

**Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Química**

Rômulo Verdolin de Sousa

**Problemas Diretos e Inversos em
Química e Física Médica**

Belo Horizonte
2009

UFMG 785
T. 329

Rômulo Verdolin de Sousa

Problemas Diretos e Inversos em Química e Física Médica

Tese de Doutorado apresentada ao Departamento de Química do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de **Doutor em Ciências - Química**.

Belo Horizonte
2009

Sousa, Rômulo Verdolin

S725p Problemas diretos e inversos em química e física
2009 médica./ Rômulo Verdolin Sousa. 2009.

xiii; 158 f. : il.

Orientador: João Pedro Braga

Co-Orientadora: Rita de Cássia de Oliveira
Sebastião.

Tese (Doutorado) - Universidade Federal de Minas
Gerais. Departamento de Química.

Bibliografia p.153-158

1.Físico-Química - Teses 2.Física médica - Teses
3.Problemas inversos- Teses I.Braga, João Pedro,
Orientador II.Sebastião, Rita de Cássia de Oliveira,
Co-orientador Título

CDU 043



PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA
DEPARTAMENTO DE QUÍMICA - ICEX
31270-901 - BELO HORIZONTE - MG
TEL.: 031 - 3409-5732
FAX: 031 - 3409-5711
E-MAIL: pgquimic@qui.ufmg.br



"Problemas Diretos e Inversos em Química e Física Médica"

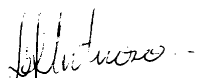
Rômulo Verdolin de Sousa

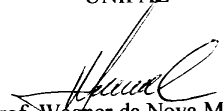
Tese aprovada pela banca examinadora constituída pelos Professores:


Prof. João Pedro Braga - Orientador
UFMG


Profª. Rita de Cássia de Oliveira Sebastião - Co-Orientadora
UFMG


Prof. Edilton de Souza Barcellos
UFV


Prof. Luciano Sindra Virtuoso
UNIFAL


Prof. Wagner da Nova Mussel
UFMG


Prof. Heitor Avelino de Abreu
UFMG

Belo Horizonte, 09 de outubro de 2009.

AGRADECIMENTOS

É com imenso prazer que expresso os meus mais sinceros agradecimentos a todos que, de alguma forma, contribuíram para que este trabalho se tornasse uma realidade:

- ◆ A Deus, que está acima de todas as coisas, e que me permitiu vencer mais uma importante etapa de minha vida;
- ◆ Ao meu professor/orientador Dr. João Pedro Braga, pela seriedade, paciência e profissionalismo com que me orientou, auxiliando-me no desenvolvimento deste trabalho;
- ◆ À minha co-orientadora Dra. Rita de Cássia de Oliveira Sebastião, por toda a ajuda prestada durante o desenvolvimento deste trabalho;
- ◆ Aos meus pais e meu irmão, por todo o carinho e amor que me deram durante toda a vida, o que me tornou uma pessoa responsável e comprometida com tudo o que faço;
- ◆ À minha maravilhosa filha Laura de Sousa Ribeiro, por todo o amor que sentimos um pelo outro, sentimento tão intenso que me dá forças para continuar lutando na vida, sempre buscando um futuro melhor para ela;
- ◆ À Elisângela Francisca da Silva, por todo carinho e paciência, e pelo amor que sentimos mutuamente;
- ◆ A todos os membros da banca examinadora, em especial ao Dr. José Domingos Fabris, por todas as sugestões que engrandeceram o conteúdo deste trabalho;
- ◆ A todo o colegiado da Química, aos professores e aos demais colaboradores do Departamento de Química da UFMG que contribuíram para que o trabalho fosse desenvolvido.

*A meu pai Rubens de Souza.
Saudades...*

“No human investigation can be called real science if it cannot be demonstrated mathematically”.

Leonardo da Vinci (1452-1519)

RESUMO

Quando tratamos de qualquer assunto relacionado a Ciências, sempre nos deparamos com situações práticas onde experimentos devem ser observados e/ou com situações teóricas onde modelamentos matemáticos devem ser desenvolvidos. Ambas as situações tornam-se rotineiras em físico-química, como no caso da deposição de dose de radiação ionizante em tecidos biológicos, assunto de grande interesse dentro da Física Médica, em especial na Física da Radioterapia. Diversas situações surgem nesse contexto, algumas sendo resolvidas por técnicas de problemas diretos e outras por técnicas de problemas inversos, sempre envolvendo ferramentas matemáticas avançadas, como métodos numéricos, cálculo integral e diferencial, regularizações e uso de redes neurais artificiais. A deposição de dose de radiação em profundidade é de grande interesse e pode ser obtida experimentalmente ou ainda através de simulações com códigos numéricos, como o MCNP. Neste trabalho desenvolve-se forma inédita de obtenção de valores de PDP através de técnicas de problemas inversos, utilizando-se a Regularização de Tikhonov. Mostra-se que esta técnica matemática sofisticada fornece excelentes resultados na resolução deste problema, podendo ser empregada futuramente em programas de planejamento, possibilitando a redução do tempo de processamento necessário. No último capítulo fala-se, de maneira introdutória, da teoria do espalhamento quântico, que futuramente poderá ser anexada ao estudo físico das técnicas de radioterapia. Com esse trabalho inicia-se uma nova linha de pesquisa dentro do Departamento de Química da UFMG, empregando-se métodos numéricos e técnicas de problemas inversos dentro da Física Médica.

ABSTRACT

When we dealt with any matter related to Sciences, we always have practical situations where experiments should be observed and/or with theoretical situations where mathematical models should be developed. Both situations become routine in physical chemistry, as in the case of the deep dose deposition of ionizing radiation in biological tissues, matter of great interest inside of the Medical Physics, especially in the Physics of the Radiotherapy. Several situations appear in that context, some being resolved with techniques of direct problems and other being resolved with techniques of inverse problems, always involving advanced mathematical tools, as numeric methods, integral and differential calculation, regularizations and use of neural artificial networks. The deep dose deposition of ionizing radiation is of great interest and it can be obtained experimentally or still through simulations with numeric codes, like MCNP. In this work we develop a new way for obtaining values of PDD through techniques of inverse problems, using Tikhonov Regularization. It is shown that this sophisticated mathematical technique supplies excellent results in the resolution of this problem, and it could be used in the future in planning programs, making possible the reduction of the time of necessary processing. In the last chapter we show, in an introductory way, the theory of the quantum scatter, and it can be used in the future in the physical study of the radiotherapy techniques. This work starts a new research line inside the Chemistry Department of UFMG, using numeric methods and techniques of inverse problems inside Medical Physics.

SUMÁRIO

Lista de Figuras	IX
Lista de Tabelas	X
Lista de Abreviaturas, Siglas e Símbolos	XI
1 – Introdução Geral	1
1.1 – O que são Problemas Inversos.....	2
1.2 – Resolução de Problemas Inversos	4
1.3 – Problemas Mal-Colocados	5
1.4 – Problemas Inversos de Primeira Ordem.....	5
1.5 – Problemas Inversos de Segunda Ordem.....	6
1.6 – Por que ensinar Problemas Inversos	6
1.7 – Problemas Inversos em Física Computacional	6
1.8 – Exemplo de um Problema Inverso	7
1.9 – Conclusão	11
2 – Equação Integral de Fredholm de Primeira Ordem	12
2.1 – A Equação	13
2.2 – Representação da Equação de Fredholm	13
2.3 – Representação em uma Base	15
2.4 – Os Quatro Subespaços Fundamentais, Rank e Índice de Condicionamento	18
2.5 – Exemplo do Uso de Programas Computacionais.....	22
2.6 – Conclusão	24
3 – Método dos Mínimos Quadrados	25
3.1 – Introdução	26
3.2 – Ortogonalidade	27
3.3 – Definição Matemática dos Mínimos Quadrados.....	27
3.4 – Uso do Método dos Mínimos Quadrados	28
3.5 – O Problema Linear dos Mínimos Quadrados.....	31

3.6 – Exemplo Matemático de Ajuste a uma Reta	32
3.7 – Qualidade do Ajuste	34
3.8 – Conclusão	36
4 – A Regularização de Tikhonov	37
4.1 – Introdução	38
4.2 – A Primeira Restrição.....	39
4.3 – A Segunda Restrição	40
4.4 – A Terceira Restrição	41
4.5 – Exemplo de Aplicação da Regularização de Tikhonov	43
4.6 – Conclusão	47
5 – O Método de Monte Carlo	48
5.1 – Introdução	49
5.2 – Descrição do Método de Monte Carlo	51
5.3 – Exemplos do Uso do Método de Monte Carlo.....	52
5.4 – O Algoritmo de Metropolis.....	63
5.5 – Conclusão	71
6 – O Código MCNP – Monte Carlo N-Particle Code.....	72
6.1 – Introdução	73
6.2 – O Método de Monte Carlo versus o Método Determinístico	74
6.3 – Características	76
6.4 – Geometria	78
6.5 – História da Partícula	81
6.6 – Seções de Choque.....	82
6.7 – Física do MCNP	83
6.8 – Cartões de Saída	91
6.9 – Exatidão e Precisão.....	92
6.10 – Erro Estimado Relativo	95
6.11 – Redução de Variância	96
6.12 – Exemplos de Uso do MCNP.....	97
6.13 – Conclusão	108

7 – Resolução de Problemas em Radioterapia	109
7.1 – Introdução	110
7.2 – Resolução com Equações Integrais	111
7.3 – Modelos de Kernel Baseados em Convolução/Superposição.....	120
7.4 – Funções de Dose	123
7.5 – Algoritmos Analíticos	128
7.6 – Conclusão	137
8 – Teoria de Espalhamento Quântico.....	139
8.1 – Introdução	140
8.2 – Formato Assintótico das Ondas	143
8.3 – Solução Utilizando Equações Integrais	143
8.4 – A Aproximação de Born.....	145
8.5 – Conclusão	148
9 – Conclusões	149
10 – Referências Bibliográficas	152

LISTA DE FIGURAS

FIGURA 1.1 –	Forma direta de processamento	02
FIGURA 1.2 –	Representação gráfica do movimento do projétil	08
FIGURA 2.1 –	Área sob a curva $y = f(x)$ dividida em n faixas, onde $h = (b-a)/n$	15
FIGURA 2.2 –	Valores exatos da matriz “ f ” gerada computacionalmente	23
FIGURA 2.3 –	Valores da matriz “ f ” obtida computacionalmente através da equação $f = K^{-1}g$	24
FIGURA 3.1 –	Valores da matriz “ f ” obtida computacionalmente através da equação $f = (K^T K)^{-1} K^T g$	30
FIGURA 3.2 –	Distância de um ponto (x_i, y_i) à reta $y = a + bx$ (Souza, 2006)	32
FIGURA 4.1 –	Valores exatos da matriz “ f ” X Valores obtidos por inversão comum de matriz	45
FIGURA 4.2 –	Valores exatos da matriz “ f ” X Valores obtidos por mínimos quadrados	45
FIGURA 4.3 –	Valores exatos da matriz “ f ” X Valores obtidos por Tikhonov	46
FIGURA 5.1 –	Circunferência de $R = 1$ cm inscrita em um quadrado de $L = 2$ cm	53
FIGURA 5.2 –	Cálculo de π utilizando o Método de Monte Carlo	54
FIGURA 5.3 –	Percentual de partículas que chegam à outra extremidade da caixa para 1000 partículas emitidas pela fonte	55
FIGURA 5.4 –	Percentual de partículas que chegam à outra extremidade da caixa para 100 interações de cada partícula	56
FIGURA 5.5 –	Percentual de partículas que chegam à outra extremidade da caixa para 100 interações de 1000 partículas emitidas pela fonte	57
FIGURA 5.6 –	Resultado da integral (5.5) utilizando Método de Monte Carlo	58
FIGURA 5.7 –	Incerteza no cálculo da integral (5.5) utilizando Método de Monte Carlo	59
FIGURA 5.8 –	Resultado da integral (5.5) utilizando Método de Monte Carlo com função peso	61
FIGURA 5.9 –	Incerteza no cálculo da integral (5.5) utilizando Método de Monte Carlo com função peso	61
FIGURA 5.10 –	Resultado comparado da integral (5.5) utilizando Método de Monte Carlo	62

FIGURA 5.11 – Comparação da incerteza no cálculo da integral (5.5) utilizando Método de Monte Carlo	62
FIGURA 5.12 – Acoplamento genético de animais usando o Algoritmo de Metropolis	71
FIGURA 6.1 – Histórico aleatório de um nêutron incidente sobre uma placa de material que pode sofrer fissão	75
FIGURA 6.2 – Fluxo de radiação (partículas/cm ²) em função da energia da fonte de fótons incidindo em uma placa de chumbo com 1,0 cm de espessura	98
FIGURA 6.3 – Fluxo de radiação (partículas/cm ²) em função da espessura da placa de chumbo	99
FIGURA 6.4 – Fluxo de radiação (partículas/cm ²) em função da distância da fonte de radiação até a placa de chumbo	100
FIGURA 6.5 – Fluxo de radiação (partículas/cm ²) em função do material da placa	101
FIGURA 6.6 – Curva de PDP simulada com o MCNP	102
FIGURA 6.7 – Curvas de isodose mostradas no certificado de calibração do GammaCell 220 do CDTN	103
FIGURA 6.8 – Geometria do GammaCell 220 do CDTN simulada no MCNP	105
FIGURA 6.9 – Comparação entre as curvas de isodose obtidas a partir do certificado de calibração, da dosimetria TLD, da dosimetria Fricke e da simulação com o MCNP (Adaptado de Rodrigues <i>et. al</i> , 2009)	107
FIGURA 7.1 – Variação na Dose Absorvida medida na profundidade de 1,5 cm quando da mudança de taxa de dose no acelerador linear CLINAC 600 C/D (adaptado de Sousa, 2009)	115
FIGURA 7.2 – Variação na Dose Absorvida medida na profundidade de 5,0 cm quando da mudança de taxa de dose no acelerador linear CLINAC 600 C/D (adaptado de Sousa, 2009)	116
FIGURA 7.3 – Variação na Dose Absorvida medida na profundidade de 15,0 cm quando da mudança de taxa de dose no acelerador linear CLINAC 600 C/D (adaptado de Sousa, 2009)	116
FIGURA 7.4 – Radiosensibilidade de acordo com a idade do indivíduo (adaptado de Bushong, 2001)	117

FIGURA 7.5 –	Três tipos mais comuns de <i>kernels</i> segundo Ahnesjö <i>et. al</i> (1999)	122
FIGURA 7.6 –	Geometria para medida do PDP (IAEA, 2003)	124
FIGURA 7.7 –	Geometria para medida do TAR (IAEA, 2003)	125
FIGURA 7.8 –	Geometria para medida do TPR (IAEA, 2003)	126
FIGURA 7.9 –	Kernel elementar de dose $h(r,r')$ (Lind, 1990)	129
FIGURA 7.10 –	PDP calculada utilizando kernel proposto X PDP obtida experimentalmente – Campo (5x5) cm ² ; distância fonte-superfície = 100,0 cm	131
FIGURA 7.11 –	PDP calculada utilizando kernel proposto X PDP obtida experimentalmente – Campo (10x10) cm ² ; distância fonte-superfície = 100,0 cm	132
FIGURA 7.12 –	PDP calculada utilizando kernel proposto X PDP obtida experimentalmente – Campo (15x15) cm ² ; distância fonte-superfície = 100,0 cm	133
FIGURA 7.13 –	PDPs obtidas experimentalmente; distância fonte-superfície = 100,0 cm	134
FIGURA 7.14 –	PDP obtida a partir da inversão de dados experimentais via regularização de Tikhonov X PDP obtida experimentalmente – Campo (5x5) cm ² ; distância fonte-superfície = 100,0 cm	135
FIGURA 7.15 –	PDP obtida a partir da inversão de dados experimentais via regularização de Tikhonov X PDP obtida experimentalmente – Campo (10x10) cm ² ; distância fonte-superfície = 100,0 cm	136
FIGURA 7.16 –	PDP obtida a partir da inversão de dados experimentais via regularização de Tikhonov X PDP obtida experimentalmente – Campo (15x15) cm ² ; distância fonte-superfície = 100,0 cm	137
FIGURA 8.1 –	Função energia potencial invertida (o) e o seu valor exato (linha cheia)	147
FIGURA 8.2 –	Seção de choque diferencial calculada a partir do potencial invertido (o) e o seu valor exato (linha cheia)	148

LISTA DE TABELAS

TABELA 2.1 –	Comparação entre os valores da matriz f obtida pela equação (2.14) e pela equação (2.3)	23
TABELA 3.1 –	Comparação entre os valores da matriz f obtida pela equação (2.14) e pela equação (3.16)	30
TABELA 3.2 –	Pontos (x_i, y_i) para determinação da reta $y = a + bx$	34
TABELA 4.1 –	Comparação entre os valores da matriz f obtida pela equação (2.14) e pela equação (4.22)	45
TABELA 5.1 –	Valores calculados para π ao longo do tempo	52
TABELA 5.1 –	Valores calculados para π ao longo do tempo (continuação)	53
TABELA 6.1 –	Cartões de superfície do MCNP	79
TABELA 6.1 –	Cartões de superfície do MCNP (continuação)	80
TABELA 6.2 –	Cartões de saída do MCNP	92
TABELA 6.3 –	Interpretação do erro relativo	96
TABELA 6.4 –	Raio do detector em anel e dose relativa	106
TABELA 6.5 –	Comparação entre as isodoses certificadas e simuladas	106

LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS

a.C	<u>a</u> ntes de <u>C</u> risto
ACTI	<u>A</u> dvanced <u>C</u> omputational <u>T</u> echnology <u>I</u> nitiative
ACTL	<u>A</u> CTivation <u>L</u> ibrary
CDTN	<u>C</u> entro de <u>D</u> esenvolvimento da <u>T</u> ecnologia <u>N</u> uclear
CLINAC	<u>C</u> linical <u>L</u> inear <u>A</u> ccelerator
d.C	<u>d</u> e <u>p</u> ois de <u>C</u> risto
ENDF	<u>E</u> valuated <u>N</u> uclear <u>D</u> ata <u>F</u> ile
ENDL	<u>E</u> valuated <u>N</u> uclear <u>D</u> ata <u>L</u> ibrary
EPDL	<u>E</u> valuated <u>P</u> hoton <u>D</u> ata <u>L</u> ibrary
IAEA	<u>I</u> nternational <u>A</u> tomio <u>E</u> nergy <u>A</u> gency
ICEx	<u>I</u> nstituto de <u>C</u> iências <u>E</u> xatas
ICRU	<u>I</u> nternational <u>C</u> ommission on <u>R</u> adiation <u>U</u> nits
KERMA	<u>K</u> inetic <u>E</u> nergy <u>R</u> eleased in <u>M</u> aterial
MATLAB	<u>M</u> ATrix <u>L</u> ABoratory
MCNP	<u>M</u> onte <u>C</u> arlo <u>N</u> - <u>P</u> article Code
MU	<u>M</u> onitor <u>U</u> nits
MV	<u>M</u> ega <u>V</u> oltagem
PDP	<u>P</u> ercentagem de <u>D</u> ose <u>P</u> rofunda
SAD	<u>S</u> ource- <u>A</u> xis <u>D</u> istance
SSD	<u>S</u> ource- <u>S</u> kin <u>D</u> istance
TAR	<u>T</u> issue- <u>A</u> ir <u>R</u> atio
TLD	<u>T</u> hermo <u>L</u> uminescent <u>D</u> osemeter
TMR	<u>T</u> issue- <u>M</u> ass <u>R</u> atio
TPR	<u>T</u> issue- <u>P</u> hantom <u>R</u> atio
TRS	<u>T</u> echnical <u>R</u> eport <u>S</u> eries
UFMG	<u>U</u> niversidade <u>F</u> ederal de <u>M</u> inas <u>G</u> erais

CAPÍTULO I

INTRODUÇÃO GERAL

1.1 – O que são Problemas Inversos

A caracterização de um problema como sendo direto ou inverso é uma questão arbitrária. Entretanto, a comunidade matemática internacional considera que a maior parte do treinamento matemático feito em currículos de graduação é dominada por **problemas diretos**, isto é, problemas que podem ser caracterizados como aqueles que contêm informações suficientemente exatas para que o estudante seja capaz de encontrar uma solução única, estável e bem-definida (Groetsch, 1999). Nesses casos, um processo é descrito em detalhes e uma entrada é fornecida ao estudante que, espera-se, encontrará uma solução para o mesmo. Em ciência esse processo seria assim descrito:

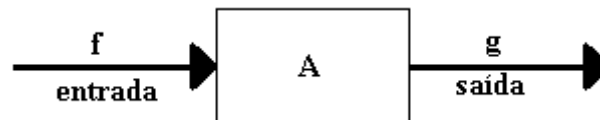


Figura 1.1 – Forma direta de processamento

Uma expressão simbólica que pode ser utilizada para representar matematicamente a figura (1.1) é descrita através da equação (1.1) a seguir:

$$g = Af \quad (1.1)$$

Para cada entrada f proveniente de certa classe de valores, ter-se-ia uma única solução possível g . Matematicamente pode-se dizer que g é função de f e que A é o operador utilizado sobre f para produzir g (Moritz, 1993). Como exemplo pode-se citar:

$$g = \cos f \quad (1.2)$$

A definição de **problemas inversos** não é simples. Dois problemas inversos podem ser imediatamente visualizados a partir de qualquer problema direto. O primeiro é a causa do problema: dado o operador A e o efeito g , encontre a causa do efeito. A solução para esse caso seria encontrada através de A^{-1} , desde que o modelo pudesse ser invertido, o que nem sempre acontece. O segundo problema inverso que surge é na identificação do modelo: dada uma informação causa-efeito, identifique o modelo, ou seja, dados f e g , encontre A . Se A for realmente um operador, isto é, uma função, então para qualquer valor

de f , um único g será encontrado. Isso significa que o problema direto tem **solução única**, diferentemente do problema inverso. Além disso, se o operador A for contínuo, a solução do problema direto será **estável** mesmo com pequenas alterações na entrada; ainda que o operador tenha uma inversão bem-definida, de forma que a causa do problema tenha solução única, é impossível assegurar que sua solução seja estável, pois o operador inverso poderá ser descontínuo (Groetsch, 1999).

Em casos gerais, o operador A pode ser definido como uma matriz ou até mesmo como um operador linear ou não-linear. Em muitos casos as soluções dos problemas diretos são representadas por **integrais médias**. No caso do problema inverso o operador A deve ser **diferenciado**. Essa é a razão pela qual as soluções de problemas inversos levam a problemas instáveis e não-lineares muito delicados para realização numérica (Anger *et. al*, 1993).

Um problema direto trivial seria, por exemplo, quando se tem uma equação e se deseja encontrar sua solução. Pode-se, nesse caso, dizer que o problema inverso seria quando se tivesse a solução e se estivesse à procura da equação. Outro problema direto que pode ser citado seria quando se tem dois números e se quer encontrar seu produto. Nesse caso, o problema inverso seria encontrar o par de números que, quando multiplicados, totalizariam o número fornecido inicialmente. Deve-se destacar que os problemas inversos podem ter uma **solução única**, **várias soluções** ou **nenhuma solução**.

Os problemas inversos têm influenciado grandemente o desenvolvimento das ciências naturais, o que não tem sido merecidamente destacado devido à ênfase dada ao ensinamento da resolução de problemas diretos. Grandes avanços na ciência e na tecnologia têm sido possíveis a partir dos problemas inversos, como na determinação de leis físicas originadas de observações e medidas indiretas, na reconstrução de eventos passados a partir da observação de dados presentes e outros.

Quando as observações não são suficientes para se determinar os parâmetros desconhecidos, informação adicional se torna necessária. Cada esquema matemático capaz de produzir uma solução única para um problema inverso a partir de dados insuficientes deve incorporar informações adicionais e é por essa razão que diferentes algoritmos, aplicados aos mesmos dados, podem fornecer respostas completamente diferentes (Anger

et. al, 1993). Na prática, se essa inserção de informações for insuficiente ou inadequada, esse feito pode se tornar perigoso. Portanto, para um dado método ou técnica envolvida na solução de um problema inverso, deve-se ter cuidado para sempre considerar seus limites de aplicabilidade e as suposições envolvidas relativas ao caso. Para os matemáticos, problemas inversos são raramente fáceis, comumente difíceis e, algumas vezes, insolúveis, mas sempre fascinantes.

Como um problema inverso geralmente não tem solução única, é extremamente desejável ter-se uma maneira de determinar a coleção de todas as soluções possíveis. Em certos casos isso é possível, mas ainda existe grande distância a ser percorrida para se conseguir esse objetivo na maioria deles. Parte da resolução dos problemas inversos está relacionada com a **intuição** e essa depende da **experiência** de cada um. Intuição é um assunto difícil e delicado, especialmente quando se trata de textos técnicos. Muitos cientistas e filósofos acreditam que uma lei da natureza ou uma fórmula física podem ser encontradas diretamente a partir da dedução lógica de um dado. Na realidade, isso é um problema inverso cuja solução está longe de ser única e de ser direta. Isso faz com que o trabalho de um cientista, como “detetive da natureza”, seja muito mais excitante, frustrante e recompensador do que o de qualquer detetive de histórias (Anger *et. al*, 1993).

1.2 – Resolução de Problemas Inversos

Para se resolver um problema inverso, os seguintes pontos são importantes (Anger *et. al*, 1993):

- Domínio experimental e teórico dos processos envolvidos;
- Possibilidade de modelamento matemático do processo;
- Domínio experimental e teórico do problema direto;
- Entendimento das informações relacionadas ao problema inverso;
- Desenvolvimento de algoritmos para solução numérica do problema inverso.

Métodos matemáticos altamente sofisticados são necessários para se resolver um problema inverso. Grande parte da matemática necessária teve que ser desenvolvida para esse propósito (Anger *et. al*, 1993). Em ciência, tecnologia e medicina, muitos resultados são obtidos em laboratórios. Dentre eles muitos parâmetros internos de um

processo especial podem ser mantidos constantes. Para se utilizar esses resultados, aparatos físicos especiais ou máquinas devem ser disponibilizados. A maioria dos resultados obtidos matematicamente pode ser provada usando métodos específicos.

1.3 – Problemas Mal-Colocados

Um problema é chamado **bem-colocado**, segundo o matemático francês Jacques Salomon Hadamard (1865-1963), se satisfizer às três seguintes condições (Anger *et. al*, 1993):

- A solução deve existir (**existência**);
- A solução deve ser unicamente determinada pelos dados (**unicidade**);
- A solução deve depender continuamente dos dados (**estabilidade**).

Isso significa que a solução deve existir para uma determinada faixa arbitrária de dados, que deve existir apenas uma única solução e que essa solução deve depender continuamente dos dados fornecidos inicialmente. Se qualquer uma dessas condições não for satisfeita, tem-se um **problema mal-colocado**.

Ao longo de muito tempo pensou-se que somente problemas bem-colocados tinham significado físico. De fato, processos determinísticos como, por exemplo, os considerados na mecânica clássica, dependem única e continuamente dos dados iniciais e correspondem, portanto, a problemas bem-colocados. Durante as últimas décadas, um grande número de questões matemáticas de relevância física teve que ser interpretada como um problema mal-colocado por não satisfazer pelo menos uma das três condições citadas anteriormente, especialmente questões relacionadas à interpretação física de dados medidos.

1.4 – Problemas Inversos de Primeira Ordem

Consistem em se resolver a equação (1.1) para f . Uma solução formal seria dada pela equação (1.3) a seguir:

$$f = A^{-1}g \quad (1.3)$$

em que g representa o dado medido (experimental) e f representa os parâmetros que descrevem o objeto. Geralmente o operador A trabalha na direção da natureza para o observador enquanto A^{-1} opera na direção oposta (Anger *et. al*, 1993).

1.5 – Problemas Inversos de Segunda Ordem

Consistem na determinação do operador A considerando que f e g sejam fornecidos. O operador A deve ser considerado como um modelo matemático ou como uma lei.

Geralmente problemas inversos de segunda ordem não possuem solução única e costumam ser de resolução mais complicada do que os de primeira ordem.

1.6 – Por que ensinar Problemas Inversos

A solução de problemas inversos é um fator importante no desenvolvimento da matemática e das ciências. Sozinha é uma razão que justifica o estudo ocasional de problemas inversos nos períodos iniciais de um currículo de graduação.

Entretanto, o mais importante no estudo de problemas inversos é o fato de desenvolver o hábito de “pensar inversamente”. Se se considerar apenas o problema direto, a situação não será analisada por todos os lados. Esse hábito de enxergar os problemas apenas do ponto de vista direto é intelectualmente limitado (Anger *et. al*, 1993).

1.7 – Problemas Inversos em Física Computacional

Muitos exemplos de problemas inversos são desenvolvidos por técnicas de física computacional, como em astrofísica, física atômica, biofísica, geofísica, física de alta energia, hidrodinâmica, física nuclear, física de plasma, física do estado sólido e mecânica estatística (Nedelkov, 1972). Algoritmos computacionais utilizados com sucesso para resolução de problemas bem-colocados falham quando se tenta sua utilização para alguns problemas inversos. Os seguintes problemas inversos em física computacional podem ser citados (Nedelkov, 1972):

- Astrofísica: determinação da distribuição da intensidade de radiação proveniente do disco de corpos pesados pela observação de seu brilho em um eclipse; determinação da distribuição de massas nas galáxias; determinação de radiobrilho;
- Física atômica: o problema inverso da espectroscopia;
- Biofísica: determinação da distribuição do biopotencial U dentro do corpo quando os valores de U na superfície corporal são conhecidos; cálculo da distribuição da concentração de substâncias radioativas nos organismos a partir de medidas experimentais;
- Geofísica: continuação analítica do potencial gravitacional e magnético; continuação analítica do potencial U de correntes elétricas dentro da Terra a partir de valores medidos em sua superfície;
- Física de alta energia: continuação analítica da amplitude de espalhamento parcial na região não-física da segunda planilha de Riemann;
- Hidrodinâmica: fluxo potencial ao redor de corpos com simetria axial; fluxo ao redor de um corpo na presença de ondas de choque;
- Física nuclear: o problema inverso do espalhamento mecânico quântico; o problema inverso da espectroscopia nuclear;
- Física de plasma: determinação da função distribuição de velocidade de elétrons a partir de curvas de dispersão;
- Física do estado sólido: determinação de espectro de cristais a partir de sua capacidade calorífica;
- Mecânica estatística: determinação do comportamento de quantidades termodinâmicas na vizinhança de temperaturas críticas onde ocorrem transições de fase;
- Decomposição térmica: estudos de decomposição térmica de sólidos baseados principalmente em teorias de crescimento e difusão de partículas;
- Radioterapia: problemas inversos em tratamentos radioterápicos de tumores malignos.

1.8 – Exemplo de um Problema Inverso

No exemplo descrito a seguir, a intenção é desenvolver-se um modelo simplificado do movimento de um projétil, através do estudo do problema do “alcance inverso”, além de

se investigar a existência e a unicidade de soluções. Para seu completo desenvolvimento, são necessários conhecimentos de pré-cálculo (álgebra e geometria de coordenadas), de matemática (parábolas, hipérbolas e quadráticas), de base científica (lei da inércia, lei de corpos em queda) e de tecnologia (cálculo de gráficos).

No século XVI, os problemas de balística eram uma preocupação técnica fundamental. Os novos canhões provavam sua eficácia mortífera em várias guerras ao longo do tempo. Estudos de balística conduziram a Itália Renascentista à moderna teoria da dinâmica e podem ser vistos como a origem da física matemática.

O primeiro livro ocidental sobre canhões, chamado *New Science*, foi publicado em 1537 por Nicolo Tartaglia (Nicolo Fontana, 1500?-1557). Tartaglia se dizia o inventor do mecanismo para medição do ângulo de elevação de um tubo de canhão com respeito à horizontal. Esse problema era tratado de forma direta: dado o ângulo de elevação, encontrar o alcance horizontal da bala do canhão para uma determinada carga de pólvora. Entretanto o problema inverso correspondente já era sugerido: encontrar o ângulo de elevação que permitirá determinado alcance da bala.

A análise do movimento de projétil feita por Tartaglia falhou, mas um século depois Galileo fez uma nova análise correta sobre o movimento de um projétil puntual no vácuo. As afirmações de Galileo eram embasadas firmemente em três pilares científicos: sua lei da inércia, sua lei de corpos em queda e a idéia da composição de movimentos.

Supõe-se um projétil puntiforme de massa unitária sendo lançado da origem com um ângulo θ em relação ao eixo-x positivo, com velocidade inicial v . A idéia de Galileo era que o movimento com velocidade v na direção θ poderia ser decomposto em dois movimentos independentes com direções x e y , com velocidades v_x e v_y respectivamente.

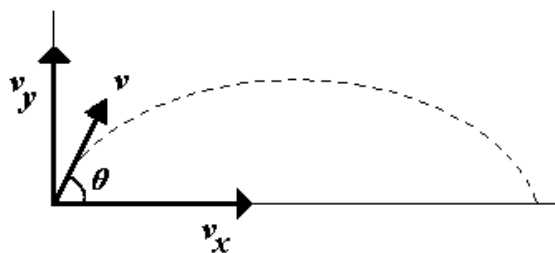


Figura 1.2 – Representação gráfica do movimento do projétil

As relações entre v e seus componentes são assim descritas:

$$\frac{v_x}{v} = \cos \theta \quad (1.4)$$

$$\frac{v_y}{v} = \sin \theta \quad (1.5)$$

Considere agora o movimento na direção do eixo-x somente. Como nenhuma força age nessa direção (desprezar resistência do ar no eixo-x e força da gravidade no eixo-y), a lei da inércia de Galileo afirma que a velocidade na direção do eixo-x é constante. Mas a velocidade inicial no eixo-x é dada por $v_x = v \cos \theta$ e, portanto, a velocidade no eixo-x será sempre v_x . Entretanto, após um tempo decorrido t , o projétil terá se movido para uma nova posição cuja coordenada x será dada por:

$$x = v_x t \quad (1.6)$$

No eixo-y duas leis têm influência no movimento do projétil. A lei da inércia mostra que o projétil atingirá uma altura $v_y t$ após um tempo decorrido t . Mas durante todo esse tempo o projétil estará sob a ação do campo gravitacional da Terra. De acordo com a lei dos corpos em queda de Galileo, depois de decorrido o tempo t o corpo terá caído $gt^2/2$ unidades de distância, onde g é a constante de aceleração da gravidade. Combinando esses dois efeitos, conclui-se que a coordenada y do projétil depois de decorrido um tempo t será dada por:

$$y = -\frac{gt^2}{2} + v_y t \quad (1.7)$$

As duas equações para x e y demonstradas anteriormente dão a representação paramétrica da curva ou a trajetória que o projétil deverá seguir.

Esse problema pode ser resolvido de forma inversa, dependendo dos dados fornecidos:

- Se for fornecido o alcance da bala e sua velocidade inicial de lançamento, pode-se calcular o ângulo de inclinação correspondente do canhão:

$$\text{sen}2\theta = \frac{x_{\max}g}{v_0^2} \quad (1.8)$$

- Se for fornecida a altura máxima da bala e sua velocidade inicial de lançamento, pode-se calcular o ângulo de inclinação correspondente do canhão:

$$\text{sen}^2\theta = \frac{y_{\max}2g}{v_0^2} \quad (1.9)$$

- Se for fornecido o tempo de subida da bala e sua velocidade inicial de lançamento, pode-se calcular o ângulo de inclinação correspondente do canhão:

$$\text{sen}\theta = \frac{t_s g}{v_0} \quad (1.10)$$

- Se for fornecido o tempo total de voo da bala e sua velocidade inicial de lançamento, pode-se calcular o ângulo de inclinação correspondente do canhão:

$$\text{sen}\theta = \frac{t_t g}{2v_0} \quad (1.11)$$

Esse problema inverso sugerido é mal-colocado uma vez que:

- Sua solução pode não existir: pode-se tentar achar, erroneamente, um ângulo de um suposto alcance que extrapola o alcance máximo possível;
- Sua solução pode não ser única: diferentes ângulos podem proporcionar o mesmo alcance do projétil;
- Sua solução pode não ser estável: dependendo de ruídos que podem surgir no lançamento do projétil, pode-se obter alcances errôneos.

1.9 – Conclusão

Embora não exista uma regra para se definir exatamente o que é um problema inverso em relação a um problema direto, é óbvio que ambos existem e que ambos são muito importantes para o desenvolvimento da ciência. Com o emprego de dados experimentais torna-se imprescindível o uso de técnicas matemáticas adequadas a fim de se modelar tais experimentos. Isso se torna mais evidente quando se tem problemas mal-colocados, bastante corriqueiros em diversas áreas da ciência, em especial na físico-química, como no caso da decomposição térmica de sólidos, entre outros. Os problemas inversos representam um novo horizonte para a caracterização de ciências experimentais, onde se consegue modelar matematicamente com precisão os fenômenos que ocorrem na natureza e em laboratórios de pesquisa. É imprescindível o estudo das técnicas associadas à resolução de problemas inversos.

CAPÍTULO II

EQUAÇÃO INTEGRAL DE FREDHOLM DE PRIMEIRA ORDEM

2.1 – A Equação

Diversos experimentos em físico-química podem ser modelados através de equações integrais lineares. Uma das equações mais utilizadas é conhecida como equação integral de Fredholm. Equações de Fredholm envolvem integrais definidas, com limites inferiores e superiores.

Uma equação inhomogênea de Fredholm de Primeira Ordem possui a seguinte estrutura:

$$\int_a^b K(x, y)f(y)dy = g(x) \quad (2.1)$$

em que $f(y)$ é uma função desconhecida a ser solucionada enquanto $g(x)$ é uma parte conhecida. A função de duas variáveis $K(x,y)$ é chamada *kernel*. Essa equação é análoga à equação matricial (2.2) a seguir:

$$Kf = g \quad (2.2)$$

A solução da equação (2.2) acima é dada pela equação (2.3):

$$f = K^{-1}g, \quad (2.3)$$

em que K^{-1} é a matriz inversa

As equações (2.1) e (2.2) possuem solução única quando g for diferente de zero e K for invertível. Casos onde g seja igual a zero dificilmente têm utilidade. Dados experimentais certamente serão utilizados na solução dessas equações.

2.2 – Representação da Equação de Fredholm

Suponha que a equação integral de Fredholm de Primeira Ordem deva ser calculada nos pontos x_1, x_2, \dots, x_n :

$$\begin{aligned}
\int_a^b K(x_1, y) f(y) dy &= g(x_1) \\
\int_a^b K(x_2, y) f(y) dy &= g(x_2) \\
&\dots \\
\int_a^b K(x_m, y) f(y) dy &= g(x_m)
\end{aligned}
\tag{2.4}$$

Resolvendo para um x_i fixo:

$$\int_a^b K(x_i, y) f(y) dy = \sum_j^n w_j K(x_i, y_j) f(y_j) = g(x_i)
\tag{2.5}$$

De maneira simplificada:

$$\sum_{j=1}^n K_{ij} f_j = g_i,
\tag{2.6}$$

em que $K_{ij} = w_j K(x_i, y_j)$, $f_j = f(y_j)$, $g_i = g(x_i)$

Na forma matricial:

$$Kf = g
\tag{2.7}$$

em que $f \in R^n$, $g \in R^m$ e $K \in R^{m \times n}$. O valor de n refere-se ao número de pontos do procedimento de discretização enquanto que o valor de m refere-se ao número de pontos experimentais.

Nesse caso, um problema mal-colocado pode ser definido como uma situação na qual pelo menos uma das condições a seguir não é atendida:

- A solução do problema, f , é única em R^n (**unicidade**);
- Para cada $f \in R^n$ existe um $g \in R^m$ tal que $Kf = g$ (**existência**);
- A dependência de f com g é contínua (**continuidade**).

Como encontrar a solução da equação integral de Fredholm? Para facilitar, considere $n = m$. Assim sendo, pode-se tentar encontrar a matriz inversa de K , que fornecerá a solução $f = K^{-1}g$.

2.3 – Representação em uma Base

Um dos métodos utilizados para a resolução de integrais definidas é a chamada **regra de integração trapezoidal**. Considere a integral definida na equação (2.8) abaixo, com representação gráfica na figura (2.1) a seguir:

$$\int_a^b f(x)dx \quad (2.8)$$

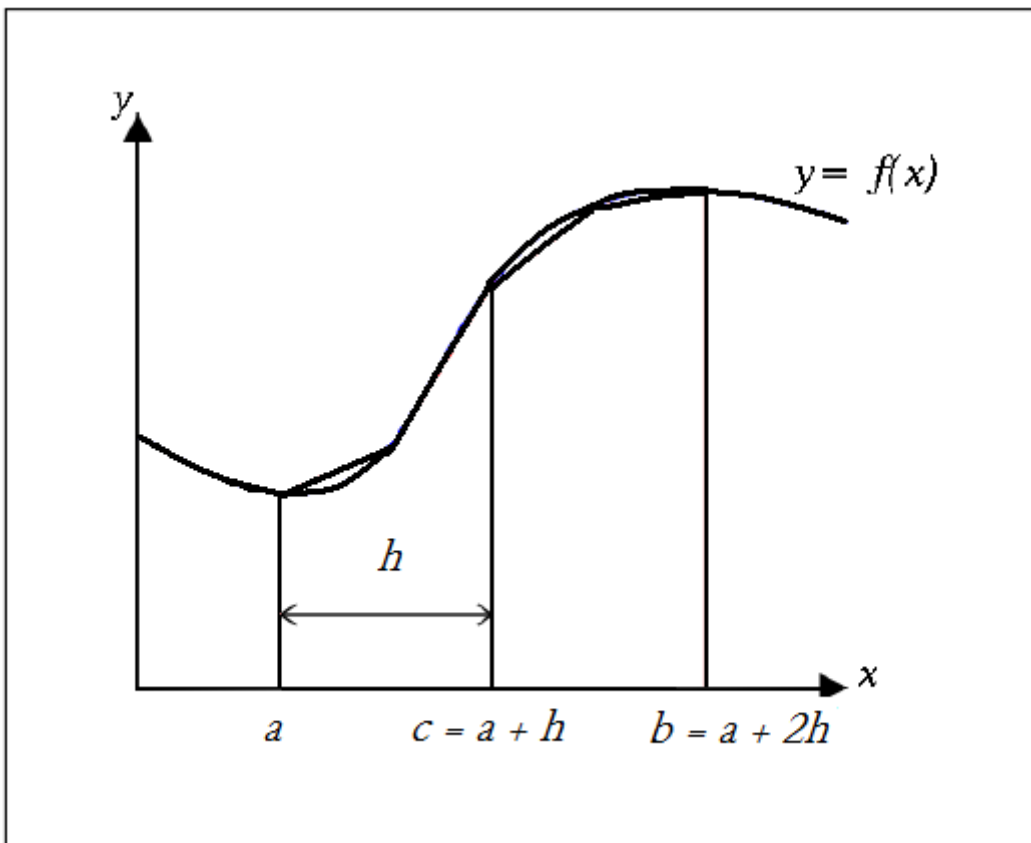


Figura 2.1 – Área sob a curva $y = f(x)$ dividida em n faixas, onde $h = (b-a)/n$

A área de cada uma das n faixas sob o gráfico tem a forma aproximada de um trapézio. O somatório de todas essas áreas trapezoidais dá uma aproximação da integral definida na equação (2.8). Como a região compreendida por cada uma das faixas tem o

formato aproximado de um trapézio, pode-se utilizar a equação de cálculo de área dessa figura para todas as faixas. Para a primeira faixa, tem-se:

$$A_1 = \frac{f(a) + f(a+h)}{2} h \quad (2.9)$$

A área para uma faixa i qualquer é dada por:

$$A_i = \frac{f(a + (i-1)h) + f(a + ih)}{2} h \quad (2.10)$$

O somatório dessas áreas dará um valor aproximado para a integral definida:

$$\int_a^b f(x) dx \approx \sum_{i=1}^n \frac{f(a + (i-1)h) + f(a + ih)}{2} h \quad (2.11)$$

Quanto maior o valor de n , maior a proximidade do valor calculado com o valor real da integral.

