobertura de um sólido T , dado um conjunto de esferas de um nó.

O índice I_{COV} , principal índice de mensuração da qualidade de um recobrimento, deve estar o mais próximo possível de 100%. Isto nem sempre é possível, pois a restrição criada pela região de segurança pode efetivamente impedir que algumas regiões de T sejam cobertas. Isto acontece, por exemplo, para T paralelepípedo, onde a menor esfera disponível posicionada junto a um canto da região de segurança pode não conseguir cobrir a região próxima ao vértice correspondente. Neste caso particular, sendo r_m o raio da menor esfera disponível, é absolutamente necessário que

$$\varepsilon \geq r_m \left(1 - \frac{\sqrt{3}}{3} \right),$$

para que o recobrimento com $I_{COV} = 100\%$ possa ser atingido.

Temos também a possibilidade da existência de regiões dentro de T que podem não ser cobertas por B . A heurística que determina a distância entre as esferas tem por objetivo reduzir esta possibilidade.

Como todos os pontos da superfície ∂T devem ser cobertos em uma cobertura com $I_{COV} = 100\%$, devemos obrigatoriamente ter $B \cap \partial T = \partial T$, ou seja, a superfície $B \cap \partial T$ deve ser um recobrimento bidimensional de ∂T .

Mesmo se tivermos $I_{COV} = 100\%$, podemos ainda ter um recobrimento ineficiente se um volume considerável da região de segurança estiver sendo coberta. O índice I_{E1} tem por objetivo justamente medir a proporção útil do volume B da união das esferas.

Podemos conseguir $I_{E1} = 100\%$ se fizermos $\varepsilon = 0$, ou seja, $T_\varepsilon = T$, e todas as esferas respeitarem a restrição $B \subset T$. No entanto, neste caso certamente teríamos I_{COV} substancialmente inferior a 100% (em particular $B \cap \partial T$ seria constituída por pontos isolados). Daí percebemos a impossibilidade de atingirmos os maiores valores possíveis de I_{COV} e I_{E1} simultaneamente.

O índice de Paddick tem a grande vantagem de reunir em um único índice os efeitos combinados de I_{COV} e I_{E1} . Desta forma, buscaremos atingir o maior valor possível de I_{PDK} , balanceando automaticamente os efeitos de I_{COV} e I_{E1} .

O índice I_{E2} está ligado à qualidade da implementação computacional e deve sempre ser 100%. Desvios deste valor implicam na necessidade de verificação da implementação.

Os índices de interseção estão também ligados à eficiência do recobrimento. Assumindo o atendimento à restrição $B \subset T_\varepsilon$, primeiramente notamos que $I_{B2} = 0$ implica em um empacotamento das esferas em T_ε . Como nosso objetivo é realizar um recobrimento, é absolutamente necessário termos $I_{B2} > 0$ quando utilizamos mais de uma esfera.

Sob um ponto de vista de dimensão topológica, temos que um recobrimento em duas dimensões utilizando círculos pode ser realizado com $I_{B2} > 0$ e $I_{B3} = 0$, bastando considerar como agrupamento básico para o recobrimento três círculos se interceptando em apenas um ponto. Um recobrimento em três dimensões utilizando esferas pode ser realizado com $I_{B2} > 0$, $I_{B3} > 0$ e $I_{B4} = 0$ considerando como agrupamento básico para o recobrimento quatro esferas se interceptando em apenas um ponto.

Quando o número de esferas dentro de T_ε cresce a tendência é o aumento do número de interseções, com o conseqüente aumento dos índices de interseção.

De uma forma geral podemos atestar a qualidade de uma cobertura quando tivermos um alto índice de Paddick acompanhado por baixos valores dos índices de interseção. Estes índices serão, portanto, as variáveis de decisão utilizadas na busca pelas melhores coberturas em uma das possíveis árvores que organizam os nós (PN) de nosso problema.

4.6.2 Árvores de decisão

Estaremos agora discutindo a organização do conjunto $N_K = \{0, 1, \dots, K\}$, indexador dos nós I_k , constituídos por seleções de esferas. Algumas observações e desenvolvimentos iniciais são necessários.

Dado o conjunto de esferas $S = \{S_i, i \in N = \{1, \dots, n\}\}$ disponíveis para o recobrimento, seja M o número total de raios considerados e O o número total de esferas para cada raio (para simplificar, estaremos considerando o mesmo número de esferas para cada raio). O número total de esferas disponíveis é, portanto, $n = MO$.

Iremos considerar que o conjunto S está ordenado pelos raios das esferas, sendo esta ordem diretamente expressa pelo conjunto de índices N . Para nosso problema escolhemos ordenar as esferas do maior para o menor raio, refletindo com isto a preferência pela utilização inicial das esferas maiores.

Sabemos que esferas iguais, quando disponibilizadas para a realização de um recobrimento, são intercambiáveis. Desta forma, a definição atual dos conjuntos I_k torna possível que, para $k_1 \neq k_2$, tenhamos nós I_{k_1} e I_{k_2} diferentes que representem, para todos os efeitos, a mesma seleção de esferas. Para resolver este problema, iremos somente considerar conjuntos I_k para os quais sempre escolheremos as primeiras esferas disponíveis para cada raio desejado, eliminando assim qualquer ambigüidade presente nos conjuntos I_k . Com esta providência, reduzimos o potencial número de nós de

$$K = 2^{MO} - 1 = 2^n - 1$$

para

$$K = (O + 1)^M - 1 = \left(\sqrt[M]{O + 1} \right)^{MO} - 1 = \left(\sqrt[M]{O + 1} \right)^n - 1,$$

uma vez que $\sqrt[O]{O+1} < 2$ para $O > 1$.

Subtraímos um nas expressões acima porque o nó I_0 (seleção nula de esferas) não nos interessa para recobrimentos. Assim, K conta somente as seleções não nulas de esferas. Apesar disto, o nó nulo será conceitualmente útil em breve.

Na Figura 4.21 representamos de forma bidimensional o conjunto de esferas S_i , considerando $M = 4$ e $O = 6$, resultando em $n = 24$.

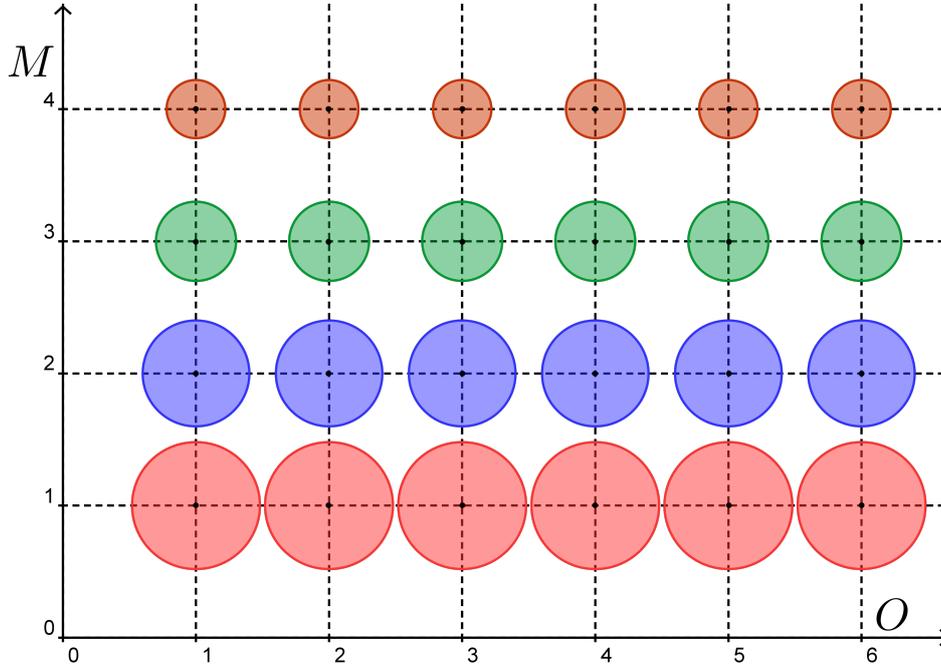


Figura 4.21: Representação do conjunto S de esferas, para $M = 4$ e $O = 6$.

Temos agora condições de definir o que chamaremos de *assinatura* do k -ésimo nó, $k \in N_K$, a qual será uma M -tupla $\eta_k = (l_1, l_2, \dots, l_M)$ constituída pelo número de esferas l_m correspondentes a cada raio, com $0 \leq l_m \leq O$, $m = 1, \dots, M$, ou seja, $\eta_k \in \{0, 1, \dots, O\}^M$. No exemplo anterior, se o nó for constituído por três esferas de maior raio, duas do raio imediatamente menor e uma do menor raio disponível, teremos um nó com assinatura $\eta_k = (3, 2, 0, 1)$. Notamos que existe uma relação biunívoca entre as assinaturas η_k e os nós I_k , uma vez resolvida a ambiguidade inicialmente presente.

Devemos lembrar que na definição de (P1) as variáveis inteiras (binárias) y_i , $i \in N$, indicam a presença ou não da esfera S_i no recobrimento. Estas variáveis podem agora ser descritas com base em nossa organização das esferas. Se considerarmos $m \in \{1, \dots, M\}$ e $o \in \{1, \dots, O\}$ como coordenadas no arranjo bidimensional das esferas apresentado acima, temos que, dados m e o , obtemos $i = o + (m - 1)O$ e, vice-versa, dado i obtemos $m = \lfloor (i - 1)/O \rfloor + 1$ e $o = i - (m - 1)O$.

Se quisermos saber se a esfera S_i pertence ao nó $\eta_k = (l_1, l_2, \dots, l_M)$, basta calcularmos os respectivos m e o em função de i e compararmos o com l_m . Teremos

$y_i = 1$ se $o \leq l_m$ e $y_i = 0$ caso contrário.

Doravante não mais recorreremos aos conjuntos I_k e nem às variáveis inteiras y_i para identificar um nó, pois seus aspectos combinatórios foram eficientemente absorvidos pela definição da assinatura de um nó, e, portanto, nos concentraremos em identificar um nó somente por meio de sua assinatura η_k .

O número total de nós pode ser bastante elevado, mesmo com M e O não muito grandes. Em nosso exemplo com $M = 4$ e $O = 6$ temos $7^4 = 2.401$ nós (com ambiguidade teríamos $2^{24} = 16.777.216$ nós). Criar uma estrutura que permita avaliar a qualidade de cada nó de forma organizada torna-se portanto uma necessidade absoluta.

Um nó genérico η_k pode ser considerado como uma M -tupla de variáveis inteiras l_m , $m \in \{1, \dots, M\}$. Para simplificar a discussão a seguir, suponhamos que cada nó tenha solução exata. Existem dois métodos tradicionalmente utilizados para a resolução exata de problemas com variáveis inteiras: o método dos planos de corte e o método *Branch and Bound*. Estes métodos organizam a busca por soluções ótimas dentro da região viável, dividindo-a em sub-regiões onde a busca deve ser priorizada, ao mesmo tempo em que eliminam as sub-regiões que não irão fornecer candidatas adequadas.

Para nosso problema resolvemos aplicar o conceito das árvores de decisão, subjacente ao método *Branch and Bound*, utilizando os índices de cobertura previamente desenvolvidos como as variáveis de decisão em cada nó. A árvore escolhida tem por origem a estrutura da assinatura de cada nó $\eta_k = (l_1, l_2, \dots, l_M)$, pois podemos ordenar naturalmente os nós se considerarmos $k = (l_M l_{M-1} \dots l_2 l_1)_O$, ou seja, k é o número, descrito na base O , formado pelos algarismos l_m partindo de l_M (mais significativo) até l_1 (menos significativo). Em outras palavras,

$$k = \sum_{j=1}^M l_j (O + 1)^{j-1}.$$

Notamos que, se a assinatura η_k gera o índice k do nó, e vice-versa por divisões e restos sucessivos, podemos trabalhar diretamente com a estrutura da assinatura de cada nó. Afinal, a ordem dada por k não necessariamente é a melhor forma de se percorrer os nós e isto ficará claro quando mencionarmos nossa estratégia de busca na árvore.

naturas, a busca sequencial direta pelas setas contínuas e as tentativas seguidas por retrocessos (*backtracking*) pelas setas pontilhadas.

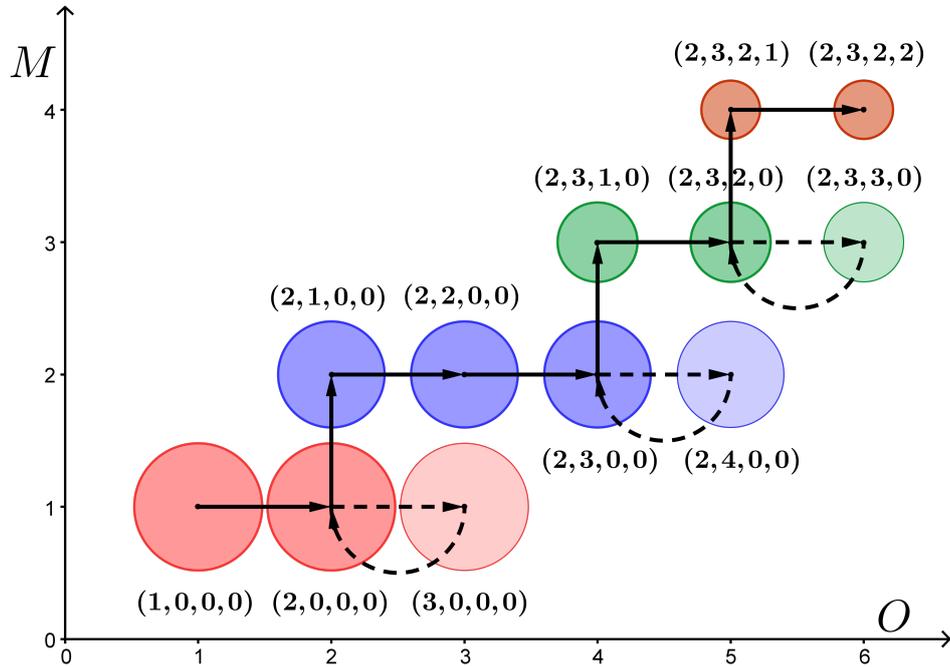


Figura 4.23: Busca p no conjunto S de esferas, para $M = 4$ e $O = 6$.

Em resumo, nossa árvore de decisão será explorada de forma linear com *backtracking*.

Quanto às decisões de ramificação, adotamos o critério de se buscar o maior índice de Paddick possível associado aos menores valores para os índices de interseção I_{B2} , I_{B3} e I_{B4} . Os índices do melhor nó corrente (*incumbent*) são, então comparados com o nó atual, em busca de melhorias. Como já mencionado, anteriormente, e graficamente exposto na Figura 4.20, a partir de um certo ponto, o aumento do número de esferas não produz uma melhoria na qualidade do recobrimento medida pelo índice de Paddick, mas produz uma redução na qualidade do recobrimento medida pelos índices de interseção.

Como estamos tratando de um problema de otimização global e nosso nó não tem garantia de otimalidade, optamos por uma estratégia de busca local randômica em cada nó, partindo-se da melhor configuração obtida no nó anterior. Desta forma garantimos o progresso no sentido de obtemos boas soluções para cada nó.

Capítulo 5

Resultados computacionais

Para testar na prática as teorias apresentadas no capítulo anterior, desenvolvemos uma implementação computacional, denominada USCS. Primeiramente iremos descrever esta implementação, com suas principais características, na Seção 5.1. Apresentaremos a seguir, na Seção 5.2, as instâncias de teste escolhidas, as quais forneceram os parâmetros para a programação do USCS. Finalizamos com a Seção 5.3, onde apresentaremos os resultados experimentais obtidos, acompanhados por uma avaliação crítica.

5.1 A implementação USCS

A implementação foi realizada no ambiente e linguagem de modelagem algébrica AIMMS (Advanced Interactive Multidimensional Modeling System), um produto da *AIMMS B.V.* (inicialmente chamada *Paragon Decision Technology*). Esta implementação recebeu o nome de USCS (*Unequal Spheres Covering System*).

O sistema USCS possui as seguintes características:

- pode ser operado nos modos manual e automático. No modo manual é possível estudar recobrimentos de tumores utilizando um conjunto fixo de esferas. No modo automático as esferas que serão utilizadas no recobrimento são selecionadas mediante a busca em árvore desenvolvida no Capítulo anterior;
- pode ser operado em ambientes bidimensionais (recobrimento com círculos) e tridimensionais (recobrimento com esferas);
- permite definir o conjunto de esferas disponíveis (balls pool) para utilização no recobrimento. Para cada tipo de esfera podem ser especificados seu raio, seu número disponível, sua cor e, por conveniência, seu bloqueio para uso;
- no modo manual as esferas são inicializadas com uma semente aleatória com distribuição próxima à origem ou uniforme dentro do corpo sólido (“tumor”).

