

V Bienal da SBM
Sociedade Brasileira de Matemática
UFPB - Universidade Federal da Paraíba
18 a 22 de outubro de 2010

Problemas Inversos e a Matemática da Tomografia Computadorizada

Adriano De Cezaro

e

Fabiana Travessini De Cezaro

Resumo

Os *Problemas Inversos* envolvem conhecimentos em várias áreas da matemática e possuem muitas aplicações em outras ciências. Entre os problemas que são foco de interesse atual de pesquisa estão: aplicações industriais, como testes não destrutivos, identificação de parâmetros em processos industriais, [2, 9, 14, 21, 28, 50]; tomografias e aplicações às ciências médicas, como detecção de tumores e fraturas, [3, 32, 33, 38, 39]; aplicações à geofísica com detecção de reservatórios de minerais, prospecção de hidrocarbonetos, [25, 43, 50, 53].

Neste trabalho, daremos atenção especial a matemática dos Problemas Inversos que envolvem a reconstrução de imagens por *Tomografia Computadorizada*. Ou seja, queremos identificar a forma e a densidade de um objeto, imerso em uma região $\Omega \subset \mathbb{R}^n$, $n = 2, 3$, através de medidas de um conjunto de intensidades dos raios-X que atravessam Ω . Nesse processo, o conjunto de parâmetros conhecidos são a intensidade inicial do raio-X e medições de sua intensidade de chegada a um detector. Fisicamente, a diferença entre a intensidade inicial e final do raio-X depende do coeficiente de absorção local e do caminho que o raio-X percorre ao atravessar Ω . Do ponto de vista matemático, esse processo é modelado pela *Transformada de Radon*.

Nosso foco é apresentar, utilizando conceitos de Álgebra Linear [12, 36, 49] e Análise [29, 30, 31], alguns métodos de obtenção de soluções estáveis para o problema em *Tomografia Computadorizada*. Observamos que a solução deste problema é complexa e envolve técnicas em matemática, física e computação científica, [3, 4, 6, 5, 38, 39]. Portanto, problemas em *Tomografia Computadorizada* envolvem conhecimentos interdisciplinares e, em particular, ferramentas matemáticas adequadas para a modelagem e análise do modelo. Ainda, exigem o desenvolvimento e estudo de métodos adequados para a obtenção de uma solução [3, 10, 25, 33, 39, 51]. Tais métodos se encaixam no estudo dos chamados *Problemas Inversos*, [1, 2, 3, 5, 6, 9, 15, 25, 53].

Dada a relevância do tópico, objetivamos despertar o interesse do público nessa área, divulgando um pouco da teoria matemática que permeia a investigação científica dos *Problemas Inversos* e a importância teórica e prática destes na reconstrução de imagens por *Tomografia Computadorizada*.

Conteúdo

Lista de Figuras	iii
Notação	iv
Introdução	5
1 Tomografia Computadorizada	8
1.1 Descrição Matemática do Processo de Tomografia Computadorizada	10
1.2 O Problema Inverso	13
1.3 Um Pouco da História da Tomografia Computadorizada	14
2 Introdução aos Problemas Inversos Lineares	16
2.1 Breve Apanhado Histórico dos Problemas Inversos	16
2.2 O que são Problemas Inversos ?	17
2.3 O Problema Inverso da Diferenciação	20
2.3.1 Reconstrução de uma Força Desconhecida	20
2.3.2 Diferenciação nos Dados	22
2.4 A Tomografia Computadorizada e a Transformada de Radon	23
2.4.1 Tomografia Computadorizada: caso contínuo	23
2.4.2 A Transformada Inversa de Radon	26
2.4.3 Tomografia Computadorizada: caso discreto	31
3 Sistemas de Equações Lineares	33
3.1 Pseudo-Inversa de Operadores Lineares	33
3.1.1 Definições e Propriedades Básicas	34
3.2 A Decomposição em Valores Singulares	40
3.2.1 Funções de Operadores: O Teorema da Aplicação Espectral	43
3.2.2 Relação entre Má-colocação e Valores Espectrais	45

4	Regularização para Problemas Inversos	48
4.1	O Conceito de Regularização	49
4.2	Resultados de Convergência	50
4.2.1	Escolha <i>a priori</i> do Parâmetro de Regularização	52
4.2.2	Escolha <i>a posteriori</i> do Parâmetro de Regularização	54
4.3	Regularização por Truncamento dos Valores Singulares	55
5	Regularização de Tikhonov	58
5.1	Problemas Lineares: Convêrgencia	58
5.2	Problemas Lineares: Semi-Convergência	62
5.2.1	Taxas de convergência	63
5.2.2	Regularização de Tikhonov e a Inversa Generalizada	63
5.3	Regularização de Tikhonov para o Problema de Tomografia Computadorizada	64
5.4	Método de Máxima Verossemelhança para Tomografia Computarizada	65
6	Regularização por Métodos Iterativos	70
6.1	Método de Landweber	71
6.1.1	Convergência	72
6.2	Steepest Descent e a Inversa Generalizada	76
6.3	Métodos tipo Kaczmarz	80
6.3.1	Método ART	80
6.3.2	Método de Landweber-Kaczmarz	82
6.4	Algoritmo EM	82
6.5	Aplicação: Tomografia Computadorizada	86
A	Definições e resultados	89
A.1	Definições e Resultados Básicos em Espaços Vetoriais.	89
A.1.1	Operadores Lineares	90
A.1.2	Transformada de Fourier	91

Lista de Figuras

Figura	Página
Tomografia Computadorizada	9
Tomografia Computadorizada do corpo humano	9
Seção Transversal	9
Atenuação do raio-X	10
Raio-X	10
Tomografia onde os raios-X e detectores são movimentados de forma paralela	12
Tomografia com o feixe de raios-X é tipo cone	12
Tomografia 3D, a partir de várias tomografias 2D	12
Tomografia 3D, construída a partir de uma tomografia sagital, coronal e axial	12
Modelo.	18
Estimativa fundamental	23
Representação de uma Tomografia em paralelo	24
Mudança de Variável	24
Parametrização da reta L que é perpendicular a w a uma distância s da origem.	25
Se $f(x, y) = 0$, quando $ (x, y) > \rho$, então $\mathcal{R}f(w, s) = 0$ quando $s > \rho$.	25
Tomografia discreta.	31
Feixe de raio-X.	31
Interpretação geométrica da inversa de uma matriz	38
Geometria do algoritmo ART	81
Phanton.	87
Dados com 4% de ruídos.	87
Solução aproximada por Landweber e steepest-descent.	88

Notação

A	operador linear
$A(\cdot)$	operador não necessariamente linear
A^*	operador adjunto de A
$A'(\cdot)$	derivada de Fréchet de $A(\cdot)$
A^\dagger	pseudo-inversa de A
δ	nível de ruídos
y^δ	dados y com nível de ruídos δ
x^\dagger	solução de norma mínima
x^δ	solução aproximada do problema para dados com ruídos y^δ
x_k^δ	k -ésimo iterado
x_α^δ	solução regularizada com parâmetro de regularização α .
$\mathcal{R}f$	Transformada de Radon da função f
$\hat{\cdot}$	Transformada de Fourier \mathcal{F}
$\check{\cdot}$	Transformada de Fourier inversa \mathcal{F}^{-1}
$\mathcal{H}f$	Transformada de Hilbert da função f
$\mathcal{H}, \mathcal{H}_1, \mathcal{H}_2$	espaços de Hilbert
$\langle \cdot, \cdot \rangle$	produto interno
$\ \cdot \ $	norma do espaço
$\ \cdot \ _\infty$	norma do espaço das funções contínuas
\rightarrow	convergência forte
\rightharpoonup	convergência fraca
$C[a, b]$	espaço das funções contínuas em $[a, b]$
$L^2[a, b]$	espaço das funções quadrado integrável em $[a, b]$
$L^1[a, b]$	espaço das funções módulo integrável em $[a, b]$
\mathbb{S}^{n-1}	esfera em \mathbb{R}^n
$H(s, w)$	Hiperplano a distância $ s $ da origem
$\mathcal{S}(\mathbb{R}^n)$	espaço de Schwartz
$\frac{\partial^k}{\partial s^k}$	derivada parcial de ordem k com relação a variável s

Introdução

O estudo de problemas inversos envolve conhecimentos em várias áreas da matemática. Entre os problemas que são foco de interesse atual de pesquisa estão: aplicações industriais (testes não destrutivos, identificação de parâmetros em processos industriais); tomografias e aplicações às ciências médicas (detecção de tumores e fraturas); aplicações a geofísica (detecção de reservatórios de minerais, prospecção de hidrocarbonetos), [3, 32, 38, 39, 43].

A base do problema de reconstrução de imagens médicas por Tomografia Computadorizada é, indubitavelmente, matemática. Em especial, tal matemática influenciou o desenvolvimento de novos métodos e técnicas matemáticas para a obtenção de uma solução para o problema, bem como, o desenvolvimento de outros campos científicos e vice-versa [3, 4, 8, 10, 28, 38, 39, 41].

O problema em Tomografia Computadorizada pode ser descrito de forma simples: Reconstruir a forma de um objeto através de medidas das intensidades dos raios-X que o atravessam, conhecendo-se a intensidade do raio-X imposto no processo e obtendo medições de sua intensidade de chegada a um detector, [3, 4, 38, 39]. Dependendo do caminho que os raios-X percorrem, estes são atenuados e a absorção local é medida por um conjunto de detectores. Para determinar a estrutura e/ou a forma do objeto, é necessário que este seja irradiado por todas as direções. A solução deste problema é complexa e envolve técnicas em matemática, física e computação científica, [4, 3, 6, 5, 36, 38, 39]. Portanto, problemas em Tomografia Computadorizada envolvem conhecimentos interdisciplinares e, em particular, ferramentas matemáticas adequadas para modelagem e análise do modelo. Ainda, exigem o desenvolvimento de métodos, tanto numéricos como matemáticos, que possam garantir a existência de uma solução estável para o problema, [3, 10, 25, 33, 36, 39, 51].

Estas notas apresentam alguns avanços da tecnologia e, principalmente, da matemática envolvida na reconstrução de imagens médicas por Tomografia Computadorizada. A solução de tais problemas se encaixam nos chamados *Problemas Inversos*, [1, 2, 3, 6, 5, 9]. Dada a relevância do tópico, objetivamos despertar o interesse do público nessa área, divulgando um pouco da teoria matemática que permeia a investigação científica dos Problemas Inversos e a importância teórica e prática destes em reconstrução de imagens por Tomografia Computadorizada.

As notas estão divididas da seguinte forma:

No Capítulo 1, abordamos aspectos teóricos e práticos da Tomografia Computadorizada, o qual é o principal foco de nossas notas. Em particular, formulamos o problema matematicamente

na Seção 1.1 e introduzimos o Problema Inverso associado a Tomografia Computadorizada na Seção 1.2. Na Seção 1.3, apresentamos um breve apanhado histórico sobre a Tomografia Computadorizada.

No Capítulo 2, apresentamos uma introdução aos Problemas Inversos Lineares. Na Seção 2.1, apresentamos um breve apanhado histórico dos Problemas Inversos. Na Seção 2.2, procuramos responder a seguinte pergunta: *O que são os Problemas Inversos?* A partir deste ponto, nas Seções 2.3 e 2.4, apresentamos exemplos de problemas inversos. Em especial, na Seção 2.4, introduzimos a Transformada de Radon aplicada a Tomografia Computadorizada e investigamos as propriedades desta transformada como forma de resolver o Problema Inverso de identificação de coeficientes de absorção associado a Tomografia Computadorizada. A formulação e análise do problema e a obtenção da Transformada de Radon inversa são apresentadas nas Subseções 2.4.1 e 2.4.2, respectivamente. Por fim, a Subseção 2.4.3 trata da formulação do problema de Tomografia Computadorizada aplicável a situações reais, i.e., quando temos acesso somente a uma pequena quantidade de informações do problema.

No Capítulo 3 fazemos uma revisão de certos conceitos e ferramentas associadas de Álgebra Linear que são usados na solução do Problema Inverso da Tomografia Computadorizada apresentado na Subseção 2.4.3. Estes conceitos são: A Pseudo-Inversa de um Operador Linear (trabalhada na Seção 3.1) e a Decomposição em Valores Singulares (SVD) (estudada na Seção 3.2). Na Subseção 3.2.2, estudamos a relação entre problemas mal postos e os valores espectrais de uma matriz.

No Capítulo 4 introduzimos de forma precisa o conceito de regularização para problemas inversos. Na Seção 4.1, introduzimos o conceito de regularização e de uma solução regularizada para o problema. Na Seção 4.2, estudamos a relação entre a escolha do parâmetro de regularização e resultados de convergência para problemas inversos. Na Seção 4.3, usamos o conceito de regularização para obter uma solução regularizada de um problema considerando o truncamento dos valores espectrais do operador.

No Capítulo 5, tratamos do método de Regularização de Tikhonov. Nas Seções 5.1 e 5.2, mostraremos que o Método de Tikhonov é um método de Regularização. Na Subseção 5.2.1 obtemos taxas de convergência de uma solução regularizada obtida pelo Método de Regularização de Tikhonov para a solução do problema. Na Subseção 5.2.2, relacionamos a regularização de Tikhonov com o conceito de Pseudo-Inversa visto no Capítulo 2. Na Seção 5.3, utilizamos o método de regularização de Tikhonov para obter uma solução regularizada para o problema da Tomografia Computadorizada. Na Seção 5.4, introduzimos conceitos estatísticos de Máxima Verossemelhança para derivar um funcional de Tikhonov relacionado com a distância de Kullback-Leibler. Ainda aplicamos esse conceito para obter uma solução regularizada para o problema de Tomografia Computadorizada.

No Capítulo 6, introduzimos alguns métodos iterativos como métodos de regularização para

problemas inversos. Na Seção 6.1, analisamos o método de Landweber para problemas lineares. Nesta seção, obtemos resultados de convergência e estabilidade para este método. Na Seção 6.2, analisamos o método de steepest descent relacionado com a Pseudo - Inversa como um método iterativo de regularização. Na Seção 6.3, incorporamos ao método de Landweber estratégias do tipo Kaczmarz para obter soluções estáveis para sistemas de equações lineares. Como exemplo de métodos tipo Kaczmarz, analisamos o famoso método ART (Algebraic Reconstruction Technique) e suas aplicações a Tomografia Computadorizada na Subseção 6.3.1. Na Subseção 6.3.2 aplicamos o método de Landweber- Kaczmarz ao problema de Tomografia Computadorizada para obter uma solução regularizada para o problema. Na Seção 6.4, analisamos o método iterativo chamado de Expectation Maximization como um método iterativo de regularização. Este método é obtido analisando a condição de otimalidade de primeira ordem para o método de Máxima Verossemelhança visto na Seção 5.4.

No Apêndice A enunciamos alguns resultados sobre espaços vetoriais e operadores que são utilizados no decorrer destas notas.

Temos ciência que estas notas não estão completas e, possivelmente, possuam falhas ou erros de escrita. Assim, deixamos o endereço **decezar@impa.br** como forma de contato para sugestões e mudanças. Todos os tipos de comentários e críticas serão muito bem vindos.

Capítulo 1

Tomografia Computadorizada

Hoje em dia a Tomografia Computadorizada (TC) tornou-se uma ferramenta indispensável em rotinas clínicas de obtenção de imagens. Este foi o primeiro método não invasivo de aquisição de imagens internas do corpo humano sem superposição de estruturas anatômicas distintas. A principal vantagem da (TC) é que ela permite o estudo de "fatias" ou seções transversais do corpo humano. Já outros procedimentos, como a radiologia convencional, consistem na representação de todas as estruturas do corpo de forma sobrepostas o que é ruim para identificar pequenas diferenças entre os tecidos. Desta forma, a utilização da (TC) em diagnósticos permite obter imagens com percepção espacial das estruturas internas, as quais ficam mais nítidas. Outra vantagem da (TC) é a maior distinção entre dois tecidos, ou seja, esta permite distinguir diferenças de densidade na ordem 0,5% entre tecidos, ao passo que na radiologia convencional este limiar situa-se nos 5%. Assim, é possível a detecção ou o estudo de anomalias, as quais não seriam possível por simples exames radiológicos ou, então, através de métodos invasivos. Sendo assim, um exame complementar de diagnóstico por (TC) é de grande valor.

O método de Tomografia Computadorizada em imagens médicas consiste em escanear uma seção transversal do corpo humano por um raio-X. A intensidade do raio-X que atravessa a seção do corpo humano é medida por um detector. Nesse processo é conhecida a intensidade do raio-X emitido e obtém-se, através de medições, a intensidade do mesmo após atravessar uma seção transversal do corpo humano. Fisicamente, a diferença entre a intensidade do raio-X emitido e das medições (supondo que não tenha espalhamento) é determinado pela absorção do tecido. Sabe-se que tecidos mais densos (tumores) absorvem maiores quantidades de radiação. Esses dados são processados e produzem uma imagem bi-dimensional da seção escaneada. As figuras¹ 1.1, 1.2 e 1.3 ilustram esse processo.

Os exames tomográficos, também conhecidos popularmente como tomografias por raio-X, utilizam feixes de raios-X que são disparados por uma máquina e são transmitidos através do paciente. Veja figura 1.2. Durante o processo, os raios-X interagem com os tecidos através do efeito

¹Figuras obtidas na Internet.

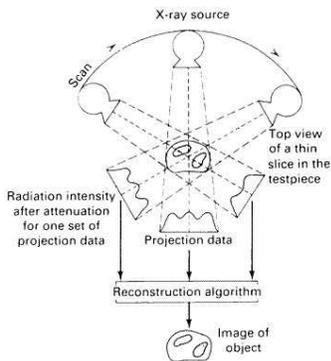


Figura 1.1: Tomografia Computadorizada.

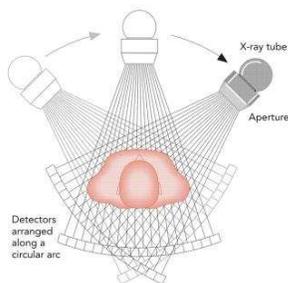


Figure 7-10 Computer tomography

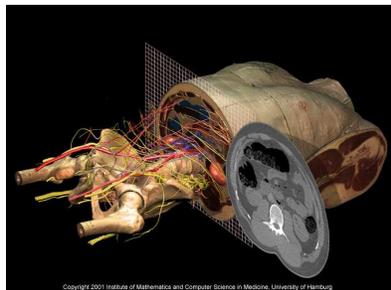


Figura 1.3: Seção Transversal.

fotoelétrico e radioativos. Como os feixes de raio-X possuem propriedades radioativas, a iteração destes feixes com os tecidos do corpo humano é prejudicial. Por isso, hoje em dia, procura-se formas mais eficazes de diagnosticar anomalias. Nessa linha, surgiram exames como a Tomografia Axial Computadorizada que, com ajuda do computador, é capaz de fornecer imagens em vários planos, de forma rápida e precisa, utilizando quantidades mínimas de radiação. Este tipo de exame é menos nocivo e fornece imagens com uma grande riqueza de detalhes. Existe uma vasta literatura sobre o assunto e fica impossível citar todas. Para os interessados, algumas das citações são [3, 15, 14, 10, 33, 32, 43, 39, 38].

Existem vários outros tipos de tomografias. Cada um destes tipos de tomografias apresenta uma forma particular de se manter competitivo no mercado de aquisição e processamento de imagens, em particular de imagens médicas. Um dos mais importantes é a *Tomografia por Ressonância Magnética (MRI - magnetic resonance imaging)*. Esse tipo de tomografia é amplamente conhecida e utilizada com o nome de ultrassonografia, [33, 39]. Outro exemplo é a *Tomografia por Emissão de Positons (PET)* que, juntamente com (MRI), tem sido amplamente usada em centros radiológicos avançados de medicina nuclear. Ainda, existe a *Tomografia por Emissão de Impedância (EIT)*. Este é um método que não usa raios radioativos para a obtenção de imagens. Ao invés de radiação uma diferença de potencial é induzida em torno do corpo da região a ser estudada, o que produz uma corrente elétrica de pequena intensidade. Técnicas para a solução de tomografia por (EIT) formam estudadas, por exemplo, em [7]. Para um apanhado geral deste tipo de tomografia e das técnicas matemáticas necessárias para estudar tais problemas, veja [3, 10, 33, 32, 39, 38].

No entanto, entre todas, a Tomografia Computadorizada é a mais usada atualmente. Existem vários motivos para tal. O primeiro deles é a simplicidade nas instalações necessárias para usar (TC) em relação as demais. O segundo é que técnicas de imagens por ressonância magnética são falhas ao examinar objetos desidratados. O terceiro, o processo de aquisição de imagens por (TC) possui clara interpretação em termos físicos e conta com o progresso na tecnologia de detectores. Também, conta com uma teoria **matemática** que explica o processo de reconstrução.

Além de imagens médicas, recentemente, aplicações da Tomografia Computadorizada em processos industriais (detecção de fraturas em estruturas), ciências florestais, arqueologia e paleontologia têm sido desenvolvidas [4, 39]. Tais aplicações buscam por métodos que permitam trabalhar com testes não destrutivos de aquisição de informações. São exemplos destes métodos a determinação de falhas em estruturas em fornos industriais ou grandes estruturas metálicas, nos quais a aquisição de informações deve ser feita indiretamente. Nessas circunstâncias, a Tomografia Computadorizada é o método escolhido para adquirir imagens.

Nesse texto, o leitor encontrará um pouco da matemática e da física que envolvem o processo de aquisição de imagens por Tomografia Computadorizada.

1.1 Descrição Matemática do Processo de Tomografia Computadorizada

Como anunciado anteriormente, o processo de aquisição de imagens por Tomografia Computadorizada é explicado por teoria matemática consistente. Nessa seção, descreveremos tal processo utilizando conceitos básicos de Cálculo.

Fisicamente, qualquer raio que passe por um obstáculo é atenuado e com o raio-X não é diferente. A atenuação da intensidade de radiação medida pelo detector que está do lado oposto à fonte de emissão com relação a um obstáculo homogêneo, pode ser modelada por um único coeficiente de absorção μ . Seja L a reta pela qual está direcionado o feixe de photons de raio-X. Chamamos de I_0 a intensidade da aplicação (input) do feixe de photons de raio-X e por I a intensidade depois de sair de Ω , veja figuras 1.4 e 1.5. Com esse modelo simples, dado pelas

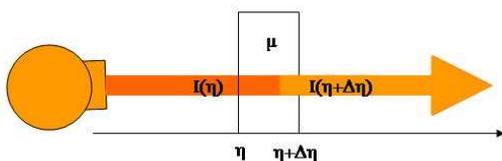


Figura 1.4: Atenuação do raio-X.

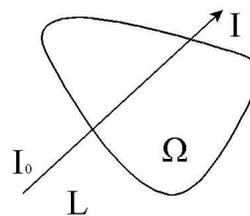


Figura 1.5: Raio-X.

figuras 1.4 e 1.5, a atenuação total de um feixe de raio-X monocromático pode ser calculada da seguinte forma: a intensidade de radiação I , após percorrer a distância $\Delta\eta$ através de um objeto, pode ser determinada por

$$I(\eta + \Delta\eta) = I(\eta) - \mu(\eta)I(\eta)\Delta\eta. \quad (1.1)$$

Reorganizando a equação (1.1), obtemos

$$\frac{I(\eta + \Delta\eta) - I(\eta)}{\Delta\eta} = -\mu(\eta)I(\eta). \quad (1.2)$$

Supondo que $\Delta\eta$ seja infinitesimal, i.e., fazendo $\Delta\eta \rightarrow 0$ em (1.2), obtemos a equação diferencial

$$\frac{dI}{d\eta} = -\mu(\eta)I(\eta). \quad (1.3)$$

Integramos (1.3) ao longo da reta L (método de separação de variáveis) e obtemos

$$I = I_0 \exp\left(-\int_L \mu(\eta)d\eta\right), \quad (1.4)$$

onde L é o comprimento do feixe. O coeficiente de absorção μ compreende os efeitos de absorção e de espalhamento do raio-X que é emitido.

A equação (1.4) é conhecida como lei de Beer, [4].

Observação: A equação (1.4) nos ensina que a intensidade de raio-X é atenuada exponencialmente ao longo de L . Para fins didáticos, consideraremos os efeitos de espalhamento desprezíveis.

De especial interesse é a quantidade

$$p(L) = -\ln\left(\frac{I}{I_0}\right) = \int_L \mu(\eta)d\eta, \quad (1.5)$$

que é a inversa da operação integral em (1.4). Esta mede a variação da intensidade de raio-X, ou seja, a atenuação do raio-X durante o processo. Denominaremos $p(L)$ em (1.5) de projeção ou projeção integral.

Na prática, as projeções $p(L)$, dadas pela equação (1.5), podem ser medidas somente sobre uma quantidade finita de retas L . Dependendo de como tais medidas são feitas, basicamente, duas formas de geometrias são obtidas. A primeira, em um modelo de escaneamento paralelo, um conjunto de integrais de linhas como (1.5) é obtido sobre um conjunto de retas L_j paralelas e igualmente espaçadas. Esse modelo requer uma única fonte de emissão e um único detector, os quais se movem e rodam (com um ângulo pré-determinado) durante o processo de escaneamento. A outra forma de escaneamento é dada por um feixe na forma de cone de raios-X (cone-beam geometry). Nesta, a fonte circula o objeto a ser escaneado e as intensidades são medidas por uma linha de detectores, simultaneamente, a cada posição da fonte. As figura 1.6 e 1.7 ilustram essas situações, respectivamente.

Em ambos os modelos descritos, o método de Tomografia Computadorizada produz somente imagens bidimensionais. No entanto, uma imagem tridimensional pode ser obtida se, de forma adequada, uma quantidade razoável de seções bidimensionais são agrupadas de maneira correta.

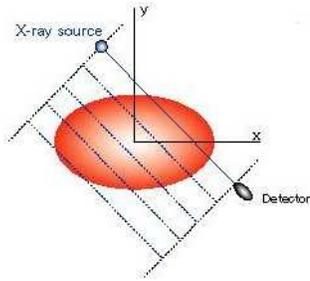


Figura 1.6: Tomografia onde os raios-X e detectores são movimentados de forma paralela.

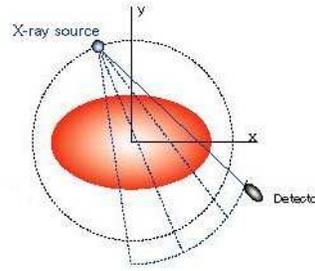


Figura 1.7: Tomografia com o feixe de raios-X é tipo cone.

Veja figura 1.8. No entanto, o procedimento de construção de uma imagem tridimensional pela



Figura 1.8: Tomografia 3D, a partir de várias tomografias 2D.

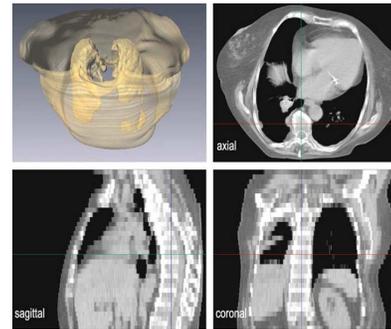


Figura 1.9: Tomografia 3D, construída a partir de uma tomografia sagittal, coronal e axial.

sobreposição de imagens bidimensionais pode ser muito custoso. Isso porque para obter uma imagem tridimensional é necessário resolver vários sub-problemas bidimensionais. Uma forma alternativa de obtenção de imagens tridimensionais é apresentada na figura 1.9. Tipicamente, três seções ortogonais são apresentadas ao radiologista: a sagittal, coronal e axial. Veja a figura 1.9. Com estas três imagens, em geral, um processo de interpolação, baseado na escala de tons de cinza, é utilizado para produzir a imagem tridimensional. Tal processo é chamado de reconstrução secundária.

Máquinas de tomografia modernas utilizam a chamada *Tomografia Computadorizada Espiral* para produzir imagens tridimensionais. Nesse caso, a fonte de emissão de raio-X é rodada em torno da região tridimensional a ser escaneada de maneira espiral. O raio-X (ou tipicamente o feixe de raio-X) é emitido e medido pelos detectores seguindo um movimento espiral em torno do paciente. Ao final desse processo, forma-se uma imagem tridimensional do objeto escaneado pelo feixe de raio-X. Para fins didáticos, nos ateremos ao caso bidimensional. Para maiores detalhes, consulte [4, 39, 38].

Os parâmetros mais importantes no processo de Tomografia Computadorizada são:

- a) a aceleração de voltagem, que determina a energia do raio-X .
- b) o tubo de corrente, que determina a intensidade de radiação.
- c) a espessura da seção, que é a espessura do feixe de raio-X da porção axial.
- d) a inclinação (angulação) da fonte emissora com relação ao eixo axial.

A relação de dependência do processo de Tomografia Computadorizada com esses parâmetros será descrita na Seção 2.4. Claro que, para tomografia do tipo espiral, o caminho espiral também é um parâmetro adicional.

Na prática, em aplicações clínicas, além de escolher um conjunto adequado de parâmetros no processo de Tomografia Computadorizada, é necessário uma etapa de planejamento para escanear as estruturas anatômicas de forma precisa, antes da aquisição da sequência de fatias dadas pela Tomografia Computadorizada. Nesta etapa de planejamento, as fatias devem ser adaptadas à situação da anatomia e, além disso, a dose de órgãos sensíveis devem ser minimizada. O planejamento é realizado com base em uma visão geral, dada por uma radiografia simples de toda a região e depois orientada a parte de maior interesse. Desta forma, a posição exata e a orientação da fatia podem ser definidas previamente.

1.2 O Problema Inverso

O problema fundamental em (TC) pode ser descrito por: *reconstruir um objeto, dada suas projeções $p(L)$* . Tais projeções são descritas por um feixe de raio-X que penetra o objeto a ser examinado, como na equação (1.5). Dependendo do caminho particular que o feixe percorre, este é atenuado quando passa através do objeto; a absorção local do raio-X é medida por um detector (esta é a projeção). É claro que, o resultado de medição em uma única direção não é suficiente para determinar a distribuição de distintas estruturas do objeto irradiado pelo raio-X. Assim, para determinar a estrutura do objeto, é necessário que o objeto seja irradiado por todas as direções. Se as diferenças entre absorção ou atenuação dos raio-X são plotadas sobre todos os ângulos de rotação, um arranjo das atenuações ou projeções são obtidas.