Para um x_i fixo, o integrando da equação de Fredholm de Primeira Ordem $K(x_i, y)f(y)$ é função da variável y e pode ser representado como $\Phi(y)$. Considerando-se o intervalo compreendido entre a e b , tem-se os pontos a , $c = a + h$ e $b = a + 2h$. Através da regra de integração trapezoidal chegar-se-á a:

$$\begin{aligned} \int_a^b \phi(y) dy &= \frac{\phi(a) + \phi(c)}{2} h + \frac{\phi(c) + \phi(b)}{2} h = \\ &= \frac{\phi(a)}{2} h + \phi(c)h + \frac{\phi(b)}{2} h = \phi(a) \frac{h}{2} + \phi(c)h + \phi(b) \frac{h}{2} \end{aligned} \quad (2.12)$$

Comparando-se as equações (2.12) e (2.5), conclui-se que:

$$w_1 = \frac{h}{2}, w_2 = h, w_3 = \frac{h}{2}, y_1 = a, y_2 = c = a + h, y_3 = b = a + 2h, h = \frac{b-a}{2}$$

Generalizando-se para n pontos:

$$\begin{aligned}
w_1 &= \frac{h}{2}, w_n = \frac{h}{2} \\
w_j &= h, j = 2, 3, \dots, n-1 \\
y_j &= a + (j-1)h, j = 1, 2, \dots, n \\
h &= \frac{b-a}{n-1}
\end{aligned}
\tag{2.13}$$

A título de exemplo, utilizando-se um trabalho de 1985, observa-se que Riele desenvolveu um programa de computador para resolver a equação integral do tipo Fredholm que representa uma situação experimental na físico-química mostrada na equação (2.14) a seguir:

$$\int_a^b K(x, y) f(y) dy = g(x) = \int_a^b \left(\frac{1}{x+y} \right) \frac{1}{y} dy = \frac{1}{x} \ln \left(\frac{1+x/a}{1+x/b} \right),
\tag{2.14}$$

em que $a \leq y \leq b$

Portanto, tem-se:

$$\begin{aligned}
K(x, y) &= \left(\frac{1}{x+y} \right) \\
f(y) &= \frac{1}{y} \\
g(x) &= \frac{1}{x} \ln \left(\frac{1+x/a}{1+x/b} \right)
\end{aligned}
\tag{2.15}$$

A equação (2.14) pode ser representada na forma matricial:

$$Kf = g
\tag{2.16}$$

Dados o *kernel* ou operador $K(x, y)$, os limites de integração a e b e a função $g(x)$, obtém-se o valor de $f(x)$ como sendo a solução da equação.

Nesse trabalho, para a resolução desse problema, utilizou-se programas computacionais desenvolvidos pelo **Grupo de Pesquisa em Problemas Inversos do Departamento de Química da UFMG**, sob coordenação do professor e pesquisador Dr.

João Pedro Braga, onde são retornadas as matrizes K , f e g utilizando-se o esquema de discretização trapezoidal a meia altura. Esse algoritmo transforma uma equação integral em equação linear, ou seja:

$$\int_a^b K(x, y)f(y)dy = g(x) \Rightarrow Kf = g \quad (2.17)$$

A equação integral de Fredholm de Primeira Ordem é uma equação integral linear. Em problemas inversos utiliza-se soluções lineares inversas.

2.4 – Os Quatro Subespaços Fundamentais, Rank e Índice de Condicionamento

Encontrar a solução de um sistema linear, aqui representado por $Kf = g$, onde $K \in R^{m \times n}$, $f \in R^n$ e $g \in R^m$, significa o mesmo que considerar que esteja ocorrendo a transformação linear K entre os espaços R^n e R^m . Cada um desses espaços vetoriais pode ser dividido em quatro subespaços fundamentais.

1. O *range* (núcleo ou dimensão do espaço) de K , denotado por $R(K)$, é definido como:

$$R(K) = \{g \in R^m \mid Kf = g, f \in R^n\} \quad (2.18)$$

Esse subespaço é também chamado de **espaço coluna de K** . O valor de g deve ser uma combinação linear das colunas de K ; caso o valor de g não pertença ao subespaço, a solução do problema somente poderá ser aproximada. Isso significa que o *range* está relacionado com a **existência** da solução do problema $Kf = g$.

2. O espaço nulo de K , denotado por $N(K)$, é definido como:

$$N(K) = \{f \in R^n \mid Kf = 0\} \quad (2.19)$$

Esse subespaço está relacionado com a **unicidade** da solução do sistema linear de equações. Cada f que pertence ao subespaço corresponde a apenas um elemento de R^n . Esse subespaço também é chamado de **kernel de K** e, caso não seja um conjunto vazio,

múltiplas soluções aparecerão para o problema. Esse fato impede a inversão tradicional de matrizes para a busca da solução.

3. Outro subespaço de R^n é o espaço das linhas de K que são as colunas de K^T , sendo definido como:

$$R(K^T) = \{f \in R^n \mid K^T g = f, g \in R^m\} \quad (2.20)$$

Para um valor de f que pertence ao espaço R^n só existem duas possibilidades: ou ele pertence a $R(K^T)$ ou pertence a $N(K)$. Se o valor de f não pertence ao espaço nulo, ele deverá obrigatoriamente pertencer ao *range* da transposta de K (K^T), o que significa que esse subespaço é o **espaço solução de K** , denotado por $S(K)$.

4. Outro subespaço de R^m é o espaço nulo de K^T , sendo definido como:

$$N(K^T) = \{g \in R^m \mid K^T g = 0\} \quad (2.21)$$

Para um valor de g que pertence ao espaço R^m só existem duas possibilidades: ou ele pertence a $R(K)$ ou pertence a $N(K^T)$.

É importante frisar que esses quatro subespaços fundamentais apresentam as seguintes propriedades:

$$N(K) \perp R(K^T) \quad , \quad N(K^T) \perp R(K) \quad (2.22)$$

, em que o símbolo \perp representa a ortogonalidade.

Como $N(K)$ e $R(K^T)$ são os únicos subespaços de R^n e como são **ortogonais** entre si, eles também são **complementares**, ou seja, qualquer elemento de R^n pode ser calculado, conhecendo-se a base destes dois subespaços, pois:

$$\begin{aligned} R^n &= N(K) \cup R(K^T) \\ R^m &= N(K^T) \cup R(K) \end{aligned} \quad (2.23)$$

Assim sendo, a **dimensão** dos subespaços fundamentais pode ser definida como:

$$\begin{aligned}\dim[N(K)] + \dim[R(K^T)] &= \dim[R^n] = n \\ \dim[N(K^T)] + \dim[R(K)] &= \dim[R^m] = m\end{aligned}\tag{2.24}$$

O **rank** ou **posto** de uma matriz é um número igual ao número máximo de colunas linearmente independentes de K :

$$\text{rank}(K) = \dim[R(K)]\tag{2.25}$$

É importante frisar que:

$$\text{rank}(K) = \text{rank}(K^T)\tag{2.26}$$

$$\text{rank}(K^T K) = \text{rank}(K)\tag{2.27}$$

O **rank** de uma matriz pode ser, no máximo, igual ao menor valor de m ou n . Com um valor menor ela é dita matriz de **rank deficiente**; caso contrário, será de **rank cheio**.

Para se compreender o **índice de condicionamento**, considere o sistema linear $Kf = g$, no qual uma pequena perturbação seja inserida nos dados g , sendo representada por δg . Se a solução for alterada por δf , pode-se escrever (considerando-se que a inversa exista):

$$K(f + \delta f) = g + \delta g \Rightarrow \delta g = K^{-1} \delta f\tag{2.28}$$

O **erro relativo** na solução da equação (2.28) será dado por:

$$\frac{\frac{\|\delta f\|}{\|f\|}}{\frac{\|\delta g\|}{\|g\|}} = \frac{\|\delta f\| \|g\|}{\|\delta g\| \|f\|}\tag{2.29}$$

Levando-se em conta que:

$$\begin{aligned}\|g\| &= \|Kf\| \leq \|K\| \|f\| \\ \|\delta f\| &= \|K^{-1} \delta g\| \leq \|K^{-1}\| \|\delta g\|\end{aligned}\tag{2.30}$$

Tem-se que:

$$\frac{\frac{\|\delta f\|}{\|f\|}}{\frac{\|\delta g\|}{\|g\|}} \leq \|K\| \|K^{-1}\|\tag{2.31}$$

O valor obtido pela equação (2.31) é o chamado **índice de condicionamento** da matriz e é uma medida da sensibilidade da solução quando de perturbações nos dados. Se o índice de condicionamento de K for grande, uma pequena alteração nos dados pode levar a uma solução inaceitável para $Kf = g$.

É razoável pensar que, se uma matriz de coeficientes for ligeiramente perturbada, por exemplo, devido a erros de medição, de arredondamento ou resultante de cálculos anteriores, o vetor solução também será apenas ligeiramente modificado. Mas nem sempre isso acontece. Existem situações em que pequenas perturbações relativas nos valores originais acabam por induzir grandes variações relativas na solução de sistemas, que nesse caso recebem o nome de mal-condicionados. A matriz então é dita **mal-condicionada**. Caso isso não aconteça, a matriz é dita **bem-condicionada**.

A título de exemplo de matriz bem-condicionada, considere o sistema $Ax=B$ e as seguintes matrizes:

$$A = \begin{bmatrix} 2,00 & -2,00 \\ 1,00 & 0,99 \end{bmatrix}; B = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Para se encontrar o valor de x , basta resolver a equação $x=A^{-1}B$:

$$Ax = B \Rightarrow x = A^{-1}B = \begin{bmatrix} 2,00 & -2,00 \\ 1,00 & 0,99 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0,7513 \\ 0,2513 \end{bmatrix}$$

Se a matriz A sofrer uma pequena alteração, como a seguir, o valor de x também será pouco modificado (0,5% de diferença nesse caso):

$$A = \begin{bmatrix} 2,00 & -2,00 \\ 1,01 & 0,99 \end{bmatrix}$$

$$Ax = B \Rightarrow x = A^{-1}B = \begin{bmatrix} 2,00 & -2,00 \\ 1,01 & 0,99 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0,7475 \\ 0,2475 \end{bmatrix}$$

A título de exemplo de matriz mal-condicionada, considere o sistema $Ax=B$ e as seguintes matrizes:

$$A = \begin{bmatrix} 2,00 & 2,00 \\ 1,00 & 0,99 \end{bmatrix}; B = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Para se encontrar o valor de x , basta resolver a equação $x=A^{-1}B$:

$$Ax = B \Rightarrow x = A^{-1}B = \begin{bmatrix} 2,00 & 2,00 \\ 1,00 & 0,99 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 50,50 \\ -50,00 \end{bmatrix}$$

Se a matriz A sofrer uma pequena alteração, como a seguir, o valor de x sofrerá uma grande modificação (50% de diferença nesse caso):

$$A = \begin{bmatrix} 2,00 & 2,00 \\ 1,01 & 0,99 \end{bmatrix}$$

$$Ax = B \Rightarrow x = A^{-1}B = \begin{bmatrix} 2,00 & 2,00 \\ 1,01 & 0,99 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 25,25 \\ -24,75 \end{bmatrix}$$

2.5 – Exemplo do Uso de Programas Computacionais

Como exemplo numérico do uso de programas computacionais para a solução de problemas, foi desenvolvido um programa pelo **Grupo de Pesquisa em Problemas Inversos do Departamento de Química da UFMG**, com coordenação do professor e

pesquisador Dr. João Pedro Braga, para se resolver a equação (2.14) de modo que $a = c = 1$, $b = d = 5$ e $n = m$. O exemplo desenvolvido possui $n = 10$.

Ainda no mesmo programa, procura-se obter a matriz f através da equação (2.3). Uma comparação entre o valor de f gerado pela equação (2.14) e o valor obtido através da inversão de K é descrita na tabela e nos dois gráficos a seguir.

Tabela 2.1 – Comparação entre os valores da matriz f obtida pela equação (2.14) e pela equação (2.3)

y	f_{exato}	f_{inv}
1,0000	1,0000	0,4060
1,4444	0,6923	2,1195
1,8889	0,5294	-6,4352
2,3333	0,4286	26,4697
2,7778	0,3600	-66,1396
3,2222	0,3103	114,2314
3,6667	0,2727	-128,1152
4,1111	0,2432	91,5996
4,5556	0,2195	-36,8550
5,0000	0,2000	13,2749

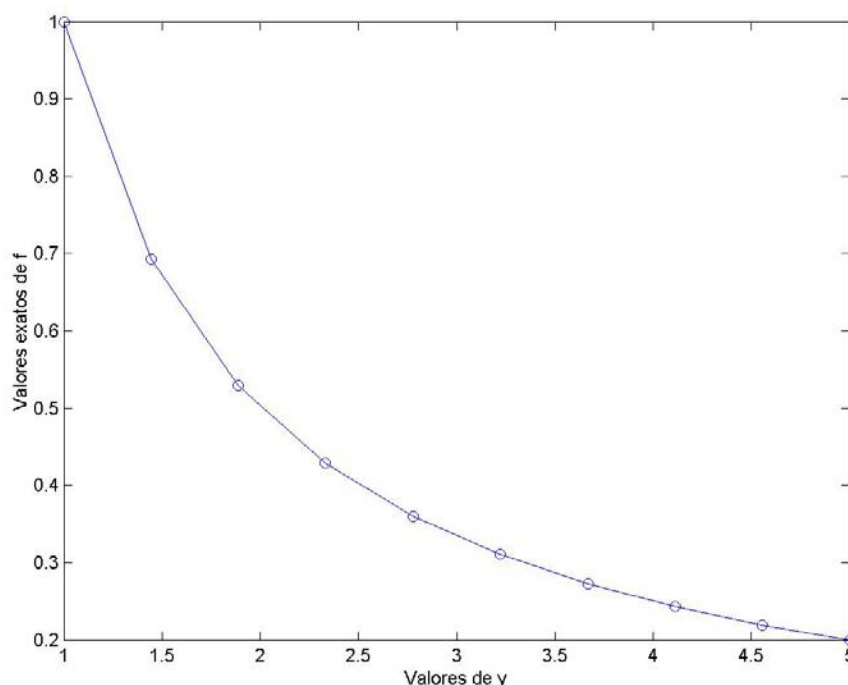


Figura 2.2 – Valores exatos da matriz “ f ” gerada computacionalmente ($f = y^{-1}$)

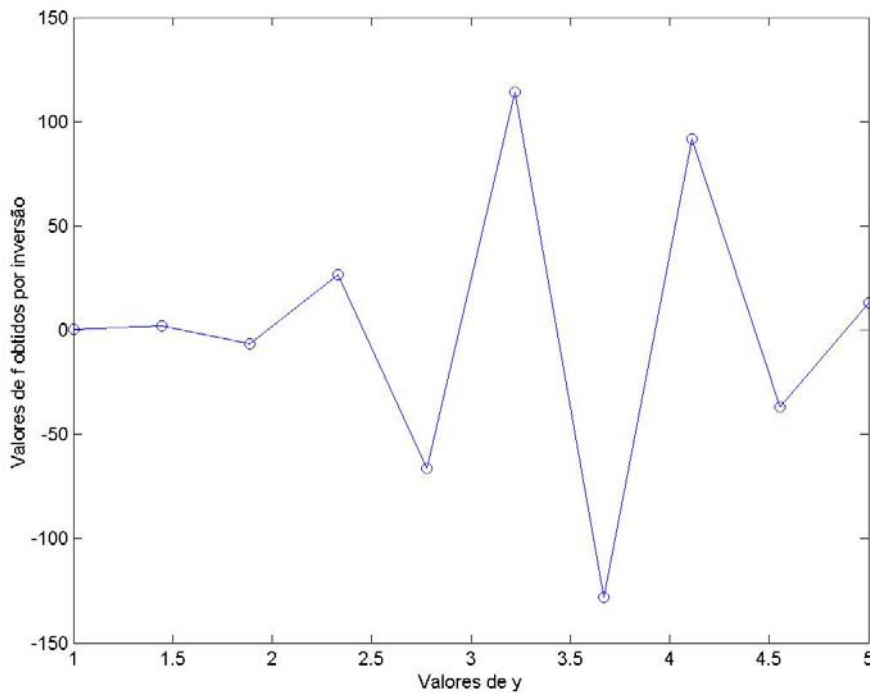


Figura 2.3 – Valores da matriz “f” obtida computacionalmente através da equação $f = K^{-1}g$

Como se pode observar, os valores obtidos através das duas equações são totalmente discrepantes. Isso pode ser explicado em função do índice de condicionamento da matriz. **Um grande valor para o índice de condicionamento indica que o uso da equação (2.3) não é confiável.** Para o exemplo descrito anteriormente, o valor encontrado foi:

$$\text{Cond}(K) = 1,4633 \times 10^{15} \quad (2.32)$$

2.6 – Conclusão

A equação integral de Fredholm de Primeira Ordem tem uma importância muito grande no estudo da ciência, uma vez que vários fenômenos podem ser modelados a partir dessa equação. Para a sua resolução, o uso de álgebra de matrizes associado a técnicas matemáticas adequadas e a poderosos softwares torna a tarefa menos árdua e menos susceptível a erros.

CAPÍTULO III

MÉTODO DOS MÍNIMOS QUADRADOS

3.1 – Introdução

O método dos mínimos quadrados é uma técnica de otimização matemática utilizada na busca do melhor ajuste para um conjunto de dados, onde se procura minimizar a soma dos quadrados das diferenças (resíduos) entre a curva ajustada e os dados utilizados para a sua construção.

A curva dos mínimos quadrados é usualmente um gráfico de um tipo padrão de função de forma linear, polinomial ou trigonométrica, por exemplo. Como os dados podem incluir incertezas de medida ou ajustes experimentais, não se pode esperar que a curva passe por todo o conjunto de dados levantados no experimento. Ao invés disso, a curva deve passar por uma **aproximação otimizada**, baseada na soma dos quadrados dos erros entre os valores das ordenadas descritas no gráfico.

A técnica dos mínimos quadrados foi desenvolvida por Adrien-Marie Legendre (1752-1833) e Carl Friedrich Gauss (1777-1855). O primeiro trabalho sobre o assunto foi publicado por Legendre em 1806, embora hajam evidências claras de que Gauss tenha feito progresso 9 anos antes, utilizando o conhecimento em seus cálculos astronômicos (Leon, 1998).

Um sistema de equações lineares pode ser classificado de acordo com a quantidade de soluções que apresenta:

1. Se o sistema **não possui solução**, ele é chamado **impossível** (há mais equações do que variáveis). Esse tipo de sistema é usualmente **inconsistente**;
2. Se o sistema possui **apenas uma solução**, ele é chamado **possível e determinado (determinação exata)**;
3. Se o sistema possui **muitas soluções**, ele é chamado **possível e indeterminado** (há menos equações do que variáveis). Esse tipo de sistema é usualmente **consistente**.

Um problema de mínimos quadrados pode ser formulado geralmente como um sistema de equações. Um requisito implícito para o método dos mínimos quadrados funcionar é que as incertezas em cada medida sejam distribuídas aleatoriamente, com função densidade gaussiana, além do fato de os resíduos terem de ser independentes.

3.2 – Ortogonalidade

Para um sistema $Ax = b$, do tipo $m \times n$, com $m > n$, geralmente, não se pode esperar encontrar um vetor $x \in \mathbb{R}^n$ para o qual Ax seja igual a b . Ao invés disso, pode-se procurar um vetor x para o qual o valor de Ax seja o mais próximo possível de b . A **ortogonalidade** desempenha, assim, um importante papel na busca do valor de x , uma vez que esse conceito está intimamente ligado à aproximação de valores.

Se $x \in \mathbb{R}^n$, então o **comprimento Euclidiano** ou **norma** de x é definido como:

$$\|x\| = (x^T x)^{1/2} = (x_1^2 + x_2^2 + \dots + x_n^2)^{1/2} \quad (3.1)$$

Para dois vetores x e y em \mathbb{R}^n diferentes de zero tem-se:

$$-1 \leq \frac{x^T y}{\|x\| \|y\|} \leq 1 \quad (3.2)$$

O ângulo θ entre os dois vetores é dado por:

$$\cos \theta = \frac{x^T y}{\|x\| \|y\|}, \quad 0 \leq \theta \leq \pi \quad (3.3)$$

Os vetores x e y são ditos **ortogonais** se $x^T y = 0$. Para representar a ortogonalidade, utiliza-se o símbolo \perp . Portanto, para se mostrar que x e y são ortogonais, escreve-se dessa forma: $x \perp y$. Na estatística utiliza-se o cosseno do ângulo entre dois vetores para se medir o quão perto esses dois vetores estão um do outro.

3.3 – Definição Matemática dos Mínimos Quadrados

Seja A uma matriz do tipo $m \times n$, com $m > n$. Para cada $b \in \mathbb{R}^m$ define-se:

$$\|b\| = \sqrt{\langle b, b \rangle} = \sqrt{b^T b} \quad (3.4)$$

Considere o sistema de equações $Ax=b$. Para cada $x \in \mathbf{R}^n$ tem-se um resíduo:

$$r(x) = b - Ax \quad (3.5)$$

A distância entre b e Ax será dada por:

$$\|b - Ax\| = \|r(x)\| \quad (3.6)$$

Deseja-se obter um vetor $x \in \mathbf{R}^n$ para o qual $\|r(x)\|$ seja mínimo. Minimizar esse valor é equivalente a minimizar o seu quadrado, ou seja, $\|r(x)\|^2$. Um vetor x que possa ser utilizado nessa equação é uma solução de mínimos quadrados para o sistema $Ax=b$.

3.4 – O Método dos Mínimos Quadrados

Nem sempre a inversão de uma matriz é suficiente para se conseguir encontrar valores a partir de dados experimentais. Pode-se utilizar o método dos mínimos quadrados na busca de $f(y)$ na equação (2.1), ou seja, procura-se obter $f(y)$ que dê o mínimo do funcional especificado na equação (3.7) a seguir:

$$\Phi(f) = \|Kf - g\|_2^2 = (Kf - g)^T (Kf - g) \quad (3.7)$$

O valor de $\Phi(f)$ pode ser assim desenvolvido:

$$\Phi(f) = \left(\sum_{j=1}^n K_{1j} f_j - g_1 \right)^2 + \left(\sum_{j=1}^n K_{2j} f_j - g_2 \right)^2 + \dots + \left(\sum_{j=1}^n K_{mj} f_j - g_m \right)^2 \quad (3.8)$$

Num sistema 2x2, por exemplo, tem-se:

$$\Phi(f) = \|Kf - g\|_2^2 = (K_{11}f_1 + K_{12}f_2 - g_1)^2 + (K_{21}f_1 + K_{22}f_2 - g_2)^2 \quad (3.9)$$

Pela regra dos mínimos quadrados, a derivada de $\Phi(f)$ em relação a f deve ser igual a zero:

$$\begin{aligned}\frac{\partial\Phi(f)}{\partial f_1} &= 2(K_{11}f_1 + K_{12}f_2 - g_1)K_{11} + 2(K_{21}f_1 + K_{22}f_2 - g_2)K_{21} = 0 \\ \frac{\partial\Phi(f)}{\partial f_2} &= 2(K_{11}f_1 + K_{12}f_2 - g_1)K_{12} + 2(K_{21}f_1 + K_{22}f_2 - g_2)K_{22} = 0\end{aligned}\tag{3.10}$$

Rearranjando os termos:

$$\begin{aligned}(K_{11}f_1 + K_{12}f_2 - g_1)K_{11} + (K_{21}f_1 + K_{22}f_2 - g_2)K_{21} &= 0 \\ (K_{11}f_1 + K_{12}f_2 - g_1)K_{12} + (K_{21}f_1 + K_{22}f_2 - g_2)K_{22} &= 0\end{aligned}\tag{3.11}$$

Redistribuindo os termos das equações:

$$\begin{aligned}(K_{11}K_{11} + K_{21}K_{21})f_1 + (K_{12}K_{11} + K_{22}K_{21})f_2 &= K_{11}g_1 + K_{21}g_2 \\ (K_{11}K_{12} + K_{21}K_{22})f_1 + (K_{12}K_{12} + K_{22}K_{22})f_2 &= K_{12}g_1 + K_{22}g_2\end{aligned}\tag{3.12}$$

Matricialmente:

$$\begin{bmatrix} K_{11}K_{11} + K_{21}K_{21} & K_{12}K_{11} + K_{22}K_{21} \\ K_{11}K_{12} + K_{21}K_{22} & K_{12}K_{12} + K_{22}K_{22} \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} = \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \end{bmatrix}\tag{3.13}$$

Simplificando:

$$\begin{aligned}\begin{bmatrix} K_{11}K_{11} + K_{21}K_{21} & K_{12}K_{11} + K_{22}K_{21} \\ K_{11}K_{12} + K_{21}K_{22} & K_{12}K_{12} + K_{22}K_{22} \end{bmatrix} &= \begin{bmatrix} K_{11} & K_{21} \\ K_{12} & K_{22} \end{bmatrix} \begin{bmatrix} K_{11} & K_{12} \\ K_{21} & K_{22} \end{bmatrix} = K^T K \\ \begin{bmatrix} K_{11} & K_{21} \\ K_{12} & K_{22} \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} &= K^T g\end{aligned}\tag{3.14}$$

Assim sendo, o mínimo da função pode ser escrito dessa forma:

$$K^T K f = K^T g\tag{3.15}$$

A solução da equação (3.15) é dada pela equação (3.16) a seguir:

$$f = (K^T K)^{-1} K^T g \quad (3.16)$$

Uma comparação entre os valores de f gerados pelas equações (2.14) e (3.16) é mostrada na tabela a seguir.

Tabela 3.1 – Comparação entre os valores da matriz f obtida pela equação (2.14) e pela equação (3.16)

y	f_{exato}	f_{mq}
1,0000	1,0000	0,4472
1,4444	0,6923	1,2859
1,8889	0,5294	0,6741
2,3333	0,4286	0,1029
2,7778	0,3600	0,1592
3,2222	0,3103	0,4287
3,6667	0,2727	2,5878
4,1111	0,2432	-0,3191
4,5556	0,2195	-1,5232
5,0000	0,2000	-2,3907

A figura (3.1) a seguir foi gerada computacionalmente através da equação (3.16) descrita anteriormente:

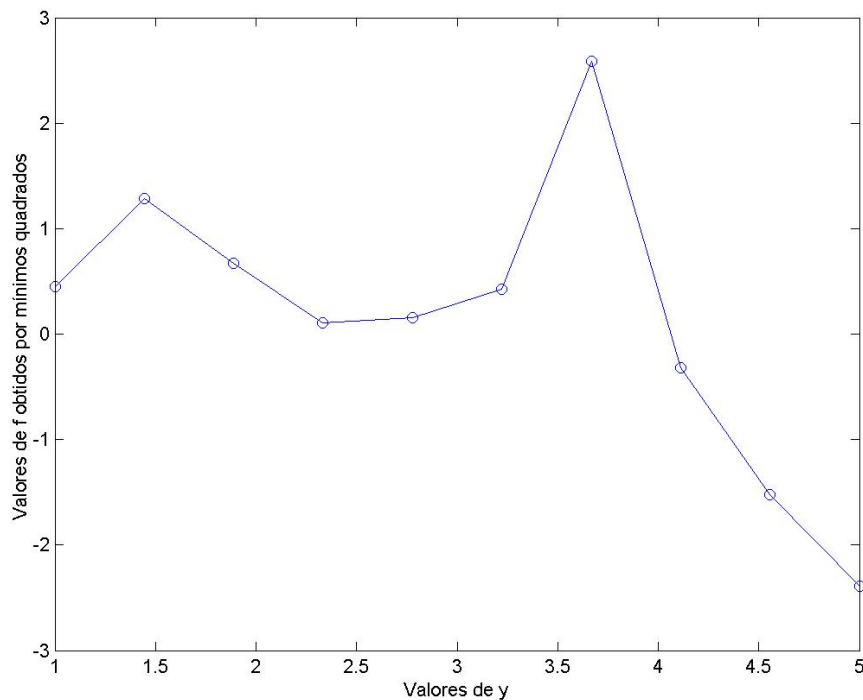


Figura 3.1 – Valores da matriz “f” obtida computacionalmente através da equação $f = (K^T K)^{-1} K^T g$

Comparando-se a figura (3.1) com a figura (2.2), pode-se observar que essa solução também não é aceitável. O índice de condicionamento de $K^T K$ indica que o uso da equação (3.16) não é confiável:

$$\text{Cond}(K^T K) = 5,0995 \times 10^{17} \quad (3.17)$$

3.5 – O Problema Linear dos Mínimos Quadrados

Considerando-se que tanto K quanto g podem depender de um parâmetro t , procura-se encontrar os melhores valores para f_1, f_2, \dots, f_n quando os valores de t_1, t_2, \dots, t_n forem levados em conta. Essa situação pode ser representada matematicamente por:

$$\begin{aligned} f_1 k_1(t_1) + f_2 k_2(t_1) + \dots + f_n k_n(t_1) &= g(t_1) \\ f_1 k_1(t_2) + f_2 k_2(t_2) + \dots + f_n k_n(t_2) &= g(t_2) \\ &\dots \\ f_1 k_1(t_m) + f_2 k_2(t_m) + \dots + f_n k_n(t_m) &= g(t_m) \end{aligned} \quad (3.18)$$

Nas equações acima, $k_1(t)$, $k_2(t)$, ..., $k_n(t)$ são funções apropriadamente definidas para cada situação particular. Definindo-se $k_{ij} = k_i(t_j)$ e $g_i = g(t_i)$, o sistema de equações (3.18) pode ser representado na forma linear $Kf = g$. Esse sistema geralmente não possui solução única e sua solução deve ser encontrada de tal forma que $\|r\|_2^2 = \|Kf - g\|_2^2$ seja mínima. Esse é o chamado **problema linear dos mínimos quadrados**.

A solução que gera o resíduo mínimo deve pertencer obrigatoriamente ao subespaço $N(K^T)$, caso contrário não será mínimo. Isso significa que:

$$K^T Kf = K^T g \quad (3.19)$$

A solução da equação (3.19) é dada pela equação (3.20):

$$f = (K^T K)^{-1} K^T g \quad (3.20)$$

A matriz $(K^T K)^{-1} K^T$ é conhecida como **matriz de Moore-Penrose** ou **pseudo-inversa de K**. Não se deve tentar encontrar a solução para um problema linear de mínimos quadrados diretamente a partir da aplicação da equação (3.20). Em muitos casos é impossível encontrá-la a partir de soluções analíticas. Caso a matriz K relacionada ao problema seja de *rank* deficiente, a matriz $K^T K$ também será.

A solução de sistemas mal-condicionados ou com *rank* deficiente não pode ser obtida por inversão de matrizes ou pelo método dos mínimos quadrados. Uma solução é remover a **singularidade** (ponto de um conjunto que não é “bem comportado”) que aparece quando se inverte a matriz K . A remoção da singularidade é chamada **regularização**.

3.6 – Exemplo Matemático de Ajuste a uma Reta

Considere uma reta do tipo $y = a + bx$, onde a e b são os parâmetros a serem determinados. A intenção é minimizar a distância de um conjunto de pontos (x_i, y_i) da tabela a cada ponto $(x_i, a + bx_i)$ da reta, conforme figura a seguir (Souza, 2006):

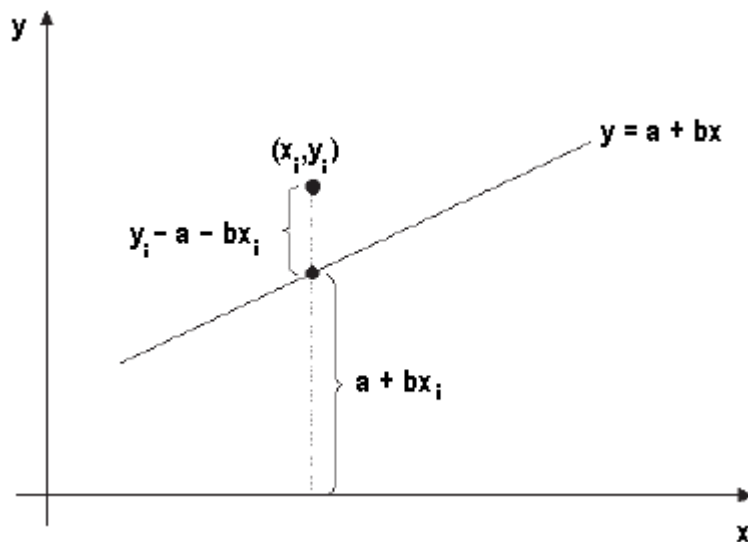


Figura 3.2 – Distância de um ponto (x_i, y_i) à reta $y = a + bx$ (Souza, 2006)

A distância entre o ponto da reta e o ponto da tabela é dada por:

$$|y_i - a - bx_i| \quad (3.21)$$

A soma dos quadrados das distâncias de todos os pontos é dada por:

$$q = \sum_{i=1}^n (y_i - a - bx_i)^2 \quad (3.22)$$

Para a função descrita pela equação (3.22), os valores que podem ser seus mínimos são aqueles onde as derivadas parciais de q em relação aos parâmetros a e b são nulas, ou seja:

$$\begin{aligned} \frac{\partial q}{\partial a} &= -2 \sum_{i=1}^n (y_i - a - bx_i) = 0 \\ \frac{\partial q}{\partial b} &= -2 \sum_{i=1}^n x_i (y_i - a - bx_i) = 0 \end{aligned} \quad (3.23)$$

Como:

$$\sum_{i=1}^n (y_i - a - bx_i) = \sum_{i=1}^n y_i - \sum_{i=1}^n a - \sum_{i=1}^n bx_i = \sum_{i=1}^n y_i - na - \left(\sum_{i=1}^n x_i \right) b \quad (3.24)$$

$$\sum_{i=1}^n x_i (y_i - a - bx_i) = \sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i \right) a - \left(\sum_{i=1}^n x_i^2 \right) b \quad (3.25)$$

Tem-se:

$$\frac{\partial q}{\partial a} = -2 \left[\sum_{i=1}^n y_i - na - \left(\sum_{i=1}^n x_i \right) b \right] = 0 \quad (3.26)$$

$$\frac{\partial q}{\partial b} = -2 \left[\sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i \right) a - \left(\sum_{i=1}^n x_i^2 \right) b \right] = 0 \quad (3.27)$$

As equações (3.26) e (3.27) são utilizadas para se escrever um sistema de equações do problema em questão, cujas incógnitas são os parâmetros a e b da equação inicial $y = a + bx$:

$$\begin{cases} na + \left(\sum_{i=1}^n x_i \right) b = \sum_{i=1}^n y_i \\ \left(\sum_{i=1}^n x_i \right) a + \left(\sum_{i=1}^n x_i^2 \right) b = \sum_{i=1}^n x_i y_i \end{cases} \quad (3.28)$$

A título de ilustração, considera-se a tabela a seguir, que mostra os pontos (x_i, y_i) . A seguir é utilizado o método dos mínimos quadrados para a determinação da equação da reta que melhor se ajusta a esses pontos.

Tabela 3.2 – Pontos (x_i, y_i) para determinação da reta $y = a + bx$

x_i	-1,0	-0,1	0,2	1,0
y_i	1,000	1,099	0,808	1,000

Aplicando-se o sistema de equações (3.28) com $n = 4$ pontos:

$$\sum_{i=1}^n x_i = -1 - 0,1 + 0,2 + 1 = 0,1$$

$$\sum_{i=1}^n x_i^2 = (-1)^2 + (-0,1)^2 + 0,2^2 + 1^2 = 2,05$$

$$\sum_{i=1}^n y_i = 1^2 + 1,099^2 + 0,808^2 + 1^2 = 3,907$$

$$\sum_{i=1}^n x_i y_i = (-1 \times 1) + (-0,1 \times 1,099) + (0,2 \times 0,808) + (1 \times 1) = 0,0517$$

$$\begin{cases} 4,0a + 0,10b = 3,9070 \\ 0,1a + 2,05b = 0,0517 \end{cases}$$

Resolvendo-se o sistema para a e b tem-se:

$$\begin{cases} a = 0,9773 \\ b = -0,0224 \end{cases}$$

Assim sendo, a equação da reta que melhor se ajusta à tabela de pontos é:

$$y = 0,9773 - 0,0224x$$

O método dos mínimos quadrados também pode ser utilizado em alguns casos com outros tipos de função com solução analítica, como exponencial, hipérbole e outros.

3.7 – Qualidade do Ajuste

Verifica-se a qualidade do ajuste feito em função do coeficiente de determinação r^2 dado por:

$$r^2 = \frac{\sum_{i=1}^n (a + bx_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (3.29)$$

$$\text{em que } \bar{y} = \frac{1}{n} \left(\sum_{i=1}^n y_i \right)$$

Quanto mais próximo o valor de r^2 estiver da unidade, melhor é a qualidade do ajuste. Esse coeficiente de determinação é uma medida da proporção da variação total dos dados em torno da média dos mesmos. O numerador representa a soma dos quadrados dos

desvios de cada ponto da reta de ajuste ao ponto médio \bar{y} dos pontos dados, enquanto que o denominador representa a soma dos quadrados dos desvios de cada ponto dado ao ponto médio \bar{y} .

Como:

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - a - bx_i)^2 + \sum_{i=1}^n (a + bx_i - \bar{y})^2 \quad (3.30)$$

Assim:

$$r^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2 - \sum_{i=1}^n (y_i - a - bx_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3.31)$$

Como:

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - 2\bar{y} \sum_{i=1}^n y_i + n \sum_{i=1}^n \bar{y}^2 = \sum_{i=1}^n y_i^2 - \frac{1}{n} \left(\sum_{i=1}^n y_i \right)^2 \quad (3.32)$$

Simplifica-se o coeficiente de determinação em:

$$r^2 = 1 - \frac{\sum_{i=1}^n (y_i - a - bx_i)^2}{\sum_{i=1}^n y_i^2 - \frac{1}{n} \left(\sum_{i=1}^n y_i \right)^2} \quad (3.33)$$

3.8 – Conclusão

O método dos mínimos quadrados é uma eficiente técnica matemática que pode ser empregada na solução de problemas em ciências, uma vez que nem sempre a técnica de inversão de matrizes retorna valores confiáveis em sua resolução. Entretanto, para matrizes com índice de condicionamento muito grande, mesmo esse método pode não ser adequado, o que exige o uso de técnicas mais avançadas, como a **regularização**.

CAPÍTULO IV

A REGULARIZAÇÃO DE TIKHONOV

4.1 – Introdução

Uma técnica muito eficiente de resolução de problemas consiste em adicionar informações extras a eles. Uma resolução de problemas com o método dos mínimos quadrados foi mostrada com a equação (3.7) utilizada para se encontrar o mínimo do funcional:

$$\Phi(f) = \|Kf - g\|_2^2 = (Kf - g)^T (Kf - g) \quad (3.7)$$

Pode-se adicionar informações extras a esse problema, transformando a equação (3.7) na equação (4.1) a seguir:

$$\Phi(f) = \|Kf - g\|_2^2 + \lambda \|Lf\|_2^2 \quad (4.1)$$

em que λ é chamado **parâmetro de regularização** e representa um valor a ser encontrado, enquanto L é um **operador** geralmente da forma mostrada na equação (4.2) a seguir:

$$\|Lf\|_2^2 = a_0 \|f - \hat{f}\|_2^2 + a_1 \left\| \frac{df}{dx} \right\|_2^2 + a_2 \left\| \frac{d^2 f}{dx^2} \right\|_2^2 \quad (4.2)$$

em que a_0 , a_1 e a_2 são valores constantes e iguais a 0 ou 1, dependendo de condições a serem impostas. A função \hat{f} é um valor inicial estimado para a solução.

A regularização é necessária na resolução de problemas inversos devido principalmente aos erros embutidos nos dados experimentais e aos de arredondamento. Praticamente todos os métodos de regularização que buscam soluções estáveis para problemas inversos envolvem uma ligação entre o tamanho da solução regularizada e a qualidade do ajuste dos dados. O que distingue os vários métodos de regularização é a maneira como eles medem e como decidem a melhor ligação entre as duas quantidades.

O melhor valor para o parâmetro de regularização λ é encontrado a partir da “**curva L**”, um gráfico paramétrico montado em escala logarítmica, onde as abscissas correspondem à norma do vetor de erros nos dados calculados e as ordenadas correspondem à norma do vetor de derivadas dos parâmetros do modelo estimado (Hansen, 1999). O “joelho” da curva representa a situação de equilíbrio entre as influências desses dois termos na função, correspondendo ao melhor modelo matemático. Esse parâmetro otimizado, selecionado no ponto da curva L onde a curvatura é máxima, proporcionará uma suavização ideal nos dados fornecendo um modelo bem próximo do que seria o verdadeiro.

A seguir, a equação (4.2) será dividida em 3 partes e analisada separadamente.

4.2 – A Primeira Restrição

Considere a primeira parte da equação (4.2) associada à constante a_0 :

$$\|Lf\|_2^2 = a_0 \|f - \hat{f}\|_2^2 \quad (4.3)$$

Essa equação pode ser reescrita da forma:

$$\|Lf\|_2^2 = a_0 \|f - \hat{f}\|_2^2 = a_0 \sum_{i=1}^n (f_i - \hat{f}_i)^2 \quad (4.4)$$

O mínimo da equação (4.4) será:

$$\frac{d \left[a_0 \sum_{i=1}^n (f_i - \hat{f}_i)^2 \right]}{df_k} = 2a_0 (f_k - \hat{f}_k) = 0, \quad (4.5)$$

em que k é um valor diferente de i

Deve-se encontrar o valor de f que minimize o funcional descrito a seguir:

$$\Phi(f) = \|Kf - g\|_2^2 + \lambda a_0 \|f - \hat{f}\|_2^2 \quad (4.6)$$

Se $a_0 = 1/2$, a solução será equivalente a resolver a equação (4.7) a seguir:

$$(K^T K + \lambda I)f = K^T g + \lambda \hat{f} \quad (4.7)$$

em que I representa uma **matriz identidade**. Se $\lambda = 0$, tem-se a solução por mínimos quadrados, descrita pela equação (3.45).

4.3 – A Segunda Restrição

Considere a definição de derivada primeira dada pela equação (4.8) a seguir:

$$\frac{df}{dx} = \frac{f(x+h) - f(x)}{h} \quad (4.8)$$

Para analisar a segunda restrição, considere inicialmente apenas 4 pontos na quadratura de Gauss (técnica de integração numérica de Gauss). Fazendo $a = y_1, y_2$ e y_3 , e $b = y_4$, tem-se:

$$\|Lf\|_2^2 = a_1 \left\| \frac{df}{dx} \right\|_2^2 = \frac{a_1}{h^2} [(f_2 - f_1)^2 + (f_3 - f_2)^2 + (f_4 - f_3)^2] \quad (4.9)$$

Calculando-se as derivadas parciais, tem-se:

$$\begin{aligned} \frac{\partial \|Lf\|_2^2}{\partial f_1} &= \frac{a_1}{h^2} [2(f_2 - f_1)(-1)] = 2 \frac{a_1}{h^2} (f_1 - f_2) = 0 \\ \frac{\partial \|Lf\|_2^2}{\partial f_2} &= \frac{a_1}{h^2} [2(f_2 - f_1) + 2(f_3 - f_2)(-1)] = 2 \frac{a_1}{h^2} (-f_1 + 2f_2 - f_3) = 0 \\ \frac{\partial \|Lf\|_2^2}{\partial f_3} &= \frac{a_1}{h^2} [2(f_3 - f_2) + 2(f_4 - f_3)(-1)] = 2 \frac{a_1}{h^2} (-f_2 + 2f_3 - f_4) = 0 \end{aligned} \quad (4.10)$$

Considerando-se os pontos na quadratura de Gauss onde se faz $a = y_1, y_2, y_3$ e y_4, y_5, y_6 , tem-se:

$$\|Lf\|_2^2 = a_2 \left\| \frac{d^2 f}{dx^2} \right\|_2^2 = a_2 \left[(f_3 - 2f_2 + f_1)^2 + (f_4 - 2f_3 + f_2)^2 + (f_5 - 2f_4 + f_3)^2 + (f_6 - 2f_5 + f_4)^2 \right] \quad (4.16)$$

Derivando-se a equação (4.16), obtém-se:

$$\begin{aligned} \frac{\partial \|Lf\|_2^2}{\partial f_1} &= 2a_2 [(f_3 - 2f_2 + f_1)] = 2a_2 (f_1 - 2f_2 + f_3) = 0 \\ \frac{\partial \|Lf\|_2^2}{\partial f_2} &= 2a_2 [(f_3 - 2f_2 + f_1)(-2) + (f_4 - 2f_3 + f_2)] = \\ &= 2a_2 (-2f_1 + 4f_2 - 2f_3 + f_2 - 2f_3 + f_4) = 2a_2 (-2f_1 + 5f_2 - 4f_3 + f_4) = 0 \\ \frac{\partial \|Lf\|_2^2}{\partial f_3} &= 2a_2 [(f_3 - 2f_2 + f_1) + (f_4 - 2f_3 + f_2)(-2) + (f_5 - 2f_4 + f_3)] = \\ &= 2a_2 (f_1 - 2f_2 + f_3 - 2f_2 + 4f_3 - 2f_4 + f_3 - 2f_4 + f_5) = \\ &= 2a_2 (f_1 - 4f_2 + 6f_3 - 4f_4 + f_5) = 0 \\ \frac{\partial \|Lf\|_2^2}{\partial f_4} &= 2a_2 [(f_4 - 2f_3 + f_2) + (f_5 - 2f_4 + f_3)(-2) + (f_6 - 2f_5 + f_4)] = \\ &= 2a_2 (f_2 - 2f_3 + f_4 - 2f_3 + 4f_4 - 2f_5 + f_4 - 2f_5 + f_6) = \\ &= 2a_2 (f_2 - 4f_3 + 6f_4 - 4f_5 + f_6) = 0 \\ \frac{\partial \|Lf\|_2^2}{\partial f_5} &= 2a_2 [(f_5 - 2f_4 + f_3) + (f_6 - 2f_5 + f_4)(-2)] = \\ &= 2a_2 (f_3 - 2f_4 + f_5 - 2f_4 + 4f_5 - 2f_6) = 2a_2 (f_3 - 4f_4 + 5f_5 - 2f_6) = 0 \\ \frac{\partial \|Lf\|_2^2}{\partial f_6} &= 2a_2 [(f_6 - 2f_5 + f_4)] = 2a_2 (f_4 - 2f_5 + f_6) \end{aligned} \quad (4.17)$$

As equações (4.17) podem ser agrupadas na forma de matrizes:

$$\int_a^b K(x,y)f(y)dy = g(x) = \int_a^b \left(\frac{1}{x+y} \right) \frac{1}{y} dy = \frac{1}{x} \ln \left(\frac{1+x/a}{1+x/b} \right), \quad (2.14)$$

em que $a \leq y \leq b$

Portanto tem-se:

$$\begin{aligned} K(x,y) &= \left(\frac{1}{x+y} \right) \\ f(y) &= \frac{1}{y} \\ g(x) &= \frac{1}{x} \ln \left(\frac{1+x/a}{1+x/b} \right) \end{aligned} \quad (2.15)$$

A equação (2.14) pode ser representada na forma matricial:

$$Kf = g \quad (2.16)$$

Dados o *kernel* ou operador $K(x,y)$, os limites de integração a e b e a função $g(x)$, obtém-se o valor de $f(x)$ como sendo a solução da equação.