O modo manual adicionalmente implementa o tumor “nulo”, o qual permite a avaliação dos agregados de esferas sem a interferência das restrições de um tumor;

- no modo automático uma busca local randômica é realizada para cada novo nó, preservando a melhor configuração obtida dos nós anteriores. O nó acrescentado é inicializado em uma posição aleatória com base em uma distribuição uniforme esférica em torno do tumor, ou em uma distribuição uniforme em um pequeno cubo centrado na origem. Se a busca local não encontra um nó viável, é realizado um *backtracking* (retrocesso) e um ramo da árvore utilizando esferas menores é selecionado;
- possibilita a utilização de algumas das principais penalizações desenvolvidas no capítulo anterior;
- permite selecionar os corpos sólidos utilizados na cobertura (em adição ao tumor nulo, utilizado somente no modo manual). No ambiente bidimensional o tumor pode ser: paralelogramo ou elipse (incluindo o círculo). No ambiente tridimensional o tumor pode ser: paralelepípedo ou elipsóide (esfera, prolato ou oblato);
- inclui uma implementação em linguagem C do quadrado da distância ao corpo paralelo nos casos em que o corpo sólido é uma elipse, uma esfera ou um elipsóide (prolato ou oblato);
- apresenta estatísticas e índices relacionados com o recobrimento, incluindo os contadores da integração quasi-Monte Carlo, os volumes calculados e os índices de cobertura, de Paddick e de interseção;
- permite a geração de uma listagem de informações, incluindo um sumário das estatísticas e índices, um histórico cronológico dos eventos de processamento e segmentos de programa para visualização dos resultados da cobertura (para CaRMetal¹, no ambiente bidimensional, e para o R statistical software², no ambiente tridimensional).

Após a realização de alguns testes preliminares, o CONOPT v 3.14 foi escolhido como o solver não-linear a ser utilizado pelo USCS. No entanto, como o USCS foi desenvolvido utilizando o ambiente AIMMS, outros solvers não-lineares podem ser empregados, se necessário.

¹O CaRMetal é um programa de geometria dinâmica.

²A linguagem R é utilizada para a computação estatística. Possui uma excelente biblioteca para apresentação gráfica tridimensional.

Apresentamos a tela de interfaceamento com o USCS na Figura E.1 do Apêndice E.1. É possível observarmos os vários parâmetros que podem ser ajustados e os resultados obtidos pós-otimização.

No Apêndice E.2 apresentamos um exemplo do sumário das estatísticas e índices, extraído da listagem de informações proveniente do recobrimento automático de um elipsóide prolato. Temos assim acesso a um registro dos contadores de pontos utilizados na integração de quasi-Monte Carlo e dos volumes calculados para o sólido (tumor) e para as esferas participantes do recobrimento, assim como dos índices de cobertura, de eficiência e de interseção. A comparação do erro entre os volumes exatos e integrados das esferas permite uma avaliação geral da qualidade da integração empregada.

No Apêndice E.3 apresentamos um exemplo do histórico cronológico dos eventos de processamento extraído da listagem de informações gerada pelo USCS. Esta listagem nos permite facilmente perceber que uma grande parte do tempo total de processamento do recobrimento (18% neste exemplo) é utilizado na fase inicial de construção da grade e integração do sólido a ser recoberto. Também notamos que as integrações intermediárias, as quais envolvem somente as esferas, tomam inicialmente mais tempo do que o solver, mas esta situação se equilibra à medida que o número de esferas aumenta.

Finalizando, gostaríamos de comentar que os segmentos de programa gerados pelo USCS foram extremamente úteis para uma avaliação visual da qualidade das coberturas geradas.

5.2 Instâncias de teste

Conforme já mencionado na seção anterior, o USCS pode trabalhar com problemas bidimensionais. No entanto, tendo em vista a aplicação do problema de recobrimento à técnica de radiocirurgia Gamma Knife, iremos apresentar resultados somente de problemas tridimensionais.

Para a realização de experimentos computacionais foi necessária a especificação dos parâmetros das esferas e dos sólidos implementados no USCS. As dimensões físicas utilizadas na definição dos sólidos foram, a princípio, especificadas em milímetros, mas a unidade de medida não será considerada uma informação relevante para a realização dos experimentos computacionais. De uma forma geral, o que realmente importa é a proporção entre as medidas das esferas e dos sólidos considerados.

Especificamos o conjunto de esferas disponibilizadas com base nos parâmetros utilizados para o tratamento de tumores cerebrais utilizando o *Gamma Knife*. A Tabela 5.1 lista as características deste conjunto de esferas.

Podemos notar que elas estão ordenadas das maiores para as menores, em con-

Tipo de esfera	Raio	Cor	Disponíveis
tipo 1	9	azul	3
tipo 2	7	verde	4
tipo 3	4	turquesa	8
tipo 4	2	vermelho	10

Tabela 5.1: Tabela com os tipos de esferas disponíveis e suas características

cordância com a heurística de alocação das esferas para o recobrimento. O número de esferas disponíveis por tipo foi previamente ajustado com base nas dimensões dos sólidos utilizados nos recobrimentos, de tal forma a termos um número de esferas por tipo inversamente proporcional às suas dimensões. Aumentar a disponibilidade destas esferas não se traduziria em nenhum benefício, pois o processo de busca iria desprezar as esferas adicionais³.

Quanto à especificação dos sólidos, optamos por dimensões inteiras e que facilitassem verificações geométricas. Desta forma, as dimensões escolhidas são as apresentadas na Tabela 5.2.

Sólido	Comprimento	Largura	Altura
Paralelepípedo	15	9	9
Cubo	12	12	12
Esfera	12	12	12
Elipsóide prolato	15	9	9
Elipsóide oblato	9	15	15

Tabela 5.2: Tabela com as dimensões dos sólidos utilizados

Para a região de segurança escolhemos $\varepsilon = 1,5$. Este valor permitiu um bom aproveitamento das esferas que interceptam a superfície do sólido. É importante comentarmos que o valor de ε limita as dimensões das maiores esferas empregadas no recobrimento. Esta verificação é realizada no USCS, com base no valor do parâmetro $DiamMin(T)$. Para defini-lo, iremos primeiramente apresentar a função suporte a um conjunto convexo não vazio $C \subset \mathbb{R}^3$:

$$\sigma_C(x) = \sup_{y \in C} \langle x, y \rangle.$$

Se tomarmos $v \in \mathbb{R}^3$ unitário (ou seja, v pertence à superfície S^2), $\sigma_C(v)$ representa a distância à origem do plano de suporte na direção v . Em nosso caso, para o sólido T podemos definir

$$DiamMin(T) = \inf_{v \in S^2} (\sigma_T(v) + \sigma_T(-v)).$$

³Esta providência também ajudou a impedir longos processamentos causados por erros, durante a depuração do USCS.

O valor de $DiamMin(T)$ representa a menor distância possível entre dois planos de suporte contendo o sólido. Este parâmetro é trivialmente calculado para os sólidos de nossos experimentos. Para que a esfera S_i possa ser utilizada em um recobrimento, a seguinte desigualdade deve ser válida:

$$r_i \leq \frac{DiamMin(T)}{2} + \varepsilon.$$

Se um tipo de esfera não atende a esta desigualdade, ele é bloqueado e não pode ser empregado no recobrimento.

O USCS realiza o pré-cálculo dos parâmetros M_{ij} , m_{ij} e δ_{ij}^* para cada par de esferas utilizadas. Estes parâmetros são empregados para ajustar o argumento da função de penalização p_u . Por outro lado, o perfil geral da curva p_u foi ajustado com base nos parâmetros $E_2 = 1$, $L_2 = 10^8$, $n = 6$ e $k = 4$, considerados satisfatórios durante a realização dos experimentos.

Para acelerar o cálculo dos volumes utilizados na construção dos índices de cobertura, a grade de integração quasi-Monte Carlo é construída somente uma vez no início do processamento do USCS, e a integração do sólido já é realizada durante este processo. Os pontos da grade localizados além de $1,5\varepsilon$ do sólido são desprezados, pois os experimentos constataram uma violação muito baixa da região de segurança (ou seja, I_{E2} muito próximo de 100%). Os pontos que permanecem são armazenados para servirem à integração das esferas. Desta forma, em cada nó somente são realizadas as computações absolutamente necessárias para ao cálculo dos volumes das esferas, dos volumes de suas interseções mútuas e dos volumes de suas interseções com o sólido.

Escolhemos processar cada nó três vezes e preservar o melhor resultado. Um nó é considerado admissível se o índice de Paddick é maior ou igual ao valor mínimo $IPdkMin$ e o índice I_{B2} é menor ou igual ao valor máximo $IB2Max$. Um nó admissível é comparado com o melhor resultado até aquele momento, que denominamos *incumbent*. Se o nó tiver um índice de Paddick superior ao do *incumbent*, ele passa a ser o *incumbent*. Ao término do processamento o *incumbent* é recuperado e apresentado como solução do problema. É possível que nenhum nó analisado durante a busca atenda aos critérios, então o USCS informa que não encontrou solução para o problema.

Exigir um nível mínimo para o índice de Paddick nos garante automaticamente um bom nível para os índices de cobertura I_{COV} e de interseção I_{E1} . Da mesma forma, exigir um nível máximo de I_{B2} nos garante limitar os valores máximos de I_{B3} e I_{B4} . Para nossos experimentos escolhemos $IPdkMin = 70\%$ e $IB2Max$ variando entre 34% e 45%, dependendo do sólido escolhido para o recobrimento. Estes valores podem ser facilmente ajustados na tela de controle do USCS.

5.3 Resultados

Nesta seção apresentaremos os resultados do USCS para o recobrimento dos sólidos escolhidos. Como o processamento envolve um componente estocástico, estaremos apresentando os resultados mais interessantes, sob um ponto de vista crítico.

A Tabela 5.3 apresenta um resumo dos resultados destes experimentos.

Sólido	Assinatura	Parâmetros		Índices de cobertura					
		IPm	$IB2M$	I_{COV}	I_{E1}	I_{B2}	I_{B3}	I_{B4}	I_{PDK}
Paralel	(3, 1, 5, 0)	70%	34%	80%	91%	33%	6%	0%	73%
	(2, 6, 0, 0)	70%	40%	85%	89%	40%	14%	2%	76%
Cubo	(4, 4, 1, 0)	70%	34%	91%	91%	26%	6%	1%	83%
Esfera	(2, 2, 8, 0)	70%	34%	91%	89%	34%	13%	3%	82%
Elip. prol.	(2, 2, 2, 0)	70%	34%	99%	77%	31%	10%	4%	76%
Elip. obl.	(3, 2, 4, 0)	70%	34%	95%	87%	34%	6%	1%	83%

Tabela 5.3: Tabela com os resultados dos recobrimentos dos sólidos

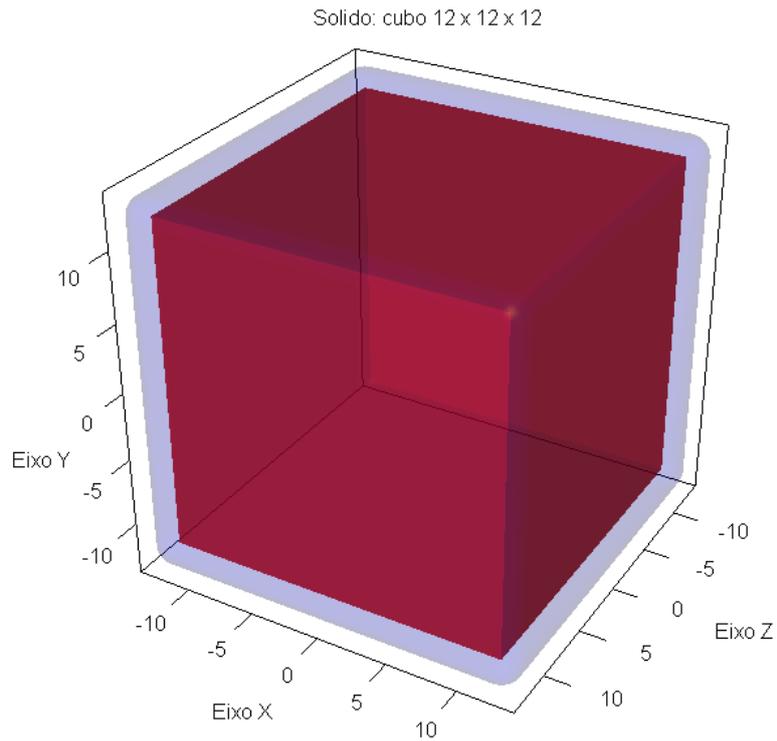
Nesta tabela, temos as seguintes colunas: sólidos, onde o paralelepípedo, o elipsóide prolato e o oblato aparecem com seus nomes abreviados, respectivamente, paralel, elip. prol e elip obl; assinatura do nó *incumbent*; parâmetro $IPdkMin$, abreviado para IPm ; parâmetro $IB2Max$, abreviado para $IB2M$ e índices I_{COV} , I_{E1} , I_{B2} , I_{B3} , I_{B4} e I_{PDK} . O índice de eficiência I_{E2} sempre apresentou um valor muito próximo ou igual a 100%, e foi suprimido da tabela.

Houve dificuldades no recobrimento do paralelepípedo utilizando o valor de $IB2Max = 34%$. Ao aumentarmos para $IB2Max$ para 40% foi possível aprimorarmos I_{COV} e I_{PDK} , ao custo de uma maior interseção entre as esferas. Interpretamos este resultado avaliando a dificuldade inerente de se realizar o recobrimento de um sólido com “quinas”. Este é o motivo de apresentarmos dois resultados para o Paralelepípedo.

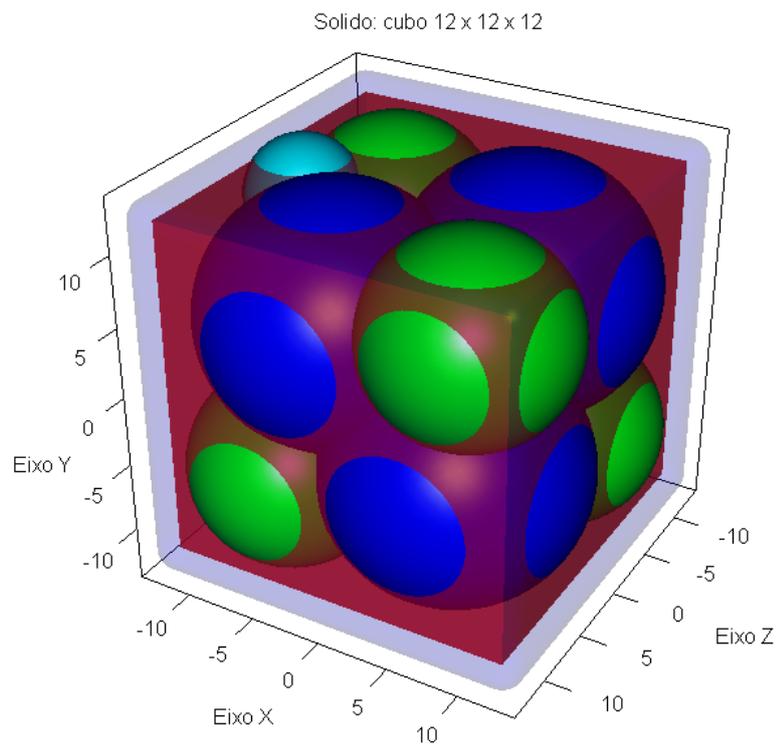
Para o cubo selecionamos um bom resultado, dentre os vários disponíveis. O motivo dos bons resultados está ligado a uma coincidência geométrica entre as dimensões do cubo e das esferas utilizadas, o que permitiu uma disposição bastante regular para a realização do recobrimento.

Para a esfera e os elipsóides prolato e oblato também conseguimos bons recobrimentos, com índices de cobertura I_{COV} superiores a 90% e índices de Paddick da ordem de 80%.

A título de ilustração, apresentamos na Figura 5.1 o resultado do recobrimento do cubo.



(a) Sólido cúbico e sua região de segurança.

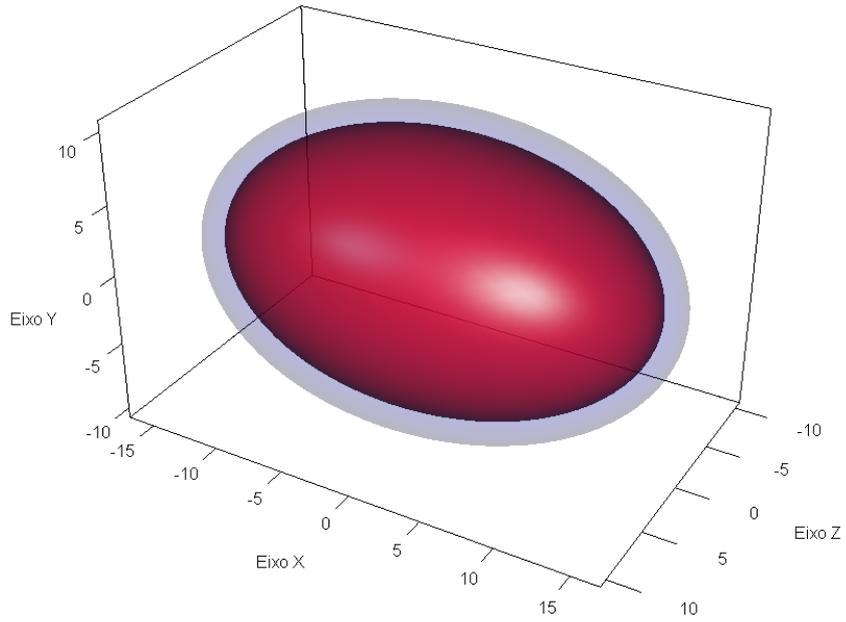


(b) Sólido cúbico e as esferas do revestimento

Figura 5.1: Resultado do revestimento do sólido com o formato de um cubo.