O **problema inverso** associado a tomografia é: Através da medida dos outputs de várias seções transversais em diferentes ângulos da região Ω , encontrar uma aproximação apropriada para a distribuição de densidade μ . Desta forma, a solução do problema inverso de reconstruir μ consiste em inverter a equação (1.5) (em um espaço de funções adequado).

O problema inverso pode ser traduzido como: É possível determinar o coeficiente de absorção μ a partir de sua integral de linha (1.5) ao longo de um conjunto de retas L ? Esta pergunta é exatamente a mesma feita por *Johann Radon* em seu famoso trabalho de 1917, [42].

A aplicação desta ideia em detecção de tumores é imediata. Suponha que um médico deseja obter informações sobre a presença de anomalias em diferentes tecidos do corpo humano. A presença de tais anomalias implicam em diferenças nos coeficientes de absorção dos tecidos. Assim, determinar o coeficiente de absorção μ , significa determinar a presença de anomalias ou não.

Exemplos deste tipo de pergunta podem ser encontrados em vários outros campos da ciência. Por exemplo, em estatística, dada todas as distribuições marginais de uma distribuição de densidade de probabilidade: É possível obter a função densidade de probabilidade? Em astrofísica, a pergunta pode ser: Observando a velocidade angular de estrelas em todas as possíveis direções, obtidas através de medidas de espectro de raios infra-vermelhos, é possível reconstruir a distribuição atual das estrelas?

A solução para o problema de Tomografia Computadorizada é complexa e envolve conhecimentos **matemáticos**, físicos e químicos. Ainda, a interpretação dos resultados envolve profissionais de medicina, etc.

Problemas como os apresentados acima são chamados de *Problemas Inversos*. Os próximos capítulos serão dedicados a descrever um pouco da teoria dos tais *Problemas Inversos* na tentativa de responder, pelo menos de maneira satisfatória, a pergunta formulada por Radon.

1.3 Um Pouco da História da Tomografia Computadorizada

Desde o descobrimento do raio-X em 1895 por Wilhelm Conrad Röntgen, muitos estudos foram feitos para entender propriedades particulares deste tipo de radiação, sua interação e interferência ao atravessar objetos. Destes estudos, novos procedimentos físicos, químicos e matemáticos foram desenvolvidos e uma quantidade significativa de respostas foram obtidas. No entanto, alguns pontos ainda permaneceram em aberto na tentativa de entender o fenômeno por completo. Estes pontos ainda são alvos de estudos de pesquisadores.

No início do século de 1900, o raidologista Italiano Alessandro Vallebona propôs um método para representar uma fatia simples de um objeto em um filme radiográfico. Este método ficou conhecido como Tomografia.

A solução matemática do *problema inverso* associado a Tomografia Computadorizada foi publicada pelo matemático Johann Radon em 1917, [42]. Dada a sua complexidade técnica e tendo em vista que a publicação foi em Alemão, os resultados só foram reconhecidos na metade do século XX.

Na Tomografia Computadorizada o significado do termo problema inverso é aparente. No processo de aquisição de imagens, a distribuição do coeficiente atenuação, que produz as projeções $p(\xi)$, não é conhecido **a priori**. Assim, para uma sequência de projeções $\{p_{\gamma_1}(\xi), \dots, p_{\gamma_n}(\xi)\}$ obtidas por medição, a distribuição espacial do coeficiente de absorção μ sobre uma seção pré-estabelecida do paciente deve ser determinada.

Em 1961, a solução deste problema foi aplicado pela primeira vez para uma sequência de projeções de raios-X de um objeto medido por diferentes direções. Allen MacLeod Cormack (1924 - 1998) e Sir Godfrey Hounsfield (1919 - 2004) foram os pioneiros em aplicações médicas da Tomografia computadorizada. Em 1979 receberam o Prêmio Nobel de Medicina pelos trabalhos na área.

A primeira máquina de tomografia foi construída no *THORN EMI Central Research Laboratories* em 1972, na Inglaterra. Esta foi construída por Godfrey Newbold Hounsfield. Uma grande parte da pesquisa foi suportada graças à contribuição da banda The Beatles, sendo considerada um dos seus maiores legados, a par com a sua música, [52].

Hoje em dia, vivemos num mundo onde impera a quarta geração da Tomografia Computadorizada, com a qual é possível obter resultados que mostram pequenos objetos e imperfeições das partes do objeto imageado. Tais gerações diferem, basicamente, pela forma em que o raio-X e os detectores são construídos e a maneira como são movimentados ao redor do paciente. Na primeira geração, um único feixe de raio-X e um único detector são usados. Estes são movidos, primeiramente, de forma paralela e depois rodados em torno do paciente por um ângulo pré-determinado. Veja figura 1.6. A segunda e terceira gerações, um feixe na forma de cone de raios-X são emitidos. Veja figura 1.7. Tal feixe é coletado por uma sequência de detectores dispostos do lado oposto do emissor, de forma a contemplar a largura total do feixe cônico. O emissor e detectores são rodados em torno do paciente, por um ângulo fixo. A diferença entre estas duas gerações está na largura do feixe cônico de raios-X emitidos e, conseqüentemente, da largura da sequência de detectores. A última geração, novamente, um feixe cônico de raios-X é emitido, mas desta vez, os detectores estão fixos em torno de todo paciente.

Um apanhado histórico muito interessante sobre a evolução do processo de Tomografia Computadorizada pode ser encontrado em [35].

Capítulo 2

Introdução aos Problemas Inversos Lineares

Neste capítulo, apresentaremos uma breve introdução aos "Problemas Inversos" e mal postos. Iniciaremos com um breve apanhado histórico de uma das áreas da matemática aplicada que despertou o interesse de muitos pesquisadores, centros de pesquisa e da indústria nas últimas décadas. Discutiremos alguns exemplos de problemas inversos, suas aplicações em áreas diversificadas do conhecimento, dando uma síntese de algumas aplicações possíveis. Os problemas que apresentaremos têm o intuito de dar ênfase aos principais pontos dos problemas teóricos e práticos envolvendo a pesquisa em problemas inversos.

Neste texto, estamos interessados, principalmente, em problemas inversos associados a Tomografia Computadorizada. Nas referências, o leitor interessado pode buscar mais informações sobre aplicações.

Esperamos que o leitor possa acompanhar os raciocínios com conhecimentos básicos de Análise (Espaços Métricos) [31, 30, 29] e Álgebra Linear [11, 36, 49] e generalizações simples desse conceito para dimensão infinita.

2.1 Breve Apanhado Histórico dos Problemas Inversos

O estudo de problemas inversos é muito novo e também muito velho. Há cerca de dois milênios atrás, no livro VII do diálogo "*República*", Platão¹ propôs o filosófico problema de reconstruir a "realidade" através da observação da imagem de objetos ou pessoas, cujas sombras eram projetadas na penumbra de uma caverna. Com a ideia de discutir aspectos filosóficos das fontes de conhecimento humano, Platão, também acabou introduzindo o primeiro exemplo de problemas inversos que se tem relatos, [40].

¹Platão viveu entre 427 – 347 a.c.

Nas ciências aplicadas, possivelmente, um dos primeiros problemas inversos data de 200 anos antes de Cristo. Eratóstenes² propôs o problema de determinar o diâmetro da terra através de medições feitas em duas cidades distintas. Eram conhecidas a distância entre as cidades, as suas latitudes e dado o ângulo que a sombra de um marco vertical (em cada uma destas cidades) fazia com a direção do sol, [14].

Em 1800, Gauss³ fez uso do método de quadrados mínimos para reconstruir a órbita de um cometa a partir de dados de órbitas anteriores. Este também é um exemplo de problemas inversos.

A transformada, que hoje em dia chamamos de *Transformada de Radon*, é uma das precursora do estudo de Tomografia computadorizada [2, 9, 53]. A solução do problema inverso associado a Transformada de Radon foi publicada por Radon⁴ em 1917. A solução deste problema inverso permite determinar o coeficiente de absorção μ de cada parte do objeto escaneado.

Nas últimas quatro décadas um novo campo de estudos na área de matemática aplicada tem conquistado uma grande quantidade de pesquisadores adeptos. Este campo trata de problemas como os formulados por Platão, Eratóstenes entre outros, cuja abordagem exige o desenvolvimento de métodos matemáticos como os apresentados por Gauss e Radon. A essa área de estudos denominamos *Problemas Inversos*.

A área de *Problemas Inversos* se desenvolveu rapidamente nos últimos anos. O súbito crescimento deve-se, certamente, ao grande número de aplicações em outras ciências e o aparato de novas técnicas e teorias matemáticas envolvidas no ataque a tais problemas.. Por exemplo, em geofísica e ciências ambientais, explorações sísmicas (como detecção de depósito de petróleo e outras riquezas ou poluentes no subsolo)[34, 53], ciências médicas e tomografias (com ênfase na reconstrução de imagens, ultasonografia)[3, 10, 33, 38, 39], em engenharia (detecção de fraturas em estruturas, testes não-destrutivos em componentes de semi-condutores e nanotecnologia)[2], física, química, biologia, finanças, entre outras.

Além da relevância das aplicações, a formulação e solução de tais problemas envolvem o conhecimento de vários campos da matemática, de ciências aplicadas e o envolvimento de profissionais dessas áreas.

Segue uma lista considerável de Livros sobre o assunto, onde os interessados em tais problemas podem basear a sua pesquisa [1, 2, 9, 13, 23, 25, 47, 53].

2.2 O que são Problemas Inversos ?

Problemas que envolvem a determinação de uma dada **causa**, observado (ou medido) um dado **efeito**, possuem uma vasta quantidade de aplicações em várias áreas da ciência.

²Eratóstenes viveu entre 284-202 a.c.

³C. F. Gauss viveu entre 1777 – 1855

⁴J. Radon viveu entre 1887 – 1956

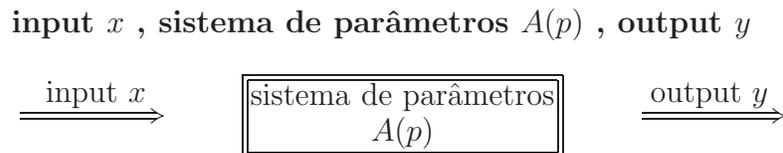
Desde o início, o leitor deve estar se perguntando: o que são os tais ”**Problemas Inversos**”? Ou ainda, a pergunta pode ser: ”**inversos**” do quê? Para J.B. Keller [24], dois problemas são o inverso um do outro, se a formulação de um envolve o conhecimento (mesmo que parcial) do outro, este último, conhecido como o ”**Problema Direto**”. Assim, a grosso modo, *problemas inversos* estão relacionados com a determinação de **causas**, através da observação (ou medida) de **efeitos**.

Do ponto de vista de aplicações, existem pelo menos duas motivações distintas para estudar ”Problemas Inversos”. A primeira é movida pela curiosidade humana de conhecer estados físicos passados ou parâmetros em um sistema físico que descreve certos modelos. Como exemplos temos os estudos de mudanças climáticas drásticas a milhões de anos atrás, através de medidas observáveis hoje nas camadas glaciais das calotas polares. A segunda, é prever os fenômenos futuros influenciados pelos estados atuais ou por parâmetros de um sistema físico. Ambas motivações são modeladas por equações matemáticas [21, 20, 50].

Associado ao estudo e solução de problemas inversos estão fatores relevantes no desenvolvimento de uma nação. Por exemplo, problemas inversos em imagens médicas influenciam em

- **fatores sociais:** técnicas de detecção de tumores implicam em prolongar a vida das pessoas.
- **fatores econômicos:** detecção de tumores implica em tratamentos mais eficazes contra o câncer, diminuindo os custos dos mesmos. Ainda, prolonga a vida ativa das pessoas que, conseqüentemente, geram mais riquezas.
- **desenvolvimento tecnológico:** desenvolvimento de novos métodos e máquinas de tomografia para a obtenção de imagens médicas.

Sintetizando as idéias acima, podemos assumir que o fenômeno físico (biológico e etc) a ser estudado é modelado por um processo que envolve três quantidades principais do modelo:



O problema direto. Dados o input (**causa**) e o sistema de parâmetros, determinar o output do modelo (**efeito**).

O problema inverso. Esse pode aparecer de duas formas.

1. *O problema de reconstrução:* Dado o sistema de parâmetros e observado o output (**efeito**), encontrar que input (**causa**) corresponde ao output.
2. *O problema de identificação.* Dados input (**causa**) e output (**efeito**), determinar o sistema de parâmetros que relaciona input/output.

De um modo geral, representamos um problema inverso por equações do tipo

$$A(p)x = y, \quad (2.1)$$

para um dado efeito y ; uma causa a ser determinada x (que pertence a um conjunto de parâmetros admissíveis $p \in \mathcal{U}$) e $A(p)$ representa o modelo que manda a causa no determinado efeito.

Em termos práticos, os dados y dificilmente são obtidos de forma precisa, dada a natureza da obtenção destes dados (medidas). Assim, costumamos denotar as medições obtidas por y^δ , das quais assumimos conhecer o nível de **ruídos** δ , de forma a satisfazer

$$\|y - y^\delta\| \leq \delta. \quad (2.2)$$

Numa formulação matemática, $A(p)$ é um operador (matriz) definido entre espaços vetoriais que, para nossos objetivos, consideraremos espaços de Hilbert \mathcal{H}_1 e \mathcal{H}_2 , com (veja Apêndice A). Para uma boa quantidade de aplicações, o operador $A(p)$ é linear e a teoria já é amplamente desenvolvida. Veja, por exemplo, [1, 9, 13, 25]. Uma grande quantidade de problemas de interesse envolvem um operador $A(p)$ não-linear, [23, 16, 9, 2]. Para operadores não-lineares, a teoria é bem mais complicada.

Nestas notas, daremos ênfase a teoria matemática envolvida na solução de problemas inversos lineares, uma vez que, o problema de interesse (tomografia) é um problema linear. O estudo de métodos de solução será feito no decorrer dos Capítulos 3, 4, 5 e 6.

Na formulação matemática dos problemas podemos caracterizar:

O problema direto: Dado $x \in \mathcal{H}_1$ e $p \in \mathcal{U}$, encontrar $y := A(p)x$.

O problema inverso: Esse aparece, pelo menos de duas formas:

1. *O problema de reconstrução:* Observado $y \in \mathcal{H}_2$ e conhecido sistema de parâmetros $A(p)$ para $p \in \mathcal{U}$, encontrar $x \in \mathcal{H}_1$ tal que $A(p)x = y$.
2. *O problema de identificação.* Dados $x \in \mathcal{H}_1$ e $y \in \mathcal{H}_2$, encontrar $p \in \mathcal{U}$ tal que $A(p)x = y$.

Uma característica que diferencia um **problema direto** de um **problema inverso** é que, o segundo, é *mal-posto* no sentido de Hadamard [18]. Um problema é dito *bem posto* no sentido de Hadamard se satisfaz as condições de existência, unicidade e dependência contínua dos dados iniciais. Caso um dos requerimentos acima não seja satisfeito, o problema é dito mal-posto.

Problemas inversos em dimensão finita aparecem naturalmente na solução de sistemas de equações, [19]. Do ponto de vista computacional [51, 19], sempre estamos tratando de problemas inversos em dimensão finita, assim, faz-se jus estudá-los com o devido interesse. Ainda, os problemas de mal-condicionamento estão intimamente ligados aos autovalores da matriz que representa o sistema [49, 36, 11]. Se os autovalores são próximos de zero ou os dados não pertencem ao espaço solução do problema, podemos enfrentar sérias complicações numéricas [49, 36, 11].

Em termos práticos, estamos sempre em dimensão finita, pois, não somos capazes de atribuir quantidades infinitas a um programa de simulação computacional, muito embora, a verdadeira solução viva em espaço de dimensão infinita. Voltaremos a esse assunto no Capítulo 3.

Exemplos importantes de problemas inversos lineares são dados por equações integrais de primeira espécie ou por algumas transformadas integrais [3, 13, 19, 38, 43, 48]. Apresentaremos o problema da tomografia por raio-X na Seção 2.4 como um problema inverso formulado por uma equação integral.

Problemas inversos de identificação de parâmetros em equações diferenciais parciais são, em geral, não-lineares, mesmo que o problema direto seja linear [9, 23, 21]. Nestas notas, não trataremos da teoria para estes problemas, muito embora sejam muito importantes. Por exemplo, o problema associado a *Tomografia Elétrica por Impedância - (EIT)* é não-linear. Interessados numa introdução ao problema de (EIT) podem consultar, por exemplo, [3, 6, 7, 39] e referências.

2.3 O Problema Inverso da Diferenciação

Dada a natureza suavizante do operador de integração, costumamos considerar a diferenciação como sendo o problema inverso da integração.

Nesta seção, apresentaremos dois exemplos que deixam claro a afirmação acima.

2.3.1 Reconstrução de uma Força Desconhecida

Considere um sistema mecânico com uma dinâmica de forças atuantes, cujas medidas não possam ser realizadas diretamente. Assim, tais medidas são dada pela instabilidade de algum tipo de dinamômetro e, portanto, sujeitas a erros.

Um modelo simplificado, onde um único grau de liberdade no sistema mecânico é considerado, pode ser descrito pela equação diferencial ordinária

$$m\ddot{x} + kx = y(t), \quad t > 0, \quad (2.3)$$

onde, m é a massa do sistema, k é a constante de rigidez, $\ddot{}$ indica derivadas, x é a função que descreve o deslocamento e y é a função de forças desconhecidas do sistema.

O **problema direto** associado é: dado a dinâmica de forças atuantes $y(t)$, encontrar x solução da EDO (2.3).

Exercício 2.3.1. *Determine a solução da E.D.O. (2.3), no caso do sistema de forças*

$$y(t) = 0.$$

O **problema inverso** associado é: encontrar a dinâmica de forças y , a partir de medidas das vibrações respostas do sistema, que são obtidas por medições e, portanto, sujeitas a erros.

Caso x seja conhecido exatamente, recuperamos y facilmente, bastando para tal, substituir x na E.D.O. (2.3).

Mas, no caso em que x é contaminado por uma função de ruídos η , isto é, só temos informações sobre $x^\delta = x + \eta$, não podemos substituir x^δ diretamente na E.D.O. (2.3), pois, não sabemos se a função η é duas vezes diferenciável. Ainda, caso η seja duas vezes diferenciável, $m\ddot{\eta}$ pode ser altamente oscilatória. Esta oscilação leva a soluções muito ruins.

Por exemplo, tome $m = k = 1$ e $x(t) = \exp(-t)$ como solução da EDO (2.3) com $y(t) = 0$. Suponha que conhecemos somente o dado perturbado

$$x^\delta(t) := \exp(-t) + a \sin(\omega t), \quad t > 0.$$

Substituindo x^δ na E.D.O. (2.3), obtemos como resposta

$$y^\delta(t) := a(1 - \omega^2) \sin(\omega t), \quad t > 0.$$

Note que a função $y^\delta(t)$ está muito longe da solução para dados sem ruídos, se ω é muito grande.

Exercício 2.3.2. Calcule o erro cometido na solução do problema acima como uma função de ω . Compare com o erro nos dados. O que podemos concluir (em função de w)? Use a norma no espaço das funções contínuas, isto é, a norma do supremo $\|\cdot\|_\infty$.

Exercício 2.3.3 (O conjunto das funções quadrado integráveis em $[a, b]$).] Denote por $C[a, b]$ o conjunto das funções contínuas no intervalo $[a, b]$ (e assim, uniformemente contínuas (prove!)). Considere a seguinte função:

$$\begin{aligned} \|\cdot\|_2 : C[a, b] \times C[a, b] &\longrightarrow \mathbb{R} \\ (x, y) &\longmapsto \|x(t) - y(t)\|_2 = \left(\int_a^b (x(t) - y(t))^2 dt \right)^{\frac{1}{2}}. \end{aligned}$$

i) Mostre que $\|\cdot\|_2$ é uma norma em $C[a, b]$.

ii) Mostre que $C[a, b]$ não é completo com a norma $\|\cdot\|_2$.

Definimos por $L^2[a, b]$ o completamento de $C[a, b]$ com relação a norma $\|\cdot\|_2$. Tal completamento sempre existe. Veja [31, 30, 29].

1- Usando os resultados do exercício acima, faça a comparação com o erro nos dados e na solução usando a norma de L^2 .

2- O que podemos concluir (em função de w)?

3- Existe diferença em medir erros com a norma $\|\cdot\|_\infty$ e $\|\cdot\|_2$?

2.3.2 Diferenciação nos Dados

Sejam, $y : [0, 1] \rightarrow \mathbb{R}$ e $y^\delta : [0, 1] \rightarrow \mathbb{R}$ funções contínuas com y^δ contaminada por ruídos de forma que

$$\|y(t) - y^\delta(t)\|_\infty \leq \delta, \quad \forall t \in [0, 1].$$

Gostaríamos de reconstruir a derivada $x = y'$ de y . Uma das estratégias (pelo menos do ponto de vista numérico) é considerar aproximações por diferenças simétricas, i.e., para qualquer $\tau \in (0, 1)$ tomar

$$x^{\delta,h}(\tau) := \frac{y^\delta(\tau+h) - y^\delta(\tau-h)}{2h}.$$

Um simples argumento com a desigualdade triangular fornece

$$\|x^{\delta,h}(\tau) - x(\tau)\|_\infty \leq \left\| \frac{y(\tau+h) - y(\tau-h)}{2h} - x(\tau) \right\|_\infty + \left\| \frac{(y^\delta - y)(\tau+h) - (y^\delta - y)(\tau-h)}{2h} \right\|_\infty.$$

Suponha que tenhamos o limitante

$$\|x'(t)\|_\infty \leq E, \quad \forall t \in [0, 1].$$

Substituindo o limitante na desigualdade acima, obtemos a estimativa de erro

$$\|x^{\delta,h}(\tau) - x(\tau)\|_\infty \leq hE + \frac{\delta}{h}. \quad (2.4)$$

A equação (2.4) é típica em problemas inversos e reaparecerá novamente em nossas notas. O que é importante, por agora, é entender o que a equação (2.4) quer nos dizer. Temos dois termos nessa estimativa de erro: um devido a aproximação da aplicação inversa e o outro devido ao erro de medida. Observe que, quanto mais refinarmos a aproximação (quanto mais próximo de zero tomarmos h) mais precisamente estamos calculando a derivada y' . Por outro lado, como os dados estão corrompidos por erros, (2.4) nos ensina que, se h for tomado muito pequeno, então $x^{\delta,h}$ pode estar longe da solução verdadeira.

O melhor que podemos fazer é escolher h de forma a balancear o lado direito de (2.4). Ou seja, tomar

$$h(\delta) := E^{-\frac{1}{2}}\delta^{\frac{1}{2}}.$$

A figura 2.1 ilustra a escolha de $h(\delta)$.

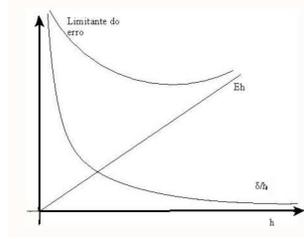


Figura 2.1: Estimativa fundamental.

2.4 A Tomografia Computadorizada e a Transformada de Radon

Entre os problemas que têm chamado a atenção da comunidade científica estão alguns problemas em Tomografia⁵ [43, 39, 38]. As aplicações percorrem os campos de ciências médicas: detecção de tumores, fraturas na estrutura óssea, etc (tomografias por raio-X), [10, 33, 32]; ciências ambientais: detecção depósitos de sedimentos, prospecção de petróleo (tomografias acústicas), [10, 33, 32, 43, 39, 38]. Estes problemas, em geral, consistem em recuperar a forma e a localização de um objeto imerso (ou de sua densidade) em uma região do plano \mathbb{R}^n a partir de medidas (parciais) sobre a fronteira da região. Essas medidas, na maioria dos casos, são adquiridas por um número reduzido de experimentos, [17]. Assim, uma característica comum destes tipos de problema é a falta de informações nos dados. Problemas como os citados fazem parte da pesquisa em Problemas Inversos, dada a grande utilidade comercial, bem como, sua complexidade matemática.

Nessas notas, nosso interesse está no tratamento matemático da Tomografia Computadorizada. Para tal, faremos uso da transformada conhecida como Transformada de Radon (bidimensional) e de sua Transformada Inversa. Também, abordaremos aspectos práticos da Tomografia Computadorizada, estudando o modelo discreto da Transformada de Radon. Para um tratamento completo dos demais tipos de tomografia, consulte, por exemplo, [10, 33, 32, 43, 39, 38].

2.4.1 Tomografia Computadorizada: caso contínuo

A Tomografia Computadorizada (TC) ou Tomografia Axial Computadorizada (TAC), auxilia na obtenção de diagnósticos baseados em imagens. Tais imagens consistem numa secção ou "fatia" do corpo. Estas são obtidas através do processamento por computador de informação recolhida após expor o corpo a uma sucessão de raios-X. No entanto, para que um processo computacional tenha sucesso, é necessário um estudo matemático detalhado do problema.

A (TC) baseia-se nos mesmos princípios que a radiografia convencional, i.e., os tecidos com diferentes composições absorvem a radiação emitida pelo raio-X de forma diferente. Ao serem

⁵Existem várias formas de Tomografias. Nessas notas, daremos ênfase especial a Tomografia Computadorizada.

atravessados por raios-X, tecidos mais densos (como o fígado) ou com elementos mais pesados (como o cálcio presente nos ossos), absorvem mais radiação que tecidos menos densos (como o pulmão, que está cheio de ar). Assim, um processo de (TC) indica a quantidade de radiação absorvida por cada parte do corpo analisada (radiodensidade), e traduz essas variações numa escala de cinzentos, produzindo uma imagem. Cada pixel da imagem corresponde à média da absorção dos tecidos na região escaneada pelo processo.

Descrição do Problema:

Uma maneira bastante intuitiva de compreender o processo de reconstrução de imagens usando (TC) é considerar o caso em que possuímos apenas um feixe de raio-X em forma de uma reta. Este feixe é movimentado de forma paralela a distribuição linear dos detectores. De forma sequencial, o conjunto de detectores é rotacionado por um ângulo γ , de forma que todo o objeto de interesse seja escaneado. O processo é representado pela figura 2.2

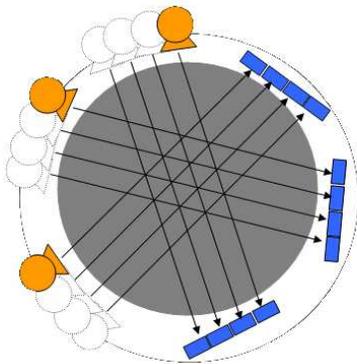


Figura 2.2: Representação de uma Tomografia em paralelo.

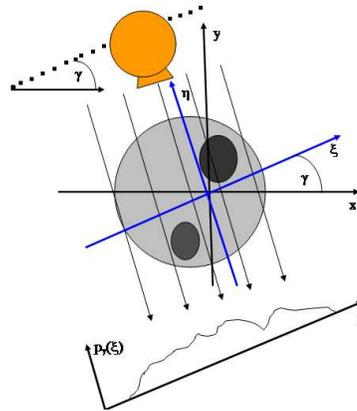


Figura 2.3: Mudança de Variável.

Assim, para um ângulo γ fixo e para uma posição em particular da fonte de raio-X, o processo é descrito pela projeção integral (1.5).

Por outro lado, existe um número de raios-X paralelos que passam pelo objeto a ser escaneado, os quais definem a seção plana ou a fatia da (TC). Desta maneira, é razoável reescrevermos a equação (1.5) em função da posição da fonte de raio-X - ou da localização do correspondente detector ξ - e do ângulo de projeção γ . Assim, consideraremos o sistema de coordenada (ξ, η) , o qual gira, juntamente com o raio-X e a fonte de detectores, ao invés do sistema de coordenadas fixo (x, y) . A figura 2.3 mostra a correspondência entre os dois sistemas.

No novo sistema de variáveis, a equação (1.5) fica escrita como

$$p_\gamma(\xi) = \int_L \mu(\xi, \eta) d\eta. \quad (2.5)$$

A equação (2.5) é uma integral de linha ao longo do segmento de reta L que descreve a posição ξ

da fonte de raio-X e do respectivo detector, com um ângulo de projeção γ em relação ao plano de coordenadas (x, y) . Veja figura 2.3. Na prática, os valores do coeficiente de atenuação devem ser dados como uma função das coordenadas fixas (x, y) . Para descrever a relação entre os sistemas de coordenadas (ξ, η) e (x, y) , defina $\mathbf{n}_\xi = (\cos(\gamma), \sin(\gamma))^T$ e $\mathbf{n}_\eta = (-\sin(\gamma), \cos(\gamma))^T$ o span do do sistema de coordenadas (ξ, η) . Com essa notação, temos que

$$\xi = x \cos(\gamma) + y \sin(\gamma) \quad (2.6)$$

$$\eta = -x \sin(\gamma) + y \cos(\gamma). \quad (2.7)$$

Substituindo (2.6) e (2.7) no coeficiente de atenuação $\mu(\xi, \eta)$, obtemos o coeficiente de atenuação nas variáveis (x, y) definido por

$$f(x, y) = \mu(x \cos(\gamma) + y \sin(\gamma), -x \sin(\gamma) + y \cos(\gamma)). \quad (2.8)$$

Do ponto de vista físico, os coeficientes de atenuação $f(x, y)$ e $\mu(\xi, \eta)$ são os mesmos. Então, se considerarmos $I_0 = 1$ (normalizado) na equação (1.5) e parametrizando a reta $L(w, s)$, que é perpendicular a w e que está a uma distância s da origem, (veja figura 2.4), obtemos

$$p_\gamma(\xi) = \mathcal{R}f(w, s) = \int_{L(w, s)} f dl, \quad s \in \mathbb{R}, \quad \|w\|_2 = 1. \quad (2.9)$$

$\mathcal{R}f$ é chamada de *Transformada de Radon* bi-dimensional de f . Esta é uma aplicação que leva funções de $x \in \mathbb{R}^2$ em funções de $(w, s) \in \mathbb{S}^1 \times \mathbb{R}$, onde \mathbb{S}^{n-1} denota a esfera unitária em \mathbb{R}^n .

