

Universidade Federal de Santa Catarina  
Curso de Pós-Graduação em Matemática e  
Computação Científica

**TÉCNICAS RESOLUTIVAS PARA  
PROBLEMAS MAL POSTOS: ESTUDO  
COMPARATIVO DE MÉTODOS DE  
REGULARIZAÇÃO**

**Altemir José Borges**

**Orientador: Prof. Dr. Fermín Sinfiriano Viloche  
Bazán**

**Florianópolis  
Junho de 2010**

Universidade Federal de Santa Catarina  
Curso de Pós-Graduação em Matemática e  
Computação Científica

**TÉCNICAS RESOLUTIVAS PARA  
PROBLEMAS MAL POSTOS: ESTUDO  
COMPARATIVO DE MÉTODOS DE  
REGULARIZAÇÃO**

Dissertação apresentada ao Curso de Pós-Graduação em Matemática e Computação Científica, do Centro de Ciências Físicas e Matemáticas da Universidade Federal de Santa Catarina, para a obtenção do grau de Mestre em Matemática, com Área de Concentração em Álgebra Linear Computacional.

Altemir José Borges

Florianópolis

Junho de 2010

**TÉCNICAS RESOLUTIVAS PARA PROBLEMAS MAL  
POSTOS: ESTUDO COMPARATIVO DE MÉTODOS DE  
REGULARIZAÇÃO**

**por**

**Altemir José Borges**

Esta dissertação foi julgada para a obtenção do Título de Mestre em Matemática, Área de Concentração em Álgebra Linear Computacional, e aprovada em sua forma final pelo Curso de Pós-Graduação em Matemática e Computação Científica.

---

**Prof. Dr. Ruy Exel Filho**

Coordenador da Pós-Graduação em Matemática

Comissão Examinadora

---

**Prof. Dr. Fermín Sinfórano Viloche Bázan**

**(UFSC-Orientador)**

---

**Prof. Dr. Marcelo V. W. Zibetti**

---

**Prof. Dr. Mário César Zambaldi**

---

**Prof. Dr. Paulo Rafael Bösing**

**Florianópolis, junho de 2010.**

# Agradecimentos

A Deus, pela força que me transmitiu durante estes anos, tornando possível a conclusão deste trabalho.

Á minha esposa Silvana por ter me incentivado a realizar este curso e também por todo o apoio durante a realização do mesmo.

Aos meus filhos, Luani e Tales, que mesmo distantes sempre estiveram em meu coração.

Aos colegas do Departamento de Matemática da Universidade Tecnológica Federal do Paraná, campus Curitiba, por terem sido solidários no momento de meu afastamento das atividades docentes para realizar este curso.

Ao casal Zenaide Claudino Possas e Marcos Elias Possas, pelo incentivo e amizade.

De maneira especial, ao professor Dr. Fermín Sinforiano Viloche Bázan não somente por sua orientação neste trabalho, mas também pela paciência e conselhos durante estes anos.

A todos os professores do Programa de Pós-Graduação em Matemática e Computação Científica da UFSC, especialmente ao então coordenador, professor Dr. Clóvis Caesar Gonzaga.



# Resumo

Este trabalho apresenta alguns dos métodos diretos e alguns dos iterativos mais comumente utilizados para resolver problemas mal postos discretos, focalizando diferentes estratégias de escolha do parâmetro de regularização. Os métodos diretos aqui abordados são o da GCV, da curva-L, do ponto fixo, da quase-otimalidade e o da discrepância. Os métodos iterativos são o LSQR, GMRES e o RRGMR. Estes métodos são aplicados na resolução dos problemas teste Heat, Baart, Deriv2, Foxgood, Gravity, Laplace, Phillips, Shaw, Tomo e Wing, da literatura. Nos métodos iterativos são adotados os critérios de parada de Morigi e da discrepância.

Para os métodos iterativos, aqui também é apresentado um novo critério de parada baseado no decrescimento da norma do resíduo e no crescimento da norma da solução. Este novo critério desempenhou melhor performance que os critérios de Morigi e da discrepância, na maioria dos problemas testes abordados.



# Abstract

This paper presents some of the direct and some of the iterative methods more commonly used for solving discrete ill-posed problems, focusing on different regularization parameter choices. The direct methods discussed here are the GCV, the curve-L, the fixed point, the quasi-optimality and the discrepancy principle of Morozov. The iterative methods are the LSQR, GMRES and RRGMRES. These methods are applied in solving some test problems from the literature, such as Heat, Baart, Deriv2, Foxgood, Gravity, Laplace, Phillips, Tomo, Shaw and Wing. For the iterative methods are adopted stopping criteria due to Morigi as well as the discrepancy principle.

For the iterative methods is also presented here a new stopping criterion based on the decrease of the norm of the residue and the increase of the norm of the solution. This new criterion has shown to outperform criteria of Morigi and the discrepancy in most test problems addressed.





# Lista de Figuras

2.1	Gráfico da função GCV $\mathcal{G}(\lambda)$ , problema <i>Deriv2</i> com $n = 512$ e erro relativo de 5%. . . . .	30
2.2	Gráfico da Curva-L, para o problema <i>Shaw</i> com $n=128$ e erro relativo de 5%. . . . .	32
2.3	Gráfico da função quase-otimalidade $\mathcal{Q}(\lambda)$ , problema <i>Shaw</i> com $n=512$ e erro relativo de 5%. . . . .	39
2.4	Interpretação geométrica do problema (2.22). . . . .	41
2.5	Gráfico de $\phi_1$ com três pontos fixos. . . . .	53
2.6	Curva-L com 2 cantos, problema <i>Heat</i> . . . . .	54
2.7	Gráfico de $\phi_1$ e $\phi_1^{-1}$ e dois passos da sequência inversa. . . . .	56
3.1	Semi-convergência da série $\sum_{k=0}^n (-1)^{k-1} \frac{(k-1)!}{t^k}$ , com $t = 10$ . . . . .	61
3.2	Semi-convergência do método LSQR, problema <i>Phillips</i> , $n = 64$ e erro relativo de 1%. . . . .	62
3.3	GMRES: minimização da norma do resíduo $r_n$ . . . . .	67

3.4	Erro relativo no passo $k$ : $\ x_k - x_{\text{exato}}\ _2 / \ x_{\text{exato}}\ _2$ para LSQR, problema <i>Heat</i> , $n = 256$ e erro nos dados de 5% (esquerda). Comportamento da sequência $\Psi_k$ definida em (3.24) (direita). . . . .	76
4.1	Solução $x_{\text{exata}}$ para o problema <i>Heat</i> de dimensão $n = 32$ , $\kappa = 1$ . . . . .	79
4.2	Solução $x_{\text{exata}}$ para o problema <i>Baart</i> de dimensão $n = 32$ . . . . .	80
4.3	Solução $x_{\text{exata}}$ para o problema <i>Deriv2</i> de dimensão $n = 32$ , nos 3 casos previstos. . . . .	82
4.4	Solução $x_{\text{exata}}$ para o problema <i>Foxgood</i> de dimensão $n = 32$ . . . . .	83
4.5	Solução $x_{\text{exata}}$ para o problema <i>Gravity</i> de dimensão $n = 32$ . . . . .	85
4.6	Solução $x_{\text{exata}}$ para o problema <i>Llaplace</i> de dimensão $n = 32$ . . . . .	86
4.7	Solução $x_{\text{exata}}$ para o problema <i>Phillips</i> de dimensão $n = 32$ . . . . .	88
4.8	Solução $x_{\text{exata}}$ para o problema <i>Shaw</i> de dimensão $n = 32$ . . . . .	89
4.9	Solução $x_{\text{exata}}$ para o problema <i>Tomo</i> de dimensão $n = 32$ . . . . .	90
4.10	Solução $x_{\text{exata}}$ para o problema <i>Wing</i> de dimensão $n = 32$ , $t_1 = 1/3$ e $t_2 = 2/3$ . . . . .	91
4.11	Solução exata, $x_{\text{exata}}$ , e solução regularizada, $x_\lambda$ , para o problema <i>Heat</i> , com erro relativo de 0.1% em $b$ . . . . .	97
4.12	Solução exata, $x_{\text{exata}}$ , e solução regularizada, $x_\lambda$ , para o problema <i>Gravity</i> , com erro relativo de 1% em $b$ . . . . .	106
4.13	Solução $x_{\text{exata}}$ para o problema <i>tomo</i> de dimensão 625. . . . .	116
4.14	Valores singulares para a matriz $A$ do problema <i>Wing</i> , coeficientes de Fourier e a razão entre eles, com $n=256$ e erro relativo=5% no vetor $b$ . . . . .	121

4.15 Solução exata,  $x_{exata}$ , e solução regularizada,  $x_\lambda$ , para o problema *Wing*, com erro relativo de 0.1% em  $b$ . . . . . 122



# Lista de Tabelas

1	Correspondência entre problemas mal postos e mal condicionados. . . . .	8
4.1	Comparação dos resultados dos métodos da Curva-L, Discrepância, Ponto fixo, Quase otimalidade e GCV, para o problema <i>Heat</i> . . . . .	95
4.2	Comparação dos resultados dos métodos LSQR e RRGMRRES, para o problema <i>Heat</i> . . . . .	96
4.3	Comparação dos resultados dos métodos da Curva-L, Discrepância, Ponto fixo, Quase otimalidade e GCV, para o problema <i>Baart</i> . . . . .	98
4.4	Comparação dos resultados dos métodos LSQR e RRGMRRES, para o problema <i>Baart</i> . . . . .	99
4.5	Comparação dos resultados dos métodos da Curva-L, Discrepância, Ponto fixo, Quase otimalidade e GCV, para o problema <i>Deriv2</i> . . . . .	101

4.6	Comparação dos resultados dos métodos LSQR e RRGM-RES, para o problema <i>Deriv2</i> . . . . .	102
4.7	Comparação dos resultados dos métodos da Curva-L, Discrepância, Ponto fixo, Quase otimalidade e GCV, para o problema <i>Foxgood</i> . . . . .	104
4.8	Comparação dos resultados dos métodos LSQR e RRGM-RES, para o problema <i>Foxgood</i> . . . . .	105
4.9	Comparação dos resultados dos métodos da Curva-L, Discrepância, Ponto fixo, Quase otimalidade e GCV, para o problema <i>Gravity</i> . . . . .	107
4.10	Comparação dos resultados dos métodos LSQR e RRGM-RES, para o problema <i>Gravity</i> . . . . .	108
4.11	Comparação dos resultados dos métodos da Curva-L, Discrepância, Ponto fixo, Quase otimalidade e GCV, para o problema <i>Llaplace</i> . . . . .	110
4.12	Comparação dos resultados dos métodos LSQR e RRGM-RES, para o problema <i>Llaplace</i> . . . . .	111
4.13	Comparação dos resultados dos métodos da Curva-L, Discrepância, Ponto fixo, Quase otimalidade e GCV, para o problema <i>Phillips</i> . . . . .	112
4.14	Comparação dos resultados dos métodos LSQR e RRGM-RES, para o problema <i>Phillips</i> . . . . .	113
4.15	Comparação dos resultados dos métodos da Curva-L, Discrepância, Ponto fixo, Quase otimalidade e GCV, para o problema <i>Shaw</i> . . . . .	114

4.16	Comparação dos resultados dos métodos LSQR e RRGM-RES, para o problema <i>Shaw</i> . . . . .	115
4.17	Comparação dos resultados dos métodos da Curva-L, Discrepância, Ponto fixo, Quase otimalidade e GCV, para o problema <i>Tomo</i> . . . . .	117
4.18	Comparação dos resultados dos métodos LSQR e RRGM-RES, para o problema <i>Tomo</i> . . . . .	118
4.19	Comparação dos resultados dos métodos da Curva-L, Discrepância, Ponto fixo, Quase otimalidade e GCV, para o problema <i>Wing</i> . . . . .	119
4.20	Comparação dos resultados dos métodos LSQR e RRGM-RES, para o problema <i>Wing</i> . . . . .	120
5.1	Resumo dos métodos de regularização mais eficientes por problema teste. . . . .	124
5.2	Comparação dos critérios de parada utilizados no método LSQR por problema teste, com base no erro relativo médio. . . . .	125





# Sumário

<b>Introdução</b>	<b>3</b>
<b>1 Aspectos teóricos</b>	<b>11</b>
1.1 SVD . . . . .	11
1.2 Fatoração QR . . . . .	17
1.3 A condição discreta de Picard . . . . .	21
1.4 Subespaços de Krylov . . . . .	22
1.5 Bidiagonalização de Lanczos . . . . .	23
<b>2 Regularização empregando métodos diretos</b>	<b>25</b>
2.1 Propriedades . . . . .	26
2.2 Regularização direta . . . . .	29
2.2.1 GCV . . . . .	29
2.2.2 Curva-L . . . . .	31
2.2.3 Ponto fixo . . . . .	35
2.2.4 Quase-otimalidade . . . . .	36

2.2.5	Discrepância . . . . .	40
2.3	Concavidade e Convexidade da Curva-L . . . . .	42
2.3.1	Sobre a condição $\phi'_1(\lambda^*) < 1$ e a hipótese $\phi''_1(\lambda^*) \neq 0$	49
2.3.2	Conclusões sobre a concavidade/convexidade da curva-L . . . . .	52
2.3.3	Observações finais sobre o método do ponto fixo . . .	54
<b>3</b>	<b>Regularização iterativa</b>	<b>59</b>
3.1	Introdução . . . . .	60
3.2	LSQR . . . . .	61
3.3	GMRES e RRGMR . . . . .	66
3.4	CrITÉrios de parada para métodos iterativos . . . . .	71
3.5	Um novo critÉrio de parada para problemas discretos mal-postos . . . . .	74
<b>4</b>	<b>Problemas testes e resultados numÉricos</b>	<b>77</b>
4.1	Descrição de problemas teste . . . . .	78
4.1.1	Heat . . . . .	78
4.1.2	Baart . . . . .	79
4.1.3	Deriv2 . . . . .	80
4.1.4	Foxgood . . . . .	81
4.1.5	Gravity . . . . .	84
4.1.6	Llaplace . . . . .	84
4.1.7	Phillips . . . . .	86
4.1.8	Shaw . . . . .	87
4.1.9	Tomo . . . . .	89

4.1.10	Wing . . . . .	90
4.2	Resultados numéricos . . . . .	92
4.2.1	Heat . . . . .	94
4.2.2	Baart . . . . .	98
4.2.3	Deriv2 . . . . .	100
4.2.4	Foxgood . . . . .	103
4.2.5	Gravity . . . . .	106
4.2.6	Llaplace . . . . .	109
4.2.7	Phillips . . . . .	112
4.2.8	Shaw . . . . .	114
4.2.9	Tomo . . . . .	116
4.2.10	Wing . . . . .	118
<b>5</b>	<b>Conclusão</b>	<b>123</b>
	<b>Referências</b>	<b>127</b>



# Abreviaturas e notações

Símbolo	Significado
CG	Gradiente conjugado
GCV	Validação cruzada generalizada
LS	Mínimos quadrados
LSQR	Mínimos quadrados via fatoração QR
SVD	Decomposição em valores singulares
TSVD	SVD truncada
GSVD	SVD generalizada
GMRES	Resíduo mínimo generalizado
RRGMRES	Resíduo mínimo generalizado de posto restrito
$A$	Matriz $m \times n$
$I$	Matriz identidade
$A^T$	Matriz transposta de $A$
$b$	Termo independente de $AX = b$
$e$	Vetor perturbação ou ruído

Símbolo	Significado
$\Sigma$	Matriz diagonal de valores singulares
$diag(\cdot)$	Matriz diagonal
$A^\#$	Inversa regularizada de $A$
$A^\dagger$	Pseudo-inversa de $A$
$\mathcal{N}(A)$	Espaço nulo da matriz $A$
$span(v_i)$	Espaço gerado pelos vetores $v_i$
$x$	Vetor solução
$x_{exata}$	Solução exata
$x^k$	k-ésima solução iterada
$It$	Iteração de parada
$\hat{k}$	Iteração ótima
$ER$	Erro relativo
$ER_o$	Erro relativo ótimo
$\lambda$	Parâmetro de regularização
$NR$	Nível de ruído
MOR	Critério de Morigi
DP	Critério da discrepância
Min_psi	Novo critério de parada
$\ x\ _2$	A norma 2 do vetor $x$
$cond(A)$	Número de condição da matriz $A$
$\mathcal{K}_n(A, b)$	Subespaço de Krylov
$\mathcal{G}(\lambda)$	Função GCV
$\sigma_i$	O i-ésimo valor singular
$f_i$	O i-ésimo fator de filtro
$tr(A)$	Traço da matriz $A$
$a e + b$	$a \times 10^b$





# Introdução

Muitos problemas das áreas tecnológicas ou biológicas podem ser expressos através de equações da forma:

$$\mathcal{A}(x) = g, \tag{1}$$

em que  $\mathcal{A}$  é um operador que relaciona dois espaços normados  $\mathcal{U}$  e  $\mathcal{V}$ ,  $x \in \mathcal{U}$  e  $g \in \mathcal{V}$ .

Os elementos  $\mathcal{A}$ ,  $x$  e  $g$  podem ser, por exemplo, interpretados da seguinte forma: o operador  $\mathcal{A}$  pode representar um modelo matemático associado a algum sistema físico, o vetor  $x$  é então o estado atual desse sistema (também denominado função de entrada ou causa) e o vetor  $g$  é o conjunto de efeitos ou respostas do sistema ou, ainda, função de dados.

Dentre os problemas expressos pelo modelo (1), existem os problemas diretos e os problemas inversos. Os problemas diretos são aqueles em que se conhece o modelo matemático e o estado atual do sistema e deseja-se calcular a resposta desse sistema, isto é, em (1) são conhecidos o operador

$\mathcal{A}$  e a causa  $x$  e tem-se como incógnita o efeito  $g$ . Já os problemas inversos, objeto deste trabalho, são os problemas que consistem em encontrar uma propriedade desconhecida de um objeto, a partir da observação de uma resposta desse objeto [24]. Ou seja, em (1) são conhecidos o efeito  $g$  e o operador  $\mathcal{A}$ , tendo como incógnita a causa  $x$ . Assim, um problema inverso corresponde à inversão do operador  $\mathcal{A}$ , caso este exista.

No estudo de um problema inverso deve ser analisado previamente se a solução caso exista é única, bem como se ela é estável em relação aos dados de entrada. Esta análise é necessária para se saber se o problema em questão é bem ou mal posto e, assim, determinar a forma mais adequada de tratar o problema.

Segundo o matemático francês Jacques Hadamard, 1915, em [27], um problema bem posto é aquele que:

1. tem solução (existência) , isto é, em (1) para cada  $g$  existe  $x$  tal que  $\mathcal{A}x = g$
2. a solução é única (unicidade) , isto é, para cada  $g$  a equação (1) possui uma única solução  $x$ .
3. a solução depende suavemente dos dados (regularidade ou estabilidade) , isto é, a solução  $x$  depende continuamente dos dados de entrada, ou seja, o operador  $\mathcal{A}$  possui uma inversa contínua.

Caso pelo menos uma das três condições anteriores não seja satisfeita o problema é chamado de *mal posto*. Há muitos casos em que um problema inverso gera um problema mal posto. Um exemplo típico de

problema inverso mal posto corresponde as equações integrais de Fredholm de primeira espécie, que podem ser escritas como

$$\int_a^b K(s, t)x(t)dt = g(s), \quad a \leq s \leq b \quad (2)$$

ou

$$\mathcal{A}x = g, \quad (3)$$

em que  $\mathcal{A}$  é um operador integral definido sobre  $L^2[a, b]$  gerado pelo núcleo  $K(s, t) \in L^2([a, b] \times [a, b])$ . Essas equações geralmente produzem um problema mal posto, pois:

- o problema não tem solução em  $L^2(a, b)$  se  $g$  não é contínua;
- tem infinitas soluções caso o núcleo tenha dimensão infinita, como pode ser observado na equação integral  $\int_{-\pi}^{\pi} s \sin(t)f(t)dt = g(s)$ . Neste exemplo a família infinita de funções definidas por  $f_n(t) = \sin(nt)$ ,  $n = 2, 3, \dots$ , pertence ao núcleo do operador, justificando o fato do núcleo  $k(s, t)$  ter dimensão infinita;
- o problema é sempre instável, exceto no caso do núcleo ser degenerado, isto é,  $K(s, t) = \sum_{i=0}^n S_i(s)T_i(t)$ ,  $n < \infty$ , em que as funções  $S_i$  e  $T_i$  são linearmente independentes, pois mesmo que exista o operador inverso  $\mathcal{A}^{-1}$  ele não é obrigatoriamente contínuo, como por exemplo, se  $\mathcal{A}$  é um operador compacto definido sobre um espaço de dimensão infinita [12].

Neste trabalho, serão resolvidos alguns problemas provenientes da discretização de problemas inversos mal postos, conhecidos na literatura como problemas testes. Estes problemas decorrem da discretização da equação

integral de Fredholm ou da equação integral de Volterra. As equações de Volterra de primeira espécie são definidas por:

$$\int_a^s K(s, t)x(t)dt = g(s). \quad (4)$$

Para a resolução destas equações há necessidade de que elas sejam discretizadas. Conforme Pedersen [23], a partir do uso de um método de quadratura como, por exemplo, as fórmulas de quadratura interpoladoras, compostas ou diretas, sendo que esta última é utilizada quando a equação de Volterra possui núcleo regular, a equação (5) passa a ter a forma

$$\int_a^s K(s, t)x(t)dt = \sum_{j=1}^n \omega_j K(s, t_j)\tilde{x}(t_j) = g(s), \quad (5)$$

em que  $t_1, t_2, \dots, t_n$  são as abscissas para a regra de quadratura escolhida e  $\omega_j$  são os pesos correspondentes. O fato da fórmula de quadratura não calcular exatamente a integral, faz com que a incógnita  $x(t)$  seja substituída por  $\tilde{x}(t)$ . Ao utilizar o método de colocação, ou seja, exigindo que a soma iguale o lado direito em uma série de pontos  $s_1, s_2, \dots, s_m$ , chega-se a:

$$\sum_{j=1}^n \omega_j K(s, t_j)\tilde{x}(t_j) = g(s_i) \quad i = 1, \dots, m.$$

Isto resulta num sistema de equações lineares da forma

$$\begin{pmatrix} \omega_1 K(s_1, t_1) & \omega_2 K(s_1, t_2) & \cdots & \omega_n K(s_1, t_n) \\ \omega_1 K(s_2, t_1) & \omega_2 K(s_2, t_2) & \cdots & \omega_n K(s_2, t_n) \\ \vdots & \vdots & \ddots & \vdots \\ \omega_1 K(s_m, t_1) & \omega_2 K(s_m, t_2) & \cdots & \omega_n K(s_m, t_n) \end{pmatrix} \begin{pmatrix} \tilde{x}(t_1) \\ \tilde{x}(t_2) \\ \vdots \\ \tilde{x}(t_n) \end{pmatrix} =$$

$$= \begin{pmatrix} g(s_1) \\ g(s_2) \\ \vdots \\ g(s_m) \end{pmatrix}$$

que pode ser representado matricialmente como

$$Ax = g, \tag{6}$$

em que  $A \in \mathbb{R}^{m \times n}$ ,  $a_{ij} = \omega_j K(s_i, t_j)$ ,  $x \in \mathbb{R}^n$ ,  $x_j = \tilde{x}(s_i)$  e  $g \in \mathbb{R}^m$  com  $g_i = g(s_i)$ .