Outro programa desenvolvido pelo **Grupo de Pesquisa em Problemas Inversos do Departamento de Química da UFMG**, com coordenação do professor e pesquisador Dr. João Pedro Braga retorna as matrizes K , f e g utilizando o esquema de discretização trapezoidal. Esse algoritmo transforma uma equação integral em equação linear, ou seja:

$$\int_a^b K(x,y)f(y)dy = g(x) \Rightarrow Kf = g \quad (2.17)$$

Para o exemplo numérico descrito a seguir, o programa foi configurado de modo que $a = c = 1$, $b = d = 5$ e $n = m = 10$. Ainda no código, procura-se obter a matriz f através de três métodos: inversão da matriz K , mínimos quadrados e regularização de Tikhonov. Os valores encontrados são mostrados na tabela a seguir:

Tabela 4.1 – Comparação entre os valores da matriz f obtida pela equação (2.14) e pela equação (4.22)

y	f_{exato}	f_{Tik}
1,2000	0,8333	0,9066
1,6000	0,6250	0,6212
2,0000	0,5000	0,4708
2,4000	0,4167	0,3853
2,8000	0,3571	0,3332
3,2000	0,3125	0,2992
3,6000	0,2778	0,2755
4,0000	0,2500	0,2579
4,4000	0,2273	0,2442
4,8000	0,2083	0,2329

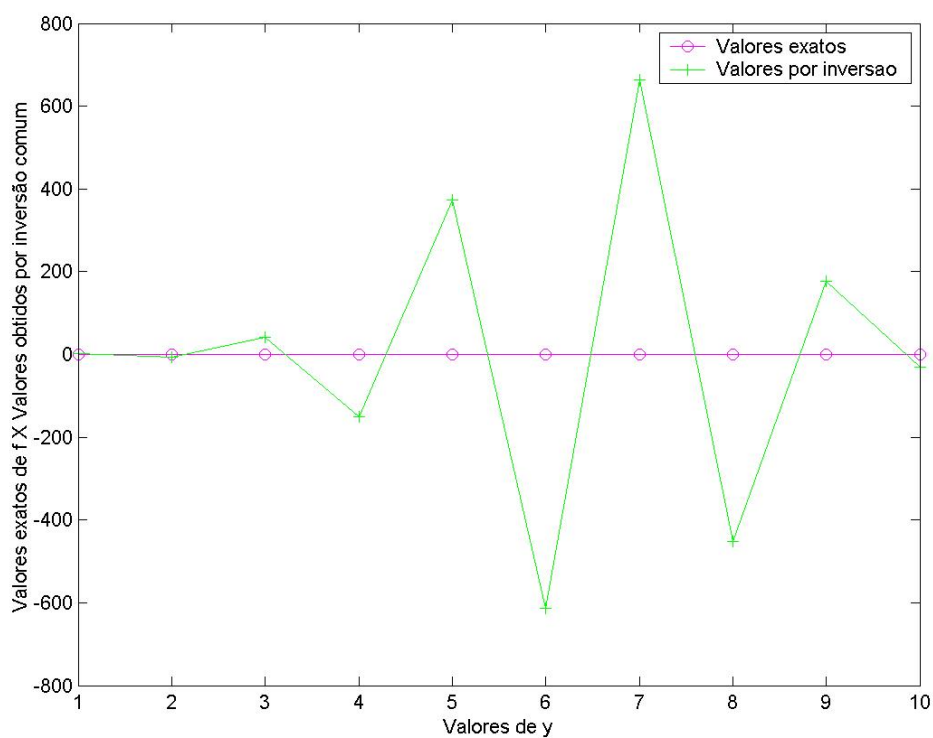


Figura 4.1 – Valores exatos da matriz “ f ” X Valores obtidos por inversão comum de matriz

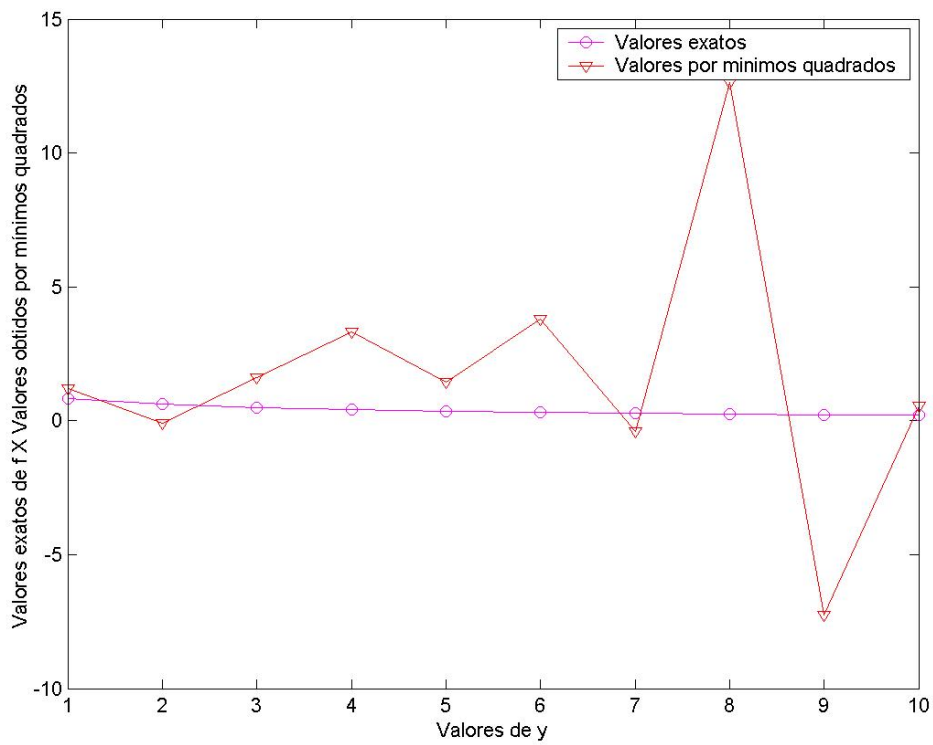


Figura 4.2 – Valores exatos da matriz “f” X Valores obtidos por mínimos quadrados

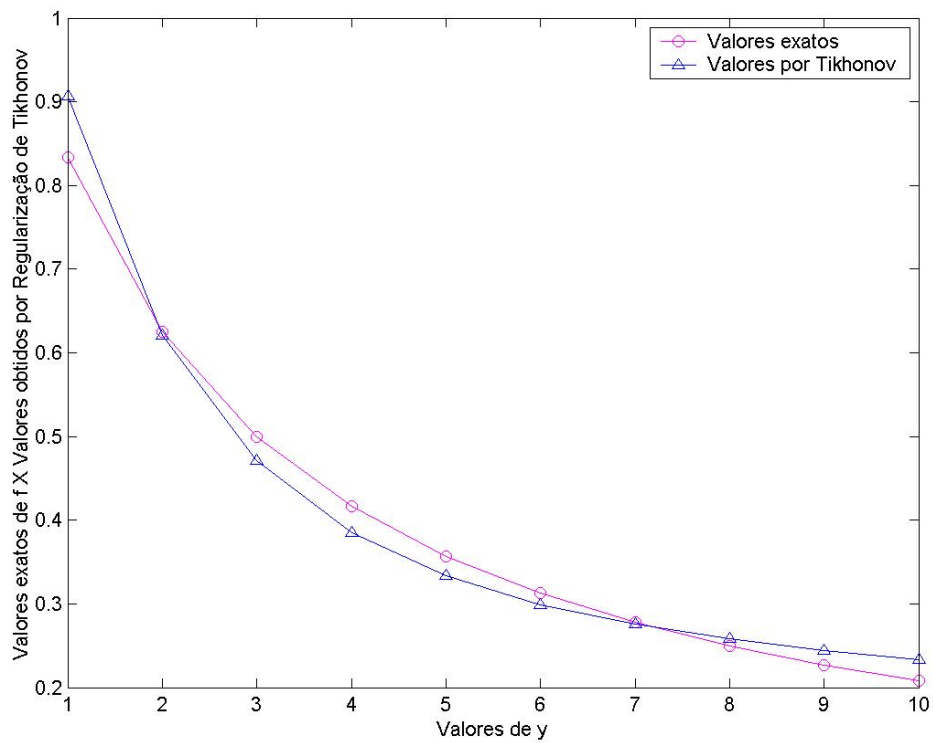


Figura 4.3 – Valores exatos da matriz “f” X Valores obtidos por Tikhonov

Como se pode observar, os valores obtidos através dos dois métodos mais simples (inversão de matrizes e mínimos quadrados) são muito discrepantes dos valores exatos encontrados para a função f , diferentemente daquele encontrado com a Regularização de Tikhonov. Isso pode ser explicado em função do índice de condicionamento da matriz ser muito alto:

$$\text{Cond}(K) = 3,9644 \times 10^{15} \quad (4.23)$$

4.6 – Conclusão

Quando da resolução de sistemas de equações, a inserção de dados adicionais pode fazer com que se consigam soluções com melhores aproximações do que nos métodos de inversão de matrizes e/ou dos mínimos quadrados. Na regularização de Tikhonov, acrescenta-se o parâmetro de regularização e um operador composto de três restrições. Utilizando-se informações adequadas, esse método de regularização é capaz de fornecer excelentes resultados na resolução desses sistemas.

CAPÍTULO V

O MÉTODO DE MONTE CARLO

5.1 – Introdução

Os **métodos numéricos** são importantes ferramentas no desenvolvimento das ciências teóricas e experimentais. O aumento da sofisticação das necessidades de conhecimentos científicos obriga o uso de métodos matemáticos menos sutis ou tradicionais. Isso acontece em áreas como na previsão de taxas de reações químicas, no comportamento de reatores nucleares e no conhecimento da energia de núcleos atômicos a partir de informações fundamentais, entre outras.

O **Método de Monte Carlo** é muito usado em ciências físicas por algumas razões (Kalos, 2007):

- permite a transformação rápida e confiável de modelos estocásticos ou naturais em código de computador;
- são métodos numéricos poderosos para se tratar problemas multidimensionais derivados de sistemas físicos importantes;
- muitos métodos naturalmente permitem o uso de computadores paralelos modernos de maneira eficiente.

Esse método desempenha um importante papel no cenário científico. Por exemplo, a previsão do comportamento de um detector de partículas de alta energia é mais fácil e precisa se analisada por esse método, onde são simulados os processos estocásticos da criação de partículas em um alvo, o transporte dessas partículas, o decaimento de umas em outras e as suas interações finais no processo de detecção. Tal cálculo é formidável e pode ser feito em computadores modernos. Uma importante característica do método de Monte Carlo é que ele oferece procedimentos naturais e eficientes para problemas numéricos em muitas dimensões, o que é impossível para métodos numéricos tradicionais (Kalos, 2007).

O método de Monte Carlo usa algoritmos para resolver vários tipos de problemas computacionais, a partir de números aleatórios ou pseudo-aleatórios desde 1953 (Amar, 2006). Suas aplicações variam da biologia à física quântica, passando por áreas como computação de imagens, arquitetura e economia, entre outras. Como a quantidade de cálculos resolvidos usualmente é muito grande, é altamente recomendável o **uso de computadores** para a aplicação do método.

Enquanto métodos determinísticos podem ser empregados com muita eficiência em resolução de funções matemáticas com uma variável, para funções de muitas variáveis os métodos estocásticos são mais aconselhados, não havendo restrições de dimensão do problema numérico. Por exemplo, para integrar numericamente uma função de um vetor N -dimensional com 10 pontos em cada dimensão, seria necessária a avaliação de 10^{100} pontos, número muito grande de cálculos mesmo para um computador. O método de Monte Carlo proporciona uma maneira muito mais rápida de efetuar tais operações: se a função for bem comportada, a solução pode ser estimada pela seleção aleatória de alguns pontos no espaço N -dimensional, sendo então feita uma média dos valores da função nesses pontos.

Com esse método é possível se obter soluções aproximadas para uma grande variedade de problemas matemáticos, através de amostragens estatísticas feitas principalmente no computador. Ele pode ser aplicado, inclusive, em problemas sem qualquer teor probabilístico. Porém, entre todos os métodos numéricos que utilizam avaliações de n pontos em espaço m -dimensional para produzir uma solução aproximada, o método de Monte Carlo apresenta erro absoluto de estimativa que diminui segundo $n^{-1/2}$, enquanto que, na ausência de estruturas especiais, todos os outros têm erros que diminuem, no máximo, segundo $n^{-1/m}$. Essa propriedade confere ao método de Monte Carlo uma considerável eficiência computacional com o aumento de m . Isso significa que, enquanto a solução exata para um problema combinatorial com m elementos tem frequentemente custo computacional que cresce exponencialmente ou superexponencialmente com o valor de m , o método de Monte Carlo frequentemente proporciona uma solução estimada com erro tolerável a um custo que aumenta não mais do que como um polinômio em m (Fishman, 1996).

Para ilustrar a importância do uso dos Métodos de Monte Carlo em sistemas com muitos graus de liberdade ou com uma quantidade muito grande de partículas, considere a função partição clássica para um gás com A átomos na temperatura $1/\beta$ interagindo em um potencial v , o que pode ser descrito pela equação a seguir:

$$Z = \int d^3 r_1 \dots d^3 r_A \exp \left[-\beta \sum_{i < j} v(r_{ij}) \right] \quad (5.1)$$

A avaliação direta de uma integral como essa seria extremamente difícil de ser efetuada, exceto para valores muito pequenos de A . Em um sistema com modestos 20 átomos ($A = 20$), se cada coordenada tiver 10 valores diferentes, em um computador capaz de realizar 10^7 avaliações por segundo, esses cálculos consumiriam 10^{53} segundos, mais que 10^{34} vezes a idade do universo (Koonin & Meredith, 1990).

5.2 – Descrição do Método de Monte Carlo

O Método de Monte Carlo é muito eficiente para efetuar avaliações de integrais multidimensionais. Considere a seguinte integral para avaliação de uma função f qualquer:

$$I = \int_0^1 f(x) dx \quad (5.2)$$

Uma alternativa para avaliação do valor de I é considerá-la como uma média dos valores de f no intervalo $[0,1]$. Para valores de x_i escolhidos aleatoriamente, com igual probabilidade dentro do intervalo $[0,1]$, é possível escrever:

$$I = \int_0^1 f(x) dx \approx \frac{1}{N} \sum_{i=1}^N f(x_i) \quad (5.3)$$

Para se estimar a incerteza associada a essa fórmula, considere $f_i \equiv f(x_i)$ como uma variável aleatória. Pelas leis usuais da estatística tem-se o cálculo da variância σ^2 :

$$\sigma_I^2 \approx \frac{1}{N} \sigma_f^2 = \frac{1}{N} \left[\frac{1}{N} \sum_{i=1}^N f_i^2 - \left(\frac{1}{N} \sum_{i=1}^N f_i \right)^2 \right] \quad (5.4)$$

Baseado na equação de cálculo da variância pode-se chegar a dois importantes aspectos dos Métodos de Monte Carlo. Primeiro, a incerteza da estimativa da integral σ_f decresce com $N^{-1/2}$. Assim, quanto maior o número de pontos utilizado, menor a incerteza. Segundo, a precisão é maior na medida em que o valor de σ_f é menor.

5.3 – Exemplos do Uso do Método de Monte Carlo

Cálculo do Número Pi (π)

O número π , possivelmente a mais antiga constante matemática que se conhece, é obtido pela razão entre a circunferência de um círculo qualquer e o seu diâmetro. O uso da letra grega para representação dessa constante só foi introduzido em 1706 por William Jones (1657-1749). O valor exato de π sempre foi interesse dos matemáticos. Arquimedes de Siracusa (287-212 a.C.) chegou ao valor 3,142857...

No século XVIII ficou provado que essa constante é um número irracional, o que, em termos práticos, significa que o número de casas decimais de π pode ser infinito. No século XIX foi demonstrado que π é um número transcendental, ou seja, não pode ser expresso como sendo a raiz de uma equação algébrica qualquer, com coeficientes racionais. A tabela a seguir apresenta os progressos feitos no cálculo do valor de π ao longo do tempo.

Tabela 5.1 – Valores calculados para π ao longo do tempo

Povo, pessoa ou ferramenta	Ano	Valor ou característica
Babilônia	~2000 a.C.	3 1/8
Egípcios	~2000 a.C.	$(16/9)^2 = 3,1605$
Chineses	~1200 a.C.	3
Antigo Testamento*	~550 a.C.	3
Arquimedes	~300 a.C.	encontra $3 \frac{10}{71} < \pi < 3 \frac{1}{7}$ usa $211875/67441 = 3,14163$
Ptolomeu	~200 d.C.	$377/120 = 3,14166...$
Chung Huing	~300 d.C.	$10^{1/2} = 3,16...$
Wang Fau	263 d.C.	$157/50 = 3,14$
Tsu Chung-Chi	~500 d.C.	$3,1415926 < \pi < 3,1415929$
Aryabhatta	~500	3,1416
Brahmagupta	~600	$10^{1/2}$
Al-Khwarizmi	820	3,1416
Fibonacci	1220	3,141818
Ludolph van Ceulen	1596	Calcula π com até 35 casas decimais
Machin	1706	100 casas decimais
Lambert	1766	Prova que π é irracional
Richter	1855	500 casas decimais
Lindeman	1882	Prova que π é transcendental
Ferguson	1947	808 casas decimais

Tabela 5.1 – Valores calculados para π ao longo do tempo (continuação)

Computador Pegasus	1957	7.840 casas decimais
IBM 7090	1961	100.000 casas decimais
CDC 6600	1967	500.000 casas decimais

* Reis, Volume I, Versículo 7, Parágrafo 23: “Fez também o mar de fundição; era redondo e media dez côvados de uma borda à outra, cinco côvados de altura e trinta de circunferência”. (Circunferência = $2\pi r \rightarrow 30 = 2 \cdot \pi \cdot 5 \rightarrow \pi \sim 3$).

Foram desenvolvidos programas computacionais em MATLAB, Fortran e Microsoft Visual Basic com o intuito de se comprovar o valor de π utilizando-se o Método de Monte Carlo. Observa-se que a precisão do cálculo aumenta na medida em que se estende o número de tentativas, o que já era previsto. Para isso considera-se uma circunferência de raio $R = 1$ cm inscrita em um quadrado de lado $L = 2$ cm.

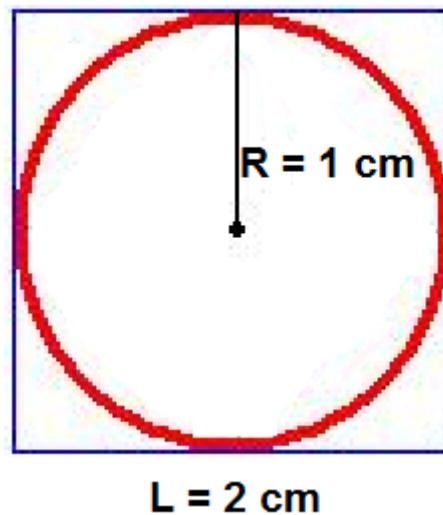


Figura 5.1 – Circunferência de $R = 1$ cm inscrita em um quadrado de $L = 2$ cm

$$A_{circ} = \pi R^2$$

$$A_{quad} = L^2$$

$$\frac{A_{circ}}{A_{quad}} = \frac{\pi R^2}{L^2} \Rightarrow \frac{A_{circ}}{A_{quad}} = \frac{\pi \times 1^2}{2^2} \Rightarrow \frac{A_{circ}}{A_{quad}} = \frac{\pi}{4} \Rightarrow \pi = 4 \frac{A_{circ}}{A_{quad}}$$

Assim, pode-se chegar ao valor de π através da relação matemática entre as áreas dessas 2 figuras, ou seja, $\pi = 4 \times A_{circ}/A_{quad}$.

O gráfico a seguir mostra o valor encontrado para π em função do número de tentativas realizadas pelo Método de Monte Carlo utilizando programas de computador:

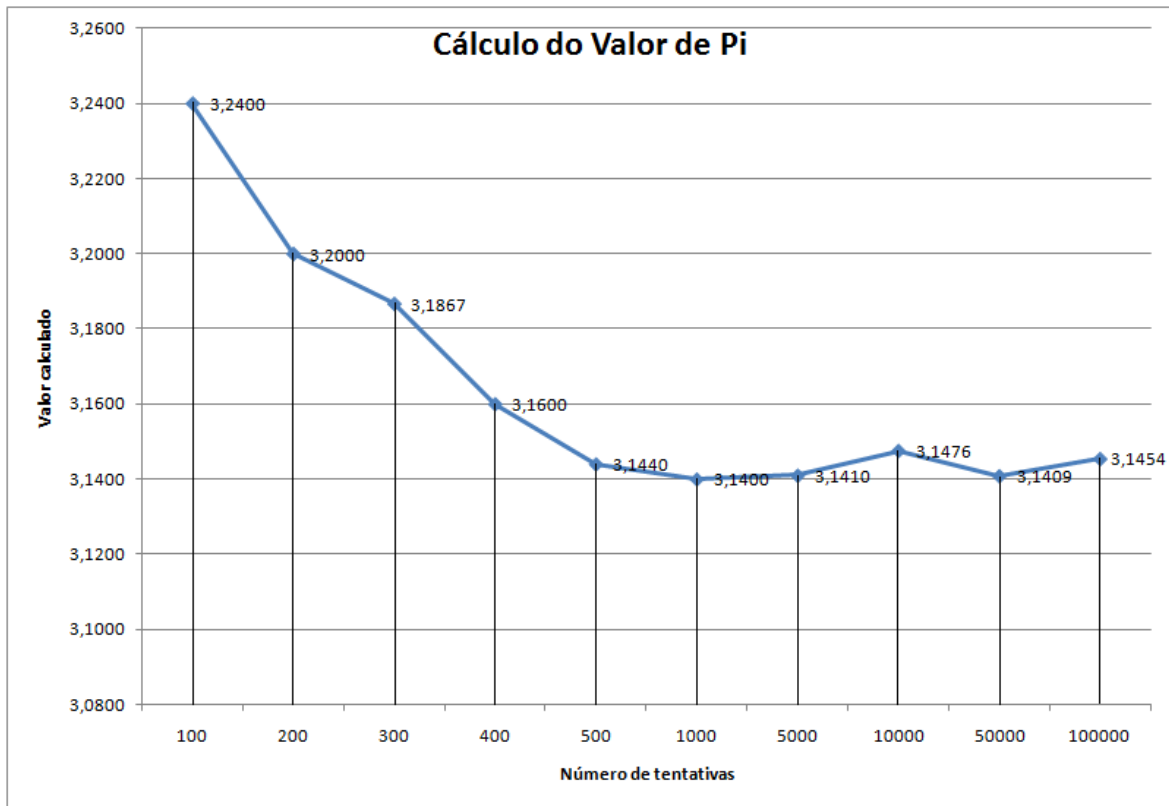


Figura 5.2 – Cálculo de π utilizando o Método de Monte Carlo

Observa-se claramente no gráfico que a precisão no cálculo do valor de π aumenta com o número de tentativas, o que pode ser variado facilmente com o Método de Monte Carlo, independentemente da linguagem utilizada na construção do programa.

Difusão de partículas em uma caixa

Outro exemplo que demonstra o poder do Método de Monte Carlo é o modelo de difusão de partículas em uma caixa. Para comprovação, foi construído um algoritmo nas três linguagens utilizadas no cálculo do valor de π onde uma fonte de partículas é colocada em uma extremidade de uma caixa e um detector de partículas é colocado na extremidade oposta. O comprimento da caixa pode ser variado dentro do programa, assim como o número de partículas emitidas e o número de interações que elas podem sofrer no seu interior. O ângulo de espalhamento é sorteado aleatoriamente pelo programa e o deslocamento se dá em função do cosseno desse ângulo.

Os gráficos a seguir mostram a variação no número de partículas que chega até a outra extremidade da caixa em função da quantidade de partículas emitidas quando se varia o número de interações, o número de partículas e o comprimento da caixa.

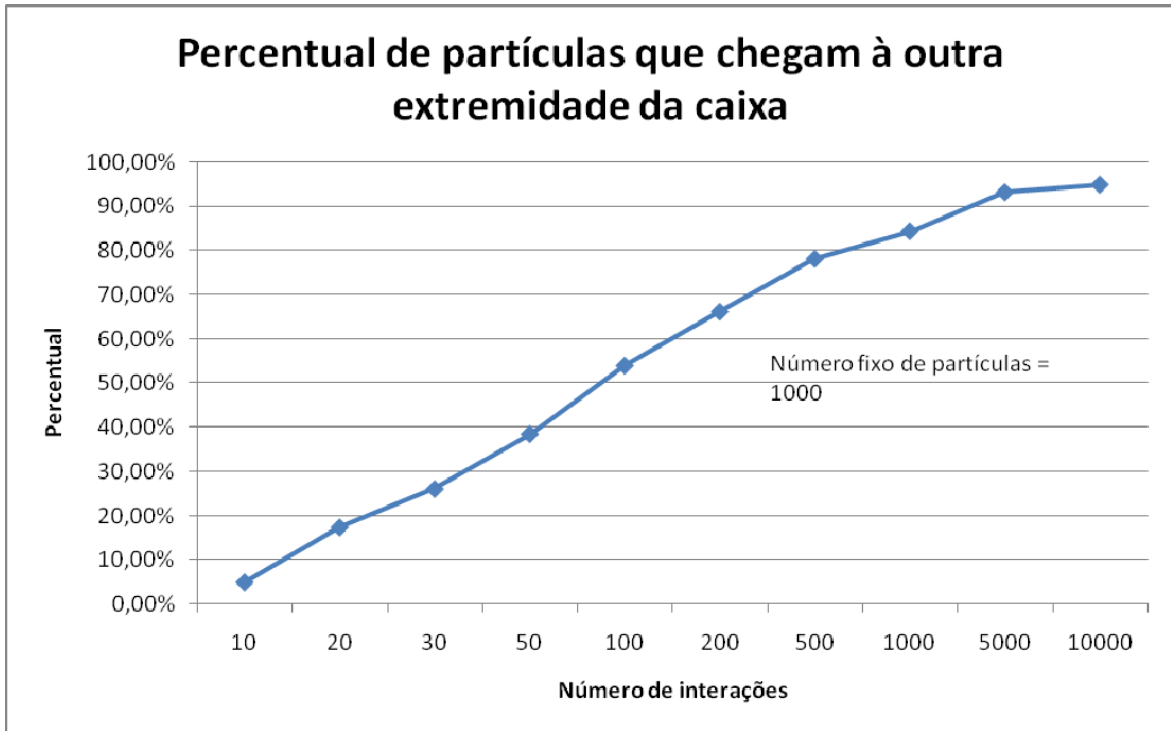


Figura 5.3 – Percentual de partículas que chegam à outra extremidade da caixa para 1000 partículas emitidas pela fonte

Observa-se que, para um número fixo de partículas, o percentual dessas que chega até a outra extremidade da caixa aumenta com a elevação do número de interações, o que já era esperado.

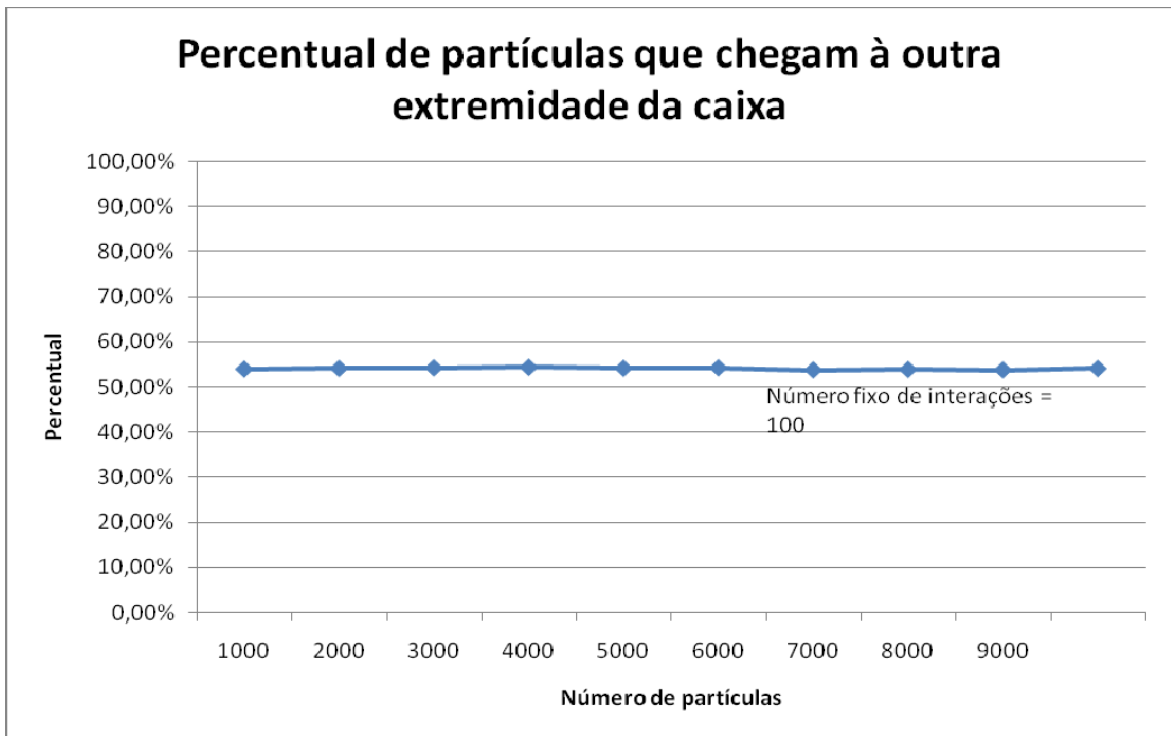


Figura 5.4 – Percentual de partículas que chegam à outra extremidade da caixa para 100 interações de cada partícula

Observa-se que, para um número fixo de interações sofridas individualmente pelas partículas, o percentual dessas que chega até a outra extremidade da caixa é praticamente o mesmo, independentemente do número de partículas emitidas pela fonte.

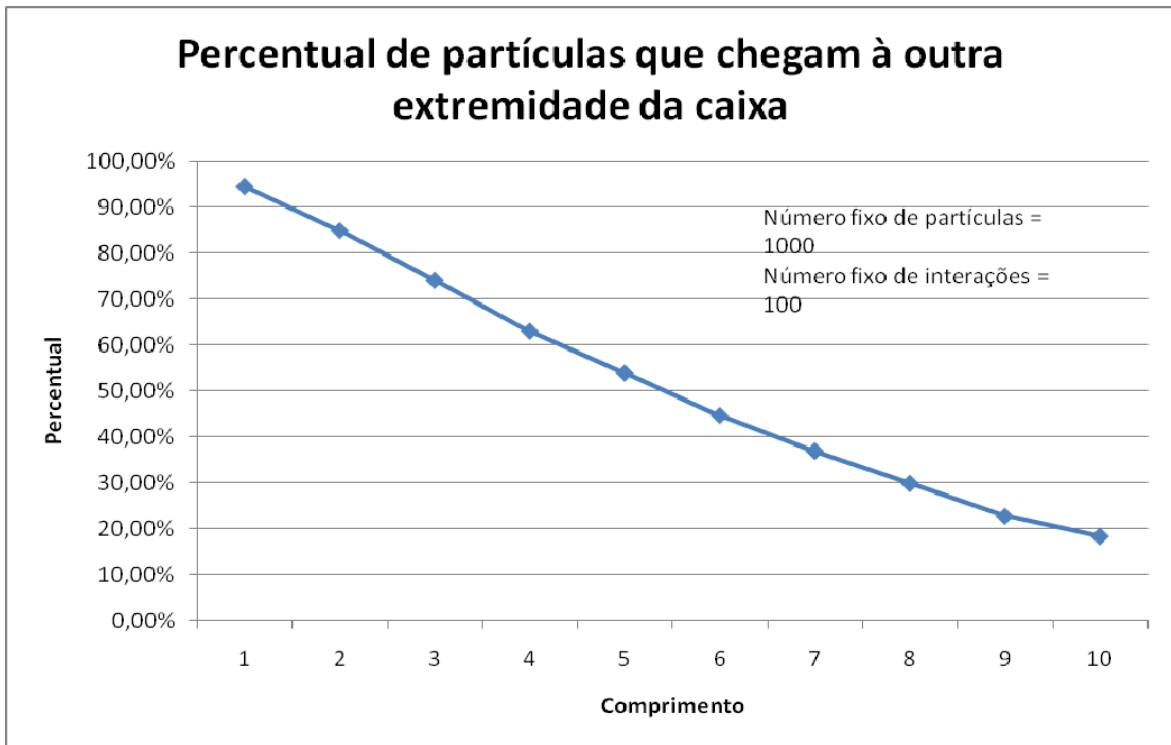


Figura 5.5 – Percentual de partículas que chegam à outra extremidade da caixa para 100 interações de 1000 partículas emitidas pela fonte

Como já era previsto, observa-se que, para um número fixo de partículas e também para um número fixo de interações sofridas por elas, o percentual dessas que chega até a outra extremidade da caixa diminui com o aumento do comprimento da trajetória descrita por elas.

Até esse momento, foram mostradas técnicas de Monte Carlo chamadas “*simple sampling*”, que podem ser utilizadas na resolução de problemas mais simples. Entretanto, em muitos problemas necessita-se de distribuições probabilísticas que reflitam a física apropriada do sistema sob estudo. Nesses casos deve-se utilizar uma técnica mais aprimorada chamada “*importance sampling*”, onde se adiciona dados que tornarão o método mais eficiente, concentrando a distribuição em pontos mais importantes, diminuindo o tempo gasto em pontos de menor importância.

Avaliação de integrais

Considere a integral a seguir, que será avaliada pelo Método de Monte Carlo:

$$\int_0^1 \frac{dx}{1+x^2} = \frac{\pi}{4} = 0,78540 \quad (5.5)$$

Através de programas de computador utilizando o Método de Monte Carlo, é possível se chegar a valores como os descritos nos gráficos a seguir:

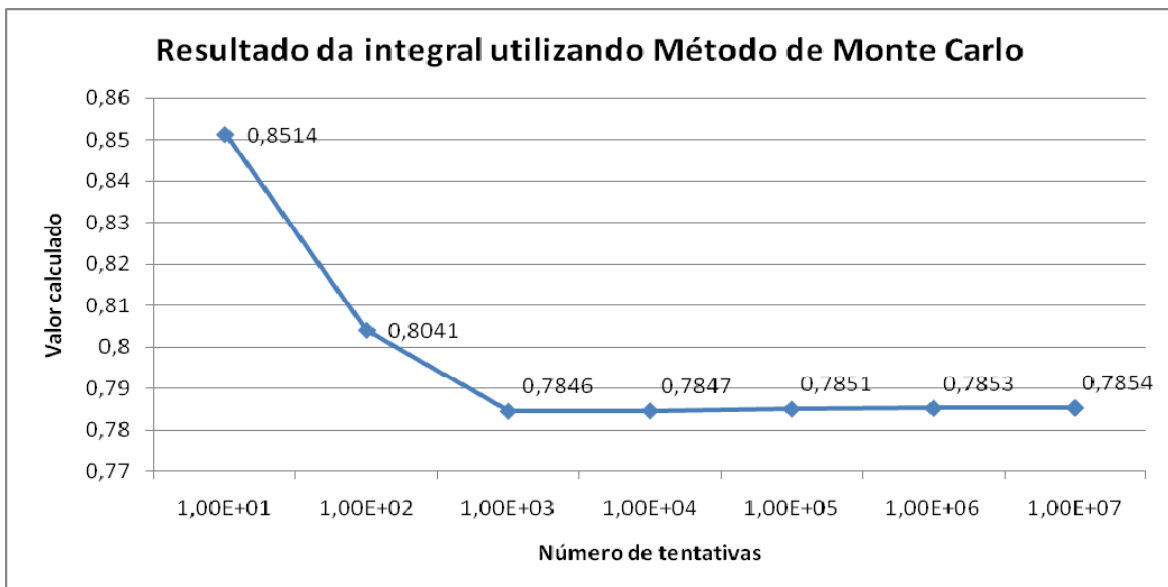


Figura 5.6 – Resultado da integral (5.5) utilizando Método de Monte Carlo

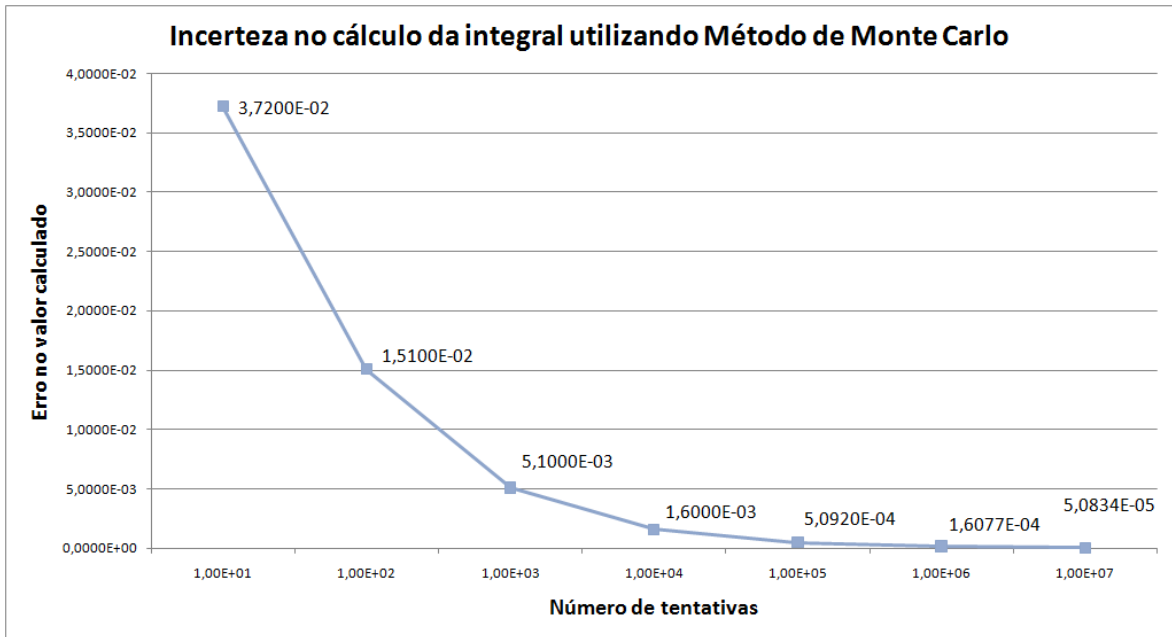


Figura 5.7 – Incerteza no cálculo da integral (5.5) utilizando Método de Monte Carlo

É possível reduzir a variância e melhorar a eficiência desse método. Imagine que o integrando seja multiplicado e dividido por uma função peso $w(x)$ normalizada, de maneira que:

$$\int_0^1 dxw(x) = 1 \quad (5.6)$$

A integral (5.6) pode ser reescrita da seguinte forma:

$$I = \int_0^1 f(x)dx = \int_0^1 dxw(x) \frac{f(x)}{w(x)} \quad (5.7)$$

Fazendo-se a troca de variáveis obtém-se:

$$y(x) = \int_0^x dx'w(x') \quad (5.8)$$

Verifica-se que:

$$\begin{aligned}\frac{dy}{dx} &= w(x) \\ y(x=0) &= 0 \\ y(x=1) &= 1\end{aligned}$$

A integral (5.2) pode ser reescrita da seguinte forma:

$$I = \int_0^1 f(x)dx = \int_0^1 dxw(x) \frac{f(x)}{w(x)} = \int_0^1 dy \frac{f[x(y)]}{w[x(y)]} \approx \frac{1}{N} \sum_{i=1}^N \frac{f[x(y_i)]}{w[x(y_i)]} \quad (5.9)$$

Essa troca de variáveis tem uma razão: escolhendo-se um valor para w que tenha comportamento aproximadamente igual ao de f haverá uma redução da variância. Uma maneira mais geral para se analisar essa mudança de variáveis é imaginar que a distribuição uniforme de pontos em y implica se a distribuição de pontos em x é dada por $dy/dx=w(x)$. Isso significa que os pontos serão concentrados em torno dos valores mais importantes de x onde w e possivelmente f são grandes. Isso faz com que o computador gaste menos tempo no cálculo de valores onde w e f sejam pequenos. Para exemplificar, considere o seguinte valor para a função $w(x)$:

$$w(x) = \frac{1}{3}(4 - 2x) \quad (5.10)$$

Aplicando-se na equação (5.10) a técnica de inversão de variáveis descrita através da equação (5.8) obtém-se:

$$\begin{aligned}y(x) &= \int_0^x dx' w(x') \Rightarrow y(x) = \int_0^x dx \frac{1}{3}(4 - 2x) = \int_0^x \left(\frac{4}{3} - \frac{2}{3}x \right) dx = \\ &= \int_0^x \frac{4}{3} dx - \int_0^x \frac{2}{3} x dx = \frac{4}{3} \int_0^x dx - \frac{2}{3} \int_0^x x dx = \frac{4}{3}x + C_1 - \frac{x^2}{3} + C_2 = \\ &= \frac{1}{3}x(4 - x) + C = \frac{1}{3}x(4 - x)\end{aligned} \quad (5.11)$$