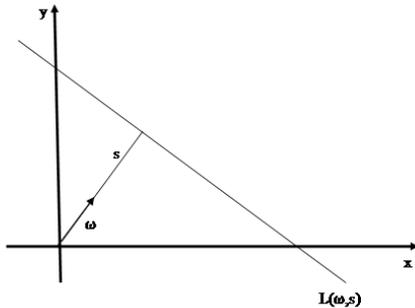


Figura 2.4: Parametrização da reta L que é perpendicular a w a uma distância s da origem.

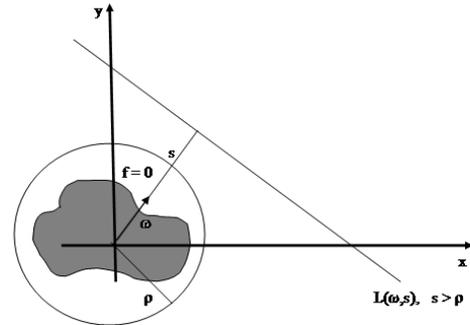


Figura 2.5: Se $f(x, y) = 0$, quando $|(x, y)| > \rho$, então $\mathcal{R}f(w, s) = 0$ quando $s > \rho$.

O **problema inverso** associado a tomografia é: Encontrar uma aproximação apropriada para a distribuição de densidade f através da medida dos outputs de várias seções transversais em diferentes ângulos da região Ω . Ou seja, a solução do problema inverso de reconstruir a densidade

f , consiste em inverter o operador \mathcal{R} (em um espaço de funções adequado).

É possível determinar o coeficiente de absorção f a partir de sua integral de linha (2.9) ao longo da reta L ? A pergunta é exatamente a mesma feita por *Johann Radon* em seu famoso trabalho de 1917 [42].

A tentativa de responder a pergunta acima, nos leva a várias outras perguntas.

1. \mathcal{R} é injetiva?
2. Qual é a imagem de \mathcal{R} ?
3. É possível encontrar uma fórmula para a inversa de \mathcal{R} ?
4. Se $\mathcal{R}f(w, s) = 0$ para $|s| > \rho$ e para todo $w \in \mathbb{S}^1$, é verdade que $f(x, y) = 0$ para todo $x \in \mathbb{R}^2$ com $|x| > \rho$. Veja figura 2.5.

Exercício 2.4.1. *Verifique que a recíproca da questão (4) é verdadeira.*

Do ponto de vista prático, é impossível obter medidas sobre todas as retas que passam por Ω . Deste fato, surge uma outra pergunta, a qual, talvez, seja a mais difícil. **Quantas medidas são necessárias para obtermos uma boa aproximação?** Na próxima seção, apresentaremos uma abordagem discreta da Tomografia Computadorizada onde aparecerá, naturalmente, tal questão. Para mais detalhes consulte [38, 39].

2.4.2 A Transformada Inversa de Radon

Agora, abordaremos as questões (1) - (4) do ponto de vista matemático. Os resultados que apresentaremos podem ser encontrados em [46] e referência. Para isso faremos uso da Transformada de Fourier. Para os que não tenham tido contato com essa importante ferramenta matemática, aqui está uma ótima oportunidade para conhecer. Recomendamos uma breve olhada no Apêndice A dessas notas.

O conjunto

$$H(w, s) = \{x \in \mathbb{R}^n : \langle x, w \rangle = s\} \quad (2.10)$$

é um hiperplano a uma distância $|s|$ da origem com vetor normal $w \in \mathbb{S}^{n-1}$. Aqui, $\langle \cdot, \cdot \rangle$ representa o produto interno de \mathbb{R}^n . Veja Apêndice.

Dada $f \in \mathcal{S}(\mathbb{R}^n)$, considere a integral de superfície da função f ao longo do hiperplano $H(w, s)$, isto é:

$$\mathcal{R}f(w, s) = \int_{H(w, s)} f dS. \quad (2.11)$$

Exercício 2.4.2. Mostre que a integral de superfície (2.11) é convergente para $f \in \mathcal{S}(\mathbb{R}^n)$.

Exercício 2.4.3. Mostre que

$$\mathcal{R}f(w, s) = \mathcal{R}f(-w, -s). \quad (2.12)$$

Note que, dado $k \in \mathbb{N}$

$$\int_{\mathbb{R}} s^k \mathcal{R}f(w, s) ds = \int_{\mathbb{R}} \int_{H(w, s)} (\langle x, w \rangle)^k f(x) dS ds. \quad (2.13)$$

Quando calculamos $(\langle x, w \rangle)^k$, obtemos um polinômio homogêneo de grau k em w , com coeficientes que também são polinômios homogêneos de grau k na variável x . Já, o resultado da integral (2.13) é um polinômio de grau k em w .

Com isso obtemos que: $F(w, s)$ está na imagem de \mathcal{R} se satisfizer:

$$F(w, s) = F(-w, -s) \quad (2.14)$$

$$\int_{\mathbb{R}} s^k F(w, s) ds = P_k(w) \quad (2.15)$$

é um polinômio homogêneo de grau k em w .

Exercício 2.4.4. Demonstre que

$$\mathcal{R}f(w, s) = \int_{w^\perp} \int_0^\infty f(sw + t\theta) t^{n-2} dt d\theta,$$

onde $w^\perp = \{\theta \in \mathbb{S}^{n-1} : \langle w, \theta \rangle = 0\}$.

Exercício 2.4.5. Uma função $f : \mathbb{R}^n \rightarrow \mathbb{R}$ é dita uma função radial, se existir uma função $g : \mathbb{R} \rightarrow \mathbb{R}$ tal que $f(x) = g(\|x\|)$, $\forall x \in \mathbb{R}^n$.

Mostre que se f é radial, então

$$\mathcal{R}f(w, s) = \int_{\mathbb{R}^{n-1}} g(\sqrt{s^2 + y^2}) dy = 2 \frac{\pi^{(n-1)/2}}{\Gamma((n-1)/2)} \int_0^\infty g(\sqrt{s^2 + t^2}) t^{n-2} dt. \quad (2.16)$$

Lema 2.4.1. Seja $f \in \mathcal{S}(\mathbb{R}^n)$. Então, a transformada de Fourier de $\mathcal{R}f(w, s)$ com relação a f satisfaz

$$\widehat{\mathcal{R}f}(w, \lambda) = \int_{\mathbb{R}} \mathcal{R}f(w, s) e^{-i\lambda s} ds = \hat{f}(\lambda w). \quad (2.17)$$

Demonstração : Seja $k \in \mathbb{N}$. Pela desigualdade de Cauchy-Schwarz, [36],

$$|s|^k = |\langle w, s \rangle|^k \leq |x|^k.$$

Assim,

$$|s|^k |\mathcal{R}f(w, s)| \leq \int_{H(w, s)} |\langle w, s \rangle|^k |f(x)| dS \leq \int_{H(w, s)} |x|^k |f(x)| dS < \infty.$$

Com isso, a integral (2.17) converge (justifique). Pelo Teorema de Fubini [31],

$$\begin{aligned} \int_{\mathbb{R}} \mathcal{R}f(w, s) e^{-i\lambda s} ds &= \int_{\mathbb{R}} e^{-i\lambda s} \left(\int_{H(w, s)} f(x) dS \right) ds = \int_{\mathbb{R}} \left(\int_{H(w, s)} e^{-i\lambda \langle w, x \rangle} f(x) dS \right) ds \\ &= \int_{\mathbb{R}^N} e^{-i\lambda \langle w, x \rangle} f(x) dx = \hat{f}(\lambda w) \end{aligned}$$

■

Exercício 2.4.6. Mostre que \mathcal{R} é uma aplicação linear.

Lema 2.4.2. Seja $f \in \mathcal{S}(\mathbb{R}^n)$. Então $\mathcal{R}f(w, s) = 0$ se e só se $f(w, s) = 0$.

Demonstração: Como a Transformada de Fourier é um isomorfismo, temos que $\mathcal{R}f(w, s) = 0$ se e só se $\widehat{\mathcal{R}f}(w, s) = 0$. Por (2.17) se e só se $\hat{f}(\lambda w) = 0$. Novamente pelo isomorfismo da Transformada de Fourier, se e só se $f = 0$. ■

Exercício 2.4.7. Seja

$$\begin{aligned} P : \mathbb{S}^{n-1} \times (0, \infty) &\longrightarrow \mathbb{R}^n \\ (w, s) &\longmapsto sw = x. \end{aligned}$$

- i) Mostre que P é um isomorfismo e $P^{-1}(x) = (x/\|x\|, \|x\|)$ é a inversa de P .
- ii) Demonstre que se $f \in C^\infty(\mathbb{R}^n)$, então $f \circ P \in C^\infty(\mathbb{S}^{n-1} \times [0, \infty))$.
- iii) Seja $F \in C^\infty(\mathbb{S}^{n-1} \times \mathbb{R})$. Mostre que $F \circ P^{-1} = F(x/\|x\|, \|x\|) \in C^\infty(\mathbb{R}^n - \{0\})$.
- iv) Seja $F \in C^\infty(\mathbb{S}^{n-1} \times \mathbb{R})$. Demonstre que $f = F \circ P^{-1} \in C^\infty(\mathbb{R}^n)$. Ainda:

$$\frac{\partial^k}{\partial s^k} F(w, s) \Big|_{s=0} = P_k(w),$$

é um polinômio homogêneo de grau k , para todo $k \in \mathbb{N}$.

Defina o conjunto

$$\mathcal{S}(\mathbb{S}^{n-1} \times \mathbb{R}) := \left\{ F \in C^\infty(\mathbb{S}^{n-1} \times \mathbb{R}) : |s|^k \left| \frac{\partial^l}{\partial s^l} \frac{\partial^{m_1}}{\partial w_1^{m_1}} \cdots \frac{\partial^{m_n}}{\partial w_n^{m_n}} F(w, s) \right| \leq C(k, l, m_1, \dots, m_n) \right\}.$$

Temos:

Teorema 2.4.1. A transformada de Radon é uma aplicação bijetora

$$\mathcal{R} : \mathcal{S}(\mathbb{R}^n) \longrightarrow M := \{ \mathcal{S}(\mathbb{S}^{n-1} \times \mathbb{R}) : F \text{ satisfaz (2.14) e (2.15)} \} \quad (2.18)$$

Demonstração : A injetividade foi provada no Lema 2.4.2. Seja $F(w, s) \in M$. Como

$$\int_{\mathbb{R}} s^k F(w, s) ds = \widehat{s^k F}(w, 0) = i^k \frac{\partial^k}{\partial s^k} \hat{F}(w, 0)$$

é, por hipótese, um polinômio homogêneo de grau k . Fazendo $\xi = \lambda w$, pelo exercício 2.4.7, $f(\xi) = \hat{F}(\xi/\|\xi\|, \|\xi\|) \in C^\infty(\mathbb{R}^n)$. Como $F \in M$, obtemos que $f \in \mathcal{S}(\mathbb{R}^n)$. De (2.17) deduzimos que $F = \mathcal{R}\mathcal{F}^{-1}f$. ■

Como a Transformada de Radon é uma bijeção, cabe a pergunta: Será que é possível encontrar uma forma analítica para \mathcal{R}^{-1} ? A idéia é utilizar a relação entre a Transformada de Radon e a Transformada de Fourier obtida na (2.17).

Exercício 2.4.8. *Mostre que:*

$$\mathcal{F}^{-1}(\lambda^{n-1} \widehat{\mathcal{R}f}(w, \lambda)) = 2\pi \frac{1}{i^{n-1}} \frac{\partial^{n-1}}{\partial s^{n-1}} \mathcal{R}f(w, s).$$

Para encontrarmos uma expressão analítica para \mathcal{R}^{-1} temos que considerar o caso n par e n ímpar. A demonstração de cada caso difere muito. Veja [46].

Teorema 2.4.2. *Seja $n > 1$ ímpar. Então*

$$f(x) = 2^{-n} (i\pi)^{1-n} \int_{\mathbb{S}^{n-1}} \frac{\partial^{n-1}}{\partial s^{n-1}} \mathcal{R}f(w, \langle x, w \rangle) dS \quad (2.19)$$

Demonstração : Usando a Transformada de Fourier inversa em coordenadas polares e o fato de $n - 1$ ser par temos que

$$f(x) = \frac{1}{(2\pi)^n} \int_{\mathbb{S}^{n-1}} \int_0^\infty e^{i\langle x, \lambda w \rangle} \hat{f}(\lambda w) \lambda^{n-1} d\lambda dS_w = \frac{1}{2(2\pi)^n} \int_{\mathbb{S}^{n-1}} \int_{-\infty}^\infty e^{i\langle x, \lambda w \rangle} \hat{f}(\lambda w) \lambda^{n-1} d\lambda dS_w.$$

Pela (2.17) e o exercício 2.4.8 concluímos a demonstração. ■

A fórmula (2.19) possui uma propriedade interessante que deve ser destacada. Para obtermos $f(x_0)$, é necessário conhecer os valores de $\mathcal{R}f(w, s)$ para $s = \langle w, x_0 \rangle$. Ou seja, não precisamos conhecer as integrais de f ao longo de todos os planos $H(w, s)$, basta obtermos informações sobre os que distam $\langle w, x_0 \rangle$ da origem.

Note que a demonstração do Teorema 2.4.2 não é verdadeira para o caso de n ser par, onde é necessário introduzir a transformada de Hilbert

Definição 2.4.1 (Transformada de Hilbert). *Seja $f \in \mathcal{S}(\mathbb{R}^n)$. Então*

$$\mathcal{H}f(s) = \mathcal{F}^{-1}(-i \cdot \text{sign}(\lambda) \hat{f}(\lambda)) \quad (2.20)$$

Teorema 2.4.3. *Seja $n > 1$ par. Então*

$$f(x) = 2^{-n}(i\pi)^{1-n} \int_{\mathbb{S}^{n-1}} \left(\frac{\partial^{n-1}}{\partial s^{n-1}} \mathcal{H}\mathcal{R}f \right) (w, \langle x, w \rangle) dS. \quad (2.21)$$

A seguir, deduziremos uma fórmula para \mathcal{R}^{-1} para o caso especial $n = 2$. Para tal, segue um resultado importante que relaciona o suporte da Transformada de Radon com o suporte da função f .

Teorema 2.4.4. *Seja $f \in C^\infty(\mathbb{R}^n)$ tal que $f(x) \rightarrow 0$ mais rapidamente que qualquer polinômio. Se $\mathcal{R}f(w, s) = 0$ para $|s| > \rho$, então $f(x) = 0$ para $\|x\| > \rho$.*

Demonstração : A figura 2.5 representa as hipóteses do Teorema. A demonstração pode ser encontrada em [46]. ■

Para nosso entendimento mais profundo, considere Ω como um círculo de raio ρ . Ainda supomos que f é axial-simétrica com respeito a origem, isto é, que existe uma função g tal que $f(x) = g(|x|)$. Assim, toda a informação que necessitamos saber está na direção $w_0 = (0, \pm 1)$.

Deste modo, podemos assumir que

$$f(w, s) = f(s), \quad 0 < s \leq \rho \quad \|w\| = 1.$$

Exercício 2.4.9. *Seja as hipóteses do Teorema 2.4.4 satisfeitas e $0 < s \leq \rho$. Use (2.16) para mostrar que $\mathcal{R}f(w_0, s)$ satisfaz uma equação integral de Abel de primeira espécie*

$$\mathcal{R}f(w_0, s) = s \int_s^\rho r f(r) / (\sqrt{r^2 - s^2}) dr. \quad (2.22)$$

Exercício 2.4.10. *Suponha que $\mathcal{R}f(w_0, \rho) = 0$. Prove que*

$$f(s) = -\pi^{-1} \int_r^\rho \frac{d/ds(\mathcal{R}f(w_0, s))}{\sqrt{s^2 - r^2}} ds. \quad (2.23)$$

O que podemos aprender de (2.19) e de (2.23) é que ambas as fórmulas envolvem derivadas de $\mathcal{R}f$. Isto é uma indicação de um problema mal-posto. Claro que, por exemplo, na equação (2.23), depois de derivarmos, fazemos uma operação suavizante novamente, a saber, integramos. Por outro lado, o kernel em (2.23) é singular e portanto, não anula totalmente a instabilidade introduzida pela diferenciação. Assim, métodos mais adequados que a invertibilidade direta da transformada de Radon devem ser considerados. Desenvolveremos alguns desses métodos nos próximos Capítulos.

2.4.3 Tomografia Computadorizada: caso discreto

Neste seção, faremos uma breve discussão do aspecto prático da Tomografia Computadorizada e da Transformada de Radon. Como comentado anteriormente, do ponto de vista prático, é impossível obter a integral de linha de f (veja (2.9)) em todas as direções. Na verdade, o conjunto de informações obtidas num processo de Tomografia são valores da transformada de Radon medidos por N detectores.

As limitações físicas do processo de medição implicam em uma discretização da imagem tomográfica. O tamanho e o número N de pixels, dentro do campo de visão, que devem ser reconstruídos consistem de um vetor de variáveis desconhecidas f_j , para $j = \{1, \dots, N\}$. Os f_j são os coeficientes de atenuação. A figura 2.6 mostra, esquematicamente, uma imagem tomográfica a ser reconstituída.

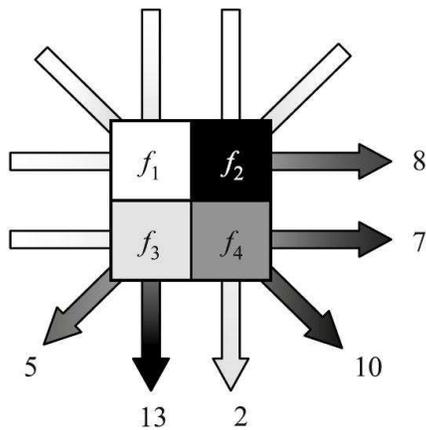


Figura 2.6: Tomografia discreta.

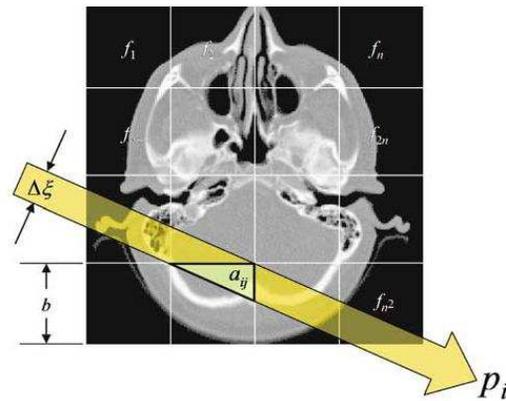


Figura 2.7: Feixe de raio-X.

Fisicamente, cada feixe de raio-X possui uma espessura. Quando o feixe de raio-X passa pela região Ω , temos que levar em conta quanto do pixel a ser reconstruído é afetado pelo feixe. Para este propósito, são introduzidos pesos que refletem a relação entre a área iluminada pelo feixe de raio-X com relação a área total do pixel. A figura 2.7 ilustra a situação.

Para um feixe de espessura $\Delta\xi$, o peso a_{ij} é determinado pela relação

$$a_{ij} = \frac{\text{área iluminada do pixel } j \text{ pelo raio } i}{\text{área total do pixel } j}. \quad (2.24)$$

Assim, para um conjunto de f_j , $j = \{1, \dots, N\}$, densidades a serem determinadas e dado um conjunto de $i = \{1, \dots, M\}$ raios-X medidos, com intensidade p_i , obtemos um sistema de M

equações lineares com N coeficientes

$$\sum_{j=1}^N a_{ij} f_j = p_i \quad i = \{1, \dots, M\}. \quad (2.25)$$

Escrevendo na forma matricial, temos que:

$$Af = p, \quad (2.26)$$

onde $A = (a_{ij})_{M \times N}$ pode ser considerada como a **caixa preta da máquina de tomografia**.

Fazendo uma comparação direta entre (2.9) e (2.26) obtemos:

$$\begin{array}{ccc} A & f & = & p \\ \uparrow & \uparrow & = & \uparrow \\ \mathcal{R} & f(w, s) & = & \int_{L(w,s)} f dl \end{array}$$

Algumas das dificuldades de reconstrução no modelo discreto são:

- O sistema (2.26) possui solução exata somente com condições ideais. Para dados reais, a presença de ruídos implica em obter apenas soluções aproximadas do sistema (2.26), mesmo quando $M = N$. No caso em que $M > N$, isto é, que temos mais informações (medidas) que o número de densidades a serem determinadas, possivelmente, obtemos reconstruções melhores da densidade.
- Tipicamente, a matriz A é singular e, em geral, não quadrada. Isto indica que o problema é mal-posto.
- A não possui uma estrutura simples. Assim, mesmo que A seja não singular, é difícil determinar uma maneira de resolver o sistema (2.26) de forma eficaz e com pouco custo computacional.
- Nos problemas práticos, a dimensão de A é muito grande, assim, métodos diretos de inversão são inapropriados, pois são computacionalmente muito intensos e custosos.

Exercício 2.4.11. Com base na figura 2.6 determine a matriz A associada. Para o vetor de intensidades p , determine uma solução f para o sistema (2.26), onde A é a matriz obtida analisando a figura 2.6.

Exercício 2.4.12. Com base no exercício acima, compare a solução $f = (f_1, f_2, f_3, f_4)$ e a figura 2.6. O que você tem a dizer?

Capítulo 3

Sistemas de Equações Lineares

Neste capítulo, faremos uso do background da Álgebra Linear e Teoria Linear de Operadores para introduzir algumas técnicas muito utilizadas em Problemas Inversos.

Os capítulos anteriores nos ensinaram que, dado um problema linear (um sistema de equações lineares, por exemplo), o operador que rege o problema nem sempre é injetivo e, mesmo que seja esse o caso, não é nada recomendável inverter esse operador.

Apresentaremos abaixo dois conceitos relacionados com a solução de problemas inversos. Estes são a *Inversa Generalizada ou Pseudo-Inversa* de operadores lineares e o *Teorema de Decomposição em Valores Singulares (SVD)*¹ ou *Teorema Espectral*. O primeiro destes conceito nos permitirá definir uma solução com uma propriedade especial, dentre todas as possíveis soluções (que podem ser infinitas), de um problema inverso. O segundo permite decompor um operador (uma matriz) como a soma de projeções sobre certos subespaços. Além disso, essa teoria nos ajuda a entender a influência dos autovalores de um operador na solução de problemas inversos.

3.1 Pseudo-Inversa de Operadores Lineares

Na verdade, tudo o que queremos em problemas inversos é: Encontrar uma maneira de aproximar um operador (o operador inverso) por uma família de operadores bem postos. Conseqüentemente, encontrar uma aproximação (a melhor possível) para a solução do problema.

Nesta seção, apresentaremos uma forma de aproximar “da melhor maneira” o inverso de um operador linear. Desenvolveremos a teoria para operadores lineares limitados que possuam imagem fechada, que é o caso de operadores em dimensão finita (matrizes) e também de operadores compactos. Assim, com a teoria linear, cobrimos uma ampla quantidade de casos interessantes. Faz jus mencionar que existem versões dos resultados apresentadas abaixo para operadores lineares limitados quaisquer. Para uma abordagem completa sobre o assunto, veja [12].

¹O Teorema de Decomposição em Valores Singulares é um dos teoremas mais fortes da matemática. Existem versões deste Teorema para operadores auto-adjuntos não limitados [26, 44]

3.1.1 Definições e Propriedades Básicas

Se o leitor não está habituado a certas definições pertinentes a teoria de operadores lineares limitados, sugerimos que faça uma breve leitura das definições e resultados contidos no Apêndice A.

Esperamos que o leitor esteja familiarizado com as definições de matrizes **Hermitianas**, **Simétricas**, **Unitárias**, **Normais**, etc. Caso contrário, uma breve leitura em livros de Álgebra Linear é recomendado. Veja, por exemplo, [36, 49]. Nosso objetivo é encurtarmos o caminho. Assim, vamos direto ao Teorema Espectral² (dimensão infinita) para operadores compactos e auto-adjuntos e obteremos, como corolário, o Teorema da SVD (dimensão finita).

Exercício 3.1.1. *Faça um paralelo entre as definições de matrizes Hermitianas, Simétricas, Unitárias, Normais e operadores adjuntos e auto-adjuntos encontrados no Apêndice A.*

Suponha que tenhamos um operador linear limitado $A : \mathcal{H}_1 \longrightarrow \mathcal{H}_2$, onde \mathcal{H}_1 e \mathcal{H}_2 denotam espaços de Hilbert. Consideraremos o problema fundamental de resolver a equação linear do tipo

$$Ax = y, \quad (3.1)$$

onde $y \in \mathcal{H}_2$.

Exemplo 3.1.1. *Exemplos da equação (3.1) são:*

Caso em que $\mathcal{H}_1 = \mathbb{R}^n$, $\mathcal{H}_2 = \mathbb{R}^m$ e $A \in M_{m \times n}(\mathbb{R})$.

Caso em que $\mathcal{H}_1 = \mathcal{H}_2 = L^2[0, 1]$ e A é um operador integral da forma

$$(Ax)(s) = \int_0^1 k(s, t)x(t)dt, \quad s \in [0, 1],$$

e $k(s, t) \in (C[0, 1] \times C[0, 1])$ é o chamado Kernel³.

Como já vimos, se o operador A possui uma inversa, então a equação (3.1) possui uma única solução $x = A^{-1}y$. Mas, nossa experiência anterior nos ensinou que "nem tudo são rosas", isto é, pode acontecer de que $\mathcal{N}(A) \neq \{0\}$ ou, talvez, $y \notin \text{Im}(A)$.

Um fato confortante é que, mesmo no caso da equação (3.1) não possuir uma solução no sentido tradicional, é possível definir uma solução generalizada do problema que é "a melhor" entre as soluções generalizadas de (3.1). Para tal, necessitamos de certas hipóteses sobre a imagem do operador A . No caso em que A é um operador compacto, a hipótese que faremos abaixo não é restritiva (veja Exercício ??)

²Teorema Espectral é o nome que se dá ao Teorema de Decomposição em Valores Singulares para operadores em dimensão infinita.

³Não confundir com o núcleo do operador A

Hipótese 3.1.1. *Suponha que A é um operador linear limitado e que $Im(A)$ é fechada em \mathcal{H}_2 .*

Seja $P : \mathcal{H}_2 \longrightarrow Im(A)$ o operador de projeção ortogonal (que está bem definido pelo Exercício 3.1.6). Assim, $Py \in Im(A)$ é o vetor mais próximo de y .

Exercício 3.1.2. *Suponha a Hipótese 3.1.1 satisfeita e P o operador de projeção ortogonal sobre $\mathcal{R}(A)$. Prove que, dado $y \in \mathcal{H}_2$, $Py - y \in (Im(A))^\perp$.*

Definição 3.1.1. *Uma solução generalizada de (3.1) é qualquer solução $u \in \mathcal{H}_1$ da equação*

$$Ax = Py. \quad (3.2)$$

Exemplo 3.1.2. *Suponha que $A = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$ e $y = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$.*

Então $Im(A) = span\{(1, -1)\}$ e $Py = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$. Portanto, o conjunto de soluções generalizadas é dada por

$$\left\{ (x_1, x_2) : x_2 = \frac{1}{2} + x_1 \right\}.$$

Geometricamente, uma solução generalizada, como na Definição 3.1.1, significa encontrar $u \in \mathcal{H}_1$ solução do problema de minimização

$$u = arg \min_{x \in \mathcal{H}_1} \frac{1}{2} \|Ax - y\|^2. \quad (3.3)$$

Mais geral:

Teorema 3.1.1. *Suponha que $A : \mathcal{H}_1 \longrightarrow \mathcal{H}_2$ seja um operador linear limitado, $y \in \mathcal{H}_2$ e a Hipótese 3.1.1 seja satisfeita. Então, as seguintes condições sobre $u \in \mathcal{H}_1$ são equivalentes:*

(i) $Au = Py$,

(ii) $u = arg \min_{x \in \mathcal{H}_1} \frac{1}{2} \|Ax - y\|^2$,

(iii) $A^*Au = A^*y$ (conhecidas como Equações Normais).

Demonstração : (i) \Rightarrow (ii): Seja $Au = Py$. Segue do exercício 3.1.2 e do Teorema de Pitágoras que, dado $x \in \mathcal{H}_1$,

$$\begin{aligned} \|Ax - y\|^2 &= \|Ax - Py\|^2 + \|Py - y\|^2 \\ &= \|Ax - Py\|^2 + \|Au - y\|^2 \geq \|Au - y\|^2. \end{aligned}$$

(ii) \Rightarrow (iii): Por hipótese, existe pelo menos um $x \in \mathcal{H}_1$ solução de (3.2). Disto e do Teorema de Pitágoras, temos que

$$\begin{aligned}\|Au - y\|^2 &= \|Au - Py\|^2 + \|Py - y\|^2 \\ &\geq \|Au - Py\|^2 + \|Au - y\|^2.\end{aligned}$$

Portanto

$$Au - y = Py - y \in (Im(A))^\perp = \mathcal{N}(A^*).$$

Assim, $A^*(Au - y) = 0$ e (iii) segue.

(iii) \Rightarrow (i): De (iii) satisfeito, obtemos que $Au - y \in (Im(A))^\perp$ e, assim,

$$0 = P(Au - y) = Au - Py. \quad \blacksquare$$

Definição 3.1.2. Um vetor $u \in \mathcal{H}_1$ satisfazendo qualquer uma das sentenças do Teorema 3.1.1 é chamado uma **solução de quadrados mínimos** da equação $Ax = y$.