Uma consequência do problema original ser mal posto é que a matriz  $A$  do problema discretizado é em geral, severamente mal condicionada, isto é, pequenos ruídos no vetor  $g$  em (6), em geral resultantes de imprecisões nos instrumentos de medida usados para obter experimentalmente os dados, produzem grandes alterações na solução  $x$ . Nestes casos a solução quando existe é dita instável. Num sistema mal condicionado, o número de condição da matriz  $A$ , dado por  $cond(A) = \|A\| \|A^{-1}\|$ , é grande, uma vez que é este número de condição que relaciona o erro relativo na solução com o erro relativo nos dados. Assim, a correspondência entre os problemas mal postos e mal condicionados, pode ser resumida de seguinte forma:

Uma maneira intuitiva para resolver um problema mal posto é através da substituição do problema original por um problema associado bem posto. Esta substituição é feita conforme a *teoria de regularização* que emprega os chamados *métodos de regularização*. A idéia básica da regularização é incorporar uma informação adicional ao problema que permita estabilizar o problema e determinar uma solução aproximada e compatível

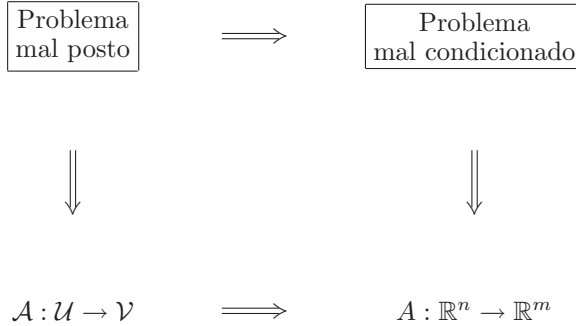


Tabela 1: Correspondência entre problemas mal postos e mal condicionados.

com os dados de entrada. Dentre os métodos de regularização existentes para o problema (6) ou para o problema mais geral

$$x_{ls} = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \|Ax - b\|_2^2, \quad (7)$$

um dos mais usuais é o método de regularização de Tikhonov. A idéia do método é definir a solução regularizada  $x_\lambda$  como o minimizador de uma combinação linear entre a norma do resíduo  $\|Ax - b\|_2$  e um funcional contínuo  $\Omega(x)$ , chamado funcional estabilizador, isto é:

$$x_\lambda = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \|Ax - b\|_2^2 + \lambda^2 \Omega(x), \quad (8)$$

em que  $\lambda > 0$  é o parâmetro de regularização. A forma mais comum do funcional estabilizador é  $\Omega(x) = \|\mathcal{L}x\|_2^2$ , sendo  $\mathcal{L}$  uma versão discreta de algum operador diferencial de primeira ou de segunda ordem. Assim (8) fica da forma:

$$x_\lambda = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \quad \|Ax - b\|_2^2 + \lambda^2 \|x\|_2^2 \quad . \quad (9)$$

Se em (9) tomarmos o operador  $\mathcal{L}$  como sendo o operador identidade  $I$ , tem-se:

$$x_\lambda = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \quad \|Ax - b\|_2^2 + \lambda^2 \|x\|_2^2 \quad . \quad (10)$$

Para obter uma solução significativa para o problema (10) deve-se escolher um parâmetro de regularização adequado, pois será ele que irá realizar um equilíbrio entre a norma do resíduo  $\|Ax - b\|_2$  e a norma da solução  $\|x\|_2$ . Se o valor do parâmetro  $\lambda$  for muito grande, a norma do resíduo será muito sensível a pequenas variações de  $\lambda$ , enquanto a norma da solução será praticamente constante. Já, se  $\lambda$  for um valor muito pequeno, mudanças na norma da solução ocorrerão muito mais rapidamente do que na norma do resíduo.

Assim, o objetivo desta dissertação é resolver, por vários métodos numéricos de regularização, diretos e iterativos, alguns problemas provenientes da discretização de problemas inversos mal postos, conhecidos na literatura como problemas testes. Nos métodos diretos usamos o método de regularização de Tikhonov onde o objetivo é construir soluções após a escolha do parâmetro de regularização. A escolha do parâmetro é feita através dos métodos da validação cruzada generalizada (GCV), da curva-L, do ponto fixo, da quase-otimalidade e da discrepância [13, 14, 15, 17, 25, 30]. Dentre os métodos iterativos usamos LSQR [22, 18], GMRES e RRGMR [6, 7, 21, 13]. Os métodos iterativos necessitam de um critério de parada e nos problemas mal postos a iteração de parada fun-

ciona como um parâmetro de regularização. Aqui são adotados os critérios de parada de Morigi e da discrepância [21, 6], bem como é apresentado um novo critério de parada baseado no decrescimento da norma do resíduo e no crescimento da norma da solução. Mostraremos que o novo critério tem melhor performance que os critérios de Morigi e da discrepância, na maioria dos problemas testes abordados. Já os problemas testes a serem resolvidos são Heat, Baart, Deriv2, Foxgood, Gravity, Llaplace, Phillips, Shaw, Tomo e Wing.

Esta dissertação é composta de 5 capítulos. O capítulo 1 aborda os aspectos teóricos necessários ao desenvolvimento deste trabalho. Nos capítulos 2 e 3 são apresentados cada um dos métodos de regularização a serem utilizados e estão baseados em literaturas recentes tais como [13, 15, 2, 3, 21, 6, 7]. No capítulo 4 cada um dos problemas testes mencionados é descrito e resolvido usando o Pacote Regularization Tools de P. C. Hansen [15] e, também são apresentadas comparações entre as soluções obtidas. Finalmente, no capítulo 5 são apresentadas as conclusões do trabalho.



# Capítulo 1

## Aspectos teóricos

O objetivo deste capítulo é apresentar alguns aspectos teóricos sobre as fatorações SVD, GSVD e QR, a condição discreta de Picard, os subespaços de Krylov e o processo de bidiagonalização de Lanczos, necessários para o entendimento dos métodos de regularização utilizados neste trabalho.

### 1.1 SVD

A SVD (*singular value decomposition*) de uma matriz  $A \in \mathbb{R}^{m \times n}$ , com  $m \geq n$  é uma fatoração da forma

$$A = U\Sigma V^T = \sum_{i=1}^n \sigma_i u_i v_i^T,$$

tais que  $U = [u_1, u_2, \dots, u_n] \in \mathbb{R}^{m \times n}$ ,  $V = [v_1, v_2, \dots, v_n] \in \mathbb{R}^{n \times n}$  são matrizes com colunas ortonormais e, ainda,  $\Sigma \in \mathbb{R}^{n \times n}$  sendo dada por  $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_n)$ , em que  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$  são os valores singulares de  $A$ . As colunas  $u_i$  de  $U$  são chamadas *vetores singulares à esquerda*. As colunas  $v_i$  de  $V$  são chamadas *vetores singulares à direita*. Caso  $m < n$ , a SVD é definida considerando  $A^T$ .

A decomposição em valores singulares (SVD) é uma ferramenta muito útil para a resolução do problema de mínimos quadrados

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2, \quad A \in \mathbb{R}^{m \times n} \text{ com posto } r, \quad b \in \mathbb{R}^m, \quad m \geq n, \quad (1.1)$$

cuja solução de norma mínima é dada por

$$x^\dagger = \sum_{i=1}^r \frac{u_i^T b}{\sigma_i} v_i, \quad (1.2)$$

pois para todo  $x \in \mathbb{R}^n$  e  $V$  ortogonal, pode-se escrever:

$$\|Ax - b\|_2^2 = \|A(VV^T)x - b\|_2^2.$$

Pelo fato de  $U$  também ser ortogonal vem:

$$\|Ax - b\|_2^2 = \|(U^T A V)(V^T x) - U^T b\|_2^2,$$

e usando a SVD de  $A$ , tem-se

$$\|Ax - b\|_2^2 = \|\Sigma(V^T x) - U^T b\|_2^2$$

chamando  $V^T x = \alpha$ , escreve-se:

$$\|Ax - b\|_2^2 = \|\Sigma\alpha - U^T b\|_2^2 = \sum_{i=1}^r (\sigma_i \alpha_i - u_i^T b)^2 + \sum_{i=r+1}^m (u_i^T b)^2.$$

Para se ter uma solução de norma mínima em (1.1), então  $\sum_{i=1}^r (\sigma_i \alpha_i - u_i^T b) = 0$ , isto é,  $\alpha_i = \frac{u_i^T b}{\sigma_i}$  para  $i = 1, 2, \dots, r$ , o que prova (1.2).

Observe que se a matriz  $A^\dagger \in \mathbb{R}^{n \times m}$  for definida por

$$A^\dagger = V \Sigma^\dagger U^T,$$

em que  $\Sigma^\dagger = \text{diag} \left( \frac{1}{\sigma_1}, \dots, \frac{1}{\sigma_r}, 0, \dots, 0 \right)$ , com  $\Sigma^\dagger \in \mathbb{R}^{n \times m}$  e sendo  $r$  o posto da matriz  $A$ , então (1.2) fica escrita como  $x^\dagger = A^\dagger b$ . A matriz  $A^\dagger$  assim definida é chamada de *pseudo-inversa da matriz  $A$* , e satisfaz as quatro condições de Moore-Penrose:

(i)  $AA^\dagger A = A$

(ii)  $A^\dagger AA^\dagger = A^\dagger$

(iii)  $(AA^\dagger)^T = AA^\dagger$

(iv)  $(A^\dagger A)^T = A^\dagger A$

Um conjunto de propriedades importantes da SVD é descrito no seguinte teorema:

**Teorema 1.1.** *Seja  $A = U \Sigma V^T$  a SVD de uma matriz  $A \in \mathbb{R}^{m \times n}$ , com  $m \geq n$ .*

(1) *Suponha que  $A$  é simétrica, com autovalores  $\lambda_i$  e autovetores ortonormais  $u_i$ . Em outras palavras  $A = U \Lambda U^T$  é uma autodecomposição de  $A$ , com  $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ ,  $U = [u_1, u_2, \dots, u_n]$  e  $UU^T = I$ . Então a SVD de  $A$  é  $U \Sigma V^T$ , em que  $\sigma_i = |\lambda_i|$  e  $v_i = \text{sign}(\lambda_i) u_i$  com  $\text{sign}(0) = 1$ .*

- (2) Os autovalores da matriz simétrica  $A^T A$  são  $\sigma_i^2$ . Os vetores singulares à direita  $v_i$  são os correspondentes autovetores ortonormais.
- (3) Os autovalores da matriz simétrica  $AA^T$  são  $\sigma_i^2$  e  $m - n$  zeros. Os vetores singulares à esquerda  $u_i$  são os correspondentes autovetores ortonormais para os autovalores  $\sigma_i^2$ . Qualquer um dos outros  $m - n$  vetores ortogonais correspondem ao autovalor zero.
- (4) Se  $A$  tem posto completo, a solução do problema  $\min_{x \in \mathbb{R}^n} \|Ax - b\|_2$  é  $x = V\Sigma^{-1}U^T b$ .
- (5)  $\|A\|_2 = \sigma_1$ . Se  $A$  é também quadrada e não singular, então  $\|A^{-1}\|_2 = \sigma_n$  e  $\|A\|_2 \cdot \|A^{-1}\|_2 = \frac{\sigma_1}{\sigma_n}$ .
- (6) Suponha  $\sigma_1 \geq \dots \geq \sigma_r > \sigma_{r+1} = \dots = 0$ . Então o posto da matriz  $A$  é igual a  $r$ . Consequentemente, o espaço nulo de  $A$ , isto é, o subespaço dos vetores  $v$  tais que  $Av = 0$ , é o espaço gerado pela  $(r + 1)$ -ésima até a  $n$ -ésima colunas de  $V$ :  $\text{span}(v_{r+1}, \dots, v_n)$ . O espaço posto de  $A$ , o subespaço dos vetores da forma  $Aw$  para todo  $w$ , é o espaço gerado pelas colunas de 1 até  $r$  de  $U$ :  $\text{span}(u_1, \dots, u_r)$ . Resultados análogos podem se obtidos para  $m < n$  [9].

### **Demonstração:**

- (1) A afirmação é verdadeira pela própria definição da SVD
- (2) Sendo  $U\Sigma V^T$  a SVD da matriz  $A$ , vem:

$$A^T A = (U\Sigma V^T)^T U\Sigma V^T = V\Sigma^T U^T U\Sigma V^T = V\Sigma^T \Sigma V^T$$

Esta é uma autodecomposição de  $A^T A$ , com os autovetores sendo

as colunas de  $V$  e os autovalores a diagonal de

$$\Sigma^T \Sigma = \begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_n^2 \end{pmatrix}$$

(3) Seja uma matriz  $\tilde{U} \in \mathbb{R}^{m \times (m-n)}$  de modo  $U, \tilde{U}$  é quadrada e ortogonal. Então

$$\begin{aligned} AA^T &= U \Sigma V^T V \Sigma^T U^T = U \Sigma \Sigma^T U^T = \\ &= U, \tilde{U} \begin{pmatrix} \Sigma \Sigma^T & 0 \\ 0 & 0 \end{pmatrix} U, \tilde{U}^T \end{aligned}$$

é uma autodecomposição de  $AA^T$ .

(4) Sendo  $U \Sigma V^T$  a SVD da matriz  $A$ , pode-se escrever

$$\|Ax - b\|_2^2 = \|U \Sigma V^T x - b\|_2^2 .$$

Como a matriz  $A$  tem posto completo, tem-se que  $\Sigma$  é inversível.

Seja, agora,  $U, \tilde{U}$  quadrada e ortogonal como no ítem anterior.

Então

$$\begin{aligned} \|U \Sigma V^T x - b\|_2^2 &= \begin{pmatrix} U^T \\ \tilde{U}^T \end{pmatrix} (U \Sigma V^T x - b) \begin{matrix} 2 \\ 2 \end{matrix} \\ &= \begin{pmatrix} \Sigma V^T x - U^T b \\ -\tilde{U}^T b \end{pmatrix} \begin{matrix} 2 \\ 2 \end{matrix} \\ &= \|\Sigma V^T x - U^T b\|_2^2 + \|\tilde{U}^T b\|_2^2 \end{aligned}$$

Para que a última expressão seja minimizada, basta fazer a primeira parcela igual a zero, ou seja,  $x = V\Sigma^{-1}U^Tb$ .

(5) Como  $U$  e  $V$  são matrizes ortogonais, então

$$\|A\|_2 = \|U^TAV\|_2 = \|\Sigma\|_2 = \sigma_1.$$

Analogamente,

$$\|A^{-1}\|_2 = \|V^TA^{-1}U\|_2 = \|\Sigma^{-1}\|_2 = \sigma_n^{-1}.$$

(6) Novamente seja uma matriz  $\tilde{U} \in \mathbb{R}^{m \times (m-n)}$ , com  $\hat{U} = U, \tilde{U}$  ortog-

nal. Desde que  $\hat{U}$  e  $V$  sejam não singulares,  $A$  e  $\hat{U}^TAV = \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} =$

$\hat{\Sigma}$  têm o mesmo posto  $r$ , pela hipótese sobre  $\Sigma$ , onde  $\Sigma \in \mathbb{R}^{n \times n}$  e  $0 \in \mathbb{R}^{(m-n) \times n}$ . Além disso,  $v$  está no espaço nulo de  $A$  se e somente se  $V^Tv$  está no espaço nulo de  $\hat{U}^TAV = \hat{\Sigma}$ , uma vez que  $Av = 0$  se e só se  $\hat{U}^TAV(V^Tv) = 0$ . Mas o espaço nulo de  $\hat{\Sigma}$  é gerado pelas colunas  $r + 1$  até  $n$  da matriz identidade de ordem  $n$ , pela própria definição de  $\hat{\Sigma}$ , então o espaço nulo de  $A$  é gerado pelas colunas  $v_{r+1}$  até  $v_n$ . Um argumento similar mostra que o espaço posto de  $A$  é o mesmo daquele gerado pelas  $r$  primeiras colunas de  $I_m$ , ou seja, de  $u_1$  até  $u_r$ . ■

Ainda sobre a *SVD* de uma matriz  $A$ , é importante citar o caso em que o número de condição de  $A$ , dado neste caso por  $cond(A) = \frac{\sigma_1}{\sigma_n}$ , é muito grande. Isso implica que uma ou mais colunas da matriz  $A$  são quase combinação linear de algumas ou de todas as colunas remanescentes. Se a isto for somado o fato de que o vetor termo independente  $b$  pode possuir

ruídos ou perturbações, então a solução (1.2) terá coeficientes  $\frac{u_i^T b}{\sigma_i}$  muito grandes, fazendo com que a solução encontrada possa não ter relação com a verdadeira solução de (1.2).

Para minorar os problemas acima citados, pode-se supor que a matriz  $A$  tenha um agrupamento de valores singulares muito pequenos e também que exista uma separação entre o maior e o menor valor singular, de modo que a série  $(\sigma_i)$ ,  $i = 1, 2, \dots, n$  possa ser truncada em um termo  $(\sigma_s)$  com  $s < n$  e próximo de  $r$ . No entanto, se  $s$  for muito pequeno algumas informações importantes do problema deixarão de ser consideradas.

A estratégia da escolha deste índice  $s$  é a chamada *SVD* truncada ou *TSVD*, cuja solução é dada por

$$x_s = \sum_{i=1}^s \frac{u_i^T b}{\sigma_i} v_i . \quad (1.3)$$

## 1.2 Fatoração QR

A fatoração QR tem como objetivo escrever toda matriz  $A \in \mathbb{R}^{m \times n}$  de modo único como

$$A = QR ,$$

com  $Q \in \mathbb{R}^{m \times m}$ , cujas colunas formam uma base ortonormal para o espaço coluna de  $A$ , e  $R = Q^T A$ ,  $R \in \mathbb{R}^{m \times n}$  é uma matriz triangular superior.

A comparação entre  $m$  e  $n$  gera 3 casos:

- i) Quando  $m = n$ , isto é, quando  $A \in \mathbb{R}^{m \times m}$ , então  $Q \in \mathbb{R}^{m \times m}$  e  $R = Q^T A$ ,  $R \in \mathbb{R}^{m \times m}$ , triangular superior.

ii) Quando  $m > n$ ,  $R$  é da forma

$$R = Q^T A = \begin{pmatrix} \hat{R} \\ 0 \end{pmatrix}$$

em que  $\hat{R} \in \mathbb{R}^{n \times n}$  é triangular superior e  $0$  é uma matriz nula pertencente a  $\mathbb{R}^{(m-n) \times n}$ .

iii) Quando  $m < n$ , a matriz  $R$  terá a forma  $R = Q^T A = \begin{pmatrix} R_{11} & R_{12} \end{pmatrix}$  com  $R_{11} \in \mathbb{R}^{n \times n}$  triangular superior e  $R_{12} \in \mathbb{R}^{n \times (n-m)}$ .

Analisando o caso ii), pode-se constatar que as últimas  $m - n$  colunas da matriz  $Q$ , em  $A = QR$ , estarão multiplicando as linhas nulas da matriz  $R$ . Assim, na fatoração QR reduzida estas  $m - n$  colunas de  $Q$ , bem como as linhas nulas de  $R$ , podem ser removidas, resultando:

$$A = \hat{Q} \hat{R} ,$$

em que  $\hat{Q} \in \mathbb{R}^{m \times n}$  e  $\hat{R} \in \mathbb{R}^{n \times n}$ .

Para a determinação das matrizes  $Q$  e  $R$ , de  $A = QR$ , inicialmente pode-se considerar a matriz  $A \in \mathbb{R}^{m \times n}$  com posto  $n$ . Assim, escrevendo  $A$  e  $Q$  por seus vetores colunas,  $A = \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix}$  e  $Q = \begin{bmatrix} q_1 & q_2 & \dots & q_n \end{bmatrix}$ , respectivamente, determinam-se vetores  $q_i$ ,  $1 \leq i \leq n$ , de modo que o espaço gerado pelos vetores unitários  $q_i$  seja igual ao espaço gerado por  $a_i$ . Isto é conseguido pelo processo de ortogonalização de Gram-Schmidt

$$v_i = a_i - \langle a_i, q_1 \rangle q_1 - \langle a_i, q_2 \rangle q_2 - \dots - \langle a_i, q_{i-1} \rangle q_{i-1} , \quad (1.4)$$

em que  $\langle u, v \rangle$  representa o produto interno canônico em  $\mathbb{R}^n$  entre os vetores  $u$  e  $v$ . Como os vetores  $v_i$  acima gerados não são unitários, é necessário dividi-los por sua norma, resultando:



$$q_i = \frac{v_i}{\|v_i\|}.$$

Reescrevendo  $A = QR$  como

$$\begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix} = \begin{bmatrix} q_1 & q_2 & \dots & q_n \end{bmatrix} \begin{pmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ 0 & r_{22} & \dots & r_{2n} \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & r_{nn} \end{pmatrix} \quad (1.5)$$

chega-se a:

$$\begin{aligned} a_1 &= r_{11}q_1 \\ a_2 &= r_{12}q_1 + r_{22}q_2 \\ a_3 &= r_{13}q_1 + r_{23}q_2 + r_{33}q_3 \\ &\vdots \\ a_n &= r_{1n}q_1 + r_{2n}q_2 + \dots + r_{nn}q_n. \end{aligned} \quad (1.6)$$

Isolando  $q_1, q_2, \dots, q_n$  em (1.6), tem-se

$$\begin{aligned} q_1 &= \frac{a_1}{r_{11}} \\ q_2 &= \frac{a_2 - q_1 r_{12}}{r_{22}} \\ q_3 &= \frac{a_3 - q_1 r_{13} - q_2 r_{23}}{r_{33}} \\ &\vdots \\ q_n &= \frac{a_n - q_1 r_{1n} - \dots - q_{n-1} r_{n-1,n}}{r_{nn}}. \end{aligned} \quad (1.7)$$

Comparando os numeradores de (1.7) com a expressão (1.4), pode-se definir os elementos  $r_{ij}$  como

$$r_{ij} = \langle q_i, a_j \rangle \quad \text{se } i \neq j .$$

Os termos  $r_{jj}$  dos denominadores de (1.7) são definidos por:

$$|r_{jj}| = \left\| a_j - \sum_{i=1}^{j-1} r_{ij} q_i \right\|_2 \quad \text{se } i = j$$

e o sinal de  $r_{jj}$  pode ser escolhido como positivo [28]. Assim, fazendo  $r_{jj} > 0$ , conclui-se a fatoração  $A = QR$  com a matriz  $R$  tendo diagonal positiva.