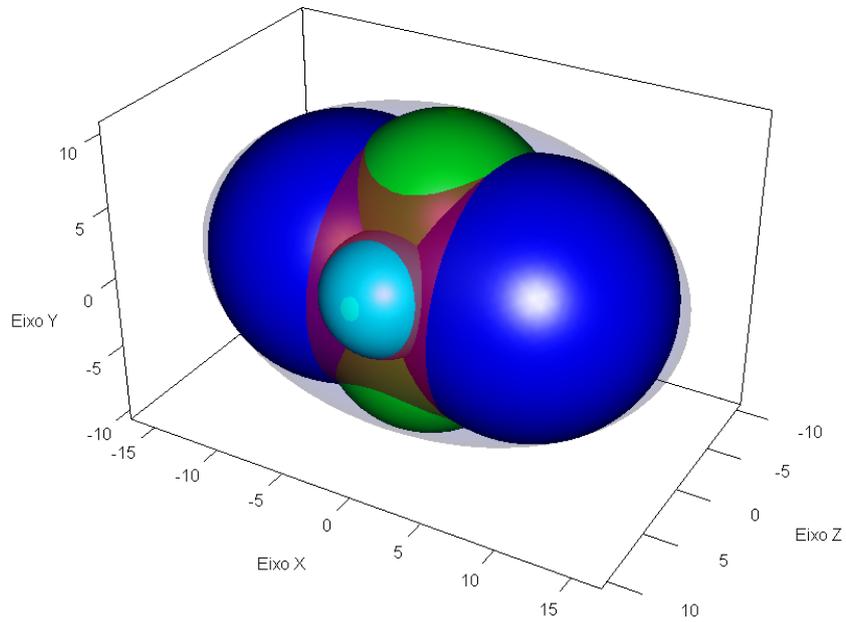
Para os casos de sólidos com fronteiras suaves, apresentamos nas Figuras 5.2 e 5.3 os resultados dos revestimentos do elipsóide prolato e do elipsóide oblato.

Sólido: Elipsoide prolato, 15 x 9 x 9



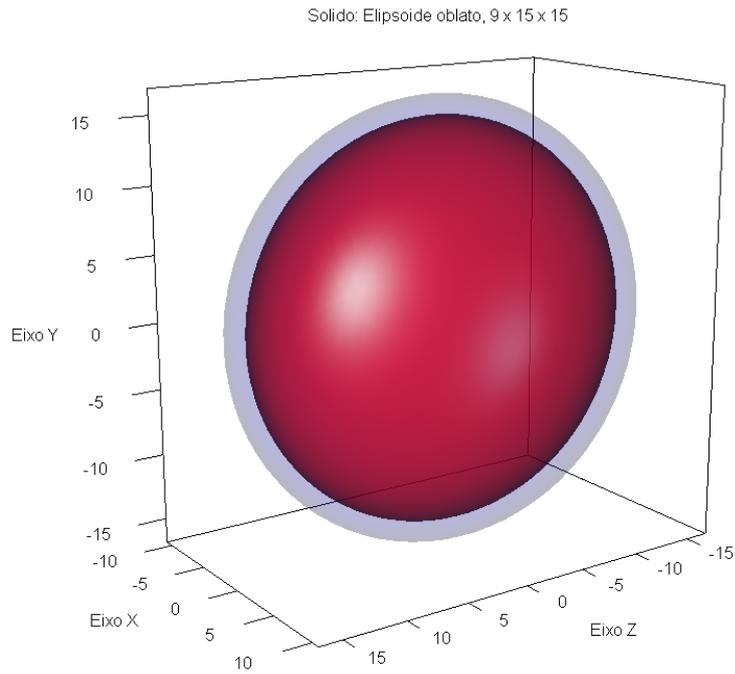
(a) Elipsóide prolato e sua região de segurança.

Sólido: Elipsoide prolato, 15 x 9 x 9

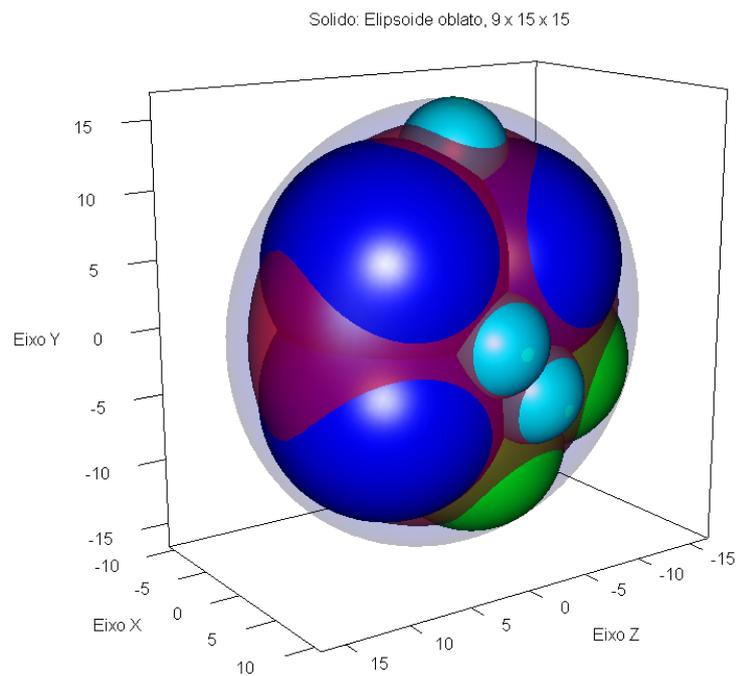


(b) Elipsóide prolato e as esferas do recobrimento.

Figura 5.2: Resultado do recobrimento do Elipsóide prolato.



(a) Elipsóide oblato e sua região de segurança.



(b) Elipsóide oblato e as esferas do recobrimento.

Figura 5.3: Resultado do recobrimento do Elipsóide oblato.

Os resultados dos experimentos foram bastante satisfatórios, considerando-se as dificuldades inerentes ao problema de recobrimento de sólidos utilizando esferas de raios diferentes.

Capítulo 6

Conclusões e trabalhos futuros

Finalizamos tecendo comentários sobre as principais conclusões deste trabalho e sugestões para a sua continuação.

6.1 Conclusões

Durante o desenvolvimento deste trabalho foi possível vivenciar as grandes dificuldades associadas ao recobrimento de corpos sólidos utilizando esferas. Obter uma solução viável, dentro das restrições impostas pelo problema, é com certeza uma tarefa excepcionalmente difícil, tanto em termos de formulação, como também da efetiva implementação e resolução por um solver. Em suma, tudo o que se poderia esperar de um problema classificado como MINLP não-convexo.

Assim sendo, concessões se tornaram absolutamente necessárias na busca por soluções que, mesmo sem ter a otimalidade comprovada, pelo menos apresentem características gerais das soluções consideradas boas. Dentro deste espírito, podemos listar as principais conclusões referentes aos métodos escolhidos na abordagem deste problema:

- o estudo de recobrimentos reticulados no \mathbb{R}^3 foi importante para a realização deste trabalho. A partir dele, obtivemos valores iniciais para os α'_{ij} s. Verificamos ainda que são necessárias no máximo quatro esferas no recobrimento reticulado de corpo centrado para cobrir um ponto e a interseção entre quatro esferas adjacentes é apenas um ponto;
- a afirmação contida no lema 4.1 nos levou a considerações heurísticas para o desenvolvimento de um método para a resolução do problema;
- o estudo de curvas e superfícies paralelas foi de extrema valia para o estabelecimento das regiões de segurança e definição de corpo paralelo;

- a utilização dos métodos de penalidades proporcionou uma abordagem alternativa bastante satisfatória para o problema ($P1$);
- a implementação de uma busca em árvore especialmente adaptada aos propósitos de um bom recobrimento dentro do contexto de aplicação ao *Gamma Knife* se mostrou muito eficiente, como é possível de se comprovar nos tempos de processamento;
- a aplicação do índice de Paddick como medida de uma boa cobertura no contexto de radiocirurgia, juntamente com o índice de interseção I_{B2} , levaram a um critério eficaz para a avaliação da qualidade de um nó;
- devido à restrição imposta pela região de segurança ao problema, dependendo do valor de ϵ e do conjunto de esferas disponível não é possível obter um recobrimento total do corpo.

De uma forma geral podemos concluir que os resultados deste trabalho foram muito satisfatórios. Soluções com bons índices de Paddick e relativamente baixos índices de interseção são facilmente obtidos com o uso do USCS.

6.2 Sugestões para trabalhos futuros

A implementação dos métodos desenvolvidos neste trabalho proporcionou uma visão crítica acerca de suas virtudes e deficiências. Ideias para aprimoramentos surgiam a cada nova etapa da implementação. Algumas destas ideias puderam ser acrescentadas ao corpo teórico deste trabalho e implementadas. Outras ideias, no entanto, se mostraram mais profundas e as restrições de tempo nos obrigaram a postergá-las como sugestões para desenvolvimentos posteriores. Dentre elas, podemos divisar:

- ampliar o escopo da busca em árvore, atualmente restrito à ordem imposta pelo tamanho das esferas utilizadas;
- explorar alternativas de curvas de penalização para distância entre as esferas, principalmente as que permitam perfil mais agressivo para as distâncias menores que a ideal;
- avaliar a possibilidade de inclusão de restrições diretamente ligadas ao aprimoramento dos índices de interseção;
- tratar de corpos não-convexos, ainda com conções geométricas simples;
- desenvolver restrições para o tratamento de corpos geometricamente mais complexos;

- avaliar a implementação da distância exata à elipse, viabilizada pela implementação de funções externas ao AIMMS escritas em linguagem C;
- incluir o valor da função objetivo no processo de escolha das melhores soluções em cada nó;
- avaliar a possibilidade de propor um único índice que considere os efeitos dos índices de interseção.

Referências Bibliográficas

- ADICKES, M., BILLO, R., NORMAN, B., et al., 2002, “Optimization of indoor wireless communication network layouts”, *IIE Transactions*, v. 34, n. 9, pp. 823–836.
- ARAÚJO, P. V., 1998, *Geometria Diferencial*. 1st ed. Rio de Janeiro, IMPA.
- BAMBAH, R. P., 1954, “On lattice coverings by spheres”, *Proc. Nat. Inst. Sci. India*, v. 20, pp. 25–52.
- BOURLAND, J. D., WU, Q. J., 1999, “Morphology-guided radiosurgery treatment planning and optimization for multiple isocenters”, *Med. Phys.*, v. 26, pp. 2151–2160.
- BOYD, S., VANDENBERGHE, L., 2003, *Convex optimization*. Cambridge, UK, Cambridge University Press.
- BRASS, P., WILLIAM, M., PACH, J., 2005, *Research Problems in Discrete Geometry*. 1st ed. New York, Springer.
- BURER, S., LETCHFORD, A. N., 2012, “Non-convex mixed-integer nonlinear programming: a survey”, *Surveys in Operations Research and Management Science*, v. 17, pp. 97–106.
- CASTILLO, I., KAMPAS, F. J., PINTER, J. D., 2008, “Solving circle packing problems by global optimization: Numerical results and industrial applications”, *European Journal of Operation Research*, v. 191, n. 3, pp. 786–802.
- CAYLEY, M. A., 1860, “Sur la courbe parallele a l’ellipse”, *Annali di Matematica Pura ed Applicata*, v. 3, pp. 311–316.
- CONWAY, J. H., SLOANE, N. J. A., 1999, *Sphere Packings, Lattices and Groups*. 3rd ed. New York, Springer-Verlag.

- CONWAY, J. H., JIAO, Y., TORQUATO, S., 2011, “New family of tilings of three-dimensional Euclidean space by tetrahedra and octahedra”, *Proceedings of the National Academy of Sciences of the United States of America*.
- DO CARMO, M. P., 1971, *Elementos de Geometria Diferencial*. 1st ed. Rio de Janeiro, Ao Livro Técnico S.A.
- DO CARMO, M. P., 1976, *Differential Geometry of Curves and Surface*. Englewood Cliffs, New Jersey, Prentice-Hall, Inc.
- DRUD, A., 1985, “CONOPT: A GRG code for large sparse dynamic nonlinear optimization problems”, *Mathematical Programming*, v. 31, pp. 153–191.
- DU, D. Z., PARDALOS, P., WANG, J., 2000, “Discrete Mathematical Problems with Medical Applications”, *American Mathematical Society*, v. 55.
- EDELSBRUNNER, H., KERBER, N., 2011, “Covering and packing with spheres by diagonal distortion in \mathbb{R}^n ”. In: Calude, C. S., Rozenberg, G., Salomaa, A. (Eds.), *Rainbow of computer science*, Springer-Verlag, pp. 20–35.
- FAROUKI, R. T., 1986, “The approximation of non degenerate offset surfaces”, *Comp.Aided Des.*, v. 3, n. 1, pp. 15–44.
- FAROUKI, R., 1985, “Exact offset procedures for simple solids”, *Computer Aided and Geometric Design*, v. 2, pp. 257–279.
- FEJES TÓTH, G., 1999, “Recent Progress on Packing and Covering”. In: *Advances in Discrete and Computational Geometry*, v. 223, Contemporary Mathematics, pp. 145–162, Providence, RI, AMS.
- FEJES TÓTH, L., 1972, *Lagerungen in der Ebene auf der Kugel und im Raum*. 2nd ed. Berlin, Springer-Verlag.
- FEJES TÓTH, L., 1953, *Lagerungen in der Ebene auf der Kugel und im Raum*. 1st ed. Berlin, Springer-Verlag.
- FEJES TÓTH, L., 1964, *Regular figures*. Oxford, Pergamon Press.
- FEJES TÓTH, L., 1940, “Über einen geometrischen Satz”, *Math. Z.*, v. 46, pp. 79–83.
- FERGUSON, S., HALES, T., 2005, “The proof of the Kepler conjecture”, *Annals. Math.* Disponível em: <<http://annals.math.princeton.edu/wp-content/uploads/annals-v162-n3-p01.pdf>>.

- FERRIS, M., SHEPARD, D., 2000, “Optimization of gamma knife radiosurgery”. In: DU, D. Z., PARDALOS, P., WANG, J. (Eds.), *Discrete Mathematical Problems with Medical Applications*, v. 55, *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, AMS, pp. 27–44.
- FERRIS, M., VOELKER, M., 2004, “Fractionation in radiation treatment planning”, *Math. Program.*, v. 101, pp. 387–413.
- FERRIS, M., LIM, J., SHEPARD, D., 2003a, “An optimization approach for the radiosurgery treatment planning”, *SIAM Journal on Optimization*, v. 13, n. 3, pp. 921–937.
- FERRIS, M., LIM, J., SHEPARD, D., 2003b, “Radiosurgery treatment planning via nonlinear programming”, *Annals of Operations Research*, v. 119, pp. 247–260.
- FEUVRET, L., NOEL, G., MAZERON, J. J., et al., 2006, “Conformity index: a review”, *Int. J. Radiation Oncology Biol. Phys.*, v. 64, n. 2, pp. 333–342.
- FLOUDAS, C. A., 1995, *Nonlinear and Mixed-Integer Optimization: Fundamentals and Applications*. Oxford University Press.
- FUKSHANSKY, L., 2011, “Revisiting the hexagonal lattice: on optimal lattice circle packing”, *Elemente der Mathematik*, v. 66, n. 1, pp. 1–9.
- GAMMA KNIFE CENTER, Florida Hospital. Disponível em: <<http://www.gamma-knife-surgery.com/gammaKnife/index.asp>>.
- GANZ, J. C., 1997, *Gamma Knife surgery*. 2nd revised ed. New York, Wien-Springer.
- GANZ, J. C., 2011, *Gamma Knife Neurosurgery*. 1st ed. New York, Wien-Springer.
- GAUSS, C. F., 1831, *Untersuchungen über die Eigenschaften der positiven ternären quadratischen Formen von Ludwig*. August Seber.
- GIAQUINTA, M., MODICA, G., “Convex Sets and Convex Functions”. In: *Mathematical Analysis: Foundations and Advanced Techniques for Functions of Several Variables*, Birkhauser, Boston.
- GOODMAN, J. E., ROURKE, J. O., 2004, *Handbook of Discrete and Computational Geometry*. 2nd ed. Boca Raton, FL, Chapman & Hall/CRC.
- GRAY, A., 1996, *Modern Differential Geometry of Curves and Surfaces with Mathematica*. 1st ed. Boca Raton, FL, USA, CRC Press, Inc.

- HALES, T. C., 2000, “Cannonballs and Honeycombs”, *Notices Amer. Math. Soc.*, v. 47, n. 4, pp. 440–449. Disponível em: <<http://www.ams.org/notices/200004/feahales.pdf>>.
- HALES, T. C., 2005, “A proof of the Kepler Conjecture”, *Annals of Mathematics, Second Series*, v. 162, n. 3, pp. 1065–1185.
- HALES, T. C., 2006, “Historical Overview of the Kepler Conjecture”, *Discrete Comput. Geom.*, v. 36, pp. 5–20.
- HIFI, M., M’HALLAH, R., 2009a, “Beam search and non-linear programming tools for the circular packing problem”, *International Journal of Mathematics in Operational Research*, v. 1, n. 4.
- HIFI, M., M’HALLAH, R., 2009b, “A Literature Review on Circle and Sphere Packing Problems: Models and Methodologies”, *Advances in Operations Research, Hindawi Publishing Corporation*, v. 2009.
- HILBERT, D., 1901, “Mathematische Probleme”, *Archiv. Math. Phys.*, v. 1, pp. 44–63.
- HORST, R., PARDALOS, P., THOAI, N. V., 1995, *Introduction to Global Optimization*. 1st ed. Dordrecht, Kluwer Academic Publishers.
- JITPRAPAIKULSARN, S., 2005, *An optimization-based treatmentplanner for gamma knife radiosurgery*. Tese de Doutorado, Case Western Reserve University.
- JONES, J. E., 1924, “On the Determination of Molecular Fields. II. From the Equation of State of a Gas”, *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, v. 106, n. 738, pp. 463–477.
- KEPLER, J., 1611, *Strena seu de nive sexangula (The six-cornered snowflake)*.
- KEPLER, J., 1966, *The six-cornered snowflake*. Oxford University Press.
- KERSHNER, R., 1939, “The number of circles covering a set”, *AMJ*, v. 61, pp. 665–671.
- KUCHERENKO, S., BELOTTI, P., LIBERTI, L., et al., 2007, “New formulations for the Kissing Number Problem”, *Discrete Applied Mathematics*, v. 155, pp. 1837–1841.