Exercício 3.1.3. Mostre que o conjunto de soluções de quadrados mínimos pode ser escrito como

$$\{u \in \mathcal{H}_1 : A^*Au = A^*b\}. \quad (3.4)$$

Também, prove que este conjunto é convexo e fechado.

Uma observação importante a ser feita é a de que, sob a Hipótese 3.1.1, uma solução de quadrados mínimos de (2.1) sempre existe $\forall b \in \mathcal{H}_2$ (veja Exercício 3.1.4). Caso $\mathcal{N}(A) \neq 0$ então, existe uma infinidade de soluções de quadrados mínimos de (2.1). De fato, se u é uma solução de quadrados mínimos e $v \in \mathcal{N}(A)$, então $u + v$ também é uma solução de quadrados mínimos.

Exercício 3.1.4. Assuma que A satisfaz a Hipótese 3.1.1. Prove que existe pelo menos uma solução de quadrados mínimos. Dê condições sobre o operador A para que a solução de quadrados mínimos seja única.

Exercício 3.1.5. Prove que ambos os exemplos apresentados acima para a equação (3.1) são operadores compactos. Sugestão: Para o caso de A ser o operador integral, comece supondo que $k(s, t) \in (C[0, 1] \times C[0, 1])$ e use o Teorema de Ascoli-Arzelá. Use a densidade de $C[0, 1]$ em $L^2[0, 1]$.

Exercício 3.1.6. Seja \mathcal{H} um espaço de Hilbert e $C \subset \mathcal{H}$ um conjunto convexo e fechado. Prove que para todo $b \in \mathcal{H}$, existe uma única projeção de b sobre C . Prove ainda que a projeção de b sobre C é o vetor de C mais próximo de b .

Estamos buscando um caminho para "inverter" o operador A , associando com cada $b \in \mathcal{H}_2$ uma única solução de quadrados mínimos. Sabemos que, se $\mathcal{N}(A) \neq 0$, não temos tal unicidade. Será que temos alguma alternativa? A resposta é afirmativa: basta para tal, escolhermos entre as (várias) possíveis soluções, uma que tenha uma característica especial. Mas, que característica especial podemos escolher num conjunto que possui, possivelmente, uma infinidade de elementos?

Vamos voltar e analisar o que temos de hipóteses e resultados.

- i) Sabemos que o conjunto de soluções de quadrados mínimos é não-vazio.
- ii) Pelo Exercício 3.1.3, o conjunto de soluções de quadrados mínimos é **convexo e fechado**.

Portanto, pelo *Teorema da Projeção* A.1.1, existe uma única solução de quadrados mínimos com norma mínima associada a cada elemento $b \in \mathcal{H}_2$. Logo, temos um caminho para encontrar uma inversa (mesmo que generalizada) para o operador A .

Definição 3.1.3. *Seja A um operador satisfazendo a Hipótese 3.1.1. A aplicação*

$$A^\dagger : \mathcal{H}_2 \longrightarrow \mathcal{H}_1$$

definida por

$$A^\dagger b = u,$$

*onde u é a **única** solução de quadrados mínimos de norma mínima da equação (3.1). Tal solução é chamada de Inversa Generalizada⁴ de A .*

Exercício 3.1.7. *Mostre que, se A possui uma inversa, então $A^\dagger = A^{-1}$.*

Existem outras definições de inversa generalizada de operadores que são equivalentes a dada acima (veja [12]).

Exercício 3.1.8. *Mostre que, se A satisfaz a Hipótese 3.1.1, então $\text{Im}(A^*)$ é fechada e $\mathcal{N}(A)^\perp = \text{Im}(A^*)$.*

Teorema 3.1.2. *Se A satisfaz a Hipótese 3.1.1, então $\text{Im}(A^\dagger) = \text{Im}(A^*) = \text{Im}(A^\dagger A)$.*

Demonstração : Seja $b \in \mathcal{H}_2$. Num primeiro momento, mostraremos que $A^\dagger b \in \mathcal{N}(A)^\perp$ e então usaremos o Exercício 3.1.8. Suponha que

$$A^\dagger b = u_1 + u_2 \in \mathcal{N}(A)^\perp \oplus \mathcal{N}(A).$$

Então, u_1 é uma solução de quadrados mínimos de $Ax = b$. De fato,

$$Au_1 = A(u_1 + u_2) = AA^\dagger b = Pb.$$

⁴Para um apanhado histórico muito interessante sobre inversa generalizada, consulte [45]

e, portanto, a afirmação está completa usando o Teorema 3.1.1.

Suponha que $u_2 \neq 0$. Então, pelo Teorema de Pitágoras,

$$\|u_1\|^2 < \|u_1 + u_2\|^2 = \|A^\dagger b\|^2$$

contradizendo o fato de $A^\dagger b$ ser uma solução de quadrados mínimos que tem a norma mínima. Logo, $A^\dagger b = u_1 \in \mathcal{N}(A)^\perp$.

Reciprocamente, sejam $u \in \mathcal{N}(A)^\perp$ e $b = Au$. Então,

$$Au = PAu = Pb$$

e, assim, u é uma solução de quadrados mínimos. Se x é qualquer outra solução de quadrados mínimos, então

$$Ax = Pb = Au$$

e, portanto, $x - u \in \mathcal{N}(A)$. Pelo Teorema de Pitágoras,

$$\|x\|^2 = \|u\|^2 + \|x - u\|^2 \geq \|u\|^2.$$

Assim, u é a solução de quadrados mínimos que tem norma mínima, i.e., $u = A^\dagger b$.

Isto prova a primeira das igualdades. Para verificar a segunda, note que, para qualquer $b \in \mathcal{H}_2$,

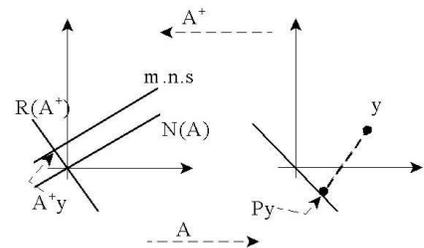
$$A^\dagger b = A^\dagger Pb \in \text{Im}(A^\dagger A). \quad \blacksquare$$

Corolário 3.1.1. *Se A satisfaz a Hipótese 3.1.1, então $A^\dagger : \mathcal{H}_2 \longrightarrow \mathcal{H}_1$ é linear e limitado.*

Exercício 3.1.9. *Demonstre o Corolário 3.1.1.*

Interpretação Geométrica da Inversa Generalizada

Do ponto de vista computacional é importante ter condições mais simples para representar o operador A^\dagger . Esse é a situação se ambos \mathcal{H}_1 e \mathcal{H}_2 têm dimensão finita. Neste caso, sabemos que A possui uma representação matricial e encontrar A^\dagger reduz-se a calcular a inversa generalizada de uma matriz. De qualquer forma, temos:



Teorema 3.1.3. *Suponha que A satisfaça a Hipótese 3.1.1.*

Então,

$$A^\dagger = (A^*A)^\dagger A^* = A^*(AA^*)^\dagger.$$

Demonstração : Faremos a demonstração da primeira das igualdades, a segunda é similar e, assim, um bom exercício.

Seja $y \in \mathcal{H}_2$. Pelo Teorema da Projeção $y = y_1 + y_2 \in \text{Im}(A) \oplus \mathcal{N}(A^*)$ e assim, $A^*y = A^*y_1 \in \text{Im}(A^*A)$. Portanto, se $b \in \mathcal{H}_2$, então

$$A^*A(A^*A)^\dagger A^*b = P_{\text{Im}(A^*A)}A^*b = A^*b.$$

Logo, $(A^*A)^\dagger A^*b$ é uma solução de quadrados mínimos (satisfaz as equações normais). Consequentemente,

$$(A^*A)^\dagger A^*b = A^\dagger b + v,$$

para algum $v \in \mathcal{N}(A)$. Como $(A^*A)^\dagger A^*b \in \text{Im}((A^*A)^\dagger)$, segue do Teorema 3.1.2 que $(A^*A)^\dagger A^*b \in \text{Im}((A^*A)) = \text{Im}(A^*) = \mathcal{N}(A)^\perp$. Portanto, $(A^*A)^\dagger A^*b = A^\dagger b$. ■

Um fato muito importante para o que segue na teoria de "regularização para problemas inversos" é consequência do seguinte resultado:

Teorema 3.1.4. *A inversa generalizada A^\dagger possui gráfico $\text{Gr}(A^\dagger)$ fechado. Consequentemente, A^\dagger é contínua se, e só se, $\text{Im}(A)$ é fechada.*

Demonstração : Uma parte deste Teorema é consequência do Corolário 3.1.1. A demonstração completa do Teorema foge ao escopo destas notas, pois, usa conceitos fortes de Análise Funcional em particular o Teorema do Gráfico Fechado [26, 44]. Interessados na demonstração podem consultar [9]. ■

Observação: O Teorema 3.1.4 reforça ainda mais a diferença entre problemas inversos em dimensão finita e infinita. Pois, no caso de dimensão finita, o operador (matriz) A sempre possui a imagem fechada. Assim, temos a garantia de existência e unicidade de uma solução de mínimos quadrados de norma mínima.

Exercício 3.1.10. *Prove que se A é um operador linear entre espaços de dimensão finita, então $\text{Im}(A)$ é fechada.*

Exercício 3.1.11. *Suponha que A satisfaça a Hipótese 3.1.1. Prove que $A^\dagger : \mathcal{H}_2 \longrightarrow \mathcal{H}_1$ é o único operador linear limitado satisfazendo*

$$AA^\dagger = P_{\text{Im}(A)} \quad e \quad A^\dagger A = P_{\text{Im}(A^\dagger)}.$$

Esta é a definição de Moore para Inversa Generalizada. (Sugestão: Consulte [12])

3.2 A Decomposição em Valores Singulares

Um dos principais resultados da Álgebra Linear é o *Teorema de Decomposição em Valores Singulares (SVD)*. Este teorema permite escrever uma matriz qualquer como uma soma de matrizes de projeção de posto 1. Mais geral, o Teorema de SVD vale para operadores lineares em espaços de Hilbert de dimensão infinita que são auto-adjuntos⁵. Neste caso, o Teorema de SVD é conhecido como *Teorema Espectral*, [26]. No caso especial em que A é um operador linear e compacto, o Teorema Espectral se traduz de forma similar ao caso de dimensão finita.

Entraremos em mais detalhes a partir de agora. Esperamos que o leitor esteja familiarizado com o conceito de autovalores e autovetores da Álgebra Linear e com o conceito de espectro e resolvente⁶ para operadores lineares. Seguem algumas referências importantes para os que querem se aprofundar no assunto [26, 49, 36, 11].

Observação: Faremos a hipótese de que o corpo de escalares do espaço vetorial é o corpo dos números complexos. Assim, se temos uma matriz $n \times n$, esta possui n autovalores. Esse fato é importante no que segue.

Exercício 3.2.1. *Prove que uma matriz quadrada A possui, no máximo n , autovalores. Dê um exemplo de uma matriz que não possui autovalores.*

Exercício 3.2.2. *Prove que, se estamos considerando o espaço vetorial sobre o corpo dos números complexos, então uma matriz quadrada $A_{(n \times n)}$ possui n autovalores.*

Exercício 3.2.3. *Dê um exemplo, em dimensão infinita, de um operador linear que não possui autovalores.*

Nosso ponto de partida é uma versão simplificada do Teorema SVD, a qual faremos a demonstração. Formulações mais gerais podem ser encontradas em [26, 49, 36, 11].

Teorema 3.2.1. [Diagonalização] *Seja A uma matriz quadrada de ordem $n \times n$ com um conjunto de n autovetores L.I. Então, A é similar a uma matriz diagonal ou diagonalizável.*

Demonstração : Construa a matriz S tendo como colunas os vetores $\{v_1, \dots, v_n\}$. Assim:

$$AS = A \begin{pmatrix} | & | & \cdots & | \\ v_1 & v_2 & \cdots & v_n \\ | & | & \cdots & | \end{pmatrix} = \begin{pmatrix} | & | & \cdots & | \\ \lambda_1 v_1 & \lambda_2 v_2 & \cdots & \lambda_n v_n \\ | & | & \cdots & | \end{pmatrix} = S \cdot \text{Diag}(\lambda_1, \lambda_2, \dots, \lambda_n).$$

⁵Um operador A entre espaços de Hilbert é Auto-Adjunto se $\mathcal{D}(A^*) = \mathcal{D}(A)$ e $\langle x, Ay \rangle = \langle Ax, y \rangle, \forall x, y \in \mathcal{D}(A)$

⁶Operadores lineares em dimensão infinita podem possuir elementos no espectro que não são autovalores [26]

Como S é invertível, temos

$$A = S \text{Diag}(\lambda_1, \lambda_2, \dots, \lambda_n) S^{-1}. \quad \blacksquare$$

Exercício 3.2.4. *Mostre que, dado um conjunto L.I. de vetores, sempre existe um conjunto ortogonal. Mostre ainda que o espaço gerado pelos dois conjuntos são iguais. Sugestão: Use o Processo de Gram-Schmidt.*

Exercício 3.2.5. *Justifique, de maneira adequada, que a matriz S no Teorema acima é de fato inversível.*

Corolário 3.2.1. *Seja A uma matriz com de ordem $n \times n$ que possui n autovalores distintos. Então, A é diagonalizável.*

Exercício 3.2.6. *Prove o Corolário 3.2.1.*

Pergunta: Será que toda matriz quadrada é diagonalizável? Não, pelo menos, no sentido do Teorema 3.2.1. O contra-exemplo é a matriz $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$.

Exercício 3.2.7. *Mostre que a matriz A acima não é diagonalizável no sentido do Teorema 3.2.1.*

O Teorema 3.2.1 é a versão mais simples do Teorema de Decomposição em valores singulares. Passaremos agora para uma versão mais geral.

Teorema 3.2.2. *Todo operador compacto possui no máximo uma quantidade enumerável de autovalores que formam uma sequência cujos valores absolutos convergem para zero. Os autovalores de um operador auto-adjunto são reais.*

Teorema 3.2.3. [Teorema Espectral⁷ - A compacto e auto-adjunto] *Seja $A : \mathcal{H} \rightarrow \mathcal{H}$ um operador compacto e auto-adjunto. Então, existe um sistema ortonormal completo $\{e_j\}$ de \mathcal{H} tal que $Ae_j = \lambda_j e_j$ e $\lambda_j \rightarrow 0$ quando $n \rightarrow \infty$.*

Demonstração : Esse é um dos problemas mais importante da Análise. Não faremos a demonstração, pois foge das nossas pretensões. Para a demonstração sugerimos que o leitor consulte [26].

■

Teorema 3.2.4. [Teorema Espectral - A compacto] *Seja $A : \mathcal{H}_1 \rightarrow \mathcal{H}_2$ um operador linear compacto. Então, existem conjuntos ortogonais (não necessariamente completos) $\{e_1, \dots, e_m\}$ de \mathcal{H}_1 e $\{f_1, \dots, f_m\}$ de \mathcal{H}_2 e de números $\sigma_1, \dots, \sigma_m$ com $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_m$, tal que*

$$Ax = \sum_{j=1}^m \sigma_j \langle x, e_j \rangle f_j, \quad x \in \mathcal{H}_1. \quad (3.5)$$

⁷O Teorema Espectral como enunciado também é conhecido como Teorema de Hilbert-Schmidt. Existem várias versões deste Teorema (Veja [26] e referências)

No caso da imagem do operador A ter dimensão infinita, temos que considerar $m \rightarrow \infty$. Neste caso, $\sigma_m \rightarrow 0$.

Demonstração : Como A é compacto sabemos que A^*A é compacto e auto-adjunto. Pelo Teorema Espectral 3.2.3, existe um conjunto ortogonal $\{e_1, \dots, e_m\}$ de \mathcal{H}_1 tal que $A^*Ae_j = \lambda_j e_j$, onde $0 \leq \lambda_j \in \mathbb{R}$. Defina $\sigma_j = \sqrt{\lambda_j}$ e $f_j = \frac{1}{\sigma_j} Ae_j$ (para $\sigma_j > 0$). Um cálculo simples mostra que $\{f_1, \dots, f_m\}$ é um conjunto ortonormal e que a equação (3.9) é satisfeita. ■

Definição 3.2.1. Os valores $\sigma_1, \dots, \sigma_m$ são chamados de **valores espectrais** de A . Chamamos de **sistema singular** de A a tripla (σ_j, e_j, f_j) .

Exercício 3.2.8. Mostre que, se λ_j satisfaz $A^*Ae_j = \lambda_j e_j$, então $\lambda_j \geq 0$ e $\lambda_j \in \mathbb{R}$.

Exercício 3.2.9. Mostre que o conjunto $\{f_1, \dots, f_m\}$ definido como no Teorema 3.2.4 é ortonormal.

Exercício 3.2.10. Mostre que se A^* é um operador linear compacto, então

$$A^*y = \sum_{j=1}^m \sigma_j \langle y, f_j \rangle e_j, \quad x \in \mathcal{H}_1. \quad (3.6)$$

Exercício 3.2.11. Mostre que se A^* é um operador linear compacto, então

$$A^\dagger y = \sum_{j=1}^m \sigma_j^{-1} \langle y, f_j \rangle f_j, \quad y \in \mathcal{D}(A^\dagger). \quad (3.7)$$

Corolário 3.2.2. [Teorema espectral em dimensão finita - SVD.] Seja $A \in \mathbb{R}^{m,n}$ com $n \leq m$. Então, existem matrizes unitárias $U \in M_{m \times m}$, $V \in M_{n \times n}$ e uma matriz diagonal com entradas não-negativas $\Sigma := \text{diag}\{\sigma_1, \dots, \sigma_n\}$ tais que

$$A = U\Sigma V^T.$$

Os passos para demonstrar o teorema SVD estão nos exercícios abaixo.

Exercício 3.2.12. Mostre que todo operador linear cuja imagem possui dimensão finita é compacto. Consequentemente, toda matriz é um operador linear compacto.

Exercício 3.2.13. Mostre que se A é uma matriz, então AA^* e A^*A são operadores compactos e auto-adjuntos.

Exercício 3.2.14. Demonstre o Corolário 3.2.2.

3.2.1 Funções de Operadores: O Teorema da Aplicação Espectral

Daremos agora uma breve introdução ao *Cálculo Funcional*, como é conhecida a teoria que trata de funções de Operadores Lineares. Essa importantíssima ferramenta matemática nos ensinará a derivar as chamadas de *Funções Filtro* que são a peça chave para o entendimento dos Métodos de Regularização dos capítulos a seguir.

Por simplicidade, daremos somente a idéia intuitiva em dimensão finita. Para interessados em detalhes mais aprofundados, recomendamos [26] e livros de Análise Funcional em geral.

Exercício 3.2.15. *Seja A uma matriz como no Teorema 3.2.1. Mostre que*

$$A^2 = S \operatorname{diag}(\lambda_1^2, \dots, \lambda_n^2) S^{-1}.$$

Use indução para mostrar que $A^p = S \operatorname{diag}(\lambda_1^p, \dots, \lambda_n^p) S^{-1}$, para qualquer $p \in \mathbb{N}$.

Sejam $t \in [0, T]$ e $g(t) = a_0 + a_1 t + \dots + a_n t^n$ um polinômio de ordem n em t .

Definição 3.2.2. *Seja A um operador linear limitado ($\|A\| \in [0, T]$), definimos um polinômio do operador A por*

$$g(A) = a_0 + a_1 A + \dots + a_n A^n$$

onde $g(t)$ é o polinômio acima.

Exercício 3.2.16. *Mostre que o polinômio $g(A)$ está bem definido, como um operador linear, para A linear limitado.*

Exercício 3.2.17. *Mostre que se A satisfaz as hipóteses do Teorema 3.2.1, então*

$$g(A) = S g(\operatorname{diag}(\lambda_1, \dots, \lambda_n)) S^{-1}.$$

Exercício 3.2.18. *O que você pode dizer no de uma função contínua f aplicada em A ?*

O que acontece com uma função contínua de um operador linear limitado? A resposta é dada pelo Teorema da Aplicação Espectral. Elucidaremos suas consequências através de um exemplo. Para interessados na demonstração consulte [26].

Exemplo 3.2.1. *Uma função de operadores muito especial é a exponencial de um operador linear limitado $\exp(A)$. Dar sentido a esse tipo de operações tem uma importância enorme na caracterização de soluções para sistemas de EDO's e na Teoria de Semigrupos associados a operadores diferenciais parciais.*

Vamos considerar o caso especial em que A é uma matriz e, mais ainda, esta satisfaz as hipóteses do Teorema 3.2.1.

Sabemos que a função $\exp(t)$ possui uma expansão em séries de potências dada por

$$\exp(t) = \sum_{j=0}^{\infty} \frac{t^j}{j!} = 1 + \frac{t}{1!} + \frac{t^2}{2!} + \cdots + \frac{t^k}{k!} + \cdots .$$

que converge uniformemente $\forall t \in \mathbb{R}$.

Pelo o Exercício 3.2.15, temos que

$$\begin{aligned} \sum_{j=0}^{\infty} \frac{A^j}{j!} &= I + \frac{A}{1!} + \frac{A^2}{2!} + \cdots + \frac{A^n}{n!} + \cdots = I + S \cdot \left(\sum_{k=1}^{\infty} \frac{\text{diag}(\lambda_1^k, \dots, \lambda_n^k)}{k!} \right) S^{-1} \\ &= S \cdot \text{diag}(\exp(\lambda_1), \dots, \exp(\lambda_n), \dots) \cdot S^{-1} =: \exp(A). \end{aligned}$$

Como o operador (a matriz) A é limitado, a série $\sum_{j=0}^{\infty} \frac{A^j}{j!}$ converge uniformemente na norma dos operadores e, assim, $\exp(A)$ está bem definida.

No caso especial em que A é uma matriz quadrada e injetiva, temos, do Teorema SVD, que

$$A^{-1} = V \Sigma^{-1} U^T, \quad (3.8)$$

onde $\Sigma^{-1} = \text{diag}\{\frac{1}{\sigma_1}, \dots, \frac{1}{\sigma_n}\}$.

Observação: Note que a inversão de uma matriz pode ser pensada como a função $f(t) = t^{-1}$ aplicada a matriz.

De fato, prova-se que este resultado não é mera coincidência. O Teorema que garante tal situação é

Teorema 3.2.5. [Teorema da Aplicação Espectral.] *Seja A um operador linear limitado e $f : \mathbb{R} \rightarrow \mathbb{R}$ uma função contínua. Denotemos por $\Sigma(A)$ o espectro de A . Então,*

$$\Sigma(f(A)) = f(\Sigma(A)).$$

Em particular, se A é compacto, então

$$f(A)x = \sum_{j=0}^m f(\sigma_j) \langle x, e_j \rangle f_j, \quad x \in \mathcal{H}.$$

O Teorema também vale se f for contínua à direita ou à esquerda.

Demonstração: Veja [26]. ■

Esse resultado é extremamente importante no entendimento, para a construção das estratégias de regularização (veja Capítulo 4) e para entender a relação existente entre mal-condicionamento de um problema com os respectivos valores espectrais do operador associado.

3.2.2 Relação entre Má-colocação e Valores Espectrais

Segue do Teorema Espectral que, se o operador linear A é compacto e possui inversa, então

$$A^{-1}y = \sum_{j=1}^m \frac{1}{\sigma_j} \langle y, f_j \rangle e_j, \quad y \in \text{Im}(A). \quad (3.9)$$

Assim, se a dimensão do problema é grande, ou melhor, se σ_m está próximo de zero, significa que $\frac{1}{\sigma_m}$ é grande. Portanto, pequenas perturbações na direção de um autovetor associado ao autovalor σ_m implicam numa grande variação na solução.

Exercício 3.2.19. *Prove que, se A é um operador linear compacto e possui inversa limitada, então a dimensão do espaço vetorial é finita.*

Explicaremos esse fenômeno de maneira mais clara através da Decomposição em Valores Singulares em dimensão finita.

Considere o problema inverso (em dimensão n) de recuperar x na equação matricial

$$Ax = y^\delta,$$

para um dado com ruídos $y^\delta = (y_1, \dots, y_n + \frac{1}{n})$. Assim, o erro nos dados é da ordem de $\frac{1}{n}$.

Suponha que a matriz A tenha sua decomposição espectral dada por (3.8) onde os valores singulares são $\sigma_j = \mathcal{O}(\frac{1}{j})^8$, $j = 1, 2, \dots, n$. Logo, A é inversível.

Da equação (3.8), temos que a solução com ruídos é dada por

$$x^\delta = A^{-1}y^\delta = V\Sigma^{-1}U^T y^\delta.$$

Como as matrizes U e V são unitárias e que todas as normas são equivalentes (dimensão finita), temos a seguinte estimativa para a solução

$$\|x - x^\delta\|^2 = \|V\Sigma^{-1}U^T y - V\Sigma^{-1}U^T y^\delta\|^2 = \sum_{j=1}^n (\mathcal{O}(j)(y_j - y_j^\delta))^2 = \left(\frac{\mathcal{O}(n)}{n}\right)^2 = (\mathcal{O}(1))^2. \quad (3.10)$$

Note que o erro na solução é muito grande, se $(\mathcal{O}(1))^2$ for grande.

Exercício 3.2.20. *Seja A um operador compacto. Considere o problema de determinar x na equação $Ax = y^\delta$. Você usaria a estratégia de minimizar o resíduo $\|Ax - y^\delta\|^2$ para solucionar o problema? Justifique adequadamente sua resposta. Sugestão: Olhe para as equações normais.*

Vamos agora a um critério de solvabilidade de equações lineares governadas por operadores compactos. Muitos autores referem-se a este resultado como sendo o **Critério de Picard** [1, 2, 9].

⁸ $\mathcal{O}(a)$ significa que $\frac{\mathcal{O}(a)}{a} = \text{constante}$.

Teorema 3.2.6. *Seja A um operador compacto e (σ_j, e_j, f_j) um sistema singular de A . Dado $y \in \mathcal{H}_2$, as seguintes condições são equivalentes:*

- a) $y \in \text{Im}(A)$,
 b) $y \in \overline{\text{Im}(A)}$; $\sum_{j=0}^{\infty} \sigma_j^{-2} |\langle y, f_j \rangle|^2 < \infty$.

Demonstração : Daremos uma ideia da prova.

a) \Rightarrow b) De $y \in \text{Im}(A) \subset \overline{\text{Im}(A)}$. Seja $x \in \mathcal{H}_1$ tal que $Ax = y$. Segue do Teorema 3.2.4 que $A^*f_j = \sigma_j e_j$. Assim, usando a desigualdade de Bessel,

$$\sum_{j=0}^{\infty} \sigma_j^{-2} |\langle y, f_j \rangle|^2 = \sum_{j=0}^{\infty} \sigma_j^{-2} |\langle Ax, f_j \rangle|^2 = \sum_{j=0}^{\infty} \sigma_j^{-2} |\langle x, A^*f_j \rangle|^2 = \sum_{j=0}^{\infty} |\langle x, e_j \rangle|^2 \leq \sum_{j=0}^{\infty} \|x\|^2 < \infty.$$

b) \Rightarrow a) Defina $x_n := \sum_{j=0}^n \sigma_j^{-1} \langle y, f_j \rangle e_j$. Portanto, para $m, n \in \mathbb{N}$ temos:

$$\|x_n - x_m\|^2 = \sum_{j=n+1}^m \sigma_j^{-2} |\langle y, f_j \rangle|^2$$

e, portanto, $\{x_n\}$ é uma sequência de Cauchy, cujo limite denotaremos por $x := \lim_{n \rightarrow \infty} x_n$. Pela continuidade de A e definição de x_n , segue que

$$Ax = \sum_{j=0}^{\infty} \langle y, f_j \rangle f_j \quad \text{e} \quad \|Ax\| \leq \|y\|.$$

Defina $z := y - \sum_{j=0}^{\infty} \langle y, f_j \rangle f_j$. Segue facilmente que

$$\|z\|^2 = \|y\|^2 - \sum_{j=0}^{\infty} |\langle y, f_j \rangle|^2; \quad \langle z, f_j \rangle = 0, \forall j \in \mathbb{N} \quad \text{e} \quad A^*z = 0. \quad (3.11)$$

Portanto, como $y \in \overline{\text{Im}(A)} = \mathcal{N}(A^*)^\perp$, temos

$$\langle z, y \rangle = \|y\|^2 - \sum_{j=0}^{\infty} |\langle y, f_j \rangle|^2 = \|z\|^2.$$

Logo, $y = \sum_{j=0}^{\infty} \langle y, f_j \rangle f_j = Ax$. ■

Exercício 3.2.21. *Preencha os detalhes da demonstração do Teorema 3.2.6.*

Observação: Note que o Teorema de Picard 3.2.6 sugere que uma tentativa de solução para a equação $Ax = y$ é

$$x = \sum_{j=0}^{\infty} \sigma_j^{-1} \langle y, f_j \rangle e_j. \quad (3.12)$$

Suponha que temos somente o dado perturbado $y^\delta = y + \delta f_k$. Substituindo na equação (3.12), obtemos como solução perturbada $x^\delta = x + \delta \sigma_k^{-1} e_k$. Assim,

$$\|x - x^\delta\|^2 = \left\| \frac{\delta}{\sigma_k} e_k \right\|^2 = \frac{\delta^2}{\sigma_k^2}. \quad (3.13)$$

Como $\sigma_j \rightarrow 0$, segue que a estimativa (3.13) pode ser muito ruim, mostrando o efeito de mal-condicionamento causado pelos valores singulares de um operador compacto.

Continuamos com o filosófico conceito de precisão. Quanto mais **precisos** procuramos ser (aproximando melhor o operador A^{-1}) mais longe ficamos de uma solução **precisa**, uma vez que, os dados não são medidos de maneira **precisa**.

Capítulo 4

Regularização para Problemas Inversos

Passaremos a desenvolver as chamadas *Estratégias ou Métodos de Regularização* para Problemas Inversos. Sempre que procuramos solucionar um Problema Inverso, temos que contornar impecilhos como instabilidade e mal-condicionamento (má-colocação). Exemplos desta situação foram apresetados na Seção 2.3.