A fatoração QR, assim como a SVD, também pode ser usada para a determinação da solução do problema (1.1), que é dada em [9] por

$$x = R^{-1} Q^T b .$$

Finalmente, é importante citar dois teoremas, que se encontram demonstrados em [28]:

**Teorema 1.2.** *Toda matriz  $A \in \mathbb{R}^{m \times n}$  ( $m \geq n$ ) tem uma fatoração QR.*

**Teorema 1.3.** *Toda matriz  $A \in \mathbb{R}^{m \times n}$  ( $m \geq n$ ) de posto completo tem uma única fatoração QR reduzida  $A = \hat{Q} \hat{R}$  com  $r_{jj} > 0$ .*

### 1.3 A condição discreta de Picard

A existência da condição de Picard para as equações integrais de primeira espécie é similar à alternativa de Fredholm para as equações de segunda espécie. Segundo [12] a condição de Picard envolve os autovalores do operador compacto auto-adjunto  $K^*K$ , em  $Kf = g$ . Observe que se  $\sigma$  é um autovalor de  $K^*K$  e  $x$  é uma autofunção com norma igual a 1, tem-se  $\sigma = \langle \sigma x, x \rangle = \langle K^*Kx, x \rangle = \|Kx\|^2 \geq 0$ , em que  $\langle \cdot, \cdot \rangle$  representa o produto interno usual em  $L^2[a, b]$ .

Portanto os autovalores não nulos de  $K^*K$  são todos positivos e podem ser enumerados como  $\sigma_1^2, \sigma_2^2, \dots$ , conforme a seção 1.1. Se  $v_1, v_2, \dots$  for a correspondente sequência de autofunções e se definirmos  $u_i$  por  $u_i = \frac{Kv_i}{\sigma_i}$ , então pode ser estabelecida a condição de Picard como se segue.

*Condição de Picard:* A fim de que exista uma solução quadrado integrável  $f$  para a equação integral  $\int_a^b K(s, t)f(t)dt = g(s)$ ,  $c \leq s \leq d$ , o lado direito  $g$  deve satisfazer

$$\sum_{i=1}^{\infty} \left( \frac{\langle u_i, g \rangle}{\sigma_i} \right)^2 < \infty. \quad (1.8)$$

Assim, a partir de um determinado índice  $i$  do somatório (1.8), os coeficientes de Fourier  $\langle u_i, g \rangle$  devem diminuir mais rapidamente do que os valores singulares  $\sigma_i$  a fim de que a solução quadrado integrável exista. A expressão (1.8) é sempre verificada para o caso finito, mas mesmo assim, a relação entre o decaimento dos valores singulares da matriz  $A$  e o decaimento dos coeficientes de Fourier  $\langle u_i, g \rangle$  é importante, pois determinarão a qualidade da solução regularizada.

No caso de um problema de mínimos quadrados, cuja solução é dada

em (1.2), se o módulo dos coeficientes de Fourier ,  $|u_i^T b|$ , decaírem mais rápido que os valores singulares  $\sigma_i$ , então a solução regularizada  $x_\lambda$  preserva aproximadamente as mesmas propriedades da solução exata  $x_{exata}$  [13].

Hansen em [13] caracteriza a *condição discreta de Picard* exigindo que os coeficientes de Fourier  $|u_i^T \hat{b}|$  decaiam em média mais rápido que os correspondentes valores singulares generalizados numericamente não nulos  $\gamma_i$ , em que  $\hat{b}$  corresponde ao vetor de dados livre de erros.

## 1.4 Subespaços de Krylov

Segundo Trefethen em [28], dada uma matriz  $A \in \mathbb{R}^{m \times m}$  e um vetor  $b \in \mathbb{R}^m$ , a sequência de Krylov associada a esta matriz  $A$  e ao vetor  $b$ , é o conjunto de vetores  $b, Ab, A^2b, A^3b, \dots$ . O correspondente subespaço de Krylov, denotado por  $\mathcal{K}_n(A, b)$ , é o espaço gerado pelos vetores da sequência de Krylov, ou seja,

$$\mathcal{K}_n(A, b) = \text{span}\{b, Ab, A^2b, A^3b, \dots, A^{n-1}b\} \text{ ,}$$

em que  $n$  é um inteiro positivo.

Ainda, a matriz de Krylov para as matrizes  $A \in \mathbb{R}^{m \times m}$  e  $b \in \mathbb{R}^m$ , denotada por  $K_n$ , é definida como

$$K_n = \begin{bmatrix} b & Ab & \dots & A^{n-1}b \end{bmatrix} \text{ .} \tag{1.9}$$

## 1.5 Bidiagonalização de Lanczos

Como será visto em 2.3.1, o processo de bidiagonalização de Lanczos é um dos fundamentos necessários para o desenvolvimento do método LSQR.

O objetivo do processo de bidiagonalização de Lanczos é escrever uma dada matriz  $A$ , na forma  $A = UBV^T$ , com  $U$  e  $V$  matrizes ortogonais e  $B$  uma matriz bidiagonal superior ou inferior. Aqui será abordada apenas a determinação de uma matriz bidiagonal inferior.

Assim, seja  $A \in \mathbb{R}^{m \times n}$  com  $m \geq n$  tal que

$$U^T AV = B, \quad (1.10)$$

com  $U \in \mathbb{R}^{m \times (n+1)}$  e  $V \in \mathbb{R}^{n \times n}$  sendo matrizes ortogonais e, ainda,  $B \in \mathbb{R}^{(n+1) \times n}$  uma matriz bidiagonal inferior da forma

$$B_{(n+1) \times n} = \begin{pmatrix} \alpha_1 & & & & & \\ \beta_2 & \alpha_2 & & & & \\ & \beta_3 & \ddots & & & \\ & & \ddots & \alpha_n & & \\ & & & \beta_{n+1} & & \end{pmatrix}.$$

A partir de (1.10), levando-se em conta que  $U$  e  $V$  são ortogonais, isto é,  $U^T U = I_{n+1}$ ,  $U U^T = I_m$  e  $V^T V = V V^T = I_n$ , pode-se escrever:

$$AV = UB. \quad (1.11)$$

Tomando a transposta em ambos os membros, tem-se:

$$V^T A^T = B^T U^T.$$

Usando novamente a ortogonalidade de  $U$  e  $V$ , obtém-se:

$$A^T U = V B^T . \quad (1.12)$$

A expressão anterior, na forma matricial, fica:

$$A^T [u_1 \ u_2 \ \cdots \ u_{n+1}] = [v_1 \ v_2 \ \cdots \ v_n] \begin{pmatrix} \alpha_1 & \beta_2 & & & & \\ & \alpha_2 & \beta_3 & & & \\ & & \ddots & \ddots & & \\ & & & & \alpha_n & \beta_{n+1} \end{pmatrix} . \quad (1.13)$$

Igualando as  $j$ -ésimas colunas de ambos os membros de (1.13), chega-se a:

$$A^T u_i = \alpha_1 v_1$$

e

$$A^T u_j = \beta_j v_{j-1} + \alpha_j v_j \quad \text{para } j = 2, \dots, n. \quad (1.14)$$

Analogamente, pode-se escrever (1.11) como:

$$A v_j = \beta_{j+1} u_{j+1} + \alpha_j u_j \quad (1.15)$$

Das expressões (1.14) e (1.15) escreve-se:

$$A^T u_j - \beta_j v_{j-1} = \alpha_j v_j = s_j, \quad \alpha_j = \|s_j\|, \quad v_j = \frac{s_j}{\alpha_j} \quad (1.16)$$

$$A v_j - \alpha_j u_j = \beta_{j+1} u_{j+1} = t_j, \quad \beta_{j+1} = \|t_j\|, \quad u_{j+1} = \frac{t_j}{\beta_{j+1}} . \quad (1.17)$$

O processo iterativo para a obtenção da matriz bidiagonal inferior  $B$  é realizado escolhendo um vetor unitário  $u_1 \in \mathbb{R}^m$ , tomando  $v_0 = 0$  e utilizando as iterações dadas em (1.16) e (1.17) com  $1 \leq j \leq n$ .

## Capítulo 2

# Regularização empregando métodos diretos

Este capítulo apresenta os métodos GCV, curva-L, ponto fixo, quase-otimalidade e discrepância para a escolha do parâmetro de regularização  $\lambda$  do método de regularização de Tikhonov dado em (10). Nesses métodos,  $\lambda$  é obtido diretamente da função que caracteriza cada método. Inicialmente são estabelecidas algumas propriedades sobre a norma do resíduo e a norma da solução em (10), necessárias ao desenvolvimento do assunto.

## 2.1 Propriedades

Para o desenvolvimento dos métodos da curva-L e do ponto fixo é necessário o conhecimento de algumas propriedades referentes a norma do resíduo,  $\|Ax - b\|_2$ , e a norma da solução,  $\|x\|_2$ , que agora serão mostradas.

No problema de minimização

$$x_\lambda = \operatorname{argmin}_{x \in \mathbb{R}^n} \|Ax - b\|_2^2 + \lambda^2 \|x\|_2^2 \quad , \quad (2.1)$$

denotaremos

$$u(\lambda) = \|Ax_\lambda - b\|_2^2 \quad (2.2)$$

e

$$v(\lambda) = \|x_\lambda\|_2^2 \quad . \quad (2.3)$$

A solução regularizada  $x_\lambda$  da expressão (2.1) satisfaz as equações normais regularizadas

$$(A^T A + \lambda^2 I)x = A^T b \quad , \quad (2.4)$$

usando a SVD da matriz  $A$  segue que

$$x_\lambda = \sum_{i=1}^r \frac{\sigma_i b_i}{\sigma_i^2 + \lambda^2} v_i \quad , \quad (2.5)$$

onde  $b_i = u_i^T b$ .

Tendo em vista que o parâmetro de regularização  $\lambda$  é positivo e se encontra no denominador, o cálculo de  $x_\lambda$  torna-se menos mal condicionado do que o cálculo de  $x$ , isto é, uma influência de erros correspondentes aos menores valores singulares torna-se muito menor.



Como  $\|x_\lambda\|^2 = x_\lambda^T x_\lambda$  e o  $v_i$  são ortonormais obtém-se

$$v(\lambda) = \|x_\lambda\|^2 = \sum_{i=1}^r \frac{\sigma_i^2 b_i^2}{(\sigma_i^2 + \lambda^2)^2}. \quad (2.6)$$

Se forem usados os fatores de filtro de Tikhonov,  $f_1, f_2, \dots, f_n$ , que dependem de  $\sigma_i$  e de  $\lambda$ , que são encontrados em [13] e dados por

$$f_i = \frac{\sigma_i^2}{\sigma_i^2 + \lambda^2} \simeq \begin{cases} 1, & \sigma_i \gg \lambda \\ \sigma_i^2/\lambda^2, & \sigma_i \ll \lambda \end{cases}, \quad (2.7)$$

a expressão (2.6) torna-se

$$\|x_\lambda\|^2 = \sum_{i=1}^r f_i^2 \left( \frac{u_i^T b}{\sigma_i} \right)^2.$$

Além disso, denotando  $b_\perp = \sum_{i=r+1}^m b_i u_i$ , tem-se

$$Ax_\lambda - b = - \sum_{i=1}^r \frac{\lambda_i^2 b_i}{\sigma_i^2 + \lambda^2} u_i + b_\perp.$$

Assim

$$u(\lambda) = \|Ax_\lambda - b\|^2 = \sum_{i=1}^r \frac{\lambda_i^4 b_i^2}{(\sigma_i^2 + \lambda^2)^2} u_i + \|b_\perp\|^2. \quad (2.8)$$

Com os fatores de filtro, (2.8) fica

$$\|Ax_\lambda - b\|^2 = \sum_{i=1}^r (1 - f_i) u_i^T b^2 + \|b_\perp\|^2.$$

**Teorema 2.1.** *Sejam  $u(\lambda)$  e  $v(\lambda)$ , respectivamente, a norma do resíduo e a norma da solução regularizada, definidos em (2.2) e (2.3), então:*

- a)  $u(\lambda)$  é estritamente crescente;
- b)  $v(\lambda)$  é estritamente decrescente;
- c)  $\frac{dv}{du} = -\frac{1}{\lambda^2}$ .

**Demonstração:** Derivando (2.8) em relação ao parâmetro  $\lambda$ , tem-se

$$\frac{du}{d\lambda} = \sum_{i=1}^r \frac{(\sigma_i^2 + \lambda^2)^2 4\lambda^3 b_i^2 - \lambda^4 b_i^2 2(\sigma_i^2 + \lambda^2) 2\lambda}{(\sigma_i^2 + \lambda^2)^4}$$

$$\frac{du}{d\lambda} = 4\lambda^3 \sum_{i=1}^r \frac{b_i^2(\sigma_i^2 + \lambda^2) - \lambda^2 b_i^2}{(\sigma_i^2 + \lambda^2)^3}.$$

Simplificando, tem-se:

$$\frac{du}{d\lambda} = 4\lambda^3 \sum_{i=1}^r \frac{\sigma_i^2 b_i^2}{(\sigma_i^2 + \lambda^2)^3}. \quad (2.9)$$

Como (2.9) é sempre positiva, conclui-se que a função  $u(\lambda)$  é estritamente crescente.

Agora, derivando (2.6) em relação ao parâmetro  $\lambda$ , vem:

$$\frac{dv}{d\lambda} = \sum_{i=1}^r \frac{(-2)\sigma_i^2 b_i^2 (2\lambda)}{(\sigma_i^2 + \lambda^2)^3}$$

$$\frac{dv}{d\lambda} = -4\lambda \sum_{i=1}^r \frac{\sigma_i^2 b_i^2}{(\sigma_i^2 + \lambda^2)^3}. \quad (2.10)$$

Como (2.10) é sempre negativa, conclui-se que  $v(\lambda)$  é estritamente decrescente.

A demonstração de c) é feita aplicando a regra da cadeia em (2.10):

$$\frac{dv}{du} = \frac{dv}{d\lambda} \frac{d\lambda}{du}$$

$$\frac{dv}{du} = -4\lambda \sum_{i=1}^r \frac{\sigma_i^2 b_i^2}{(\sigma_i^2 + \lambda^2)^3} \cdot \frac{1}{4\lambda^3} \sum_{i=1}^r \frac{(\sigma_i^2 + \lambda^2)^3}{\sigma_i^2 b_i^2},$$

que simplificada resulta em

$$\frac{dv}{du} = -\frac{1}{\lambda^2} < 0 . \quad (2.11)$$

■

## 2.2 Regularização direta

### 2.2.1 GCV

O método GCV (*Generalized cross-validation*), desenvolvido por Golub, Heat e Wahba [13], é um método bem usual para a escolha do parâmetro de regularização. Este método baseia-se em considerações estatísticas, com a idéia que um bom parâmetro de regularização deverá prever a falta ou a retirada de dados de entrada, ou seja, modificações no vetor  $b$ . Por exemplo, se uma componente arbitrária  $b_i$  do vetor  $b$  não é considerado, então a correspondente solução regularizada deve prever esta condição, e a escolha do parâmetro de regularização deve ser independente de uma transformação ortogonal de  $b$ , incluindo permutações dos elementos de  $b$ .

A GCV é um método que procura minimizar o erro quadrático médio  $\|Ax_\lambda - \hat{b}\|_2^2$ . Mas, como  $\hat{b}$  é desconhecido, no método GCV trabalha-se com a função GCV

$$\mathcal{G}(\lambda) = \frac{\|Ax_\lambda - b\|_2^2}{\text{tr}(I_m - A.A^\#)^2} ,$$

em que

$$\text{tr}(I_m - A.A^\#) = m - (n - p) - \sum_{i=1}^p f_i$$

e  $A^\#$ , chamada de *inversa regularizada*, é tal que a solução regularizada  $x_\lambda$  pode ser escrita como

$$x_\lambda = A^\# b, \text{ em que } A^\# = (A^T A + \lambda^2 I)^{-1} A^T$$

Assim o parâmetro de regularização proposto pelo método é aquele que minimiza a função  $\mathcal{G}(\lambda)$ .

O gráfico da função GCV,  $\mathcal{G}(\lambda)$ , para o problema teste *Deriv2* quando  $n = 512$  e erro relativo de 5% no vetor  $b$ , está exibido na figura 2.1.

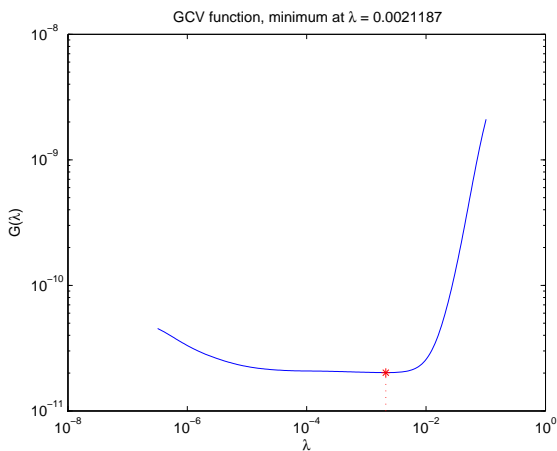


Figura 2.1: Gráfico da função GCV  $\mathcal{G}(\lambda)$ , problema *Deriv2* com  $n = 512$  e erro relativo de 5%.

Em particular, quando  $x_{reg}$  puder ser escrito em termos da SVD da matriz  $A$  e do conjunto de fatores de filtro  $f_1, f_2, \dots, f_n$ , a matriz  $A^\#$  é

escrita como

$$A^\# = VF\Sigma^\dagger U^T, \quad F = \text{diag}(f_1, \dots, f_n).$$

O ponto de mínimo da função  $\mathcal{G}(\lambda)$ , em alguns casos, pode ser difícil de determinar, tendo em vista o fato que o gráfico  $\mathcal{G}(\lambda)$  possui um trecho quase horizontal, conforme pode ser visto na figura 2.1.

## 2.2.2 Curva-L

Sendo  $u(\lambda) = \|Ax_\lambda - b\|_2^2$  e  $v(\lambda) = \|x_\lambda\|_2^2$  pode-se definir a curva-L numa escala  $f$  como o conjunto de pontos

$$L_f = f(u(\lambda), v(\lambda)), \quad (2.12)$$

em que  $f = \ln$ . Isto é, a curva-L é um gráfico bi-log (log-log) da norma do resíduo  $\|Ax - b\|_2^2$  pela norma do termo de regularização  $\|x\|_2^2$ . O nome deste método vem da forma do gráfico de  $L_f$  definido em (2.12) que quando plotado em escala bi-log possui, em geral, partes vertical e horizontal bem definidas [25].

Na figura 2.2 tem-se o gráfico da curva-L, para o problema *Shaw* com  $n = 128$  e um erro relativo aplicado no termo independente  $b$  igual a 5%.

O método da curva-L para determinação do parâmetro de regularização  $\lambda$  está baseado na observação que na parte vertical da curva-L, pequenas mudanças em  $\lambda$  correspondem à variações muito grandes da solução regularizada e à pouquíssimas mudanças na norma do resíduo, enquanto que na parte horizontal, grandes valores de  $\lambda$  correspondem à lentas variações na norma da solução regularizada e à grandes alterações na norma do resíduo.

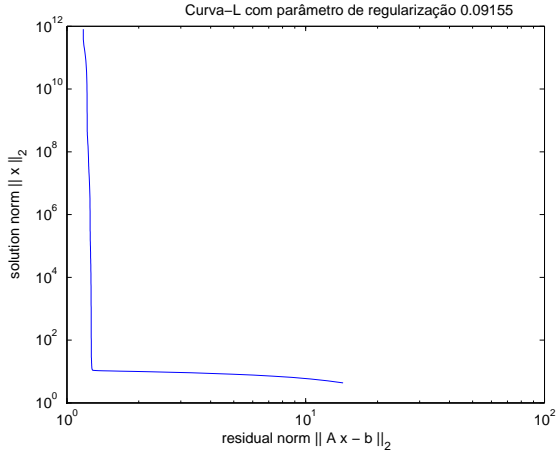


Figura 2.2: Gráfico da Curva-L, para o problema *Shaw* com  $n=128$  e erro relativo de 5%.

Assim uma razoável solução regularizada pode estar associada a um parâmetro na vizinhança do “canto” da curva-L. Para termos a localização desse canto é necessário conhecer as regiões onde a curva-L é côncava e onde é convexa.

Em [25] é observado que:

- (i) A curva-L permanece decrescente para qualquer escala  $f$  que seja diferenciável e estritamente monótona. Empregando a notação  $\eta = f(v(\lambda))$  e  $\xi = f(u(\lambda))$ , tem-se:

$$\frac{d\eta}{d\xi} = \frac{f'(v) \cdot dv/d\lambda}{f'(u) \cdot du/d\lambda} = \frac{-1}{\lambda^2} \cdot \frac{f'(v)}{f'(u)}, \quad (2.13)$$

$$\text{em que } f'(v) = \frac{df}{dv} \text{ e } f'(u) = \frac{df}{du}.$$

(ii) A mudança da escala  $f$  não preserva a convexidade da curva-L.

Com respeito à concavidade da curva-L, temos o seguinte resultado obtido por Regińska [25].