Invertendo-se a equação (5.11) obtém-se:

$$x = 2 - \sqrt{(4 - 3y)} \quad (5.12)$$

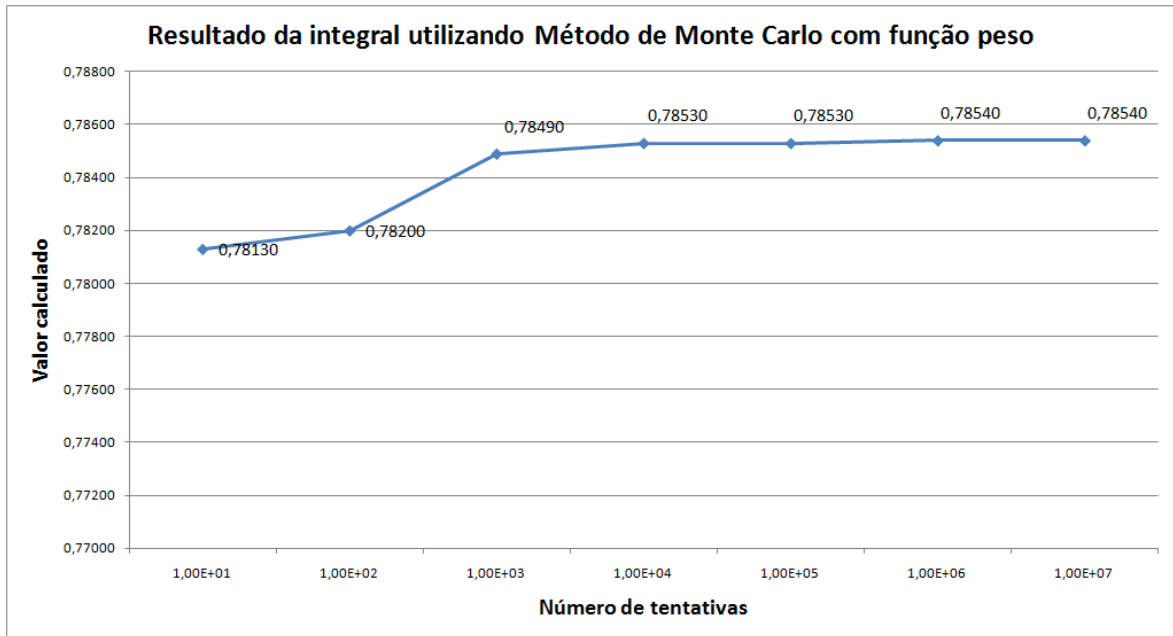


Figura 5.8 – Resultado da integral (5.5) utilizando Método de Monte Carlo com função peso

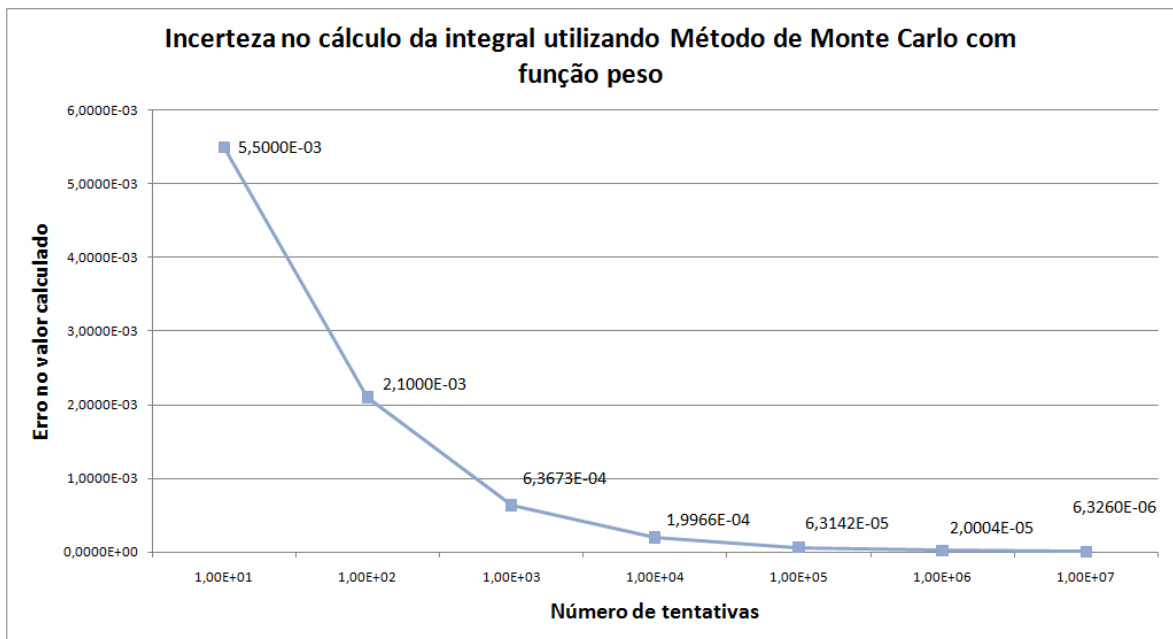


Figura 5.9 – Incerteza no cálculo da integral (5.5) utilizando Método de Monte Carlo com função peso

Os gráficos a seguir mostram uma comparação dos valores obtidos pelo Método de Monte Carlo para determinação do valor da integral (5.5) utilizando-se as duas técnicas demonstradas anteriormente:

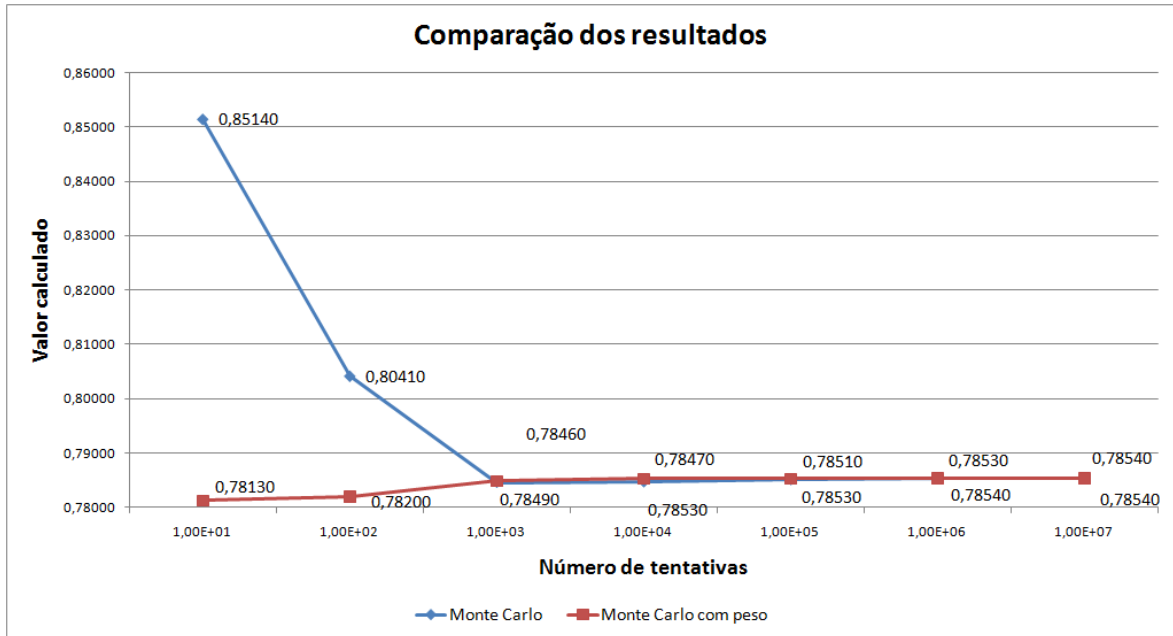


Figura 5.10 – Resultado comparado da integral (5.5) utilizando Método de Monte Carlo

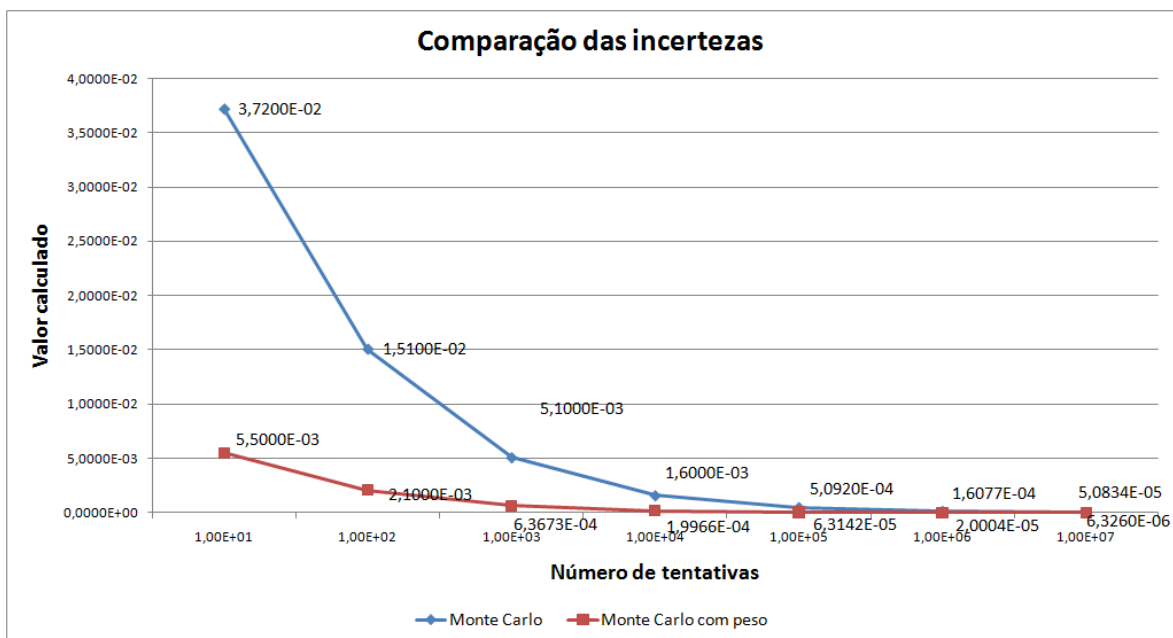


Figura 5.11 – Comparação da incerteza no cálculo da integral (5.5) utilizando Método de Monte Carlo

5.4 – O Algoritmo de Metropolis

O método descrito anteriormente utiliza números aleatórios cuja geração pelo computador pode ser otimizada de acordo com distribuições específicas que podem ser muito eficientes, mas pode ser muito difícil ou até mesmo impossível gerar uma amostragem de números com função-peso complicada para várias dimensões, sugerindo que uma aproximação alternativa deva ser usada.

Uma maneira geral de se produzir variáveis aleatórias com uma dada distribuição de probabilidade de forma arbitrária é conhecida como **Algoritmo de Metropolis**. Em uma publicação de 1953, Nick Metropolis, Marshall e Arianna Rosenbluth, e Edward e Mici Teller descreveram pela primeira vez a técnica de Monte Carlo, que passou a ser conhecida com esse nome (Newman & Barkema, 2001; Danese *et. al*, 1993). Esse algoritmo é o mais utilizado hoje em dia juntamente às técnicas de Monte Carlo, e requer somente a habilidade para se calcular a **função-peso** para um dado valor das variáveis de integração. Uma das aplicações mais difundidas desse algoritmo é na resolução de problemas de mecânica estatística, onde a função-peso do “*ensemble*” canônico pode ser uma função muito complicada das coordenadas do sistema e, por essa razão, pode não ser corretamente descrita e avaliada por outros métodos (Koonin & Meredith, 1990).

Pode-se implementar esse algoritmo de várias maneiras, sendo uma delas descrita a seguir. Imagine que se queira gerar um conjunto de pontos no espaço a (possivelmente multidimensional) de variáveis X distribuídas com densidade de probabilidade $w(X)$. O algoritmo de Metropolis gera uma sequência de pontos X_0, X_1, \dots , de maneira tal que, quanto maior o número de pontos, maior a chance de se chegar à distribuição desejada. Imagine que tenha sido gerado o ponto X_n . Para geração do ponto X_{n+1} é feito um ensaio para o novo ponto X_t , que pode ser escolhido de qualquer maneira que seja conveniente, por exemplo, uniformemente aleatório, dentro de um cubo multidimensional de lado com pequeno tamanho. Esse ensaio é, então, aceito ou rejeitado, de acordo com a razão:

$$r = \frac{w(X_t)}{w(X_n)} \quad (5.13)$$

Se o valor de r for maior do que 1, o passo é aceito, *i.e.*, faz-se $X_{n+1} = X_t$. Se o valor de r for menor do que 1 o passo é aceito com probabilidade r , o que pode ser conseguido de maneira conveniente através da comparação do valor de r com um número aleatório η distribuído uniformemente no intervalo $[0,1]$, aceitando-se esse passo se $\eta < r$. Se o passo não for aceito, então ele é rejeitado, e o algoritmo deve fazer $X_{n+1} = X_n$. Essa operação gera o valor de X_{n+1} , abrindo caminho para a geração de X_{n+2} pelo mesmo processo. Qualquer ponto arbitrário pode ser usado como ponto inicial para o passo aleatório.

Agora considere um grande número de pontos começando de diferentes valores iniciais dentro do espaço X . Se $N_n(X)$ é a densidade desses pontos em X após n passos, então o número líquido de passos com movimentação do ponto X ao ponto Y no próximo passo será:

$$\begin{aligned}
\Delta N(X) &= N_n(X)P(X \rightarrow Y) - N_n(Y)P(Y \rightarrow X) \Rightarrow \\
\Rightarrow \frac{\Delta N(X)}{N_n(Y)P(X \rightarrow Y)} &= \frac{N_n(X)P(X \rightarrow Y)}{N_n(Y)P(X \rightarrow Y)} - \frac{N_n(Y)P(Y \rightarrow X)}{N_n(Y)P(X \rightarrow Y)} \Rightarrow \\
\Rightarrow \frac{\Delta N(X)}{N_n(Y)P(X \rightarrow Y)} &= \frac{N_n(X)}{N_n(Y)} - \frac{P(Y \rightarrow X)}{P(X \rightarrow Y)} \Rightarrow \\
\Delta N(X) &= N_n(Y)P(X \rightarrow Y) \left[\frac{N_n(X)}{N_n(Y)} - \frac{P(Y \rightarrow X)}{P(X \rightarrow Y)} \right]
\end{aligned} \tag{5.14}$$

em que $P(X \rightarrow Y)$ é a probabilidade de movimentação de X até Y , enquanto que $P(Y \rightarrow X)$ é a probabilidade de movimentação de Y até X . A equação (5.14) mostra que há equilíbrio quando:

$$\frac{N_n(X)}{N_n(Y)} = \frac{N_e(X)}{N_e(Y)} \equiv \frac{P(Y \rightarrow X)}{P(X \rightarrow Y)} \tag{5.15}$$

Além disso, alterações em $N(X)$ quando o sistema não está em equilíbrio tendem a levá-lo a esse estado, *i.e.*, $N(X)$ é positivo se existirem muitos passos em X ou se $N_n(X)/N_n(Y)$ for maior que o valor de equilíbrio. Conclui-se que, após um grande número de passos, será atingido o estado de equilíbrio N_e .

As probabilidades de transição do algoritmo de Metropolis levam a uma distribuição de equilíbrio $N_e(X) \sim w(X)$. A probabilidade de deslocamento de X até Y é dada por:

$$P(X \rightarrow Y) = T(X \rightarrow Y)A(X \rightarrow Y) \quad (5.16)$$

em que T é a probabilidade de efetuar um ensaio de X até Y , enquanto A é a probabilidade de se aceitar esse passo. Se Y pode ser alcançado a partir de X através de um único passo, então:

$$T(X \rightarrow Y) = T(Y \rightarrow X) \quad (5.17)$$

Assim, a distribuição de equilíbrio satisfaz a equação a seguir:

$$\frac{N_e(X)}{N_e(Y)} = \frac{A(Y \rightarrow X)}{A(X \rightarrow Y)} \quad (5.18)$$

Se $w(X) > w(Y)$, então:

$$\begin{aligned} A(Y \rightarrow X) &= 1 \\ A(X \rightarrow Y) &= \frac{w(Y)}{w(X)} \end{aligned}$$

Se $w(X) < w(Y)$, então:

$$\begin{aligned} A(X \rightarrow Y) &= 1 \\ A(Y \rightarrow X) &= \frac{w(X)}{w(Y)} \end{aligned}$$

Em ambos os casos, a população de equilíbrio deve satisfazer a:

$$\frac{N_e(X)}{N_e(Y)} = \frac{w(X)}{w(Y)}$$

Ao longo do algoritmo, pode-se usar quaisquer transições e regras de aceitação que satisfaçam à equação seguinte:

$$\frac{w(X)}{w(Y)} = \frac{T(Y \rightarrow X)A(Y \rightarrow X)}{T(X \rightarrow Y)A(X \rightarrow Y)} \quad (5.19)$$

Como saber o tamanho do passo? Suponha que X_n esteja no máximo de w , o lugar mais provável para ele. Se o passo for muito grande, então $w(X_l)$ será muito menor do que $w(X_n)$ e a maioria dos ensaios será rejeitada, levando a uma amostragem ineficiente de w . Se o passo for muito pequeno, a maioria dos ensaios será aceita, mas o deslocamento nunca será muito grande, prejudicando a distribuição. Uma boa regra é que o tamanho do passo deve ser escolhido de maneira que a metade dos ensaios seja aceita (Koonin & Meredith, 1990).

Um problema apresentado pelo algoritmo de Metropolis é o fato de novos pontos serem sempre dependentes de pontos anteriores. Assim, é provável que o ponto X_{n+1} esteja nas vizinhanças de X_n . Outro problema é a escolha do valor inicial X_0 . A princípio ele pode ser qualquer um, sendo os resultados independentes dessa escolha após certo número de passos. Na prática esse ponto inicial deve ser escolhido onde w for grande.

No trabalho original, de 1953, foi descrito um método geral, apropriado para computadores, onde é possível calcular propriedades de qualquer substância que possa ser considerada como um composto onde moléculas individuais interagem entre si. Foi utilizada estatística clássica, forças bidimensionais de corpo a corpo, e o potencial de cada molécula foi considerado como esfericamente simétrico. Essas três considerações são usuais em teorias de líquidos. Esse método não é restrito a qualquer faixa de temperatura ou de densidade. No mesmo trabalho foram apresentados resultados preliminares de cálculos bidimensionais para sistemas de esferas rígidas.

Considere um número finito N de partículas incidindo sobre uma superfície quadrada. A fim de se minimizar os efeitos nessa, considera-se que ela é periódica, ou seja, é composta de muitas superfícies quadradas, cada qual interagindo com N partículas, na mesma configuração. Define-se d_{AB} como a menor distância possível entre as partículas A e B do sistema onde o potencial contribui substancialmente para sua interação. A energia potencial do sistema pode ser facilmente calculada se forem conhecidas as posições das N partículas e o potencial V entre elas:

$$E = \frac{1}{2} \sum_{i=1}^N \sum_{\substack{j=1 \\ i \neq j}}^N V(d_{ij}) \quad (5.20)$$

Para se calcular uma quantidade F de interesse, pode-se utilizar a equação a seguir:

$$\overline{F} = \frac{\int F e^{-\frac{E}{kT}} d^{2N} p d^{2N} q}{\int e^{-\frac{E}{kT}} d^{2N} p d^{2N} q} \quad (5.21)$$

em que $d^{2N} p d^{2N} q$ é o elemento de volume de um espaço com quatro dimensões. Como as forças entre as partículas são independentes das velocidades, a integração pode ser feita em um espaço de dimensão $2N$. Com o uso de técnicas de Monte Carlo, evita-se a integração de espaços com centenas de dimensões, executando-se apenas amostragem aleatória de pontos para efetuar esses cálculos.

Uma maneira bastante simples para se efetuar esses cálculos seria colocar cada uma das N partículas em uma posição aleatória qualquer dentro da superfície quadrada (dimensão $2N$), calcular a energia de interação entre elas utilizando-se a equação (5.20) e dar um peso $e^{-E/kT}$ para essa configuração. Esse método não é aconselhável porque pode-se dar uma alta probabilidade para situações em que o peso $e^{-E/kT}$ é pequeno. Ao invés disso, pode-se escolher configurações com probabilidade $e^{-E/kT}$ e dar peso uniformemente a ela.

Assim, coloca-se as N partículas em uma configuração qualquer, movendo-as sucessivamente de acordo com as seguintes equações:

$$\begin{aligned} X &\rightarrow X + \alpha\xi_1 \\ Y &\rightarrow Y + \alpha\xi_2 \end{aligned} \quad (5.22)$$

Nas equações anteriores, α é o deslocamento máximo permitido (arbitrário), enquanto que ξ_1 e ξ_2 são números aleatórios entre -1 e 1. Assim, após a movimentação de uma partícula, é igualmente provável que ela esteja em qualquer lugar dentro de um quadrado de lado 2α , centrado na sua posição original.

Calcula-se, então, a alteração de energia no sistema ΔE , ocasionada pelo movimento da partícula. Se $\Delta E < 0$, isto é, se o movimento causar uma diminuição na quantidade de energia do sistema, o movimento é permitido e a partícula é aceita na sua nova posição. Se $\Delta E > 0$, o movimento é aceito com probabilidade $e^{-\Delta E/kT}$, isto é, toma-se um número aleatório ζ_3 entre 0 e 1 e, se $\zeta_3 < e^{-\Delta E/kT}$, então a partícula é movida para sua posição; caso contrário, ela retorna à sua posição anterior. Independentemente de se aceitar a movimentação ou não, a nova configuração do sistema é aceita para o propósito de se efetuar os cálculos dos valores médios. Assim, considerando-se F_j como o valor da propriedade F do sistema após o j -ésimo movimento, tem-se:

$$\bar{F} = \frac{1}{M} \sum_{j=1}^M F_j \quad (5.23)$$

Esse processo descrito anteriormente é repetido para as demais partículas. Como uma partícula pode se mover para qualquer ponto dentro da superfície quadrada de lado 2α com uma probabilidade finita, é perceptível que um enorme número de movimentos permitirá que se atinja qualquer um dos pontos de toda a superfície quadrada. Isso é válido para todas as partículas do sistema.

Considere agora um “ensemble” muito grande de sistemas. Para simplificar, considere que existam apenas poucos estados (configurações) do sistema. O número V_r representa o número de sistemas do “ensemble” no estado r . Deve-se provar que, após muitos movimentos, o “ensemble” tende à distribuição:

$$V_r \propto e^{-\frac{E_r}{kT}} \quad (5.24)$$

Antes de se considerar a probabilidade $e^{-\Delta E/kT}$, assumamos que o sistema vá do estado r para o estado s com uma probabilidade qualquer P_{rs} . Inicialmente, as probabilidades P_{rs} e P_{sr} são iguais, pois a partícula pode se mover igualmente em qualquer direção dentro da superfície quadrada de lado $2a$, centrada na sua posição original. Considere agora $E_r > E_s$. O número de sistemas se movendo do estado r para o estado s será dado por $v_r P_{rs}$, já que qualquer movimento para um estado de menor energia é permitido. O número de sistemas se movendo em sentido contrário será dado por $v_s P_{sr} e^{-(E_r - E_s)/kT}$, pois devemos utilizar o fator exponencial para dar “peso” à essa distribuição, utilizando a diferença de energia entre os dois estados. Assim, o número líquido de sistemas se movendo de s para r será dado por:

$$v_s P_{sr} e^{-\frac{E_r - E_s}{kT}} - v_r P_{rs} = P_{rs} \left(v_s e^{-\frac{E_r - E_s}{kT}} - v_r \right) \quad (5.25)$$

Assim, se entre dois estados r e s for válida a equação a seguir, na média mais sistemas se moverão do estado r para o estado s .

$$\frac{v_r}{v_s} > \frac{e^{-\frac{E_r}{kT}}}{e^{-\frac{E_s}{kT}}} \quad (5.26)$$

A seguir é mostrado um exemplo do uso do Algoritmo de Metropolis para resolução de um problema de inferência bayesiana, descrito na referência Martinez & Martinez, 2002. Essa inferência baseia-se em uma técnica estatística na qual evidências ou observações são utilizadas para atualizar ou inferir uma probabilidade onde uma hipótese pode ser verdadeira. São utilizados aspectos de métodos científicos, que envolvem evidências coletadas que podem ser consistentes ou inconsistentes com uma dada hipótese. Na medida em que as evidências se acumulam, o grau de confiança na hipótese se modifica.

O exemplo mostrado trata de acoplamento genético de 197 animais, divididos em 4 categorias com as frequências abaixo discriminadas:

$$Z = (z_1, z_2, z_3, z_4) = (125, 18, 20, 34)$$

As probabilidades celulares correspondentes são:

$$\left(\frac{1}{2} + \frac{\theta}{4}, \frac{1}{4}(1-\theta), \frac{1}{4}(1-\theta), \frac{\theta}{4} \right)$$

A distribuição posterior de Φ a partir dos valores de Z será dada por:

$$P(\theta | Z) = \pi(\theta) \propto (2 + \theta)^{z_1} (1 - \theta)^{z_2 + z_3} \theta^{z_4}$$

Essa função probabilidade é, então, simulada no MATLAB para se observar o comportamento ou valores prováveis de Φ a partir desses dados. O passo é gerado através de distribuição uniforme no intervalo $(-a, a)$. Foram dados 1100 passos para a construção do gráfico a seguir, que mostra um histograma estimativo do valor de Φ :

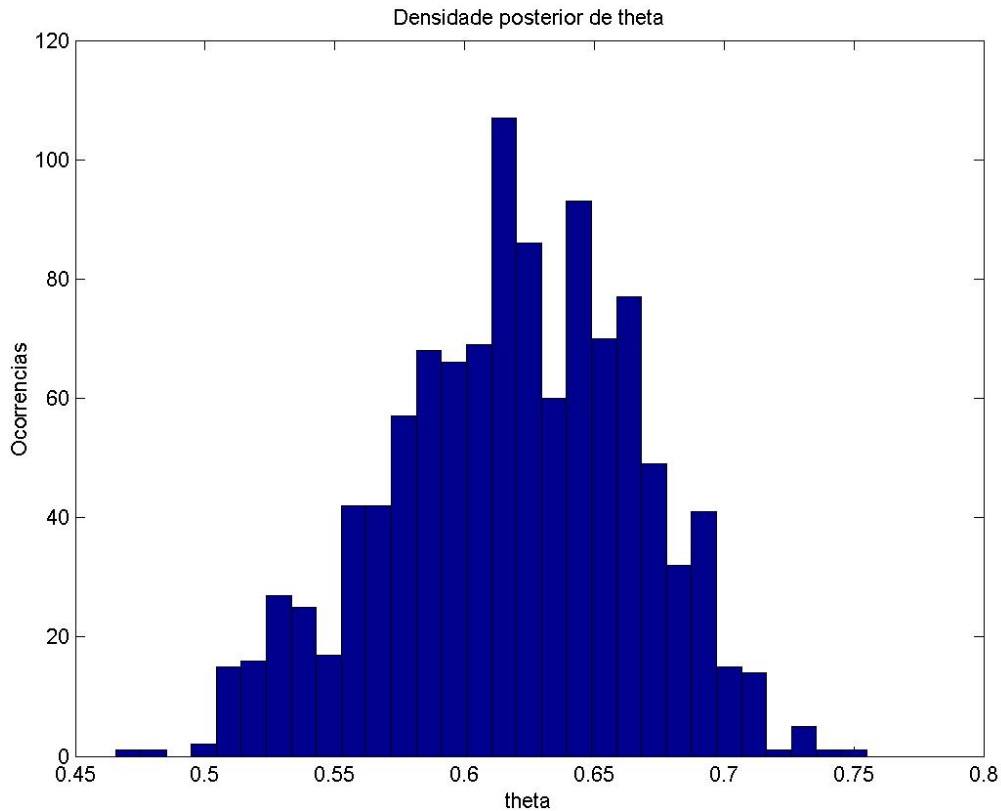


Figura 5.12 – Acoplamento genético de animais usando o Algoritmo de Metropolis

5.5 – Conclusão

O Método de Monte Carlo é uma importante ferramenta matemática para o desenvolvimento e o entendimento das mais diversas ciências existentes. Todas as suas técnicas, começando das mais simples, como a “*simple sampling*”, até as mais sofisticadas, como o “*importance sampling*” e o Algoritmo de Metropolis, têm aplicação dentro da físico-química, podendo ser empregadas principalmente quando se torna muito difícil a solução analítica de determinados problemas, justificando-se o emprego de técnicas de amostragem. Dependendo da experiência do pesquisador, o Método de Monte Carlo a ser empregado pode ser menos ou mais sofisticado, apresentando maior ou menor estatística de erro.

CAPÍTULO VI

O CÓDIGO MCNP MONTE CARLO N-PARTICLE CODE

6.1 – Introdução

MCNP é um programa de computadores para transporte de Monte Carlo para uso geral, de energia contínua, com geometria generalizada, dependente do tempo e que funciona para simulações com nêutrons, fótons e elétrons (X-5 Monte Carlo Team, 2003). O usuário cria um arquivo de entrada que é lido pelo MCNP; esse arquivo contém informações sobre o problema a ser simulado, como:

- Especificação geométrica;
- Descrição de materiais e seleção das seções de choque;
- Localização e caracterização da fonte de nêutrons, fótons ou elétrons;
- Tipo de respostas ou de registros desejados;
- Quaisquer técnicas de redução de discrepâncias para melhorar sua eficiência.

Cinco regras são sempre consideradas no uso desse programa (X-5 Monte Carlo Team, 2003):

- Deve-se definir e provar muito bem a geometria e a fonte;
- Não se pode recuperar informação perdida;
- Deve-se questionar a estabilidade e a confiabilidade dos resultados;
- Deve-se ser conservativo e cuidadoso com influências na redução de variações;
- O número de histórias (execuções) não é indicativo da qualidade da resposta.

O MCNP é um programa que se baseia em **técnicas de Monte Carlo**, ou seja, em estatísticas e ensaios aleatórios repetitivos para resolver problemas físicos e matemáticos. Uma função de distribuição de probabilidade é usada para modelar o fenômeno a ser simulado. Assim, estimam-se as quantidades procuradas utilizando-se médias estatísticas. Um extenso banco de dados com informações experimentais pode ser usado para representar e substituir os ensaios aleatórios. Para minimizar os custos computacionais para armazenar esse banco de dados, o número de ensaios aleatórios é determinado por algoritmo computacional, que gera dados satisfatórios à aleatoriedade, sendo fundamentado em fórmulas de recorrência, onde cada evento é simulado individualmente (Nunes, 2006).

Um arquivo de entrada do MCNP tem basicamente a seguinte estrutura:

1. Cartão de título;
2. Cartões de células geométricas;
3. Linha delimitadora em branco;
4. Cartões de superfícies;
5. Linha delimitadora em branco;
6. Cartões de dados;
7. Linha delimitadora final em branco;
8. Qualquer observação extra (opcional).

A palavra “cartão” é empregada no MCNP para descrever linhas com até 80 caracteres no arquivo de entrada. No cartão de título, informa-se o título do problema a ser resolvido com o MCNP. Esse nome será apresentado posteriormente em alguns lugares dentro do arquivo de saída. Os cartões de células geométricas e os de superfícies são utilizados para descrição da geometria do problema. Os cartões de dados contêm informações gerais, como modo de transporte, materiais empregados, especificação da fonte de radiação e números de histórias, entre outros. As informações localizadas após a última linha delimitadora em branco não são processadas e podem conter dados que serão utilizados em novas versões da simulação, por exemplo.

Qualquer linha no arquivo de entrada que comece com uma letra “C” localizada entre as colunas 1 e 5 e seguida por um espaço em branco é considerada comentário e não é processada pelo programa. Um sinal “\$” no interior de uma linha também indica que tudo o que vier à sua direita é comentário e não é processado.

6.2 – O Método de Monte Carlo versus o Método Determinístico

O método de Monte Carlo é muito diferente dos métodos determinísticos de transporte. Os métodos determinísticos, entre os quais o mais comum é o método de ordenadas discretas, resolvem a equação de transporte para o comportamento médio da partícula, enquanto que o Monte Carlo não resolve uma equação explícita, mas obtém preferencialmente respostas através da simulação de partículas individuais, registrando

alguns aspectos do seu comportamento médio. O comportamento médio das partículas no sistema físico é, então, inferido a partir do comportamento médio das partículas simuladas.

A figura 6.1 a seguir mostra o histórico aleatório de um nêutron incidente sobre uma placa de material que pode sofrer fissão (Sousa, 2004):

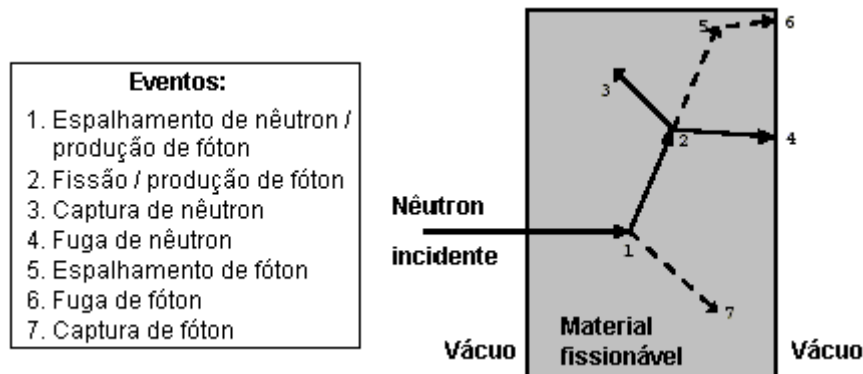


FIGURA 6.1 – *Histórico aleatório de um nêutron incidente sobre uma placa de material que pode sofrer fissão*

Números entre 0 e 1 são selecionados aleatoriamente para determinar qual interação pode acontecer e onde pode acontecer, baseado em regras (físicas) e probabilidades (dados de transporte) que governam os processos e os materiais envolvidos. Nesse exemplo particular, a interação do nêutron ocorre no evento 1. O nêutron é espalhado na direção mostrada, que é selecionada aleatoriamente a partir da distribuição de espalhamento físico. Um fóton é, então, produzido e temporariamente armazenado para análise posterior. No evento 2 ocorre a fissão, resultando na morte do nêutron inicial e no nascimento de 2 novos nêutrons, além de 1 fóton. Um nêutron e o fóton são armazenados para análise futura. O primeiro nêutron oriundo da fissão é capturado no evento 3 e morre. O nêutron armazenado é, então, analisado e, por amostragem aleatória, escapa da placa no evento 4. O fóton produzido na fissão interage no evento 5 e escapa no evento 6. O fóton remanescente gerado no evento 1 é agora seguido, sendo capturado no evento 7. O MCNP recupera os dados das partículas armazenadas previamente, de forma que a última partícula armazenada é a primeira a ser analisada. O histórico do nêutron está completo. Esse histórico é feito para todas as partículas, sendo que o usuário informa ao código a quantidade desejada de partículas, e na resposta, o usuário é informado de todo o histórico das partículas, além da estimativa de incerteza da precisão estatística dos resultados.

O método de Monte Carlo e os métodos determinísticos, além de resolverem problemas de maneiras diferentes, também constituem soluções diferentes. Os métodos determinísticos fornecem tipicamente resultados completos (por exemplo, fluxo) sobre o problema, enquanto que o método de Monte Carlo fornece apenas as informações solicitadas pelo usuário. Quando o método de Monte Carlo e os métodos de ordenadas discretas são comparados, diz-se frequentemente que o método de Monte Carlo resolve a equação de transporte integral, enquanto os métodos de ordenadas discretas resolvem a equação de transporte integro-diferencial (Los Alamos National Laboratory, 2000).

6.3 – Características

São utilizados dados do Nuclear Physics Group de Los Alamos, além de bibliotecas de dados atômicos e nucleares de energia contínua cujas fontes primárias de dados compiladas em Livermore são (X-5 Monte Carlo Team, 2003):

- ENDF (Evaluated Nuclear Data File);
- ACTI (Advanced Computational Technology Initiative);
- ENDL (Evaluated Nuclear Data Library);
- EPDL (Evaluated Photon Data Library);
- ACTL (ACTivation Library).

Existem tabelas de dados para interações de nêutrons, de fótons induzidos por nêutrons, de fótons, de dosimetria ou ativação de nêutrons e de espalhamento térmico de partículas. Cada tabela de dados disponível para o MCNP está listada em um diretório de arquivos chamado XSDIR. O usuário pode selecionar tabelas de dados específicas por meio de identificadores únicos para cada tabela, chamados ZAIDs. Esses identificadores geralmente contêm o número atômico Z , a massa atômica A e o identificador ID da biblioteca específica.

Aproximadamente 836 tabelas de interações neutrônicas estão disponíveis para cerca de 100 diferentes isótopos e elementos. Novas tabelas de interações para nêutrons são adicionadas constantemente, além de avaliações de revisão. Tabelas de interação de fótons estão disponíveis para $Z = 1$ até $Z = 100$. Esses dados permitem que o MCNP considere espalhamento coerente e incoerente, absorção fotoelétrica com a possibilidade de

emissão fluorescente e produção de pares. Seções de choque para aproximadamente 2000 reações de ativação envolvendo cerca de 400 núcleos alvo em estado fundamental e/ou excitado fazem parte do pacote de dados do MCNP. Essas seções de choque podem ser usadas como funções-resposta dependentes da energia para determinação das taxas de reação, mas não podem ser usados como seções de choque de transporte. Tabelas de dados térmicos são próprias para uso em reações de espalhamento. Esses dados incluem ligações químicas e efeitos cristalinos que são importantes quando a energia dos nêutrons se torna muito baixa. Dados em várias temperaturas estão disponíveis para água leve ou pesada, berílio metálico ou óxido, benzeno, grafite, polietileno, zircônio e hidrogênio em hidreto de zircônio (X-5 Monte Carlo Team, 2003).

O usuário deve criar uma fonte especificando seus detalhes, como tipo de partículas emitidas, geometria, energia e direção, entre vários outros parâmetros. O MCNP disponibiliza algumas funções analíticas para distribuição de probabilidade de partículas da fonte, como espectro Gaussiano, Maxwelliano e de Watt. A energia dos nêutrons emitidos pela fonte do MCNP pode variar de 10^{-11} MeV a 20 MeV para todos os isótopos e pode chegar até a 150 MeV para alguns isótopos. No caso dos fótons, a energia pode variar de 1 keV a 100 GeV, enquanto que, para elétrons, a variação é de 1 keV a 1 GeV.

Os dados de saída podem ser configurados através de cartões relacionados à corrente de partículas, fluxo de partículas e deposição de energia. Correntes podem ser configuradas como função de direção em superfícies, segmentos de superfícies ou somatório de superfícies. Fluxos através de superfícies, segmentos de superfícies, somatório de superfícies, interior de células, segmentos de células ou somatório de células também podem ser disponibilizados. Cartões relacionados a calor e fissão dão a deposição de energia em células específicas. Todos os cartões são função de tempo e energia, especificados pelo usuário, sendo normalizados por partícula emitida.

Além das informações contidas nos cartões, o arquivo de saída contém tabelas com informações padrão sumarizadas, dando ao usuário uma idéia melhor sobre como o programa rodou. Essas informações ajudam no entendimento da física do problema e na adequação da simulação de Monte Carlo. Se acontecer algum erro durante a execução da simulação do problema, um relatório detalhado de diagnóstico é dado no arquivo de saída.

Para cada cartão de saída é dado também um erro relativo estatístico, correspondente a um desvio padrão. As informações relativas à estimativa de erro são computadas após o término de cada histórico de partida. Esse erro relativo estatístico pode ser utilizado para formar intervalos de confiança sobre a média estimada, permitindo fazer afirmações sobre a veracidade dos resultados. O Teorema Limite Central mostra que na medida em que N (número de histórias) se aproxima do infinito, há uma chance de 68% de que o resultado verdadeiro esteja na faixa da média estimada (média mais erro relativo estatístico) e de 95% de que esteja na faixa da média mais duas vezes o erro relativo estatístico. É extremamente importante notar que essa afirmação se refere somente à precisão dos cálculos de Monte Carlo e não à sua exatidão de resultado se comparado com o valor físico real (X-5 Monte Carlo Team, 2003).

6.4 – Geometria

A geometria no MCNP trata de configurações 3D arbitrárias, utilizando materiais definidos pelo usuário em células geométricas. As células são definidas através de interseções, uniões e complementos de regiões existentes em superfícies definidas. As superfícies são descritas através de equações analíticas ou, para certos tipos de superfícies, através de pontos conhecidos. O MCNP também possibilita o uso de macro-corpos, onde superfícies básicas como esferas, caixas e cilindros podem ser combinadas através de operadores booleanos.

O MCNP apresenta uma geometria mais geral do que aquela disponibilizada na maioria dos códigos geométricos combinatórios. O usuário tem ainda a flexibilidade de definir regiões geométricas e de combiná-las entre si. O código faz checagem interna extensa à procura de erros. Além disso, o programa conta ainda com “interface” gráfica que ajuda o usuário a visualizar o problema a ser simulado. As células são tratadas em um sistema de coordenadas cartesianas.

As células são definidas em cartões de células. Cada uma é descrita através de um número, um número de material e a densidade desse. A seguir aparece uma lista de operadores e de números de superfícies que contornam a célula. Caso o sinal da superfície seja positivo, esse pode ser omitido. O número do material e sua densidade podem ser substituídos por 0, que indica vácuo. O número da célula deve ser colocado entre as

colunas 1 e 5. O restante vem a seguir, com espaços em branco para separação de dados. A descrição da geometria define a célula através de interseções, uniões e/ou complementos de regiões listadas.

As superfícies são definidas através de mnemônicos como C/Z, que indica um cilindro paralelo ao eixo-z. Existem 29 mnemônicos disponíveis no MCNP versão 5, sendo todos encontrados no manual do usuário. Há basicamente dois caminhos para se especificar parâmetros de uma superfície: através do fornecimento de coeficientes apropriados para satisfazer a equação de superfície e através da especificação de pontos geométricos conhecidos de uma superfície rotacionalmente simétrica em um eixo de coordenadas. A tabela 6.1 a seguir mostra os cartões de superfície que podem ser utilizados no MCNP:

Tabela 6.1 – Cartões de superfície do MCNP

Mnemônico	Tipo	Descrição	Equação	Entrada
P		Geral	$Ax + By + Cz = 0$	A B C D
PX	Plano	Normal ao eixo-x	$x - D = 0$	D
PY		Normal ao eixo-y	$y - D = 0$	D
PZ		Normal ao eixo-z	$z - D = 0$	D
S0		Centralizada na origem	$x^2 + y^2 + z^2 - R^2 = 0$	R
S		Geral	$(x - \bar{x})^2 + (y - \bar{y})^2 + (z - \bar{z})^2 - R^2 = 0$	\bar{x} \bar{y} \bar{z} R
SX	Esfera	Centralizada no eixo-x	$(x - \bar{x})^2 + y^2 + z^2 - R^2 = 0$	\bar{x} R
SY		Centralizada no eixo-y	$x^2 + (y - \bar{y})^2 + z^2 - R^2 = 0$	\bar{y} R
SZ		Centralizada no eixo-z	$x^2 + y^2 + (z - \bar{z})^2 - R^2 = 0$	\bar{z} R
C/X	Cilindro	Paralelo ao eixo-x	$(y - \bar{y})^2 + (z - \bar{z})^2 - R^2 = 0$	\bar{y} \bar{z} R
C/Y		Paralelo ao eixo-y	$(x - \bar{x})^2 + (z - \bar{z})^2 - R^2 = 0$	\bar{x} \bar{z} R
C/Z		Paralelo ao eixo-z	$(x - \bar{x})^2 + (y - \bar{y})^2 - R^2 = 0$	\bar{x} \bar{y} R
CX		No eixo-x	$y^2 + z^2 - R^2 = 0$	R
CY		No eixo-y	$x^2 + z^2 - R^2 = 0$	R
CZ	No eixo-z	$x^2 + y^2 - R^2 = 0$	R	

Tabela 6.1 – Cartões de superfície do MCNP (continuação)

Mnemônico	Tipo	Descrição	Equação	Entrada
K/X		Paralelo ao eixo-x	$\sqrt{(y-\bar{y})^2 + (z-\bar{z})^2} - t(x-\bar{x}) = 0$	$\bar{x} \quad \bar{y} \quad \bar{z}$ $t^2 \pm 1$
K/Y		Paralelo ao eixo-y	$\sqrt{(x-\bar{x})^2 + (z-\bar{z})^2} - t(y-\bar{y}) = 0$	$\bar{x} \quad \bar{y} \quad \bar{z}$ $t^2 \pm 1$
K/Z	Cone	Paralelo ao eixo-z	$\sqrt{(x-\bar{x})^2 + (y-\bar{y})^2} - t(z-\bar{z}) = 0$	$\bar{x} \quad \bar{y} \quad \bar{z}$ $t^2 \pm 1$
KX		No eixo-x	$\sqrt{y^2 + z^2} - t(x-\bar{x}) = 0$	$\bar{x} \quad t^2 \pm 1$
KY		No eixo-y	$\sqrt{x^2 + z^2} - t(y-\bar{y}) = 0$	$\bar{y} \quad t^2 \pm 1$
KZ		No eixo-z	$\sqrt{x^2 + y^2} - t(z-\bar{z}) = 0$	$\bar{z} \quad t^2 \pm 1$
SQ	Elipsóide, hiperbolóide, parabolóide	Eixo paralelo ao eixo-x, eixo-y ou eixo-z	$A(x-\bar{x})^2 + B(y-\bar{y})^2 + C(z-\bar{z})^2 + 2D(x-\bar{x}) + 2E(y-\bar{y}) + 2F(z-\bar{z}) + G = 0$	A B C D E F G $\bar{x} \quad \bar{y} \quad \bar{z}$
GQ	Cilindro, cone elipsóide, hiperbolóide, parabolóide	Eixo não paralelo ao eixo-x, eixo-y ou eixo-z	$Ax^2 + By^2 + Cz^2 + Dxy + Eyz + Fzx + Gz + Hy + Jz + K = 0$	A B C D E F G H J K
TX			$(x-\bar{x})^2 / B^2 + (\sqrt{(y-\bar{y})^2 + (z-\bar{z})^2} - A)^2 / C^2 - 1 = 0$	$\bar{x} \quad \bar{y} \quad \bar{z}$ A B C
TY	“Torus” elíptico ou circular	Eixo paralelo ao eixo-x, eixo-y ou eixo-z	$(y-\bar{y})^2 / B^2 + (\sqrt{(x-\bar{x})^2 + (z-\bar{z})^2} - A)^2 / C^2 - 1 = 0$	
TZ			$(z-\bar{z})^2 / B^2 + (\sqrt{(x-\bar{x})^2 + (y-\bar{y})^2} - A)^2 / C^2 - 1 = 0$	
XYZP		Superfícies definidas por pontos		

Os problemas de transporte de fluxo de partícula em Monte Carlo são estimados frequentemente como o comprimento de rasto por unidade de volume ou o número de partículas que cruzam uma superfície por unidade de área. Entretanto, saber os volumes e as áreas de superfície das regiões geométricas em um problema de Monte Carlo é essencial. Conhecer os volumes é útil no cálculo das massas e densidades das células e, assim, pode-se calcular o aquecimento mássico ou volumétrico. Além disso, cálculo da massa de uma geometria frequentemente é uma boa checagem na exatidão da geometria quando a massa for conhecida através de outros meios.