A maneira natural de solucionar um problema do tipo (2.1) (para dados exatos ou para dados perturbados por ruídos) é inverter o operador A . Mas, como foi exemplificado e como acontece na prática, nos deparamos muitas vezes (quase sempre) com operadores que não possuem inversa. Ou, se a inversa existe, esta é mal-condicionada (no caso de problemas em dimensão finita) ou ilimitada (no caso de dimensão infinita). Assim, o axioma de dependência contínua dos dados falha, produzindo soluções inadequadas para o problema.

Como vimos no Capítulo 3, mesmo que o operador A não possua inversa, podemos nos ater em aproximar a exata usando a solução de quadrados mínimos que possua norma mínima, ou seja, aproximar a solução $x^\dagger = A^\dagger y$ do problema (2.1). Relembrando, os dados y para o problema, em geral, não são conhecidos exatamente. Na prática, que temos em mão são dados aproximados y^δ , obtidos por medições, com o nível de ruídos δ , satisfazendo

$$\|y - y^\delta\| \leq \delta. \quad (4.1)$$

No caso em que A não possui, necessariamente, uma imagem fechada, o Teorema 3.1.4 nos ensina que $A^\dagger y^\delta$ não é uma boa aproximação para $A^\dagger y$, pois A^\dagger não é limitado.

Portanto, temos que elaborar outras estratégias para solucionar problemas inversos. É isso que faremos nesse capítulo.

4.1 O Conceito de Regularização

Denominamos por *estratégia de regularização* o artifício matemático¹ de obtermos uma solução aproximada (digamos x_α^δ) de maneira **estável** e que **convirja** (em topologias adequadas), quando o nível de ruídos converge para zero, para a solução x^\dagger do problema inverso considerado. Além disso, o parâmetro α deve ser escolhido de maneira apropriada (seja lá o que isso signifique). Em termos gerais, uma *estratégia ou método de regularização* consiste em aproximar uma solução x^\dagger de um problema mal posto (2.1) por uma família (a um parâmetro α) de problemas bem postos. Mais precisamente:

Definição 4.1.1. *Sejam $A : \mathcal{H}_1 \longrightarrow \mathcal{H}_2$ um operador linear e limitado e $\alpha_0 \in (0, +\infty)$. Para todo $\alpha \in (0, \alpha_0)$, seja*

$$R_\alpha : \mathcal{H}_2 \longrightarrow \mathcal{H}_1$$

*um operador contínuo (não necessariamente linear). A família $\{R_\alpha\}$ é chamada de uma **regularização** ou **uma família de operadores de regularização** (para A^\dagger) se, para todo $y \in \mathcal{D}(A^\dagger)$, existir uma regra para escolha do parâmetro $\alpha := \alpha(\delta, y^\delta)$ tal que*

$$\limsup_{\delta \rightarrow 0} \{\|R_\alpha y^\delta - A^\dagger y\| : y^\delta \in \mathcal{H}_2, \|y - y^\delta\| \leq \delta\} = 0 \quad (4.2)$$

é satisfeita para $\alpha := \alpha(\delta, y^\delta) \xrightarrow{\delta \rightarrow 0} 0$.

Observação: Note que não estamos requerendo que a família de operadores de regularização $\{R_\alpha\}$ seja de operadores lineares. No caso em que $\{R_\alpha\}$ é linear, então dizemos que o método de regularização é linear.

Definição 4.1.2. *Uma estratégia de regularização (R_α, α) é dita **convergente** se $x_\alpha := R_\alpha y$ converge para x^\dagger .*

A arte de aplicar métodos de regularização está sempre relacionada com o compromisso entre precisão e estabilidade. Ou seja, procuramos por aproximações x_α^δ de x^\dagger , que dependam continuamente dos dados com ruídos y^δ (estabilidade) e que convirjam para x^\dagger , se o nível de ruídos δ convergir para zero. Aliado a isso tudo, o parâmetro de regularização α deve ser escolhido de forma adequada. Existem basicamente duas formas de escolha do parâmetro de regularização. Estas formas de escolha para α ficarão mais claras logo abaixo.

Queremos enfatizar que um método de regularização consiste:

- a) de uma estratégia para aproximar o operador inverso A^{-1} de maneira a evitar o mal condicionamento

¹Nós, matemáticos, gostamos de denominar os truques, as estratégias e outros artifícios por **métodos**.

- b) de uma regra para escolha de parâmetros de regularização, no sentido que, se o parâmetro de regularização é escolhido de acordo com essa regra, então a solução regularizada converge (em alguma norma) para a solução do problema, quando o nível de ruído tende para zero.
- c) do conceito de solução que estamos considerando e da topologia em que esse conceito de solução está imerso.

4.2 Resultados de Convergência

Diante ao apresentado até então, surgem as seguintes questões:

- (i) Como construir uma família de operadores de regularização?
- (ii) Como obter uma escolha de parâmetros para que um tal método de regularização convirja?
- (iii) É possível obtermos alguma performance “ótima” nesse caminho?

Para responder as questões (i) e (ii) acima, apresentaremos, logo abaixo, alguns métodos de regularização, divididos em duas classes: contínuos (por exemplo o Método de Tikhonov) e iterativos (por exemplo o Método de Landweber). Estes respondem, pelo menos em parte, as primeiras duas questões.

Para responder a (iii), ou seja, para assegurar que um método de regularização aplicado ao problema inverso (2.1) converge a uma solução e para expressar essa convergência em termos de taxas, é necessário obtermos algumas informações *a-priori* sobre a solução exata x^\dagger ou sobre y . Essas informações *a-priori* são formuladas em termos das *condições de fonte* (*source conditions*).

Como o próprio nome sugere, uma condição de fonte é algum tipo de informação *a priori* sobre a solução do problema. Em geral, aparece na forma de uma representação da solução x^\dagger em termos da imagem do operador A^* (ou A), ou como potências da imagem do mesmo [1, 2, 13, 9, 25].

No nosso contexto, podemos dizer o seguinte:

Teorema 4.2.1. *Seja $A : \mathcal{H}_1 \longrightarrow \mathcal{H}_2$ linear e limitado.*

i) Se $x \in \text{Im}(A^)$ e $\|Ax\| \leq \tau$, então*

$$\|x\| \leq C_1 \tau^{\frac{1}{2}}. \quad (4.3)$$

*ii) Se $x \in \text{Im}(A^*A)$ e $\|Ax\| \leq \tau$, então*

$$\|x\| \leq C_2 \tau^{\frac{2}{3}}. \quad (4.4)$$

Demonstração : *i)* De $x \in \text{Im}(A^*)$, segue que existe $y \in \mathcal{H}_2$, com $x = A^*y$ e $\|y\| \leq C_1^{\frac{1}{2}}$. Assim,

$$\|x\|^2 = \langle x, x \rangle = \langle x, A^*y \rangle = \langle Ax, y \rangle \leq \|Ax\| \|y\| \leq C_1^{\frac{1}{2}} \tau.$$

ii) De $x \in \text{Im}(A^*A)$, segue que existe $z \in \mathcal{H}_1$, com $x = A^*Az$ e $\|z\| \leq C_2^{\frac{1}{2}}$. Portanto,

$$\|x\|^2 = \langle x, A^*Az \rangle = \langle Ax, Az \rangle \leq \tau \|Az\| = \tau (\langle Az, Az \rangle)^{\frac{1}{2}} = \tau (\langle z, x \rangle)^{\frac{1}{2}} \leq \tau (\|z\| \|x\|)^{\frac{1}{2}}. \quad \blacksquare$$

Uma interpretação da condição i) no Teorema acima é a de que a inversa de $A|_{K_C} : K_C \rightarrow A(K_C)$ é contínua em y , onde $K_C := \{x : x = A^*y, \|y\| \leq C\}$. Note que a limitação em ii) é melhor pois estamos assumindo mais condições na solução x .

Exercício 4.2.1. Um operador B é dito ser semi-definido positivo se $\langle z, Bz \rangle \geq 0 \forall z \in \mathcal{D}(B)$. Prove que se A é um operador linear limitado entre espaços de Hilbert, então A^*A e AA^* são operadores lineares semi-definido positivos.

Exemplo 4.2.1 (Diferenciação nos dados). Considere o operador

$$\begin{aligned} A : L^2[0, 1] &\longrightarrow L^2[0, 1] \\ x &\longmapsto (Ax)(t) = \int_0^t x(s) ds = y(t), \quad t \in [0, 1]. \end{aligned} \quad (4.5)$$

Suponha que tenhamos uma informação a priori de que $x \in \{v \in AC[0, 1] : v(1) = 0 \text{ e } v' \in L^2[0, 1]\}$ e que a norma da primeira derivada seja estimada por:

$$\|x'\|_{L^2[0,1]}^2 \leq C.$$

Claramente, o operador adjunto é dado por

$$\begin{aligned} A^* : L^2[0, 1] &\longrightarrow L^2[0, 1] \\ y &\longmapsto (A^*y)(s) = - \int_s^1 y(r) dr. \end{aligned} \quad (4.6)$$

Assim, se $\|Ax\| \leq \tau$, então

$$\|x\|_{L^2[0,1]}^2 = - \int_0^1 x'(t) \int_0^t x(s) ds dt = - \int_0^1 x'(t) Ax(t) dt \leq \|x'\| \|Ax\| \leq C\tau.$$

Exercício 4.2.2. Prove que o operador adjunto do operador A definido pela equação (4.5) é o operador dado pela equação (4.6).

Exercício 4.2.3. Suponha A um operador linear limitado com $\text{Im}(A)$ fechada. Construa um operador de regularização para o problema inverso $Ax = y$. Que regra foi usada na construção do parâmetro de regularização? (Sugestão: Usar o Teorema 3.1.4)

Observação: O exercício acima mostra que, no caso em que o operador A possui imagem fechada, a **pseudo-inversa** é uma possível regularização para o problema inverso.

Existem basicamente duas formas de escolha do parâmetro de regularização: uma escolha *a-priori*, ($\alpha = \alpha(\delta)$) ou uma escolha *a-posteriori* ($\alpha = \alpha(\delta, y^\delta)$), dependendo do método de regularização utilizado. Passaremos a estudar cada um dos casos com um pouco mais de detalhes.

Uma pergunta que pode ser feita é: Existe uma escolha de parâmetros de regularização α que dependa somente dos dados y^δ e não dependa do nível de ruídos δ ? A resposta é a seguinte:

Proposição 4.2.1. *Seja $A : \mathcal{H}_1 \longrightarrow \mathcal{H}_2$ um operador linear limitado. Suponha que existe $\{R_\alpha\}$ uma regularização para A^\dagger , com uma escolha de parâmetros α que dependa somente de y^δ mas não dependa de δ , tal que o método de regularização (R_α, α) seja convergente, para todo $y \in \mathcal{D}(A^\dagger)$. Então A^\dagger , é limitado.*

Demonstração : Seja $\alpha = \alpha(\delta)$. Pela definição de método de regularização convergente, temos que

$$\limsup_{\delta \rightarrow 0} \{ \|R_{\alpha(y^\delta)} y^\delta - A^\dagger y\| : y^\delta \in \mathcal{H}_2, \|y - y^\delta\| \leq \delta \} = 0, \quad (4.7)$$

e, portanto, $R_{\alpha(y)} y = A^\dagger y$ para todo $y \in \mathcal{D}(A^\dagger)$. Logo, de (4.7), segue que para qualquer sequência $\{y_n\} \in \mathcal{D}(A^\dagger)$ que converge para $y \in \mathcal{D}(A^\dagger)$,

$$A^\dagger y_n = R_{\alpha(y_n)} y_n \longrightarrow A^\dagger y,$$

e, assim, A^\dagger é limitado no $\mathcal{D}(A^\dagger)$. Pelo Teorema 3.1.4, temos que $\mathcal{D}(A^\dagger) = \mathcal{H}_2$. ■

Exercício 4.2.4. *Justifique a última afirmação na Proposição 4.2.1.*

A Proposição 4.2.1 nos ensina que uma escolha de parâmetros para um método de regularização convergente para problemas mal postos deve, obrigatoriamente, levar em conta o nível de ruídos δ .

4.2.1 Escolha *a priori* do Parâmetro de Regularização

Uma escolha do parâmetro de regularização *a-priori* ($\alpha = \alpha(\delta)$) é, teoricamente, feita antes de qualquer cálculo numérico na tentativa de resolver o problema inverso. Desta forma, não depende do cálculo atual, digamos o resíduo $\|Ax_\alpha^\delta - y^\delta\|$.

Por enquanto, para o caso de A ser um operador linear, temos:

Proposição 4.2.2. *Para todo $\alpha > 0$, suponha que R_α um operador contínuo. Então a família R_α é uma regularização para A^\dagger se*

$$R_\alpha \longrightarrow A^\dagger \quad \text{pontualmente no } \mathcal{D}(A^\dagger) \quad \text{quando } \alpha \longrightarrow 0. \quad (4.8)$$

Neste caso, existe, para cada $y \in \mathcal{D}(A^\dagger)$, uma escolha a priori para α tal que $x_\alpha := R_\alpha y$ convirja para uma solução de $Ax = y$.

Demonstração : Fixe $y \in \mathcal{D}(A^\dagger)$ qualquer. Pela hipótese, existe um função monótona $\gamma : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ com $\lim_{\varepsilon \rightarrow 0} \gamma(\varepsilon) = 0$ tal que, para todo $\varepsilon > 0$,

$$\|R_{\gamma(\varepsilon)}y - A^\dagger y\| \leq \frac{\varepsilon}{2}.$$

Da continuidade de $R_{\gamma(\varepsilon)}$, para cada $\varepsilon > 0$ existe um $\rho(\varepsilon)$ tal que

$$\|z - y\| \leq \rho(\varepsilon) \quad \Rightarrow \quad \|R_{\gamma(\varepsilon)}z - R_{\gamma(\varepsilon)}y\| \leq \frac{\varepsilon}{2}.$$

Sem perda de generalidade, podemos assumir que a função $\rho(\varepsilon)$ é estritamente monótona, contínua e $\lim_{\varepsilon \rightarrow 0} \rho(\varepsilon) = 0$. Portanto, a inversa ρ^{-1} existe na imagem de ρ , é estritamente monótona, contínua e satisfaz $\lim_{\delta \rightarrow 0} \rho^{-1}(\delta) = 0$.

Defina $\alpha := \gamma(\rho^{-1}(\delta))$. Note que α é monótona e satisfaz $\lim_{\delta \rightarrow 0} \alpha(\delta) = 0$. Como $\|y - y^\delta\| \leq \delta$, uma simples desigualdade triangular mostra que

$$\|R_{\alpha(\delta)}y^\delta - A^\dagger y\| \leq \varepsilon.$$

Isso demonstra a nossa afirmação. ■

Exercício 4.2.5. *Mostre que se uma família de operadores lineares converge uniformemente para um operador linear, então este limite é contínuo. Use isto para justificar porquê não podemos requerer a convergência uniforme na Proposição 4.2.2. Ainda, mostre que, no caso em que $Im(A)$ não é fechada, então $\|R_\alpha\| \xrightarrow{\alpha \rightarrow 0} +\infty$.*

Quando estamos resolvendo problemas inversos, temos que ter sempre em mente o quanto queremos, de fato, aproximar a solução do problema inverso. Vamos ser mais específicos.

Suponha que $Im(A)$ não é fechada, assim, A^\dagger é não limitada (pelo Teorema 3.1.4). Seja $\{R_{\alpha(\delta)}\}$ uma estratégia linear de regularização para o problema $Ax = y^\delta$. Seja $y^\delta \in \mathcal{H}_2$ satisfazendo $\|y - y^\delta\| \leq \delta$, então

$$\begin{aligned} \|R_{\alpha(\delta)}y^\delta - A^\dagger y\| &\leq \|x_{\alpha(\delta)} - A^\dagger y\| + \|R_{\alpha(\delta)}y^\delta - x_{\alpha(\delta)}\| \\ &= \|x_{\alpha(\delta)} - A^\dagger y\| + \|R_{\alpha(\delta)}y^\delta - R_{\alpha(\delta)}y\| \leq \|x_{\alpha(\delta)} - A^\dagger y\| + \delta \|R_{\alpha(\delta)}\|. \end{aligned} \quad (4.9)$$

Notamos que, na estimativa (4.9) temos dois efeitos competindo. O primeiro termo é o efeito da regularização: quanto menor for $\alpha(\delta)$, melhor é a solução aproximada $x_{\alpha(\delta)}$ para x^\dagger . O segundo termo é o efeito da má-colocação do problema inverso: quando $\alpha(\delta) \rightarrow 0$, $\|R_{\alpha(\delta)}\| \rightarrow \infty$ (pelo exercício 4.2.5). A figura ?? ilustra esta situação.

O caso em que temos igualdade na equação (4.9) é, sem sombra de dúvidas, o **pior caso**. Mas, temos que trabalhar com a hipótese de que o pior caso aconteça. Assim, a importância de escolher α de forma apropriada (e positivo) fica evidente, mesmo que tenhamos que abrir mão de aproximar, o tanto quanto queríamos, o problema original.

Algumas técnicas de escolha *a priori* para o parâmetro α são bem conhecidas e amplamente usadas. Uma delas é a chamada de **curva L**. Não entraremos em detalhes aqui. Para interessados sugerimos [2, 9].

Exercício 4.2.6. *Faça uma comparação entre a estimativa (4.9) e a estimativa (2.4). Qual o papel de h em (2.4)?*

Assim, para obtermos uma estratégia ótima de solução para problemas inversos com uma escolha *a priori* do parâmetro de regularização, temos que ser capazes de escolher valores apropriados para α de forma a balancearmos a equação (4.9). Uma maneira de construir uma família de regularização adequadamente será apresentada logo mais na Seção 4.3 em termos dos valores espectrais. Essa técnica está baseada na construção das chamadas *funções filtro* [1, 2, 25] ou nas funções de truncamento. Nos Capítulos 5 e 6 apresentaremos outras formas de escolher o parâmetro de regularização.

Da estimativa (4.9), segue que:

Proposição 4.2.3. *Seja $\{R_\alpha\}$ uma estratégia linear de regularização. Para cada $y \in \mathcal{D}(A^\dagger)$, seja α uma escolha *a priori* para o parâmetro de regularização. Então, (R_α, α) é uma estratégia de regularização convergente se e só se*

$$\lim_{\alpha \rightarrow 0} \alpha(\delta) = 0 \quad e \quad \lim_{\delta \rightarrow 0} \delta \|R_{\alpha(\delta)}\| = 0.$$

Exercício 4.2.7. *Demonstre a Proposição 4.2.3.*

4.2.2 Escolha *a posteriori* do Parâmetro de Regularização

Uma escolha *a posteriori* do parâmetro de regularização é feita via uma comparação entre o resíduo (ou a **discrepância**), i.e.,

$$\|Ax_\alpha^\delta - y^\delta\| \leq \tau\delta \tag{4.10}$$

e o nível de ruídos δ . Esta escolha é chamada de *Princípio da Discrepância*.

Observação: Uma motivação heurística para tal escolha é a seguinte: Queremos resolver $Ax = y$, mas só conhecemos y^δ com $\|y - y^\delta\| \leq \delta$. Assim, pedir que, para uma solução aproximada x_α^δ , com $\|Ax_\alpha^\delta - y^\delta\| < \delta$, não faz sentido. Ou seja, o melhor que podemos esperar é que tenhamos um resíduo da ordem de δ .

Voltando a analisar a equação (4.9), vemos que quanto menor o parâmetro de regularização, pior é a estabilidade. Assim, devemos escolher α *a posteriori* o **maior** possível tal que a *discrepância*

(4.10) seja satisfeita. Notamos que, se $\delta = 0$, então o princípio da discrepância nunca é atingido. Neste caso, tomamos $\alpha := \alpha(y, \delta = 0) = +\infty$. Disto segue o Teorema:

Teorema 4.2.2. *Um método de regularização (R_α, α) , onde $\alpha := (\delta, y^\delta)$ é escolhido de acordo com o princípio da discrepância (4.10), é convergente $\forall y \in Im(A)$.*

Demonstração : Veja [9, Teorema 4.17]. ■

4.3 Regularização por Truncamento dos Valores Singulares

Nesta seção construiremos o primeiro **método de regularização** específico destas notas. Nos deteremos aos detalhes do caso particular em que o operador A é linear e compacto. Para o caso em que A é um operador linear limitado qualquer, os resultados são muito bem apresentados em [9].

Vamos direto ao problema a ser considerado.

Seja A um operador linear e compacto com um sistema singular (σ_n, e_n, f_n) . Considere o problema de encontrar x na equação

$$Ax = y^\delta.$$

Como y^δ pode não pertencer a $Im(A)$, temos que nos contentar em encontrar uma melhor aproximação $x^\dagger = A^\dagger y^\delta$ da solução exata x . Ou, equivalentemente, encontrar entre as soluções das equações normais

$$A^*Ax = A^*y^\delta,$$

a solução x^\dagger que tem a menor norma.

Do Teorema de Picard 3.2.6, temos que uma possibilidade de solução seria

$$x^\delta = \sum_{j=1}^{\infty} \sigma_j^{-2} \langle A^*y^\delta, e_j \rangle e_j. \quad (4.11)$$

Vimos na Observação 3.2.2 que usar a equação (4.11) não é uma boa alternativa para calcular uma aproximação para a solução x^\dagger do problema inverso acima, haja visto que A^*A também é compacto e, assim, $\sigma_j \rightarrow \infty$.

Para $\alpha \in (0, \alpha_0)$ e $\lambda \in [0, \|A\|^2]$, defina a função (contínua a direita)

$$f_\alpha(\lambda) = \begin{cases} \frac{1}{\lambda}, & \text{se } \lambda \geq \alpha \\ 0, & \text{se } \lambda < \alpha. \end{cases}$$

Portanto, pelo Teorema da Aplicação Espectral,

$$\begin{aligned} x_\alpha^\delta &:= f_\alpha(A^*A)A^*y^\delta = \sum_{\substack{j=1 \\ \sigma_j^2 \geq \alpha}}^{\infty} \sigma_j^{-2} \langle A^*y^\delta, e_j \rangle e_j = \sum_{\substack{j=1 \\ \sigma_j^2 \geq \alpha}}^{\infty} \sigma_j^{-2} \langle y^\delta, Ae_j \rangle e_j \\ &= \sum_{\substack{j=1 \\ \sigma_j^2 \geq \alpha}}^{\infty} \sigma_j^{-2} \langle y^\delta, \sigma_j f_j \rangle e_j = \sum_{\substack{j=1 \\ \sigma_j^2 \geq \alpha}}^{\infty} \sigma_j^{-1} \langle y^\delta, f_j \rangle e_j. \end{aligned} \quad (4.12)$$

A definição de x_α^δ como na equação (4.12) pode ser vista como uma versão *truncada* da expansão em valores singulares (3.7).

Definição 4.3.1. *O método dado pela equação (4.12) é chamado de **expansão truncada em valores singulares**.*

Exercício 4.3.1. *Mostre que x_α^δ dado em (4.12) pode ser calculado como $x_\alpha^\delta = A_\alpha^\dagger y^\delta$, onde A_α é um operador com imagem de dimensão finita definido por*

$$A_\alpha x = \sum_{\substack{j=1 \\ \sigma_j^2 \geq \alpha}}^{\infty} \sigma_j \langle x, e_j \rangle f_j.$$

Observação: Calcular x_α^δ por $x_\alpha^\delta = A_\alpha^\dagger y^\delta$ é um método de projeção sobre os auto-espacos de A^*A . Ainda, o nível de truncamento α , que decide quando os valores singulares são trocados por 0, age como uma escolha *a priori* do parâmetro de regularização.

Teorema 4.3.1. *O método de expansão truncada em valores singulares é um método de regularização.*

Demonstração : Basta notarmos que A_α satisfaz a Proposição 4.2.2. ■

Exercício 4.3.2. *Preencha os detalhes da demonstração do Teorema 4.3.1.*

Exercício 4.3.3 (Taxas de convergência). *Seja A um operador linear compacto com sistema singular (σ_j, e_j, f_j) . Suponha que a solução x^\dagger de $Ax = y$ satisfaça a condição de fonte*

$$x^\dagger \in \text{Im}((A^*A)^\nu) \quad \nu > 0. \quad (4.13)$$

i) Mostre que a condição de fonte (4.13) é satisfeita se e só se

$$\sum_{j=1}^{\infty} \frac{|\langle y, f_j \rangle|^2}{\sigma_j^{2+4\nu}} < \infty \quad (4.14)$$

ii) Suponha que tenhamos somente dados perturbados y^δ com $\|y - y^\delta\| \leq \delta$. Dê uma estimativa da taxa de convergência de x_α^δ par x^\dagger em função dos valores singulares.

Capítulo 5

Regularização de Tikhonov

Com a mesma filosofia da teoria geral de regularização para problemas inversos, a **regularização de Tikhonov** é um compromisso entre precisão e estabilidade. O objetivo deste capítulo é estabelecermos tal compromisso. Nestas notas, trataremos exclusivamente do caso em que o operador A é linear. O estudo do método de regularização de Tikhonov para problemas não lineares pode ser encontrado em [2, 6, 9, 47].

5.1 Problemas Lineares: Convêrgencia

Nesta seção, consideraremos o caso em que o operador A é linear e limitado. Com essa hipótese, estamos interessados em encontrar, de forma estável, uma aproximação para a solução do problema inverso

$$Ax = y^\delta,$$

para medidas conhecidas do erro $\|y - y^\delta\| \leq \delta$.

Assim, como nos métodos iterativos estudados nos Capítulos acima, uma solução regularizada requer uma estratégia mais adequada que tentar resolver as equações normais

$$A^*Ax = A^*y^\delta. \tag{5.1}$$

Ou, de forma equivalente, encontrar um mínimo para o problema variacional de quadrados mínimos

$$J(x) = \frac{1}{2} \|Ax - y^\delta\|^2.$$

Lembrando, tudo o que queremos é "inverter" o operador A de maneira estável. Mais que isso, não queremos errar muito ao fazer essa inversão, i.e., queremos manter o resíduo $\|Ax - y^\delta\|$ controlado. Veja a estimativa (4.9). Então, formalmente, gostaríamos de inverter de forma estável

o operador A^*A . Pelo Teorema da Aplicação Espectral 3.2.5, isto equivale a calcular $g(A^*A)$, onde $g(\lambda) = \frac{1}{\lambda}$, $\lambda \neq 0$. Portanto, uma solução aproximada para a equação (5.1) é dada por $x^\delta = g(A^*A)A^*y^\delta$.

Antes de continuarmos, vamos a algumas propriedades importantes, que enunciaremos em forma de exercício.

Exercício 5.1.1. *Dê um exemplo de um operador linear que possui espectro não-vazio e que não possui autovalores. Isso pode acontecer no caso em que A é uma matriz, isto é, em dimensão finita?*

Exercício 5.1.2. *Seja A um operador linear limitado entre espaços de Hilbert. Mostre que o espectro de A^*A é real e positivo.*

Exercício 5.1.3. *Dê um exemplo de um operador linear limitado A tal que 0 seja um elemento do espectro de A . Existe operadores auto-adjuntos tais que 0 seja um autovalor? Se afirmativo, dê exemplos. Sugestão: procure exemplos nos espaços de seqüências.*

Aprendemos, dos exercícios acima, que o espectro de A^*A pode estar muito próximo de ZERO e até mesmo conter o 0. Portanto, a estratégia de calcular x^δ , como $x^\delta = g(A^*A)A^*y^\delta$ não é possível, ou, se é possível, é muito instável.

Exercício 5.1.4. *Mostre que se λ é um autovalor de A , então λ^2 é um autovalor de A^*A . Seja A seja uma matriz inversível. Mostre que $1/\lambda$ é um autovalor de A^{-1} .*

Exercício 5.1.5. *Suponha que A^*A seja inversível. Considere $x^\dagger = g(A^*A)A^*y$ e $x^\delta = g(A^*A)A^*y^\delta$. Use o exercício acima para mostrar que*

$$\|x^\dagger - x^\delta\|^2 \leq (g(A^*A))^2 \|A^*\|^2 \|y - y^\delta\|^2.$$

*O que acontece com $\|x^\dagger - x^\delta\|$ como uma função dos autovalores de A^*A , para dados com ruído, satisfazendo $\|y - y^\delta\| \leq \delta$.*

Qual é a estratégia para obter uma solução aproximada de (2.1) de maneira estável? Afastar o espectro de A^*A de zero.

Seja $0 < \alpha \in [0, \alpha_0]$, defina

$$f_\alpha(\lambda) := g(\lambda + \alpha) = \frac{1}{\lambda^2 + \alpha}. \quad (5.2)$$

A função $f_\alpha(\cdot)$ é dita ser a *função filtro* para o método de regularização de Tikhonov.

Exercício 5.1.6. *Mostre que $f_\alpha(\cdot)$ é contínua pela direita.*

Segue, do Teorema da Aplicação Espectral, que

$$f_\alpha(A^*A) = (A^*A + \alpha I)^{-1}. \quad (5.3)$$

Exercício 5.1.7. *Seja A um operador linear e limitado entre espaços de Hilbert. Mostre que, para todo $0 < \alpha \in \mathbb{R}$, o operador $A^*A + \alpha I$ é linear, limitado, injetivo e sobrejetivo e, assim, $(A^*A + \alpha I)^{-1}$ existe e é limitada. Sugestão: Use o Teorema do Gráfico Fechado, [26], para mostrar a limitação da inversa.*

Segue do exercício acima que a escolha de x_α^δ , da forma

$$x_\alpha^\delta = (A^*A + \alpha I)^{-1}A^*y^\delta \quad (5.4)$$

é uma solução regularizada, definida via a equação linear

$$(A^*A + \alpha I)x_\alpha^\delta = A^*y^\delta. \quad (5.5)$$

Esta última equação pode ser pensada como uma regularização para as equações normais (5.1). Este método é chamado de **Regularização de Tikhonov**¹.