**Lema 2.2.** *Se  $b_{\perp} = 0$ , a curva-L, em escala logarítmica, é estritamente côncava para  $\lambda \in (0, \sigma_r)$ , em que  $\sigma_r$  é o menor valor singular de  $A$ , e para  $\lambda \in (\sigma_1, \infty)$ , sendo  $\sigma_1$  o maior valor singular de  $A$ . Se  $b_{\perp} \neq 0$ , então a curva-L é convexa para  $\lambda \in (0, \epsilon)$  em que  $\epsilon$  é suficientemente pequeno e permanece côncava para  $\lambda \in (\sigma_1, \infty)$ .*

**Demonstração:** Calculando a derivada segunda de  $\eta$  em relação a  $\xi$  e tomando  $f = \ln$  tem-se

$$\begin{aligned} \frac{d^2\eta}{d\xi^2} &= \frac{d \, d\eta/d\xi}{d\xi} = -\frac{d \, \frac{1}{\lambda^2} \cdot \frac{e^\xi}{e^\eta}}{d\xi} \\ \frac{d^2\eta}{d\xi^2} &= -\frac{\lambda^2 e^\eta e^\xi \cdot 1 - e^\xi \, 2\lambda\lambda' e^\eta + \lambda^2 e^\eta \, - u/\lambda^2 v}{\lambda^4 e^{2\eta}} \\ \frac{d^2\eta}{d\xi^2} &= -\frac{\lambda^2 uv - u(2\lambda v\lambda' - u)}{\lambda^4 v^2} \\ \frac{d^2\eta}{d\xi^2} &= -\frac{u}{\lambda^4 v^2} \lambda^2 v + u - 2\lambda v \frac{1}{d\eta/d\lambda} \quad . \end{aligned}$$

Como

$$\frac{d\eta}{d\lambda} = \frac{d\eta}{du} \frac{du}{d\lambda},$$

utilizando (2.9) e  $\frac{d\eta}{du} = \frac{d(\ln(u))}{du} = \frac{1}{u}$  tem-se

$$\frac{d\eta}{d\lambda} = \frac{1}{u} 4\lambda^3 z$$

$$\text{com } z = \sum_{i=1}^r \frac{\sigma_i^2 b_i^2}{(\sigma_i^2 + \lambda^2)^3}.$$

Assim:

$$\begin{aligned} \frac{d^2 \eta}{d\xi^2} &= -\frac{u}{\lambda^4 v^2} \lambda^2 v + u - \frac{2\lambda v u}{4\lambda^3 z} \\ \frac{d^2 \eta}{d\xi^2} &= -\frac{u}{\lambda^4 v^2} \lambda^2 v + u - \frac{v u}{2\lambda^2 z} \\ \frac{d^2 \eta}{d\xi^2} &= -\frac{u^2}{\lambda^6 v} \frac{\lambda^4}{u} + \frac{\lambda^2}{v} - \frac{1}{2z}. \end{aligned} \quad (2.14)$$

Se  $\lambda \leq \sigma_r$ , então  $\frac{2\sigma_i^2}{\sigma_i^2 + \lambda^2} \geq 1$ ,  $i = 1, \dots, r$ , e para  $\lambda \geq \sigma_1$  tem-se  $\frac{2}{\sigma_i^2 + \lambda^2} \geq \frac{1}{\lambda^2}$ ,  $i = 1, \dots, r$ .

Para  $\lambda \leq \sigma_r$  e  $b_{\perp} = 0$  tem-se  $2z \geq \frac{u}{\lambda^4}$ , enquanto que para  $\lambda \geq \sigma_1$  independentemente de  $b_{\perp}$  tem-se  $2z \geq \frac{v}{\lambda^2}$ . Isto implica que, em ambos os casos, tem-se  $\frac{\lambda^4}{u} + \frac{\lambda^2}{v} - \frac{1}{2z} > 0$ , ou seja,  $\frac{d^2 \eta}{d\xi^2} < 0$ .  $\blacksquare$

O método da curva-L consiste então em selecionar o parâmetro  $\lambda$  que maximiza a curvatura da curva-L. Essa curvatura, denotada por  $\kappa$ , como uma função de  $\lambda$ , é dada em [14] por:

$$\kappa = \frac{\hat{\rho}' \hat{\eta}'' - \hat{\rho}'' \hat{\eta}'}{((\hat{\rho}')^2 + (\hat{\eta}')^2)^{3/2}},$$

em que  $\hat{\eta} = \ln \|Ax_{\lambda} - b\|_2$  e  $\hat{\rho} = \ln \|x_{\lambda}\|_2$ .

As vantagens do método da curva-L são a robustez e a capacidade para tratar perturbações decorrentes de erros não independentes. No entanto, determinar o ponto de máxima curvatura pode não ser uma tarefa das mais fáceis. Outro agravante é quando a curva-L apresenta dois cantos.

Vários procedimentos para a determinação do canto da curva-L estão apresentados em [14].



### 2.2.3 Ponto fixo

O método do ponto fixo utiliza um procedimento iterativo para determinar o parâmetro de regularização  $\lambda$  onde, conforme [25],  $\lambda$  que minimiza a curvatura da curva-L numa escala bi-log está relacionado com o mínimo local da função

$$\psi_\mu(\lambda) = u(\lambda)v(\lambda)^\mu, \quad \mu > 0. \quad (2.15)$$

Em [25], é mostrado que um ponto de mínimo local da função (2.15) tem uma interpretação geométrica muito similar ao canto da curva-L. Ali também é provado que se a curvatura da curva-L é maximizada no ponto  $\lambda = \lambda^*$  e se a tangente à curva-L no ponto  $(\ln u(\lambda^*), \ln v(\lambda^*))$  possui inclinação  $\frac{-1}{\mu}$ , então  $\psi_\mu$  tem um mínimo local em  $\lambda = \lambda^*$  e a curva-L é localmente convexa em  $\lambda = \lambda^*$ .

Derivando (2.15) em relação à variável  $\lambda$ , obtém-se

$$\psi'_\mu(\lambda) = \frac{d\psi_\mu}{d\lambda} = v(\lambda)^\mu v'(\lambda) \left( \mu \frac{u(\lambda)}{v(\lambda)} + \frac{u'(\lambda)}{v'(\lambda)} \right). \quad (2.16)$$

Observando que  $v(\lambda)^\mu v'(\lambda) \neq 0$  e utilizando a expressão (2.11) tem-se que os pontos críticos de  $\psi_\mu$  devem satisfazer

$$\lambda^{*2} = \mu \frac{u(\lambda^*)}{v(\lambda^*)}.$$

Portanto, se  $\eta(\lambda) = \sqrt{u(\lambda)}$ ,  $\xi(\lambda) = \sqrt{v(\lambda)}$  e ainda se  $\psi_\mu$  atingir um mínimo local em  $\lambda = \lambda^* \neq 0$ , o parâmetro  $\lambda^*$  deve ser um ponto fixo da função  $\phi_\mu : \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$ , definida por

$$\phi_\mu(\lambda) = \sqrt{\mu} \frac{\eta(\lambda^*)}{\xi(\lambda^*)} = \sqrt{\mu} \frac{\|Ax_\lambda - b\|_2}{\|x_\lambda\|_2}, \quad 0 < \lambda < \infty. \quad (2.17)$$

Em [3] são estabelecidos critérios para localizar os pontos fixos da função  $\psi_\mu$ , quando eles existirem, e também para fornecer a natureza dos extremos da função  $\psi_\mu(\lambda)$ .

Em forma resumida, o método do ponto fixo pode ser descrito da seguinte maneira:

$$\left\{ \begin{array}{l} \lambda_0 : \text{chute inicial,} \\ \text{Para } k \geq 1 \text{ faça} \\ \lambda_{k+1} = \phi_\mu(\lambda_k) \end{array} \right.$$

A sequência será calculada até que um determinado critério de parada, escolhido previamente, seja atingido, detalhes podem ser vistos em [3].

## 2.2.4 Quase-otimalidade

O método quase-otimalidade tem como objetivo escolher o parâmetro de regularização  $\lambda$  baseado na estimativa do erro dada por

$$\|x_{\text{exato}} - x_\lambda\|_2 = \|b^T(AA^T + \lambda I_m)^{-4}AA^T\|_2^{1/2}$$

e, conforme [13], a minimização desta estimativa conduz ao problema de minimizar a função

$$\begin{aligned} \mathcal{Q}(\lambda) &= \lambda^2 \frac{dx_\lambda}{d(\lambda^2)} = \frac{1}{2} \lambda \frac{dx_\lambda}{d(\lambda)} \\ &= \left( \sum_{i=1}^p f_i(1 - f_i) \frac{u_i^T b}{\sigma_i} \right)^{1/2}. \end{aligned} \quad (2.18)$$

Para minimizar a referida função, usa-se  $\frac{dx_\lambda}{d(\lambda^2)} = - (A^T A + \lambda^2 I_m)^{-1} x_\lambda$ , obtida da forma básica de regularização de Tikhonov na qual (10) se es-

creve

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2 + \lambda^2 \|Lx\|_2^2, \quad (2.19)$$

com duas importantes formulações básicas:

$$(A^T A + \lambda^2 L^T L)x = b \quad e \quad \min \begin{pmatrix} A \\ \lambda L \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix}_2.$$

A partir destas formulações, vê-se que se  $\mathcal{N}(A) \cap \mathcal{N}(L) = 0$ , ou seja, se os espaços nulos de  $A$  e  $L$  possuem somente o vetor nulo como elemento comum, tal que a matriz coeficiente tem posto completo, então a solução de Tikhonov  $x_\lambda$  é única e formalmente definida por:

$$x_{\lambda,L} = A_\lambda^\# b.$$

com

$$A_\lambda^\# = (A^T A + \lambda^2 L^T L)^{-1} A^T.$$

Tomando  $L = I_n$ , vem

$$x_\lambda = (A^T A + \lambda^2 I_n)^{-1} A^T b. \quad (2.20)$$

Derivando (2.20) em relação a  $\lambda^2$ , vem

$$\frac{dx_\lambda}{d(\lambda^2)} = - (A^T A + \lambda^2 I_n)^{-1} x_\lambda.$$

A partir da iteração de Landweber que tem a forma

$$x^{(k)} = x^{(k-1)} + \omega A^T r^{(k-1)}, \quad k = 1, 2, \dots$$

em que  $x^{(0)}$  é o vetor inicial,  $\omega$  é um parâmetro real que satisfaz  $0 < \omega < 2\|A^T A\|_2^{-1}$  e  $r^{(k)} = b - Ax^{(k)}$  é o vetor resíduo correspondente a  $x^{(k)}$ , obtém-se a seguinte forma generalizada por Strand:

$$x^{(k)} = x^{(k-1)} + \mathcal{F}(A^T A) A^T r^{(k-1)}, \quad k = 1, 2, \dots$$

em que  $\mathcal{F}(A^T A) = \omega$  é um funcional de  $A^T A$ . Uma escolha particular do funcional  $\mathcal{F}$ , a saber,

$$\mathcal{F}(A^T A) = (A^T A + \lambda^2 I_m)^{-1},$$

conduz ao esquema iterativo

$$x^{(k)} = (A^T A + \lambda^2 I_m)^{-1}(A^T b + \lambda^2 x^{(k-1)}) \quad (2.21)$$

que é geralmente conhecido como *iteração de regularização de Tikhonov* quando toma-se  $x^{(0)} = 0$  [13].

É possível constatar que, em (2.21),  $x^{(1)}$  é a solução de Tikhonov  $x_\lambda$ , dada em (2.20), e no segundo passo da iteração de regularização de Tikhonov tem-se:

$$x_\lambda^{(2)} = (A^T A + \lambda^2 I_m)^{-1}(A^T b + \lambda^2 x_\lambda)$$

$$x_\lambda^{(2)} = x_\lambda - \lambda^2 \frac{dx_\lambda}{d(\lambda^2)}$$

tal que a minimização de  $\mathcal{Q}(\lambda)$  em (2.18) minimiza a correção para  $x_\lambda$  nesta solução.

Por conseguinte, o valor de  $\lambda$  procurado é aquele que minimiza a função  $\mathcal{Q}(\lambda)$ . No entanto, na prática a minimização de  $\mathcal{Q}(\lambda)$  é complicada pelo fato de  $\mathcal{Q}(\lambda)$  admitir vários mínimos locais, conforme pode ser visto na figura 2.3, na qual o gráfico de  $\mathcal{Q}(\lambda)$  foi obtido a partir do problema teste *Shaw* para  $n=512$  e erro relativo igual a 5%. Mas possui a vantagem de ser robusta ao nível de erro relativo [1].

Hansen em [13] relaciona a minimização de  $\mathcal{Q}(\lambda)$  às normas da perturbação e erros de regularização em  $x_\lambda$ . Os fatores de filtro satisfazem as

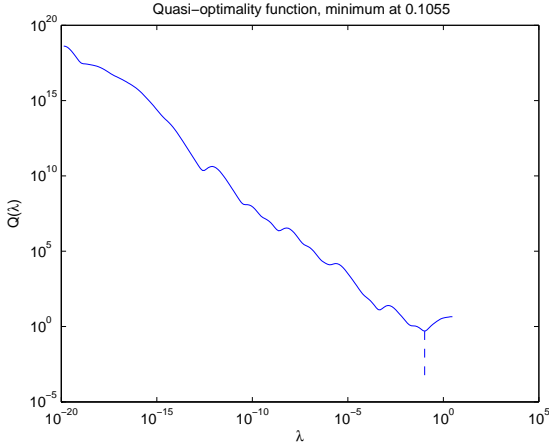


Figura 2.3: Gráfico da função quase-otimalidade  $\mathcal{Q}(\lambda)$ , problema *Shaw* com  $n=512$  e erro relativo de 5%.

relações  $f_1 \approx 1$  para  $\sigma_i \gg \lambda$  e  $f_1 \approx 0$  para  $\sigma_i \ll \lambda$ , obtendo-se

$$\mathcal{Q}(\lambda)^2 \approx \sum_{\sigma_i > \lambda} (1 - f_i)^2 \left( \frac{u_i^T b}{\sigma_i} \right)^2 + \sum_{\sigma_i < \lambda} f_i^2 \left( \frac{u_i^T b}{\sigma_i} \right)^2.$$

Assumiui-se que as condições discretas de Piccard foram satisfeitas e que os erros provenientes da matriz são pequenos tais que o limite de resolução, dado por  $\eta_{res} = \|u_{i_b}^T x_{exato}\| \approx \frac{\tau_b}{\sigma_{i_b}}$ , é determinado pelos erros do lado direito. Se  $i_b$  é o número de componentes da SVD nas quais  $u_i^T b_{exato}$  predomina, então  $u_i^T \hat{b} \approx u_i^T b$  e  $u_i^T e < u_i^T b$  para  $\sigma_i \gg \sigma_{i_b}$ , quando  $i \leq i_b$ , enquanto  $u_i^T \hat{b} \ll u_i^T b$  e  $u_i^T e \approx u_i^T b$  para  $\sigma_i \ll \sigma_{i_b}$ , quando  $i > i_b$ . Usando esta aproximações, obtém-se as seguintes relações para a regularização e

os erros de perturbação [13]:

$$\|A^\dagger - A^\# \hat{b}\|_2^2 = \sum_{i=1}^n (1 - f_i)^2 \left( \frac{u_i^T \hat{b}}{\sigma_i} \right)^2 \approx \sum_{\sigma_i \geq \lambda} (1 - f_i)^2 \left( \frac{u_i^T b}{\sigma_i} \right)^2$$

$$\|A^\# e\|_2^2 = \sum_{i=1}^n f_i^2 \left( \frac{u_i^T e}{\sigma_i} \right)^2 \approx \sum_{\sigma_i < \lambda} f_i^2 \left( \frac{u_i^T b}{\sigma_i} \right)^2.$$

Assim escolhendo  $\lambda \approx \sigma_{i_b}$ , obtém-se a seguinte expressão para a função quase-otimalidade:

$$\mathcal{Q}(\lambda)^2 \approx \|A^\dagger - A^\# \hat{b}\|_2^2 + \|A^\# e\|_2^2.$$

Desde que  $\|A^\dagger - A^\# \hat{b}\|_2$  cresce com  $\lambda$ , enquanto  $\|A^\# e\|_2$  decresça com  $\lambda$ , a minimização de  $\mathcal{Q}(\lambda)$  procura balancear o erro de regularização com o erro de perturbação enquanto  $\lambda$  está próximo do valor ótimo,  $\lambda \approx \sigma_{i_b}$ .

## 2.2.5 Discrepância

O método da discrepância ou critério de Morozov é, segundo Hansen em [13], o método mais difundido baseado na norma do vetor perturbação,  $\|e\|_2 = \|b - \hat{b}\|_2$ , para escolha do parâmetro de regularização. Se o problema é consistente no sentido que  $Ax_{exato} = \hat{b}$  vale exatamente, então a idéia para a escolha do parâmetro de regularização  $\lambda$  é que a norma do resíduo não deve exceder uma cota superior para o erro nos dados, isto é,

$$\|Ax_\lambda - b\|_2 = \delta_e, \quad \|e\|_2 \leq \delta_e. \quad (2.22)$$

Aplicando a SVD da matriz  $A$  na equação do resíduo  $r_\lambda = Ax_\lambda - b$ , tem-se  $r_\lambda = \sum_{i=1}^r (1 - f_i) u_i^T b u_i + b_\perp$ , em que  $r = \text{posto}(A)$  e  $f_i$  são os

fatores de filtro de Tikhonov dados em (2.7) e, ainda,  $b_{\perp} = \sum_{i=r+1}^n u_i^T b u_i$  é o componente do vetor  $b$  que não pertence ao espaço coluna da matriz  $A$ . Percebe-se que resolver (2.22) é equivalente a determinar a interseção entre a curva definida pela norma do resíduo e pela reta correspondente a  $\delta_e$  conforme visto na figura 2.4

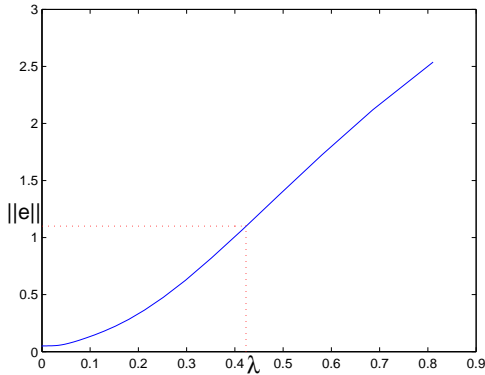


Figura 2.4: Interpretação geométrica do problema (2.22).

Quando uma boa estimativa é conhecida para  $\|e\|_2$ , este método produz um bom parâmetro de regularização para a solução regularizada imediatamente à direita do canto da curva-L. Devido a parte íngreme da curva-L, vê-se que uma estimativa por baixo de  $\|e\|_2$  é provável que se venha a produzir uma solução pouco regularizada de norma muito grande. Por outro lado, uma super estimativa de  $\|e\|_2$  produz uma solução super-regularizada com um grande erro de regularização.

Uma generalização do método da discrepância pode ser vista em Wang [30],

onde o parâmetro é determinado pelo uso do princípio amortecido da discrepância, isto é,  $\|Ax_\lambda - b\|^2 + \lambda\|x_\lambda\|^2 = \delta^2$ . Em [30] é afirmado ainda que este princípio é mais eficiente para maiores níveis de erros relativos.

## 2.3 Concavidade e Convexidade da Curva-L

O objetivo desta seção é descrever propriedades geométricas da curva-L introduzidos recentemente na literatura. Mostraremos que todas as propriedades dependem do comportamento da função  $\phi_1$ , que é obtida a partir de 2.17 fazendo  $\mu = 1$ , empregada pelo método de ponto fixo. Os resultados estão baseados em [3, 25]. No caso do lema 2.3, ele permite identificar os pontos fixos de  $\phi_\mu$ , dada em 2.17, quando existirem, e auxilia na classificação dos extremos de  $\psi_\mu$ , dada por 2.15.

**Lema 2.3.** *Assuma que  $\mu = 1$ . Sejam  $\underline{\sigma} = \min_i \sigma_i$  e  $\bar{\sigma} = \max_i \sigma_i$ . Se  $\|b_\perp\|_2 = 0$  e  $0 \leq \lambda \leq \underline{\sigma}$ , então*

$$0 \leq \phi_1(\lambda) \leq \lambda. \quad (2.23)$$

*Além disto, se  $\lambda \geq \bar{\sigma}$ , independente de  $\|b_\perp\|_2$  tem-se*

$$\phi_1(\lambda) \geq \lambda. \quad (2.24)$$

**Demonstração:** Partindo de  $\lambda \leq \underline{\sigma}$  e tomando-se  $\alpha_i = |u_i^T b|^2$ , ou seja o quadrado do coeficiente de Fourier de  $b$ , pode-se escrever  $\lambda^4 \alpha_i \leq \sigma^4 \alpha_i$  e então (2.8), no caso em que  $\|b_\perp\|_2 = 0$ , fica

$$u(\lambda) \leq \underline{\sigma}^4 \sum_{i=1}^p \frac{\alpha_i}{(\sigma_i^2 + \lambda^2)^2}. \quad (2.25)$$



Para,  $\lambda \geq 0$  (2.6) fica

$$v(\lambda) \geq \underline{\sigma}^2 \sum_{i=1}^p \frac{\alpha_i}{(\sigma_i^2 + \lambda^2)^2}. \quad (2.26)$$

Então para  $0 \leq \lambda \leq \underline{\sigma}$  e com o uso de (2.25) e (2.26) tem-se  $0 \leq \phi_1(\lambda) \leq \underline{\sigma}$ . Agora para  $\tilde{\lambda}$  fixo e  $0 \leq \lambda \leq \tilde{\lambda} \leq \underline{\sigma}$  tem-se

$$u(\lambda) \leq \tilde{\lambda}^4 \sum_{i=1}^p \frac{\alpha_i}{(\sigma_i^2 + \lambda^2)^2}. \quad (2.27)$$

A partir da desigualdade (2.27) e utilizando (2.26) chega-se a  $\phi_1(\lambda)^2 \leq \tilde{\lambda}^4 / \underline{\sigma}^2 \leq \tilde{\lambda}^2$ , provando assim a desigualdade (2.23). Analogamente prova-se a desigualdade (2.24) para  $\|b_{\perp}\|_2 = 0$ .