Calcular volumes e áreas de superfície em modernos códigos de transporte de Monte Carlo não é trivial. O MCNP permite a construção de células a partir de interseções e/ou de uniões de regiões definidas por uma combinação arbitrária de superfícies de complexas. Essas superfícies podem ter orientações diferentes ou podem estar segmentadas

por algum propósito. Embora tal generalidade aumente a flexibilidade de MCNP, calcular volumes de células e áreas de superfície requer métodos computacionais muito elaborados.

O MCNP calcula volumes e áreas de células poliédricas automaticamente e de células ou superfícies geradas por superfícies de revolução sobre qualquer eixo, até mesmo um eixo torcido. Se um cartão é segmentado, os volumes ou áreas segmentadas serão calculados.

6.5 – História da Partícula

Para uma dada história é configurado um número sequencial randômico e o número da história, NPS, é incrementado. O valor do parâmetro IPT é configurado para o tipo de partícula: 1 para nêutron, 2 para fóton e 3 para elétron. Algumas matrizes e variáveis são inicializadas em zero. O valor de NODE é fixado em 1.

A seguir a rotina da fonte apropriada é acionada. As opções de fonte são fonte fixa padrão, fonte de superfície, fonte de criticalidade e fonte definida pelo usuário. Todos os parâmetros necessários para descrição da fonte são incluídos nessas rotinas, incluindo posição, direção de emissão, energia, peso, tempo e célula inicializadora. Várias checagens são feitas nesse momento para verificar se as partículas estão na célula correta ou na superfície correta.

Nesse momento, são impressos os parâmetros iniciais das primeiras 50 histórias de partículas. São feitas, então, checagens relacionadas à energia, tempo e peso. Um número de checagem de erros é feito. Contribuições de detector são marcadas. Partículas geradas pela rotina DXTRAN (se for o caso) são criadas em esferas. As partículas são salvas em um banco para posterior acompanhamento.

A partícula atual a ser transportada é iniciada. Para uma fonte de elétrons, as partículas são rodadas separadamente. Para uma fonte de nêutrons ou de fótons, a interseção da trajetória da partícula com cada superfície da célula é calculada. As seções de choque são calculadas; a seção de choque total é modificada por transformação exponencial, se necessário. A distância para a próxima colisão é determinada. Algumas

informações são sumarizadas. Os parâmetros da partícula (tempo, posição e energia) são atualizados.

Se a distância para colisão for menor do que a distância até a superfície, ou se uma partícula multi-grupo carregada alcançar a distância até a fronteira de energia, a partícula sofre uma colisão. A energia e a direção das partículas resultantes são determinadas. A análise da colisão trata de colisões fotonucleares, absorção fotoelétrica, captura, espalhamento Compton, espalhamento Thomson e produção de pares. Partículas secundárias de colisões fotonucleares (fótons ou nêutrons) são tratadas utilizando as mesmas rotinas de colisões neutrônicas inelásticas. Elétrons são gerados por espalhamento incoerente, produção de pares e absorção fotoelétrica.

Após cruzar a superfície ou sofrer a colisão, o transporte continua com o cálculo da distância até a fronteira da célula, e assim por diante. Se uma partícula envolvida na colisão for extinta por captura ou redução de variância, o banco é checado para buscar partículas anteriores. Caso não existam partículas, a história é terminada. O sumário de informações é incrementado e os cartões dessa história particular são adicionados aos dados totalizadores.

6.6 – Seções de choque

O pacote do MCNP é incompleto sem as tabelas de dados nucleares associados. Existem dois objetivos para se preparar essas tabelas para uso no programa: os dados disponíveis para o MCNP devem reproduzir os valores originais obtidos na prática e novos dados devem ser incorporados com facilidade para que o programa esteja sempre utilizando valores atualizados.

Existem nove classes de tabelas de dados no MCNP (X-5 Monte Carlo Team, 2003):

- Dados de interação neutrônica de energia contínua;
- Dados de interação neutrônica de reação discreta;
- Dados de interação fotoatômica de energia contínua;
- Dados de interação fotonuclear de energia contínua;

- Seções de choque de dosimetria neutrônica;
- Dados térmicos neutrônicos;
- Nêutron multigrupo, acoplamento nêutron-fóton e partículas carregadas mascarando nêutrons;
- Fóton multigrupo;
- Dados de interação de elétrons.

Dados fotoatômicos e elétricos são atômicos por natureza, *i.e.*, uma tabela elemental é aceita para qualquer isótopo do elemento. Tabelas neutrônicas e fotonucleares são nucleares ou isotópicas por natureza, *i.e.*, cada isótopo requer sua própria tabela. As tabelas de dados disponíveis para o MCNP são identificadas por um ZAID, cuja forma geral é ZZZAAA.nnX, em que ZZZ é o número atômico, AAA é a massa atômica, nn é um identificador único e X indica a classe de dados. No caso de avaliação elemental, o valor de AAA é 0.

6.7 – Física do MCNP

Peso

Em um nível bastante fundamental pode-se dizer que o peso é um multiplicador de cartão, ou seja, um ajuste para o desvio de uma simulação física direta do processo de transporte. Note que, se um código de Monte Carlo sempre utilizar as mesmas distribuições como faz a natureza, então esse código terá a mesma média e a mesma variância observadas naturalmente. A variância natural é inaceitavelmente alta e o código de Monte Carlo a modifica utilizando algumas formas de redução, o que reduz o tempo requerido no computador para se obter resultados com precisão suficiente. Os métodos de redução de variância usam esquemas de peso para produzir a mesma média que os processos naturais de transporte, mas com discrepância calculável mais baixa.

Rastos de partícula

Quando uma partícula parte de uma fonte, seu rasto é criado. Se aquele rasto for dividido em 2 durante uma interação com uma superfície ou uma colisão, um segundo rasto é criado e os dois rastos da partícula original serão acompanhados, cada um com

metade do peso. Se um dos rastros estiver relacionado com uma reação (n,2n), mais um rastro é criado, totalizando 3. Um rastro é relacionado a cada componente de uma partícula da fonte durante sua história. Os cartões de comprimento de rastro usam essa medida em determinada célula para calcular uma quantidade de interesse, como fluência, fluxo ou deposição de energia. São usadas superfícies de cruzamento de rastros para calcular fluência, fluxo ou estimativas de superfície. Rastros relacionados a colisões são usados para se calcular multiplicação e criticalidade.

Dentro de uma determinada célula ou composição fixa, o método de provar uma colisão ao longo do rastro é determinado usando a teoria seguinte. A probabilidade de uma primeira colisão para uma partícula localizada entre l e $l + dl$ ao longo de sua linha de trajetória é dada por:

$$p(l)dl = e^{-\sum_t l} \sum_t dl \quad (6.1)$$

em que Σ_t é a seção de choque macroscópica total do meio e é interpretada como a probabilidade por unidade de comprimento de uma colisão. Assume-se ζ como um número aleatório localizado no intervalo [0,1], ou seja:

$$\zeta = \int_0^l e^{-\sum_t s} \sum_t ds = 1 - e^{-\sum_t l} \quad (6.2)$$

Assim:

$$l = -\frac{1}{\sum_t} \ln(1 - \zeta) \quad (6.3)$$

Mas, como $1 - \zeta$ é distribuída da mesma maneira que ζ e, portanto, pode ser substituída por esse valor na equação acima, então obtém-se a expressão para o cálculo da distância da colisão:

$$l = -\frac{1}{\sum_t} \ln(\zeta) \quad (6.4)$$

Interações de nêutrons

Quando uma partícula colide com um núcleo, a seguinte sequência acontece (X-5 Monte Carlo Team, 2003):

1. O núcleo colidido é identificado;
2. Ou o tratamento $S(\alpha,\beta)$ para espalhamento térmico é usado ou a velocidade do núcleo é testada com nêutrons de baixa energia;
3. Fótons são gerados opcionalmente para transporte posterior;
4. Captura neutrônica (*i.e.*, desaparecimento de nêutron por qualquer processo) é modelada;
5. Exceto quando o tratamento $S(\alpha,\beta)$ é usado, ou espalhamento elástico ou inelástico (incluindo fissão) é selecionado e a nova energia e direção do rasto resultante são determinadas;
6. Se a energia do nêutron é baixa o bastante e uma tabela $S(\alpha,\beta)$ apropriada está presente, a colisão é modelada pelo tratamento de $S(\alpha,\beta)$ ao invés do passo anterior.

Um espalhamento entre um nêutron e um átomo é afetado pelo movimento térmico do átomo e, na maioria dos casos, a interação também é afetada pela presença de outros átomos. O movimento térmico não pode ser ignorado em muitas aplicações de MCNP sem erro sério. Os efeitos de átomos localizados nas proximidades também são importantes em algumas aplicações. O MCNP usa um tratamento térmico baseado na aproximação de gás livre para considerar o movimento térmico. Tem também uma capacidade explícita de $S(\alpha,\beta)$ que considera os efeitos de ligação química e estrutura cristalina para energias de nêutron incidentes abaixo de aproximadamente 4 eV, mas está disponível para somente um número limitado de substâncias e temperaturas.

O tratamento térmico de gás livre no MCNP considera que o meio é um gás livre e também que, na gama de peso atômico e energia de nêutron onde efeitos térmicos são significativos, a seção de choque de espalhamento elástico à temperatura zero é quase independente da energia do nêutron, e que as seções de choque de reação são quase independentes de temperatura. Essas suposições permitem ao MCNP dar um tratamento térmico de colisões neutrônicas que roda no computador quase tão rapidamente quanto um tratamento completamente não-térmico e isso é adequado para a maioria dos problemas

práticos. Com as suposições anteriores, o tratamento térmico do gás livre consiste em ajustar a seção de choque elástica e levar em conta a velocidade do núcleo designado quando estão sendo calculadas as cinemáticas de uma colisão. O tratamento térmico de gás livre do MCNP só é aplicado efetivamente em espalhamentos elásticos.

A seção de choque de espalhamento efetivo em um sistema de laboratório para um nêutron de energia cinética E é dada por:

$$\sigma_s^{eff}(E) = \frac{1}{v_n} \iint \sigma_s(v_{rel}) v_{rel} p(V) dv \frac{d\mu_t}{2} \quad (6.5)$$

em que v_{rel} é a velocidade relativa entre o nêutron se movendo com uma velocidade escalar v_n e o núcleo-alvo se movendo com uma velocidade escalar V ; μ_t é o cosseno do ângulo entre os vetores do nêutron e da direção de movimentação do alvo. A equação para v_{rel} é:

$$v_{rel} = \sqrt{(v_n^2 + V^2 - 2v_n V \mu_t)} \quad (6.6)$$

A seção de choque de espalhamento na velocidade relativa é escrita como $\sigma_s(v_{rel})$ e $p(V)$ é a função densidade de probabilidade para a distribuição Maxwelliana de velocidades do alvo:

$$p(V) = \frac{4}{\sqrt{\pi}} \beta^3 V^2 e^{-\beta^2 V^2} \quad (6.7)$$

em que β é definido como:

$$\beta = \sqrt{\frac{AM_n}{2kT}} \quad (6.8)$$

em que A é a massa do núcleo-alvo em unidades de massa neutrônica, M_n é a massa neutrônica em MeV-sh²/cm² e kT é a temperatura de equilíbrio dos núcleos-alvo em MeV. A velocidade escalar V para os núcleos alvos é $1/\beta$, que corresponde à energia cinética de kT para os núcleos-alvo. Essa não é a energia cinética média dos núcleos, que é $3kT/2$. A

quantidade que o MCNP espera na entrada é kT e não apenas T . Notar que kT não é uma função da massa da partícula e sim a energia cinética na velocidade mais provável das partículas de qualquer massa.

A distribuição de probabilidade para uma velocidade V do alvo e um cosseno μ_t é:

$$P(V, \mu_t) = \frac{\sigma_s(v_{rel})v_{rel}P(V)}{2\sigma_s^{eff}(E)v_n} \quad (6.9)$$

Assume-se que a variação de $\sigma_s(v)$ com a velocidade do alvo pode ser ignorada porque, para núcleos leves, $\sigma_s(v_{rel})$ varia lentamente com a velocidade, enquanto que para núcleos pesados pode variar rapidamente, mas o efeito moderador do espalhamento é pequeno e as consequências da aproximação são desprezíveis. Como resultado, a distribuição de probabilidade pode ser escrita como:

$$P(V, \mu_t) \propto \sqrt{v_n^2 V^2 - 2Vv_n\mu_t} V^2 e^{-\beta^2 V^2} \quad (6.10)$$

Interações de fótons

Colisões com núcleo, captura analógica, captura implícita e muitos outros fatores das interações de fótons como redução de variância são os mesmos utilizados para nêutrons, mas a física da colisão é completamente diferente.

Quando uma partícula colide há uma probabilidade de que ela seja absorvida por um núcleo. Na absorção analógica, a partícula é morta com essa probabilidade. Na absorção implícita, também conhecida como captura implícita, a partícula nunca morre por absorção; ao invés disso, seu peso é reduzido pela probabilidade de absorção a cada colisão. É permitido que partículas importantes sobrevivam não sendo perdidas em absorções. Se partículas já não são consideradas úteis depois de sofrer algumas colisões, a absorção analógica é eficiente na eliminação dessas. O MCNP possui dois modelos de interação para fótons: simples e detalhado.

O tratamento físico simples ignora o espalhamento coerente (Thomson) e os fótons fluorescentes de absorção fotoelétrica. Ele é utilizado em problemas com fótons de alta energia ou em problemas onde elétrons são livres e também importantes para o próximo evento estimativo como detectores puntiformes, onde o espalhamento pode ser aproximadamente retilíneo para frente e elástico. Um detector puntiforme é uma estimativa determinística (do ponto de evento corrente) do fluxo em um ponto do espaço. O tratamento físico simples usa captura implícita a menos que se modifique o programa de entrada, onde seria utilizada captura analógica.

A seção de choque total σ_t é o somatório de 3 componentes:

$$\sigma_t = \sigma_{pe} + \sigma_{pp} + \sigma_s \quad (6.11)$$

em que σ_{pe} é a seção de choque de efeito fotoelétrico, σ_{pp} é a seção de choque de produção de pares e σ_s é a seção de choque de espalhamento Compton.

O tratamento físico detalhado inclui o espalhamento coerente (Thomson) e considera os fótons fluorescentes após absorção fotoelétrica. São usados fatores de forma e perfis Compton para considerar efeitos de elétrons ligados. Captura analógica sempre é usada. O tratamento físico detalhado é o melhor para a maioria das aplicações, particularmente para núcleos de alto Z ou problemas de penetração em profundidade.

A seção de choque diferencial é dada pela fórmula de Klein-Nishina:

$$K(\alpha, \mu)d\mu = \pi r_0^2 \left(\frac{\alpha'}{\alpha} \right)^2 \left[\frac{\alpha'}{\alpha} + \frac{\alpha}{\alpha'} + \mu^2 - 1 \right] d\mu \quad (6.12)$$

em que r_0 é o raio eletrônico clássico ($2,817938 \times 10^{-13}$ cm), α e α' são as energias incidente e final do fóton em unidades de 0,511 MeV [$\alpha = E/(mc^2)$] (m é a massa do elétron e c é a velocidade da luz), e $\alpha' = \alpha/[1 + \alpha(1-\mu)]$.

Na versão 5 do MCNP é incluída a física fotonuclear para interações colisionais de fótons. Uma interação fotonuclear começa com a absorção de um fóton por um núcleo. Há

vários mecanismos pelos quais isso pode acontecer. Os dados nucleares disponíveis atualmente estão na gama de energia até 150 MeV para o fóton incidente. O valor de 150 MeV foi escolhido porque está logo abaixo do limiar para a produção de pions e da necessidade subsequente de modelamento nuclear mais complicado.

Interações de elétrons

O transporte de elétrons e de outras partículas carregadas é fundamentalmente diferente de nêutrons e fótons. A interação de partículas neutras é caracterizada por colisões isoladas relativamente infrequentes, com movimentação livre entre as colisões. Em contraste, o transporte de elétrons é dominado por força de Coulomb, resultando em várias interações pequenas. Como um exemplo, um nêutron no alumínio que reduz a velocidade de 0,5 MeV para 0,0625 MeV terá aproximadamente 30 colisões, enquanto um fóton nas mesmas circunstâncias experimentará menos que dez. Um elétron que realiza a mesma perda de energia sofrerá aproximadamente 10^5 interações individuais. Esse grande aumento em complexidade computacional faz com que o Monte Carlo não seja muito interessante para a maioria das situações de interesse prático.

Trabalho teórico considerável tem sido feito para desenvolver uma variedade de teorias analíticas e semi-analíticas de múltiplos espalhamentos para o transporte de partículas carregadas. Essas teorias tentam usar as seções de choque fundamentais e a natureza estatística do processo de transporte para prever distribuições de probabilidade para quantidades significantes, como perda de energia e desvio angular. Para seguir um elétron durante uma perda significativa de energia, é necessário desmembrar o seu caminho em muitos passos. Esses passos são escolhidos com tamanho grande o bastante para cercar muitas colisões (de forma que as teorias de múltiplo espalhamento sejam válidas) e ao mesmo tempo curto o bastante para que a perda de energia em qualquer passo seja pequena (de forma que as aproximações necessárias para as teorias de múltiplo espalhamento sejam satisfeitas). A perda de energia e o desvio angular do elétron durante cada um dos passos podem ser determinados baseados em teorias apropriadas de múltiplo espalhamento.

O poder de parada colisional do elétron, *i.e.*, a energia perdida por unidade de comprimento de trajetória em colisões que resultam em transferência fracional de energia menor que a energia máxima arbitrária pode ser determinado pela equação a seguir:

$$-\left(\frac{dE}{dS}\right)_{\varepsilon_m} = NZC \left\{ \ln \frac{E^2(\tau+2)}{2I^2} + f^-(\tau, \varepsilon_m) - \delta \right\} \quad (6.13)$$

em que:

$$f^-(\tau, \varepsilon_m) = -1 - \beta^2 + \left(\frac{\tau}{\tau+1}\right)^2 \frac{\varepsilon_m^2}{2} + \frac{2\tau+1}{(\tau+1)^2} \ln(1-\varepsilon_m) + \ln[4\varepsilon_m(1-\varepsilon_m)] + \frac{1}{1-\varepsilon_m} \quad (6.14)$$

Nas equações anteriores, ε e ε_m representam transferência de energia na forma de frações da energia cinética E do elétron; I é o potencial médio de ionização em unidades de E ; $\beta=v/c$; τ é a energia cinética do elétron em unidades de massa de repouso; δ é a correção de efeito de densidade relacionada à polarização do meio; Z é o número atômico médio do meio; N é a densidade atômica do meio em cm^{-3} e o coeficiente C é dado pela equação a seguir:

$$C = \frac{2\pi e^4}{mv^2} \quad (6.15)$$

em que m é a massa de repouso do elétron, e é sua carga e v é a sua velocidade.

O poder de parada radiativo é dado pela equação a seguir:

$$-\left(\frac{dE}{dS}\right)_{rad} = 10^{24} Z \left(Z + \bar{\eta} \right) (\alpha r_e^2) (T + mc^2) \Phi_{rad}^{(n)} \quad (6.16)$$

em que $\Phi_{rad}^{(n)}$ é a seção de choque de perda de energia radiativa da interação elétron-núcleo, $\bar{\eta}$ é um parâmetro que considera o efeito de bremsstrahlung elétron-elétron, α é a constante da estrutura, mc^2 é a energia mássica de um elétron e r_e é o raio eletrônico clássico.

A deflexão angular de um elétron é dada de acordo com a distribuição a seguir:

$$F(s, \mu) = \sum_{l=0}^{\infty} \left(l + \frac{1}{2} \right) \exp(-sG_l) P_l(\mu) \quad (6.17)$$

em que s é o tamanho do subpasso, $\mu = \cos\theta$ é a deflexão angular da direção do início do subpasso, $P_l(\mu)$ é o l -ésimo polinômio de Legendre e G_l é dado por:

$$G_l = 2\pi N \int_{-1}^{+1} \frac{d\sigma}{d\Omega} [1 - P_l(\mu)] d\mu \quad (6.18)$$

6.8 – Cartões de Saída

O MCNP cria informação sumária padrão que dá ao usuário uma visão melhor da física do problema e a adequação da simulação de Monte Carlo inclui automaticamente: uma contabilidade completa da criação e da perda de todos os rastos e de sua energia; o número de rastos entrando e reentrando em uma célula mais a população de rasto na célula; o número de colisões em uma célula; o peso médio, livre caminho médio e a energia de rastos em uma célula; a atividade de cada nuclídeo em uma célula (*i.e.*, como partículas interagiram com cada nuclídeo, não a radioatividade); e um equilíbrio de peso completo para cada célula.

O MCNP também provê sete tipos padrão de cartões. Esses incluem sete padrões para nêutrons, seis padrões para fótons e quatro padrões para elétrons. Esses cartões básicos podem ser modificados em muitas formas pelo usuário. Todos os cartões são normalizados por partícula inicializadora, exceto em problemas de criticalidade KCODE, onde são normalizados por nêutron de fissão. Os cartões são descritos na tabela a seguir:

Tabela 6.2 – Cartões de saída do MCNP

Mnemônico			Descrição
F1:N ou	F1:P ou	F1:E	Corrente na superfície
F2:N ou	F2:P ou	F2:E	Fluxo na superfície
F4:N ou	F4:P ou	F4:E	Estimativa do comprimento do rasto no fluxo da célula
F5a:N ou	F5a:P		Fluxo em um ponto ou detector de anel
F6:N ou	F6:P ou	F6:N,P	Estimativa do comprimento do rasto da deposição de energia
F7:N			Estimativa do comprimento do rasto da deposição de energia de fissão
F8:N ou	F8:P ou	F8:E ou F8:P,E	Pulsos de energia em um detector

Os cartões existentes na tabela 6.2 representam o básico no MCNP. Para se criar mais cartões de um determinado tipo, deve-se adicionar múltiplos de 10 ao número do cartão desejado. Por exemplo: F1, F11, F21, ..., F991, todos para o cartão F1. O tipo de partícula desejada é mostrado explicitamente após os dois pontos, onde “N” representa nêutron, “P” representa fóton e “E” representa elétron.

6.9 – Exatidão e Precisão

Os resultados obtidos com o Monte Carlo representam uma média das contribuições de muitas histórias simuladas durante o curso do problema. Uma quantidade importante para o Monte Carlo é o erro estatístico ou a incerteza associada com o resultado. A importância desse erro e o seu comportamento versus o número de histórias não podem ser sobre-enfatizados porque o usuário não só ganha perspicácia na qualidade do resultado, mas também pode determinar se um cartão aparece estatisticamente bem comportado. Se um cartão não for bem comportado, o erro calculado geralmente associado com o resultado não refletirá o verdadeiro intervalo de confiança do resultado e, assim, a resposta poderia ser completamente errônea. O MCNP contém várias quantidades que ajudam o usuário a avaliar a qualidade do intervalo de confiança.

Suponha que $f(x)$ seja a função densidade de probabilidade com a história de x , onde x pode ser, por exemplo, a energia depositada em um evento aleatório. O valor verdadeiro esperado de x , $E(x)$, é encontrado através da equação a seguir:

$$E(x) = \int xf(x)dx \quad (6.19)$$

O valor médio é estimado com a equação a seguir:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (6.20)$$

em que x_i é o valor do x selecionado em $f(x)$ para a i -ésima história e N é o número de histórias calculadas para o problema.

A variância da população de valores de x é a medida da dispersão desses valores e é dada pela equação a seguir:

$$\sigma^2 = \int (x - E(x))^2 f(x)dx = E(x^2) - (E(x))^2 \quad (6.21)$$

A raiz quadrada da variância é chamada desvio padrão da população e é representada no Monte Carlo por “ S ”. Para grandes valores de N o desvio padrão é obtido através da equação a seguir:

$$S^2 = \frac{\sum_{i=1}^N 1^{(x_i - \bar{x}^2)}}{N - 1} \approx \overline{x^2} - \bar{x}^2 \quad (6.22)$$

em que:

$$\overline{x^2} = \frac{1}{N} \sum_{i=1}^N x_i^2 \quad (6.23)$$

A quantidade S é um desvio padrão estimado da população de x baseada em valores de x_i que foram amostrados recentemente. A variância estimada de \bar{x} é dada por:

$$S_{\bar{x}}^2 = \frac{S^2}{N} \quad (6.24)$$

Existe uma diferença extremamente importante entre precisão e exatidão em cálculos matemáticos. **Precisão** é a incerteza causada por flutuações estatísticas enquanto que **exatidão** é a medida do quão perto se está do valor esperado de \bar{x} . Estimativas de erro ou incerteza nos resultados de Monte Carlo se referem somente à precisão do resultado e não à sua exatidão.

A exatidão pode ser afetada por três fatores: o algoritmo de código, o modelamento do problema e o usuário.

Os fatores de código compreendem as características físicas incluídas no cálculo assim como o modelo matemático usado; as incertezas nos dados, como as seções de choque de transporte e de reação, o número de Avogadro, os pesos atômicos, etc; a qualidade da representação das seções de choque diferencial em energia e ângulo; e erros no código (“*bugs*”). Toda a física aplicável deve ser incluída em um cálculo para produzir resultados exatos. Embora as avaliações não sejam perfeitas, uma representação mais fiel dos dados deve produzir resultados mais exatos. Erros no código sempre podem acontecer porque nenhum programa é imune a esse tipo de problema. Porém, o MCNP é um código muito maduro, fortemente usado há muitos anos. Seu uso crescente durante os anos diminui a probabilidade de um erro sério de codificação.

O modelamento do problema é um fator que contribui muito frequentemente para uma diminuição na exatidão de um cálculo. Muitos cálculos produzem resultados aparentemente pobres porque o modelo da energia e da distribuição angular da fonte de radiação não é adequado. Dois outros problemas de modelamento que afetam a exatidão são a descrição geométrica e as características físicas dos materiais do problema.

A exatidão envolve também erros de usuário na entrada ou em subrotinas do MCNP. O usuário também pode abusar das técnicas de redução de discrepância de tal modo que interfira nos resultados. Uma checagem cuidadosa na entrada e na saída pode ajudar a aliviar esse problema. Fatores como eficiência de detector, redução de dados e interpretação etc, devem ser entendidos completamente e devem ser incluídos no cálculo, caso contrário a comparação não será significante.

A precisão de um resultado de Monte Carlo é afetada através de quatro escolhas controladas pelo usuário: cálculo direto versus cálculo adjunto, tipo de conta, técnicas de redução de discrepância, e número de histórias.

A escolha de um cálculo direto ou adjunto depende principalmente dos tamanhos relativos da fonte e das regiões de detecção. Partículas emitidas de uma região pequena são fáceis de lidar, enquanto que partículas transportadas para uma região pequena são de manipulação mais difícil. Como cálculos diretos transportam partículas da fonte para regiões de detector, esses cálculos são preferíveis quando a região do detector (ou cartão) é grande e a região de fonte é pequena. Reciprocamente, como cálculos adjuntos transportam partículas de volta da região do detector para a região da fonte, esses cálculos são preferíveis quando a região da fonte (ou cartão) é grande e a região de detector é pequena. O MCNP pode ser executado em modo adjunto multi-grupo.

Quanto menor a região do cartão, mais difícil é adquirir boas estimativas. A dimensionalidade do cartão é extremamente importante. Um cartão unidimensional é tipicamente 10 a 100 vezes mais fácil para se calcular do que uma conta bidimensional, que é 10 a 100 vezes mais fácil que uma conta tridimensional.

Podem ser usadas técnicas de redução de discrepância para melhorar a precisão de um determinado cartão aumentando a eficiência de cartões não zerados e diminuindo a dispersão das contagens de história não zeradas. Maiores detalhes podem ser obtidos na referência X-5 Monte Carlo Team, 2003.

6.10 – Erro Relativo Estimado

Todos os cartões padrão do MCNP são normalizados por história de partícula emitida pela fonte (exceto para alguns cálculos de criticalidade) e são mostrados no arquivo de saída com um segundo número, chamado erro relativo estimado, definido por:

$$R \equiv S_{\bar{x}} / \bar{x} \quad (6.25)$$

O erro relativo é um número conveniente porque representa a precisão estatística como um resultado fracional com relação à média estimada.

Através da combinação das equações (6.20), (6.22), (6.23) e (6.24), chega-se a:

$$R = \left[\frac{1}{N} \left(\frac{\overline{x^2}}{\bar{x}^2} - 1 \right) \right]^{1/2} = \left[\frac{\sum_{i=1}^N 1^{x_i^2}}{\left(\sum_{i=1}^N 1^{x_i} \right)^2} - \frac{1}{N} \right]^{1/2} \quad (6.26)$$

Algumas observações importantes devem ser feitas. Se todos os x_i 's forem diferentes de zero e tiverem o mesmo valor, R será zero. Se todos os x_i 's forem iguais a zero, R será definido como zero. Se apenas um dos x_i 's for diferente de zero, R se aproxima da unidade na medida em que N aumenta de valor. O limite para valores de R com x_i 's de mesmo sinal está entre zero e um. A tabela a seguir mostra a qualidade do cartão baseada no valor de R :

Tabela 6.3 – Interpretação do erro relativo

Valor de R	Qualidade do cartão
0,5 a 1	Péssimo
0,2 a 0,5	Ruim
0,1 a 0,2	Questionável
< 0,10	Geralmente confiável, exceto para detectores puntiformes
< 0,05	Geralmente confiável para detectores puntiformes

6.11 – Redução de Variância

Técnicas de redução de variância em cálculos de Monte Carlo reduzem o tempo de computação necessário para se obter resultados com precisão suficiente. É importante salientar que a precisão é somente um dos requisitos para um bom cálculo de Monte Carlo. Mesmo um cálculo com variância zerado pode não ter exatidão necessária se outras fontes de erro não forem minimizadas.

Há quatro classes de técnicas de redução de variância que vão do trivial ao sofisticado.

Métodos de **truncamento** são os mais simples. Eles aceleram os cálculos truncando partes de espaço de fase que não contribui significativamente para a solução. O exemplo mais simples é o truncamento de geometria na qual partes sem importância simplesmente

não são modeladas. Métodos de truncamento específicos disponíveis em MCNP são corte de energia e de tempo.

Métodos de **controle de população** usam divisão de partículas e roleta russa para controlar o número de amostras em várias regiões de espaço de fase. Em regiões importantes são localizadas muitas amostras de baixo peso, enquanto que em regiões sem importância são usadas poucas amostras com alto peso. Um ajuste de peso é feito para assegurar que a solução do problema permaneça imparcial. Métodos de controle de população específicos disponíveis em MCNP são divisão de geometria e roleta russa, divisão / roleta de energia, divisão / roleta de tempo, corte de peso e janelas de peso.

Métodos de **modificação de amostra** alteram a amostragem estatística de um problema para aumentar o número de cartões por partícula. A amostragem é feita de distribuições que enviam partículas em direções desejadas ou em outras regiões desejadas de espaço de fase como tempo ou energia, ou muda o local ou tipo de colisões. Os métodos de modificação de amostra em MCNP incluem transformação exponencial, captura implícita, colisões forçadas, influência da fonte, e influência de fóton nêutron-induzido.

Métodos **parcialmente determinísticos** são a classe mais complicada de métodos de redução de variância. Eles evitam o processo de caminho aleatório usado normalmente, utilizando técnicas do tipo determinísticas, como estimativas do próximo evento, ou controlando a sequência de números aleatórios. Em MCNP esses métodos incluem detectores puntiformes, DXTRAN (utilizado quando uma pequena região está sendo inadequadamente amostrada porque partículas têm probabilidade muito pequena de espalhar para aquela região), e amostragem correlata.

6.12 – Exemplos de Uso do MCNP

Exemplos simples ilustrativos

Para ilustrar o uso do MCNP, foram desenvolvidos vários arquivos de entrada para resolver um problema básico: confeccionar uma barreira de proteção para uma determinada fonte de radiação ionizante. Foi criado um “universo” e em seu interior foram

inseridos uma fonte radioativa, uma placa de material atenuante e um detector de radiação. Durante as simulações foram feitas algumas modificações no *layout* projetado.

Em todas as simulações realizadas o universo foi definido como sendo uma caixa de ar com dimensões (60 cm x 60 cm x 60 cm). Dentro dessa caixa foi inserida uma placa de material atenuante com 50 cm de altura e 50 cm de largura, tendo sua espessura variada de acordo com os requisitos das simulações. O material da placa também pode ser modificado. Um detector esférico com 1,0 cm de raio foi colocado do outro lado da placa em relação à fonte emissora de radiação ionizante, em distâncias variadas. Foi medido o fluxo médio de radiação dentro desse instrumento. Em todas as simulações foi utilizado um valor de 10^7 partículas emitidas pela fonte, o que garante uma boa estatística.

No primeiro conjunto de simulações foi planejada uma fonte monoenergética puntiforme de fótons, com energia variando de 1 MeV até 10 MeV, localizada a 1,0 cm de distância de uma placa de chumbo com 1,0 cm de espessura. Um detector de radiação foi colocado a 1,0 cm de distância da placa. O erro estatístico máximo encontrado pelo MCNP durante essas simulações foi mantido abaixo de 0,25%.

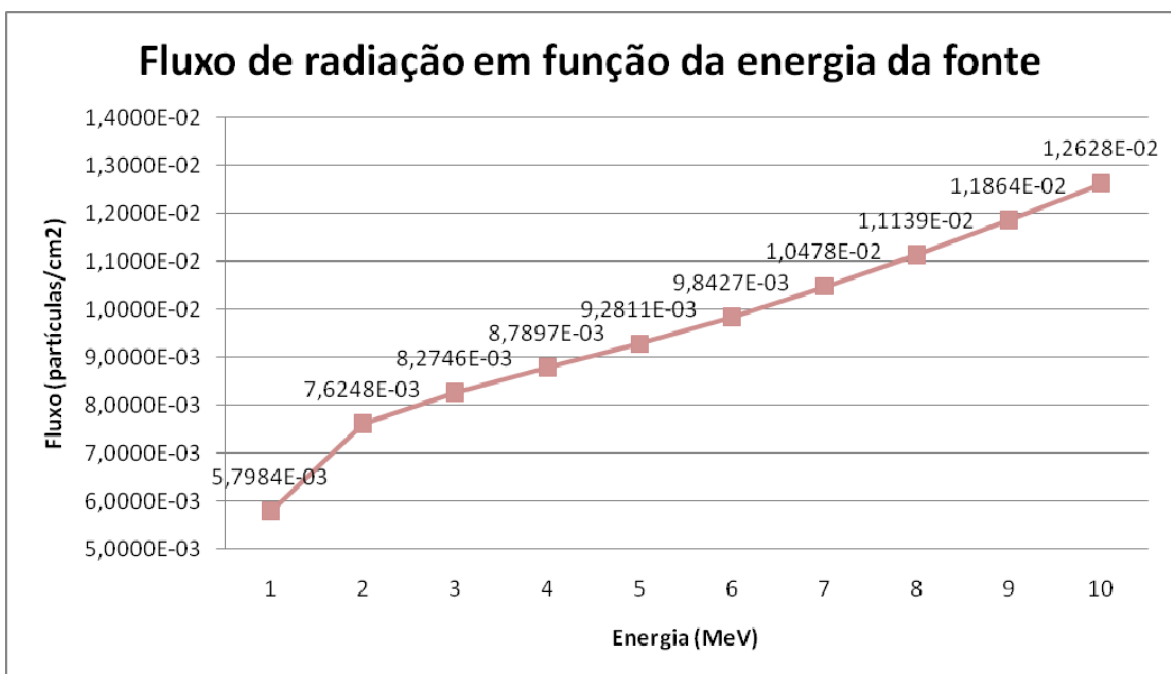


FIGURA 6.2 – Fluxo de radiação (partículas/cm²) em função da energia da fonte de fótons incidindo em uma placa de chumbo com 1,0 cm de espessura

Como esperado, quanto maior a energia da fonte, maior também o fluxo de partículas que consegue atravessar a placa de chumbo localizada entre a fonte e o detector.

No segundo conjunto de simulações foi planejada uma fonte monoenergética puntiforme de fótons, com energia fixa de 10 MeV, localizada a 1,0 cm de distância de uma placa de chumbo com espessura variando de 1,0 cm a 10,0 cm. Um detector de radiação foi colocado a 1,0 cm de distância da placa. O erro estatístico máximo encontrado pelo MCNP durante essas simulações foi mantido abaixo de 4,0%.

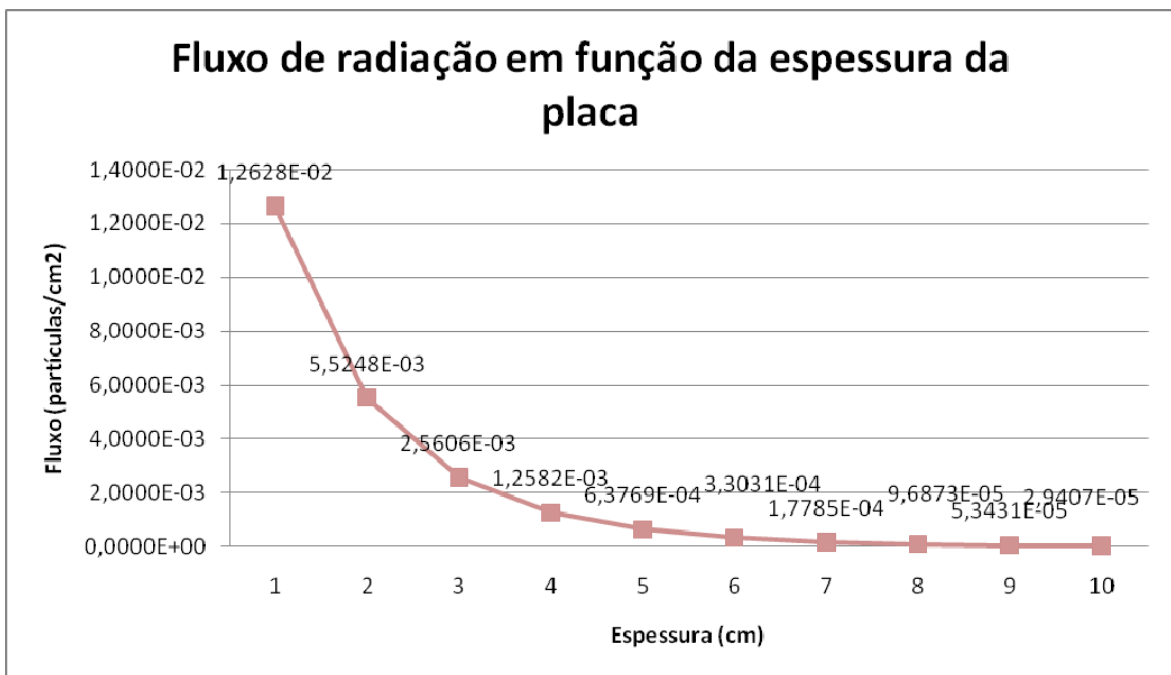


FIGURA 6.3 – Fluxo de radiação (partículas/cm²) em função da espessura da placa de chumbo

Como esperado, quanto maior a espessura da placa, menor o fluxo de partículas que consegue atravessá-la.

No terceiro conjunto de simulações foi planejada uma fonte monoenergética puntiforme de fótons, com energia fixa de 10 MeV, localizada a uma distância variável de 1,0 cm a 10,0 cm de uma placa de chumbo com espessura de 1,0 cm. Um detector de radiação foi colocado a 1,0 cm de distância da placa. O erro estatístico máximo encontrado pelo MCNP durante essas simulações foi mantido abaixo de 1,0%.

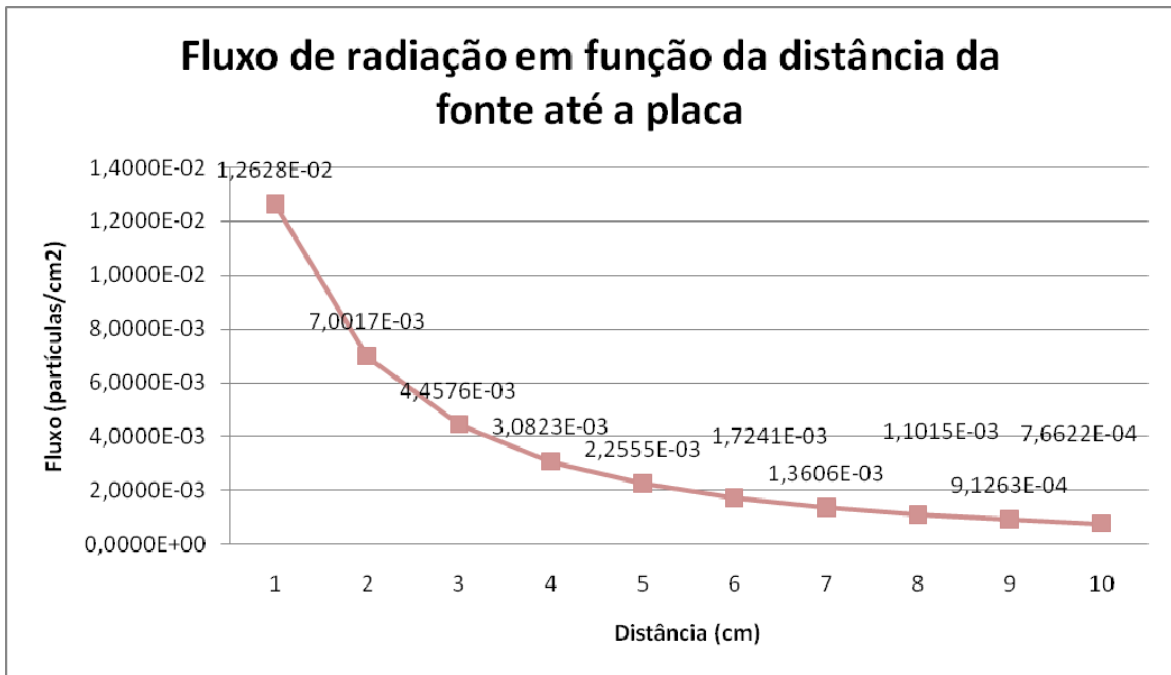


FIGURA 6.4 – Fluxo de radiação (partículas/cm²) em função da distância da fonte de radiação até a placa de chumbo

Como esperado, quanto maior a distância da fonte até placa, menor o fluxo de partículas que consegue atravessá-la.