Exercício 5.1.8. *Seja A um operador linear e compacto entre espaços de Hilbert com um sistema singular dado por (σ_j, e_j, f_j) . Mostre que a solução regularizada x_α^δ na equação (5.4) tem a forma*

$$x_\alpha^\delta = \sum_{j=1}^{\infty} \frac{\sigma_j}{\sigma_j^2 + \alpha} \langle y^\delta, f_j \rangle e_j. \quad (5.6)$$

Use o mesmo raciocínio para mostrar que $x^\delta = g(A^*A)A^*y^\delta$ satisfaz

$$x^\delta = \sum_{j=1}^{\infty} \frac{1}{\sigma_j} \langle y^\delta, f_j \rangle e_j. \quad (5.7)$$

Observação: Fazendo uma comparação entre (5.7) e (5.6), vemos claramente o resultado de estabilidade da equação (5.6): o erro em $\langle y^\delta, f_j \rangle$ é propagado com um fator de $\frac{\sigma_j}{\sigma_j^2 + \alpha}$, o qual é sempre limitado quando $j \rightarrow \infty$. Em (5.7), o fator de propagação é $\frac{1}{\sigma_j}$.

Como nem sempre estamos trabalhando com operadores compactos e, mesmo se esse for o caso, a determinação de um sistema singular de um operador é uma tarefa muito custosa (tanto do ponto de vista numérico quanto matemático). Seria ideal termos uma outra forma de determinar uma solução pela regularização de Tikhonov. Isto é dado pelo teorema abaixo que trata de uma versão variacional da regularização de Tikhonov:

¹Muitas vezes é chamado de *Regularização de Tikhonov-Phillips*.

Teorema 5.1.1. *Seja x_α^δ como na equação (5.4). Então x_α^δ é o único minimizador do **funcional de Tikhonov***

$$J_\alpha(x) := \|Ax - y^\delta\|^2 + \alpha\|x\|^2. \quad (5.8)$$

Demonstração: Para $\alpha > 0$, o funcional J_α é estritamente convexo e coercivo. Assim, J_α possui um único minimizador que deve satisfazer a condição necessária (e neste caso também suficiente) de primeira ordem

$$J'_\alpha(x).h = 0 \quad \text{para todo } h \in \mathcal{H}_1. \quad (5.9)$$

Disto, segue que

$$\begin{aligned} 0 = J'_\alpha(x).h &= 2\langle Ax - y^\delta, Ah \rangle + 2\alpha\langle x, h \rangle \\ &= 2\langle A^*Ax - A^*y^\delta + \alpha Ix, h \rangle \end{aligned} \quad (5.10)$$

para todo $h \in \mathcal{H}_1$. Portanto, (5.10) é equivalente a (5.4). ■

Observação: Note que, qualquer minimizador do funcional de Tikhonov (5.8) pertence a $\mathcal{N}(A)^\perp$. Com efeito, pois, caso contrário, poderíamos fazer a segunda parte do funcional crescer, mantendo a primeira parte constante.

Exercício 5.1.9. *Um funcional $J : \mathcal{H} \rightarrow \mathbb{R}$ é dito ser **coercivo** se $\lim_{\|x\| \rightarrow \infty} J(x) = +\infty$. Mostre que, se J é convexo e coercivo, então, este atinge um mínimo. Ainda, se J é estritamente convexo e coercivo, mostre que o mínimo é único. Sugestão: Consulte [22].*

Exercício 5.1.10. *Mostre que dado um funcional J como no exercício acima, então a condição de otimalidade de primeira ordem é necessária e também suficiente para x ser um minimizante de J . Sugestão: Consulte [22].*

O parâmetro α no funcional (5.8) é o parâmetro de regularização. Minimização em (5.8) é um compromisso entre minimizar a norma do resíduo $\|Ax - y^\delta\|$ e tomar o tamanho do **termo de penalização** $\|x\|$ pequeno e, assim, forçar a estabilidade. A escolha apropriada do parâmetro α é ainda um problema e deve ser feita *a priori*. Como α é o parâmetro que estabelece este compromisso, muitos autores sugerem sua escolha através da chamada **curva L**. Um boa referência para a construção da curva L é [2].

Observação: Notemos que a definição de x_α^δ como em (5.4) só tem sentido para operadores lineares. Mas, o problema de minimização (5.8) pode ser formulado para operadores não-lineares [2, 6, 9, 47].

5.2 Problemas Lineares: Semi-Convergência

A definição da solução regularizada, pela minimização do funcional de Tikhonov (5.8), nos dá diretamente resultados de convergência e estabilidade, como:

Teorema 5.2.1. *Seja x_α^δ definida por (5.4), $y \in \text{Im}(A)$ com $\|y - y^\delta\| \leq \delta$. Se $\alpha := \alpha(\delta)$ é tal que*

$$\lim_{\delta \rightarrow 0} \alpha(\delta) = 0 \quad e \quad \lim_{\delta \rightarrow 0} \frac{\delta^2}{\alpha(\delta)} = 0, \quad (5.11)$$

então

$$\lim_{\delta \rightarrow 0} x_{\alpha(\delta)}^\delta = A^\dagger y. \quad (5.12)$$

Demonstração : Seja $\delta_n \rightarrow 0$ qualquer. Definimos $\alpha_n := \alpha(\delta_n)$ e $x_n := x_{\alpha_n}^{\delta_n}$. Seja J_n o funcional de Tikhonov (5.8) com $\alpha = \alpha_n$ e x_n seu correspondente minimizante (que existe e é único, pelo Teorema 5.1.1). Seja $x^\dagger := A^\dagger y$. Então, por definição de x_n , temos

$$\begin{aligned} \alpha_n \|x_n\|^2 &\leq J_n(x_n) \leq J_n(x^\dagger) = \|Ax^\dagger - y^{\delta_n}\|^2 + \alpha_n \|x^\dagger\|^2 \\ &\leq \delta_n^2 + \alpha_n \|x^\dagger\|^2, \end{aligned}$$

e, assim,

$$\|x_n\|^2 \leq \frac{\delta_n^2}{\alpha_n} + \|x^\dagger\|^2. \quad (5.13)$$

Portanto, por hipótese, $\{x_n\}$ é uniformemente limitada. Pelo Teorema de Banach-Alaoglu A.1.2, $\{x_n\}$ possui uma subsequência que converge fraco² para $z \in \mathcal{H}$. Como A é linear e limitado,

$$Ax_{n_k} \rightharpoonup Az. \quad (5.14)$$

Novamente, a definição de x_{n_k} implica que

$$\|Ax_{n_k} - y^{\delta_{n_k}}\|^2 \leq J_{n_k}(x_{n_k}) \leq \delta_{n_k}^2 + \alpha_{n_k} \|x^\dagger\|^2 \xrightarrow{k \rightarrow \infty} 0.$$

Segue de (5.14) que

$$Az = y. \quad (5.15)$$

²Uma sequência em um espaço de Hilbert é **fracamente convergente** para $z \in \mathcal{H}$ se para todo $h \in \mathcal{H}$ temos $\langle x_n, h \rangle \rightarrow \langle z, h \rangle$. Notação: $x_n \rightharpoonup z$.

Da Observação 5.1, temos que $x_n \in \mathcal{N}(A)^\perp$ e, assim, $z \in \mathcal{N}(A)^\perp$ (Prove). Pelo Teorema 3.1.2 e sua demonstração, obtemos $z = x^\dagger$. Assim, $x_{n_k} \rightharpoonup x^\dagger$. Aplicando o mesmo argumento para todas as subsequências, obtemos que

$$x_n \rightharpoonup x^\dagger. \quad (5.16)$$

Afirmção: $x_n \rightarrow x^\dagger$.

Assuma que exista $\varepsilon > 0$ tal que $\|x_{n_k}\| \leq \|x^\dagger\| - \varepsilon$. Então, o mesmo argumento de extração de subsequências acima implica que $\|z\| \leq \|x^\dagger\| - \varepsilon$, contradizendo (5.16). Logo,

$$\liminf \|x_n\| \geq \|x^\dagger\|. \quad (5.17)$$

De (5.13), temos

$$\liminf \|x_n\| \leq \|x^\dagger\|. \quad (5.18)$$

Juntando as últimas três estimativas, temos que $x_n \rightarrow x^\dagger$. ■

Exercício 5.2.1. *Nas hipóteses do Teorema acima, mostre que (5.16) - (5.18) implicam que $x_n \rightarrow x^\dagger$.*

5.2.1 Taxas de convergência

Segue da definição de solução regularizada pelo método de Tikhonov que

$$\|x_\alpha^\delta - x^\dagger\| \leq \sup_{\lambda \in \Sigma(A)} |\lambda f_\alpha(\lambda)| \|y - y^\delta\| \leq \frac{\delta}{\sqrt{\alpha}}. \quad (5.19)$$

Assim, se $\alpha \sim \delta$, obtemos a seguinte ordem de convergência

$$\|x^\dagger - x_\alpha^\delta\| = \mathcal{O}(\sqrt{\delta}). \quad (5.20)$$

Exercício 5.2.2. *Verifique a desigualdade em (5.19) no caso em que A é compacto e com um sistema singular dado por (σ_j, e_j, f_j) .*

5.2.2 Regularização de Tikhonov e a Inversa Generalizada

De acordo com a definição variacional, $A^\dagger y$ é o vetor $x \in \mathcal{H}_1$ de menor norma que minimiza o funcional $\|Ax - y\|$. Por outro lado, a ideia da regularização de Tikhonov é minimizar ambos: o funcional, $\|Ax - y\|$, e a norma, $\|x\|$, através da minimização do funcional de Tikhonov (5.8).

Pelo Teorema 5.1.1, a única solução x_α satisfaz

$$x_\alpha = (A^*A + \alpha I)^{-1}A^*y.$$

Segue, do Teorema 5.2.1, que

$$\lim_{\alpha \rightarrow 0} x_\alpha = A^\dagger y.$$

Com esse resultados obtemos:

Corolário 5.2.1. *Seja A um operador linear e limitado satisfazendo a hipótese 3.1.1. Então,*

$$A^\dagger = \lim_{\alpha \rightarrow 0} (A^*A + \alpha I)^{-1}A^*,$$

uniformemente no espaço $\mathcal{L}(\mathcal{H}_2, \mathcal{H}_1)$.

Exercício 5.2.3. *Demonstre o Corolário 5.2.1. Sugestão: Use o que você sabe da função $f_\alpha(x) = (\alpha + x)^{-1}$, $\alpha > 0$, e o Teorema da Aplicação Espectral.*

Para obtermos um limitante para o erro com este método, considere $\lambda \in \Sigma(A^*A)$. Então temos que

$$|\lambda f_\alpha(\lambda) - 1| = \frac{\alpha}{\lambda + \alpha} \leq \frac{\alpha}{\|A^\dagger\|^{-2} + \alpha}.$$

Assim, segue que

$$\|(A^*A + \alpha I)^{-1}A^* - A^\dagger\| \leq \frac{\alpha \|A^\dagger\|^3}{1 + \alpha \|A^\dagger\|^2}. \quad (5.21)$$

Exercício 5.2.4. *Preencha os detalhes na demonstração da estimativa (5.21).*

5.3 Regularização de Tikhonov para o Problema de Tomografia Computadorizada

Como vimos antes, o método de regularização de Tikhonov produz soluções estáveis e convergentes para a solução do problema $Ax = y^\delta$, como função do nível de ruídos nos dados.

Nesta seção, adaptaremos a teoria desenvolvida anteriormente para o problema da Tomografia Computadorizada. Para tal, assumiremos que os dados p_i^δ são obtidos por medidas, tal que, $p_j = \mathcal{R}_j f$ na norma de $L^2(\Omega)$, para um conjunto de raios-X com $j = 1, \dots, N$. De acordo com a Subseção 2.4.3, temos que

$$\mathcal{R}_j f = (A(:, j), f) \quad j = 1, \dots, N,$$

onde (\cdot, \cdot) representa o produto interno em $L^2(\Omega)$ e $A(:, j)$ representa a coluna j da matriz A , dada na Subseção 2.4.3.

Lema 5.3.1. *Definindo, $\mathcal{R} : L^2(\Omega) \longrightarrow \mathbb{R}^N$ por*

$$\mathcal{R} = \begin{pmatrix} \mathcal{R}_1 \\ \vdots \\ \mathcal{R}_M \end{pmatrix}, \quad (5.22)$$

temos que o operador adjunto $\mathcal{R}^ : \mathbb{R}^N \longrightarrow L^2(\Omega)$ é dado por*

$$\mathcal{R}^* r = \sum_{j=1}^N A(:, j) r_j. \quad (5.23)$$

Demonstração : A demonstração fica como exercício para o leitor. ■

Exercício 5.3.1. *Mostre que $S = \mathcal{R}\mathcal{R}^* = (A(:, k), A(:, j))$ para $k, j = 1, \dots, N$.*

Como \mathcal{R} é linear, o funcional de Tikhonov (5.8) consiste em minimizar

$$\|\mathcal{R}f - p^\delta\|^2 + \alpha\|f\|^2.$$

Pelo Teorema 5.1.1, a única solução regularizada para o problema de Tomografia Computarizada, dada pelo método de regularização de Tikhonov, é

$$f_\alpha = \mathcal{R}^* r \quad (S + \alpha I)r = p^\delta. \quad (5.24)$$

Exercício 5.3.2. *Use o método de regularização de Tikhonov para encontrar uma solução aproximada para o problema do Exercício 2.4.11. Como deve ser escolhido α nesse caso?*

5.4 Método de Máxima Verossemelhança para Tomografia Computarizada

O *Método de Máxima Verossemelhança* é um método estatístico para estimar a imagem obtida como a melhor possível, levando em conta as medidas estatísticas dos raios-X projetados nos detectores. Mais precisamente, o procedimento de medidas de dados deve ser modelado como um processo estocástico cujos parâmetros f^* tem que ser estimados, dada uma amostra aleatória p (que são os valores da transformada de Radon).

Este é um método de reconstrução de imagens completamente diferente dos descritos anteriormente. Este método é tipicamente usado na situação onde o número de informações nos detectores

é muito pequeno ou com muitos erros. Neste caso, a reconstrução é muito afetada por ruído nos dados. Portanto, algum tipo de regularização deve ser usada.

Nesta seção, assumiremos que a quantidade de raio-X medida em cada detector é dada por uma distribuição de Poisson. Isto é, o decaimento, por unidade de tempo de cada pixel f_j , do objeto é representado por uma variável aleatória cuja distribuição de Poisson possui um valor esperado f_j^* . Assim, a probabilidade de medir um certo número $N_j = f_j$ do decaimento para um pixel com um valor esperado f_j^* pode ser modelado por

$$P(N_j|f_j) = \frac{(f_j^*)^{f_j}}{f_j!} e^{-f_j^*}. \quad (5.25)$$

Toda combinação linear

$$\sum_{j=1}^N a_{ij} f_j = p_i \quad (5.26)$$

de N pixels com valor esperado

$$\sum_{j=1}^N a_{ij} f_j^* = p_i^* \quad (5.27)$$

é novamente Poisson distribuída e estatisticamente independente. Enquanto isso, os pixels são as variáveis estatisticamente independentes, f_j , de modo que para a projeção M ($M_i = p_i$ em (5.26)), a distribuição de Poisson

$$P(M_i|p_i) = \frac{(p_i^*)^{p_i}}{p_i!} e^{-p_i^*}, \quad (5.28)$$

seja satisfeita. Para obter a probabilidade de todos os valores de projeção estatisticamente independentes, i.e., a probabilidade de observar o vetor $p = (p_1, \dots, p_M)^T$ para um dado valor esperado $p^* = (p_1^*, \dots, p_M^*)^T$, multiplicamos as probabilidades simples tal que

$$P(p|p^*) = \prod_{i=1}^M \frac{(p_i^*)^{p_i}}{p_i!} e^{-p_i^*}. \quad (5.29)$$

Substituindo (5.26) em (5.30), obtemos que

$$P(p|p^*) = \prod_{i=1}^M \frac{(\sum_{j=1}^N a_{ij} f_j^*)^{p_i}}{p_i!} e^{-\sum_{j=1}^N a_{ij} f_j^*} = P(p|f^*), \quad (5.30)$$

onde $f^* = (f_1^*, \dots, f_N^*)$.

Deste modo, podemos olhar para a equação (5.30) como uma função do vetor de valores esperados f^* . Isto é, o valor esperado da distribuição de atividades espacial é

$$L(f^*) = \prod_{i=1}^M \frac{(\sum_{j=1}^N a_{ij} f_j^*)^{p_i}}{p_i!} e^{-\sum_{j=1}^N a_{ij} f_j^*}. \quad (5.31)$$

A equação (5.31) é dita de função de verossemelhança.

O *Método de Máxima Verosselhança* consiste em variar todos os valores esperados f_j^* , com $j \in \{1, \dots, N\}$, de forma a maximizar o funcional L .

A distribuição f^* , que maximiza o funcional L , é chamada de *solução de máxima verossemelhança* para o problema de reconstrução. Esta pode ser interpretada como a solução mais provável para um dado conjunto de dados do problema e, portanto, a que deve ser usada.

Para evitar dificuldades técnicas, no que segue assumiremos que $p_i, p_i^*, a_{ij} > 0$ para todo i, j .

Aplicando logaritmo no funcional L , obtemos que (5.31) é equivalente a

$$\ln(L(f^*)) = \sum_{i=1}^M \left(p_i \ln \left(\sum_{j=1}^N a_{ij} f_j^* \right) - \ln(p_i!) - \sum_{j=1}^N a_{ij} f_j^* \right). \quad (5.32)$$

Como (5.32) não depende de p_i , definimos

$$l(f^*) = \sum_{i=1}^M \left(p_i \ln \left(\sum_{j=1}^N a_{ij} f_j^* \right) - \sum_{j=1}^N a_{ij} f_j^* \right) = \sum_{i=1}^M (p_i \ln (Af^*)_i - (Af^*)_i). \quad (5.33)$$

Assim, o problema de encontrar a distribuição f^* pode ser reescrito como

$$\max_{f^*} l(f^*) \quad \text{sujeito a} \quad f^* \in \Gamma, \quad (5.34)$$

onde Γ é o conjunto de todos as distribuições admissíveis.

Exercício 5.4.1. *Justifique a afirmação: Os funcionais $L(\cdot)$ em (5.31) e $l(\cdot)$ em (5.34) são equivalentes. Sugestão: Use propriedades da função \ln .*

O funcional de máxima verossemelhança (5.31) é sabido ser instável para resolver problemas inversos. Isto significa que temos que utilizar algum tipo de regularização. Ou seja, é necessário introduzir um termo de penalidade que controle o compromisso entre estabilidade e precisão para o problema.

A introdução de um funcional de regularização pode (e deve) ser interpretada como uma informação *a priori* para o problema. Nos problemas de recuperação de imagens, na presença de ruído nos dados, é natural requerermos que os tons de cinza em cada pixel não sejam muito dis-

tantes dos tons de cinza nos pixels vizinhos. Assim, é razoável supormos que os tons dos pixels a serem recuperados não difiram tanto da média, o que pode ser considerado como **a priori** f_0 .

Na prática, imagens são modeladas por processos de Markov. Para tais processos estocásticos, a distribuição condicional de observar f^* , dado f_0 , satisfaz a distribuição de Gibbs dada por

$$R(f^*, f_0) = \frac{1}{Z} \exp \left\{ -\alpha \sum_{j=1}^N |f_j^* - f_{0,j}|^2 \right\}, \quad (5.35)$$

para um processo Gaussiano generalizado. Z é uma constante de normalização e pode ser considerada $Z = 1$.

A motivação para a escolha da penalização (5.35) é a seguinte:

Exercício 5.4.2. *Mostre que um processo é Gaussiano se, e somente se, a distribuição de probabilidade satisfaz uma expressão do tipo (5.35).*

Assim, o funcional de máxima verossemelhança regularizado consiste em incorporar ao funcional (5.33) a penalização $\ln(R(f^*, f_0))$. Assim, estamos interessados em

$$\max_{f^*} J_\alpha(f^*), \quad (5.36)$$

onde

$$J_\alpha(f^*) = l(f^*) + \ln(R(f^*, f_0)) = \sum_{i=1}^M (p_i \ln((Af^*)_i) - (Af^*)_i) - \alpha \sum_{j=1}^N |f_j^* - f_{0,j}|^2. \quad (5.37)$$

Pelo teorema de dualidade para funcionais convexos, (5.37) é equivalente a minimizar o funcional de Tikhonov

$$F_\alpha(f^*) = \sum_{i=1}^M ((Af^*)_i - p_i \ln((Af^*)_i)) + \alpha \sum_{j=1}^N |f_j^* - f_{0,j}|^2. \quad (5.38)$$

Notemos que o funcional (5.38) é um funcional de Tikhonov generalizado, e o termo que mede o resíduo (5.33) é dado pelo funcional de *Kullback-Leibler*. O funcional de Kullback-Leibler pode ser considerado uma pseudo-distância, a qual tem propriedades que se assemelham com a norma. Para uma demonstração da existência de mínimos para o funcional (5.38), veja [47].

Exercício 5.4.3. *Seja $f : \mathcal{D}(f) \subset \mathbb{R}^N \rightarrow \mathbb{R}$ uma função convexa. Mostre que $-f$ é uma função concava.*

Exercício 5.4.4. *Seja f como no exercício anterior. Mostre que todo mínimo local de f , se existir, é um mínimo global.*

Exercício 5.4.5. *Mostre que, para uma função convexa as condições*

$$f'(x) = 0 \quad e \quad f''(x) \geq 0 \quad (5.39)$$

*são necessárias e suficientes para mostrar que x é um mínimo global de f . **Sugestão:** Consulte [22]. As condições dada na equação (5.39) são ditas condições de Karush-Kuhn-Tucker (KKT) para um problema de minimização convexa.*

Capítulo 6

Regularização por Métodos Iterativos

Como vimos no Capítulo 2, solucionar o problema inverso (2.1) de forma estável, isto é, de forma que condiz com a realidade dos dados obtidos, implica em encontrar uma forma de aproximar o operador inverso, por uma família de operadores contínuos. Ou seja, necessitamos de um método de (**regularização**).

Uma alternativa para regularização de operadores são os *métodos iterativos de regularização*. A vantagem de tais métodos é que eles possuem propriedades regularizantes [23, 25, 2, 9].

Gauss demonstrou que, a melhor maneira de determinar um parâmetro desconhecido de uma equação do tipo (2.1), é minimizando a soma dos quadrados dos resíduos¹, isto é,

$$\min_{x \in \mathcal{H}_1} \frac{1}{2} \|A(x) - y\|^2. \quad (6.1)$$

Assumindo algumas propriedades do operador A , prova-se que, o minimizador (caso exista) de (6.1) deve satisfazer a condição necessária de primeira ordem²

$$A'(x)^* A(x) = A'(x)^* y. \quad (6.2)$$

Exercício 6.0.6. *Seja A uma transformação linear de \mathbb{R}^n para \mathbb{R}^m . Mostre que a equação (6.2) é a condição de otimalidade (de primeira ordem) para o problema de mínimos quadrados (6.1). Generalize para o caso em que A é um operador linear entre espaços de Hilbert.*

Uma possibilidade para encontrar uma solução de (6.2) é interpreta-la como uma iteração de ponto fixo

$$x_{k+1} = \Phi(x_k) \quad (6.3)$$

¹Tal método é conhecido hoje em dia como **método de Mínimos Quadrados**.

²A equação (6.2) é chamada de **Equação Normal**.

para o operador

$$\Phi(x) = x + A'(x)^*(y - A(x)). \quad (6.4)$$

No entanto, Φ , em geral, não é contrativo.

Os métodos iterativos para aproximação de pontos fixos de operadores não expansivos Φ , isto é,

$$\|\Phi(x) - \Phi(\tilde{x})\| \leq \|x - \tilde{x}\|, \quad x, \tilde{x} \in \mathcal{D}(\Phi)$$

têm sido considerados atualmente [23]. Neste caso, maior ênfase é dada numa prova construtiva de pontos fixos para Φ .

Em muitos exemplos práticos (e até mesmo teóricos) é quase impossível verificar analiticamente quando o operador Φ é contrativo ou não. No contexto de problemas não-lineares, esta dificuldade é ainda maior, [23, 9].

Muitos métodos iterativos para resolver (2.1) são baseados na solução da equação normal (6.2), via sucessivas iterações, partindo de um chute inicial x_0 . Observamos que tal chute, em particular para problemas não-lineares, costuma conter informações *a priori* sobre a solução do problema.

Nesse capítulo, nos dedicaremos a alguns métodos do tipo gradiente, mais especificamente, ao método de Landweber (clássico), ao método steepest descent e algumas estratégias do tipo Kaczmarz (ART e Landweber Kaczmarz). No final, trataremos do método Expectation Maximization (EM) que é uma estratégia de ponto fixo para a equação de primeira ordem do funcional (5.33).

Nestas notas, nos deteremos ao estudo de métodos iterativos, considerando o caso em que o operador A é linear. O caso em que A é não linear possui tratamento detalhado em [6, 9, 23].

6.1 Método de Landweber

Nessa seção, assumiremos que o operador A na equação (2.1) é linear e limitado. Com essa hipótese, uma maneira de resolvermos (6.2) é considerarmos a iteração

$$x_{k+1} = x_k + \gamma A^*(y - Ax_k), \quad k = 0, 1, 2, \dots \quad (6.5)$$

em que $\|A\|^{-2} \geq \gamma > 0$ é um parâmetro de relaxação, de forma que a iteração tenha a propriedade de descida. Esse foi o ponto de partida de Landweber [27], em 1951, ao estudar equações integrais de primeira espécie, quando propôs o método que hoje leva o nome de **Método de Regularização de Landweber**.

No caso de dados com ruído y^δ , denotando as iterações por x_k^δ , chegamos a *iteração de Landweber*

$$x_{k+1}^\delta = x_k^\delta + A^*(y^\delta - Ax_k^\delta). \quad (6.6)$$

Observação: Note que a equação (6.6) é equivalente a pré-multiplicar a equação $Ax = y^\delta$ por

$\gamma^{\frac{1}{2}}$ e, então, iterar como em (6.5). Dada a limitação de γ , sem perda de generalidade, podemos supor que $\|A\| \leq 1$ e iterar como em (6.6).

Nesta seção, a menos que façamos menção em contrário, admitiremos que $\|A\| \leq 1$.

Abaixo, apresentaremos resultados de convergência da iteração de Landweber (6.6). As técnicas utilizadas na demonstração de convergência são relativamente simples e bem conhecidas de métodos iterativos para problemas diretos. Observamos ainda que, obtemos a mesma iteração (6.6) partindo de um chute inicial $x_0 = 0$.

6.1.1 Convergência

Nesta subseção, provaremos que a sequência $\{x_k\}$ de iterados pelo método de Landweber converge para uma solução aproximada do problema (2.1). Para tal, temos que definir o que entendemos por uma solução aproximada. Denotaremos por $x^\dagger := A^\dagger y$, como a solução de quadrados mínimos com norma mínima para o problema (2.1). Aqui A^\dagger denota a pseudo-inversa de A . Veja o Apêndice 3 para maiores detalhes e para uma interpretação geométrica da solução x^\dagger .

Começaremos dando condição necessárias e suficientes para a iteração (6.5) convergir.

Teorema 6.1.1. *Se $y \in \mathcal{D}(A^\dagger)$, então a sequência x_k gerada pela iteração de Landweber (6.5) converge para $x^\dagger = A^\dagger y$ quando $k \rightarrow \infty$. Se $y \notin \mathcal{D}(A^\dagger)$, então $\|x_k\| \rightarrow \infty$ quando $k \rightarrow \infty$.*

Demonstração : De forma recursiva, podemos escrever x_k em (6.5) como

$$x_k = \sum_{j=0}^{k-1} (I - A^* A)^j A^* y. \quad (6.7)$$

Como $y \in \mathcal{D}(A^\dagger)$, então $A^* y = A^* A x^\dagger$. Assim,

$$x^\dagger - x_k = x^\dagger - A^* A \sum_{j=0}^{k-1} (I - A^* A)^j x^\dagger = (I - A^* A)^k x^\dagger. \quad (6.8)$$

Definimos $r_k(\lambda) = (1 - \lambda)^k$. Como, por hipótese $\|A\| \leq 1$, segue que o espectro de $A^* A$ é um subconjunto de $(0, 1]$. Notemos que, para $\lambda \in \Sigma(A^* A) \subset (0, 1]$, $r_k(\lambda)$ converge para zero uniformemente quando $k \rightarrow \infty$. Pelo Teorema da Aplicação Espectral [26, 49, 36],

$$x^\dagger - x_k = r_k(A^* A) x^\dagger. \quad (6.9)$$

Logo, $x_k \xrightarrow{k \rightarrow \infty} x^\dagger$. ■

Exercício 6.1.1. *Preencha os detalhes da demonstração do Teorema 6.1.1.*

Exercício 6.1.2. *Mostre que se $y \in \mathcal{D}(A^\dagger)$, então $A^*y = A^*Ax^\dagger$. Sugestão: use a inversa generalizada. Consulte o Apêndice 3.*

O Teorema 6.1.1 nos ensina que, $\{x_k\}$ gerada pela iteração de Landweber converge para uma solução de quadrados mínimos da equação (2.1) quando $y \in \mathcal{D}(A^\dagger)$. Como, em geral, dados perturbados y^δ são tal que $y^\delta \notin \mathcal{D}(A^\dagger)$, então, ainda do Teorema 6.1.1 sabemos que a sequência x_k^δ diverge. A pergunta é: Qual é o fator de propagação destes erros?