Para todo  $\lambda > 0$  a desigualdade  $2\sigma_i\lambda \leq \sigma_i + \lambda^2$  leva a

$$\frac{\sigma_i^2}{(\sigma_i^2 + \lambda^2)^2} \leq \frac{1}{4\lambda^2}$$

e para  $\lambda \geq \bar{\sigma}$  tem-se

$$\sum_{i=1}^p \frac{\sigma_i^2 \alpha_i}{(\sigma_i^2 + \lambda^2)^2} \leq \frac{1}{4\bar{\sigma}^2} \sum_{i=1}^p \alpha_i. \quad (2.28)$$

Também para  $\lambda \geq \bar{\sigma}$  tem-se  $\lambda^4 \geq \bar{\sigma}^2 \sigma_i^2$  que implica em

$$\sum_{i=1}^p \frac{\lambda^4 \alpha_i}{(\sigma_i^2 + \lambda^2)^2} \geq \bar{\sigma}^2 \sum_{i=1}^p \frac{\sigma_i^2 \alpha_i}{(\sigma_i^2 + \lambda^2)^2}. \quad (2.29)$$

Para  $\lambda \geq \bar{\sigma}$  as desigualdades (2.28) e (2.29) fornecem

$$\phi_1(\lambda)^2 \geq \bar{\sigma}^2 \left( 1 + 4 \frac{\|b_{\perp}\|_2}{\sum_{i=1}^p \alpha_i} \right) \geq \bar{\sigma}^2.$$

Similarmente, demonstra-se (2.24). ■

**Teorema 2.4.** *Assuma que  $\mu = 1$ . Sejam  $I_1$  e  $I_2$  dois intervalos abertos tais que  $\phi_1(\lambda) < \lambda, \forall \lambda \in I_1$  e  $\phi_1(\lambda) > \lambda, \forall \lambda \in I_2$ . Então  $\psi_\mu$  é crescente em  $I_1$  e decrescente em  $I_2$ . Além disso, valem as seguintes afirmações:*

- (a) *Se  $\|b_\perp\|_2 = 0$  e  $\alpha_i = |u_i^T b|^2 \neq 0, i = 1, \dots, r$ , existe  $\tilde{\lambda} \in ]\underline{\sigma}, \bar{\sigma}[$  tal que  $\psi_\mu$  possui um máximo local em  $\tilde{\lambda}$ . Ainda, se  $\tilde{\lambda}$  é um ponto fixo de  $\phi$  mais próximo de zero e  $\psi_\mu$  tem um mínimo local em  $\lambda^*$ , então  $\tilde{\lambda} < \lambda^*$  e  $\psi_\mu$  possui outro máximo local em  $]\lambda^*, \bar{\sigma}[$ .*
- (b) *Se  $\|b_\perp\|_2 \neq 0$  e  $\psi_\mu$  possui um mínimo local em  $\lambda^*$ , existe um parâmetro  $\tilde{\lambda}$  no intervalo  $]\lambda^*, \bar{\sigma}[$  no qual  $\psi_\mu$  tem um máximo local.*
- (c) *Seja  $\phi$  com um ponto fixo  $\lambda^*$  e seja  $P$  um ponto da curva-L na escala bi-log associado a  $\lambda^*$ . Então a curva-L é convexa na vizinhança de  $P$  se e só se  $\lambda^*$  minimiza localmente  $\psi_\mu$  e ela é côncava numa vizinhança de  $P$  se e só se  $\lambda^*$  maximiza localmente  $\psi_\mu$ .*

**Demonstração:** A derivada de (2.15) em relação à variável  $\lambda$ , já vista em (2.16), é dada por  $\psi_\mu(\lambda)' = \frac{d\psi_\mu}{d\lambda} = v(\lambda)^\mu v'(\lambda) \left( \mu \frac{u(\lambda)}{v(\lambda)} + \frac{u'(\lambda)}{v'(\lambda)} \right)$  e utilizando (2.11), vem

$$\psi_\mu(\lambda)' = v(\lambda)^\mu v'(\lambda) \left( \mu \frac{u(\lambda)}{v(\lambda)} - \lambda^2 \right).$$

Mas  $\mu \frac{u(\lambda)}{v(\lambda)} = \phi_\mu(\lambda)^2$  então

$$\psi_\mu(\lambda)' = v(\lambda)^\mu v'(\lambda) \left( \phi_\mu(\lambda)^2 - \lambda^2 \right). \quad (2.30)$$

Logo por (2.24) conclui-se que  $\psi_\mu(\lambda)' > 0$  sempre que  $\lambda \in I_1$ , e também que  $\psi_\mu$  é crescente para  $\lambda \in I_1$ . Analogamente mostra-se que  $\psi_\mu$  é decrescente para  $\lambda \in I_2$ .

Para provar (a), deve-se observar inicialmente que se  $b_\perp = 0$  nem  $\underline{\sigma}$  nem  $\bar{\sigma}$  podem ser pontos fixos de  $\phi_\mu$ , pelo menos não no contexto de problemas discretos mal postos. De fato, se  $\lambda^*$  é um ponto fixo de  $\phi_\mu$  então

$$\begin{aligned} u(\lambda^{*2}) &= \lambda^* v(\lambda) \Leftrightarrow \\ \Leftrightarrow \sum_{i=1}^p \frac{\lambda^{*4} \alpha_i}{(\sigma_i^2 + \lambda^{*2})^2} &= \lambda^{*2} \sum_{i=1}^p \frac{\sigma_i^2 \alpha_i}{(\sigma_i^2 + \lambda^{*2})^2} \Leftrightarrow \sum_{i=1}^p \frac{\lambda^{*2} - \sigma_i^2}{(\sigma_i^2 + \lambda^{*2})^2} \alpha_i = 0. \end{aligned} \tag{2.31}$$

A última igualdade em (2.31) permanecerá válida com  $\lambda^* = \underline{\sigma}$  somente quando  $\sigma_i = \underline{\sigma}$  para  $i = 1, \dots, p$ . Mas isso não pode acontecer neste contexto pois assumiu-se que os valores singulares generalizados caem para zero de modo contínuo. Além disso  $\underline{\sigma}$  não pode ser um ponto fixo de  $\phi_\mu$ . Analogamente conclui-se que  $\bar{\sigma}$  não pode ser um ponto fixo de  $\phi_\mu$ . A partir desta observação e do lema 2.3 segue que  $\phi_\mu$  tem um ponto fixo no intervalo  $]\underline{\sigma}, \bar{\sigma}[$ . As afirmações em (a) seguem usando este resultado juntamente com (2.23), (2.24) e (2.30).

Para provar (b) deve-se observar também que neste caso  $\bar{\sigma}$  não pode ser um ponto fixo de  $\phi_\mu$ . Veja agora que por causa do lema 2.3 e do fato que  $v(\lambda) \rightarrow 0$  quando  $\lambda \rightarrow \infty$ , que é imediato a partir de (2.6), segue que  $\psi_\mu$  é monotonicamente decrescente quando  $\lambda > \bar{\sigma}$  e  $\psi_\mu \rightarrow 0$  quando  $\lambda \rightarrow \infty$ . Isto mostra que o extremo local de  $\psi_\mu$  pode estar dentro do intervalo aberto  $]0, \bar{\sigma}[$ . As afirmações descrita em (b) são consequência imediata desta observação e das propriedades de monotonicidade de  $\psi_\mu$ .

A primeira afirmação em (c) é o teorema 1 em [25]. Para provar a segunda afirmação tome  $f = \ln$  em (2.13) para chegar a

$$\frac{d\eta}{d\xi} = -\frac{\phi_1(\lambda)^2}{\lambda^2} \quad (2.32)$$

que derivando vem

$$\frac{d^2\eta}{d\xi^2} \frac{d\xi}{d\lambda} = -\frac{2}{\lambda^2} \phi_1(\lambda) \phi_1'(\lambda) - \frac{\phi_1(\lambda)}{\lambda} \quad . \quad (2.33)$$

Assim chega-se em  $\phi_1'(\lambda^*) > 1$  numa vizinhança (suficientemente pequena) de  $\lambda^*$  se e só se  $\lambda^*$  maximiza  $\psi_\mu$ . A partir desse resultado e lembrando que  $\frac{d\xi}{d\lambda} = \frac{u'(\lambda)}{u(\lambda)} > 0$  a segunda afirmação de (c) é determinada com o uso de (2.33). ■

**Teorema 2.5.** *Sejam  $\hat{\eta} = \ln u(\lambda)$ ,  $\hat{\rho} = \ln v(\lambda)$  e  $m_L(\lambda)$  a inclinação da curva-L, em escala bi-log, no ponto  $(\hat{\eta}, \hat{\rho})$ . Então a curva-L é convexa na vizinhança de  $\tilde{\lambda}$  se e somente se para todo  $\lambda$  nesta vizinhança vale a relação*

$$\phi_1'(\lambda) \leq \frac{\phi_1(\lambda)}{\lambda} \quad . \quad (2.34)$$

**Demonstração:**

$$\begin{aligned} m_L(\lambda) \frac{d\hat{\rho}}{d\hat{\eta}} &= \frac{d\hat{\rho}}{d\lambda} \cdot \frac{d\lambda}{d\hat{\eta}} \\ &= \frac{v'(\lambda)}{v(\lambda)} \cdot \frac{u(\lambda)}{u'(\lambda)} = \frac{v'(\lambda)}{u'(\lambda)} \cdot \frac{u(\lambda)}{v(\lambda)} \\ &= \frac{-1}{\lambda^2} \cdot \phi_1(\lambda)^2 = -\frac{\phi_1(\lambda)^2}{\lambda^2} \quad . \end{aligned}$$

Derivando ambos os membros da sentença anterior em relação a  $\lambda$  segue que

$$\begin{aligned} \frac{d^2 \hat{\rho}}{d\hat{\eta}^2} \frac{u'(\lambda)}{u(\lambda)} &= m'_L(\lambda) \\ &= -2 \frac{\phi_1(\lambda)}{\lambda} \left[ \frac{\lambda \phi_1'(\lambda) - \phi_1(\lambda)}{\lambda^2} \right] \end{aligned}$$

Assumindo que  $m'_L(\lambda) \geq 0$ , para termos a convexidade da curva-L, e lembrando que  $\frac{u'(\lambda)}{u(\lambda)}$  é sempre positivo, conforme (2.8) e (2.9), segue que  $\lambda \phi_1'(\lambda) - \phi_1(\lambda) \leq 0$ , comprovando imediatamente a tese do teorema. ■

**Corolário 2.6.** *Assuma que a função  $\zeta(\lambda) = \phi_1(\lambda)/\lambda$ ,  $\lambda > 0$ , atinja dois extremos relativos em dois pontos críticos consecutivos de  $\zeta$ , chamados  $\lambda_1$  e  $\lambda_2$ , com  $\lambda_1 < \lambda_2$ . Então a curva-L é convexa no intervalo  $(\lambda_1, \lambda_2)$  se  $\lambda_1$  é um ponto de máximo local e a curva-L é côncava sobre  $(\lambda_1, \lambda_2)$  se  $\lambda_1$  é um ponto de mínimo local.*

**Teorema 2.7.** *Assuma que  $\lambda^*$  é um ponto fixo de  $\phi_1(\lambda)$  satisfazendo  $\phi_1''(\lambda) \neq 0$ . Então a curva-L é convexa na vizinhança de  $\lambda^*$  se e só se  $-\lambda^* \frac{v'(\lambda^*)}{v(\lambda^*)} < 1$ . Além disso, quando esta igualdade vale, tem-se  $\lambda^* < \frac{\sqrt{3}}{3} \sigma_1$ , e existe  $\lambda_* \in (\lambda^*, \sigma_1)$  que é um ponto fixo de  $\phi_1(\lambda)$  no qual a curva-L é localmente côncava.*

**Demonstração:** Derivando ambos os membros da identidade  $\phi_1(\lambda)^2 = \frac{u(\lambda)}{v(\lambda)}$  em relação à variável  $\lambda$  obtém-se

$$2\phi_1(\lambda)\phi_1'(\lambda) = \frac{v(\lambda)u'(\lambda) - u(\lambda)v'(\lambda)}{v(\lambda)^2}$$

$$= -\frac{\lambda^2 v(\lambda) + u(\lambda)}{v(\lambda)^2} v'(\lambda)$$

onde foi usado o fato de que  $u'(\lambda) = -\lambda^2 v'(\lambda)$  (veja (2.11)). Assim tem-se

$$\phi_1'(\lambda^*) = -\frac{\lambda^{*2} + \phi_1(\lambda^*)^2}{2\phi_1(\lambda^*)} \frac{v'(\lambda^*)}{v(\lambda^*)}.$$

Como  $\lambda^*$  é um ponto fixo de  $\phi_1(\lambda)$ , isto é,  $\frac{\lambda^{*2} + \phi_1(\lambda^*)^2}{2\phi_1(\lambda^*)} = \lambda^*$ , vem

$$\phi_1'(\lambda^*) = -\frac{\lambda^* v'(\lambda^*)}{v(\lambda^*)}.$$

Assumindo  $\phi_1'(\lambda^*) = -\lambda^* \frac{v'(\lambda^*)}{v(\lambda^*)} < 1$ , segue que existe uma vizinhança de  $\lambda^*$  tal que  $\phi_1'(\lambda) \leq \frac{\phi_1(\lambda)}{\lambda}$ , para todo  $\lambda$  nesta vizinhança, e o teorema 2.5 assegura que a curva-L é convexa nesta vizinhança.

Reciprocamente, se a curva-L é convexa nesta vizinhança de  $\lambda^*$ , pelo teorema 2.5 tem-se que

$$-\lambda^* \frac{v'(\lambda^*)}{v(\lambda^*)} \leq 1,$$

só faltando provar que a igualdade não é verificada na desigualdade anterior.

De fato, se tivéssemos  $\phi_1'(\lambda^*) = 1$ , então  $\lambda^*$  seria um ponto crítico de  $\zeta$  e assim  $\lambda^* \zeta''(\lambda^*) > 1$  e desde que  $\phi''\lambda^* \neq 0$  por hipótese, segue que  $\lambda^*$  é um ponto extremo de  $\zeta$  e, conseqüentemente pelo corolário 2.6, um ponto onde a curva-L muda de concavidade, o que é uma contradição.

Agora para provar que  $\lambda^* < \frac{\sqrt{3}}{3} \sigma_r$  deve-se observar que, similarmente ao caso convexo, uma condição necessária e suficiente para a curva-L ser localmente côncava em  $\lambda_*$  é  $\phi_1'(\lambda_*) > 1$ . Tomando esta observação e

assumindo agora que a sentença a ser provada não é verdadeira, isto é, assumindo que  $\lambda^* \geq \frac{\sqrt{3}}{3}\sigma_r$ , então devido a (2.6), (2.8), (2.9) e (2.10), vem

$$-\lambda^* v'(\lambda^*) - v(\lambda^*) = \sum_{i=1}^r \frac{3\lambda^{*2} - \sigma_i^2}{\sigma_i^2 + \lambda^{*2}} \frac{\sigma_i^2 \alpha_i}{3} > 0 \Rightarrow -\lambda^* \frac{v'(\lambda^*)}{v(\lambda^*)} > 1,$$

assim implicando que a curva-L é localmente côncava em  $\lambda^*$ , o que é uma contradição. Consequentemente  $\lambda^{*2} < \frac{\sigma_r^2}{3}$ .

Para provar a última parte deve-se recordar de [3] que os pontos fixos de  $\phi_1$  quando existem eles pertencem ao intervalo  $]0, \sigma_r[$  e, também, para todo  $\lambda > \sigma_r$  tem-se  $\phi_1(\lambda) \geq \lambda$ . Disto segue que a expressão  $\lambda - \phi_1(\lambda)$  muda de sinal no intervalo  $]\lambda^*, \sigma_r[$  e assim  $\phi_1(\lambda_*) = \lambda_*$ . Completando esta demonstração vê-se que a condição  $\phi_1'(\lambda) > 1$  é satisfeita numa vizinhança suficientemente pequena de  $\lambda_*$ . ■

Uma conclusão importante extraída do teorema 2.7, diz respeito a escolha da iteração inicial (chute inicial) utilizado no algoritmo do método do ponto fixo. Este chute deverá ser o maior possível, mas nunca superior a  $\sigma_1/\sqrt{3}$ .

### 2.3.1 Sobre a condição $\phi_1'(\lambda^*) < 1$ e a hipótese

$$\phi_1''(\lambda^*) \neq 0$$

Assumindo que  $\lambda^*$  é um ponto fixo de  $\phi_\mu$ , ou seja,  $\lambda^* = \sqrt{\mu}\phi_1(\lambda^*)$ , pode-se analisar o papel que a condição  $\phi_1'(\lambda^*) < 1$  e a hipótese  $\phi''(\lambda^*) \neq 0$  desempenham na minimização do funcional  $\psi_\mu$  de Regińska e na convexidade da curva-L. Assim, é necessário determinar as derivadas de primeira

e de segunda ordem do funcional dado em (2.15), no ponto  $\lambda^*$ , obtendo-se

$$\begin{aligned}
 \psi'_\mu(\lambda^*) &= u'(\lambda^*)v^\mu(\lambda^*) + u(\lambda^*)\mu v^{\mu-1}(\lambda^*)v'(\lambda^*) \\
 \psi'_\mu(\lambda^*) &= u'(\lambda^*)v^\mu(\lambda^*) + \frac{u}{v}\mu v^\mu(\lambda^*)v'(\lambda^*) \\
 \psi'_\mu(\lambda^*) &= -\lambda^{*2}v'(\lambda^*)v^\mu(\lambda^*) + \phi_1^2(\lambda^*)\mu v^\mu(\lambda^*)v'(\lambda^*) \\
 \psi'(\lambda^*) &= -\lambda^{*2} + \phi_1^2(\lambda^*)\mu v^\mu(\lambda^*)v'(\lambda^*) = 0. \quad (2.35)
 \end{aligned}$$

A derivada segunda de (2.15) é igual a

$$\begin{aligned}
 \psi''_\mu(\lambda) &= -2\lambda + 2\phi_1(\lambda)\phi'_1(\lambda)\mu v^\mu(\lambda)v'(\lambda) + -\lambda^2 + \\
 &+ \phi_1^2(\lambda)\mu \mu v^{\mu-1}(\lambda)v'^2(\lambda) + v^\mu(\lambda)v''(\lambda) .
 \end{aligned}$$

Substituindo  $\lambda$  por  $\lambda^*$  na expressão anterior e vendo que  $-\lambda^{*2} + \phi_1^2(\lambda^*)\mu = 0$  vem

$$\begin{aligned}
 \psi''_\mu(\lambda^*) &= -2\lambda^* + 2\frac{\lambda^*}{\sqrt{\mu}}\phi'_1(\lambda^*)\mu v^\mu(\lambda^*)v'(\lambda^*) \\
 \psi''_\mu(\lambda^*) &= -2\lambda^* 1 - \sqrt{\mu}\phi'_1(\lambda^*) v^\mu(\lambda^*)v'(\lambda^*) . \quad (2.36)
 \end{aligned}$$

Sabe-se que para minimizar  $\psi$  é suficiente ocorrer  $\psi'' > 0$ , e como  $v'(\lambda^*) < 0$  resulta a partir de (2.36)

$$1 - \sqrt{\mu}\phi'_1(\lambda^*) > 0$$

ou, ainda,  $\sqrt{\mu}\phi'_1(\lambda^*) < 1$ .

Isto acontece quando  $\phi'_1(\lambda^*) < 1$  e  $\mu \leq 1$  ou quando  $\phi'_1(\lambda^*) \leq 1$  e  $\mu < 1$ . Mas no caso  $\phi'_1(\lambda^*) = \frac{1}{\sqrt{\mu}}$ , isto é,  $\psi''_\mu(\lambda^*) = 0$  devem ser analisadas



as derivadas de maior ordem. Restringindo à terceira e quarta derivadas, tem-se

$$\psi_\mu'''(\lambda^*) = 2\sqrt{\mu}\lambda^* \phi_1''(\lambda^*)v^\mu(\lambda^*)v'(\lambda^*), \quad (2.37)$$

e

$$\begin{aligned} \psi_\mu''''(\lambda^*) = & 2\sqrt{\mu} \lambda^* \phi_1'''(\lambda^*) + 3\phi_1''(\lambda^*) v^\mu(\lambda^*)v'(\lambda^*) + \\ & + 4\sqrt{\mu} \lambda^* \phi_1''(\lambda^*) \mu v^{\mu-1}(\lambda^*)v'^2(\lambda^*) + v^\mu(\lambda^*)v''(\lambda^*) \quad , \end{aligned} \quad (2.38)$$

onde, como antes,  $v^\mu(\lambda^*)v'(\lambda^*) < 0$  é sempre válido.

Algumas informações sobre  $\phi_1''(\lambda^*)$  podem ser obtidas a partir de (2.37) e (2.38), como as que seguem:

- (i) Se  $\phi_1''(\lambda^*) \neq 0$ , então  $\lambda^*$  é um ponto de inflexão de  $\psi_\mu$ .
- (ii) Se  $\phi_1''(\lambda^*) > 0$  e  $\mu = 1$ , então  $\lambda^*$  é um mínimo local de  $\zeta$ . Isto ocorre porque  $\zeta''(\lambda^*) > 0$  e  $\zeta'(\lambda^*) = \frac{\lambda^* \phi_1'(\lambda^*) - \phi_1(\lambda^*)}{\lambda^{*2}} = 0$ , pois  $\phi_1'(\lambda^*) = 1$  e  $\phi_1(\lambda^*) = \lambda^*$ . Como  $\phi_1'(\lambda^*) = 1$ , então a reta tangente ao gráfico de  $z = \phi_1(\lambda)$ , no ponto  $(\lambda^*, \phi_1(\lambda^*))$ , é igual a  $z = \lambda$ . Logo  $\phi_1'(\lambda) \geq \lambda$ , para todo  $\lambda \in I$  contendo  $\lambda^*$ . Assim, pelo Corolário 1, a curva-L muda a concavidade dentro deste intervalo  $I$ . O mesmo raciocínio se aplica para o caso em que  $\phi_1''(\lambda^*) < 0$ .
- (iii) Se  $\phi_1''(\lambda^*) = 0$ , então utilizando (2.37) tem-se  $\psi_\mu'''(\lambda^*) = 0$  e por (2.38) verifica-se que  $\psi_\mu''''(\lambda^*) = 2\sqrt{\mu}\lambda^* \phi_1'''(\lambda^*)$ . Assim, se  $\phi_1'''(\lambda^*) = 0$  e  $\phi_1''''(\lambda^*) < 0$ , então  $\lambda^*$  é um ponto de mínimo de  $\psi_\mu$ .