- LEE, S., GROSSMANN, I. E., 2000, “New Algorithms for Nonlinear Generalized Disjunctive Programming”, *Comput. Chem. Eng.*, v. 24, n. 9-10, pp. 2125–2141.
- LEE, S., GROSSMANN, I. E., 2001, “A global optimization algorithm for non-convex generalized disjunctive programming and applications to process systems”, *Comput. Chem. Eng.*, v. 25, pp. 1675–1697.
- LEECH, J., 1956, “The Problem of the thirteen spheres”, *Math. Gaz.*, v. 40 (Feb), pp. 22–23.
- LEKSELL, L., 1968, “Cerebral radiosurgery”, *Acta Chirurgica Scandinavica*, v. 134, n. 8, pp. 585–595.
- LEKSELL, L., 1951, “The stereotactic method and radiosurgery of the brain”, *Acta Chirurgica Scandinavica*, v. 102, pp. 316–319.
- LIBERTI, L., MACULAN, N. (Eds.), 2006, *Global Optimization: from theory to Implementation*. Berlin, Springer.
- LIBERTI, L., MACULAN, N., ZHANG, Y., 2009, “Optimal configuration of gamma ray machine radiosurgery units: the sphere covering subproblem”, *Optimization Letters*, v. 3, pp. 109–121.
- LIM, J., 2002, *Optimization in radiation treatment planning*. Tese de Doutorado, University of Wisconsin, Madison.
- LINDSEY, J. H., 1986, “Sphere-packing in \mathbb{R}^3 ”, *Math.*, v. 33, n. 1, pp. 137–147.
- LOCATELLI, M., SCHOEN, F., 2003, “Efficient algorithms for large scale global optimization: Lennard-Jones clusters”, *Computational Optimization and Applications*, v. 26, n. 9, pp. 173–190.
- LUBKE, D. C., 2014, *Cobertura de corpos por esferas utilizando suavização hiperbólica*. Tese de Mestrado, COPPE/UFRJ, Rio de Janeiro.
- LUENBERGER, D. G., YE, Y., 2008, *Linear and Nonlinear Programming*. 3rd ed. New York, Springer Science+Business Media.
- MACULAN, N., FAMPA, M. H. C., 2006, *Otimização linear*. Unb.
- MAEKAWA, T., 1998a, “Self-intersections of offsets of quadratic surfaces: Part I, explicit surfaces”, *Engineering with Computers*, v. 14, pp. 1–13.
- MAEKAWA, T., 1998b, “Self-intersections of offsets of quadratic surfaces: Part II, implicit surfaces”, *Engineering with Computers*, v. 14, pp. 14–22.

- MAEKAWA, T., 1999, “An overview of offset curves and surfaces”, *Computer Aided Geometric Design*, v. 39, pp. 165–173.
- MATEMATICASVISUALES, “Truncated octahedron cube”, Disponível em: <<http://matematicasvisuales.com/html/espacio/truncatedoctahedroncube.html>>.
- MECKE, K. R., 2002, “The shape of parallel surfaces: porous media, fluctuating interfaces and complex fluids”, *Physica A: Statistical Mechanics and its Applications*, v. 314, pp. 655–662.
- MLADNOVIC, N., PLASTIA, F., UROSEVIC, D., 2005, “Reformulation descent applied to circle packing problems”, *Computer & Operations Research*, v. 32, n. 9, pp. 2419–2434.
- MOROKOFF, W. J., CAFLISCH, R. E., 1995, “Quasi-Monte Carlo Integration”, *Journal of Computational Physics*, v. 122, pp. 218–230.
- MUDER, D. J., 1993, “A new bound on the local density of sphere packings”, *DCG*, v. 10, pp. 351–375.
- NASCIMENTO, R. Q., MACAMBIRA, A. F. U. S., CABRAL, L. F., et al., 2010, “The Ellipsoid Covering Problem: an Continuous Approach”. In: *2nd International Conference on Engineering Optimization*, Lisboa.
- OSTERGARD, P. R. J., 2008, “Book Review”, *Computers and Operations Research*, v. 36, pp. 276–278.
- PADDICK, I. I., 2000, “A simple scoring ratio to index the conformity of radio-surgical treatment plans. Technical note”, *J. Neurosurg.*, v. 93, Suppl 3, pp. 219–222.
- PARDALOS, P. M., COLEMAN, T. F., 2009, “Lectures on Global Optimization”, *American Mathematical Society*, v. 55.
- PARDALOS, P. M., EDWIN, H. (Eds.), 2009, *Handbook of Optimization in Medicine, Series Springer Optimization and Its Applications*, v. 26. Springer.
- PATRIKALAKIS, N. M., MAEKAWA, T., 2010, “Offset Curves and Surfaces”. In: *Shape Interrogation for Computer Aided Design and Manufacturing*, Springer Science+Business Media, cap. 11. Disponível em: <<http://web.mit.edu/hyperbook/Patrikalakis-Maekawa-Cho/node210.html>>.
- PHAM, B., 1992, “Offset curves and surfaces: a brief survey”, *Computer Aided Design*, v. 24, n. 4.

- PINTO, R. V., 2011, *Um enfoque para o problema de recobrimento de sólidos por esferas de diâmetros diferentes*. Tese de Mestrado, COPPE/UFRJ, Rio de Janeiro.
- PINTO, R. V., 2015, *O problema de recobrimento de sólidos por esferas de diâmetros diferentes*. Tese de Doutorado, COPPE/UFRJ, Rio de Janeiro.
- PRIX, R., 2007, “Template-based searches for gravitational waves: efficient lattice covering of flat parameter spaces”, *Classical and Quantum Gravity*, v. 24, pp. 5481–5490.
- ROGERS, C. A., 1958, “The packing of equal spheres”, *Proceedings of the London Mathematical Society, Third Series*, v. 8, n. 4, pp. 609–620.
- ROGERS, C. A., 1964, *Packing and Covering, Cambridge Tracts in Mathematics and Mathematical Physics*, v. 54. England, Cambridge University Press.
- SCHURMAN, A., VALLENTIN, F., 2006, “Computational Approaches to Lattice Packing and Covering Problems”, *Discrete and Computational Geometry*, v. 35, n. 1, pp. 73–116.
- SHAW, E., 1993, “Radiation Therapy Oncology Group: Radiosurgery quality assurance guidelines”, *Int. J. Radiat. Oncol., Biol., Phys.*, v. 27, pp. 1231–1239.
- SHEPARD, D. M., FERRIS, M. C., OVE, R., et al., 2000, “Inverse Treatment Planning for Gamma Knife Radiosurgery”, *Medical Physics*, v. 27, pp. 2748–2756.
- SOUTOU, A., DAI, Y., 2002, “Global optimization approach to unequal sphere packing problems in 3D”, *J. Optim. Theory Appl.*, v. 114, n. 3, pp. 671–694.
- SPIRAKIS, P. G., 1985, “The volume of the union of many spheres and point inclusion problems”, *LNCS, STACS 85*, v. 128, pp. 328–338.
- STOYAN, Y. G., YAS’KOV, G., 2004, “A mathematical model and a solution method for the problem of placing varioussized circles into a strip”, *European Journal of Operational Research*, v. 156, n. 3, pp. 590–600.
- SZABÓ, P. G., MARKÓT, M. C., SCENDES, T., et al., 2007, *New Approaches to Circle Packings in a Square; with Program Codes*, v. 6. Berlin, Springer.

- SZPIRO, G. G., 2003, *Kepler's conjecture. How some of the greatest minds in history helped solve one of the oldest math problems in the world*. John Wiley & Sons.
- TAWARMALANI, M., SAHINIDIS, N., 2002, *Convexification and Global Optimization in Continuous and Mixed-Integer Nonlinear Programming; Theory, Algorithms, Software, and Applications*. Dordrecht, Kluwer Academic Publishers.
- THUE, A., 1892, "Om ongle geometrisk taltheortiske Theoremer", *Forand. Skand. Natur.*, v. 14, pp. 352–353.
- TORQUATO, S., 2010, "Reformulation of the covering and quantizer problems as ground states of interacting particles", *Physical Review E*, v. 82, 056109.
- TORQUATO, S., JIAO, Y., 2010, "Exact constructions of a family of dense periodic packings of tetrahedra", *Physical Review E*, v. 81, 041310.
- UHLER, C., WRIGHT, S. J., 2013, "Packing ellipsoids with overlap", *Society for Industrial and Applied Mathematics*, v. 55, n. 4.
- VALLENTIN, F., 2003, *Sphere Coverings, Lattices and Tilings (in low dimensions)*. Tese de Doutorado, Technical University Munich, Germany.
- WANG, J., 1999, "Packing of unequal spheres and automated radiosurgical treatment planning", *Journal of Combinatorial Optimization*, v. 3, n. 4, pp. 453–463.
- WEISSTEIN, E. W., "Kissing Number", *MathWorld-A Wolfram Web Resource*. Disponível em: <<http://mathworld.wolfram.com/KissingNumber.html>>.
- WIKIPEDIA, "Convex uniform honeycomb", Disponível em: <http://en.wikipedia.org/wiki/Convex_uniform_honeycomb>.
- WU, A., 1992, "Physics and dosimetry of the gamma knife", *Neurosurgery Clinics of North America*, v. 3, pp. 35–50.
- WU, Q. J., WESSELS, B. W., EINSTEIN, D. B., et al., 2003, "Quality of coverage: conformity measures for stereotactic radiosurgery", *Journal of Applied Clinical Medical Physics*, v. 4, n. 4, pp. 374–381.
- WU, Q. R., 1996, *Treatment planning optimization for gamma unit radiosurgery*. Tese de Doutorado, The Mayo Graduate School.

- XAVIER, A. E., OLIVEIRA, A. A. F., 2005, “Optimal Covering of Plane Domains by Circles Via Hyperbolic Smoothing”, *Journal of Global Optimization*, v. 31, pp. 403–504.
- XAVIER, A. E., MAGALHÃES, P. C., SILVA, L. P., 1989, “Calibração Automática de Modelos Chuva-Vazão: Um Método Assintótico”, *Pesquisa Operacional*, v. 10, n. 1, pp. 1–10.
- YAN, Y., SHU, H., BAO, X., 1997, “Clinical treatment planning optimization by Powell’s method for Gamma unit treatment system”, *International Journal of Radiation Oncology, Biology and Physics*, v. 39, pp. 247–254.
- YU, G. S., YAS’KOV, G., 2004, “A mathematical model and a solution method for the problem of placing various-sized circles into a strip”, *European Journal of Operation Research*, v. 156, n. 3, pp. 590–600.
- ZONG, C., TALBOT, J., 1999, *Sphere Packings*. New York, Springer-Verlag.

Apêndice A

Curvas e superfícies paralelas

Em muitas situações precisamos construir uma curva ou superfície que esteja localizada a uma certa distância de uma curva ou superfície dada. É o caso, por exemplo, de aplicações industriais tais como o planejamento do trajeto de robôs e a geração de caminhos para o controle numérico de ferramentas nas indústrias têxtil, automobilísticas e de calçados.

Em termos formais, dada uma curva C no \mathbb{R}^2 , desejamos obter uma nova curva C_d onde cada um de seus pontos está a uma distância fixa d ao longo da normal da curva original C . Chamaremos C_d de curva paralela à C . Analogamente, dada uma superfície S no \mathbb{R}^3 , podemos obter outra superfície S_d , situada a uma distância fixa d ao longo da normal da superfície original S . A superfície S_d é chamada de superfície paralela à S .

As curvas são geralmente representadas na forma paramétrica, uma vez que o parâmetro fornece uma correspondência entre a curva dada e sua curva paralela.

Uma curva plana parametrizada e diferenciável é uma aplicação $C : I \subset \mathbb{R} \rightarrow \mathbb{R}^2$ dada por $C(t) = (x(t), y(t))$, com x e y diferenciáveis. Diz-se que C é uma *curva regular* se o vetor tangente $C'(t) = (x'(t), y'(t)) \neq 0$ para todo $t \in I$, e parametrizada pelo comprimento de arco se $|C'(t)| = 1$ para todo $t \in I$.

Dada uma curva $C : I \rightarrow \mathbb{R}^2$ regular parametrizada pelo comprimento de arco a equação de uma *curva paralela* C_d com uma distância d de C é descrita por:

$$C_d(t) = C(t) \pm d\mathbf{N}(t),$$

onde o sinal mais (menos) é usado se a curva está à esquerda (à direita) da curva original. $\mathbf{N}(t)$ é o vetor normal unitário à curva $C(t)$ no ponto (x, y) dado pela expressão:

$$\mathbf{N}(t) = \left(\frac{-y'(t)}{\sqrt{(x'(t))^2 + (y'(t))^2}}, \frac{x'(t)}{\sqrt{(x'(t))^2 + (y'(t))^2}} \right).$$

Mostra-se que que o lugar geométrico dos parâmetros não regulares das paralelas

C_d é a evoluta de C . Mais ainda, se a curva C for regular, a curva C_d também será regular para valores pequenos de d (GRAY, 1996).

Em geral, uma curva paralela não é do mesmo tipo da curva geradora, como mostram os dois exemplos apresentados nas s A.1 e A.2. Observe que, embora as curvas paralelas à elipse com distâncias d muito próximas da curva geradora pareçam ter a forma de uma elipse, elas não são elipses. As retas e as circunferências são exemplos especiais de curvas cujas paralelas são do mesmo tipo das curvas geradoras.

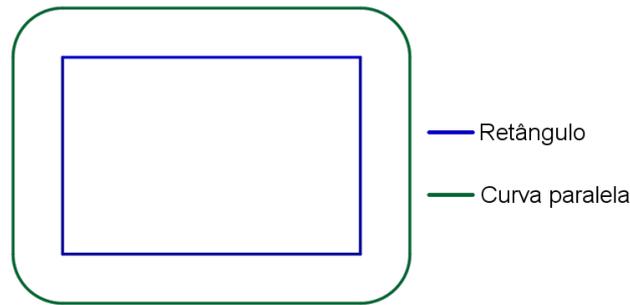


Figura A.1: Um retângulo e uma curva paralela.

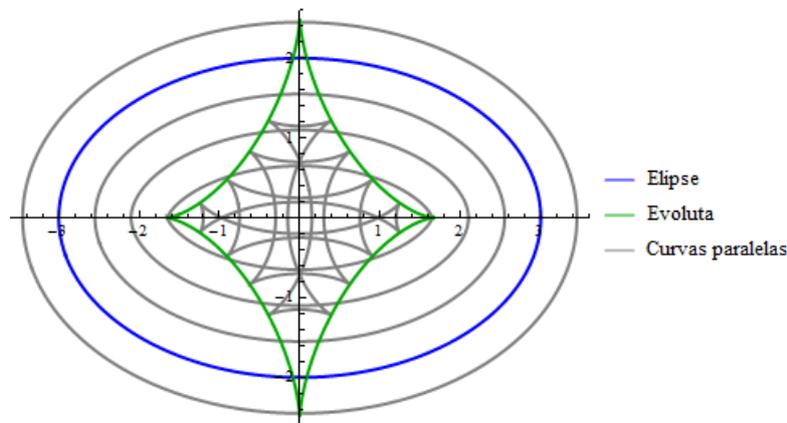


Figura A.2: Uma elipse com $a = 3$ e $b = 2$ e algumas de suas curvas paralelas.

Por analogia com a definição de curva parametrizada, definimos superfície parametrizada como uma aplicação diferenciável $\mathbf{x}(u, v) = (x(u, v), y(u, v), z(u, v))$ de um aberto conexo $U \subset \mathbb{R}^2$ em \mathbb{R}^3 , com $(u, v) \in U$, cujo vetor gradiente é não-nulo em todos os pontos (u, v) .

DO CARMO (1976) e ARAÚJO (1998) fazem uma boa apresentação das propriedades elementares das superfícies no \mathbb{R}^3 . Citaremos apenas as definições e resultados necessários para o desenvolvimento desse trabalho.

A grosso modo, uma *superfície regular* do \mathbb{R}^3 é obtida a partir de pedaços do plano, deformando-os e colando-os entre si, de tal modo que a figura resultante não apresenta pontas, arestas ou auto-interseções, e que tenha sentido falar em plano

tangente nos pontos dessa figura. Formalmente, podemos definir uma *superfície regular* como (ARAÚJO, 1998; DO CARMO, 1971):

Definição A.1. Uma superfície S contida no \mathbb{R}^3 é regular em um ponto $p \in S$ se existem $U \subset \mathbb{R}^2$ e $V \subset \mathbb{R}^3$ abertos, com $q \in U$, e uma aplicação $\mathbf{x} : U \rightarrow V \cap E$, com $\mathbf{x}(q) = p$, tal que \mathbf{x} é diferenciável, \mathbf{x} é um homeomorfismo entre U e $V \cap S$ e a diferencial $d\mathbf{x}_q$ é biunívoca. Se S é regular em todos os seus pontos $p \in S$, então S é uma *superfície regular*.