No quarto conjunto de simulações foi planejada uma fonte monoenergética puntiforme de fótons, com energia fixa de 10 MeV, localizada a uma distância fixa de 1,0 cm de uma placa de material atenuante de composição variável com espessura de 1,0 cm. Um detector de radiação foi colocado a 1,0 cm de distância da placa. O erro estatístico máximo encontrado pelo MCNP durante essas simulações foi mantido abaixo de 0,25%.

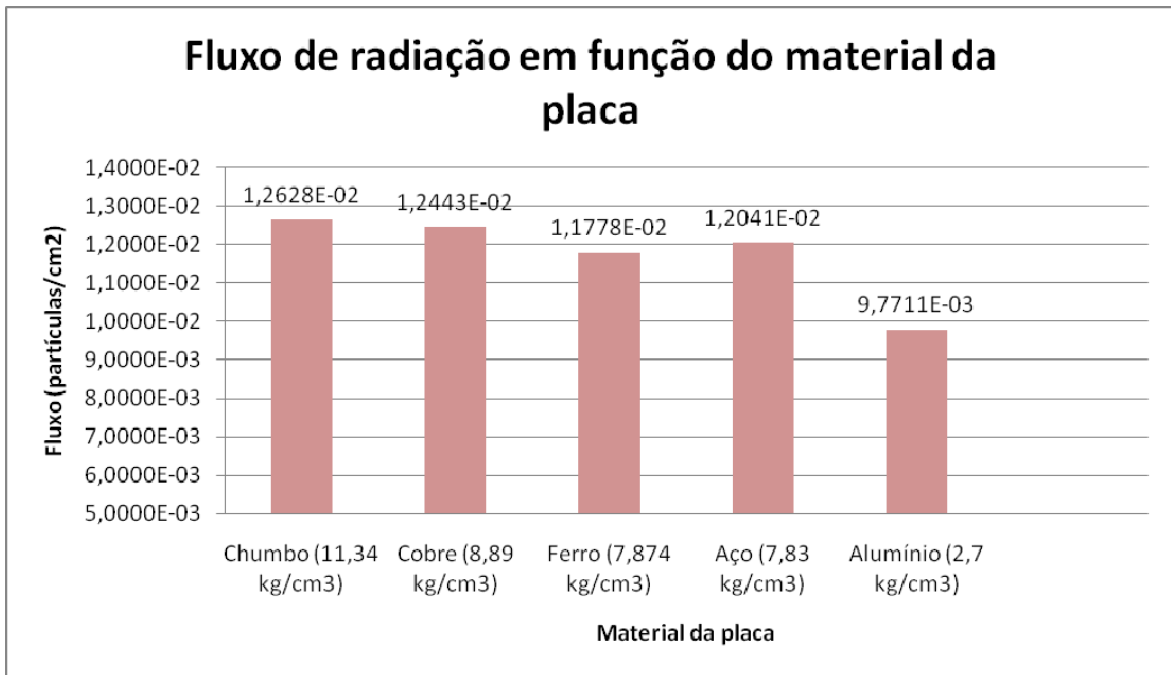


FIGURA 6.5 – Fluxo de radiação (partículas/cm²) em função do material da placa

Observa-se, a partir do gráfico anterior, uma proximidade na atenuação proporcionada por materiais de densidades de valores aproximadamente iguais.

Obtenção de curvas de PDP para radioterapia

Um exemplo mais complexo, mostrado a seguir, simula uma curva de percentagem de dose profunda (PDP) em radioterapia, para um feixe de fótons proveniente de um acelerador linear de 6 MV de energia máxima, colimadores com abertura (5 x 5) cm², distância fonte-superfície de 100,0 cm, incidindo num *phantom* (caixa de acrílico contendo água).

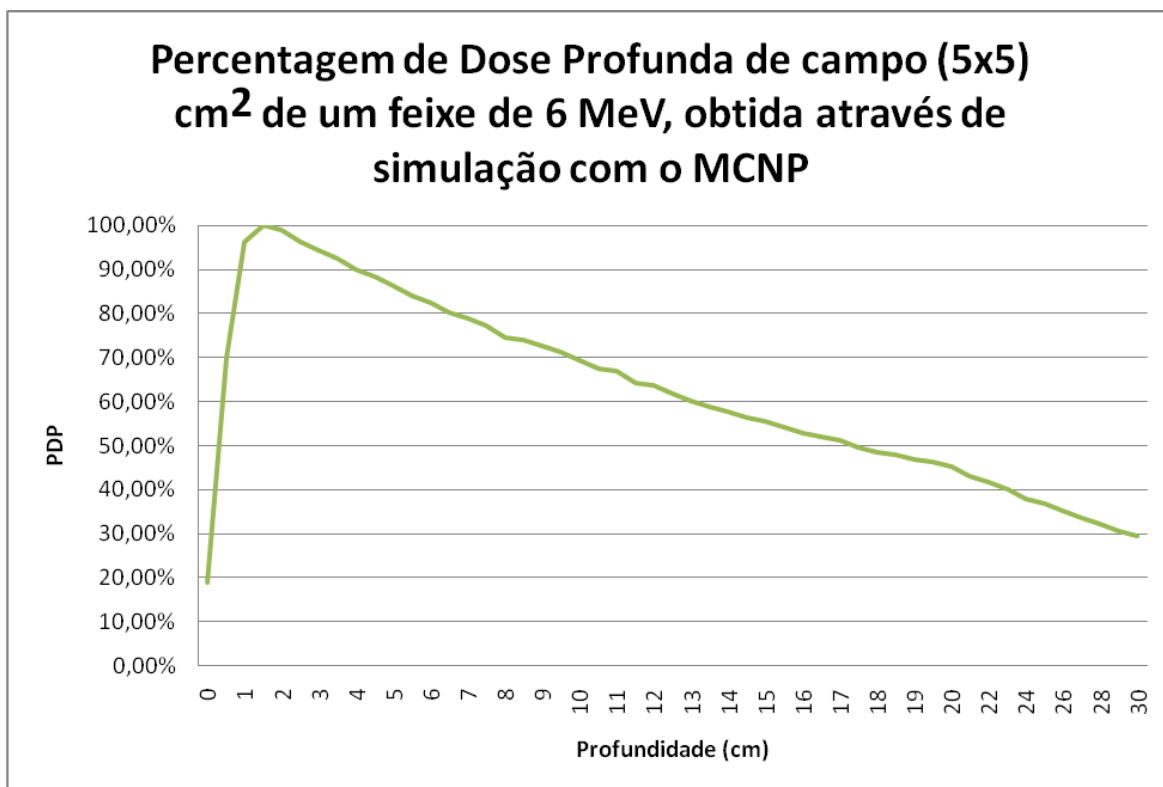


FIGURA 6.6 – Curva de PDP simulada com o MCNP

Com o uso do MCNP, pode-se descrever corretamente a geometria de qualquer equipamento utilizado em tratamentos radioterápicos, simulando a deposição de dose absorvida decorrente da irradiação de qualquer corpo quando submetido a feixes provenientes do mesmo. Pode-se simular equipamentos de telecobaltoterapia e aceleradores lineares, entre outros. O material a ser irradiado pode ser qualquer um que tenha bibliotecas de seção de choque disponíveis para o programa.

Gammacell 220

Em outra simulação mostrada a seguir a título de exemplo para uso do MCNP, foi feito um modelamento do irradiador industrial Gammacell 220, de propriedade do CDTN (Centro de Desenvolvimento de Tecnologia Nuclear). Esse equipamento foi desenvolvido pela empresa Atomic Energy of Canada Limited, atualmente parte da MDS Nordion International, para pesquisas e estudos onde se empregam grandes taxas de dose de radiação gama. Ele pode ser utilizado em uma enorme variedade de áreas de pesquisa, como irradiação de alimentos, geologia, área médica, pesquisa de materiais, efeitos da

radiação em sistemas biológicos, esterilização de organismos e solos, e alterações e melhorias nas propriedades de polímeros, entre outros (Rodrigues *et. al*, 2009).

Esse equipamento utiliza uma fonte de cobalto-60 duplamente encapsulada em tubos de aço inoxidável e alumínio, com proteções biológicas de chumbo afixadas para que o operador não seja exposto à radiação. Ele possui uma câmara de geometria cilíndrica com 20,3 cm de altura e 7,5 cm de raio para colocação de amostras. O certificado de calibração atesta uma taxa de dose de $(4,43 \pm 0,16) \times 10^5 \text{ cGy.h}^{-1}$ em 28 de agosto de 1962, no centro de exposição da câmara. A figura a seguir mostra as curvas de isodose descritas no certificado de calibração.

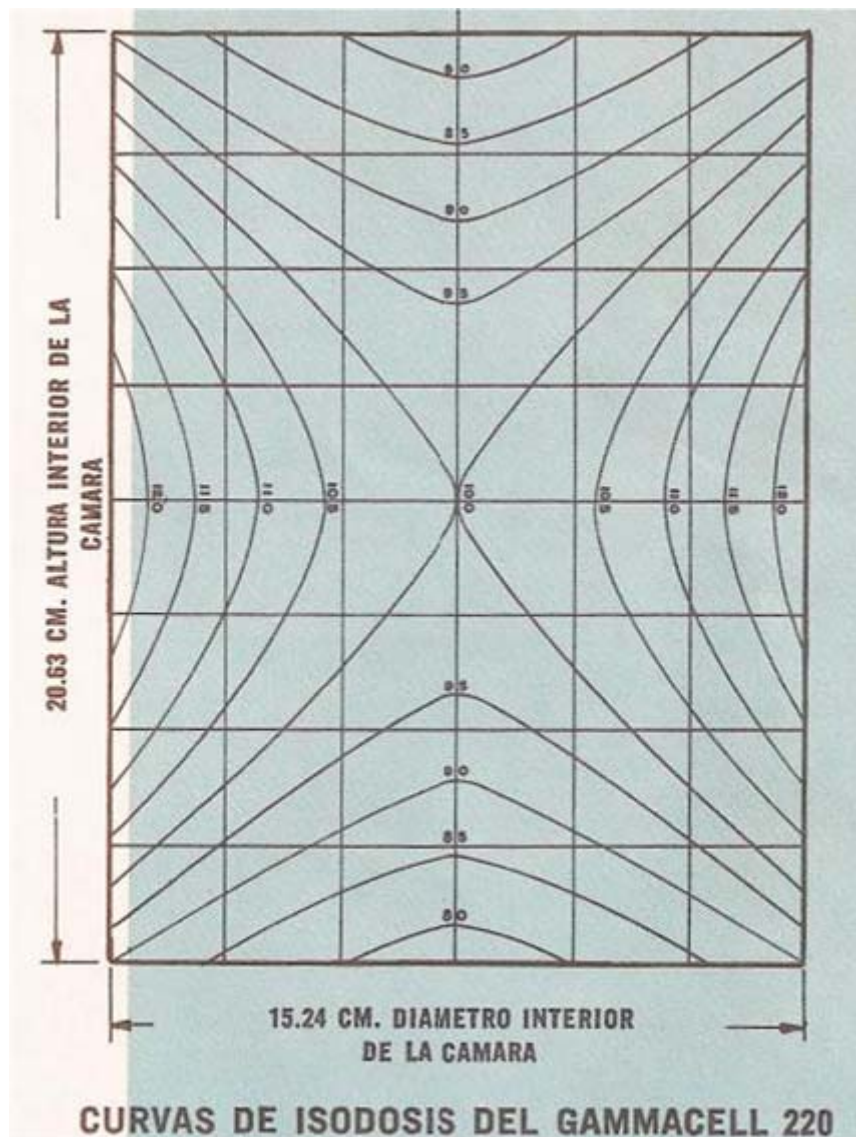


FIGURA 6.7 – Curvas de isodose mostradas no certificado de calibração do GammaCell 220 do CDTN

Para uso correto do equipamento, ele deve ser calibrado e suas curvas de isodose são usualmente medidas através do uso de dosímetros distribuídos ao longo de sua extensão. Esse procedimento consome um número elevado de dosímetros, além de muito tempo de irradiação e de muito trabalho para operadores do equipamento. O uso de métodos matemáticos e de códigos de transporte pode reduzir significativamente esse problema, embora seja necessário o uso de códigos complexos e um considerável tempo de processamento de dados em computadores. O MCNP é uma ótima opção para simulação dessas curvas.

A geometria simulada foi descrita de maneira a se reproduzir o mais fielmente possível o “*setup*” original. Todas as distâncias e todos os tamanhos utilizados na prática foram repetidos nas diversas simulações com o MCNP. O arquivo de entrada utilizado contém 240 superfícies e 90 células, com 54 fontes de cobalto-60 com 0,3594 cm de raio e 24,6276 cm de comprimento, distribuídas geometricamente da mesma forma que no equipamento real. As fontes cilíndricas foram separadas da câmara de exposição por placas de alumínio e de aço inoxidável com 0,2 cm de espessura e raio interno de 7,5 e 7,7 cm respectivamente. Uma carapaça externa de chumbo envolve o equipamento.

O decaimento radioativo foi levado em conta durante as simulações. O espectro de emissão utilizado proporcionou uma distribuição apropriada e os cartões utilizados para saída do programa mediram fluxo médio e pulsos nas células - fótons/cm² e pulsos, respectivamente. Os resultados da simulação foram validados com os experimentais obtidos através de dosimetria Fricke e TLD.

A figura a seguir mostra a geometria simulada no MCNP.

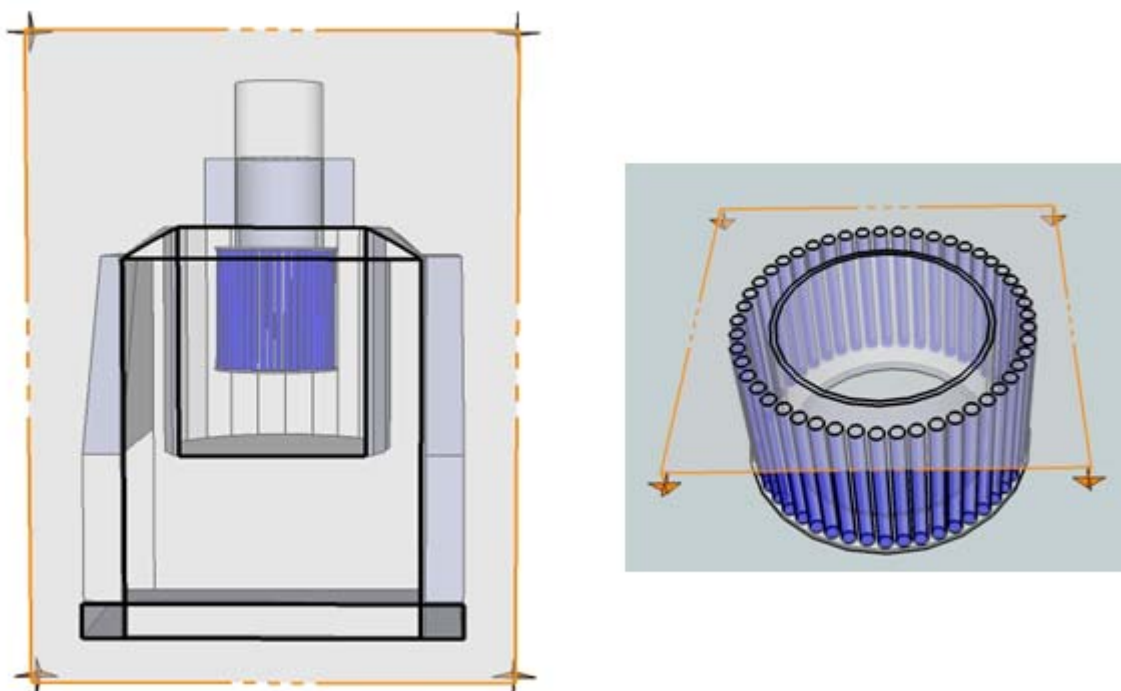


FIGURA 6.8 – Geometria do GammaCell 220 do CDTN simulada no MCNP

Foi utilizado o cartão *sdef* do MCNP para descrever a fonte, de onde os fótons são emitidos a partir do centro dos 54 cilindros. A descrição do espectro de emissão de fótons seguiu uma distribuição Maxwelliana apropriada para esse tipo de problema. Considerou-se também os dois picos de emissão de raios-gama do cobalto-60, de 1,17321 e 1,332491 MeV, com probabilidade de 0,9986 e 0,9998 respectivamente.

A parte experimental do trabalho foi feita em duas etapas. Na primeira utilizou-se 49 detectores TLD-800 acoplados em um disco cilíndrico, onde cada detector possui dimensões de aproximadamente $(3 \times 3 \times 1) \text{ mm}^3$. Esses dosímetros foram irradiados com uma dose de 4,0 Gy. Posteriormente os dosímetros foram avaliados em um leitor do tipo Harshaw 4500. Esse procedimento foi repetido 4 vezes. Na segunda etapa utilizou-se dosímetros Fricke, compostos por uma solução aquosa de ácido sulfúrico e sulfato ferroso. A irradiação desses dosímetros causa a radiólise da água e uma reação secundária, onde os íons ferrosos são oxidados em íons férricos. Posteriormente os dosímetros foram analisados em espectrômetros do tipo Shimadzu UV1240. Foram feitas 5 irradiações com doses de 50, 90, 150, 200 e 250 Gy. A seguir, utilizando-se uma geometria cilíndrica, 8 amostras foram irradiadas na extremidade do raio, 8 na metade do raio e 1 no centro da geometria, todas com dose de 160 Gy.

A simulação das curvas de isodose do GammaCell 220 foi feita utilizando-se detectores em anel com 1 cm de comprimento, 0,5 cm de espessura e raio variável, conforme tabela a seguir. Os detectores foram posicionados a meia altura na câmara de exposição e um histórico de 10^6 partículas foi utilizado nas simulações.

Tabela 6.4 – Raio do detector em anel e dose relativa

Raio do detector (cm)	Dose (%)
0,25	100
0,75	102
1,25	103
2,75	104
3,25	106
3,75	108
4,25	109
4,75	113
5,75	117
6,25	119
6,75	123

A eficácia do MCNP está intimamente relacionada com o arquivo de entrada e pode ser verificada a partir do erro relativo experimental calculado. A tabela a seguir efetua uma comparação entre o valor certificado e o obtido nas simulações.

Tabela 6.5 – Comparação entre as isodoses certificadas e simuladas

Raio do detector (cm)	Isodoses certificadas (Gy)	Dose simulada (Gy)	Erro relativo (%)
3,00	105	104	0,95
4,55	110	109	0,90
5,81	115	117	1,74
6,90	120	123	2,50

A figura a seguir mostra uma comparação entre os 4 conjuntos de resultados, obtidos a partir do certificado de calibração, da dosimetria TLD, da dosimetria Fricke e da simulação com o MCNP.

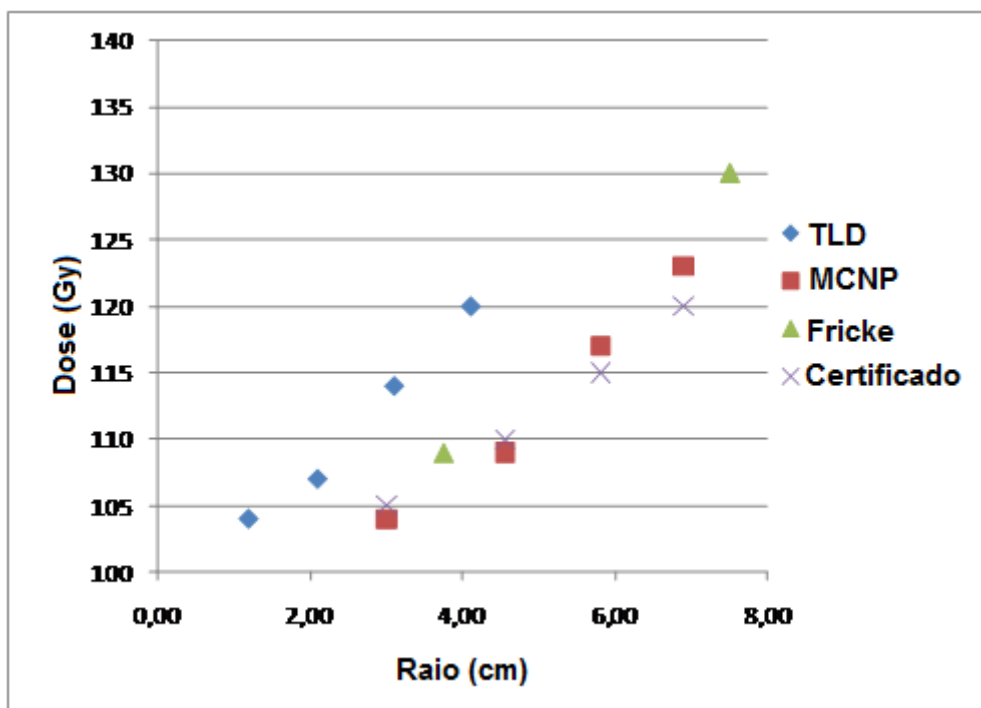


FIGURA 6.9 – Comparação entre as curvas de isodose obtidas a partir do certificado de calibração, da dosimetria TLD, da dosimetria Fricke e da simulação com o MCNP (Adaptado de Rodrigues et. al, 2009)

A dosimetria feita com TLD's e Fricke mostrou-se incoerente com os resultados simulados, com erros relativos variando entre 1 e 9%. Esses resultados também não estão de acordo com os valores certificados e novos estudos deverão ser feitos futuramente para se tentar descobrir a razão dessa diferença. Uma das prováveis razões seria uma avaria no equipamento após sua certificação. A simulação feita com o MCNP atingiu seu objetivo, pois os valores encontrados ficaram bem próximos dos certificados pelo fabricante, o que demonstra a potencialidade do método.

No CDTN, a avaliação de dose é feita geralmente a partir da irradiação de toda a massa de ar que ocupa a câmara do GammaCell 220. Essa rotina não considera peculiaridades de cada amostra, especialmente com relação a volume, capacidade e densidade, parâmetros essenciais para determinação da dose depositada. O uso do MCNP nessas simulações também pode corrigir esse problema, trazendo melhorias consideráveis no uso desse equipamento.

6.13 – Conclusão

O MCNP é um eficiente software que demonstra toda a capacidade do Método de Monte Carlo para efetuar simulações relacionadas a fontes radioativas. Uma grande dificuldade no uso do MCNP é a correta descrição da geometria no arquivo de entrada. Todo o “*setup*” tem que ser descrito corretamente, incluindo-se as superfícies, as células, os materiais, as fontes radioativas e todos os demais dados necessários para o processamento. Além disso, é necessário que existam bibliotecas de seção de choque para todos os elementos químicos utilizados. Outro fator agravante do uso do MCNP está no tempo necessário para processamento. Quanto mais complexa for a geometria e quanto mais partículas se usar para as simulações, maior será o tempo necessário para conseguir-se uma saída com estatística de erro confiável. O MCNP pode ser utilizado para se simular desde casos simples até casos muito complexos, como os exemplos descritos ao longo do capítulo.

CAPÍTULO VII

RESOLUÇÃO DE PROBLEMAS INVERSOS EM RADIOTERAPIA

7.1 – Introdução

A **dose absorvida** é a quantidade física fundamental de interesse em tratamentos radioterápicos, uma vez que ela é o principal critério para a busca da cura e para a minimização de lesões em tecidos sadios adjacentes ao tumor/leito tumoral. Mas a medida direta da dose absorvida nos tecidos irradiados não é uma tarefa trivial. **Calorímetros e câmaras de ionização** podem ser usados para a medida de dose absorvida em termos absolutos, mas não estão disponíveis para dosimetria *in vivo*. **Dosímetros termoluminescentes e diodos** podem ser colocados na superfície do paciente ou mesmo dentro de seu organismo a fim de se determinar a dose absorvida em determinados pontos, *i.e.*, nos pontos onde o feixe de radiação entra ou sai do corpo; entretanto, não é possível determinar-se a distribuição da dose absorvida ao longo da trajetória do feixe de radiação utilizando apenas esses dados. Assim, a quantificação da distribuição de dose em pacientes é feita a partir de modelos de cálculo, sendo extrapolada posteriormente para os tecidos humanos irradiados.

Métodos empíricos são limitados quanto à precisão e frequentemente falham no modelamento do tratamento. Embora a equação de transporte e as seções de choque de interação sejam bem conhecidas, nenhum algoritmo analítico de cálculo de dose para fótons de radiação ionizante foi desenvolvido de maneira que seja **suficientemente geral** para todos os planejamentos radioterápicos com a necessária eficiência (Ahnesjö, 1999). **Métodos analíticos** podem funcionar se considerarem as interações que precedem a completa absorção da energia pelo tecido irradiado ou sua saída do sistema. **Métodos de Monte Carlo** podem ser implementados para simular o processo básico em vários propósitos na área médica, inclusive em radioterapia. Infelizmente esses métodos apresentam a limitação de exigirem recursos demasiados dos computadores utilizados, o que torna essa tarefa **extremamente lenta**.

Modelos de cálculo de dose são utilizados, do ponto de vista do ambiente de um sistema de planejamento, para fornecerem resultados rápidos e confiáveis para todos os tipos relevantes de configurações para um tratamento radioterápico. Nesses modelos, são importantes vários aspectos como generalidade, flexibilidade, logística, eficiência e outros.

A base da radioterapia é o fato de que as células cancerígenas são mais sensíveis à radiação ionizante do que as células normais. Seguindo-se recomendações gerais de eficiência na liberação da dose, da ordem de 5% (ICRU, 1976), obtém-se valores de dose extremamente benéficos para o paciente, capazes de eliminar totalmente as células cancerígenas, poupando ao máximo as células normais.

Processos não-lineares dominam a interface entre o tumor e o tecido sadio, que é uma estrutura dinâmica extremamente complexa, **não existindo uma estrutura teórica completamente compreensível**. Os efeitos potenciais da manipulação desses sistemas são baseados em experiências em outras áreas da ciência, que demonstram que esses tipos de sistemas requerem formulação matemática complexa para que se consiga expressar concisamente a dinâmica das interações e se obter um claro entendimento acerca dos processos envolvidos. Entretanto, a medicina clínica não tem integrado análise teórica ao entendimento da biologia tumoral (Ringor *et. al*, 1998). Dessa forma, o progresso da oncologia clínica é prejudicado porque uma imensidão de dados gerados pela biologia molecular e por outras tecnologias não tem sido utilizada de forma coerente e compreensiva.

Devido à ausência de uma base teórica adequada, o projeto e a avaliação de terapias têm sido feitos de forma empírica e geralmente direcionados a estratégias de redução de quantidade de células neoplásicas, apesar dos resultados limitados para a maioria dos tumores sólidos, mesmo após várias décadas de intensas pesquisas. A maioria dos problemas em radioterapia é resolvida com **equações integrais**.

7.2 – Resolução com Equações Integrais

O uso de feixes de radiação eletromagnética ionizante (*i.e*, raios-x e raios-gama) para o tratamento de doenças, principalmente tumores malignos (cânceres) começou dias após a descoberta dos raios-x por Wilhelm Röntgen, em 28 de dezembro de 1895. Para tratamento de lesões em profundidade, utiliza-se amplamente feixes dessa radiação, embora, para profundidades menores, utilize-se também feixes de radiações corpusculares, como elétrons, prótons, nêutrons e outras partículas.

O planejamento de um tratamento radioterápico começa com a obtenção de imagens anatômicas tridimensionais do paciente, obtidas a partir de tomografia computadorizada (TC) e/ou de ressonância magnética nuclear (RMN). Nessa fase, é feita a determinação do volume tumoral, que engloba o tumor propriamente dito e uma margem de segurança variável, além dos tecidos normais adjacentes ao volume tumoral, que devem ser, na medida do possível, protegidos dos feixes de radiação.

Quando um feixe de radiação penetra no organismo do paciente, acontecem diversas interações entre a radiação e as partículas existentes no corpo do paciente, *i.e.*, os elétrons e núcleos constituintes das moléculas celulares. Nessa interação, há transferência de energia da radiação para as partículas do meio (paciente), que são, então, colocadas em movimento. O transporte de partículas de radiação no meio é descrito por um sistema de equações de transporte de Boltzmann. A solução desse sistema é um vetor de densidades numéricas espaciais de fase, *i.e.*, espaço posição-direção-energia. Diferentes componentes desse vetor correspondem a diferentes tipos de partículas existentes. Mesmo que o feixe de radiação utilizado seja exclusivamente de um tipo de radiação (normalmente raios-x), com a interação ocorrida no corpo do paciente, novas partículas serão colocadas em movimento, o que exigirá cuidadosos cálculos para todas. Como as interações das partículas da radiação entre si podem ser desprezadas, as equações de transporte relevantes serão lineares.

A velocidade de um feixe de radiação eletromagnética é a velocidade da luz e para um feixe de radiação corpuscular é uma fração significativa desse valor. Isso faz com que a transferência de energia e o início da movimentação das partículas do meio aconteçam num intervalo de tempo muito pequeno. As equações de transporte relevantes não possuirão derivadas no tempo.

O tratamento radioterápico é sempre **fracionado**, ou seja, a dose total é sempre dividida em intervalos temporais, de forma a maximizar os efeitos sobre o volume tumoral e minimizá-los sobre os tecidos sadios que são irradiados juntamente. A cada sessão diária de tratamento podem ser empregados diversos feixes de radiação com configurações diferentes (angulação da entrada, distância da fonte até o paciente, uso de acessórios para modificar a distribuição de dose, etc). A intenção é não sobrecarregar os tecidos normais por onde a radiação deve passar antes de chegar ao volume tumoral e nem os tecidos

normais localizados após esse, caso contrário o paciente poderá apresentar efeitos colaterais durante ou após o término do tratamento. Esses efeitos podem, inclusive, ser tão prejudiciais quanto o próprio tumor que está sendo tratado.

A **dose absorvida** é uma medida da quantidade de radiação empregada em determinado corpo, seja ela direta ou indiretamente ionizante. Para radiações indiretamente ionizantes (por exemplo, raios-x e raios gama) a energia é distribuída na matéria segundo dois passos. No primeiro passo (resultando no KERMA) a radiação indiretamente ionizante transfere sua energia como energia cinética para partículas carregadas secundárias. No segundo passo essas partículas carregadas transferem parte da sua energia cinética ao meio (resultando em dose absorvida) e outra parte é dissipada na forma de bremsstrahlung. A dose absorvida está relacionada com a quantidade de energia armazenada estocasticamente. Ela é definida como o quociente da energia média de radiação ionizante depositada na matéria por sua massa m em um volume finito V :

$$D(\text{Gy}) = \frac{d\bar{E}(\text{J})}{dm(\text{kg})} \quad (7.1)$$

O parâmetro mais observado no tratamento é a **dose total**, ou seja, a quantidade total de energia por unidade de massa onde a radiação se deposita ao longo de todo o tratamento, como resultado de excitações e ionizações das moléculas das células irradiadas. Entretanto, além da dose absorvida, outros parâmetros são importantes num tratamento radioterápico. Os efeitos biológicos causados pela radiação dependem do tipo e da energia da radiação utilizada, da taxa de dose empregada, do esquema de fracionamento, da dose diária empregada, de questões biológicas relacionadas ao paciente como sexo, idade, predisposição genética, fatores hospedeiros (outras doenças, por exemplo), nível de oxigenação do volume tumoral, etc. Na prática, entretanto, é comum considerar-se que para um determinado tipo de tratamento, sua eficácia poderá ser determinada previamente no planejamento, baseada somente na **distribuição de dose**, embora grandes mudanças na taxa de dose possam ser significativas (Sousa, 2009).

A fim de se comprovar prováveis alterações na dose absorvida quando da mudança na taxa de dose nos tratamentos teleterápicos realizados através do uso de aceleradores lineares, foi desenvolvido um trabalho experimental onde um acelerador linear CLINAC 600 C/D de 6 MV da Varian Medical Systems irradiou um *phantom* em várias profundidades, com vários tamanhos de campo e taxas de dose diferentes, sendo a dose absorvida medida em vários pontos utilizando-se câmaras de ionização.

O CLINAC 600 C/D é um acelerador linear de partículas utilizado no tratamento de doenças, principalmente tumores malignos (cânceres). Essa máquina gera um feixe de raios-x de 6 MV de energia, com campos retangulares simétricos e assimétricos. Pode ser feita terapia estacionária (onde o “*gantry*” não se move durante a irradiação) ou terapia de arco (onde o “*gantry*” se move continuamente durante a irradiação ao redor da região a ser tratada no corpo do paciente). A taxa de dose para terapia estacionária nesse equipamento é disponibilizada dentro de uma faixa que varia de 100 MU min⁻¹ a 600 MU min⁻¹ (Varian Medical Systems, 1996). Tipicamente 1 MU corresponde a uma dose de 1 cGy depositada em um *phantom*, na profundidade de dose máxima, no raio central do feixe para um campo de (10 x 10) cm² a uma distância fonte-superfície de 100,0 cm (Sousa, 2009).

A dose absorvida na água é a principal quantidade utilizada em radioterapia, pois ela descreve com bastante veracidade os efeitos da radiação em tecidos biológicos, podendo ser determinada através de vários métodos, especialmente aqueles que empregam câmaras de ionização. A Agência Internacional de Energia Atômica (IAEA) estabeleceu diversos protocolos para dosimetria de feixes de radiação ionizante. O IAEA TRS-398 é um protocolo que pode ser utilizado com a maioria dos tipos de radiação e faixas de qualidade de feixes, incluindo fótons de alta energia gerados por elétrons com energias que variam de 1 a 50 MeV. A obtenção da dose absorvida se dá a partir da leitura de corrente ou carga elétrica gerada no volume sensível de uma câmara de ionização exposta ao feixe de radiação ionizante. Além dessa leitura, utiliza-se no cálculo o fator de calibração da câmara, temperatura, pressão atmosférica, correção de polaridade do eletrômetro e correção para recombinação de íons (Sousa, 2009).

Os seguintes equipamentos foram utilizados nesse trabalho:

- Um acelerador linear CLINAC 600 C/D da Varian Medical Systems, com potencial acelerador nominal de 6 MV e qualidade do feixe ($TPR_{20,10}$) igual a 0,657;
- Um *phantom* com dimensões $(40 \times 40 \times 40) \text{ cm}^3$;
- Um eletrômetro CNMC 1100;
- Uma câmara de ionização Farmer Wellhöffer IC 70, com câmara com paredes de grafite com espessura de $0,068 \text{ g.cm}^{-2}$, a prova d'água.

Seguindo instruções do IAEA TRS-398, as medidas foram feitas respeitando-se todas as correções para as quantidades de influência. Foram irradiados campos de $(5 \times 5) \text{ cm}^2$, $(10 \times 10) \text{ cm}^2$ e $(20 \times 20) \text{ cm}^2$, nas profundidades de 1,5 cm, 5,0 cm e 15,0 cm, com 100 unidades monitoras (MU), equivalendo a 100 cGy. As taxas de dose empregadas foram 100 cGy min^{-1} , 200 cGy min^{-1} , 300 cGy min^{-1} , 400 cGy min^{-1} , 500 cGy min^{-1} e 600 cGy min^{-1} . As leituras foram feitas em uma temperatura ambiente de 22° C e com pressão atmosférica de 92,4 kPa. Os resultados são mostrados nas figuras a seguir.

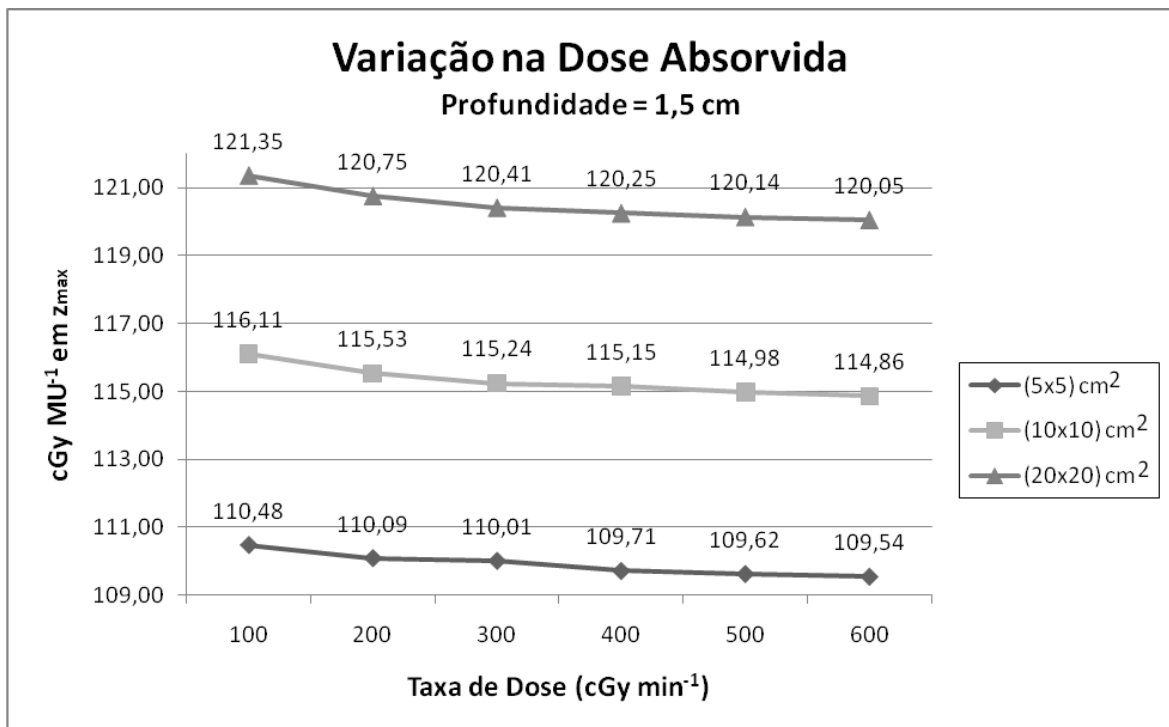


Figura 7.1 – Variação na Dose Absorvida medida na profundidade de 1,5 cm quando da mudança de taxa de dose no acelerador linear CLINAC 600 C/D (adaptado de Sousa, 2009)

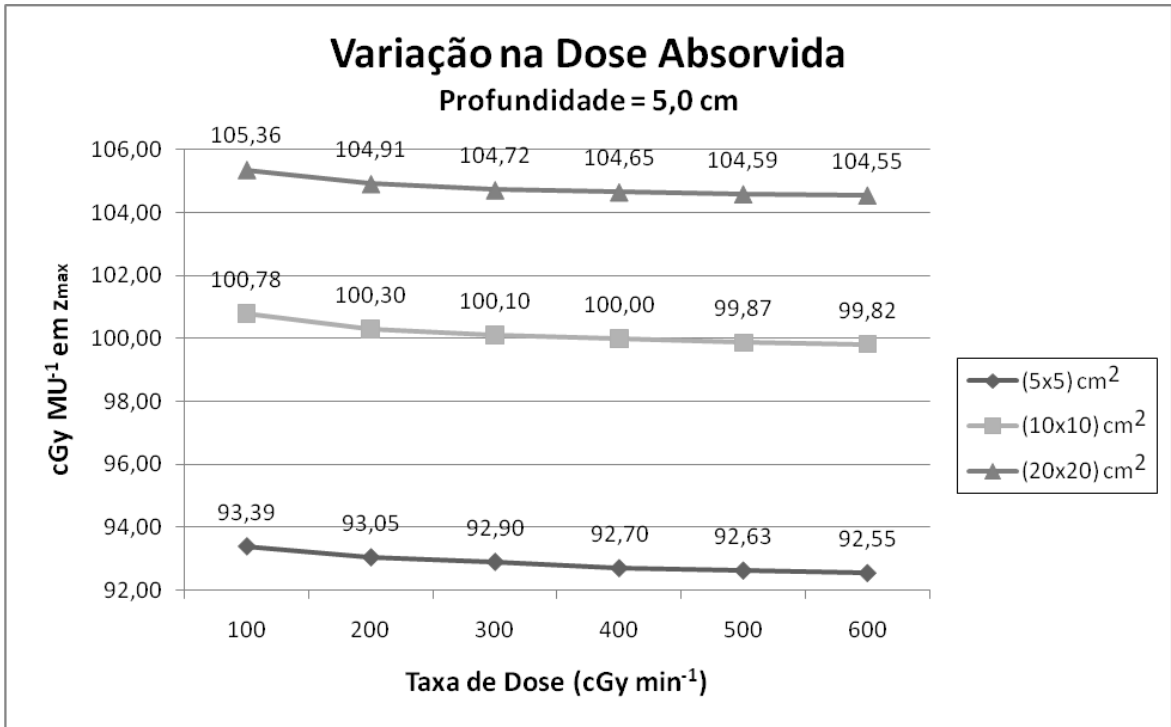


Figura 7.2 – Variação na Dose Absorvida medida na profundidade de 5,0 cm quando da mudança de taxa de dose no acelerador linear CLINAC 600 C/D (adaptado de Sousa, 2009)

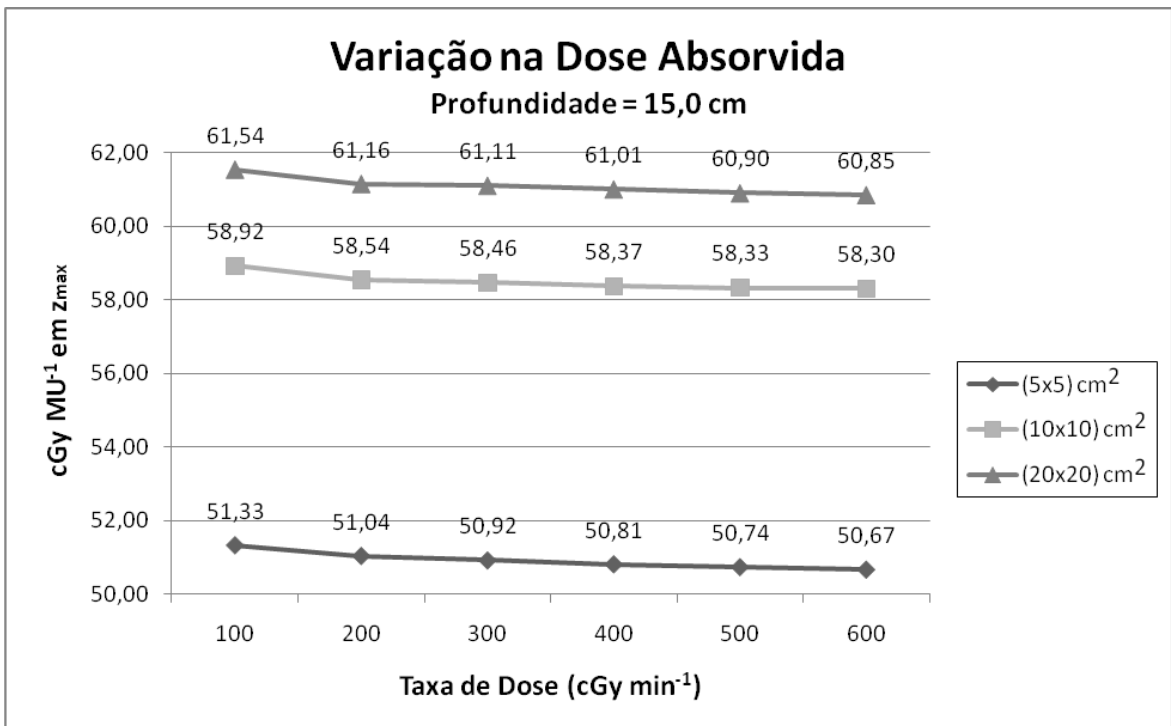


Figura 7.3 – Variação na Dose Absorvida medida na profundidade de 15,0 cm quando da mudança de taxa de dose no acelerador linear CLINAC 600 C/D (adaptado de Sousa, 2009)

Através deste trabalho, conclui-se que a variação da taxa de dose em um acelerador linear de partículas CLINAC 600 C/D, do seu valor mínimo de 100 cGy min^{-1} até o seu valor máximo de 600 cGy min^{-1} pode modificar a dose absorvida da ordem de 1% para os tamanhos de campo e profundidades estudadas (Sousa, 2009). Essa variação é muito pequena, mesmo considerando-se a sugestão do ICRU (International Commission on Radiation Units and Measurements) para variação máxima da ordem de $\pm 5\%$ (ICRU, 1976).