Lema 6.1.1. *Sejam y, y^δ com $\|y - y^\delta\| \leq \delta$ e x_k e x_k^δ obtidos pelas respectivas iterações de Landweber (6.5) e (6.6). Então,*

$$\|x_k - x_k^\delta\| \leq \sqrt{k}\delta, \quad k \geq 0. \quad (6.10)$$

Demonstração : Pela linearidade de A , temos que

$$x_k - x_k^\delta = \sum_{j=0}^{k-1} (I - A^*A)^j A^*(y - y^\delta).$$

Como $\|A\| \leq 1$, segue que $(I - A^*A)$ é um operador semi-definido positivo com $\|I - A^*A\| \leq 1$. Assim,

$$\left\| \sum_{j=0}^{k-1} (I - A^*A)^j A^* \right\|^2 = \left\| \sum_{j=0}^{k-1} (I - A^*A)^j (I - (I - A^*A)^k) \right\| \leq \left\| \sum_{j=0}^{k-1} (I - A^*A)^j \right\| \leq k$$

e o Lema segue. ■

Exercício 6.1.3. *Preencha os detalhes da demonstração do Lema 6.1.1.*

Note que, na presença de erro nos dados temos

$$\|A^\dagger y - x_k^\delta\| \leq \|A^\dagger y - x_k\| + \|x_k - x_k^\delta\|. \quad (6.11)$$

Esta é a estimativa fundamental para a iteração de Landweber.

A estimativa (6.11) nos mostra que o erro total possui duas componentes, um *erro de aproximação* que diminui lentamente e um *erro nos dados* que cresce na ordem de no máximo $\sqrt{k}\delta$. Isso nos leva a seguinte conclusão: Para valores de k pequenos, o erro nos dados é desprezível e a iteração parece convergir para a solução exata $A^\dagger y$. Quando $\sqrt{k}\delta$ atinge a magnitude da ordem do erro de aproximação, o erro propagado nos dados torna-se grande e a aproximação tende a piorar. Veja figura 2.1.

Assim, segue que a propriedade de regularização por métodos iterativos, para problemas mal postos, depende fortemente de um **critério de parada** que detecte a transição entre convergência

e divergência do método. O índice da iteração faz o papel do parâmetro de regularização. Já, o critério de parada faz, o papel da regra de escolha do parâmetro. Conseqüentemente, um critério de parada apropriado deve levar em conta a informação adicional do nível de ruídos δ .

Lema 6.1.2. *A norma do resíduo $y^\delta - Ax_k^\delta$ é sempre monótona não-crescente durante a iteração.*

Demonstração : De fato, pela definição da iteração de Landweber

$$y^\delta - Ax_k^\delta = y^\delta - A(x_{k-1}^\delta + A^*(y^\delta - Ax_{k-1}^\delta)) = (I - A^*A)(y^\delta - Ax_{k-1}^\delta).$$

Como $\|I - A^*A\| \leq 1$, o Lema segue. ■

Por outro lado, se $y^\delta \notin \mathcal{D}(A)$, pelo Teorema 6.1.1, a iteração x_k^δ diverge para infinito. Portanto, um resíduo pequeno não implica que a aproximação para solução é melhor. Veja a estimativa (6.11) e a figura 2.1.

Exercício 6.1.4. *Faça um gráfico comparando os resultados obtidos nos Lemas 6.1.1 e 6.1.2 com a estimativa (6.11). Compare com a figura 2.1.*

Uma alternativa para a escolha do critério de parada é o **princípio da discrepância**: a iteração é parada no índice $k_* = k(\delta, y^\delta)$ quando, pela primeira vez,

$$\|y^\delta - Ax_{k(\delta, y^\delta)}\| \leq \tau\delta, \quad \tau > 2 \text{ fixo.} \quad (6.12)$$

O próximo Teorema garante que, enquanto a discrepância (6.12) não é atingida, a aproximação para a solução não piora.

Teorema 6.1.2. *[Monotonia] Sejam $y \in \mathcal{D}(A)$, x^\dagger a solução de norma mínima de (2.1) e y^δ satisfazendo $\|y - y^\delta\| \leq \delta$. Se (6.12) é satisfeita, então*

$$\|x_{k+1}^\delta - x^\dagger\| \leq \|x_k^\delta - x^\dagger\|. \quad (6.13)$$

Demonstração : Fazendo

$$\begin{aligned} \|x_{k+1}^\delta - x^\dagger\|^2 - \|x_k^\delta - x^\dagger\|^2 &= 2\langle x_k^\delta - x^\dagger, x_{k+1}^\delta - x_k^\delta \rangle + \|x_{k+1}^\delta - x_k^\delta\|^2 \\ &= 2\langle x_k^\delta - x^\dagger, A^*(y^\delta - Ax_k^\delta) \rangle + \langle A^*(y^\delta - Ax_k^\delta), A^*(y^\delta - Ax_k^\delta) \rangle \\ &= 2\langle Ax_k^\delta \pm y^\delta - y, y^\delta - Ax_k^\delta \rangle + \langle y^\delta - Ax_k^\delta, AA^*(y^\delta - Ax_k^\delta) \rangle \\ &= 2\langle y^\delta - y, y^\delta - Ax_k^\delta \rangle - \|Ax_k^\delta - y^\delta\|^2 - \langle y^\delta - Ax_k^\delta, (I - AA^*)(y^\delta - Ax_k^\delta) \rangle. \end{aligned}$$

Como $I - A^*A$ é semi-definido positivo, segue que

$$\|x_{k+1}^\delta - x^\dagger\|^2 - \|x_k^\delta - x^\dagger\|^2 \leq \|Ax_k^\delta - y^\delta\|(2\|y - y^\delta\| - \|Ax_k^\delta - y^\delta\|).$$

Como y^δ satisfaz (2.2) e $k < k_*$, a afirmativa segue. ■

Exercício 6.1.5. *Faça os detalhes na demonstração do Teorema 6.1.2.*

Como comentado anteriormente, no caso de dados corrompidos por ruídos, a iteração de Landweber deve ser parada após uma quantidade finita de passos. O próximo Teorema mostra que o princípio da discrepância implica nessa importante propriedade para a iteração de Landweber.

Teorema 6.1.3. *Seja $\tau > 1$ em (6.12). Então, o princípio de discrepância determina um índice de parada $k_* = k_*(\delta, y^\delta)$, é finito para a iteração de Landweber, com $k(\delta, y^\delta) = \mathcal{O}(\delta^{-2})$.*

Demonstração : Seja $\{x_k^\delta\}$ dada por (6.5). Como no Teorema 6.1.2,

$$\|x^\dagger - x_k^\delta\|^2 - \|x^\dagger - x_{k+1}^\delta\|^2 = \|y - Ax_k^\delta\|^2 - \langle y - Ax_k^\delta, (I - AA^*)(y - Ax_k^\delta) \rangle \geq \|y - Ax_k^\delta\|^2.$$

Somando sobre $k = 1$ até j e levando em conta a monotonia dos resíduos dado pelo Lema 6.1.2, temos

$$\|x^\dagger - x_j^\delta\|^2 - \|x^\dagger - x_{j+1}^\delta\|^2 \geq \sum_{k=1}^j \|y - Ax_k^\delta\|^2 \geq k \|y - Ax_j^\delta\|^2.$$

Indutivamente, podemos escrever $y - Ax_j^\delta = (I - AA^*)^j(y - Ax_0)$, segue que

$$\|(I - AA^*)^j(y - Ax_0)\| = \|y - Ax_j^\delta\| \leq k^{-\frac{1}{2}} \|x^\dagger - x_1^\delta\|.$$

Assim,

$$\|y^\delta - Ax_j^\delta\| = \|(I - AA^*)^j(y^\delta - Ax_0)\| \leq \|(I - AA^*)^j(y^\delta - y)\| + \|(I - AA^*)^j(y - Ax_0)\| \leq \delta k^{-\frac{1}{2}} \|x^\dagger - x_1^\delta\|.$$

Consequentemente, o lado direito é menor que $\tau\delta$, se $k > \frac{\|x^\dagger - x_1\|^2}{(\tau - 1)^2\delta^2}$. Logo, $k(\delta, y^\delta) \leq c\delta^{-2}$, onde c só depende de τ . ■

Para obtermos taxas é preciso fazermos hipóteses sobre a solução x^\dagger . Tais hipóteses são conhecidas como **condições de fonte**. Condições de fonte mais comuns na literatura impõem que x^\dagger pertença a imagem do operador A^* ou pertença a imagem de potências do operador A^*A . Por exemplo, considere a condição de fonte

$$x^\dagger \in \text{Im}((A^*A)^\theta), \quad \theta \in \mathbb{R}_*^+. \quad (6.14)$$

Tal condição impõem que a solução x^\dagger pertença a imagem do operador $(A^*A)^\theta$.

Exercício 6.1.6. *Mostre que o operador $(A^*A)^{\frac{1}{2}}$ está bem definido. O domínio e imagem de*

$(A^*A)^{\frac{1}{2}}$ são subconjuntos de que espaços vetoriais? Interprete geometricamente a condição de fonte (6.14).

Teorema 6.1.4. *Suponha a condição de fonte (6.14) satisfeita para $\theta = 1/2$. Se $y \in \text{Im}(A)$ e o princípio de discrepância (6.12) é válido, então, a iteração de Landweber possui ordem de convergência $k(\delta, y^\delta) = \mathcal{O}(\delta^{-1})$.*

Demonstração : Veja [9, Teorema 6.5]. ■

6.2 Steepest Descent e a Inversa Generalizada

Agora, vamos nos ater no método conhecido como **Steepest Descent** proposto por Cauchy³, por volta de 1847, para resolver sistemas de equações não-lineares. Com a ideia de representar a solução de um sistema de equações não-lineares por meio do mínimo de um funcional não negativo, Cauchy construiu uma sequência, de forma iterativa, que passa do iterado corrente para o seguinte na direção em que o funcional decresce mais rápido. Por essa propriedade, o método de Cauchy ou Steepest Descent também é conhecido como método de Máxima Descida.

Aqui, apresentaremos alguns dos resultados, devido a Nashed, [37], para aproximarmos $A^\dagger y$, onde A é um operador que satisfaz a Hipótese 3.1.1. Nashed aplicou o método de Steepest descent juntando as equações lineares como uma única equação de operador linear $Ax = y$. Antes disso, vamos a algumas considerações iniciais.

Seja $J : \mathcal{H} \rightarrow \mathbb{R}$ um funcional não-negativo. Denotemos por $x^* \in \mathcal{H}$ um ponto tal que

$$J(x^*) = \inf\{J(x) : x \in \mathcal{H}\}. \quad (6.15)$$

Suporemos que J seja Fréchet diferenciável em cada ponto de \mathcal{H} .

Exercício 6.2.1. *Prove que se A é um operador linear e limitado, então $J(x) = \frac{1}{2}\|Ax - y\|^2$ é Fréchet diferenciável e $\nabla J(x) = A^*(Ax - y)$.*

Dado um ponto x_0 , queremos minimizar J andando na direção em que J decresce de forma mais rápida. Assim, devemos escolher uma direção $z \in \mathcal{H}$ tal que a derivada direcional

$$DJ(x_0, z) = \langle z, \nabla J(x_0) \rangle,$$

é a menor possível.

³Augustin Louis Cauchy (1789 -) matemático Francês. Foi quem começou a formular e provar os teoremas do cálculo de maneira rigorosa e assim se tornou um dos pioneiros da análise matemática.

Pela desigualdade de Cauchy-Schwarz⁴,

$$- \|z\| \|\nabla J(x_0)\| \leq \langle z, \nabla J(x_0) \rangle DJ(x_0, z)$$

e a igualdade só acontece se z é um múltiplo positivo de $-\nabla J(x_0)$. Assim, começando em x_0 , a direção de maior descida de J é $z = -\nabla J(x_0)$. Se $\alpha > 0$ é o tamanho do passo na direção $-\nabla J(x_0)$, obtemos uma nova aproximação para x^* por

$$x_1 = x_0 - \alpha_0 \nabla J(x_0).$$

O parâmetro α_0 é escolhido de forma que x_1 minimize J sobre a reta que passa por x_0 na direção $-\nabla J(x_0)$ e, assim, deve satisfazer

$$\frac{d}{d\alpha} J(x_0 - \alpha \nabla J(x_0))|_{\alpha=\alpha_0} = 0.$$

A sequência iterativa $\{x_k\}$ gerada pelo método steepest descent é dada por

$$x_{k+1} = x_k - \alpha_k \nabla J(x_k),$$

onde

$$\frac{d}{d\alpha} J(x_k - \alpha \nabla J(x_k))|_{\alpha=\alpha_k} = 0.$$

Suponha que o operador A satisfaz a Hipótese 3.1.1. Pelo Teorema 3.1.1, a solução de quadrados mínimos de $Ax = y$ é justamente o mínimo do funcional $J(x) = \frac{1}{2} \|Ax - y\|^2$. Desta forma, somos levados a aplicar o método de steepest descent para o funcional J .

Exercício 6.2.2. Defina $r := \nabla J(x) = A^*(Ax - y)$ (que está bem definido pelo exercício 6.2.1). Mostre que o valor ótimo de α é dado por

$$\alpha = \frac{\|r\|^2}{\|Ar\|^2}.$$

⁴Para o caso de somas a desigualdade foi publicada por Augustin Cauchy (1821), enquanto a correspondente desigualdade para integrais foi primeiro estabelecida por Viktor Yakovlevich Bunyakovsky (1859) e redescoberta por Hermann Amandus Schwarz (1888).

Assim, a sequência $\{x_k\} \subset \mathcal{H}_1$ pelo método steepest descent é dada por

$$\begin{aligned}x_{k+1} &= x_k - \alpha_k r_k \\r_k &= A^*(Ax_k - y) \\ \alpha_k &= \frac{\|r_k\|^2}{\|Ar_k\|^2}.\end{aligned}\tag{6.16}$$

Observação: Observamos que se $r_k = 0$, então x_k é uma solução de quadrados mínimos e o método pára em x_k . Se tivermos um critério para garantir que x_k é a solução de menor norma entre todas as soluções de quadrados mínimos, obteremos que $x_k = x^\dagger$.

Exercício 6.2.3. *Mostre que se $Ar_k = 0$, então $r_k = 0$. Assim, a sequência gerada pelo algoritmo (6.16) está bem definida.*

Seguiremos os passos de Nashed, [37] na demonstrações abaixo.

Lema 6.2.1. *Seja A satisfazendo a Hipótese (3.1.1). Então,*

$$\lim_{k \rightarrow \infty} r_k = 0.$$

Demonstração : De

$$J(x_{k+1}) = \frac{1}{2} \|Ax_k - \alpha_k Ar_k - y\|^2 = J(x_n) - \frac{1}{2} \frac{\|r_k\|^2}{\|Ar_k\|^2}.$$

Recursivamente, obtemos que

$$J(x_{k+1}) = J(x_0) - \frac{1}{2} \sum_{j=0}^k \frac{\|r_j\|^2}{\|Ar_j\|^2}.$$

Como $J(x) \geq 0$, $\forall x$, temos que

$$\|A\|^{-2} \sum_{j=0}^{\infty} \|r_j\|^2 \leq \sum_{j=0}^{\infty} \frac{\|r_j\|^2}{\|Ar_j\|^2} \leq 2J(x_0).$$

Portanto, $\|r_k\| \rightarrow 0$ quando $k \rightarrow \infty$. ■

Exercício 6.2.4. *Mostre o Lema 6.2.1 sem a hipótese de A possuir imagem fechada.*

Teorema 6.2.1. *Seja A satisfazendo a Hipótese 3.1.1. Então, a sequência gerada pelo método steepest descent converge para uma solução de quadrados mínimos de $Ax = y$, para qualquer $x_0 \in \mathcal{H}_1$. A sequência converge a x^\dagger se e somente se, $x_0 \in \text{Im}(A^*)$.*

Demonstração : Escrevemos a iteração (6.16) de forma recursiva. Então, temos que

$$x_{k+1} = x_0 - \sum_{j=0}^k \alpha_j r_j. \quad (6.17)$$

Mostraremos que $\{x_n\}$ é uma sequência de Cauchy. De fato $m > n$, então

$$x_m - x_n = - \sum_{j=n}^k \alpha_j r_j \in \text{Im}(A^*) = \mathcal{N}(A)^\perp. \quad (6.18)$$

Portanto, existe $\eta > 0$ tal que

$$\eta^2 \|x_m - x_n\|^2 \leq \langle A^* A(x_m - x_n), x_m - x_n \rangle = \|A(x_m - x_n)\|^2.$$

Assim,

$$\begin{aligned} \langle A^* A(x_m - x_n), x_m - x_n \rangle &\leq (\|A^* A x_m - A^* y\| + \|A^* A x_n - A^* y\|) \|x_m - x_n\| \\ &\leq \eta^{-1} (\|A^* A x_m - A^* y\| + \|A^* A x_n - A^* y\|) \|A(x_m - x_n)\|. \end{aligned}$$

Como $r_j \rightarrow 0$, temos que $r_m - r_n = A^* A(x_m - x_n) \rightarrow 0$ quando $m, n \rightarrow \infty$. Como $\text{Im}(A) = \mathcal{N}(A^*)^\perp$ é fechada, segue do Teorema do Gráfico Fechado [26, 44], que A^* possui inversa limitada quando restrito a $\text{Im}(A) = \mathcal{N}(A^*)^\perp$.

Portanto, $\|A(x_m - x_n)\| \leq M$ e

$$\eta^3 \|x_m - x_n\|^2 \leq M(r_m + r_n) \rightarrow 0 \quad m, n \rightarrow \infty.$$

Logo, $\{x_k\}$ é Cauchy e, assim, converge para algum $u \in \mathcal{H}_1$. Pela continuidade de A , temos

$$\|A^* A u - A^* y\| = \lim_{k \rightarrow \infty} \|A^* A x_k - A^* y\| = \lim_{k \rightarrow \infty} r_k = 0$$

donde concluímos que u é uma solução de quadrados mínimos.

Para finalizar, lembremos que qualquer solução de quadrados mínimos é da forma $x^\dagger \oplus \mathcal{N}(A)$ (Teorema 3.1.1). Como $\text{Im}(A^*) = \mathcal{N}(A)^\perp$, segue que x^\dagger é a única solução de quadrados mínimos na $\text{Im}(A^*)$. Se $x_0 \in \text{Im}(A^*)$, então, de (6.17), segue que $x_k \in \text{Im}(A^*)$ para todo k . Como $\text{Im}(A^*)$ é fechada, $u = x^\dagger$.

Caso $x_0 \notin \text{Im}(A^*)$, então $x_0 = x'_0 + P_{\mathcal{N}(A)} x_0$, onde $x'_0 \in \mathcal{N}(A)^\perp = \text{Im}(A^*)$ e $P_{\mathcal{N}(A)} x_0 \neq 0$. Como

$A^*AP_{\mathcal{N}(A)}x_0 = 0$, de (6.17), temos:

$$x_k = x'_0 - \sum_{j=0}^{k-1} \alpha_j r_j + P_{\mathcal{N}(A)}x_0 \longrightarrow x^\dagger + P_{\mathcal{N}(A)}x_0.$$

Isto completa a demonstração. ■

Exercício 6.2.5. *Faça os detalhes da demonstração do Teorema 6.2.1.*

Exercício 6.2.6. *Suponha que existam constantes $C, M, m > 0$ tais que $m\|x\|^2 \leq \langle A^*Ax, x \rangle \leq M\|x\|^2$, $\forall x \in \text{Im}(A^*)$ e que A satisfaça a Hipótese 3.1.1. Mostre que o método steepest descent possui um limitante de erro dado por*

$$\|x^\dagger + P_{\mathcal{N}(A)}x_0 - x_n\| \leq C \left(\frac{M - m}{M + m} \right)^2.$$

6.3 Métodos tipo Kaczmarz

Dedicaremos esta seção a um método de regularização para resolver problemas inversos que aparecem na forma de sistemas de equações

$$A_i(x) = y_i \quad i \in \{1, \dots, M\}. \quad (6.19)$$

Uma maneira de resolver o problema (6.19) é considerarmos o operador $\mathbb{A} := (A_1, \dots, A_M)$ e $Y = (y_1, \dots, y_M)$ e resolver a equação

$$\mathbb{A}(x) = Y,$$

usando as técnicas que foram apresentadas acima.

A principal ideia de métodos tipo Kaczmarz é resolver o sistema (6.19), de maneira cíclica, onde cada equação do sistema é considerado em separado. Nas próximas subseções apresentaremos o método ART (Algebraic Reconstruction Techniques) e o método de Landweber-Kaczmarz como exemplos de tais estratégias.

6.3.1 Método ART

Uma das técnicas mais usadas em diagnósticos médicos para Tomografia Computadorizada até pouco tempo, é o método ART. Este método representa a ideia de métodos do tipo Kaczmarz de uma maneira simples de entender. Faremos a construção do método iterativo ART para o caso em que $A_i = (a_{i1}, \dots, a_{iN})$, onde $i \in \{1, \dots, M\}$, representa uma linha de um sistema linear escrito na forma (6.19). Desta maneira, a iteração ART usa conceitos básicos de Álgebra Linear.

Iteração: Dado \vec{x}_0 um ponto inicial, projetamos \vec{x}_0 ortogonalmente sobre o hiperplano determinado pela primeira equação do sistema (6.19), isto é, sobre $a_{11}x_1 + \cdots + a_{1N}x_N = y_1$. Ao vetor projeção ortogonal, chamamos de primeiro iterado \vec{x}_1 . De posse do vetor \vec{x}_1 projetamos ortogonalmente sobre a segunda linha do sistema (6.19) e obtemos \vec{x}_2 . Aplicamos este procedimento, sucessivamente, até M , obtemos \vec{x}_M . De maneira cíclica, projetamos $\vec{x}_{mod(M)}$ sobre a equação $mod(M)$ do sistema (6.19). A figura

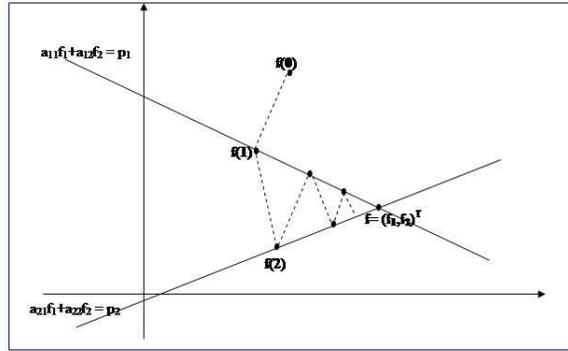


Figura 6.1: Geometria do algoritmo ART.

mostra a iteração ART para um sistema quadrado de ordem 2.

Assim, a iteração é dada por

$$\vec{x}_n = \vec{x}_{n-1} - \left(\frac{\langle A_i, \vec{x}_{n-1} \rangle - y_i}{\|A_i\|^2} \right) (A_i)^T \quad (6.20)$$

Exercício 6.3.1. Considere a reta de equação $a_{i1}x_1 + \cdots + a_{iN}x_N = y_i$ em \mathbb{R}^N . Mostre que a projeção ortogonal de $v \in \mathbb{R}^N$ sobre a reta acima satisfaz a equação (6.20).

Exercício 6.3.2. Mostre que $\langle A_i, \vec{x}_n \rangle = y_i$. Interprete esse resultado geometricamente.

Exercício 6.3.3. Use a iteração (6.20) para obter uma solução do problema de tomografia descrita pela figura 2.6.

Observamos que um dos defeitos do método ART é seu alto custo computacional. Ainda, a taxa de convergência pode ser muito lenta ou até mesmo divergir. Assim, às vezes é necessária a introdução de parâmetros de correção na equação (6.20) de forma a melhorar a performance do método⁵.

⁵Hoje em dia, novos métodos ganharam terreno, [39, 38]. Atualmente, o método Filtered Backprojection é o mais usado em Tomografia Computadorizada. Este consiste em uma estratégia para resolver a Transformada de Radon Inversa.

6.3.2 Método de Landweber-Kaczmarz

O método de Landweber-Kaczmarz consiste em aplicar o método de Landweber para resolver (6.19) de forma cíclica. Isto é, a iteração é definida como

$$x_{k+1}^\delta = x_k^\delta + \omega_k A'_{[k]}(x_k^\delta)(y^\delta - A_{[k]}(x_k^\delta)), \quad (6.21)$$

onde $[k] := k \bmod(M) \in \{0, \dots, M-1\}$ e $i = [k] + 1$. O parâmetro ω_k é definido como

$$\omega_k := \begin{cases} 1 & \text{se } \|A_{[k]}(x_k^\delta) - y_{[k]}^\delta\| > \tau\delta, \\ 0 & \text{caso contrário.} \end{cases}$$

O parâmetro ω_k determina o critério de parada para o método de Landweber-Kaczmarz. A iteração é parada no primeiro índice $k_* = k_*(\delta, y_{[k]}^\delta)$ tal que

$$\omega_{k_*+j} = 0 \quad \text{para } j = \{0, \dots, M-1\}. \quad (6.22)$$

Notamos que, um tal critério de parada dado pela equação (6.22) significa dizer que

$$x_{k_*}^\delta = x_{k_*+1}^\delta = \dots = x_{k_*+M-1}^\delta, \quad (6.23)$$

isto é, k_* é escolhido como o índice que faz com que x_k^δ seja igual em um ciclo.

Convergência para a iteração de Landweber-Kaczmarz segue similar ao feito para o método de Landweber. As hipóteses são similares as feitas para o método de Landweber para cada $A_{[k]}$.

Exercício 6.3.4. *Mostre que o método de Landweber-Kaczmarz (6.21) com o critério de parada (6.22) é um método de regularização. Sugestão: consulte [23].*

Exercício 6.3.5. *Use a iteração de Landweber-Kaczmarz (6.21) para encontrar uma solução do problema de tomografia proposto na Figura 2.6. Compare com o método ART.*

6.4 Algoritmo EM

O algoritmo EM (*Expectation Maximization*) é um método iterativo para maximizar o funcional de máxima verossimelhança (5.33).

A condição necessária de primeira ordem para maximizar o funcional $l(f)$ em (5.33) é que a derivada de primeira ordem seja nula, isto é,

$$\nabla l(f) = A^T \left(\frac{p}{Af} - 1 \right) = 0,$$

onde $\mathbf{1}$ representa o vetor com todas as componentes iguais a 1. Portanto, cada maximizador $f \geq 0$ de l satisfaz

$$fA^T \left(\frac{p}{Af} - 1 \right) = 0. \quad (6.24)$$

Normalizando A de forma que cada coluna tenha soma igual a 1, isto é, de forma que $A^T \mathbf{1} = \mathbf{1}$, temos da equação (6.24)

$$f = fA^T \frac{p}{Af}. \quad (6.25)$$

O algoritmo EM é o método iterativo mais simples para resolver (6.25), i.e.,

$$f_{k+1} = f_k A^T \frac{p}{A f_k}, \quad k = 0, 1, \dots. \quad (6.26)$$

Exercício 6.4.1. *Mostre que o funcional (5.33) é concavo e portanto todo maximizador local é global. Mostre ainda que a condição de otimalidade de primeira ordem (6.24) é também suficiente para determinar o maximizante de do funcional (5.33).*

Exercício 6.4.2. *Mostre que f é um maximizador global do funcional (5.33) se, e somente se satisfaz as condições de Karush-Kuhn-Tucker [22]*

$$\begin{aligned} \frac{\partial l}{\partial f_j}(f) &= 0 \quad \text{para } f_j > 0, \\ \frac{\partial l}{\partial f_j}(f) &\geq 0 \quad \text{para } f_j = 0. \end{aligned} \quad (6.27)$$

Exercício 6.4.3 (Desigualdade de Jensen). *Sejam $\alpha_j, \beta_j \geq 0$ com $\sum_{j=1}^n \alpha_j = 1$. Mostre que*

$$\ln \left(\sum_{j=1}^n \alpha_j \beta_j \right) \geq \sum_{j=1}^n \alpha_j \ln(\beta_j). \quad (6.28)$$

Antes de provarmos que o algoritmo EM em (6.26) converge para um maximizador do funcional (5.33), é necessário alguma preparação. Definimos a distância de Kullback-Leiber

$$KL(x, y) = \sum_{j=1}^n \left(x_j \ln \left(\frac{x_j}{y_j} \right) + y_j - x_j \right), \quad (6.29)$$

para todo $x, y \in \mathbb{R}^n$ com $x, y \geq 0$ e de forma que $t \ln(t) = 0$ se $t = 0$.

Exercício 6.4.4. *Defina a função $\phi_u(t) = t - u \ln(t)$.*

1. *Mostre que, se $u \geq 0$, então $\phi_u(t) \geq 0$ para todo $t \geq 0$.*

2. Mostre que, $\phi_u(t)$ possui um único mínimo em $t = u$.
3. Mostre que, se $\phi_u(t) \rightarrow \phi_u(u)$ para u fixo, então $t \rightarrow u$.
4. Mostre que, $KL(x, y) = \sum_{j=1}^n (\phi_{x_j}(y_j) - \phi_{x_j}(x_j))$
5. Use os itens acima para mostrar que $KL(x, y) \geq 0$. Mostre ainda que $KL(x, \cdot)$ possui um único mínimo em $y = x$. Por fim, conclua que, $y \rightarrow x$ se $KL(x, y) \rightarrow KL(x, x)$

Exercício 6.4.5. Mostre que

$$\sum_{i=1}^n p_i \ln(Af)_i = \sum_{i=1}^n p_i \sum_{j=1}^m \frac{a_{ij} h_j}{(Ah)_i} \left[\ln(a_{ij} f_j) - \ln \left(\frac{a_{ij} f_j}{(Af)_i} \right) \right], \quad (6.30)$$

vale para todo $f, h \in \mathbb{R}^m$ com $f, h > 0$.