Denotemos por x_u e x_v as derivadas parciais de $x(u, v)$ com respeito a u e v , respectivamente. A condição de regularidade da superfície S em um ponto p (diferencial $d\mathbf{x}_q$ é biunívoca) pode ser expressa pelo produto vetorial $\mathbf{x}_u \times \mathbf{x}_v \neq 0$, sendo p chamado de *ponto regular*.

O vetor normal unitário de uma superfície regular é dado por

$$\mathbf{N}(u, v) = \frac{\mathbf{x}_u \times \mathbf{x}_v}{\|\mathbf{x}_u \times \mathbf{x}_v\|}.$$

A direção da normal, por convenção, é dada pela regra da mão direita entre os vetores \mathbf{x}_u , \mathbf{x}_v e \mathbf{N} .

Podemos, agora, definir uma superfície paralela a uma distância d ao longo de uma normal a uma superfície:

Definição A.2. Dada uma superfície regular S parametrizada por $\mathbf{x}(u, v)$ a equação de uma *superfície paralela* S_d com uma distância d de S é descrita por:

$$\mathbf{x}_d(u, v) = \mathbf{x}(u, v) \pm d\mathbf{N}(u, v)$$

onde a distância d pode ser um número real positivo ou negativo.

Uma questão importante é como determinar os valores de d para os quais C_d e S_d sejam regulares. O raio de curvatura da curva C é preponderante para determinar um d_0 , chamada distância de deslocamento crítica, a partir do qual, se $-d_0 < d < d_0$, C_d será regular. No caso das superfícies, FAROUKI (1986) apresenta uma análise detalhada da curvatura que identifica os casos de degenerescência. As degenerescências do tipo *ridges* nas superfícies paralelas, ocorrem quando uma das duas curvaturas principais é igual a $-1/d$. Quando ambas as curvaturas principais são iguais a $-1/d$, resultam em *cúspides* (*cusps*). Assim, para se obter boas aproximações para as superfícies paralelas, a distância de deslocamento d é geralmente escolhida sendo significativamente menor do que ambas as curvaturas principais e, portanto, superfícies paralelas com degenerescências não são induzidas. Assim, uma curva ou superfície paralela pode se degenerar mesmo que a curva ou superfície geradora seja regular. Dessa forma, para gerar uma superfície paralela é necessário

estudar os casos de degenerescências. MAEKAWA (1998a,b) faz um estudo das degenerescências em superfícies paralelas às superfícies quádricas. A Figura A.3 ilustra um exemplo de um elipsóide e três de suas superfícies paralelas, enquanto a Figura A.4 mostra duas superfícies paralelas degeneradas ao mesmo elipsóide com distância de deslocamento $d = -3$ e $d = -6$.

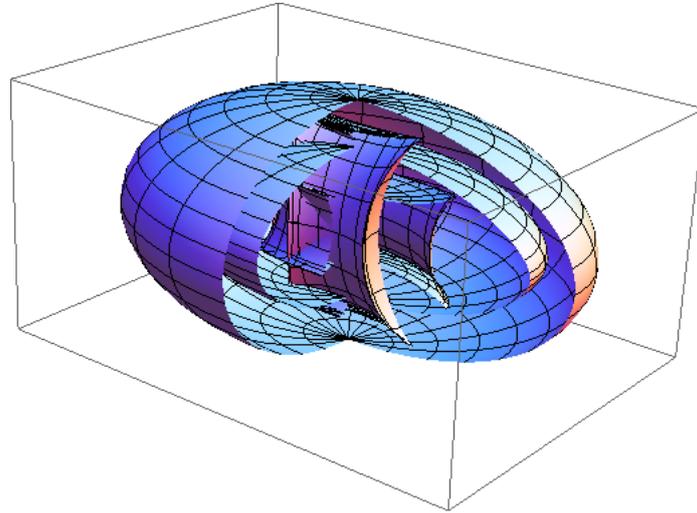


Figura A.3: Um elipsóide com $a = 8$, $b = 6$ e $c = 4$ e três de suas superfícies paralelas internas.

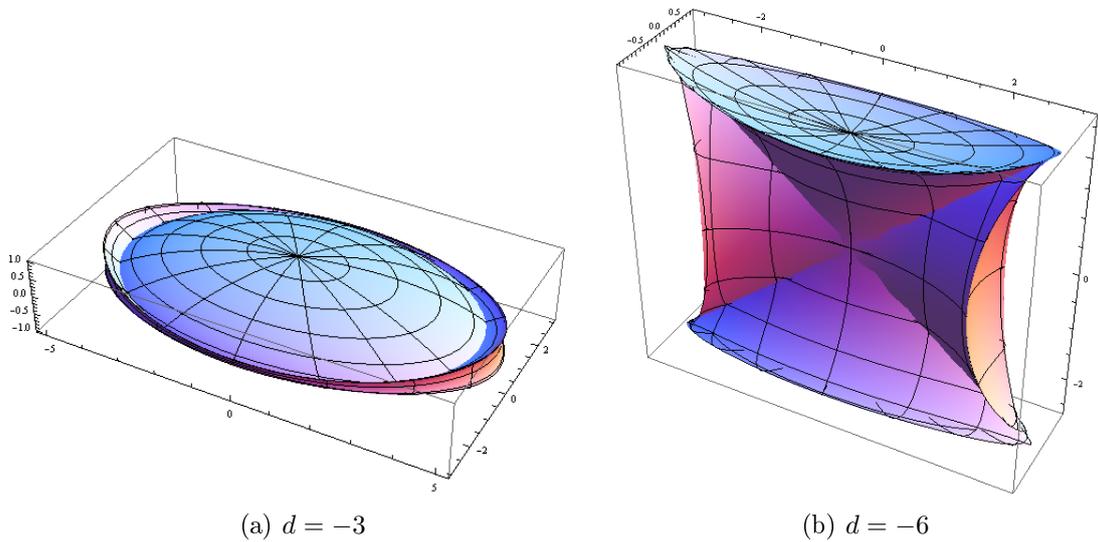


Figura A.4: Superfícies paralelas degeneradas a uma distância de deslocamento d do elipsóide, com $a = 8$, $b = 6$ e $c = 4$.

Apêndice B

Sobre a distância de um ponto a uma elipse e a um elipsóide.

B.1 O envelope das normais a uma elipse

Apresentaremos agora a equação do envelope das normais à elipse. A curva que representa este envelope terá um papel central na determinação da distância de um ponto à elipse.

Consideremos a função f de \mathbb{R}^2 em \mathbb{R} definida por

$$f(x, y) = \frac{x^2}{a^2} + \frac{y^2}{b^2} - 1,$$

com a e b reais, $a > b$.

Tomando $f(x, y) = 0$, a equação algébrica

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - 1 = 0,$$

representa uma elipse centrada na origem sendo a e b os comprimentos dos semi-eixos horizontal e vertical, respectivamente. Como o gradiente de f em um ponto (x, y) é ortogonal às curvas de nível $f(x, y) = k$, k real, (caso regular) e a elipse é representada pela curva de nível $f(x, y) = 0$, as normais à elipse podem ser expressas baseando-se no gradiente

$$\nabla f(x, y) = 2 \left(\frac{x}{a^2}, \frac{y}{b^2} \right).$$

Utilizaremos a seguinte parametrização:

$$\begin{aligned} x(\theta) &= a \cos \theta, \\ y(\theta) &= b \sin \theta. \end{aligned}$$

Desta forma, temos também que o gradiente, em função do parâmetro θ , é dado

por

$$\nabla f(\theta) = 2 \left(\frac{x}{a^2}, \frac{y}{b^2} \right) = 2 \left(\frac{\cos \theta}{a}, \frac{\sin \theta}{b} \right).$$

A equação da reta normal passando por $(x(\theta), y(\theta))$, para um determinado θ fixo, é portanto dada pela equação simétrica da reta

$$\frac{x - a \cos \theta}{\frac{\cos \theta}{a}} = \frac{y - b \sin \theta}{\frac{\sin \theta}{b}},$$

que pode ser desenvolvida como

$$a \sin \theta (x - a \cos \theta) = b \cos \theta (y - b \sin \theta),$$

ou, ainda,

$$ax \sin \theta - by \cos \theta - c^2 \sin \theta \cos \theta = 0,$$

onde $c^2 = a^2 - b^2$.

Para cada valor do parâmetro θ , a equação acima representa a normal à elipse passando pelo ponto $(x(\theta), y(\theta))$. Se denotarmos por $g(x, y, \theta)$ o membro esquerdo desta equação, o cálculo do envelope das normais à elipse nos leva a considerar a equação adicional

$$\frac{\partial g}{\partial \theta} = 0 \Rightarrow ax \cos \theta + by \sin \theta - c^2 (\cos^2 \theta - \sin^2 \theta) = 0.$$

Obtemos, portanto, o seguinte sistema de equações:

$$ax \sin \theta - by \cos \theta - c^2 \sin \theta \cos \theta = 0 \tag{B.1}$$

$$ax \cos \theta + by \sin \theta - c^2 (\cos^2 \theta - \sin^2 \theta) = 0 \tag{B.2}$$

Se agora multiplicarmos (B.1) por $\cos \theta$, (B.2) por $\sin \theta$ e subtrairmos a primeira equação da segunda equação obtemos

$$-\frac{by}{c^2} = \sin^3 \theta. \tag{B.3}$$

Por outro lado, se multiplicarmos (B.1) por $\sin \theta$, (B.2) por $\cos \theta$ e somarmos a primeira equação com a segunda equação temos que

$$\frac{ax}{c^2} = \cos^3 \theta. \tag{B.4}$$

Combinando as equações (B.3) e (B.4) obtemos finalmente a equação do envelope das normais à elipse:

$$(ax)^{\frac{2}{3}} + (by)^{\frac{2}{3}} = (c^2)^{\frac{2}{3}}.$$

Esta equação representa uma astróide, e coincide com a evoluta da elipse, que é o lugar geométrico dos centros de curvatura determinados por cada ponto da elipse.

Para $y = 0$ temos $x = \pm c^2/a$, e para $x = 0$ temos $y = \pm c^2/b$. A Figura B.1 apresenta o gráfico de uma elipse e de sua evoluta.

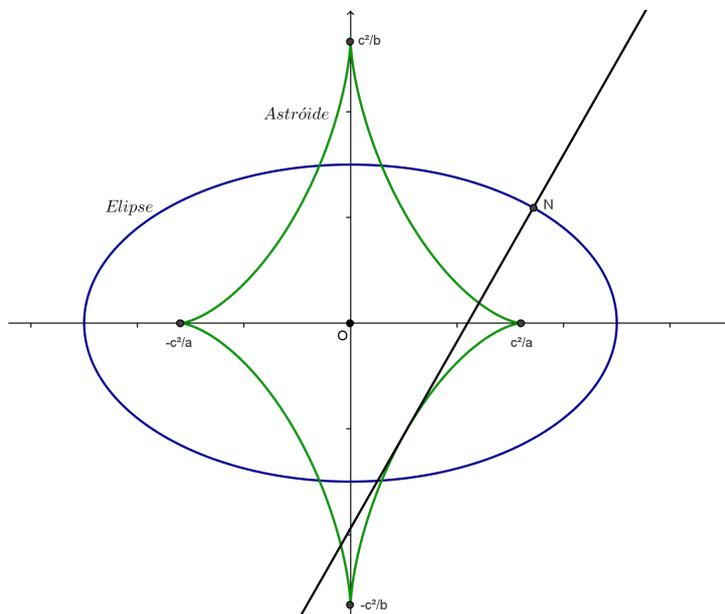


Figura B.1: Uma elipse acompanhada por uma reta normal passando pelo ponto N e por sua evoluta, a astróide.

Notemos que, por definição do envelope de uma curva, as retas normais são tangentes à astróide. Por consequência disto, as normais passando pelos vértices horizontais são tangentes horizontais à astróide e as normais passando pelos vértices verticais são tangentes verticais à astróide. Desta forma, podemos interpretar o gráfico da astróide como sendo constituído de quatro ramos iguais dispostos em cada quadrante do plano, simétricos com relação à origem.

B.2 Determinação da distância de um ponto a uma elipse

Formulação do problema e considerações iniciais

Consideremos novamente uma elipse centrada na origem, que denotaremos por E , dada pela equação

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

que também pode ser escrita como

$$b^2x^2 + a^2y^2 - a^2b^2 = 0,$$

onde a e b , $a > b$, representam os comprimentos dos semi-eixos horizontal e vertical, respectivamente. Desejamos calcular a distância de um ponto arbitrário $P(u, v)$ à elipse. Como a distância de um ponto P a um conjunto fechado de pontos M é definida como

$$d(P, M) = \min_{m \in M} d(P, m),$$

podemos, então, definir a distância do ponto $P(u, v)$ à elipse E como um problema de otimização

$$d(P, E) = \min_{(x, y) \in E} \sqrt{(x - u)^2 + (y - v)^2}.$$

Como a raiz quadrada é uma função crescente, a solução (x^*, y^*) para o problema acima pode ser calculada buscando-se o mínimo do quadrado da distância, e portanto este problema pode ser expresso sob a forma

$$d(P, E)^2 = \min_{(x, y) \in E} (x - u)^2 + (y - v)^2.$$

Esta versão do problema pode ser resolvida analiticamente mediante a utilização de multiplicadores de Lagrange. Se definirmos as funções de \mathbb{R}^2 em \mathbb{R}

$$f(x, y) = (x - u)^2 + (y - v)^2$$

e

$$g(x, y) = b^2x^2 + a^2y^2 - a^2b^2,$$

podemos, então, formular o problema como

$$\begin{aligned} g(x, y) &= 0, \\ \nabla f(x, y) &= \lambda \nabla g(x, y), \end{aligned}$$

e buscar os pontos críticos deste sistema.

A segunda equação caracteriza as soluções como sendo aquelas onde a direção de maior variação da distância ao ponto $P(u, v)$, que é uma direção cuja reta suporte contém o ponto $P(u, v)$, coincide com a direção ortogonal às curvas de nível da função $g(x, y)$. Portanto, as soluções para este sistema são pontos cujas retas normais à elipse passam pelo ponto $P(u, v)$.

Utilizando as definições de $f(x, y)$ e $g(x, y)$, o problema pode ser reescrito como:

$$\begin{aligned} b^2x^2 + a^2y^2 - a^2b^2 &= 0 \\ 2(x - u) &= 2\lambda b^2x, \\ 2(y - v) &= 2\lambda a^2y. \end{aligned}$$

Fazendo as simplificações possíveis e multiplicando a 1ª equação por a^2y e a 2ª, por b^2x , este sistema de equações, se reduz à forma:

$$\begin{aligned} b^2x^2 + a^2y^2 - a^2b^2 &= 0 \\ a^2y(x - u) &= \lambda a^2b^2xy, \\ b^2x(y - v) &= \lambda a^2b^2xy. \end{aligned}$$

Eliminando λ nas duas últimas equações, obtemos a seguinte equação:

$$a^2y(x - u) = b^2x(y - v).$$

Esta última expressão pode ser finalmente simplificada para

$$c^2xy + b^2vx - a^2uy = 0,$$

onde $c^2 = a^2 - b^2$. Portanto, o ponto (x, y) pertencente à elipse que minimiza a distância ao ponto $P(u, v)$ (ou, de forma equivalente, cuja normal à elipse passa pelo ponto $P(u, v)$), pode ser determinado pelo sistema de equações:

$$b^2x^2 + a^2y^2 - a^2b^2 = 0, \tag{B.5}$$

$$c^2xy + b^2vx - a^2uy = 0. \tag{B.6}$$

A equação (B.5) é a própria equação da elipse e simplesmente restringe o ponto (x, y) a pertencer à elipse. Os polinômios em x e y que aparecem nas equações (B.5) e (B.6) são ambos do segundo grau e portanto, pelo Teorema de Bézout para curvas algébricas, podemos ter até quatro soluções reais para a interseção das curvas algébricas associadas a cada um destes polinômios.

Mas qual curva algébrica está associada à equação (B.6)? Notemos que, se considerarmos $c \neq 0$, esta equação pode ser reescrita como:

$$xy + \frac{b^2v}{c^2}x - \frac{a^2u}{c^2}y = 0,$$

ou

$$xy + \frac{b^2v}{c^2}x - \frac{a^2u}{c^2}y - \frac{a^2b^2uv}{c^4} = -\frac{a^2b^2uv}{c^4},$$

e, fatorando a expressão acima, obtemos:

$$\left(x - \frac{a^2u}{c^2}\right) \left(y + \frac{b^2v}{c^2}\right) = -\frac{a^2b^2uv}{c^4}.$$

Se denotarmos $x_h = \frac{a^2u}{c^2}$ e $y_h = -\frac{b^2v}{c^2}$, esta equação pode ser reescrita como:

$$(x - x_h)(y - y_h) = x_h y_h,$$

a qual representa uma hipérbole equilátera com centro no ponto $H(x_h, y_h)$ e nível $x_h y_h$.

Notemos que

$$(0 - x_h)(0 - y_h) = x_h y_h,$$

e, portanto, a origem $O(0, 0)$ pertence a esta hipérbole. Por outro lado,

$$\left(u - \frac{a^2u}{c^2}\right) \left(v + \frac{b^2v}{c^2}\right) = \left(\frac{c^2u - a^2u}{c^2}\right) \left(\frac{c^2v + b^2v}{c^2}\right) = \left(-\frac{b^2u}{c^2}\right) \left(\frac{a^2v}{c^2}\right) = x_h y_h,$$

e, portanto, o ponto $P(u, v)$ também pertence à hipérbole.