Outro fator muito importante relacionado à radiosensibilidade celular diz respeito à idade do indivíduo exposto, conforme demonstrado na figura a seguir.

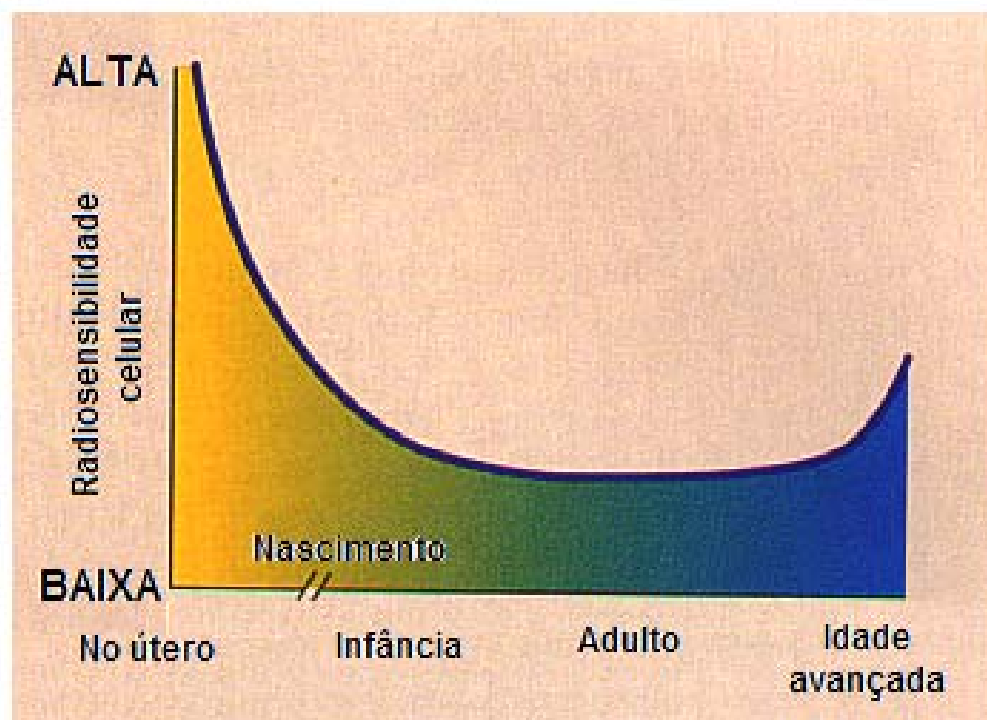


Figura 7.4 – Radiosensibilidade de acordo com a idade do indivíduo (adaptado de Bushong, 2001)

De acordo com a figura acima, observa-se que a radiosensibilidade celular é muito grande a partir da fecundação do óvulo, diminuindo rapidamente ao longo da gravidez. Após o nascimento, na medida em que o indivíduo envelhece, sua radiosensibilidade continua diminuindo até atingir o valor mínimo na fase adulta, voltando a crescer em idades muito avançadas.

O cálculo da distribuição da dose é feito, normalmente, através de **física computacional** baseada num entendimento suficiente de princípios físicos que regem a deposição de dose de radiação em tecidos biológicos. Essa distribuição de dose é feita por um **físico médico**, profissional treinado e certificado por instituição reconhecida, que utiliza sua experiência e, às vezes, intuição, para elaborar o melhor tratamento possível para cada um dos pacientes do serviço de radioterapia. Seu trabalho consiste em estimar, baseado em dados clínicos e em modelos radiobiológicos, as probabilidades de certos eventos, como erradicação do tumor, destruição parcial de tecidos saudáveis, minimização de efeitos colaterais indesejados e palição de dor como resultado da diminuição do tamanho do volume tumoral, entre outras inúmeras atividades.

O cálculo da dose é feito através de informações macroscópicas relacionadas à solução de um sistema de equações lineares de Boltzmann. Deriva-se, a seguir, uma equação para um caso especial de partículas que se movimentam através de um meio homogêneo, onde sofrem **espalhamentos** e **absorções**.

Considere uma partícula em movimento em um meio. Durante seu deslocamento, as partículas sofrerão diversas colisões (espalhamentos) em intervalos de tempo aleatórios, sendo que esses intervalos são completamente independentes uns dos outros e distribuídos exponencialmente de modo que haja um tempo médio esperado dependente da energia cinética da partícula antes da colisão. Nessas interações, acontecem mudanças de direção e de energia das partículas da radiação incidente.

Normalmente, em trabalhos sobre cálculos de dose na literatura de física médica, não se menciona explicitamente o uso de equações lineares de Boltzmann. Com os softwares utilizados atualmente para computação da dose na prática clínica, obtém-se diretamente a dose a ser empregada, sem preocupação com a solução de tais equações. Embora exista uma grande quantidade de programas, todos compartilham as mesmas idéias básicas. O feixe de radiação é tratado como um número finito de feixes estreitos (*pencil beams*), de espessura infinitesimal, de direção única e monoenergéticos. Os cálculos são feitos através de combinação de experimentos laboratoriais, simulações numéricas com métodos de Monte Carlo e análises matemáticas.

Na determinação da dose, um **problema direto** levantado refere-se ao cálculo da dose absorvida total em cada ponto do volume a ser irradiado no paciente, considerando-se todos os parâmetros de cada feixe de radiação empregado e as informações relativas aos tecidos biológicos do paciente. Tira-se, então, um **problema inverso**: dada a descrição dos tecidos do paciente, a dose prescrita para o volume tumoral e a dose máxima permissível para os tecidos sadios envolvidos no processo, calcular a configuração externa e a intensidade relativa dos feixes de radiação necessários para que haja tal distribuição de dose (Censor, 1999), cujo objetivo seria encontrar a correta configuração para tratamento do paciente nestas condições. Enquanto a distribuição de dose que resolve o problema direto é sempre obtida a partir de um feixe de radiação de intensidade específica, o problema inverso pode não ter solução, uma vez que os efeitos descritos podem não ser obtidos a partir de qualquer feixe de radiação possível.

O tratamento radioterápico de tumores malignos pode ser otimizado com feixes externos de **intensidade modulada**, onde a distribuição de dose nos tecidos do paciente é feita a partir da conformação da dose absorvida com o volume-alvo planejado, visando uma melhor proteção dos tecidos sadios adjacentes. Pesquisas na área de planejamento radioterápico inverso usualmente visam dois tópicos: como selecionar uma função objetiva que produza soluções com as propriedades desejadas e como desenvolver algoritmos interativos eficientes para se encontrar essas soluções rapidamente (Chvetsov *et. al*, 2005).

O planejamento inverso é instável quando sujeito a pequenas perturbações nos dados de entrada e no algoritmo utilizado porque é baseado em **soluções numéricas** da Equação Integral de Primeira Ordem de Fredholm, que é, notoriamente, mal-colocada. Para que um planejamento inverso seja robusto, é necessária uma técnica de **regularização** eficiente. Uma ótima técnica empregada é regularizar o problema de otimização através da adição de um termo de estabilização à função objetiva quadrada (Chvetsov *et. al*, 2005). Outra técnica interessante se baseia no uso de derivadas, pois uma derivada é uma característica universal da função, além de poder ser avaliada localmente utilizando um pequeno número de pontos adjacentes.

O planejamento inverso é um tipo de problema de transporte de radiação baseado na solução numérica da Equação Integral de Fredholm de Primeira Ordem, mostrada a seguir:

$$D(r) = \int_a^b f(E) d(r, E) dE \quad (7.2)$$

em que $D(r)$ é a distribuição desejada de dose absorvida no tecido, $f(E)$ é a função intensidade do feixe e $d(r, E)$ é a distribuição de dose de um feixe estreito de radiação ionizante.

A equação (7.2) surge naturalmente em muitos problemas da física e da matemática, devido à inversão da equação diferencial:

$$Dg(x) = f(x) \quad (7.3)$$

em que f é uma função dada e g é desconhecida. Nessa equação, D é um operador diferencial, como por exemplo:

$$D = \frac{d^2}{dx^2} \quad (7.4)$$

Um método geral para solução de tais equações é empregar as Funções de Green, através da solução da equação:

$$DK(x, y) = \delta(x - y), \quad (7.5)$$

em que $\delta(x)$ é a função delta de Dirac

A solução desejada para a equação diferencial é escrita na forma da equação (7.2), ou seja, na forma da equação integral de Fredholm. A função $K(x, y)$ é conhecida como uma Função de Green ou como o *kernel* da integral, e as variáveis x e y são dois pontos de uma superfície, como uma linha real ou um espaço Euclidiano m -dimensional.

7.3 – Modelos de *Kernel* Baseados em Convolução/Superposição

Modelos de *kernel* baseados em convolução/superposição formam uma família de modelos com raízes no mundo das imagens (Ahnesjö, 1999). A **convolução** é um operador

matemático que, a partir de duas funções, produz uma terceira, o que pode ser demonstrado matematicamente através da equação (7.6) a seguir:

$$(f * g)(t) = \int_0^t f(\tau)g(t - \tau)d\tau, \quad (7.6)$$

em que $(f * g)(t)$ representa a convolução das duas funções f e g , e t e τ são os argumentos das duas funções

Análogo à formação de imagens, a deposição de dose é vista como uma superposição de respostas com pesos apropriados (*kernels*) nos pontos irradiados (Ahnesjö, 1999). Os *kernels*, representando o transporte de energia e a deposição de dose do feixe de partículas secundárias no ponto irradiado, não são usualmente acessíveis através de medições experimentais, mas podem ser simplesmente calculados utilizando, por exemplo, códigos de transporte de partículas baseados em **métodos de Monte Carlo**.

Fótons podem viajar grandes distâncias sem sofrer interações e a energia e a direção de fótons primários é, portanto, independente do local onde ocorrem essas interações. A deposição de energia de partículas secundárias que sofreram interações com esses fótons em um meio homogêneo também independe da localização da interação primária e pode ser descrita através de um *kernel*. *Kernels* de deposição de energia são definidos como a distribuição da energia depositada aos elementos de volume (por unidade de volume) no meio, comumente água, devido à incidência de um feixe de fótons elementar na origem das coordenadas do *kernel* (Ahnesjö, 1999). *Kernels* de deposição de energia são divididos em categorias de acordo com a geometria do feixe elementar da energia incidente. Existem basicamente três geometrias diferentes de *kernels* de deposição de energia. O “*point kernel*” descreve o padrão de energia depositada em um meio infinito por uma interação de um fóton primário; o “*pencil kernel*” descreve a deposição de energia em um meio semi-infinito a partir de um feixe puntiforme monodirecional; o “*planar kernel*” descreve a energia espalhada para frente e para trás a partir das interações primárias localizadas em um plano lateralmente orientado em um feixe infinito largo. Às vezes é mencionado um quarto tipo, chamado “*rotated kernel*”, que descreve a deposição de energia devido à irradiação convergente de um *phantom* rotacionalmente simétrico (Ahnesjö, 1999). Segundo Ahnesjö *et. al* (1999), o “*pencil kernel*” é um modelo mais rápido e mais simples de ser utilizado do que o “*point kernel*” e, portanto, é utilizado como parte dos algoritmos de otimização para se encontrar modulações de feixe de formatos de

campo adequados aos tratamentos. Os três tipos mais comuns de *kernels* são mostrados na figura (7.5) a seguir.

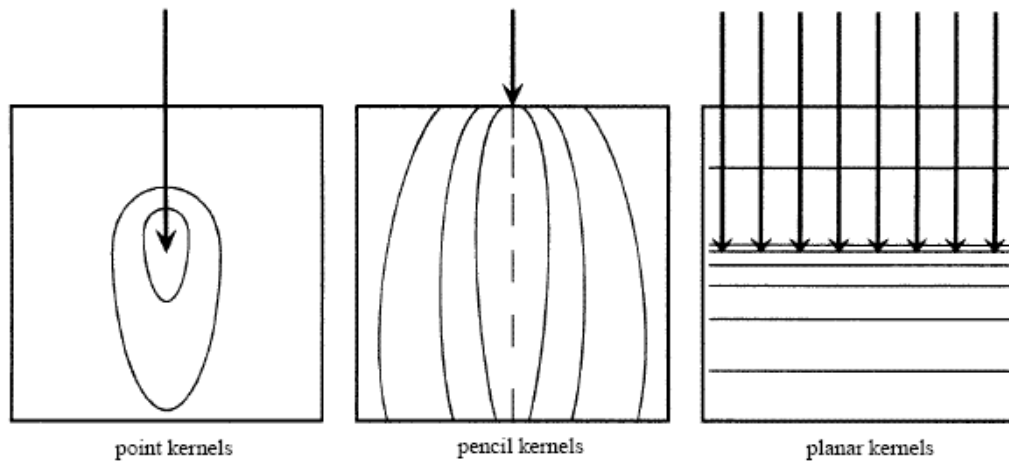


Figura 7.5 – Três tipos mais comuns de kernels segundo Ahnesjö et. al (1999)

Há várias possibilidades de normalização de *kernels*, dependendo da formulação da equação de dose da qual eles fazem parte. A aproximação comum é normalizar a energia radiante a ser distribuída pelo *kernel*, conforme a equação (7.7) a seguir:

$$h(r) = \frac{d\varepsilon}{RdV} \Rightarrow \iiint_{\infty} h(r)dV \equiv 1 \quad (7.7)$$

em que $d\varepsilon$ é a energia média depositada num elemento de volume dV devido às interações que a energia radiante R sofre antes de ser depositada como dose absorvida. A complexidade do transporte do acoplamento fóton-pósitron/elétron **limita as possibilidades para métodos analíticos** de cálculo de *kernels* (Ahnesjö, 1999). Um dos métodos padrão para cálculo de *kernels* é o **uso de simulações diretas com métodos de Monte Carlo para o transporte de partículas**, que apresenta o inconveniente de consumir **tempo demasiado de processamento em computadores**. Mesmo com o rápido e contínuo processo de modernização dos computadores, que possibilita o surgimento de máquinas cada vez mais rápidas, ainda assim a aplicação direta de superposição de *kernels* é uma tarefa **que consome muito tempo de computação**, o que impede sua total implementação na rotina clínica.

O “*pencil kernel*” descreve a energia depositada em um meio semi-infinito a partir de um feixe puntiforme monodirecional. Para o propósito de otimização de tratamento, a formulação geral a seguir descreve o problema de cálculo de dose em radioterapia:

$$D(r) = \iiint_s \iiint_E \iiint_\Omega \sum_m \Psi_{E,\Omega}^m(s) \frac{p^m}{\rho}(E, \Omega, s, r) d^2\Omega dE d^2s \quad (7.8)$$

em que $\Psi_{E,\Omega}^m(s)$ é a fluência diferencial de energia para a energia E e direção Ω para uma modalidade m de feixe, e $\frac{p^m}{\rho}(E, \Omega, s, r)$ é o “*pencil kernel*” correspondente para a deposição de energia por unidade de massa em r devido às partículas primárias que entram no corpo do paciente em s (Ahnesjö, 1999).

Para a rotina de cálculo de dose, a equação (7.8) anterior é **impraticável** uma vez que o “*pencil kernel*” é variável, devendo ser obtido através de métodos de Monte Carlo, diretamente ou através de superposição de “*point kernels*”. Como fótons têm um número limitado de interações antes de serem absorvidos, torna-se possível simular todas as suas interações diretamente, embora essa tarefa seja extremamente lenta, mesmo em computadores modernos. Para o cálculo de dose de um feixe de fótons, os livres caminhos médios entre as interações significam que a energia é transferida para o meio em frações locais dentro de grandes volumes e são necessárias simulações de grandes números de históricos para os fótons, a fim de se reduzir as incertezas estatísticas, o que torna o processo **muito demorado**.

7.4 – Funções de Dose

Diversos métodos são utilizados para se calcular a dose absorvida no corpo do paciente. Eles dependem de funções apropriadas que, por sua vez, dependem da técnica a ser utilizada quando da irradiação. Basicamente se utiliza duas técnicas na rotina clínica diária de um serviço de radioterapia: distância fonte-pele constante e isocêntrica.

Na técnica **distância fonte-pele constante**, a distância da fonte de radiação até a pele de entrada do paciente é mantida fixa, independente das outras configurações do equipamento. A cada novo campo, toda a configuração do equipamento deve ser alterada,

de maneira que a distância seja ajustada ao valor fixo, que varia de máquina para máquina. Nas atuais, essa distância costuma ser de 100 cm.

Já na técnica **isocêntrica**, a distância da fonte até a pele de entrada do paciente varia a cada campo; a única distância que é mantida constante é aquela chamada fonte-isocentro. **Isocentro** é um ponto imaginário que fica sempre à mesma distância da fonte de radiação, independentemente das configurações do aparelho, ao redor do qual toda a parte móvel gira. Na técnica isocêntrica, esse ponto imaginário é usualmente colocado no centro da região a ser irradiada, de maneira tal que, ao girarmos qualquer componente do aparelho, esse ponto não é modificado.

Na técnica distância fonte-pele constante é comum se utilizar as funções de dose intituladas **percentagem de dose profunda (PDP)** e **razão tecido-ar (TAR)**. Entretanto, essas funções apresentam limitações. A PDP, por exemplo, exige que a distância fonte-superfície seja sempre a mesma, o que impede o seu uso nas técnicas isocêntricas de tratamento, que são mais rápidas e seguras do que aquelas onde a distância fonte-pele é mantida constante. Já a TAR apresenta limitação quanto à energia, pois só pode ser utilizada para energias até da ordem do ^{60}Co (Khan, 1994).

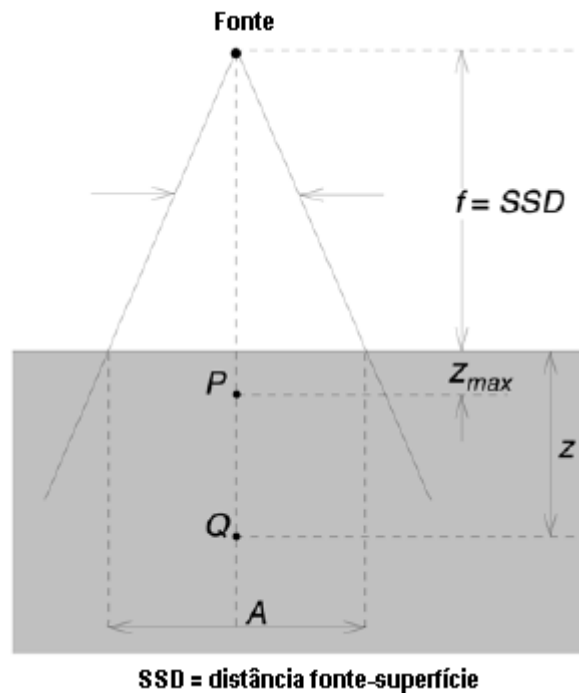


Figura 7.6 – Geometria para medida do PDP (IAEA, 2003)

Na figura (7.6) anterior, a região mais escura representa um *phantom*, P é o ponto de referência, Q é o ponto onde se deseja medir a PDP, A é o tamanho do campo, z_{max} é a profundidade de dose máxima, z é a profundidade do ponto Q e f é a distância da fonte até a superfície do *phantom*.

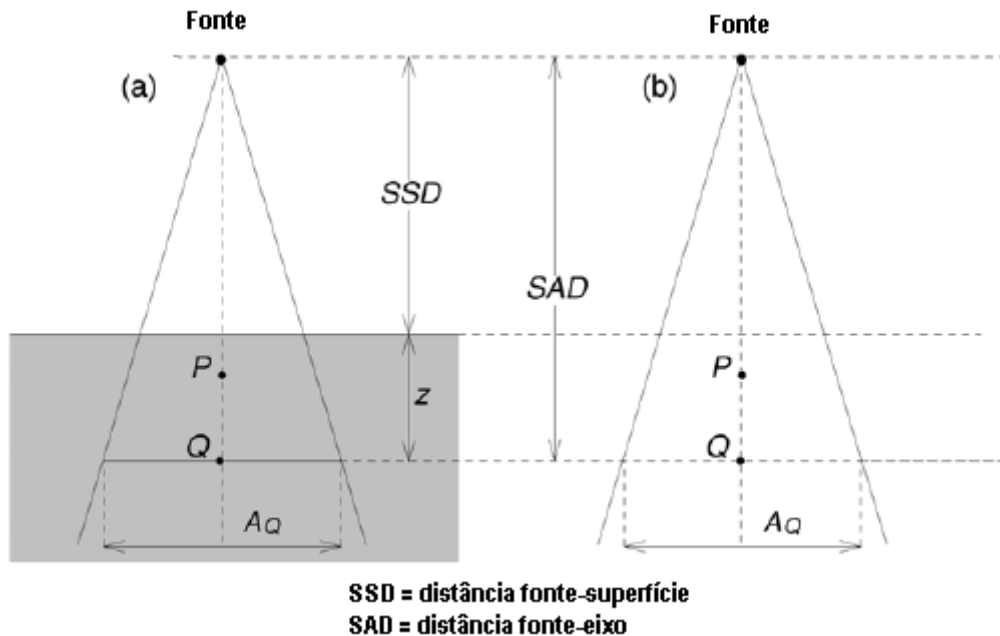


Figura 7.7 – Geometria para medida do TAR (IAEA, 2003)

Na figura (7.7) acima, a região mais escura representa um *phantom*, P é o ponto de referência, Q é o ponto onde se deseja medir o TAR, A_Q é o tamanho do campo na profundidade z , z é a profundidade do ponto Q , SSD é a distância da fonte até a superfície do *phantom* e SAD é a distância da fonte até o isocentro. Na figura da direita, o *phantom* foi retirado.

A fim de se superar limitações, foi introduzido o conceito de TPR (**razão tecido-phantom**). Essa função é similar à TAR, mas, diferentemente dessa onde as medidas são feitas no ar, no caso do TPR as medidas são feitas em um *phantom* ou objeto simulador. Um caso especial de TPR é definido para a profundidade de referência onde a dose absorvida tem seu valor máximo e é referido como TMR (**razão tecido-máximo**) (IAEA, 2003, Hendee *et. al*, 1996; Khan, 1994).

O TPR é definido como a razão entre a dose em um dado ponto no *phantom* e a dose no mesmo ponto a uma profundidade de referência fixa, usualmente 5 cm (IAEA, 2003, Hendee *et. al*, 1996; Khan, 1994). O TPR é uma função geral que pode ser normalizada em qualquer profundidade de referência. O ponto de dose máxima, localizado na direção do raio central do feixe de radiação, pode ser utilizado a fim de se simplificar o cálculo da dose em sistemas computacionais. Caso esse ponto seja adotado como referência fixa, o TPR dará origem ao TMR. Assim, TMR é um caso particular e especial do TPR, que pode ser definido como a razão da dose num dado ponto no *phantom* e a dose no mesmo ponto na profundidade de dose máxima.

A figura (7.8) a seguir mostra a geometria para obtenção da função TPR. Para a obtenção do TMR, a única diferença é o fato de que a profundidade de referência z_{ref} será igual à profundidade de dose máxima z_{max} .

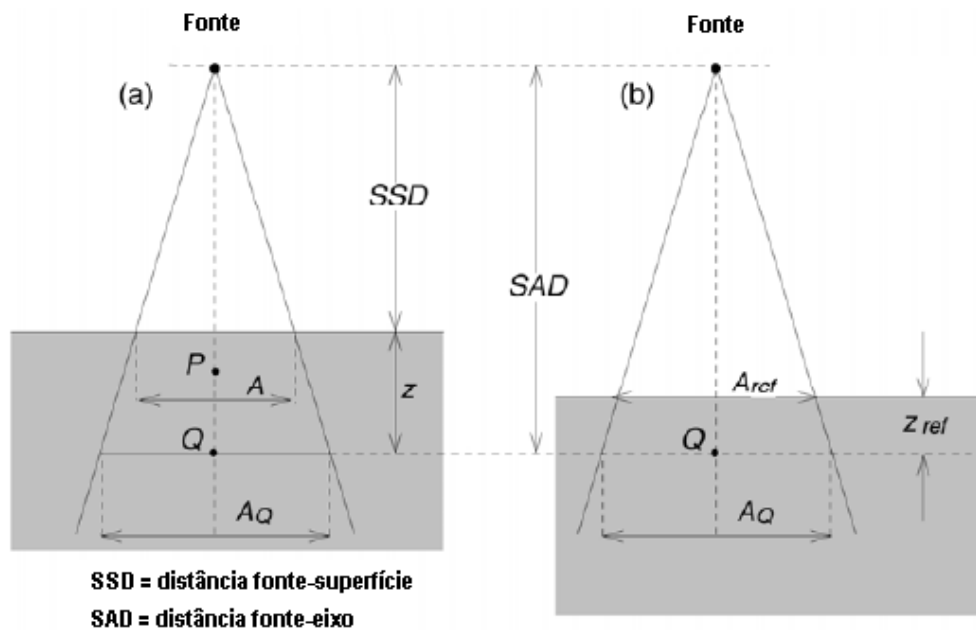


Figura 7.8 – Geometria para medição do TPR (IAEA, 2003)

Na figura (7.8) anterior, a região mais escura representa um *phantom*, P é o ponto de referência, Q é o ponto onde se deseja medir a função, A é o tamanho do campo na superfície do *phantom*, A_Q é o tamanho do campo na profundidade z , A_{ref} é o tamanho do campo na profundidade de referência, z é a profundidade do ponto Q , z_{ref} é a profundidade de referência, SSD é a distância da fonte até a superfície do *phantom* e SAD é a distância da fonte até o isocentro.

A equação (7.9) a seguir define matematicamente a função PDP:

$$PDP(z, A, f, h\nu) = 100 \frac{D_Q}{D_P} = 100 \frac{\dot{D}_Q}{\dot{D}_P} \quad (7.9)$$

em que z é a profundidade no *phantom*, A é a área do campo na superfície, f é a distância da fonte até a superfície, $h\nu$ é a energia do feixe, D_Q e \dot{D}_Q são respectivamente a dose e a taxa de dose na profundidade z no *phantom* na direção do raio central, D_P e \dot{D}_P são respectivamente a dose e a taxa de dose no ponto de dose máxima no *phantom* na direção do raio central.

A equação (7.10) a seguir define matematicamente a função TPR:

$$TPR(z, A_Q, h\nu) = \frac{D_Q}{D_{Q_{ref}}} = \frac{\dot{D}_Q}{\dot{D}_{Q_{ref}}} \quad (7.10)$$

em que z é a profundidade no *phantom*, A_Q é a área do campo no ponto Q arbitrário, $h\nu$ é a energia do feixe, D_Q e \dot{D}_Q são respectivamente a dose e a taxa de dose no ponto Q arbitrário no *phantom* na direção do raio central, $D_{Q_{ref}}$ e $\dot{D}_{Q_{ref}}$ são respectivamente a dose e a taxa de dose no ponto localizado na profundidade de referência (tipicamente 5 ou 10 cm) no *phantom* na direção do raio central.

O conceito de TMR é baseado na suposição de que a contribuição fracional do espalhamento para dose em profundidade num ponto é independente da divergência do feixe e depende somente do tamanho do campo na profundidade do ponto e da profundidade dos tecidos adjacentes (IAEA, 2003, Hendee *et. al*, 1996; Khan, 1994). Esse princípio torna essa função praticamente independente da distância fonte-superfície.

A equação (7.11) a seguir define matematicamente a função TMR:

$$TMR(z, A_Q, h\nu) = \frac{D_Q}{D_{Q_{max}}} = \frac{\dot{D}_Q}{\dot{D}_{Q_{max}}} \quad (7.11)$$

em que z é a profundidade no *phantom*, A_Q é a área do campo no ponto Q arbitrário, $h\nu$ é a energia do feixe, D_Q e \dot{D}_Q são respectivamente a dose e a taxa de dose no ponto Q arbitrário no *phantom* na direção do raio central, $D_{Q_{max}}$ e $\dot{D}_{Q_{max}}$ são respectivamente a dose e a taxa de dose no ponto localizado na profundidade de máximo (varia de acordo com a energia dos fótons) no *phantom* na direção do raio central.

As funções TPR e TMR dependem de três parâmetros: profundidade (z), área do campo na profundidade (A_Q) e energia do feixe ($h\nu$). Elas não dependem da distância fonte-pele nem da distância fonte-isocentro (Khan, 1994).

7.5 – Algoritmos Analíticos

A dose depositada em um volume infinitesimal centrado no ponto r para um campo de irradiação convergente num ponto r' no volume V para uma dada energia ou qualidade de radiação e configuração do feixe incidente é descrita pela função $h(r, r')$ (Lind, 1990). Essa função é chamada *kernel* elementar de dose e é mostrada na figura (7.9) a seguir:

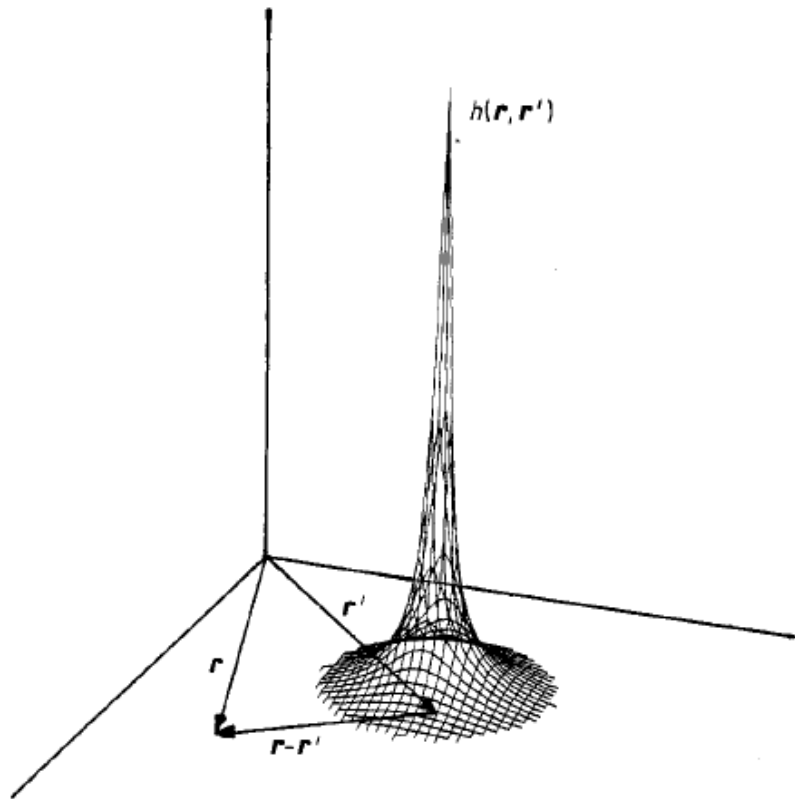


Figura 7.9 – Kernel elementar de dose $h(r, r')$ (Lind, 1990)

A contribuição de dose para todos os pontos r' irradiados pode ser adicionada, assumindo uma função de peso ou densidade de irradiação $f(r')$ a ser aplicada a cada um dos *kernels* de distribuição de dose elementar. A dose absorvida total nos pontos r de um volume V a partir dos *kernels* de dose h aplicados com uma densidade f é descrita pela equação (7.12) a seguir:

$$d(r) = \int_V f(r')h(r, r')d^3 r' \quad (7.12)$$

A equação (7.12) anterior é, claramente, uma **equação integral de Fredholm de Primeira Ordem** e é muito útil no modelamento de numerosos problemas em todos os assuntos tratados pela Física Matemática. Assumindo-se nessa equação que f e h são funções conhecidas, a determinação de d é chamada de **problema direto**, enquanto que a solução para f ou para h , conhecendo-se d e h ou d e f , é chamada de **problema inverso**. A solução do **problema direto** é um planejamento de dose convencional, corriqueiro, baseado em aproximações de convolução. Ao tentar-se otimizar um planejamento de dose, é necessário considerar as dificuldades matemáticas envolvidas na resolução de **problemas**

inversos, principalmente por serem mal-colocados, o que faz com que pequenas perturbações influenciem muito na resolução.

Uma pequena aproximação simplifica muito a equação (7.12). Assume-se que o campo de irradiação e o seu *kernel* de dose $h(r,r')$ sejam independentes da posição do ponto de convergência do campo de irradiação centrado em r' dentro do volume V , *i.e.*, $h = h(r - r')$. Assim, a equação (7.12) considera a forma da equação (7.13) a seguir, que é de convolução direta. Os erros introduzidos por essa aproximação são, na maioria dos casos, muito pequenos e podem ser quase completamente corrigidos utilizando-se técnicas apropriadas.

$$d(r) = \int_V f(r')h(r - r')d^3r' \quad (7.13)$$

Para solução da equação (7.13) anterior em f , algumas restrições físicas devem ser impostas: a integral de $f(r)$ em um volume finito deve ser limitada e todas as projeções ponderadas de $f(r)$ de uma fonte de radiação efetiva devem ser não-negativas devido à natureza física da dose absorvida e fluência como energia distribuída pela radiação ionizante por unidade de massa (Ling, 1990). Essas condições não são simples e exigem o uso de métodos matemáticos mais elaborados a fim de se resolver a equação, como expansão em valores singulares, técnicas de regularização e métodos iterativos.

Dada a **complexidade** de resolução da equação (7.13), propõe-se nesse momento a resolução do problema de deposição de dose em radioterapia através de um **kernel gaussiano** na equação de Fredholm de Primeira Ordem apresentada em (2.1):

$$\int_a^b K(x, y)f(y)dy = g(x) \quad (2.1)$$

Os gráficos a seguir mostram a deposição de dose em profundidade (PDP) obtida através de programa computacional onde se propõe um *kernel* gaussiano interno para resolução do problema direto de deposição de dose em radioterapia (Sousa *et. al*, 2008a; Sousa *et. al*, 2008b):

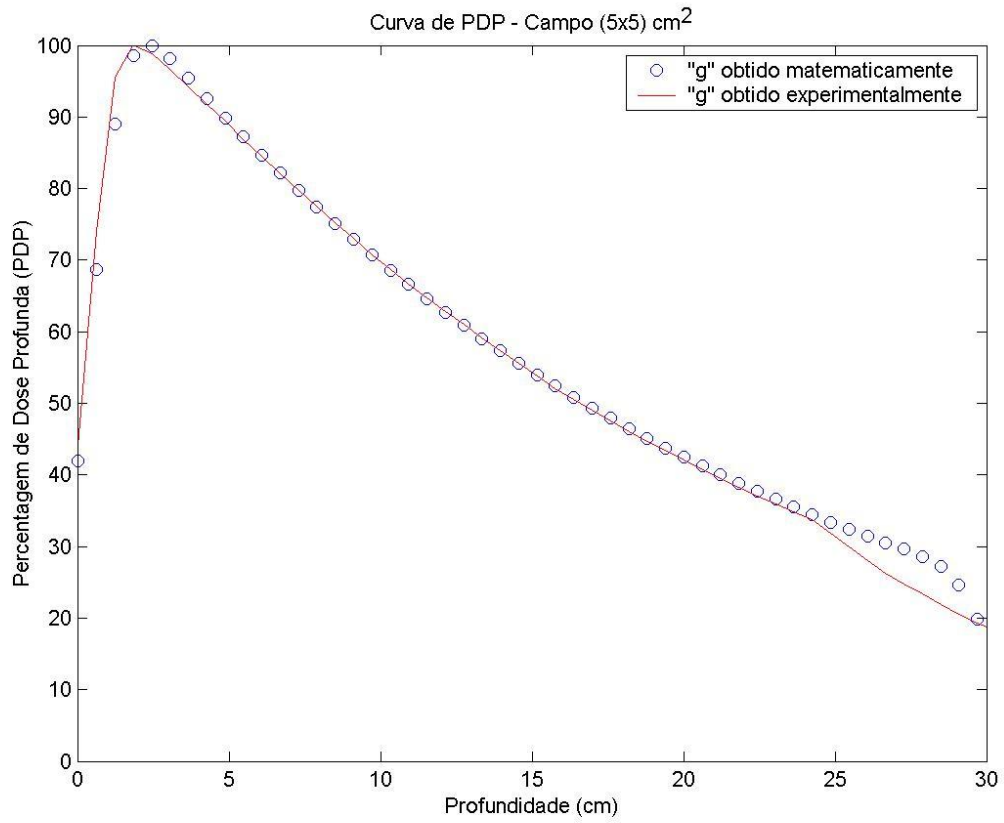


Figura 7.10 – PDP calculada utilizando kernel proposto X PDP obtida experimentalmente – Campo (5x5) cm²; distância fonte-superfície = 100,0 cm

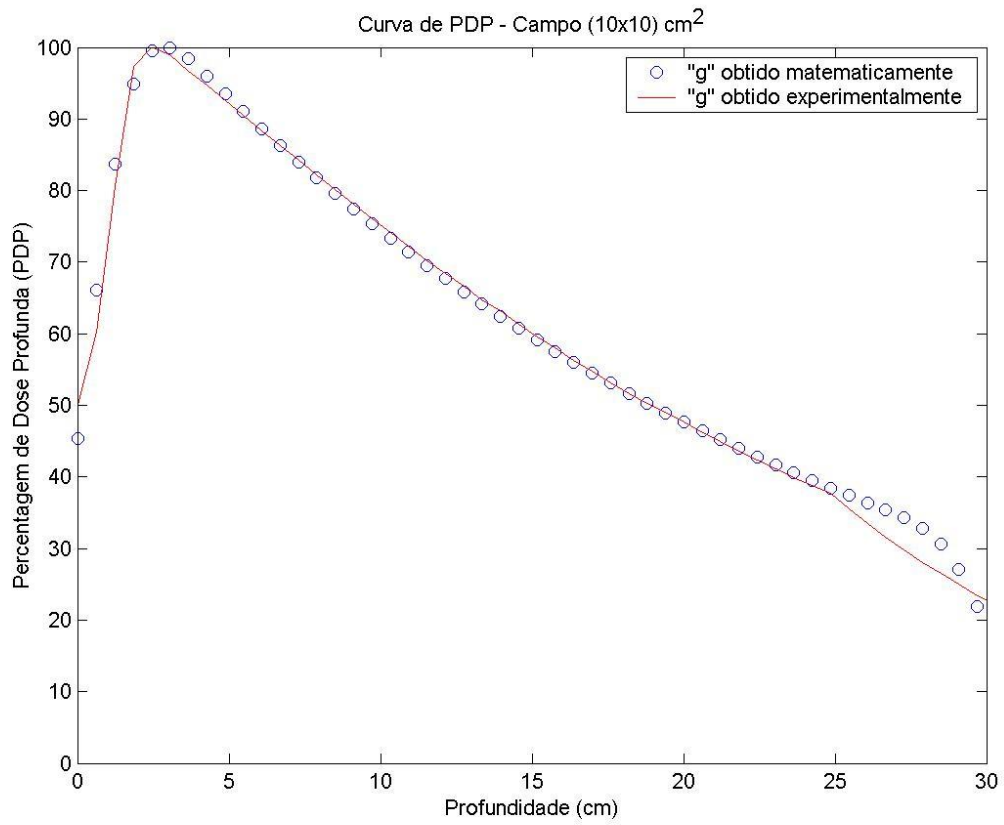


Figura 7.11 – PDP calculada utilizando kernel proposto X PDP obtida experimentalmente – Campo (10x10) cm²; distância fonte-superfície = 100,0 cm

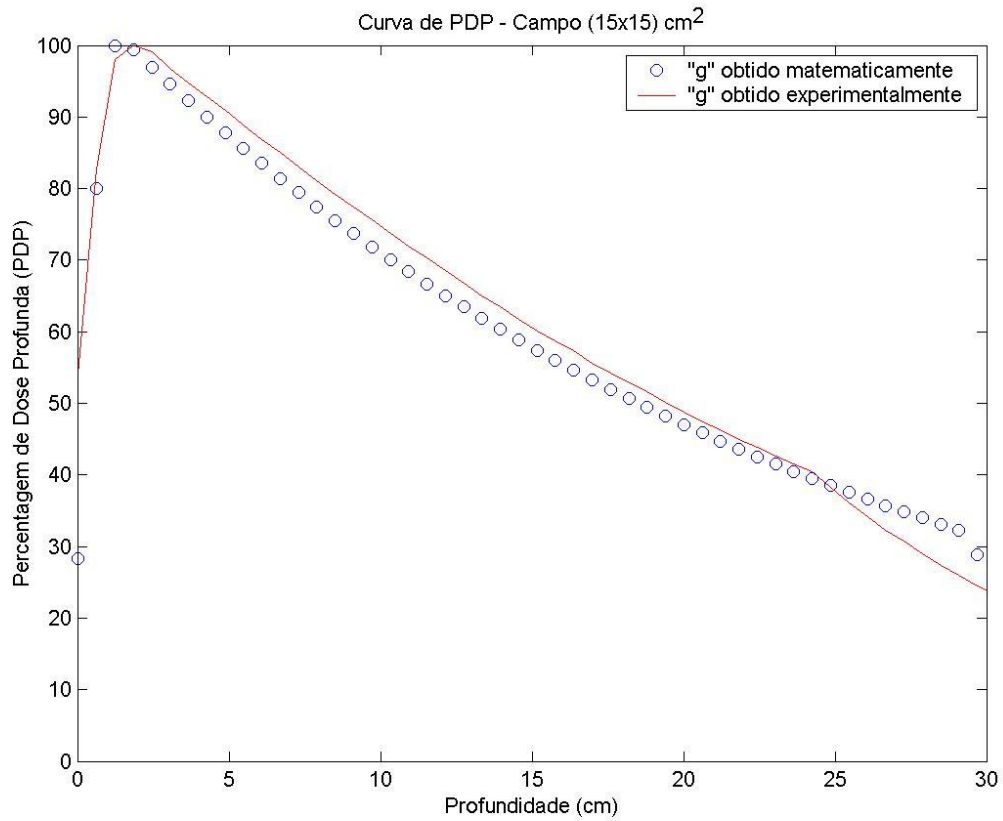


Figura 7.12 – PDP calculada utilizando kernel proposto X PDP obtida experimentalmente – Campo (15x15) cm²; distância fonte-superfície = 100,0 cm

O gráfico a seguir, reproduzido a partir das pastas de cálculo do Departamento de Física Médica do Serviço de Radioterapia do Instituto de Oncologia do Hospital São João de Deus, localizado em Divinópolis-MG, mostra curvas características de PDP:

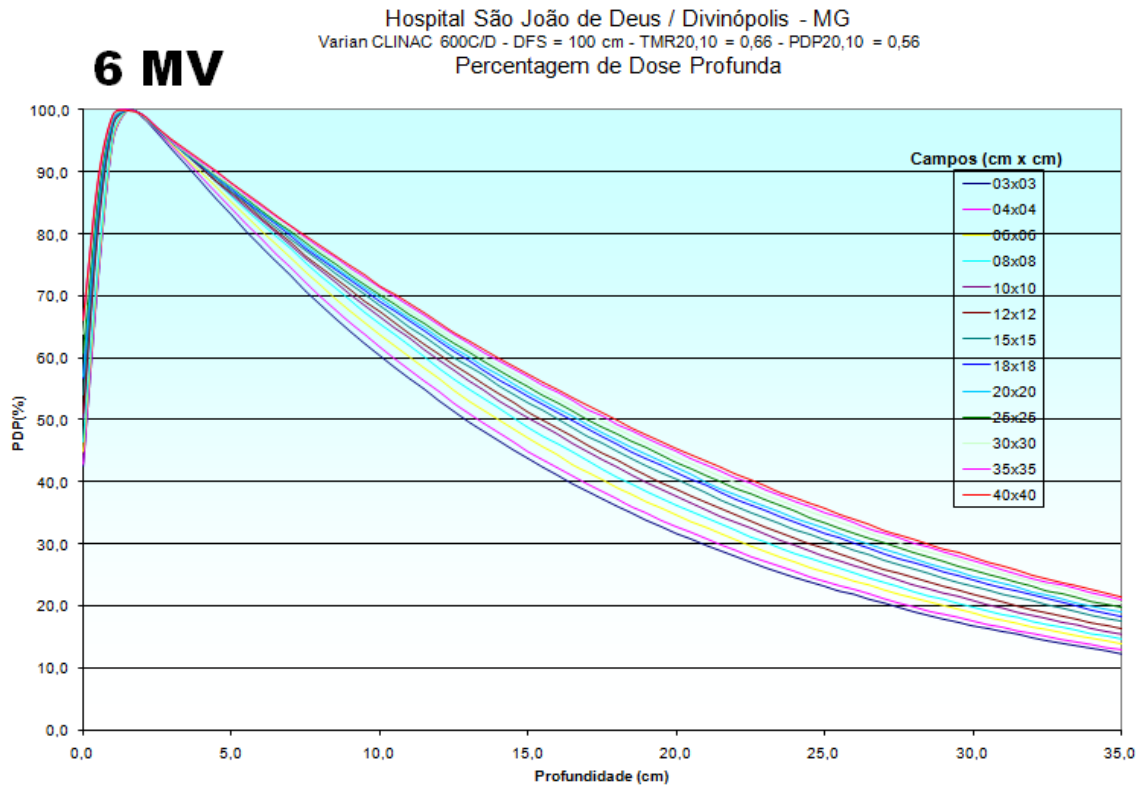


Figura 7.13 – PDPs obtidas experimentalmente; distância fonte-superfície = 100,0 cm

Comparando-se a figura (7.13) com as figuras (7.10), (7.11) e (7.12), observa-se que a resolução do problema direto de deposição de dose em radioterapia através dos programas elaborados apresenta resultados bastante parecidos. No programa utilizado para a construção dos gráficos mostrados nas figuras (7.10), (7.11) e (7.12) foi resolvido um problema direto, ou seja, a equação (2.1) foi resolvida para $g(x)$. É importante salientar que o valor de $g(x)$ é obtido **experimentalmente**. Assim sendo, a equação (2.1) deve ser resolvida através de técnicas de resolução de problemas inversos, uma vez que o valor que normalmente é procurado é o $f(y)$, que representa a intensidade do feixe de radiação. O cálculo desse valor é muito importante porque a distribuição de dose é altamente dependente da quantidade de radiação empregada. Caso se tenha a intensidade da radiação e a função de distribuição de dose, é possível o uso de técnicas matemáticas a fim de se encontrar o *kernel* da função.