Usando os resultados dos exercícios acima, podemos concluir que:

Teorema 6.4.1. *Seja $f_0 > 0$. Então a iteração dada por (6.26) converge para um maximizador do funcional (5.33).*

Demonstração : A demonstração segue os passos de [39, Teorem 5.4]. O primeiro passo é mostrar que o funcional l definido em (5.33) é monótono crescente sobre a iteração (6.26), i.e.,

$$l(f_{k+1}) \geq l(f_k), \quad k = 1, 2, \dots \quad (6.31)$$

Como, para $k > 0$ temos que

$$\sum_{i=1}^n (Af_k)_i = \sum_{i=1}^n p_i, \quad (6.32)$$

temos que

$$l(f_{k+1}) - l(f_k) = \sum_{i=1}^n p_i \ln(Af_{k+1})_i - \sum_{i=1}^n p_i \ln(Af_k)_i.$$

Usando (6.30) com $h = f_k$ e $f = f_k$, f_{k+1} respectivamente, na equação acima, obtemos que

$$\begin{aligned} l(f_{k+1}) - l(f_k) &= \sum_{i=1}^n p_i \sum_{j=1}^m \frac{a_{ij}(f_k)_j}{(Af_k)_i} \left[\ln \frac{(f_{k+1})_j}{(f_k)_j} - \ln \left(\frac{(f_{k+1})_j (Af_k)_i}{(f_k)_j (Af_{k+1})_i} \right) \right] \\ &= \sum_{j=1}^m (f_{k+1})_j \ln \frac{(f_{k+1})_j}{(f_k)_j} - \sum_{i=1}^n p_i \sum_{j=1}^m \frac{a_{ij}(f_k)_j}{(Af_k)_i} \ln \left(\frac{(f_{k+1})_j (Af_k)_i}{(f_k)_j (Af_{k+1})_i} \right). \end{aligned}$$

Aplicando a desigualdade de Jensen no último termo da soma acima, temos que

$$\begin{aligned} l(f_{k+1}) - l(f_k) &\geq \sum_{j=1}^m (f_{k+1})_j \ln \frac{(f_{k+1})_j}{(f_k)_j} - \sum_{i=1}^n p_i \ln \left(\sum_{j=1}^m \frac{a_{ij} (f_{k+1})_j}{(f_k)_j (Af_{k+1})_i} \right) \\ &= KL(f_{k+1}, f_k) - \sum_{i=1}^n p_i \ln(1) \geq 0. \end{aligned}$$

Nosso próximo passo é mostrar que, para cada limite f^* da sequência f_k , temos que

$$KL(f^*, f_{k+1}) \leq KL(f^*, f_k), \quad k = 1, 2, \dots \quad (6.33)$$

Por (6.31) temos que

$$\begin{aligned} KL(f^*, f_k) - KL(f^*, f_{k+1}) &\geq l(f_k) - l(f^*) + KL(f^*, f_k) - KL(f^*, f_{k+1}) \\ &= \sum_{i=1}^n p_i \ln \frac{(Af_k)_i}{(Af^*)_i} + \sum_{j=1}^m f_j^* \ln \frac{(f_{k+1})_j}{(f_k)_j} \\ &= \sum_{j=1}^m f_j^* \sum_{i=1}^n a_{ij} \frac{p_i}{(Af^*)_i} \ln \frac{(Af_k)_i (f_{k+1})_j}{(Af^*)_i (f_k)_j}. \end{aligned}$$

Basta mostrar que o último termo na desigualdade acima é positivo. Note que, como

$$f^* = f^* A^T (p / Af^*),$$

seque que $(A^T (p / Af^*))_j = 1$, para todo $f_j^* > 0$.

Sejam x_j e y_j vetores cujas componetes são

$$(x_j)_i = \frac{a_{ij} p_i / (Af^*)_i}{(A^T (p / Af^*))_j}, \quad (y_j)_i = \frac{a_{ij} p_i / (Af_k)_i}{(A^T (p / Af_k))_j} \quad i = 1, \dots, n.$$

Como $\sum_{i=1}^n (x_j)_i = \sum_{i=1}^n (y_j)_i = 1$, temos que, para todo $f_j^* > 0$

$$\begin{aligned} 0 &\leq \sum_{j=1}^m f_j^* KL(x_j, y_j) = \sum_{j=1}^m f_j^* \sum_{i=1}^n (x_j)_i \ln \frac{(x_j)_i}{(y_j)_i} \\ &= \sum_{j=1}^m f_j^* \sum_{i=1}^n \frac{a_{ij} p_i / (Af^*)_i}{(A^T (p / Af^*))_j} \ln \frac{(Af_k)_i (A^T (p / Af_k))_j}{(Af^*)_i (A^T (p / Af^*))_j} \\ &= \sum_{j=1}^{\tilde{m}} f_j^* \sum_{i=1}^n \frac{a_{ij} p_i}{(Af^*)_i} \ln \frac{(Af_k)_i (f_{k+1})_j}{(Af^*)_i} \sum_{j=1}^{\tilde{m}} f_j^* \sum_{i=1}^n \frac{a_{ij} p_i}{(Af^*)_i} \ln \frac{(Af_k)_i (f_{k+1})_j f_j^*}{(Af^*)_i f_j^* (f_k)_j} \end{aligned}$$

onde a soma sobre \tilde{m} indica que estamos considerando somente os índices j tais que $f_j^* > 0$.

Disto, (6.33) segue.

Para finalizar, seja f^* qualquer ponto de acumulação da sequência (f_k) . Portanto, existe uma subsequência (f_{k_l}) convergindo para f^* . Portanto, $KL(f^*, f_{k_l}) \rightarrow 0$. Como $KL(f^*, f_k)$ é monótona, segue que $KL(f^*, f_k) \rightarrow 0$. Pelo exercício acima, $f_k \rightarrow f^*$.

Para mostrar que f^* é de fato um máximo para o funcional (5.33), basta verificar que f^* satisfaz as condições de KKT. Estas são obviamente satisfeitas para todos os índices j tais que $f_j^* > 0$. Seja agora j tal que $f_j^* = 0$. Portanto

$$(f_{k+1})_j = (f_0)_j \left(A^T \frac{g}{Af_0} \right)_j \cdot \dots \cdot (f_k)_j \left(A^T \frac{g}{Af_k} \right)_j \rightarrow 0$$

Como $\left(A^T \frac{g}{Af_k} \right)_j \rightarrow \left(A^T \frac{g}{Af^*} \right)_j$ temos que, obrigatoriamente, $\left(A^T \frac{g}{Af^*} \right)_j \leq 1$. Portanto, a condição de KKT também é satisfeita para $f_j^* = 0$. ■

6.5 Aplicação: Tomografia Computadorizada

Como aplicação, apresentaremos os resultados obtidos nas Seções 6.3 e 6.2 para o problema de *Tomografia Computadorizada* com uma quantidade limitada de dados, [38, 39].

Seja $L^2(D)$ o espaço de Hilbert de todas as funções quadrado integráveis no disco unitário $D \subset \mathbb{R}^2$ e $\mathcal{H}_2 := \{y : [0, 2] \rightarrow \mathbb{R} : \|y\|^2 := \int_0^2 y(t)t dt < \infty\}$.

Exercício 6.5.1. *Mostre que \mathcal{H}_2 é um espaço de Hilbert.*

Consideraremos o sistema

$$A_j x = y_j, \quad j = 0, \dots, N-1, \quad (6.34)$$

onde cada $A_j := L^2(D) \rightarrow \mathcal{H}_2$, é dada por

$$(A_j x) := \frac{1}{\pi} \int_{S^1} x(\xi_j + t\theta) d\Omega(\theta), \quad t \in [0, 2], \quad (6.35)$$

A equação (6.35) corresponde a uma versão circular da Transformada de Radon que foi introduzida na Seção 2.4.1. Resolver (6.35) é um passo crucial em Tomografia Computadorizada. Consideraremos o caso em que os dados são obtidos por integração numérica. Ainda, supomos que o centro de integração ξ_j corresponde a posição de cada detector.

Nesta aplicação, consideraremos somente o caso especial em que cada

$$\xi_j = \left(\sin\left(\frac{\pi j}{N-1}\right), \cos\left(\frac{\pi j}{N-1}\right) \right)$$

está uniformemente distribuído no semi-círculo $S_+^1 := \{\xi = (\xi^1, \xi^2) \in \partial D : \xi^1 \geq 0\}$. Assim, os dados são medidos em uma única parte da fronteira de D (veja a figura 6.5). Portanto, temos poucos dados. É provado, por exemplo em [39, 38], que certos detalhes de x fora da região de detecção não podem ser reconstruídos. Tais resultados são conhecidos como invisibilidade.

Exercício 6.5.2. *Prove que cada operador A_j é linear, limitado e satisfaz $\|A_j\| \leq 1$. Mostre ainda que o operador adjunto é dado por $(A_j^*y)(\xi) = y \frac{(|\xi_j - \xi|)}{\sqrt{\pi}}$.*

Temos duas possibilidades. A primeira é considerar o operador

$$\begin{aligned} \mathbb{A} &= (A_0, \dots, A_{N-1}) : L^2(D)^N \longrightarrow \mathcal{H}_2^N, \\ \mathbb{A}x &:= \begin{pmatrix} A_0 \\ \vdots \\ A_{N-1} \end{pmatrix} x = \begin{pmatrix} y_0 \\ \vdots \\ y_{N-1} \end{pmatrix} = y. \end{aligned} \quad (6.36)$$

Com essa configuração podemos aplicar a iteração de Landweber linear (6.6), a iteração de Landweber-Kaczmarz ou a iteração de steepest descent (6.16).

A segunda possibilidade é aplicar o método de Landweber-Kaczmarz para o sistema de equações lineares (6.35). Para nosso teste numérico, optamos pela segunda.

A solução x^\dagger é mostrada no lado esquerdo da figura 6.5. Do lado direito da figura 6.5, vemos a figura que consiste da superposição de funções características e um kernel Gaussiano que representa os dados do problema, com $N = 50$ medidas. Os dados $y_j = A_j x^\dagger$ são calculados por integração numérica de forma a adicionar 4% de ruído.

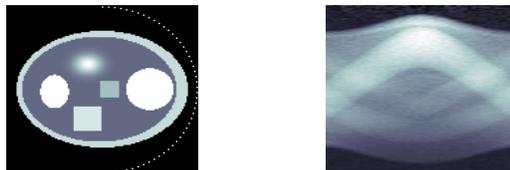


Figura 6.2: Do lado direito a solução e do esquerdo os dados com 4% de ruídos.

A figura 6.5 mostra a solução regularizada $x_{k_*}^\delta$, os métodos Landweber-Kaczmarz e steepest descent com uma estratégia tipo Kaczmarz.

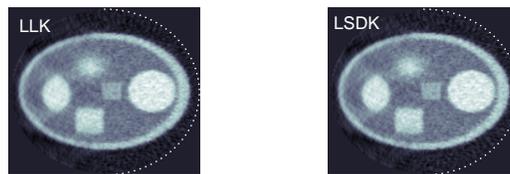


Figura 6.3: Do lado direito a solução aproximada pela regularização de Landweber e do esquerdo por steepest-descent, com estratégias tipo Kaczmarz.

Apêndice A

Definições e resultados

Aqui, postaremos algumas definições importantes e Teoremas de Análise Funcional que são utilizados com frequência nas notas. Não apresentaremos demonstrações, mas, encorajamos o leitor que as desconhece, a dar-lhes a devida atenção, na forma de exercícios complementares. As referências os guiarão a um material muito rico sobre o assunto.

A.1 Definições e Resultados Básicos em Espaços Vetoriais.

Seja \mathcal{H} um espaço vetorial sobre o corpo dos números complexos \mathbb{C} .

Definição A.1.1. Uma *norma* em \mathcal{H} é um funcional

$$\begin{aligned} \|\cdot\| : \mathcal{H} &\longrightarrow \mathbb{R} \\ x &\longmapsto \|x\|, \end{aligned}$$

que, para todo $u, v, w \in \mathcal{H}$ satisfaz:

(i) $\|u\| \geq 0$ e $\|u\| = 0$ se, e somente se $u = 0$.

(ii) $\|\alpha u\| = |\alpha| \|u\|$

(iii) $\|u + w\| \leq \|u + v\| + \|v + w\|$.

Ao par, $(\mathcal{H}, \|\cdot\|)$ chamamos de espaço vetorial normado. Se \mathcal{H} é completo, com relação a norma $\|\cdot\|$, chamamos \mathcal{H} de espaço de Banach.

Definição A.1.2. Um *produto interno* em \mathcal{H} é uma aplicação

$$\begin{aligned} \langle \cdot, \cdot \rangle : \mathcal{H} \times \mathcal{H} &\longrightarrow \mathbb{C} \\ (u, v) &\longmapsto \langle u, v \rangle, \end{aligned}$$

que satisfaz os seguintes axiomas:

$$(i) \langle \alpha u + \beta w, v \rangle = \alpha \langle u, v \rangle + \beta \langle w, v \rangle,$$

$$(ii) \langle u, v \rangle = \overline{\langle v, u \rangle}, \quad \forall u, v, w \in \mathcal{H}, \quad \forall \alpha, \beta \in \mathbb{C}.$$

Exercício A.1.1. Mostre que, em \mathbb{R}^n , a aplicação $\langle u, v \rangle = \sum_{j=1}^n u_j \bar{v}_j$ é um produto interno.

Exercício A.1.2. Seja \mathcal{H} um espaço com produto interno. Defina a aplicação

$$\begin{aligned} \|\cdot\| : \mathcal{H} &\longrightarrow \mathbb{R} \\ x &\longmapsto \|x\| := \sqrt{\langle x, x \rangle}. \end{aligned}$$

Mostre que a aplicação está bem definida. Mostre que $\|\cdot\|$ é uma norma em \mathcal{H} . Assim, $(\mathcal{H}, \|\cdot\|)$ é um espaço vetorial normado. Chama-mos $(\mathcal{H}, \|\cdot\|)$ de um espaço pré-Hilbert.

Definição A.1.3. Seja \mathcal{H} um espaço com produto interno. Se \mathcal{H} é completo como relação a norma $\|x\| := \sqrt{\langle x, x \rangle}$, então \mathcal{H} é dito ser um espaço de Hilbert.

A.1.1 Operadores Lineares

Definição A.1.4. Seja $A : \mathcal{H}_1 \longrightarrow \mathcal{H}_2$ uma aplicação entre os espaços de Hilbert \mathcal{H}_1 e \mathcal{H}_2 . A é dito ser um **operador linear** se

$$A(\alpha x + y) = \alpha Ax + Ay \quad \text{para qualquer } x, y \in \mathcal{H}_1, \quad e \quad \alpha \in \mathbb{C}.$$

O operador A é dito ser **limitado**, se existir $C > 0$ tal que

$$\|Ax\| \leq C\|x\| \quad \forall x \in \mathcal{H}_1.$$

Definição A.1.5. Seja $A : \mathcal{H}_1 \longrightarrow \mathcal{H}_2$ um operador linear limitado. A é dito ser compacto se para qualquer $S \subset \mathcal{H}_1$ subconjunto limitado, $A(S) \subset \mathcal{H}_2$ é um conjunto pré-compacto.

Definição A.1.6. Seja $A : \mathcal{H}_1 \longrightarrow \mathcal{H}_2$ um operador linear limitado. O operador adjunto de Hilbert¹, denotado por A^* , é definido por

$$\langle Ax, y \rangle = \langle x, A^*y \rangle, \quad \forall x \in \mathcal{H}_1, y \in \mathcal{H}_2.$$

O operador $A : \mathcal{H} \longrightarrow \mathcal{H}$ é dito **auto-adjunto** se

$$\langle Ax, y \rangle = \langle x, Ay \rangle, \quad \forall x, y \in \mathcal{H}.$$

¹Existe uma definição de adjunto para operadores entre espaços de Banach. Ainda, a definição acima pode ser estendida a operadores lineares, não necessariamente limitados.

Exercício A.1.3. *Seja A limitado. Mostre que o adjunto de A também é um operador limitado. O que você pode dizer sobre a norma do operador adjunto.*

Exercício A.1.4. *Sejam A um operador compacto e B um operador limitado. Mostre que, se os produtos AB e BA fazem sentido, então ambos os produtos são operadores compactos. Em particular, se A é um operador compacto, então, mostre que A^*A e AA^* são operadores compactos e auto-adjuntos.*

Observação: Operadores compactos e/ou auto-adjuntos possuem propriedades muito importante e já foram amplamente estudados. Veja [26].

Definição A.1.7. *Um operador linear P é dito uma projeção ortogonal se $P^2 = P$ e P é auto-adjunto.*

O próximo Teorema é dito Teorema da Projeção.

Teorema A.1.1. *Seja \mathcal{H} um espaço de Hilbert e M um subespaço fechado de \mathcal{H} . Então, para qualquer $x \in \mathcal{H}$ existe um único $m_0 \in M$ tal que*

$$\|x - m_0\| \leq \|x - m\|, \quad \forall m \in M.$$

Demonstração : Veja [26].

Exercício A.1.5. *Mostre que, se M é um subespaço fechado de um espaço de Hilbert \mathcal{H} , então $\mathcal{H} = M \oplus M^\perp$. Além disso, dado qualquer $x \in \mathcal{H}$, $x = m + m^\perp$ de forma única, onde $m \in M$ e $m^\perp \in M^\perp$. Assim, fica bem definido o operador $P : \mathcal{H} \rightarrow M$ como $Px = m$. Mostre que P é um operador de projeção.*

Teorema A.1.2. *[Banach-Alaoglu] Seja \mathbb{X} um espaço de Banach. Então a bola $B[0, 1]$ é compacta na topologia fraca* de \mathbb{X} . Em particular, se \mathbb{X} é um espaço reflexivo, então a bola $B[0, 1]$ é compacta na topologia fraca.*

Demonstração : Veja textos de Análise Funcional, como [26].

Teorema A.1.3. *Todo espaço de Hilbert é reflexivo.*

Demonstração : Veja [26].

A.1.2 Transformada de Fourier

A transformada de Fourier tem um papel importante na apresentação da transformada de Radon no Capítulo 2. Faremos aqui um breve resumo da Análise de Fourier, que por si só, pode ser considerada como uma área da Matemática.

Denotamos por

$$\mathcal{S}(\mathbb{R}^n) := \{f \in C^\infty(\mathbb{R}^n) : |f|_{k,l} = \sup_{x \in \mathbb{R}^n} |x^k \partial^l f(x)| < \infty, \forall k, l \in \mathbb{N}\}, \quad (1.1)$$

como o conjunto de todas as funções que são infinitamente diferenciáveis em \mathbb{R}^n e cujas derivadas de todas as ordens convergem para zero, mais rápido que qualquer polinômio. Este é um espaço vetorial, o qual não é normado, ao qual chamamos de Espaço de Schwarz.

Para toda função $f \in \mathcal{S}(\mathbb{R}^n)$, a Transformada de Fourier de f é definida como

$$\hat{f}(\xi) = \mathcal{F}(f)(\xi) := (2\pi)^{-n/2} \int_{\mathbb{R}^n} f(x) e^{-ix \cdot \xi} dx. \quad (1.2)$$

Exercício A.1.6. *Seja $f \in \mathcal{S}(\mathbb{R}^n)$. Mostre que $\hat{f} \in \mathcal{S}(\mathbb{R}^n)$.*

Exercício A.1.7. *Mostre que $\mathcal{F} : \mathcal{S}(\mathbb{R}^n) \longrightarrow \mathcal{S}(\mathbb{R}^n)$ é uma transformação linear, limitada.*

Exercício A.1.8. *Seja $f \in \mathcal{S}(\mathbb{R}^n)$. Mostre que se $\mathcal{F}(f)(\xi) = 0$ para todo $\xi \in \mathbb{R}^n$, então $f = 0$. Isto é uma transformação de Fourier é injetiva.*

Portanto, esta transformação possui uma inversa, que chamamos transformada de Fourier inversa definida por

$$\check{f}(x) = \mathcal{F}^{-1}(f)(x) := (2\pi)^{-n/2} \int_{\mathbb{R}^n} f(\xi) e^{\xi \cdot x} d\xi. \quad (1.3)$$

Da definição da transformada de Fourier inversa, segue que

$$\mathcal{F}(\mathcal{F}^{-1}(f)(x)) = \mathcal{F}^{-1}(\mathcal{F}(f)(x)) = f(x) \quad \forall f \in \mathcal{S}(\mathbb{R}^n).$$

Isto é, a transformada de Fourier é um isomorfismo em $\mathcal{S}(\mathbb{R}^n)$.

A transformada de Fourier se estende como um isomorfismo isométrico ao espaço vetorial das funções mensuráveis e quadrado integráveis $L^2(\mathbb{R}^n)$, onde vale a identidade de Parseval

$$\|f\|_{L^2} = \|\hat{f}\|_{L^2}.$$

A transformada de Fourier (1.2), também é bem definida para funções mensuráveis e módulo integráveis, isto é, para $f \in L^1(\mathbb{R}^n)$.

Exercício A.1.9. *Mostre que, para $f \in L^1(\mathbb{R}^n)$, a transformada de Fourier está bem definida e que*

$$\|\hat{f}\|_{L^\infty} \leq c \|f\|_{L^1}.$$

A transformada de Fourier possui as seguintes propriedades

Exercício A.1.10. *Seja $f \in \mathcal{S}(\mathbb{R})$. Prove que:*

1. $f(x - c) = e^{-ic\xi} \hat{f}(\xi)$.

2. $f(cx) = c^{-1} \hat{f}(c^{-1}\xi)$.

3. $\hat{f}'(\xi) = i\xi \hat{f}(\xi)$.

4. $\hat{f}^{(n)}(\xi) = (i\xi)^n \hat{f}(\xi)$.

Bibliografia

- [1] J. Baumeister. *Stable solution of inverse problems*. Advanced Lectures in Mathematics. Friedr. Vieweg & Sohn, Braunschweig, 1987.
- [2] J. Baumeister and A. Leitão. *Topics in inverse problems*. Publicações Matemáticas do IMPA. (IMPA), Rio de Janeiro, 2005. 25º Colóquio Brasileiro de Matemática.
- [3] M. Bertero and P. Boccacci. *Introduction to inverse problems in imaging*. Institute of Physics Publishing, Bristol, 1998.
- [4] T.M. Buzug. *Computer Tomography - From Photom Statistics to Modern Cone-Beam CT*. Springer, Berlin Heidelberg, 2008.
- [5] A. De Cezaro and A. Leitão. Introdução aos Problemas Inversos Lineares. In *IV Bienal da Sociedade Brasileira de Matemática - Maringá - PR*, pages 1–60. online: <http://www.dma.uem.br/bienalsbm>, 2008.
- [6] A. De Cezaro and A. Leitão. Problemas Inversos: Uma Introdução. In *I Colóquio de Matemática da Região Sul - Santa Maria - RS*, pages ii–160. <http://w3.ufsm.br/colmatsul/>, 2010.
- [7] A. De Cezaro and A. Leitão (Orientador). *Métodos de Regularização tipo Level Set para Problemas Inversos*. Universidade Federal de Santa Catarina, Florianópolis -SC, 2006.
- [8] H. W. Engl. *Inverse problems*, volume 8 of *Aportaciones Matemáticas: Textos [Mathematical Contributions: Texts]*. Sociedad Matemática Mexicana, México, 1995. With an introduction in Spanish by Roberto Martínez Villa.
- [9] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1996.
- [10] C. L. Epstein. *Introduction to the Mathematics of Medical Imaging*. Pearson Prentice Hall, Upper Saddle River, NJ, 2003.

- [11] G. H. Golub and C. F. Van Loan. *Matrix computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, third edition, 1996.
- [12] C. W. Groetsch. *Generalized inverses of linear operators: representation and approximation*. Marcel Dekker Inc., New York, 1977. Monographs and Textbooks in Pure and Applied Mathematics, No. 37.
- [13] C. W. Groetsch. *The theory of Tikhonov regularization for Fredholm equations of the first kind*, volume 105 of *Research Notes in Mathematics*. Pitman (Advanced Publishing Program), Boston, MA, 1984.
- [14] C. W. Groetsch. *Inverse problems in the mathematical sciences*. Vieweg Mathematics for Scientists and Engineers. Friedr. Vieweg & Sohn, Braunschweig, 1993.
- [15] C. W. Groetsch. *Inverse problems*. Mathematical Association of America, Washington, DC, 1999. Activities for undergraduates.
- [16] C. W. Groetsch. *Stable approximate evaluation of unbounded operators*, volume 1894 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2007.
- [17] F. A. Grünbaum. Some mathematical problems suggested by limited angle tomography. In *Inverse problems (New York, 1983)*, volume 14 of *SIAM-AMS Proc.*, pages 65–77. Amer. Math. Soc., Providence, RI, 1984.
- [18] J. Hadamard. *Lectures on Cauchy's problem in linear partial differential equations*. Dover Publications, New York, 1953.
- [19] P. C. Hansen. *Rank-deficient and discrete ill-posed problems*. SIAM Monographs on Mathematical Modeling and Computation. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1998. Numerical aspects of linear inversion.
- [20] V. Isakov. *Inverse source problems*, volume 34 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 1990.
- [21] V. Isakov. *Inverse problems for partial differential equations*, volume 127 of *Applied Mathematical Sciences*. Springer, New York, second edition, 2006.
- [22] A. Izmailov and M. Solodov. *Otimização - volume 1: Condições de Otimalidade, Elementos de Análise Convexa e Dualidade*. IMPA, Rio de Janeiro, 2005.
- [23] B. Kaltenbacher, A. Neubauer, and O. Scherzer. *Iterative regularization methods for nonlinear ill-posed problems*, volume 6 of *Radon Series on Computational and Applied Mathematics*. Walter de Gruyter GmbH & Co. KG, Berlin, 2008.

- [24] J. B. Keller. Inverse problems. *Amer. Math. Monthly*, 83(2):107–118, 1976.
- [25] A. Kirsch. *An introduction to the mathematical theory of inverse problems*, volume 120 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1996.
- [26] E. Kreyszig. *Introductory functional analysis with applications*. Wiley Classics Library. John Wiley & Sons Inc., New York, 1989.
- [27] L. Landweber. An iteration formula for Fredholm integral equations of the first kind. *Amer. J. Math.*, 73:615–624, 1951.
- [28] M. M. Lavrentiev, A. V. Avdeev, M. M. Lavrentiev, Jr., and V. I. Priimenko. *Inverse problems of mathematical physics*. Inverse and Ill-posed Problems Series. VSP, Utrecht, 2003.
- [29] E. L. Lima. *Curso de análise. Vol. 1*, volume 1 of *Projeto Euclides [Euclid Project]*. Instituto de Matemática Pura e Aplicada, Rio de Janeiro, 1976.
- [30] E. L. Lima. *Espaços métricos*, volume 4 of *Projeto Euclides [Euclid Project]*. Instituto de Matemática Pura e Aplicada, Rio de Janeiro, 1977.
- [31] E. L. Lima. *Curso de análise. Vol. 2*, volume 13 of *Projeto Euclides [Euclid Project]*. Instituto de Matemática Pura e Aplicada, Rio de Janeiro, 1981.
- [32] A. K. Louis. Inverse problems in medicine. In *Applications of mathematics in industry and technology (Siena, 1988)*, pages 277–287. Teubner, Stuttgart, 1989.
- [33] A. K. Louis. Application of the approximate inverse to 3D X-ray CT and ultrasound tomography. In *Inverse problems in medical imaging and nondestructive testing (Oberwolfach, 1996)*, pages 120–133. Springer, Vienna, 1997.
- [34] A.K. Louis and E.T. Quinto. Local tomographic methods in sonar. In *Surveys on solution methods for inverse problems*, pages 147–154. Springer, Vienna, 2000.
- [35] SIEMENS medical. Computer Tomography: Its History and Technology. pages 1–35, 2009.
- [36] C. Meyer. *Matrix analysis and applied linear algebra*. (SIAM), Philadelphia, PA, 2000.
- [37] M. Z. Nashed. Steepest descent for singular linear operator equations. *SIAM J. Numer. Anal.*, 7:358–362, 1970.
- [38] F. Natterer. *The mathematics of computerized tomography*, volume 32 of *Classics in Applied Mathematics*. (SIAM), Philadelphia, PA, 2001. Reprint of the 1986 original.
- [39] F. Natterer and F. Wübbeling. *Mathematical methods in image reconstruction*. SIAM Monographs on Mathematical Modeling and Computation. (SIAM), Philadelphia, PA, 2001.

- [40] Platão. *A República: Coleção A Obra - Prima de cada Autor*. Martin Claret, São Paulo SP, 2002.
- [41] A. I. Prilepko, D. G. Orlovsky, and I. A. Vasin. *Methods for solving inverse problems in mathematical physics*, volume 231 of *Monographs and Textbooks in Pure and Applied Mathematics*. Marcel Dekker Inc., New York, 2000.
- [42] J. Radon. Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten [on the determination of functions by their integral values along certain manifolds]. In *75 years of Radon transform (Vienna, 1992)*, Conf. Proc. Lecture Notes Math. Phys., IV, pages 324–339. Int. Press, Cambridge, MA, 1994.
- [43] A. G. Ramm and A. I. Katsevich. *The Radon transform and local tomography*. CRC Press, Boca Raton, FL, 1996.
- [44] M Reed and B. Simon. *Methods of modern mathematical physics. I*. Academic Press Inc., New York, second edition, 1980. Functional analysis.
- [45] W. T. Reid. Generalized inverses of differential and integral operators. In *Proc. Sympos. Theory and Application of Generalized Inverses of Matrices (Lubbock, Texas, 1968)*, pages 1–25. Texas Tech. Press, Lubbock, Tex., 1968.
- [46] A. Sá Barreto. Introdução as transformadas de Radon. In *Lectures Notes: Symposium on Spectral and Scattering Theory, Serrambi, de 11 a 22 de Agosto de 2003*, pages 1–22. online: <http://www.math.purdue.edu/~sabarre/papers.html>, 2003.
- [47] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational methods in imaging*, volume 167 of *Applied Mathematical Sciences*. Springer, New York, 2009.
- [48] T. Schuster. *The method of approximate inverse: theory and applications*, volume 1906 of *Lecture Notes in Mathematics*. Springer, Berlin, 2007.
- [49] G. Strang. *Linear algebra and its applications*. Academic Press [Harcourt Brace Jovanovich Publishers], New York, second edition, 1980.
- [50] A. Tarantola. *Inverse problem theory and methods for model parameter estimation*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2005.
- [51] C. R. Vogel. *Computational methods for inverse problems*, volume 23 of *Frontiers in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2002.

-
- [52] Wikipedia. *X-ray computed tomography*. http://en.wikipedia.org/wiki/X-ray_computed_tomography.
- [53] J. P. Zubelli. *An introduction to inverse problems. Examples, methods and questions*. 22^o Colóquio Brasileiro de Matemática. (IMPA), Rio de Janeiro, 1999. Appendix A by the author and Luis Orlando Castellano Pérez.