Observemos que

$$\begin{aligned} 0 < u &\Rightarrow u < \frac{a^2u}{c^2} = x_h, \\ u < 0 &\Rightarrow x_h = \frac{a^2u}{c^2} < u, \end{aligned}$$

e também que

$$\begin{aligned} 0 < v &\Rightarrow y_h = -\frac{b^2v}{c^2} < 0, \\ v < 0 &\Rightarrow y_h = -\frac{b^2v}{c^2} > 0. \end{aligned}$$

Podemos, portanto, concluir que a origem O e o ponto P sempre pertencem ao mesmo ramo da hipérbole.

Dependendo da posição do ponto P poderemos ter a hipérbole intersectando a elipse em dois pontos, três ou no máximo quatro pontos. Estas diferentes situações estão diretamente relacionadas com a posição relativa do ponto P em relação à astróide da elipse. Para simplificar, vamos considerar que o ponto P esteja localizado no primeiro quadrante, Q_1 , sendo os outros quadrantes denominados Q_2 , Q_3 e Q_4 . Caso o ponto P esteja localizado em outro quadrante as adaptações necessárias são triviais.

1. Ponto P exterior à astróide (Figura B.2). Neste caso, somente um ramo da hipérbole intersecta a elipse nos pontos $P_1 \in Q_1$ e $P_2 \in Q_3$, pelos quais é possível traçar duas normais à elipse. que são tangentes à astróide nos quadrantes Q_2 e Q_4 . Os pontos P_1 e P_2 correspondem à mínima e a máxima distâncias ao ponto P .

2. Ponto P sobre a astróide (Figura B.3). Neste caso, um ramo da hipérbole intersecta a elipse nos pontos $P_1 \in Q_1$ e $P_2 \in Q_3$ enquanto o outro ramo da hipérbole tangencia a elipse no ponto $P_3 \in Q_4$. Por estes pontos é possível traçar duas normais à elipse que são tangentes à astróide nos quadrantes Q_2 e Q_4 , como no primeiro caso, e uma normal à elipse que é tangente à astróide no quadrante Q_1 . Os pontos P_1 e P_2 correspondem à mínima e máxima distância ao ponto P , enquanto o ponto P_3 corresponde a um ponto de inflexão em relação às distâncias calculadas de um ponto da elipse ao ponto P .

3. Ponto P interior à astróide (Figura B.4). Neste caso, um ramo da hipérbole intersecta a elipse nos pontos $P_1 \in Q_1$ e $P_2 \in Q_3$ enquanto o outro ramo da hipérbole intersecta a elipse nos pontos $P_3, P_4 \in Q_4$. Por estes pontos é possível traçar duas normais à elipse que são tangentes à astróide nos quadrantes Q_2 e Q_4 , como no primeiro caso, e duas normais à elipse que são tangente à astróide no quadrante Q_1 . Os pontos P_1 e P_2 correspondem à mínima e máxima distância ao ponto P , enquanto o ponto P_3 e P_4 correspondem a pontos de mínimo e máximo locais com relação às distâncias calculadas de um ponto da elipse ao ponto P .

A exposição acima descreve completamente o problema da determinação da distância à elipse em termos geométricos. Para encontrarmos as soluções algebricamente, teremos que resolver o sistema formado pelas equações (B.5) e (B.6).

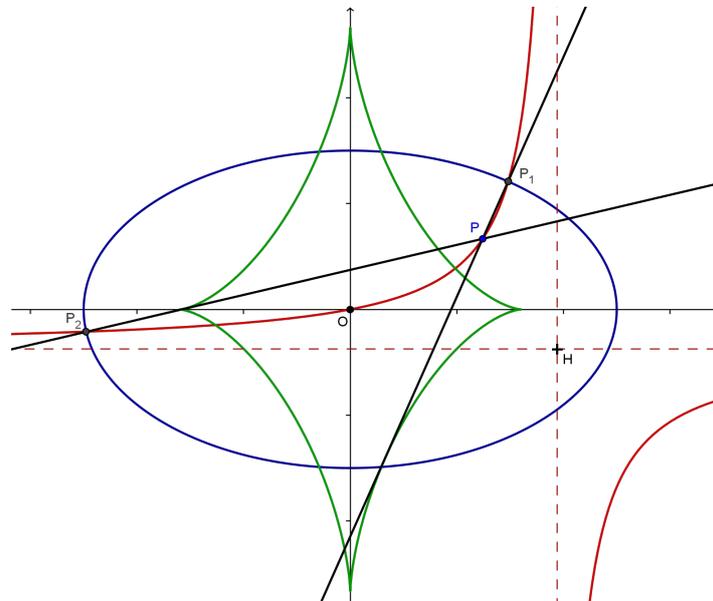


Figura B.2: Ponto P exterior à astróide e as duas normais à elipse passando pelos pontos P_1 e P_2 .

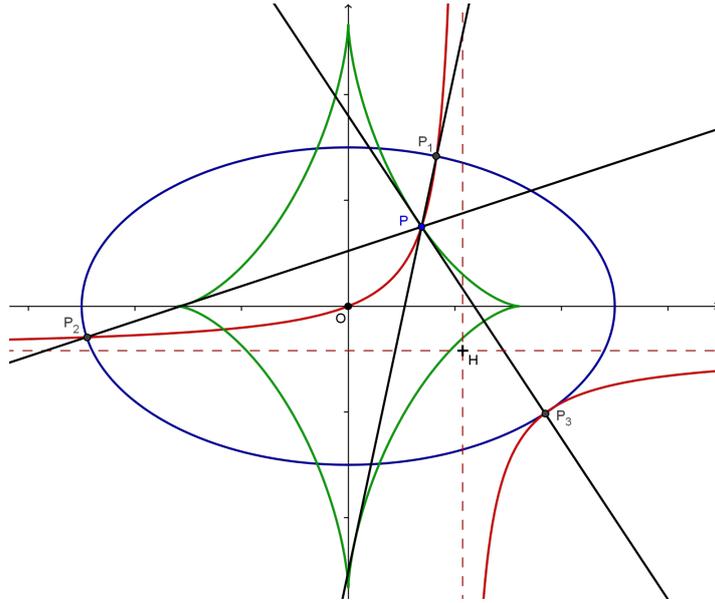


Figura B.3: Ponto P sobre a astróide e as três normais à elipse passando pelos pontos P_1 , P_2 e P_3 .

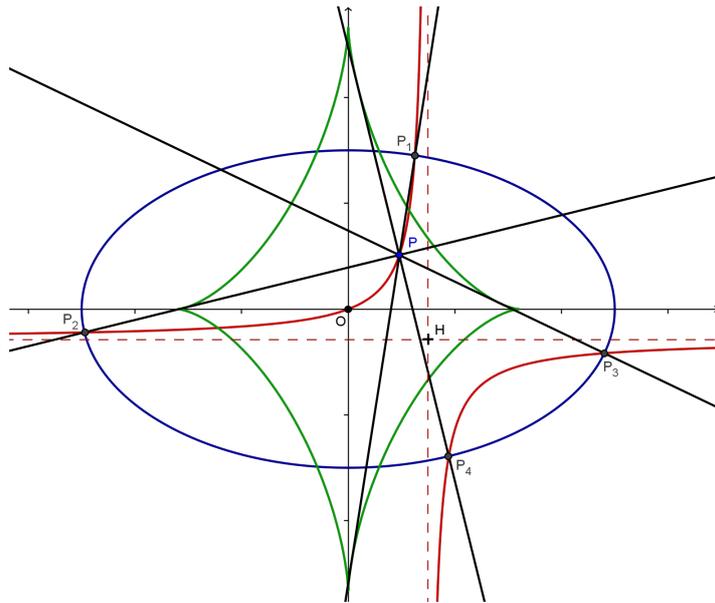


Figura B.4: Ponto P interior à astróide e as quatro normais à elipse passando pelos pontos P_1 , P_2 , P_3 e P_4 .

Uma reformulação do problema através de parametrização e substituição racional

Voltando ao sistema:

$$\begin{cases} b^2x^2 + a^2y^2 - a^2b^2 = 0 \\ c^2xy + b^2vx - a^2uy = 0 \end{cases},$$

$$b^2x^2 + a^2y^2 - a^2b^2 = 0 \quad (\text{B.7})$$

$$c^2xy + b^2vx - a^2uy = 0, \quad (\text{B.8})$$

utilizaremos, novamente, a parametrização para a elipse dada por

$$x(\theta) = a \cos \theta$$

$$y(\theta) = b \operatorname{sen} \theta$$

Esta parametrização transforma a equação (B.7) em uma identidade, o que nos permite descartá-la. Substituindo as variáveis x e y por suas parametrizações na equação (B.8) obtemos

$$c^2a \cos \theta b \operatorname{sen} \theta + b^2va \cos \theta - a^2ub \operatorname{sen} \theta = 0,$$

que pode ser simplificada, dividindo-a por $-ab$:

$$au \operatorname{sen} \theta - bv \cos \theta - c^2 \operatorname{sen} \theta \cos \theta = 0. \quad (\text{B.9})$$

A resolução da equação acima não é trivial. Para torná-la algebricamente mais tratável, substituiremos senos e cossenos por versões racionais utilizando o parâmetro $t = \operatorname{tg}(\theta/2)$, $t \in (-\infty, \infty)$, dada por

$$\operatorname{sen} \theta = \frac{2t}{1+t^2},$$

$$\cos \theta = \frac{1-t^2}{1+t^2}.$$

Devemos notar que o único valor de θ que não pode ser expresso por esta parametrização é o ângulo de π radianos, pois, neste caso, o valor de t não está definido. Esta limitação não será um problema para nosso desenvolvimento, pois normalmente estaremos transformando todos os ângulos de tal forma a localizá-los no primeiro quadrante.

Realizando a substituição das versões racionais de $\operatorname{sen} \theta$ e $\cos \theta$ na equação (B.9) obtemos

$$au \left(\frac{2t}{1+t^2} \right) - bv \left(\frac{1-t^2}{1+t^2} \right) - c^2 \left(\frac{2t}{1+t^2} \right) \left(\frac{1-t^2}{1+t^2} \right) = 0,$$

que resulta em

$$bvt^4 + 2a\left(u + \frac{c^2}{a}\right)t^3 + 2a\left(u - \frac{c^2}{a}\right)t - bv = 0. \quad (\text{B.10})$$

Considerando-se agora que o ponto $P(u, v)$ seja tal que $v \neq 0$ e $u \neq c^2/a$, vamos

definir

$$g_1 = \frac{\frac{bv}{a}}{u + \frac{c^2}{a}},$$
$$g_2 = \frac{\frac{bv}{a}}{u - \frac{c^2}{a}},$$

o que nos permite reescrever a equação (B.10) como

$$t^4 + \frac{2}{g_1}t^3 + \frac{2}{g_2}t - 1 = 0. \quad (\text{B.11})$$

Esta é uma equação algébrica de quarto grau em t . Apesar de não ser completa, ela é suficientemente complexa para não permitir simplificações triviais (como é o caso com as equações biquadráticas).

Não iremos tentar resolver a equação (B.11) pois o número de soluções reais depende do valor de seu discriminante e este fato dificulta a aplicação direta da fórmula de resolução de equações de quarto grau na formulação de problemas de programação matemática.

Um fato interessante, no entanto, é digno de ser comentado: os pontos $P(u, v)$ que anulam o discriminante da equação (B.11) são justamente os pontos que satisfazem a equação da evoluta da elipse. Isto faz total sentido pois cada ponto sobre a evoluta corresponde a duas tangentes idênticas sobrepostas, o que por sua vez corresponde a uma raiz real repetida e portanto, por definição, temos o discriminante igual a zero.

B.3 Aplicação do método de Newton na resolução do problema de determinar a distância de um ponto a uma elipse

Ao término da seção anterior concluímos que não era interessante buscarmos a resolução algébrica da equação (B.11). Uma forma alternativa de resolver esta equação é por aproximações sucessivas utilizando um método iterativo de rápida convergência. O método de Newton possui convergência quadrática quando estamos suficientemente próximos da raiz procurada. Nesta seção, iremos desenvolver a iteração de Newton para encontrar a raiz que nos interessa na equação (B.11).

Primeiramente, devemos notar que se o ponto $P(u, v)$ pertence ao primeiro quadrante a raiz que nos interessa satisfaz $t^* \in [0, 1]$. Como estaremos sempre providenciando uma realocação de $P(u, v)$ de tal forma que ele pertença ao primeiro quadrante, iremos tratar somente deste caso. Na seção seguinte, trataremos especi-

ficamente do problema de encontrar uma boa aproximação inicial t_0 para a raiz real que nos interessa da equação (B.11).

Se denotarmos por $P(t)$ o polinômio que corresponde ao lado esquerdo da equação (B.11) temos

$$P(t) = t^4 + \frac{2}{g_1}t^3 + \frac{2}{g_2}t - 1.$$

Por definição do método de Newton, temos a seguinte expressão para a iteração de Newton $N(t)$ utilizando $P(t)$:

$$t_{k+1} = N(t_k) = t_k - \frac{P(t_k)}{P'(t_k)}.$$

Podemos, portanto, definir a iteração de Newton $N(t)$ com base em $P(t)$ e sua derivada $P'(t)$:

$$N(t) = t - \frac{t^4 + \frac{2}{g_1}t^3 + \frac{2}{g_2}t - 1}{4t^3 + \frac{6}{g_1}t^2 + \frac{2}{g_2}},$$

que pode ser simplificado para

$$N(t) = \frac{3t^4 + \frac{4}{g_1}t^3 + 1}{4t^3 + \frac{6}{g_1}t^2 + \frac{2}{g_2}}. \quad (\text{B.12})$$

Esta é uma expressão racional que pode ser facilmente calculada, dados g_1 e g_2 , os quais dependem de $P(u, v)$.

Temos, assim, definida uma fórmula para a iteração de Newton bastante prática e facilmente implementável em linguagem de programação matemática. Note que se a aproximação inicial t_0 for suficientemente boa, não serão necessárias muitas iterações para se conseguir uma excelente aproximação para a raiz t^* .

Desenvolvimento de uma boa aproximação inicial t_0

Iremos agora considerar a geração de uma aproximação inicial t_0 que seja suficientemente próxima de t^* para ser aplicada à iteração definida por (B.12).

A ideia aqui desenvolvida parte de uma aproximação parabólica para o ramo da astróide que se localiza no quadrante Q_4 . A reta n , normal à elipse passando pelo ponto P , e que no problema original é tangente à astróide em Q_4 , será aproximada pela reta r , passando pelo ponto P e tangente à aproximação parabólica da astróide em Q_4 . Os pontos de interseção destas retas com a elipse serão denominados N e M , respectivamente, como apresentado na Figura B.5.

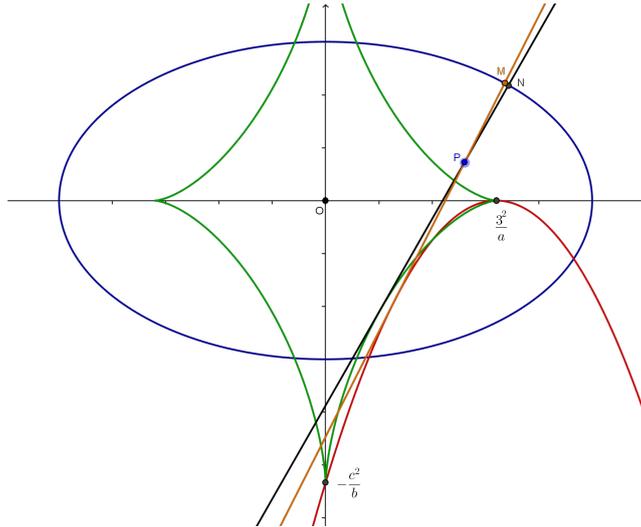


Figura B.5: Retas tangentes à astróide e à parábola passando pelo ponto P . As interseções com a elipse são os pontos N e M , respectivamente.

A determinação de t_0 será feita em três fases. Primeiro iremos determinar o coeficiente angular m da reta r . Com isto poderemos determinar $\sin \theta_0$ e $\cos \theta_0$, para $t_0 = \text{tg}(\theta_0/2)$. Finalmente, determinaremos t_0 com base em $\sin \theta_0$ e $\cos \theta_0$.

Determinação do coeficiente angular m da reta r

A aproximação parabólica utilizada irá passar pelos pontos extremos da astróide em Q_4 , que são os pontos $(0, -c^2/b)$ e $(c^2/a, 0)$, e terá seu vértice neste último ponto, onde a astróide possui tangente horizontal. Estas três condições são suficientes para definir completamente a parábola que buscamos. Iremos denominar esta parábola de $p(x)$. As duas últimas condições nos permitem afirmar que

$$p(x) = K \left(x - \frac{c^2}{a} \right)^2,$$

com K a ser determinado. Mas a primeira condição nos diz que

$$p(0) = -\frac{c^2}{b}.$$

Após substituição, K fica determinado e temos

$$p(x) = -\frac{(ax - c^2)^2}{bc^2},$$

e sua derivada

$$p'(x) = -\frac{2a}{bc^2}(ax - c^2).$$

Seja (x_1, y_1) o ponto de tangência da reta r com a parábola. Devemos ter,

portanto:

$$m = \frac{v - y_1}{u - x_1} = -\frac{2a}{bc^2}(ax_1 - c^2). \quad (\text{B.13})$$

Como $y_1 = p(x_1)$, a equação acima, após algumas manipulações algébricas, nos fornece

$$x_1 = u - \frac{\sqrt{(au - c^2)^2 + bc^2v}}{a},$$

que substituído em (B.13) finalmente resulta em

$$m = \frac{2a}{bc^2} \left(\sqrt{(au - c^2)^2 + bc^2v} - (au - c^2) \right). \quad (\text{B.14})$$

Podemos observar que, se o ponto P estiver sobre a parábola, então

$$v = -\frac{(au - c^2)^2}{bc^2}.$$

Substituindo este valor de v na equação (B.14) para m temos

$$m = -\frac{2a}{bc^2}(au - c^2) = p'(u),$$

que é a inclinação de r nesta situação, como seria de se esperar. Por outro lado, a reta n tangente à astróide quando P está na origem é vertical, e seu coeficiente de inclinação não é definido neste caso. Contudo, a expressão (B.14) nos fornece

$$m = \frac{4a}{b},$$

o que demonstra claramente o caráter aproximativo da parábola.