### 2.3.2 Conclusões sobre a concavidade/ convexidade da curva-L

Em [2] são apresentadas conclusões sobre a convexidade e a concavidade da curva-L. Para tanto, é tomado  $\lambda_*$  como sendo o maior ponto fixo de  $\phi_1$  e, no caso em que  $\delta_0 = 0$  e  $\zeta$  tem um extremo relativo, é adotado como hipóteses simplificadoras que  $\zeta$  não possui mais que um extremo relativo nem entre dois pontos consecutivos de  $\phi_1$  nem no intervalo  $(\lambda_*, \sigma_1)$ . Quando  $\delta_0 \neq 0$ , segundo o autor existe um ponto fixo  $\lambda^*$  que minimiza  $\psi_1$ , assumindo que não é permitido para  $\zeta$  ter mais do que dois extremos relativos no intervalo  $(0, \lambda^*)$ . Conforme [2]:

- (i) O caso  $b_\perp = 0$  é fácil de ser analisado ao utilizar-se a aproximação do ponto fixo. Como o teorema 1 garante sempre a existência de no mínimo um ponto fixo não nulo sobre o intervalo  $(\sigma_n, \sigma_1)$ , algumas conclusões podem ser tiradas se for conhecida a quantidade de pontos fixos de  $\phi_1$ . Supondo que  $\lambda_*$  seja o único ponto fixo de  $\phi_1$ , então a curva-L será côncava sobre todo o domínio de  $\zeta$  contanto que  $\zeta$  não tenha extremos no intervalo  $(\sigma_n, \sigma_1)$ .

Se por outro lado for assumido que  $\phi_1$  tem um único ponto fixo no qual a curva-L é localmente convexa (isto é,  $\psi_1$  tem um único minimizador), que significa que  $\phi_1$  tem três pontos fixos (veja o teorema 1 em [2]), a conclusão é que a curva-L muda a concavidade no mínimo três vezes, conforme a figura 2.5. Mais precisamente a curva-L será côncava entre a origem e o máximo relativo de  $\zeta$ , convexa entre dois extremos relativos de  $\zeta$  e côncava à direita do

mínimo relativo de  $\zeta$ .

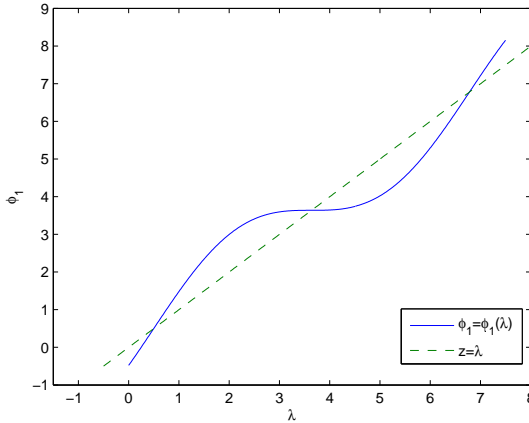


Figura 2.5: Gráfico de  $\phi_1$  com três pontos fixos.

- (ii) Se  $b_{\perp} \neq 0$ , os pontos fixos de  $\phi_1$  não estarão sempre garantidos e os intervalos de concavidade da curva-L não podem ser determinados utilizando informações do ponto fixo. Então a resposta definitiva sobre a questão dependerá finalmente do número de extremos relativos de  $\zeta$ . No entanto, sabendo que  $\zeta$  sempre tem um mínimo próximo da origem (conforme análise feita por Regińska) algumas conclusões podem ser obtidas: se  $\zeta$  possui um máximo local em  $\lambda_1$ , além do mínimo local próximo da origem, existirá  $\lambda_2$ , localizado à direita de  $\lambda_1$ , para o qual  $\zeta$  tem outro mínimo local e, se  $\zeta$  não tem um extremo relativo extra, a conclusão é que a curva-L é convexa sobre o intervalo  $(0, \epsilon)$ , para  $\epsilon$  pequeno, côncava sobre  $(\epsilon, \lambda_1) \cup (\lambda_2, \infty)$  e convexa em  $(\lambda_1, \lambda_2)$ .

### 2.3.3 Observações finais sobre o método do ponto fixo

O caso onde  $\zeta$  tem dois máximos locais e três mínimos locais, aparece em geral em problemas onde a função  $\psi_1$  tem quatro pontos fixos e a curva-L possui dois “cantos”, um deles associado com o menor  $\lambda$ . Um exemplo desta situação é o problema teste *Heat*, cujo gráfico da curva-L, para uma matriz  $256 \times 256$  e erro relativo de 5% no termo independente  $b$ , encontra-se na figura 2.6.

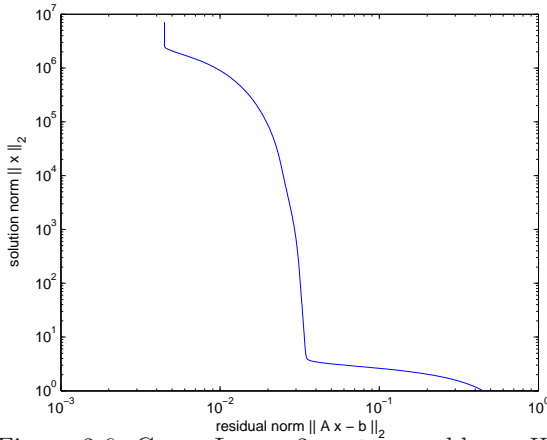


Figura 2.6: Curva-L com 2 cantos, problema *Heat*

com  $n=256$  e 5% de erro relativo em  $b$ .

Ainda conforme [2] os problemas nos quais a curva-L possui vários cantos convexos, tanto o método da curva-L como o algoritmo do ponto fixo tendem a falhar. Tal fato leva a introduzir duas variações no algoritmo.

Na primeira variação do algoritmo, o maior ponto fixo que minimiza  $\phi_1$  pode ser sempre calculado se o algoritmo do ponto fixo é iniciado razoavelmente. Mais precisamente, utilizando a conclusão do Teorema 3, é razoável esperar que  $\phi_1(\lambda_0) < \lambda_0$  para  $\lambda_0 \approx \frac{\sqrt{3}}{3}\sigma_1$ . Quando este é o caso,  $\lambda_{k+1} = \phi_1(\lambda_k)$ ,  $k \geq 0$  é uma sequência decrescente, enquanto  $\phi_1$  é crescente em relação a  $\lambda$  (em que  $\lambda$  converge para um ponto fixo localizado à esquerda de  $\lambda_0$  quando  $b_\perp \neq 0$  ou converge para zero quando  $b_\perp = 0$ ) e  $\phi_1$  tem um ponto fixo não nulo no intervalo  $(0, \lambda_0)$ . Mesmo que o último caso seja muito difícil de acontecer no conjunto dos problemas discretos mal postos, o algoritmo do ponto fixo é dito divergente se ele convergir para zero.

Assim, assumindo que  $\lambda_0 \approx \frac{\sqrt{3}}{3}\sigma_1$  satisfazendo  $\phi_1(\lambda_0) < \lambda_0$ , a primeira variação do algoritmo começa pela escolha  $\mu = 1$  como um valor inicial e continua o cálculo das iterações seguintes até que a convergência seja alcançada ou até encontrar uma divergência. Quando a convergência é atingida, um ponto fixo  $\lambda^*$  é calculado e um teste de convexidade da curva  $L$  é feito em  $\lambda^*$ . Quando diverge ou quando a convergência é alcançada e o teste da convexidade não é satisfeito, o parâmetro  $\mu$  é ajustado e as iterações recomeçam.

Na segunda variação do algoritmo, deve-se observar que enquanto  $\tilde{\lambda} > \sigma_1$  a função  $\phi_1$  satisfaz  $\phi_1(\tilde{\lambda}) > \tilde{\lambda}$  e para sua inversa  $\phi_1^{-1}$  tem-se  $\phi_1^{-1}(\tilde{\lambda}) < \tilde{\lambda}$ . Considere agora a sequência, dita inversa,

$$\tilde{\lambda}_{k+1} = \phi_1^{-1}(\tilde{\lambda}_k), \quad k = 0, 1, 2, \dots \quad (2.39)$$

Então devido ao fato de que  $\phi_1$  é estritamente crescente e sua inversa  $\phi_1^{-1}$  também, segue que a sequência inversa (2.39) está bem definida e  $\tilde{\lambda}_{k+1}$  é

a única solução da equação não linear  $v_k(\lambda) \equiv \phi_1(\lambda) - \tilde{\lambda}_k = 0$ .

Ainda, se  $\phi_1^{-1}(\tilde{\lambda}) > \tilde{\lambda}$ , então  $\tilde{\lambda}_{k+1}$  está localizado à esquerda de  $\tilde{\lambda}_k$ , conforme pode ser visto na figura 2.7.

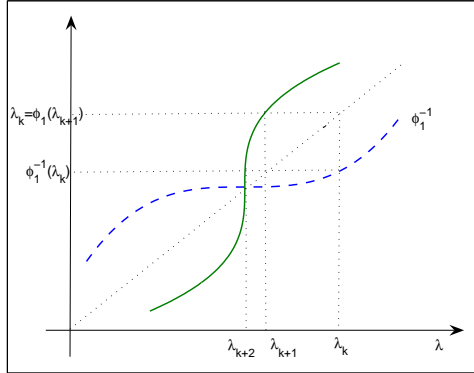


Figura 2.7: Gráfico de  $\phi_1$  e  $\phi_1^{-1}$  e dois passos da sequência inversa.

Uma imediata generalização deste resultado é que se existe garantia de  $\phi_1$  ter pontos fixos, então a sequência inversa converge para um ponto fixo de  $\phi_1$  que está à esquerda de  $\tilde{\lambda}_0$  se  $\phi_1(\tilde{\lambda}_0) > \tilde{\lambda}_0$  ou, então, converge para um ponto fixo de  $\phi_1$  que está à direita de  $\tilde{\lambda}_0$  se  $\phi_1(\tilde{\lambda}_0) < \tilde{\lambda}_0$ . Por isso, se  $\phi_1$  tem no mínimo um ponto fixo e a sequência inversa começa com  $\tilde{\lambda}_0 > \sigma_n$ , então  $\lim_{x \rightarrow \infty} \tilde{\lambda}_k = \lambda_*$  onde  $\lambda_*$  é o maior ponto fixo de  $\phi_1$ .

O cálculo de  $\tilde{\lambda}_{k+1}$  pode ser feito de várias maneiras. Dentre elas, a escolhida em [2] para o desenvolvimento do aperfeiçoamento do algoritmo foi o método Regula-Falsi. Resumidamente, esta segunda variação,

propõe o uso da sequência inversa (2.39) para determinar o maior ponto fixo côncavo de  $\phi_1$  e continua utilizando a primeira variação do algoritmo para chegar ao ponto fixo desejado.

A vantagem do método do ponto fixo em relação ao método da curva-L, é que praticamente é necessário calcular somente a norma da solução e a norma do resíduo, enquanto que no método da curva-L é preciso calcular a SVD (ou GSVD) ou, ainda, calcular a derivada da norma da solução em relação ao parâmetro de regularização. Isto faz o método do ponto fixo mais simples e potencialmente mais adequado para problemas de grande porte. A dificuldade existente neste método reside no fato que ele possui uma forte dependência do chute inicial, pois uma escolha incorreta deste pode conduzir a um ponto fixo de  $\phi_\mu(\lambda)$  que venha a maximizar  $\psi_\mu$ .





## Capítulo 3

# Regularização iterativa

Os métodos iterativos para a resolução de sistemas de equações lineares  $Ax = b$  e para problemas de mínimos quadrados  $\operatorname{argmin}_{x \in \mathbb{R}^n} \|Ax - b\|_2^2$  baseiam-se em esquemas que geram vetores iterativos  $x^{(k)}$ ,  $k = 1, 2, \dots$  que convergem para a solução desejada. Quando  $Ax = b$  ou  $\operatorname{argmin}_{x \in \mathbb{R}^n} \|Ax - b\|_2^2$  provém de problema mal posto, há a necessidade de que cada solução iterada  $x^{(k)}$  seja considerada uma solução regularizada, onde  $k$  cumpre o papel do parâmetro de regularização.

Este capítulo apresenta os métodos iterativos LSQR, GMRES e RRMRES para a escolha do parâmetro de regularização  $\lambda$  do método de regularização de Tikhonov, dado em (10), bem como os critérios de parada de Morigi e da discrepância. Aqui também é apresentado um novo critério de parada para problemas discretos mal postos.

## 3.1 Introdução

Nos problemas mal postos, a questão é saber em qual iteração  $k$  deve-se parar, pois no início das iterações a norma do erro  $\|x^{(k)} - \hat{x}\|$  (em que  $\hat{x}$  é a solução desejada do sistema livre de erros  $Ax = \hat{b}$ ) decresce, mas devido ao erro  $e = \hat{b} - b$  no vetor de dados e ao mal condicionamento de  $A$ , a norma do erro  $\|x^{(k)} - \hat{x}\|$  cresce rapidamente quando  $k$  é grande, isto é, a solução regularizada  $x^{(k)}$  começa a incorporar os componentes dos ruídos existentes e, assim, converge para uma solução não desejada que em geral é a solução de mínimos quadrados, dada por  $x_{LS} = A^\dagger b$  [13, 21]. Este fato geralmente é denominado de semiconvergência e as sequências  $\eta_n$  com essas características são chamadas de *sequências semiconvergentes*.

Um exemplo de sequência semiconvergente é  $s_1(t), s_2(t), \dots, s_3(t), \dots$ , obtida a partir da série  $s_k(t) = \frac{1}{t} - \frac{1!}{t^2} + \frac{2!}{t^3} + \dots + (-1)^{k-1} \frac{(k-1)!}{t^k} + \dots$ , para grandes valores positivos de  $t$ . Calculando os termos  $s_k(t)$  para, por exemplo,  $t = 10$ , observa-se que a sequência é aparentemente convergente até o termo  $k = 23$  e a partir de  $k = 29$  começa a divergir. Tal fato pode ser observado na figura 3.1. Também é importante notar que para  $k = 26$  tal sequência atinge um valor mínimo que, segundo Morigi [21], é usualmente escolhido como o melhor truncamento da série.

A idéia de semi-convergência também pode ser observada na figura 3.2, para as soluções iteradas do problema teste *Phillips*, quando  $n = 64$  e o erro relativo  $e = 1\%$ , utilizando o método LSQR.

Assim, nos métodos numéricos de regularização seguintes, por serem iterativos, além dos conceitos serão fornecidos critérios de parada adequados.

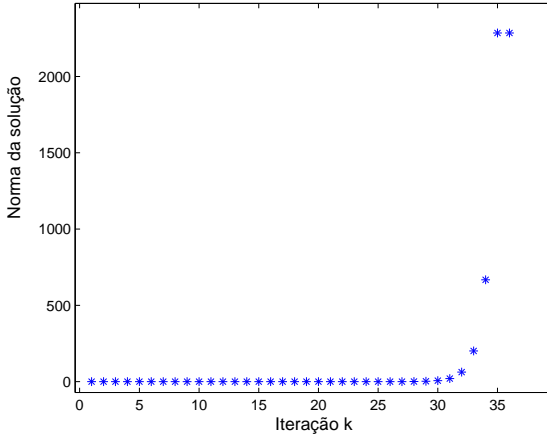


Figura 3.1: Semi-convergência da série  $\sum_{k=0}^n (-1)^{k-1} \frac{(k-1)!}{t^k}$ , com  $t = 10$ .

Recomenda-se o emprego dos métodos iterativos para problemas de grande porte, pelo fato que nos métodos iterativos a matriz  $A$  aparece na formação de multiplicações matriz-vetor com  $A$  e  $A^T$ , enquanto que nos métodos diretos o cálculo da SVD de  $A$  é inviável devido ao alto custo computacional.

## 3.2 LSQR

O LSQR (*least square QR*) é um método iterativo de regularização que está apoiado na fatoração QR e no processo de bidiagonalização de

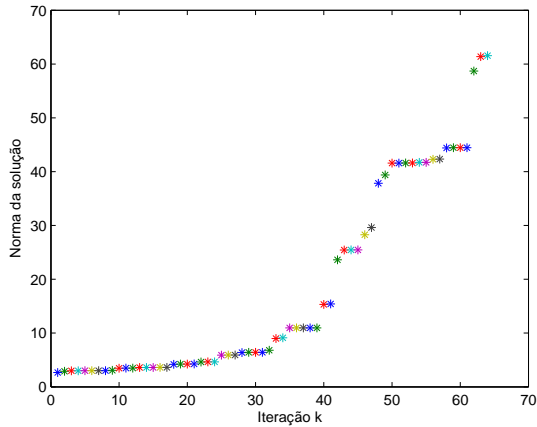


Figura 3.2: Semi-convergência do método LSQR, problema *Phillips*,  $n = 64$  e erro relativo de 1%.

Lanczos, para a resolução de  $Ax = b$  e de  $\operatorname{argmin}_{x \in R^n} \|b - Ax\|_2^2$ , quando a matriz  $A$  é grande e esparsa.

Resumidamente, a idéia do método LSQR pode ser descrita como: A matriz  $A$  em  $x = \operatorname{argmin}_{x \in R^n} \|b - Ax\|_2^2$  é bidiagonalizada segundo o processo de bidiagonalização de Lanczos ( $A = UBV^T$ , em que  $U$  e  $V$  são ortogonais). Baseado na matriz  $B$ , bidiagonal, o problema é reformulado e resolvido pela fatoração QR. Segundo Paige em [22], este método é algebricamente semelhante ao método do gradiente conjugado para equações normais, só que suas propriedades numéricas são mais eficientes, principalmente no caso da matriz  $A$  ser mal condicionada.

Desta forma, dado um sistema de equações lineares  $Ax = b$ , deve-

se bidiagonalizar a matriz  $A$ , através do processo de bidiagonalização de Lanczos. Para isto serão utilizadas as iterações dadas em (1.16) e (1.17), escolhendo o vetor unitário  $u_1 = \frac{b}{\|b\|}$  como uma iteração inicial (chute inicial) do processo iterativo de Lanczos e utilizando  $\beta_1 u_1 = b$  e  $\alpha_1 v_1 = A^T u_1$ .

Levando em conta também que se  $x_k$  é uma solução aproximada de  $x = \operatorname{argmin}_{x \in \mathbb{R}^n} \|b - Ax\|_2^2$  a ser determinada, com  $x_k \in \mathcal{K}_k(A^T A, A^T b)$ , isto é,  $x_k$  é uma combinação linear dos vetores  $v_k$ , então pode-se escrever

$$x_k = V_k y_k, \quad (3.1)$$

para algum  $y_k \in \mathbb{R}^k$ . O resíduo para um vetor  $x_k$  é dado por

$$\gamma_k = \beta_1 U_{k+1} e_1 - U_{k+1} B_k y_k. \quad (3.2)$$

Assim, utilizando (3.1) (3.2) a equação  $x = \operatorname{argmin}_{x \in \mathbb{R}^n} \|b - Ax\|_2^2$  será escrita, na iteração  $k$ , como

$$y_k = \operatorname{argmin}_{x \in \mathbb{R}^n} \|\beta_1 u_1 - A V_k y_k\|_2^2 \quad \text{ou}$$

$$y_k = \operatorname{argmin}_{x \in \mathbb{R}^n} \|\beta_1 U_{k+1} e_1 - U_{k+1} B_k y_k\|_2^2$$

ou, ainda,

$$y_k = \operatorname{argmin}_{x \in \mathbb{R}^n} \|\beta_1 e_1 - B_k y_k\|_2^2, \quad (3.3)$$

que forma a base do método LSQR.

Computacionalmente é vantajoso resolver a equação (3.3) utilizando a fatoração QR de  $B_k$ , pois  $B_k$  é bidiagonal. Assim, a fatoração QR toma a forma:

$$Q_k \quad B_k \quad \beta_1 e_1 = \begin{pmatrix} R_k & f_k \\ & \phi_{k+1} \end{pmatrix} =$$

$$= \begin{pmatrix} \rho_1 & \theta_1 & & & & & \vdots & \phi_1 \\ & \rho_2 & \theta_2 & & & & \vdots & \phi_2 \\ & & \cdot & \cdot & & & \vdots & \cdot \\ & & & \cdot & \cdot & & \vdots & \cdot \\ & & & & \cdot & \cdot & \vdots & \cdot \\ & & & & & \rho_{k-1} & \theta_{k-1} & \phi_{k-1} \\ & & & & & & \rho_k & \phi_k \\ \dots & \dots & \dots & \dots & \dots & \dots & \vdots & \dots \\ & & & & & & \vdots & \check{\phi}_{k+1} \end{pmatrix} \quad (3.4)$$

Para a determinação da matriz  $Q_k$  são aplicados produtos de rotações de Givens, chegando a

$$Q_k = G(k, k+1, \theta_k)G(1, 2, \theta_1) \dots G(k-1, k, \theta_{k-1}),$$

em que as matrizes  $G(i, k, \theta_i)$  têm o objetivo de eliminar os elementos  $\beta_2$  da matriz  $B_k$ .

As matrizes  $G(i, k, \theta)$  são as rotações de Givens e têm a forma:

$$G(i, k, \theta) = \begin{pmatrix} 1 & \dots & 0 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \dots & c & \dots & s & \dots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \dots & -s & \dots & c & \dots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & 0 & \dots & 1 \end{pmatrix} \begin{matrix} i \\ \\ k \\ \\ \end{matrix},$$

$i \qquad k$

sendo  $c = \cos(\theta)$  e  $s = \sin(\theta)$  para algum  $\theta$ . As rotações de Givens são ortogonais e a premultiplicação por  $G(i, k, \theta)^T$  executa uma rotação no sentido horário de  $\theta$  radianos no plano  $x_i x_k$ .