Os gráficos a seguir mostram uma comparação entre os valores de $f(y)$ obtidos diretamente e através de técnicas de problemas inversos onde se empregou a regularização de Tikhonov (o segundo pode ser melhorado otimizando-se o parâmetro de regularização) (Sousa *et. al*, 2008a; Sousa *et. al*, 2008b):

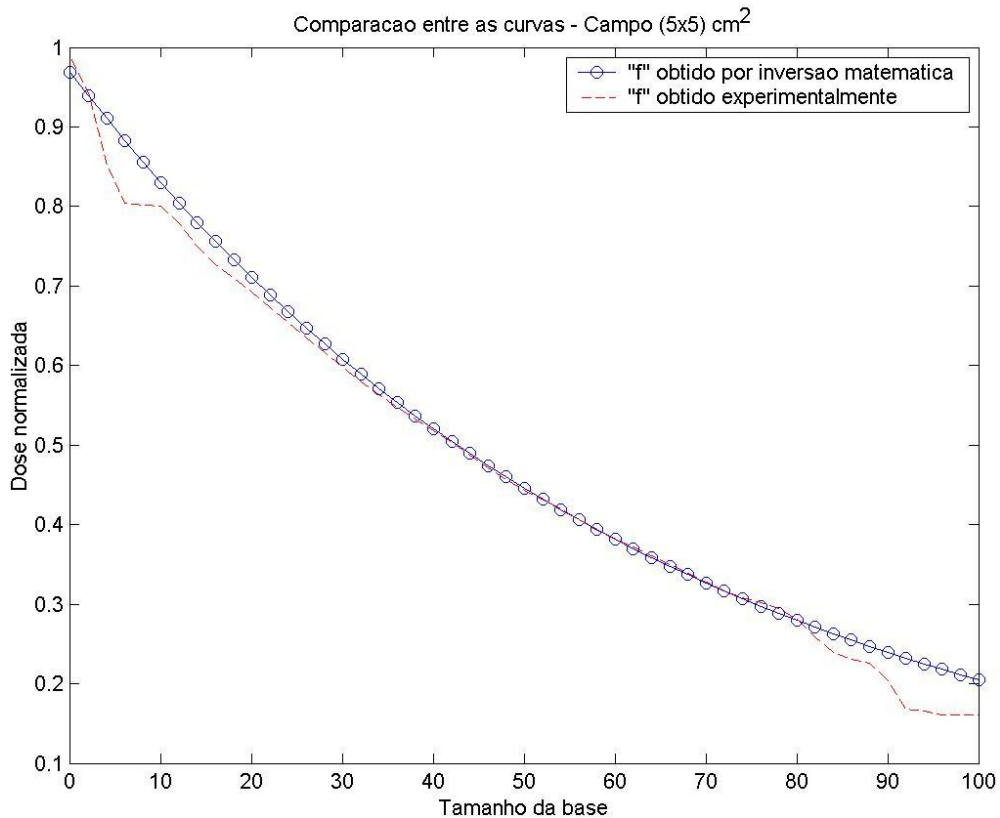


Figura 7.14 – PDP obtida a partir da inversão de dados experimentais via regularização de Tikhonov X PDP obtida experimentalmente – Campo (5x5) cm²; distância fonte-superfície = 100,0 cm

Na figura acima, nota-se que o comportamento da função matemática utilizada nos programas de simulação é satisfatório para a maior parte da curva. Entretanto, tanto no início quanto no final da curva, nota-se variações consideráveis, o que possivelmente poderá ser minimizado melhorando-se o *kernel* proposto internamente nos programas.

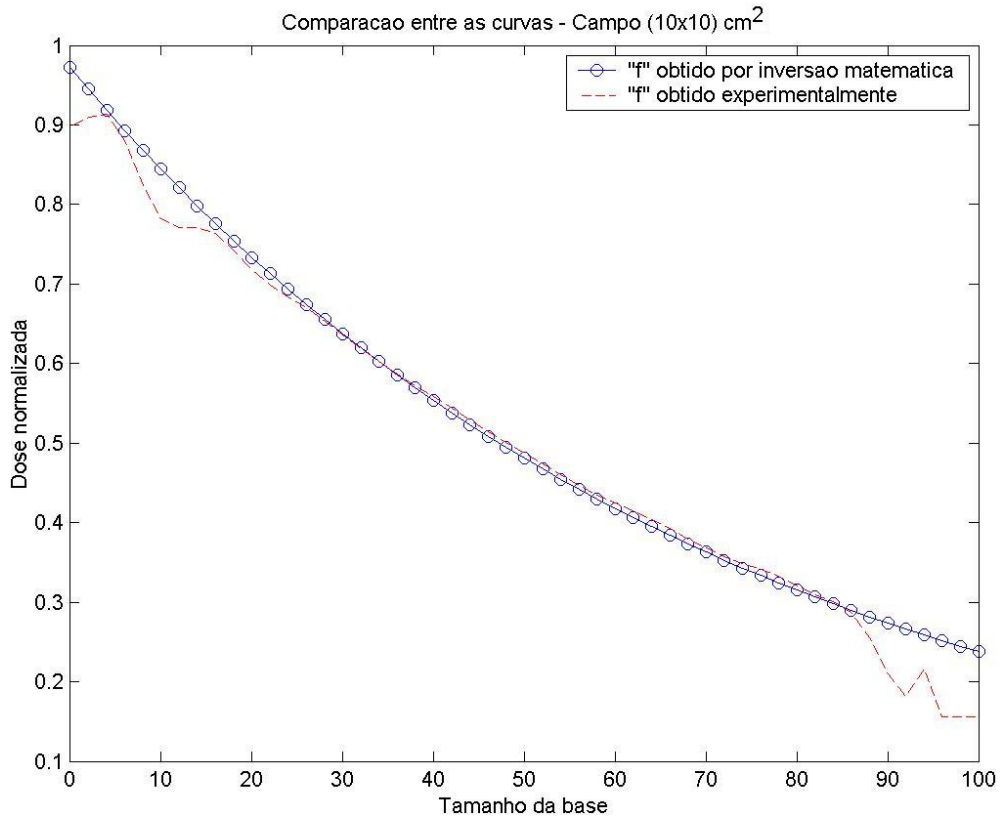


Figura 7.15 – PDP obtida a partir da inversão de dados experimentais via regularização de Tikhonov X PDP obtida experimentalmente – Campo (10x10) cm²; distância fonte-superfície = 100,0 cm

Na figura acima, nota-se que o comportamento da função matemática utilizada nos programas de simulação é satisfatório para a maior parte da curva. Entretanto, tanto no início quanto no final da curva, nota-se variações consideráveis, o que possivelmente poderá ser minimizado melhorando-se o *kernel* proposto internamente nos programas.

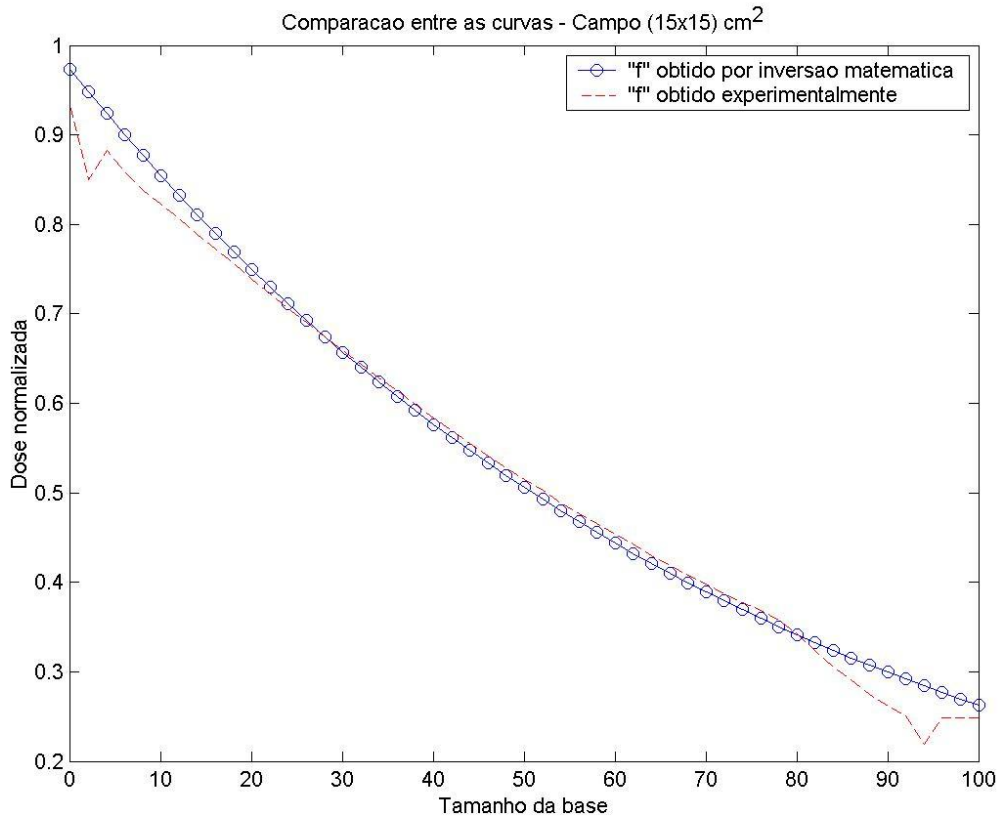


Figura 7.16 – PDP obtida a partir da inversão de dados experimentais via regularização de Tikhonov X PDP obtida experimentalmente – Campo (15x15) cm²; distância fonte-superfície = 100,0 cm

Na figura acima, nota-se que o comportamento da função matemática utilizada nos programas de simulação é satisfatório para a maior parte da curva. Entretanto, tanto no início quanto no final da curva, nota-se variações consideráveis, o que possivelmente poderá ser minimizado melhorando-se o *kernel* proposto internamente nos programas.

7.6 – Conclusão

A dose absorvida de radiação ionizante – principalmente feixes de raios-x e de raios-gama – empregada no tratamento de doenças como o câncer é, normalmente, obtida através de cálculos que envolvem métodos matemáticos que consomem muito tempo, como no caso do método de Monte Carlo. Aqui se propõe o uso de um *kernel* gaussiano, utilizando-se a regularização de Tikhonov, onde se observa através de gráficos que a distribuição de dose no tecido pode ser modelada através de técnicas de problemas inversos. A grande vantagem do uso da regularização de Tikhonov para resolver o

problema inverso da radioterapia está no ganho de tempo, uma vez que os programas baseados em métodos de Monte Carlo são muito lentos. Programas de cálculo de deposição de dose poderiam ser elaborados empregando-se técnicas similares, o que traria benefícios para o planejamento, como rapidez nos cálculos e maior número de pacientes planejados no mesmo intervalo de tempo. As simulações desenvolvidas ao longo desse capítulo podem e devem ser melhoradas, a partir do uso de novos valores para o *kernel*, assim como o uso de outros tipos de programas de computador, como redes neurais.

CAPÍTULO VIII

TEORIA DE ESPALHAMENTO QUÂNTICO

8.1 – Introdução

Quase todas as informações existentes atualmente acerca do núcleo atômico e das partículas elementares foram descobertas a partir de experimentos com espalhamentos. As primeiras descobertas foram feitas por Ernest Rutherford (1871-1937), físico e químico inglês que distinguiu as partículas alfa e beta e que, trabalhando em conjunto com Frederick Soddy (1877-1956), outro químico inglês, desenvolveu a teoria da radioatividade espontânea. Rutherford também foi o primeiro a verificar que a carga positiva de um átomo está concentrada em seu núcleo. Em uma escala muito mais reduzida, descobertas mais recentes mostraram que os prótons e os nêutrons são constituídos de partículas menores, denominadas *quarks*.

O modelo mais simples de experimento de espalhamento é dado através da solução da equação de Schrödinger para uma onda plana incidindo sobre um potencial localizado, que representa o quão rápido um elétron se choca no alvo. O cenário básico para o experimento consiste em um feixe de partículas, todas com a mesma energia, onde se detecta quantas dessas são defletidas no alvo, considerando-se os ângulos de deflexão. Normalmente se assume que as partículas incidentes são representadas por pacotes de ondas de mesmo formato e de mesmo tamanho; assim sendo, a equação de Schrödinger dependente do tempo é solucionada para esses pacotes e as probabilidades de interação para as ondas refletidas em diferentes direções após certo tempo são encontradas. Uma simples aproximação pode tornar a tarefa menos árdua: pode-se assumir que o pacote de ondas tem uma energia bem definida e, portanto, um momento bem definido. Isso significa que, por um período de tempo, o espalhamento é independente do tempo; dessa forma, a solução do problema pode ser bem aproximada com o uso da equação de Schrödinger independente do tempo.

Assume-se que a interação entre a partícula espalhada e o meio espalhador pode ser representada por uma função energia potencial $V(r)$, onde r é um vetor formado entre a partícula espalhada e o centro da força ($r = 0$), e que representa as coordenadas relativas das partículas. Se a massa do espalhador for muito grande comparada à da partícula espalhada, assume-se que o espalhador permanece em repouso durante todo o processo de espalhamento. Todas as considerações a serem feitas nesse capítulo referem-se a

espalhamentos elásticos, ou seja, sistemas onde a energia cinética não é alterada em função da interação.

De uma maneira geral, a função energia potencial $V(r)$ diminui em magnitude na medida em que o valor absoluto da distância r aumenta. Assim sendo, torna-se conveniente a escolha do valor arbitrário $V = 0$ quando $r = \infty$.

A energia total da partícula é dada pela equação (8.1) a seguir:

$$E = \frac{1}{2}mv_{\infty}^2 = \frac{\hbar^2}{2m}k^2, \quad (8.1)$$

em que v_{∞} é a velocidade em $r = \infty$, k é o número da onda correspondente e \hbar é a constante de Planck dividida por 2π

O valor da energia E é muito maior do que o de $V(r)$ para grandes valores absolutos de r . Se $V(r)$ tende a zero rapidamente quando r tende a ∞ , então a partícula pode ser considerada essencialmente livre para grandes valores de r .

Em experimentos sobre espalhamentos obtém-se o número de partículas espalhadas por tempo numa certa direção e ângulo sólido. Essas partículas podem ser contadas por um detector. O número de partículas espalhadas será proporcional ao fluxo incidente, ao número de centros espalhadores e ao ângulo sólido:

$$dn = \sigma(\theta, \phi)nF_i d\Omega, \quad (8.2)$$

em que dn é o número de partículas espalhadas, $\sigma(\theta, \phi)$ é a constante de proporcionalidade (seção de choque diferencial), n é o número de centros espalhadores, F_i é o fluxo incidente e Ω é o ângulo sólido

A seção de choque diferencial pode ser calculada como o quadrado do módulo da amplitude de espalhamento:

$$\sigma(\theta, \phi) = |f_k(\theta, \phi)|^2 \quad (8.3)$$

O estudo do problema do espalhamento pode ser simplificado utilizando-se o conceito de massa reduzida, descrito pela equação a seguir:

$$\frac{1}{\mu} = \frac{1}{m_1} + \frac{1}{m_2} \quad (8.4)$$

em que m_1 e m_2 se referem às massas da partícula incidente e do centro espalhador, respectivamente.

A equação de Schrödinger tem infinitas soluções, mas dentre essas procura-se as que têm energia bem definida (estados estacionários). A equação de Schrödinger independente do tempo que descreve uma partícula sujeita à influência de um potencial $V(r)$ é dada pela equação (8.5) a seguir:

$$\left[-\frac{\hbar^2}{2\mu} \nabla^2 + V(r) \right] \psi_E(r) = E \psi_E(r), \quad (8.5)$$

em que $\psi_E(r)$ são autofunções associadas aos autovalores E , ∇^2 é o operador laplaciano e r é o vetor cujo módulo é igual à distância entre as partículas

As equações (8.6) e (8.7) a seguir podem ser utilizadas para simplificação da escrita da equação (8.5):

$$k^2 = \frac{2\mu E}{\hbar^2} \quad (8.6)$$

$$U = \frac{2\mu V(r)}{\hbar^2} \quad (8.7)$$

Assim, pode-se reescrever a equação (8.5) na forma (8.8):

$$(\nabla^2 + k^2) \psi_E(r) = U \psi_E(r) \quad (8.8)$$

A solução geral da equação (8.8) tem a forma da equação (8.9):

$$\psi_E(r) = \phi_0(r) + \phi(r) \quad (8.9)$$

8.2 – Formato Assintótico das Ondas

A descrição matemática das ondas incidentes e espalhadas pode ser mais simples ou mais complexa, dependendo da distância à zona de influência do potencial. Próximo a essa zona essa descrição é muito complexa, pois seu formato depende do potencial (Oliveira, 2008). Já sua forma assintótica (longe da ação do potencial) pode ser descrita matematicamente sem maiores problemas.

A onda incidente praticamente não sofre interação com o potencial até atingir determinada distância. Até esse momento ela pode ser considerada plana e, de acordo com o estudo das ondas, sua função apresenta formato exponencial e^{ikz} . A onda espalhada, quando considerada em grandes distâncias após a interação, assume forma esférica, se o alvo for puntiforme, com função da forma e^{ikr}/r ; se o alvo possui determinada estrutura interna, a onda espalhada não deverá ser de forma esférica, devendo ser modulada por uma função angular $f(\theta)$ (Oliveira, 2008).

Tem-se, portanto, uma equação diferencial onde duas funções de onda diferentes são soluções da mesma. Assim, sua combinação linear também será solução (Boyce & DiPrima, 2006). Portanto, a solução assintótica para a equação (8.5) é dada pela equação (8.10) a seguir.

$$\psi_E(r) = e^{ikz} + f(\theta) \frac{e^{ikr}}{r}, \quad (8.10)$$

em que $f(\theta)$ é a amplitude de espalhamento.

8.3 – Solução Utilizando Equações Integrais

Para se tentar solucionar a equação (8.8), pode-se empregar diversos métodos matemáticos. Entretanto, há algumas dificuldades devido à natureza mal-colocada do problema do espalhamento quântico. Uma tentativa de solução é encontrar funções que satisfaçam a equação (8.11) a seguir:

$$(\nabla^2 + k^2)G(r) = \delta(r) \quad (8.11)$$

em que $G(r)$ é chamado função de Green do operador $(\nabla^2 + k^2)$, enquanto que $\delta(r)$ é o delta de Dirac (Oliveira, 2008).

A equação (8.12) a seguir é uma tentativa de solução da equação (8.11):

$$\phi(r) = \phi_0(r) + \int_a^b d^3 r' G(r - r') U(r') \phi(r'),$$

em que r é o vetor distância entre a posição da partícula e um ponto localizado a grande distância e r' é o vetor distância entre a posição da partícula e um ponto localizado dentro da zona de influência do potencial (8.12)

em que $\phi_0(r)$ é a solução da equação (8.13) a seguir:

$$(\nabla^2 + k^2)\phi_0(r) = 0 \quad (8.13)$$

Na hipótese de operador $(\nabla^2 + k^2)$ poder ser colocado dentro da integral apresentada na equação (8.12), essa torna-se:

$$\begin{aligned} (\nabla^2 + k^2)\phi(r) &= (\nabla^2 + k^2) \left[\phi_0(r) + \int_a^b d^3 r' G(r - r') U(r') \phi(r') \right] \Rightarrow \\ (\nabla^2 + k^2)\phi(r) &= (\nabla^2 + k^2)\phi_0(r) + \int_a^b d^3 r' (\nabla^2 + k^2) G(r - r') U(r') \phi(r') \end{aligned} \quad (8.14)$$

Adaptando-se a equação (8.14) à equação (8.11) obtém-se:

$$(\nabla^2 + k^2)\phi(r) = \int_a^b d^3 r' \delta(r - r') U(r') \phi(r') \quad (8.15)$$

A equação (8.16) a seguir é apresentada e provada na referência Boyce & DiPrima (2006), podendo ser utilizada para simplificação da equação (8.15).

$$\int d^3 r \delta(r - r_0) f(r) = f(r_0) \quad (8.16)$$

Reescrevendo a equação (8.15) utilizando-se (8.16) obtém-se:

$$(\nabla^2 + k^2)\phi(r) = \int_a^b d^3 r' \delta(r - r') U(r') \phi(r') = U(r) \phi(r) \quad (8.17)$$

A função de Green pode ser subdividida em G_+ , conhecida como função de Green de saída e G_- , conhecida como função de Green de entrada (Oliveira, 2008). A equação a seguir é válida para essa função:

$$G_{\pm}(r) = -\frac{1}{4\pi} \frac{e^{\pm ikr}}{r} \quad (8.18)$$

Para espalhamento no estado estacionário utiliza-se $\phi_0(r) = e^{ikz}$ e $G(r) = G_+(r)$. Assim, a equação-solução (8.12) pode ser reescrita da forma:

$$\phi(r) = e^{ikz} + \int_a^b d^3 r' G_+(r - r') U(r') \phi(r') \quad (8.19)$$

Como $|r'| \ll |r|$ tem-se:

$$\phi(r) = e^{ikz} - \frac{1}{4\pi} \frac{e^{ikr}}{r} \int_a^b d^3 r' e^{-kr} U(r') \phi(r') = e^{ikz} + f(\theta) \frac{e^{ikr}}{r} \quad (8.20)$$

8.4 – A Aproximação de Born

Faz-se a seguinte mudança de notação na equação (8.19):

$$r \rightarrow r', r' \rightarrow r'', \dots$$

Assim, reescreve-se a equação (8.19) como:

$$\phi(r') = e^{ikz} + \int_a^b d^3 r'' G_+(r' - r'') U(r'') \phi(r'') \quad (8.21)$$

Definindo-se o vetor de onda incidente k_i como sendo o vetor de módulo k dirigido ao longo do eixo Z de tal modo que $e^{ikz} = e^{ik_i r}$ tem-se:

$$\phi(r) = e^{ikz} + \int_a^b d^3 r' G_+(r-r') U(r') e^{ik_i r'} + \int_a^b d^3 r' \int_a^b d^3 r'' G_+(r'-r'') U(r'') \phi(r'') \quad (8.22)$$

Se essa mudança de notação for continuada várias vezes, obtém-se uma sequência de integrais, válida para potenciais de grande alcance. Essa equação é conhecida como **expansão de Born**.

Para potenciais de curto alcance, pode-se truncar a equação (8.21) e utilizar-se apenas o seu primeiro termo, que é chamado **aproximação de Born**. Essa aproximação pode ser muito útil na resolução do problema inverso da obtenção da função energia potencial intermolecular de um sistema químico a partir de dados experimentais de seção de choque diferencial.

A referência Lemes *et. al* (2008) trata do uso da aproximação de Born para obtenção da função energia potencial a partir da seção de choque diferencial experimental de um sistema Ar-Ar. Foi utilizada uma rede neural recorrente nas simulações. Os dados experimentais utilizados foram baseados no potencial de Aziz. Os resultados gráficos são apresentados a seguir.

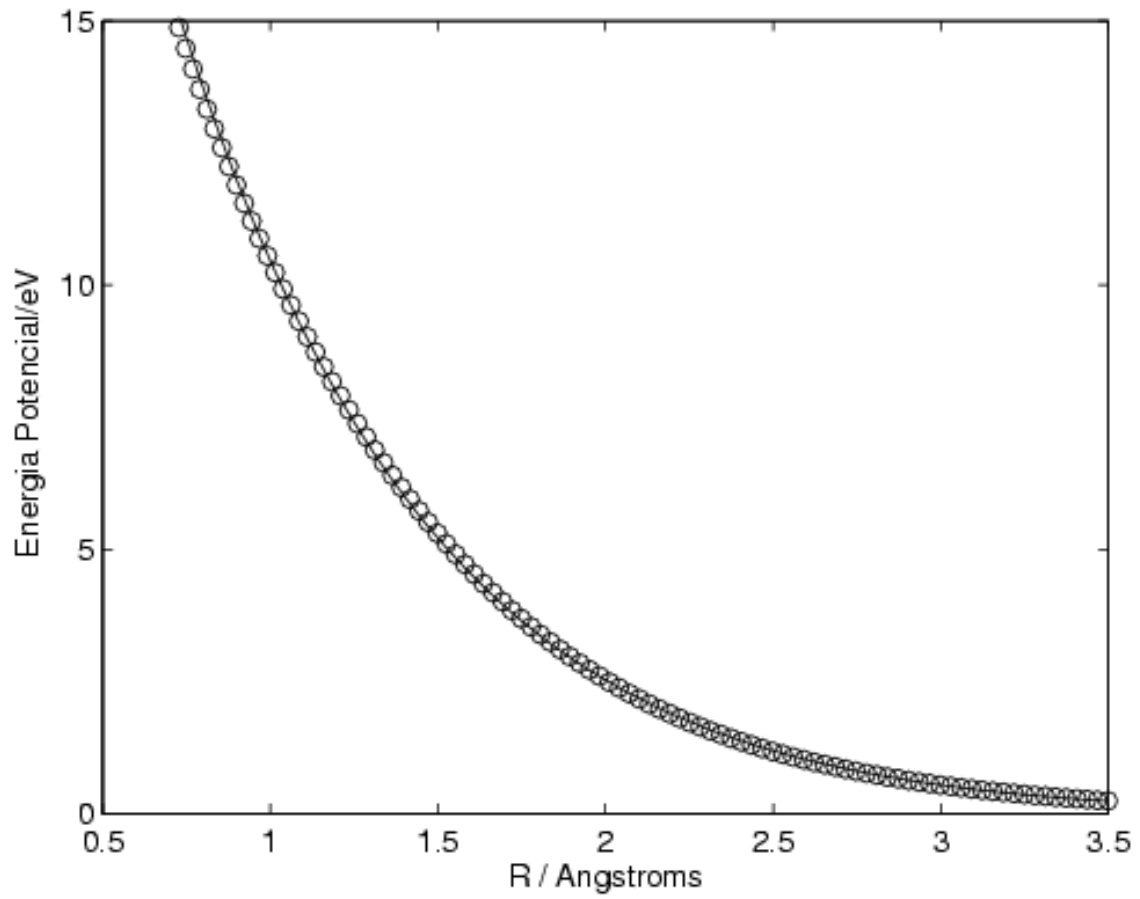


Figura 8.1 – Função energia potencial invertida (o) e o seu valor exato (linha cheia) – adaptado de Lemes et. al (2008)

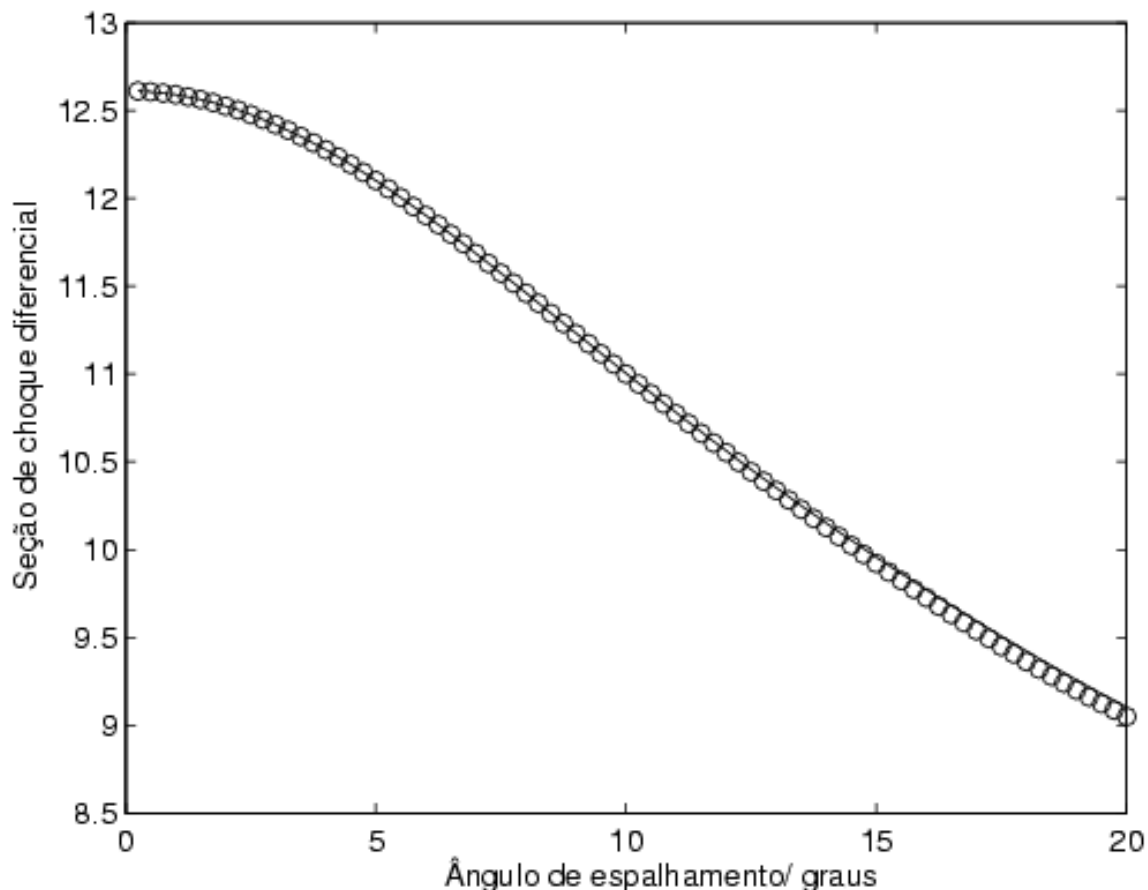


Figura 8.2 – Seção de choque diferencial calculada a partir do potencial invertido (o) e o seu valor exato (linha cheia) - adaptado de Lemes et. al (2008)

O algoritmo escrito para essas simulações apresentou um erro muito pequeno, o que assegura a precisão dos resultados invertidos. O potencial obtido a partir da seção de choque experimental reproduz com bastante precisão os valores exatos, o que também pode ser demonstrado a partir da obtenção da seção de choque via inversão do potencial.

8.5 – Conclusão

O estudo baseado em espalhamento de partículas, tanto mecânica quanto quanticamente, é essencial à compreensão de muitos eventos que ocorrem na natureza, especialmente nos casos microscópicos. O uso de cálculo integral e diferencial associado a técnicas de resolução de problemas inversos pode ser feito para o modelamento matemático de experimentos que envolvem tais espalhamentos. No caso de espalhamento quântico, o uso da aproximação de Born pode ajudar nesse modelamento, pois simplifica a matemática a ser empregada.

CAPÍTULO IX

CONCLUSÕES

Em ciência, especialmente em se tratando de físico-química, diversos experimentos são feitos diariamente em laboratórios e/ou observados no dia-a-dia. Normalmente é necessário o modelamento matemático de tais observações, que podem se tratar de problemas mal-colocados. O uso de técnicas de problemas inversos é uma opção poderosa e a regularização de Tikhonov é uma importante técnica de regularização matemática utilizada na resolução desses problemas, com aplicações em situações como na deposição de dose em radioterapia, retornando sempre valores melhores do que outras técnicas mais simples, como a inversão de matrizes ou o método de mínimos quadrados. Nesse trabalho, apresentou-se técnicas de problemas inversos como uma tentativa de resolução de problemas onde se consuma menos tempo de computação, especialmente em radioterapia.

Foram introduzidos os métodos de Monte Carlo para modelamento de experimentos realizados em físico-química. Em radioterapia pode-se empregá-los para modelagem de feixes de radiação ionizante na obtenção de valores relacionados à distribuição da dose absorvida em tecidos humanos. O uso do MCNP, poderoso software baseado no método de Monte Carlo, também foi introduzido nesse trabalho para resolver problemas relacionados à radiação ionizante. Mostrou-se a capacidade desse programa através de exemplos simples, como uma pequena fonte de radiação incidindo um feixe sobre uma placa. Com arquivos de entrada bem simples é possível fazer-se o estudo de transmissão de radiação através da placa, assegurando-se se a mesma proporciona uma blindagem adequada ou não. Com exemplos mais sofisticados, como a obtenção de curvas teóricas de PDP ou a obtenção de curvas de isodose para equipamentos como o GammaCell, é possível compreender todo o potencial desse programa para solucionar problemas relacionados a deposição de dose, blindagem de fontes, curvas de isodose e outros.

Falou-se também, de forma sucinta, sobre a teoria do espalhamento quântico, mostrando-se um uso da aproximação de Born na obtenção de curvas de potencial a partir de seções de choque. O estudo de espalhamento também é muito importante dentro da físico-química, especialmente quando relacionado à interação da radiação ionizante em meios como o tecido humano, objeto de constante estudo em radioterapia.

As técnicas de problemas inversos necessitam de uma matemática muito elaborada, de difícil entendimento e que consomem muito tempo para assimilação das teorias envolvidas. Após o entendimento das técnicas matemáticas utilizadas, todos os algoritmos

devem ser resolvidos via programação de computadores, a fim de se reduzir o tempo gasto para solução, assim como para minimizar possíveis erros. O uso de técnicas de problemas inversos em sistemas dinâmicos é muito novo, especialmente em medicina.

CAPÍTULO IX

REFERÊNCIAS BIBLIOGRÁFICAS

- AHNESJÖ, A. ASPRADAKIS, M.M. **Dose Calculations for External Photon Beams in Radiotherapy**. Physics, Medical and Biology, 44 (1999) R99-R155.
- AMAR, JACQUES G. **The Monte Carlo Method in Science and Engineering**. Computing in Science and Engineering. IEEE Xplore. USA, 2006.
- ANGER, G.; GORENFLO, R.; JOCHMANN, H.; MORITZ, H.; WEBERS, W. **Inverse Problems: Principles and Applications in Geophysics, Technology and Medicine**. Proceedings of the International Conference held in Potsdam, August 30 – September 3, 1993.
- BAUCHSPIESS, A. **Introdução aos Sistemas Inteligentes**. Aplicações em Engenharia de Redes Neurais Artificiais, Lógica Fuzzy e Sistemas Neuro-Fuzzy. Universidade de Brasília, 2004.
- BOYCE, W.E.; DIPRIMA, R.C. **Equações Diferenciais Elementares e Problemas de Valores de Contorno**. 8ª edição. LTC, 2006.
- BUSHONG, STEWART C. **Radiologic Science for Technologists – Physics, Biology and Protection**. 7th Edition. Mosby Inc. USA, Houston, 2001.
- CENSOR, Y. **Mathematical Aspects of Radiation Therapy Treatment Planning: Continuous Inversion Versus Full Discretization and Optimization Versus Feasibility**. Department of Mathematics, University of Haifa, Mt, Carmel, Haifa 31905, Israel, 1999.
- CHVETSOV, A.V.; CALVETTI, D.; SOHN, J.W.; KINSELLA, T.J. **Regularization of Inverse Planning for Intensity-Modulated Radiotherapy**. Medical Physics 32(2), February 2005.
- DANESE, G.; DE LOTTO, I.; DOTTI, D.; LANTERNA, D.; LEPORATI, F.; LOMBARDI, R.; ROMASO, S. **Mapping MonteCarlo-Metropolis Algorithm onto a Double Ring Architecture**. International Conference on Application-Specific Array Processors. 1993.

- ENFUREÇA, G. **Basic Principles of Inverse Problems**. Rathausstr, 13, Wg. 11/09. D-10178 Berlin.
- FISHMAN, S. GEORGE. **Monte Carlo. Concepts, Algorithms and Applications**. Springer –Verlag New York, Inc. 1996.
- GROETSCH, C.W. **Inverse Problems – Activities for Undergraduates**. The Mathematical Association of America. USA, 1999.
- HANSEN, P.C. **Numerical Tools for Analysis and Solution of Fredholm Integral Equations of the First Kind**. *Inverse Problems* 8 (1992) 849-872.
- HANSEN, P.C. **The L-curve and its Use in the Numerical Treatment of Inverse Problems**. Tech. Report, IMM-REP 99-15, Department of Mathematical Modeling, Technical University of Denmark, 1999.
- HENDEE, W.R.; IBBOTT, G.S. **Radiation Therapy Physics**. Second Edition. Mosby-Year Book, Inc. USA, 1996.
- HULBERT, S.F. **Models for Solid-State Reactions in Powdered Compacts: A Review**. *J.B.C.S.*, 6, 1, 1971.
- IAEA. **Review of Radiation Oncology Physics: A Handbook for Teachers and Students**. Editor Ervin B. Podgorsak. International Atomic Energy Agency. Canada, 2003.
- ICRU. **Determination of Absorbed Dose in a Patient Irradiated by Beams of X or Gamma Rays in Radiotherapy Procedures**. ICRU Publication 24 (Bethesda, MD: International Commission on Radiation Units and Measurements), 1976.
- JACOBS, P.W.M.; TOMPKINS, F.C. **Classification and Theory of Solid Reactions**. Chapter 7; in: *Chemistry of the Solid State*. W.E. Garner, Butterworths, 1955.

- KALOS, MALVIN H. **Monte Carlo Methods in the Physical Sciences**. Proceeding of the 2007 Winter Simulation Conference. Lawrence Livermore National Laboratory. USA, 2007.
- KHAN, F.M. **The Physics of Radiation Therapy**. Second Edition. Williams & Wilkins. USA, 1994.
- KOONIN, STEVEN E. & MEREDITH, DAWN C. **Computational Physics – Fortran Version**. Addison-Wesley Publishing Company, Inc. 1990.
- LEMES, N.H.T.; BORGES, E.; SOUSA, R.V.; BRAGA, J.P. **Potential Energy Function from Differential Cross-Section Data: An Inverse Quantum Scattering Theory Approach**. International Journal of Quantum Chemistry, Vol. 108, pp. 2623-2627, Wiley Periodicals, Inc., 2008.
- LEON, S.J. **Linear Algebra with Applications**. 5th Edition. University of Massachusetts Dartmouth. Prentice Hall, Upper Saddle River, New Jersey 07458, 1998.
- LIND, B.K. **Properties of an Algorithm for Solving the Inverse Problem in Radiation Therapy**. Inverse Problems 6 (1990) 415-426.
- LOS ALAMOS NATIONAL LABORATORY. **MCNPTM – A General Monte Carlo N-Particle Transport Code**. Version 4C. LA-13709-M. USA, March 2000.
- MARTINEZ, W.L. & MARTINEZ, A.R. **Computational Statistics Handbook with MATLAB**. Chapman & Hall/CRC. USA, 2002.
- MORITZ, H. **General Considerations Regarding Inverse and Related Problems**. Technische Universität Graz. Abteilung Physikalische Geodäsie. A-8010 Graz. 1993.

- NEDELKOV, I.P. **Improper Problems in Computational Physics**. Computer Physics Communications 4 (1972) 157-164. North-Holland Publishing Company.
- NEWMAN, M.E.J. & BARKEMA, G.T. **Monte Carlo Methods in Statistical Physics**. Clarendon Press. Oxford. USA, 2001.
- NUNES, W.V. **Detecção de Minas Terrestres por Radiação Penetrante**. Tese de Doutorado. Departamento de Engenharia Nuclear, UFRJ, 2006.
- OLIVEIRA, J.M. **Estudos em Espalhamento Elástico Quântico, Espalhamento Clássico e Dinâmica Molecular**. Dissertação de Mestrado. Departamento de Química da UFMG. 2008.
- OSÓRIO, F. **Redes Neurais – Aprendizado Artificial**. Fórum de Inteligência Artificial. 1999.
- RIELE, H.J.J. te. **A Program for Solving First Kind Fredholm Integral Equations by Means of Regularization**. Computer Physics Communications 36 (1985) 423-432. North-Holland, Amsterdam.
- RINGOR, M.; PAPIEZ, L. **Inverse Planning and Optimization: a Comparison of Solutions**. Radiation Physics and Chemistry 53 (1998) 263-274.
- RODRIGUES, R.R.; SOUSA, R.V.; RIBEIRO, M.A.; GRYNBERG, S.E.; FERREIRA, A.V.; BELO, L.C.M.; SQUAIR, P.L.; SEBASTIÃO, R.C.O.; **Gamma Irradiator Dose Mapping: A Monte Carlo Simulation and Experimental Measurements**. International Nuclear Atlantic Conference (INAC) 2009. Rio de Janeiro, 2009.
- SHARP, J.H.; BRINDLEY, G.W.; ACHAR, B.N.N. **Numerical Data for Some Commonly Used Solid State Reaction Equations**. J.A. Ceram. Soc., 49, 379, 1966.

- SOUSA, R.V. **Avaliação Preliminar de um Sistema de Transmutação de Rejeitos Altamente Radioativos Usando Fontes de Nêutrons Rápidos.** Dissertação de Mestrado. Departamento de Engenharia Nuclear, UFMG, 2004.
- SOUSA, R.V. **Dose Rate Influence on Deep Dose Deposition Using a 6 MV X-Ray Beam From a Linear Accelerator.** Brazilian Journal of Physics. Vol. 39, no. 2. June, 2009.
- SOUSA, R.V.; BORGES, E.; BRAGA, J.P. **Ill-posed Inverse Problems in Medical Physics: Tikhonov Regularization in Obtaining Percentage Depth Dose in Radiation Therapy.** In: Conference on Computational Physics, 2008a, Ouro Preto-MG.
- SOUSA, R.V.; BORGES, E.; LEMES, N.H.T.; BRAGA, J.P. **Obtenção de Curvas de Percentagem de Dose Profunda em Radioterapia via Inversão de Dados Experimentais por Técnicas de Regularização de Tikhonov.** In: XIII Congresso Brasileiro de Física Médica, 2008b, Belo Horizonte-MG.
- SOUZA, M.J.F. **Ajuste de Curvas pelo Método dos Quadrados Mínimos.** Notas de Aula de Métodos Numéricos. Departamento de Computação / ICEB / UFOP. 2006.
- SPINOZZI, F.; GAZZILLO, D.; GIACOMETTI, A.; MARIANI, P.; CARSUGHI, F. **Interaction of Proteins in Solution from Small-Angle Scattering: A Perturbative Approach.** Biophysical Journal. Volume 82. April 2002.
- TERVO, J.; KOLMONEN, P. **Data Fitting Model for the Kernel of Integral Operator from Radiation Therapy.** Mathl. Comput. Modelling Vol. 28, No. 3, pp. 59-77, 1998.
- Varian Medical Systems, Inc. **Low Energy C-Series CLINAC. Systems Manual.** P/N 01102658-01 3/96. USA, 1996.

- X-5 Monte Carlo Team. **MCNP – A General Monte Carlo N-Particle Transport Code, Version 5**. Volume I – Overview and Theory. Los Angeles National Laboratory, University of California, USA, 2003.
- YEREMIN, F.N. **The Foundations of Chemical Kinetics**. Mir Publishers. Moscow, 1979.