Determinação de $\sin \theta_0$ e $\cos \theta_0$

Com a inclinação m da reta r calculada, queremos agora determinar o ponto $M(x_2, y_2)$ de interseção desta reta com a elipse resolvendo o sistema

$$b^2x^2 + a^2y^2 - a^2b^2 = 0, \quad (\text{B.15})$$

$$y - v = m(x - u). \quad (\text{B.16})$$

A primeira equação é simplesmente a equação da elipse e a segunda equação é a equação da reta r . Substituindo y da segunda equação na primeira equação obtemos

$$b^2x^2 + a^2(m^2(x - u)^2 + 2mv(x - u) + v^2) - a^2b^2 = 0,$$

que pode ser simplificada para

$$(a^2m^2 + b^2)x^2 - 2a^2m(mu - v)x + a^2((mu - v)^2 - b^2) = 0.$$

Resolvendo-se esta equação quadrática e tomando a raiz que nos interessa, temos que

$$x_2 = a \left(\frac{am(mu - v) + b\sqrt{a^2m^2 + b^2 - (mu - v)^2}}{a^2m^2 + b^2} \right).$$

Utilizando (B.16) podemos calcular

$$y_2 = b \left(\frac{-b(mu - v) + am\sqrt{a^2m^2 + b^2 - (mu - v)^2}}{a^2m^2 + b^2} \right).$$

No entanto, por definição de θ_0 , temos que $x_2 = a \cos \theta_0$ e $y_2 = b \sin \theta_0$. Logo,

$$\cos \theta_0 = \frac{am(mu - v) + b\sqrt{a^2m^2 + b^2 - (mu - v)^2}}{a^2m^2 + b^2}, \quad (\text{B.17})$$

$$\sin \theta_0 = \frac{-b(mu - v) + am\sqrt{a^2m^2 + b^2 - (mu - v)^2}}{a^2m^2 + b^2}. \quad (\text{B.18})$$

Determinação de t_0

Para determinarmos t_0 iremos utilizar a identidade trigonométrica da tangente do arco metade

$$\text{tg} \left(\frac{\theta}{2} \right) = \frac{\text{sen } \theta}{1 + \cos \theta}.$$

Utilizando agora as equações B.17 e B.18 temos:

$$t_0 = \frac{-b(mu - v) + am\sqrt{a^2m^2 + b^2 - (mu - v)^2}}{a^2m^2 + b^2 + am(mu - v) + b\sqrt{a^2m^2 + b^2 - (mu - v)^2}}. \quad (\text{B.19})$$

Devemos notar que esta expressão depende dos parâmetros a e b da elipse, das coordenadas u e v do ponto P e do coeficiente angular m , que por sua vez depende de u e v . A substituição de m em função de u e v na expressão para t_0 iria torná-la bastante complexa. Na prática, esta substituição não se faz necessária pois a modelagem algébrica permite a definição de variáveis intermediárias, o que facilita a programação e aumenta a legibilidade do código.

B.4 Distância aproximada à elipse

Os desenvolvimentos anteriores forneceram as ferramentas necessárias para o cálculo aproximado da distância de um ponto P à elipse. Se considerarmos um valor t_k como

uma aproximação suficientemente boa para t^* , teremos

$$d_k(P, E) \approx \sqrt{\left(u - a \frac{1 - t_k^2}{1 + t_k^2}\right)^2 + \left(v - b \frac{2t_k}{1 + t_k^2}\right)^2},$$

onde $d_k(P, E)$ representa a distância do ponto P à elipse calculada utilizando a k -ésima iteração do método de Newton (B.12), começando com t_0 dado por (B.19).

Nesta seção estaremos calculando numericamente algumas distâncias e comparando-as com as distâncias exatas visando verificar a precisão das aproximações apresentadas. Iremos tomar como exemplo uma elipse com $a = 5$, $b = 3$, e conseqüentemente $c = 4$. Os pontos com $v = 0$ e $u > c^2/a$ não precisam ser considerados pois nestes casos as tangentes à astróide e à parábola coincidem e são horizontais, resultando em distâncias exatas. Alguns pontos arbitrários foram escolhidos e suas distâncias calculadas, resultando na Tabela B.1.

Ponto P		$k = 0$				$k = 1$	
		$d(P, E)$	$d_0(P, E)$	Erro		$d_1(P, E)$	Erro
				Absoluto	%		Absoluto
u	v						
0, 1	0, 1	2, 8991	2, 9162	+0, 0171	+0, 5904	2, 8994	+0, 0003
1, 0	0, 1	3, 1057	3, 1235	+0, 0179	+0, 6162	3, 1053	-0, 0004
2, 0	0, 1	3, 7281	3, 6550	-0, 0730	-2, 5206	3, 7235	-0, 0045
3, 0	0, 1	4, 5596	4, 6082	+0, 0486	+1, 6777	4, 5535	-0, 0061
0, 1	1, 0	2, 8991	2, 9056	+0, 0065	+0, 2229	2, 8993	+0, 0002
1, 0	1, 0	3, 0454	3, 0628	+0, 0174	+0, 6001	3, 0451	-0, 0004
2, 0	1, 0	3, 4931	3, 4665	-0, 0270	-0, 9194	3, 4926	-0, 0006
3, 0	1, 0	4, 0634	4, 0424	-0, 0210	-0, 7251	4, 0630	-0, 0004
4, 0	1, 0	4, 5246	4, 5288	+0, 0042	+0, 1434	4, 5246	-0, 0000
0, 1	2, 0	2, 8993	2, 9000	+0, 0007	+0, 0239	2, 8993	+0, 0000
1, 0	2, 0	3, 0041	3, 0133	+0, 0092	+0, 3186	3, 0040	-0, 0001
2, 0	2, 0	3, 3391	3, 3340	-0, 0050	-0, 1790	3, 3391	-0, 0000
3, 0	2, 0	3, 7829	3, 7764	-0, 0060	-0, 2240	3, 7829	-0, 0000

Tabela B.1: Tabela de comparação das distâncias exatas e aproximadas de um ponto P à elipse.

Podemos perceber que a aproximação inicial t_0 já é suficientemente boa, pois o maior desvio percentual calculado foi de apenas $-2,52\%$, sendo que a grande maioria dos pontos apresenta desvios inferiores a 1% . Para a aproximação t_1 os desvios percentuais são sempre bem inferiores a 1% e portanto não foram apresentados na tabela.

Apesar de não terem sido considerados nesta tabela, devido aos objetivos deste trabalho, pontos exteriores à elipse também apresentam bons resultados quanto às aproximações utilizadas.

A Figura B.6 apresenta o gráfico da elipse considerada acima, juntamente com

a parábola que é uma aproximação da astróide associada a esta elipse, e também as curvas exata e a aproximação por t_0 dos pontos que distam da elipse de uma unidade. É notória a excelente aproximação conseguida no primeiro quadrante.

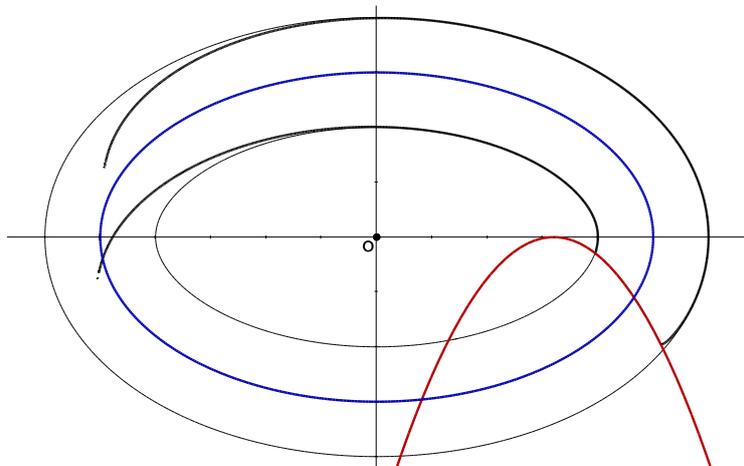


Figura B.6: Uma elipse, a aproximação parabólica da astróide no quarto quadrante, e as curvas exata (traço fino) e aproximada (traço grosso) dos pontos com distância à elipse igual a 1 unidade.

A curva de distâncias exatas utilizada foi apresentada por CAYLEY (1860). Se denominarmos

$$\begin{aligned} A(x, y) &= x^2 + y^2 - d^2 - a^2 - b^2, \\ B(x, y) &= b^2x^2 + a^2y^2 - a^2d^2 - b^2d^2 - a^2b^2, \end{aligned}$$

então a equação implícita

$$A(x, y)^2B(x, y)^2 + 4B(x, y)^3 + 4a^2b^2d^2A(x, y)^3 + 18a^2b^2d^2A(x, y)B(x, y) - 27a^4b^4d^4 = 0,$$

representa a curva dos pontos localizados à distância d da elipse com semi-eixos de tamanhos a e b .

Iremos agora comentar sobre o caso de P não pertencer ao primeiro quadrante. Dadas as simetrias da elipse, basta considerar o ponto equivalente do primeiro quadrante, obtido com as coordenadas iguais ao valor absoluto das coordenadas correspondentes do ponto original. Deve-se notar que a função valor absoluto pode apresentar problemas dependendo do ambiente computacional utilizado para resolver problemas de programação matemática. No entanto, os resolvedores não lineares costumam dar suporte a esta função. Quando não, podemos simplesmente utilizar

$$|x| = \sqrt{x^2},$$

o que costuma funcionar a contento.

B.5 Distância de um ponto a um elipsóide

Um elipsóide é uma superfície quádrlica que na forma canônica é dada por uma equação do tipo

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1,$$

onde a , b e c são números reais positivos.

No caso do elipsóide prolato o semi-eixo diferente é maior que os dois semi-eixos iguais. Vamos denominar a ao semi-eixo maior e b aos dois semi-eixos menores. Já no caso do elipsóide oblato temos o semi-eixo diferente menor que os dois semi-eixos iguais e denominaremos b ao semi-eixo menor e a aos dois semi-eixos maiores.

Neste trabalho, consideramos apenas os casos dos elipsóides prolato e oblato. Para estes elipsóides sempre temos dois semi-eixos iguais e um semi-eixo diferente.

Procedendo desta forma, sempre teremos $a > b$ e a teoria das distâncias desenvolvida para as elipses no espaço \mathbb{R}^2 pode ser facilmente aplicada para os elipsóides pertencentes a essas duas classes no espaço \mathbb{R}^3 . Consideremos, portanto, o espaço \mathbb{R}^3 , com os eixos x , y e z , e um elipsóide centrado na origem com os dois semi-eixos iguais dispostos nos eixos y e z e o semi-eixo diferente disposto no eixo x . A Figura B.7a ilustra um elipsóide prolato e a Figura B.7b, um elipsóide oblato.

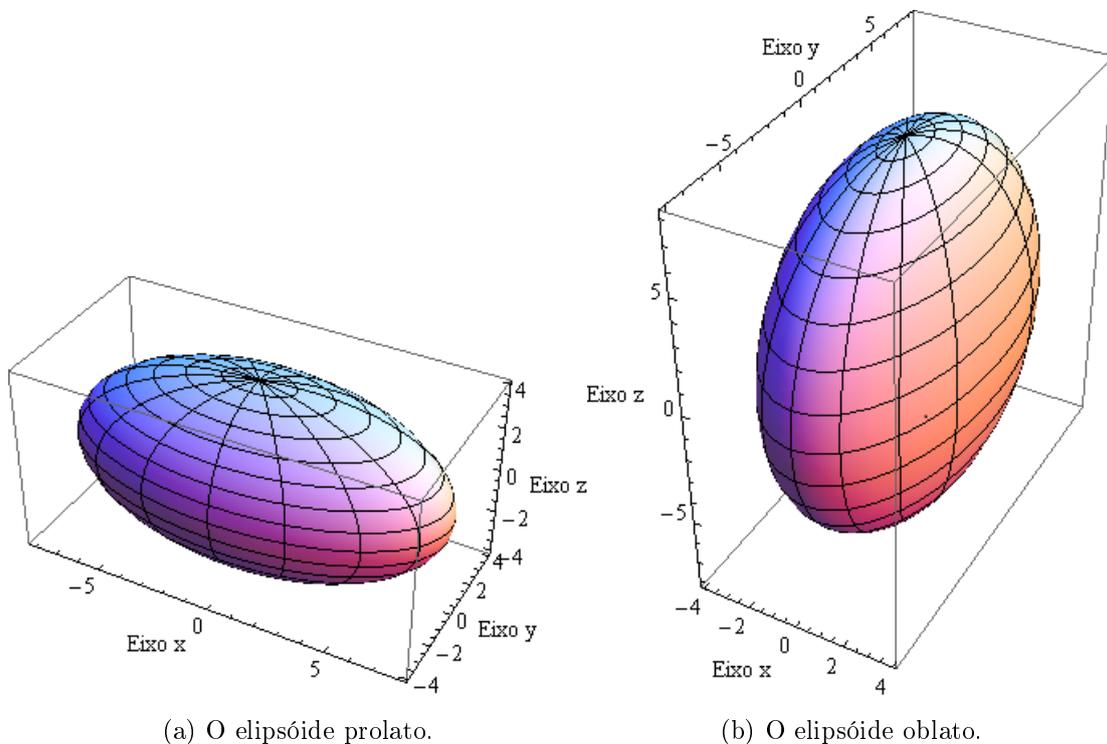


Figura B.7: Elipsóides de revolução.

Seja $R(x, y, z)$ um ponto dentro de um elipsóide prolato ou oblato para o qual

desejamos calcular sua distância ao elipsóide. Facilmente percebemos que a seção produzida pelo plano π que contém o eixo x e o ponto R será uma elipse com semi-eixos a e b .

No caso do elipsóide prolato, podemos tratar o plano π como sendo o plano \mathbb{R}^2 com uma origem O' e um eixo x' que coincidem com a origem O e o eixo x originais do \mathbb{R}^3 , e com um eixo y' ortogonal ao eixo x' passando por O' . Assim, teremos a elipse resultante da seção do elipsóide centrada na origem O' e com os semi-eixos a localizado no eixo x' e b localizado no eixo y' . Seja P o ponto correspondente ao ponto R neste novo sistema de coordenadas. Suas coordenadas podem ser trivialmente calculadas com base nas coordenadas de R como

$$P(u, v) = \left(x, \sqrt{y^2 + z^2} \right).$$

Já no caso do elipsóide oblato, também podemos tratar o plano π como sendo o plano \mathbb{R}^2 com uma origem O' e um eixo y' que coincidem com a origem O e o eixo x originais do \mathbb{R}^3 , e com um eixo x' ortogonal ao eixo y' passando por O' . Assim, mais uma vez teremos a elipse resultante da seção do elipsóide centrada na origem O' e com os semi-eixos a localizado no eixo x' e b localizado no eixo y' . Seja P o ponto correspondente ao ponto R neste novo sistema de coordenadas. Suas coordenadas podem ser calculadas com base nas coordenadas de R como

$$P(u, v) = \left(\sqrt{y^2 + z^2}, x \right).$$

A exposição acima demonstra que o cálculo da distância de um ponto R a um elipsóide, para os casos dos elipsóides prolato e oblato, pode ser reduzido trivialmente ao caso do cálculo da distância de um ponto à uma elipse mediante uma reinterpretação das coordenadas originais do ponto R . Estes resultados bastam para os propósitos almejados por esta tese.

Apêndice C

Índices de cobertura e o *Gamma Knife*

O sucesso das radiocirurgias utilizando o *Gamma Knife* depende da precisão com que a dose de radiação é efetivamente aplicada ao tumor alvo, em contraste com a quantidade de radiação aplicada ao tecido sadio nas proximidades do tumor. Visando medir esta efetividade, vários índices que descrevem a conformidade da dose com o tumor alvo foram propostos na literatura (FEUVRET *et al.*, 2006; WU *et al.*, 2003).

O índice de conformidade mais conhecido e amplamente utilizado em centros de radiocirurgia é o descrito no "Radiation Therapy Oncology Group (RTOG) guidelines for radiosurgery (diretrizes para radiocirurgia do grupo de tratamento oncológico por radiação)" proposto por SHAW (1993). Ele é definido como

$$I_S = \frac{V_P}{V_T},$$

onde V_P é o volume da região em que a dose de radiação é aplicada e V_T é o volume do alvo. Este índice apresenta um valor menor que 1 para tratamentos abaixo do necessário, igual a 1 quando existe plena conformidade, e maior que 1 quando o tratamento supera a necessidade.