Logo a solução de  $x = \operatorname{argmin}_{x \in \mathbb{R}^n} \|b - Ax\|_2^2$  é dada a partir de

$$R_k y_k = f_k .$$

Observe que se  $y_k$  for uma solução de (3.3) o resíduo associado será dado por

$$\gamma_k = Q_k^T \begin{pmatrix} 0 \\ \check{\phi}_{k+1} \end{pmatrix} ,$$

e por conseguinte

$$\|\gamma_k\|_2 = |\check{\phi}_{k+1}| .$$

Agora, supondo conhecida a fatoração QR da matriz  $B_{k-1}$  e calculando a fatoração QR da matriz  $B_k$ , utilizando as rotações de Givens, chega-se a seguinte expressão para o resíduo

$$\|r_k\|_2 = |\check{\phi}_{k+1}| = |s_k \check{\phi}_k| = \dots = |s_k s_{k-1} \dots s_1 \beta_1| . \quad (3.5)$$

Substituindo  $y_k$  por  $y_k = R_k^{-1} f_k$  em (3.1) tem-se

$$x_k = V_k y_k = V_k (R_k^{-1} f_k) = (V_k R_k^{-1}) f_k = Z_k f_k , \quad (3.6)$$

em que  $Z_k$  satisfaz o sistema triangular inferior  $R_k^T Z_k^T = V_k^T$

**Lema 3.1.** *Sejam  $x_k$ , com  $k = 1, 2, \dots$  as soluções iteradas do método LSQR, então a norma do resíduo é uma função decrescente, isto é,  $\|r_{k+1}\|_2 \leq \|r_k\|_2$ ,  $k = 1, 2, \dots$ .*

**Demonstração:** É imediata a partir (3.5), pois pode-se escrever

$$\|\gamma_{k+1}\|_2 = |\check{\phi}_{k+2}| = |s_{k+1}\check{\phi}_{k+1}| \leq |\check{\phi}_{k+1}| = \|\gamma_k\|_2,$$

ficando provado que a norma do resíduo é decrescente. ■

**Lema 3.2.** *Sejam  $x_k$ , com  $k = 1, 2, \dots$  as soluções iteradas do método LSQR, então a norma da solução é uma função crescente, ou seja,  $\|x_{k+1}\|_2 \geq \|x_k\|_2$ ,  $k = 1, 2, \dots$ .*

**Demonstração:** A demonstração que a norma da solução é uma função crescente é feita reduzindo a matriz  $R_k$  a uma matriz bidiagonal  $\bar{B}_k$  empregando uma fatoração ortogonal  $R_k \bar{Q}_k^T = \bar{B}_k$ , sendo  $\bar{Q}_k$  o produto de rotações de Givens. A partir do sistema  $\bar{B}_k \bar{z}_k = f_k$  escreve-se  $x_k = (V_k R_k^{-1}) f_k = (V_k \bar{Q}_k^T) \bar{z}_k = \bar{W}_k \bar{z}_k$ . Mas como  $V_k$  é ortogonal vem  $\|x_k\|_2 = \|\bar{z}_k\|_2$ . Do fato que a parte superior das matrizes  $\bar{B}_k$ ,  $\bar{Q}_k$ ,  $\bar{W}_k$  e  $\bar{z}_k$  não muda após a iteração  $k$ , o vetor  $\bar{z}_k$  adquire um elemento a mais da  $k$ -ésima iteração para a iteração  $k + 1$ , isto permite escrever

$$\|x_k\|_2 = \|\bar{z}_k\|_2 \leq \|\bar{z}_{k+1}\|_2 = \|x_{k+1}\|_2.$$

■

### 3.3 GMRES e RRGMR

O método GMRES (*generalized minimal residual*), proposto por Saad e Schultz (1986), é um método iterativo utilizado para resolver sistemas de equações lineares  $Ax = b$ , com  $A \in \mathbb{C}^{m \times m}$  não singular e  $b \in \mathbb{C}^m$ .



Basicamente este método aproxima a solução  $x$  de  $Ax = b$  por um vetor  $x_n$  pertencente ao subespaço de Krylov,  $\mathcal{K}_n(A, b)$ , de modo que  $x_n$  minimize a norma do resíduo  $r_n = b - Ax_n$ .

Na figura 3.3, conforme [28], tem-se a representação geométrica do problema de mínimos quadrados utilizado na determinação da solução  $x_n$ .

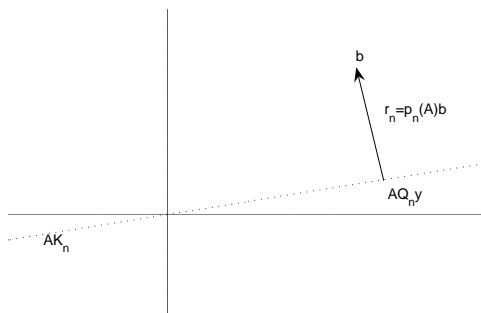


Figura 3.3: GMRES: minimização da norma do resíduo  $r_n$

Este método utiliza a iteração de Arnoldi, idealizada por W. E. Arnoldi em 1951, para reduzir uma matriz  $A \in \mathbb{C}^{n \times n}$  para a forma de Hessenberg, aqui representada pela matriz  $H$ , por uma transformação ortogonal, assim:

$$Q^T A Q = H \quad \text{ou} \quad A Q = Q H . \quad (3.7)$$

Tomando de modo particular  $Q = \begin{bmatrix} q_1 & q_2 & \dots & q_n \end{bmatrix}$ ,  $Q \in \mathbb{C}^{m \times n}$  e comparando as  $k$ -ésimas colunas de ambos os membros de  $AQ = QH$ ,

tem-se:

$$Aq_k = h_{1k}q_1 + h_{2k}q_2 + \dots + h_{kk}q_k + h_{k+1,k}q_{k+1} = \sum_{i=1}^{k+1} h_{ik}q_i .$$

Isolando o último termo do somatório anterior, vem

$$h_{k+1,k}q_{k+1} = Aq_k - \sum_{i=1}^k h_{ik}q_i := \tau_k \quad (3.8)$$

com  $h_{ik} = q_i^* Aq_k$  para  $i = 1, 2, \dots, k$ . Disto segue que se  $\tau_k \neq 0$ , então tem-se

$$q_{k+1} = \frac{\tau_k}{h_{k+1,k}}$$

com  $h_{k+1,k} = \|\tau_k\|_2$ . Estas equações definem o processo iterativo de Arnoldi.

O algoritmo de Arnoldi, conforme Trefethen em [28], para redução à forma de Hessenberg, é dado por:

$$b = \text{arbitrário}, q_1 = \frac{b}{\|b\|}$$

for  $k = 1, 2, \dots$

$$v = Aq_k$$

for  $j = 1 : k$

$$h_{jk} = q_j^T v$$

$$v = v - h_{jk}q_j$$

$$h_{k+1,k} = \|v\|$$

$$q_{k+1} = \frac{v}{h_{k+1,k}}$$

Os vetores  $q_n$  são chamados *vetores de Arnoldi* e eles definem uma base ortonormal para o subespaço  $\mathcal{K}_k(A, b)$  de Krylov, isto é,

$$\text{span}\{q_1, q_2, \dots, q_k\} = \text{span}\{q_1, Aq_1, \dots, A^{k-1}q_1\} .$$

Após  $k$  iterações do processo de fatoração de Arnoldi, as equações (3.7) e (3.8) ficam:

$$AQ_k = Q_k H_k + \tau_k e_k^T \quad (3.9)$$

em que  $Q_k = \begin{bmatrix} q_1 & q_2 & \cdots & q_k \end{bmatrix}$ ,  $e_k = I_k(:, k)$  e

$$H_k = \begin{pmatrix} h_{1,1} & h_{1,2} & \cdots & \cdots & h_{1,k} \\ h_{2,1} & h_{2,2} & \cdots & \cdots & h_{2,k} \\ 0 & h_{3,2} & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & h_{k,k-1} & h_{k,k} \end{pmatrix} .$$

A partir da fatoração de Arnoldi é possível agora efetuarmos um melhor estudo sobre o método GMRES. Seja  $K_k$  a matriz de Krylov, referente às matrizes  $A$  e  $b$ , do sistema linear  $Ax = b$ . Assim conforme (1.9) pode-se escrever

$$AK_k = \begin{bmatrix} Ab & A^2b & \cdots & A^k b \end{bmatrix} . \quad (3.10)$$

Desta maneira o problema reduz-se em determinar um vetor  $c \in \mathbb{R}^k$  tal que

$$\| AK_k c - b \| = \text{mínimo} \quad (3.11)$$

o que pode ser feito utilizando a fatoração QR de  $Ak_n$ . Determinado o valor do vetor  $c$ , chega-se a  $x_k = K_k c$  .

O procedimento descrito no parágrafo anterior é numericamente instável, por isto será usada a iteração de Arnoldi para a construção de uma sequência de matrizes de Krylov  $Q_k$  cujas colunas  $q_1, q_2, \dots$  geram os sucessivos espaços de Krylov  $\mathcal{K}_k(A, b)$ . Deste modo pode-se escrever  $x_k = Q_k y$  em vez de  $x_k = K_k c$ . Agora (3.11) torna-se

$$\| A Q_k y - b \| = \text{mínimo} \quad (3.12)$$

em que  $y \in \mathbb{C}^n$  é o vetor a ser determinado. Utilizando (3.9), com  $\tau_k = 0$ , em (3.12) chega-se a

$$\| Q_{k+1} H_n y - b \| = \text{mínimo} \quad (3.13)$$

Como a matriz  $Q_{k+1}$  é ortonormal pode-se escrever (3.13), sem alterar a sua norma em (3.13), como

$$\| H_k y - Q_{k+1}^T b \| = \text{mínimo} \quad (3.14)$$

Verificando que  $Q_{k+1}^T b = \| b \| e_1$ , pela construção das matrizes  $Q_k$  de Krylov, em que  $e_1 = (1, 0, 0, \dots)^T$ , a expressão (3.14), que é a forma final do GMRES, torna-se

$$\| H_n y - \| b \| e_1 \| = \text{mínimo} . \quad (3.15)$$

Na  $k$ -ésima iteração do método GMRES, o valor de  $y$  é determinado e a  $k$ -ésima solução aproximada construída pelo método GMRES é dada por: solução

$$x_k = Q_k y. \quad (3.16)$$

Com respeito aos lemas 3.1 e 3.2 do método LSQR, para o método GMRES apenas o lema 3.1 continua sendo verdadeiro, isto é,  $\| r_{k+1} \|_2 \leq \| r_k \|_2, k = 1, 2, \dots, n-1$ . Enquanto o lema 3.2 pode ou não ser verificado.

Quanto ao método RRGMRÉS, este se diferencia do GMRES exclusivamente pelo fato de que no RRGMRÉS o vetor  $x_k$ , da  $k$ -ésima iteração, em vez de pertencer ao subespaço  $\mathcal{K}_k(A, b)$  de Krylov pertence a um novo subespaço  $\mathcal{K}_k(A, Ab)$  de Krylov. Ou seja, o chute inicial do RRGMRÉS é o vetor unitário na direção de  $Ab$ , enquanto no GMRES é na direção de  $b$  [6].

Na prática, tem sido visto que o método RRGMRÉS produz melhores aproximações do que o GMRES quando o termo independente  $b$  está contaminado por ruídos e a solução procurada é suave. Quando a matriz  $A$  é simétrica, as iterações são ortogonais ao espaço nulo da  $A$ . Para matrizes  $A$  grandes, o trabalho para calcular  $x_k$  está na avaliação de  $k + 1$  produtos de matriz vetor por  $A$  [21].

### 3.4 Critérios de parada para métodos iterativos

Nesta seção serão descritos os critérios de parada usados neste trabalho conforme propostos por Morigi [21]. O seguinte resultado (em que  $\hat{x}$  denota a solução desejada do sistema de equações lineares consistente e livre de erros  $Ax = \hat{b}$ , ou seja,  $\hat{b}$  também é livre de erros) sugere um critério de parada das iterações nos métodos GMRES e LSQR:

**Teorema 3.3.** *Seja  $x_k$  a  $k$ -ésima solução aproximada do sistema  $Ax = \hat{b}$*

e  $\hat{k}$  o primeiro índice  $k$  tal que

$$\hat{k} = \operatorname{argmin}\{\|x_k - \hat{x}\|_2\}. \quad (3.17)$$

Então

$$\|x_k - \hat{x}\| < \|x_{k-1} - \hat{x}\| \Rightarrow \|x_{k-1} - \hat{x}\| > \frac{1}{2} \|x_k - x_{k-1}\|, \quad (3.18)$$

$$\|x_k - \hat{x}\| \leq \|x_{k+1} - \hat{x}\| \Rightarrow \|x_{k+1} - \hat{x}\| \geq \frac{1}{2} \|x_{k+1} - x_k\|. \quad (3.19)$$

Além disso, seja  $\rho_{\hat{k}} = \frac{\|x_{\hat{k}+1} - \hat{x}\|}{\|x_{\hat{k}} - \hat{x}\|}$ . Então:

$$\|x_{\hat{k}} - \hat{x}\| \geq \frac{1}{1 + \rho_{\hat{k}}} \|x_{\hat{k}+1} - x_{\hat{k}}\|. \quad (3.20)$$

**Demonstração:** Visto que  $\|x_k - x_{k-1}\| = \|x_k - \hat{x} + \hat{x} - x_{k-1}\|$ , usando a desigualdade triangular vem:

$$\|x_k - x_{k-1}\| \leq \|x_k - \hat{x}\| + \|\hat{x} - x_{k-1}\|.$$

Utilizando a hipótese, tem-se  $\|x_k - x_{k-1}\| < 2 \|x_{k-1} - \hat{x}\|$ , que é (3.18). A prova de (3.19) é feita analogamente.

Por outro lado, usando novamente a desigualdade triangular tem-se

$$\begin{aligned} \|x_{\hat{k}+1} - x_{\hat{k}}\|_2 &\leq \|x_{\hat{k}+1} - \hat{x}\|_2 + \|\hat{x} - x_{\hat{k}}\|_2 \\ &= (1 + \rho_{\hat{k}}) \|\hat{x} - x_{\hat{k}}\|_2, \end{aligned}$$

que prova (3.20). ■

Muitas iterações satisfazem a equação (3.18). Em particular, a iteração  $x_{\hat{k}}$  definida por  $\|x_{\hat{k}} - \hat{x}\| = \min_{k \geq 0} \|x_k - \hat{x}\|$  satisfaz as equações (3.18) e (3.19) com  $k = \hat{k}$ . A constante  $\rho_{\hat{k}} \geq 1$  está próxima da unidade para

muitos sistemas da forma  $Ax = b$ . Por isso  $\frac{1}{2} \|x_{k+1} - x_k\|$  sustenta um limite inferior de  $\|x_k - \hat{x}\|$  para várias iterações  $x_k$ . Isto sugere que uma iteração  $x_{k'}$ , tal que  $\|x_{k'+1} - x_{k'}\|$  é pequeno, pode garantir uma adequada aproximação de  $\hat{x}$ . Especificamente seja  $k'$  o menor índice, tal que

$$\|x_{k'+1} - x_{k'}\| = \min_{k \geq 0} \|x_{k+1} - x_k\| . \quad (3.21)$$

Experimentos numéricos mostrados em [21] sugerem que  $x_{k'}$ , em (3.21), é em geral uma razoável aproximação de  $\hat{x}$ . No entanto, é possível obter uma melhor aproximação de  $\hat{x}$  do que de  $x_{k'}$ , através de uma modificação no critério (3.21), levando-se em conta a incorporação de uma informação auxiliar.

Seja  $\mathbb{M}$  o conjunto dos índices  $k$  para os quais a função

$$k \rightarrow \|x_{k+1} - x_k\| \quad (3.22)$$

atinge um mínimo local. Assim, escolhe-se uma iteração  $x_k$  com  $k \in \mathbb{M}$  com o auxílio de um outro critério, tal como o princípio da discrepância ou o da curva-L, e assume-se que uma estimativa de  $\|e\|$  seja adequada. Então, o princípio da discrepância sugere que a primeira iteração  $x_k$  que satisfaça

$$\|Ax_k - b\| \leq \eta \|e\| \quad (3.23)$$

seja escolhida como uma aproximação de  $\hat{x}$ , onde  $\eta \geq 1$ . Seja  $k_d$  o índice da iteração determinada pelo princípio da discrepância. Pode-se então selecionar a iteração  $x_{k'_d}$ , em que  $k'_d \in \mathbb{M}$  e  $k'_d \geq k_d$ , como uma aproximação para  $\hat{x}$ .

Alternativamente, o critério da curva L também pode ser utilizado para

determinar a iteração de parada. Neste caso, a sugestão de Morigi [21] é escolher  $k_L$  como o índice do ponto do vértice da curva-L. A dificuldade aqui é que o vértice da curva-L pode ser de difícil determinação.

Assim, no capítulo 4, o critério de parada utilizado para os métodos LSQR e GMRES é o que foi apresentado anteriormente, isto é, o algoritmo para na iteração  $k$  quando  $\|x_{k+1} - x_k\|$  atinge um mínimo local e, simultaneamente, satisfaz a condição  $\|Ax_k - b\| \leq \eta\|e\|$ , em que  $\eta$  foi tomado como (1.01).

### 3.5 Um novo critério de parada para problemas discretos mal-postos

Conforme os lemas 3.1 e 3.2, a função  $\psi_\mu(\lambda)$  empregada como base para o método de ponto fixo é o produto de duas quantidades, uma crescente e a outra decrescente. Consequentemente, o parâmetro de regularização é determinado ao buscar-se obter o mínimo da função  $\psi$ . Assim, considerando a sequência  $\Psi_k$  definida por

$$\Psi_k = \|r_k\|_2 \|x_k\|_2, \quad k = 1, 2, \dots \quad (3.24)$$

e o fato de que  $\|r_k\|_2$  decresce com  $k$  enquanto  $\|x_k\|_2$  cresce (ou seja, a sequência  $\Psi_k$  se comporta analogamente como a função contínua  $\psi_1(\lambda)$ , dada em (2.16)) aqui se propõe como critério de parada escolher a iteração de parada  $\widehat{k}$  como o primeiro índice que minimiza a sequência (3.24). Isto é,

$$\widehat{k} = \operatorname{argmin}\{\Psi_k\}, \quad k = 1, 2, \dots \quad (3.25)$$



A figura 3.4 ilustra o comportamento do erro relativo associado a cada iteração do método LSQR, bem como da sequência  $\Psi_k$ .

Quando o mínimo de  $\Psi_k$  é bem visível o critério funciona bem e minimizador de  $\Psi_k$  é o mesmo que minimiza o erro  $\|x_k - \hat{x}\|_2$  o muito próximo dele, veja Figura 3.4. Porém, para alguns problemas onde o mínimo de  $\Psi_k$  se encontra num *patamar*, o processo iterativo pode-se estender desnecessariamente. Para estes casos usa-se a sugestão de J. Chung [8, 4] e escolhe-se a iteração de parada como o primeiro  $k$  que satisfaz a condição

$$|\psi_{k+1} - \psi_k| < \psi_1 \epsilon, \quad (3.26)$$

onde  $\epsilon$  é alguma tolerância.

Em síntese, o critério de parada aqui proposto escolhe o primeiro  $k$  tal que  $\psi_k$  é minimizado ou o  $k$  que satisfaz a condição (3.26).

No capítulo 4, em todos os problemas testados este novo critério de parada parou na iteração ótima ou ficou muito próximo desta. Em comparação com os outros dois critérios de parada aqui adotados, o de Morigi e o da discrepância, dentre os 30 resultados obtidos este novo critério apresentou melhor ou igual solução em 80% dos casos. Nos únicos três casos em que o novo critério se mostrou menos eficiente, tal fato ocorreu porque a função  $\psi_k$  possui um trecho em que o produto da norma de cada solução iterada pela norma do resíduo correspondente permanece quase constante, com o aumento de  $k$ .

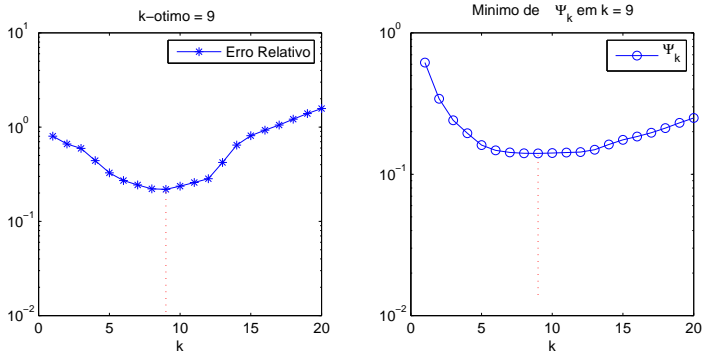


Figura 3.4: Erro relativo no passo  $k$ :  $\|x_k - x_{\text{exato}}\|_2 / \|x_{\text{exato}}\|_2$  para LSQR, problema *Heat*,  $n = 256$  e erro nos dados de 5% (esquerda). Comportamento da seqüência  $\Psi_k$  definida em (3.24) (direita).

## Capítulo 4

# Problemas testes e resultados numéricos

Neste capítulo são descritos e resolvidos alguns problemas teste da literatura a fim de comparar a eficiência dos métodos de regularização descritos nos capítulos 2 e 3. Estes problemas teste estão disponíveis no pacote *Regularization Tools* [15].

## 4.1 Descrição de problemas teste

### 4.1.1 Heat

Segundo Hansen em [15] a equação inversa *Heat* utilizada neste problema teste é a equação integral de Volterra de primeira ordem com  $[0, 1]$  sendo o intervalo de integração. Este problema foi discretizado com o emprego do método da quadratura.

Neste problema teste o núcleo da equação integral de Volterra, em (4), é dado por:

$$K(s, t) = k(s - t)$$

com

$$k(t) = \frac{t^{-3/2}}{2kappa\sqrt{\pi}} \exp \frac{1}{kappa^2 t} \quad .$$

Aqui o parâmetro *kappa* controla o mal condicionamento da matriz *A*:

*kappa* = 5 para uma matriz bem condicionada

*kappa* = 1 para uma matriz mal condicionada.

A figura 4.1 representa a solução exata do problema *Heat*, no caso em que a matriz é de ordem  $32 \times 32$ .

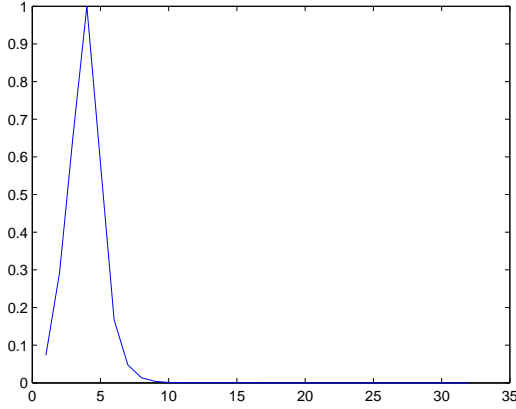


Figura 4.1: Solução  $x_{exata}$  para o problema *Heat* de dimensão  $n = 32$ ,  $\kappa = 1$ .

### 4.1.2 Baart

O problema teste *Baart* é, conforme [15], uma discretização da equação integral de primeira ordem de Fredholm, dada em (2), por meio do método de Garlekin, em que o lado direito  $g$  e o núcleo  $K$  são dados por

$$K(s, t) = \exp(s \cos(t)), \quad g(s) = 2 \frac{\sin(s)}{s},$$

com intervalos de integração  $s \in [0, \frac{\pi}{2}]$  e  $t \in [0, \pi]$ . A solução é dada por  $f(t) = \sin(t)$ .