PADDICK (2000) menciona críticas ao índice C_S por não levar em consideração a posição relativa entre os volumes V_P e V_T , fato que ele considera uma "falha fundamental", e propõe um novo índice

$$I_P = \frac{(V_T \cap V_P)^2}{V_T V_P} = \frac{V_T \cap V_P}{V_T} \frac{V_T \cap V_P}{V_P} = \frac{V_X}{V_T} \frac{V_X}{V_P} = TC \times PS,$$

onde $V_X = V_T \cap V_P$, o fator $TC = V_X/V_T$ é chamado de "fator de cobertura do alvo (*target coverage*)" e o fator $PS = V_X/V_P$ é a "seletividade do plano (*plan selectivity*)". Observemos que o volume V_X corresponde ao volume do alvo que é

irradiado e, portanto, o índice I_P leva em consideração a posição relativa entre os volumes V_P e V_T . Por outro lado, sempre teremos

$$TC = V_X/V_T \leq 1,$$

e

$$PS = V_X/V_P \leq 1.$$

Isto significa que $I_P \leq 1$ e toma-se o valor unitário como representativo de plena conformidade. Valores menores do que 1 significam ou que o alvo não foi totalmente coberto, ou que tecido sadio foi incluído na cobertura proposta, ou um efeito combinado destas duas possibilidades.

O índice de Paddick já está sendo considerado seriamente no planejamento de radiocirurgias com o *Gamma Knife*. Por exemplo, este índice foi incorporado à versão 10.0 do Leksell Gamma Plan treatment planning software (LGP 10.0) da Elekta Instrument A/B.

Conforme mencionado acima, outras propostas de índices existem, mas, dadas as suas virtudes, o índice de Paddick foi escolhido nesta tese como base para a seleção das melhores coberturas de um determinado alvo.

Apêndice D

As técnicas de integração de Monte Carlo e quasi-Monte Carlo

O cálculo de volumes no espaço \mathbb{R}^3 pode muitas vezes ser analiticamente difícil. Existe no entanto uma alternativa de computação numérica destes volumes chamada *Integração de Monte Carlo* que se baseia na utilização de números randômicos. Os métodos de integração de quasi-Monte Carlo utilizam os mesmos princípios, mas as invés de números randômicos, utilizam números quasi-randômicos gerados por sequências de baixa discrepância.

Apesar destes métodos poderem ser aplicados a integrações em espaços \mathbb{R}^n de dimensões maiores do que três, iremos focar somente em integrações no \mathbb{R}^3 , o que basta para nossos propósitos.

A ideia básica destes métodos é incluir o volume a ser calculado dentro de um paralelepípedo, cujo volume possa ser facilmente calculado. É, então, gerado um conjunto relativamente grande de pontos aleatórios dentro do paralelepípedo e parte-se do princípio que estes pontos estejam distribuídos de forma relativamente uniforme. Podemos, então, afirmar que

$$V_T \approx \frac{N_T}{N_R} V_R,$$

onde V_T é o volume desejado, V_R é o volume do paralelepípedo, N_R é o número total de pontos aleatórios gerados e N_T é o número de pontos localizados dentro do volume V_T . Esta definição é bastante intuitiva e nos diz que a proporção do volume V_T dentro de V_R é aproximadamente igual à proporção do número de pontos N_T que estão localizados dentro de V_T comparado com o número total de pontos N_R que estão localizados dentro de V_R .

A aproximação acima é tanto melhor quanto mais uniforme for a distribuição dos pontos aleatórios. Na integração de Monte Carlo tradicional são utilizados pontos oriundos de um gerador de números pseudo-aleatórios. Esta distribuição de pontos somente atinge um nível satisfatório de uniformidade dentro dos volumes a serem

calculados quando o número de pontos é muito grande.

Uma alternativa para os números pseudo-aleatórios na geração de pontos aleatórios é a utilização de sequências de baixa discrepância ("low discrepancy sequences" ou LDS). Estas sequências são, na verdade, determinísticas, mas projetadas para apresentar uma distribuição bastante uniforme dentro do intervalo $[0, 1]$.

Um interessante estudo comparativo entre as técnicas de integração de Monte Carlo e quasi-Monte Carlo é apresentada por MOROKOFF e CAFLISCH (1995). Neste estudo, as opções à integração de Monte Carlo tradicional são os métodos de integração de quasi-Monte Carlo que utilizam as sequências LDS de Halton, Sobol e Faure. A conclusão deste estudo é que para dimensões até 6 a sequência mais apropriada é a de Halton. Este é um resultado relevante no contexto deste trabalho pois o algoritmo para a geração de uma sequência de Halton tridimensional é bastante simples, exigindo pouco esforço computacional, diferentemente de sequências mais sofisticadas tais como as de Sobol e Faure.

D.1 Sequências de Halton

As sequências de Halton são generalizações para dimensões maiores das sequências unidimensionais de van der Corput. Estas, por sua vez, são construídas pela reversão, com relação à vírgula, da representação na base n da sequência dos números naturais. Por exemplo, se tomarmos $n = 10$, a sequência decimal de van der Corput é dada por

$$\{0,1; 0,2; \dots 0,9; 0,01; 0,11; 0,21; \dots 0,91; 0,02; 0,12; 0,22; 0,32; \dots\},$$

enquanto que a sequência binária de van der Corput é dada por

$$\{0,1_2; 0,01_2; 0,11_2; 0,001_2; 0,101_2; 0,011_2; 0,111_2; 0,0001_2; 0,1001_2; 0,0101_2; \dots\},$$

que também pode ser representada como

$$\left\{ \frac{1}{2}; \frac{1}{4}; \frac{3}{4}; \frac{1}{8}; \frac{5}{8}; \frac{3}{8}; \frac{7}{8}; \frac{1}{16}; \frac{9}{16}; \frac{5}{16}; \frac{13}{16}; \frac{3}{16}; \frac{11}{16}; \frac{7}{16}; \frac{15}{16}; \dots \right\}.$$

Da forma como são definidas, as sequências de van der Corput se distribuem uniformemente dentro do intervalo unitário $[0, 1]$. Adicionalmente, uma sequência de van der Corput possui a importante propriedade de ser densa neste intervalo, ou seja, para todo número real x_0 pertencente ao intervalo $[0, 1]$ existe uma subsequência que converge para x_0 .

Nas sequências de Halton são utilizados números primos distintos como base

N	Dim1	Dim2	Dim3
1	0,5000	0,33333333	0,20
2	0,2500	0,66666667	0,40
3	0,7500	0,11111111	0,60
4	0,1250	0,44444444	0,80
5	0,6250	0,77777778	0,04
6	0,3750	0,22222222	0,24
7	0,8750	0,55555556	0,44
8	0,0625	0,88888889	0,64
9	0,5625	0,03703704	0,84
10	0,3125	0,37037037	0,08

Tabela D.1: Os primeiros 10 termos de uma sequência de Halton tridimensional utilizando as bases 2, 3 e 5.

para cada uma das dimensões consideradas. Isto reduz substancialmente a correlação entre as sequências geradas para cada dimensão. Em dimensões mais elevadas, a correlação entre as sequências utilizadas para cada dimensão torna-se um problema sério e as sequências de Sobol tornam-se mais indicadas (MOROKOFF e CAFLISCH, 1995).

Para o caso tridimensional, é natural utilizar os três primeiros primos como base para as sequências em cada dimensão. Os primeiros termos da sequência de Halton tridimensional são apresentados na Tabela D.1.

Apêndice E

Exemplos extraídos do USCS

Apresentamos a seguir a tela de interfaceamento com o USCS e alguns segmentos de texto de sua listagem de informações.

E.1 Tela de interfaceamento com o USCS

A Figura E.1 apresenta a tela de interfaceamento com o USCS.

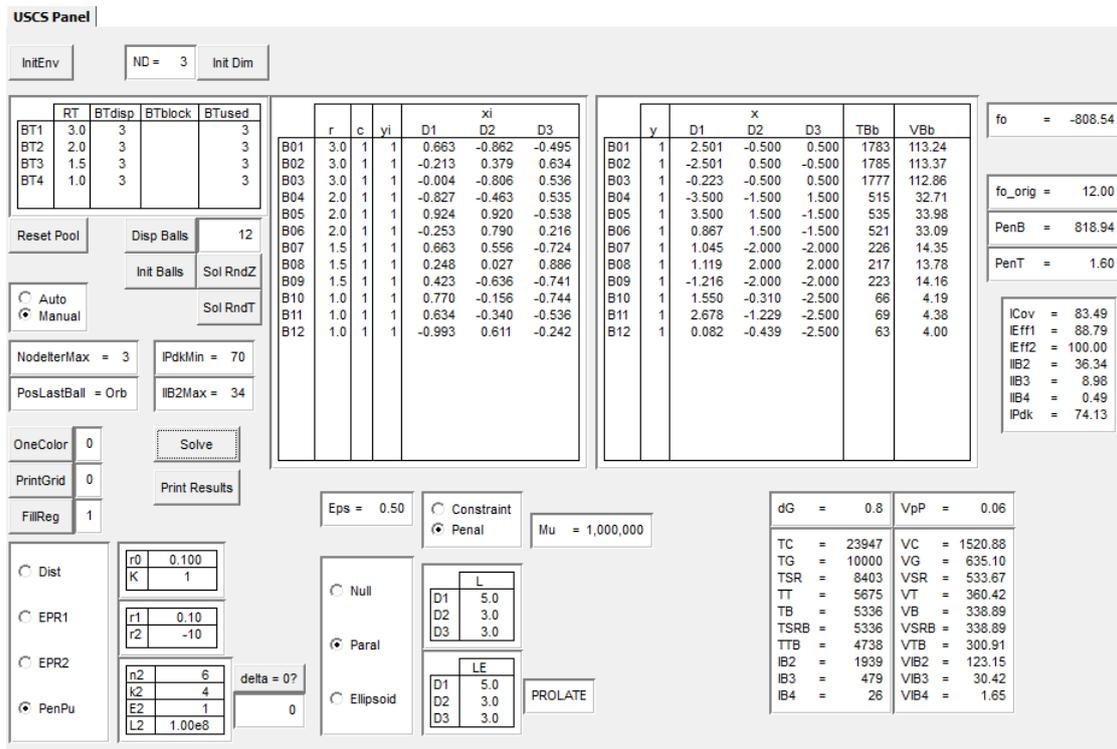


Figura E.1: Tela de interface com o USCS.

E.2 Sumário das estatísticas e índices

```

=====
Tumor: Ellipsoid prolate,   a: 15.00   b:  9.00   c:  9.00
=====
Grid points      Volumes:  Aprox.    Exact   Error%  Add. Info.
-----
TC   =  44638   VC   =  41063.63  41063.63   0.00%  VpP = 0.9199
TG   =  10000   VG   =   9199.25
TSR  =   8332   VSR  =   7664.82
TT   =   5527   VT   =   5084.43  5089.38  -0.10%
TB   =   7084   VB   =   6516.75
TSRB =   7084   VSRB =   6516.75
TTB  =   5447   VTB  =   5010.83

TB( 1)=  3337   VB( 1)=  3069.79   3053.63   0.53%  R( 1)= 9.00
TB( 2)=  3317   VB( 2)=  3051.39   3053.63  -0.07%  R( 2)= 9.00
TB( 3)=  1563   VB( 3)=  1437.84   1436.76   0.08%  R( 3)= 7.00
TB( 4)=  1565   VB( 4)=  1439.68   1436.76   0.20%  R( 4)= 7.00
TB( 5)=   297   VB( 5)=   273.22    268.08   1.92%  R( 5)= 4.00
TB( 6)=   283   VB( 6)=   260.34    268.08  -2.89%  R( 6)= 4.00

IB2  =   2222   VIB2 =   2044.07
IB3  =    706   VIB3 =    649.47
IB4  =    293   VIB4 =    269.54

Covering Index =  98.55%
Eff. Index1    =  76.89%
Eff. Index2    = 100.00%
Inters. Index2 =  31.37%
Inters. Index3 =   9.97%
Inters. Index4 =   4.14%
Paddick Index  =  75.78%
=====

```

E.3 Histórico cronológico dos eventos de processamento

```

=====
NTE  Description          Elapsed Time  %Total Time
---  -
  1  Grid/T creation       1.74s         18.35 %
  2  SR/Tum cnt/vol       0.11s         1.16 %
  3  Solver                0.06s         0.63 %
  4  Balls cnt/vol        0.11s         1.16 %
  5  Save 1st Iter        0.02s         0.21 %
  6  Solver                0.00s         0.00 %
  7  Balls cnt/vol        0.09s         0.94 %
  8  Solver                0.02s         0.21 %
  9  Balls cnt/vol        0.08s         0.84 %
 10  Recov Best Iter      0.01s         0.10 %
 11  Low IPdk node        0.00s         0.00 %
 12  Nxt nod forward      0.00s         0.00 %
 13  Solver                0.02s         0.21 %
 14  Balls cnt/vol        0.12s         1.26 %
 15  Save 1st Iter        0.02s         0.21 %
 16  Solver                0.01s         0.10 %
 17  Balls cnt/vol        0.13s         1.37 %
 18  Sav Better Iter      0.01s         0.10 %
 19  Solver                0.00s         0.00 %
 20  Balls cnt/vol        0.14s         1.47 %
 21  Sav Better Iter      0.00s         0.00 %
 22  Recov Best Iter      0.00s         0.00 %
 23  Low IPdk node        0.00s         0.00 %
 24  Nxt nod forward      0.00s         0.00 %
 25  Solver                0.02s         0.21 %
 26  Balls cnt/vol        0.17s         1.79 %
 27  Save 1st Iter        0.02s         0.21 %
 28  Solver                0.00s         0.00 %
 29  Balls cnt/vol        0.18s         1.89 %
 30  Sav Better Iter      0.02s         0.21 %
 31  Solver                0.02s         0.21 %
 32  Balls cnt/vol        0.15s         1.58 %
 33  Recov Best Iter      0.02s         0.21 %

```

34	Valid node	0.00s	0.00 %
35	Save 1st incumb	0.00s	0.00 %
36	Nxt nod forward	0.00s	0.00 %
37	Solver	0.01s	0.10 %
38	Balls cnt/vol	0.21s	2.21 %
39	Save 1st Iter	0.01s	0.10 %
40	Solver	0.02s	0.21 %
41	Balls cnt/vol	0.20s	2.10 %
42	Solver	0.03s	0.31 %
43	Balls cnt/vol	0.23s	2.42 %
44	Recov Best Iter	0.00s	0.00 %
45	High IB2 node	0.00s	0.00 %
46	BckTrk, Nxt nod	0.00s	0.00 %
47	Solver	0.04s	0.42 %
48	Balls cnt/vol	0.20s	2.10 %
49	Save 1st Iter	0.01s	0.10 %
50	Solver	0.02s	0.21 %
51	Balls cnt/vol	0.22s	2.32 %
52	Solver	0.03s	0.31 %
53	Balls cnt/vol	0.20s	2.10 %
54	Recov Best Iter	0.02s	0.21 %
55	High IB2 node	0.00s	0.00 %
56	BckTrk, Nxt nod	0.00s	0.00 %
57	Solver	0.08s	0.84 %
58	Balls cnt/vol	0.20s	2.10 %
59	Save 1st Iter	0.01s	0.10 %
60	Solver	0.08s	0.84 %
61	Balls cnt/vol	0.20s	2.10 %
62	Sav Better Iter	0.02s	0.21 %
63	Solver	0.06s	0.63 %
64	Balls cnt/vol	0.22s	2.32 %
65	Recov Best Iter	0.00s	0.00 %
66	Valid node	0.00s	0.00 %
67	Save better inc	0.00s	0.00 %
68	Nxt nod forward	0.00s	0.00 %
69	Solver	0.13s	1.37 %
70	Balls cnt/vol	0.23s	2.42 %
71	Save 1st Iter	0.02s	0.21 %
72	Solver	0.09s	0.94 %

73	Balls cnt/vol	0.23s	2.42 %
74	Solver	0.13s	1.37 %
75	Balls cnt/vol	0.23s	2.42 %
76	Recov Best Iter	0.00s	0.00 %
77	High IB2 node	0.00s	0.00 %
78	BckTrk, Nxt nod	0.00s	0.00 %
79	Solver	0.10s	1.05 %
80	Balls cnt/vol	0.23s	2.42 %
81	Save 1st Iter	0.02s	0.21 %
82	Solver	0.10s	1.05 %
83	Balls cnt/vol	0.24s	2.53 %
84	Solver	0.09s	0.94 %
85	Balls cnt/vol	0.25s	2.63 %
86	Recov Best Iter	0.00s	0.00 %
87	Valid node	0.00s	0.00 %
88	Keep incumbent	0.00s	0.00 %
89	Nxt nod forward	0.00s	0.00 %
90	Solver	0.10s	1.05 %
91	Balls cnt/vol	0.26s	2.74 %
92	Save 1st Iter	0.02s	0.21 %
93	Solver	0.50s	5.27 %
94	Balls cnt/vol	0.26s	2.74 %
95	Solver	0.16s	1.68 %
96	Balls cnt/vol	0.26s	2.74 %
97	Recov Best Iter	0.02s	0.21 %
98	High IB2 node	0.00s	0.00 %
99	BckTrk, Nxt nod	0.00s	0.00 %
100	End of search	0.00s	0.00 %
101	Balls cnt/vol	0.20s	2.10 %
102	Recov Incumbent	0.00s	0.00 %
103	End Computation	0.00s	0.00 %

Total Elapsed time: 9.48s

=====