A figura 4.2 representa a solução exata do problema *Baart*, no caso em

que a matriz é de ordem  $32 \times 32$ .

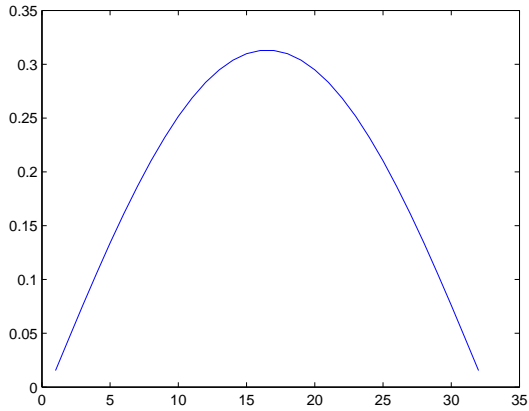


Figura 4.2: Solução  $x_{exata}$  para o problema *Baart* de dimensão  $n = 32$ .

### 4.1.3 Deriv2

O problema teste *Deriv2* é, segundo [15], um problema mal posto brando, isto é, os valores singulares decrescem lentamente para zero. Este problema é uma discretização da equação integral de primeira ordem de Fredholm, realizada com o método de Garlekin, onde em (2) o lado direito  $g$  e o núcleo  $K$ , que é a função de Green para a segunda derivada, dado

por

$$K(s, t) = \begin{cases} s(t-1), & s < t \\ t(s-1), & s \geq t \end{cases}$$

são conhecidos, sendo  $f$  a solução a ser determinada.

Ambos os intervalos de integração em (2), para  $s$  e para  $t$ , são  $[0, 1]$ , em que  $g$  e a correspondente solução  $f$  podem ser escolhidas entre as seguintes opções:

$$(1) \quad g(s) = \frac{s^3-s}{6}, \quad f(t) = t,$$

$$(2) \quad g(s) = \exp(s) + (1-e)s - 1, \quad f(t) = \exp(t),$$

$$(3) \quad g(s) = \begin{cases} \frac{(4s^3-3s)}{24}, & s < \frac{1}{2} \\ \frac{(-4s^3+12s^2-9s+1)}{24}, & s \geq \frac{1}{2} \end{cases}, \quad f(t) = \begin{cases} t, & t < \frac{1}{2} \\ 1-t, & t \geq \frac{1}{2} \end{cases}$$

A figura 4.3 representa a solução exata do problema *Deriv2*, no caso em que a matriz é de ordem  $32 \times 32$ .

#### 4.1.4 Foxgood

Este é um problema teste severamente mal posto que não satisfaz a condição discreta de Picard para valores singulares muito pequenos. Ele é uma discretização da equação integral de Fredholm de primeira ordem, dada em (2), para um sistema discreto de equações lineares  $Ax = b$ , em que  $A$  é uma matriz  $n \times n$ , com intervalos de integração iguais a  $[0, 1]$ , cujo núcleo  $K$  e o segundo membro  $g$  são dados respectivamente por:

$$K(s, t) = (s^2 + t^2)^{\frac{1}{2}},$$

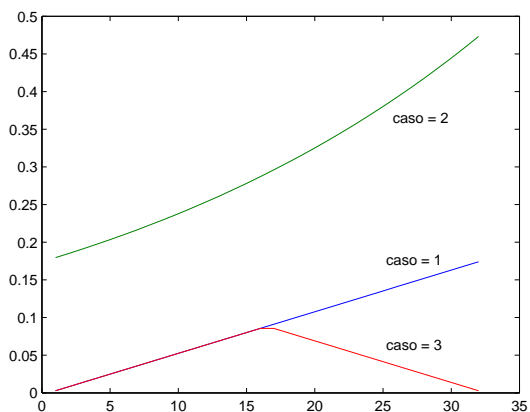


Figura 4.3: Solução  $x_{exata}$  para o problema *Deriv2* de dimensão  $n = 32$ , nos 3 casos previstos.

$$g(s) = \frac{1}{3} (1 + s^2)^{\frac{3}{2}} - s^3 \quad ,$$

e com solução  $f = t$ . A discretização foi realizada com o método da quadratura. (Cf. Hansen, em [15])

Usando o *software* MATLAB para calcular o número de condição da matriz  $A$  para diferentes parâmetros de discretização, tem-se:



n	número de condição
2	9.9
10	$1.9 \times 10^{10}$
20	$3.2 \times 10^{17}$
50	$9.1 \times 10^{19}$

A figura 4.4 representa a solução exata do problema *Foxgood*, no caso em que a matriz é de ordem  $32 \times 32$ .

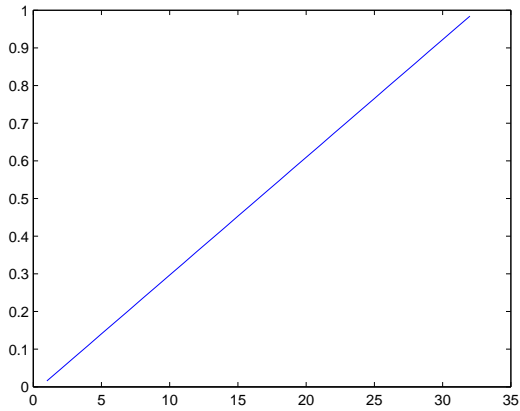


Figura 4.4: Solução  $x_{exata}$  para o problema *Foxgood* de dimensão  $n = 32$ .

### 4.1.5 Gravity

Trata-se da discretização, realizada com o método da quadratura, de um problema uni-dimensional em estudos gravitacionais, no qual a distribuição de massa  $f(t)$  está situada a uma profundidade  $d$ , enquanto a componente vertical do campo de gravidade é medido sobre a superfície. O problema resultante é uma equação integral de Fredholm de primeira espécie cujo núcleo é dado por:

$$k(s, t) = d(d^2 + (s - t)^2)^{-3/2} \quad .$$

Os seguintes exemplos estão implementados em [15]:

- (1)  $f(t) = \sin(\pi t) + 0,5\sin(2\pi t)$ ,
- (2)  $f(t) =$  função linear por partes,
- (3)  $f(t) =$  função constante por partes.

O intervalo de integração para  $t$  é fixado como  $[0, 1]$ , enquanto que o intervalo de integração de  $s$  é  $[a, b]$ . O intervalo usual para  $s$  é também  $[0, 1]$ , conduzindo para uma matriz simétrica de Toeplitz. O parâmetro  $d$  considerado padrão é igual a 0,25.

A figura 4.5 representa a solução exata do problema *Gravity*, no caso em que a matriz é de ordem  $32 \times 32$ .

### 4.1.6 Ilaplace

É a discretização da transformada inversa de Laplace (uma equação integral de Fredholm de primeira espécie) por meio da quadratura de Gauss-

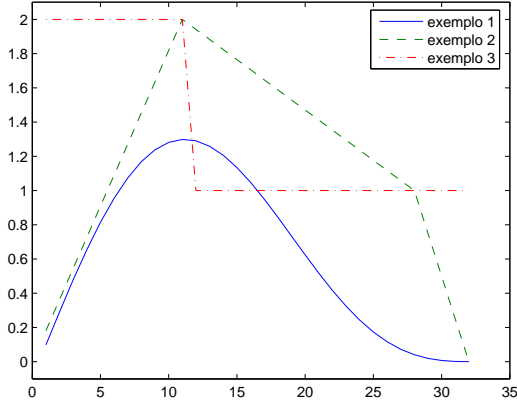


Figura 4.5: Solução  $x_{exata}$  para o problema *Gravity* de dimensão  $n = 32$ .

Laguerre. O núcleo  $K$ , em (2), é dado por

$$K(s, t) = \exp(-st)$$

e ambos os intervalos de integração são  $[0, \infty)$ .

Os seguintes exemplos estão implementados em [15], onde  $f$  denota a solução e  $g$  denota o correspondente lado direito de (2):

- (1)  $f(t) = \exp(-t/2)$ ,  $g(s) = \frac{1}{s + 1/2}$ ,
- (2)  $f(t) = 1 - \exp(-t/2)$ ,  $g(s) = \frac{1}{s} - \frac{1}{s + 1/2}$ ,
- (3)  $f(t) = t^2 \exp(-t/2)$ ,  $g(s) = \frac{2}{(s + 1/2)^3}$ ,

$$(4) \quad f(t) = \begin{cases} 0, & t \leq 2 \\ 1, & t > 2 \end{cases}, \quad g(s) = \frac{\exp(-2s)}{s}$$

A figura 4.6 representa a solução exata do problema *L\_laplace*, no caso em que a matriz é de ordem  $32 \times 32$ .

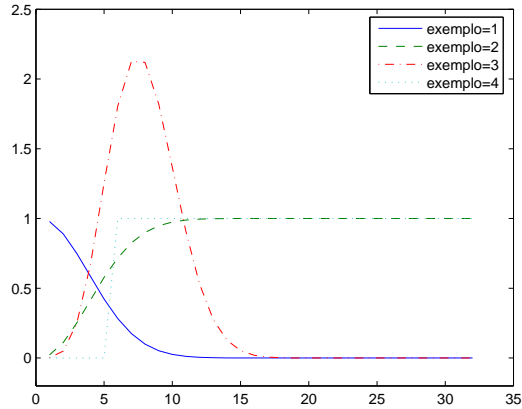


Figura 4.6: Solução  $x_{exata}$  para o problema *L\_laplace* de dimensão  $n = 32$ .

### 4.1.7 Phillips

Este problema teste, esquematizado por D. L. Phillips, é descrito em [15] como uma discretização, mediante o uso do método de Garlekin, da equação integral de Fredholm de primeira ordem, dada em (2), com núcleo suave, gerando um sistema de equações lineares  $Ax = b$ , com  $A \in \mathbb{R}^{m \times n}$ ,

$x \in \mathbb{R}^n$  e  $b \in \mathbb{R}^m$ .

Neste problema severamente mal condicionado, a matriz  $A$  possui posto mal determinado, podendo até  $A$  ser singular.

Define-se a função

$$\phi(x) = \begin{cases} 1 + \cos \frac{\pi x}{3}, & |x| < 3 \\ 0, & |x| \geq 3 \end{cases}.$$

Então, o núcleo  $K$ , a solução  $f$  e a função  $g$  em (2) são respectivamente dados por:

$$K(s, t) = \phi(s - t)$$

$$f(t) = \phi(t)$$

$$g(s) = (6 - s) \left( 1 + \frac{1}{2} \cos \frac{\pi s}{3} \right) + \frac{9}{2\pi} \sin \frac{\pi |s|}{3},$$

com ambos os intervalos de integração iguais a  $[-6, 6]$ .

A figura 4.7 representa a solução exata do problema *Phillips*, no caso em que a matriz é de ordem  $32 \times 32$ .

#### 4.1.8 Shaw

Em [15], este problema teste é descrito como uma discretização, pelo método da quadratura, da equação integral de Fredholm de primeira ordem, dada em (2), com  $[-\frac{\pi}{2}, \frac{\pi}{2}]$  em ambos os intervalos de integração. A equação integral é um modelo uni-dimensional de um problema de reconstrução de imagens.

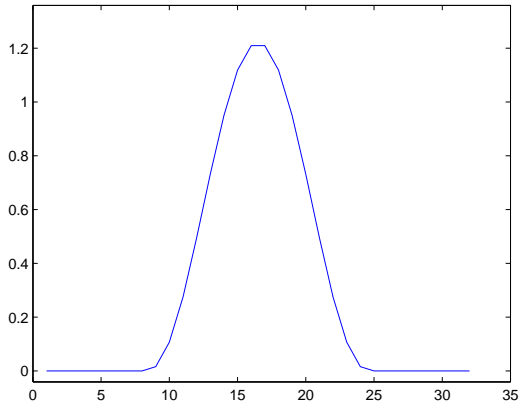


Figura 4.7: Solução  $x_{exata}$  para o problema *Phillips* de dimensão  $n = 32$ .

O núcleo  $K$  e a solução  $f$  são dados por:

$$K(s, t) = \cos(s) + \cos(t) - 2 \left( \frac{\sin(u)}{u} \right)^2, \text{ com}$$

$$u = \pi \sin(s) + \sin(t) \text{ e}$$

$$f(t) = \alpha_1 \exp -c_1(t - t_1)^2 + \alpha_2 \exp -c_2(t - t_2)^2 .$$

Os parâmetros  $\alpha_1, \alpha_2, \dots$  são constantes que determinam a forma da solução  $f$  e, para implementação, foram utilizados  $\alpha_1 = 2, \alpha_2 = 1, c_1 = 6, c_2 = 2, t_1 = 0.8$  e  $t_2 = -0.5$  resultando uma função  $f$  com dois máximos locais.

O núcleo e a solução são discretizados colocando-se  $n$  pontos para produzir  $A$  e  $x$ . Então o lado direito discreto  $g(s)$  em (2) é representado por  $b = Ax$ .

O gráfico da figura 4.8 representa a solução do problema teste *Shaw* para uma matriz de ordem  $32 \times 32$ .

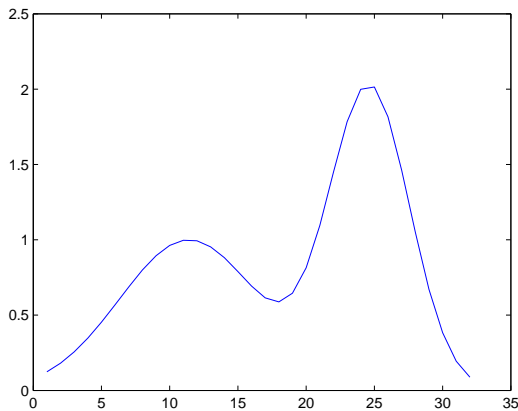


Figura 4.8: Solução  $x_{exata}$  para o problema *Shaw* de dimensão  $n = 32$ .

### 4.1.9 Tomo

Este é um problema teste de tomografia bi-dimensional, que foi discretizado através do método da quadratura [15].

Um domínio bi-dimensional  $[0, n] \times [0, n]$  é dividido em  $n^2$  células de dimensão unitária e um total de  $n^2$  raios em direções aleatórias penetram

neste domínio. A matriz  $A$  é esparsa.

A figura 4.9 representa a solução do problema teste *Tomo* para uma matriz de ordem  $32 \times 32$ .

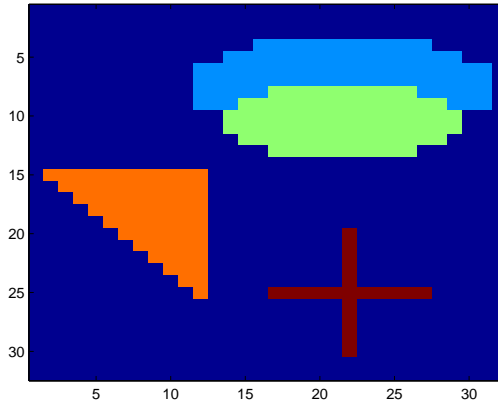


Figura 4.9: Solução  $x_{\text{exata}}$  para o problema *Tomo* de dimensão  $n = 32$ .

#### 4.1.10 Wing

Em [15] o problema teste *Wing* é conceituado como uma discretização, realizada pelo método de Garlekin, da equação integral de Fredholm de primeira ordem, dada em (2), com ambos os intervalos de integração iguais a  $[0, 1]$ , com o núcleo e o lado direito dados por:

$$K(s, t) = t \exp(-st^2) ,$$



$$g(s) = \frac{\exp(-st_1^2) - \exp(-st_2^2)}{2s},$$

e cuja solução  $f$  é dada por

$$f(t) = \begin{cases} 1, & t_1 < t < t_2 \\ 0, & \text{nos outros casos.} \end{cases}$$

Aqui,  $t_1$  e  $t_2$  são constantes satisfazendo  $0 < t_1 < t_2 < 1$ . Caso estas constantes não sejam especificadas, são utilizados os valores  $t_1 = \frac{1}{3}$  e  $t_2 = \frac{2}{3}$ .

A figura 4.10 representa a solução do problema teste *Wing* para uma matriz de ordem  $32 \times 32$ .

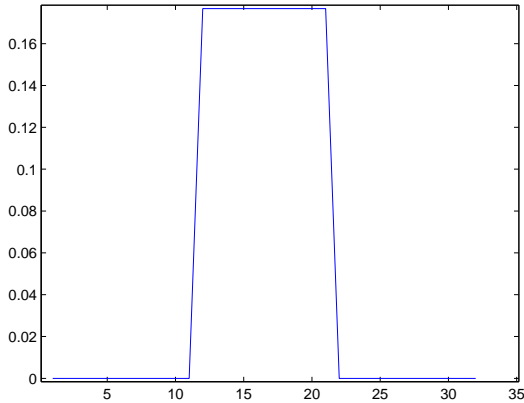


Figura 4.10: Solução  $x_{exata}$  para o problema *Wing* de dimensão  $n = 32$ ,  $t_1 = 1/3$  e  $t_2 = 2/3$ .

## 4.2 Resultados numéricos

A aplicação dos métodos da GCV, da curva-L, do ponto fixo, da quase-otimalidade, da discrepância, da LSQR, GMRES e RRGMRES aos problemas testes *Heat*, *Baart*, *Deriv2*, *Foxgood*, *Gravity*, *Ll- place*, *Phillips*, *Shaw*, *Tomo* e *Wing*, permitem concluir sobre a vantagem ou desvantagem de um método em relação aos outros, em determinado tipo de problema. Nesta seção, são apresentados alguns resultados numéricos que ilustram a eficiência dos métodos de regularização diretos e dos iterativos, abordados neste trabalho. Para os métodos diretos, a determinação dos parâmetros de regularização foi feita usando rotinas do pacote *Regularization Tools* [15]. Para o método iterativo LSQR, três critérios de parada foram utilizados: o de Morigi, o da discrepância e o proposto por nós no capítulo anterior. Nos resultados numéricos, esses 3 critérios são identificados pela notação *Crit1*, *Crit2*, e *Crit3*, respectivamente. Já para o método RRGMRES, apenas *Crit1*, *Crit2* são utilizados, visto que a propriedade descrita no lema 3.1 que fundamenta o *Crit3*, não é satisfeita para o GMRES nem para o RRGMRES.

Em  $Ax = b$  de cada um dos problemas testes discretizados, o vetor de dados  $b$  está sendo perturbado por um ruído  $e$ , tal que  $\tilde{b} = b + e$ , em que  $e$  é um vetor randômico não nulo. Os níveis de erros relativos utilizados foram de 0, 1%, a 1% e a 5% aplicados no vetor de dados  $b$ . A medida do desempenho de cada um dos métodos diretos será feita pela comparação do erro relativo, dado por

$$ER_o = \frac{\|x_{exato} - x_\lambda\|}{\|x_{exato}\|},$$

em que  $x_{exato}$  é a solução exata de cada problema e  $x_\lambda$  é a solução aproximada para o problema. Nos métodos iterativos, além da comparação do erro relativo será comparada a iteração de parada com a iteração ótima.

Para se ter melhores condições de analisar a performance de cada um dos métodos de regularização utilizados neste trabalho, cada problema teste foi resolvido 50 vezes, ou seja, foram gerados 50 vetores randômicos e distintos. Os resultados encontram-se nas tabelas 4.1 a 4.20, onde foram utilizadas as seguintes notações:

$ER$	Erro relativo médio
$ER_{max}$	Erro relativo máximo
$ER_{min}$	Erro relativo mínimo
$ER_o$	Erro relativo ótimo
$\lambda$	Parâmetro de regularização médio
$\lambda_{max}$	Parâmetro de regularização máximo
$\lambda_{min}$	Parâmetro de regularização mínimo
$It$	Iteração máxima de parada
$\hat{k}$	Iteração máxima ótima
$NR$	Nível de ruído
MOR	Critério de parada de Morigi
DP	Critério de parada da Discrepância
Min_psi	Novo critério de parada

### 4.2.1 Heat

Neste problema teste foi utilizada uma matriz  $A \in \mathbb{R}^{512 \times 512}$ . de número de condição  $5.0036e + 070$ . Os resultados estão nas tabelas 4.1 e 4.2. Na figura 4.11 encontra-se a solução regularizada,  $x_\lambda$ , obtida pelo método LSQR.

Com respeito aos métodos diretos, o método do ponto fixo forneceu melhores soluções para baixos níveis de erros  $NR$  no vetor  $b$ . O aumento desses níveis de erros fez com que todos os métodos originassem soluções semelhantes, exceto o da quase otimalidade que apresentou soluções inadequadas em relação ao erro relativo médio  $ER$ , sob quaisquer níveis de erros  $NR$ . Já em relação aos métodos iterativos, o LSQR com o novo critério de parada mostrou-se mais eficiente, tanto em relação ao erro relativo médio como em relação à iteração ótima, independente do aumento do nível de erro  $NR$  no vetor  $b$ .

Tabela 4.1: Comparação dos resultados dos métodos da Curva-L, Discrepância, Ponto fixo, Quase otimalidade e GCV, para o problema *Heat*.

	Curva-L	Discrep.	Ponto fixo	Quase otim	GCV
	NR = 0.1%				
<i>ER</i>	0.0926	0.1098	0.0296	9.6246e+048	0.0426
<i>ER<sub>max</sub></i>	0.1214	0.1365	0.0373	2.7949e+049	0.0710
<i>ER<sub>min</sub></i>	0.0678	0.0856	0.0235	2.3577e+047	0.0257
$\lambda$	0.00021	0.00018	0.001007	3.2660e-063	0.000478
$\lambda_{max}$	0.00022	0.00018	0.00114	3.2660e-063	0.00055
$\lambda_{min}$	0.00019	0.00017	0.00086	3.2660e-063	0.00031
	NR = 1%				
<i>ER</i>	0.1060	0.0992	0.0866	9.6246e+049	0.1206
<i>ER<sub>max</sub></i>	0.1413	0.1325	0.1152	2.7949e+050	0.3430
<i>ER<sub>min</sub></i>	0.0774	0.0699	0.0656	2.3577e+048	0.0712
$\lambda$	0.00169	0.001855	0.003439	3.2660e-063	0.001633
$\lambda_{max}$	0.00181	0.00189	0.00409	3.2660e-063	0.00198
$\lambda_{min}$	0.00154	0.00182	0.00280	3.2660e-063	0.00068
	NR = 5%				
<i>ER</i>	0.1700	0.1950	0.1782	4.8123e+050	0.2108
<i>ER<sub>max</sub></i>	0.2270	0.2363	0.2364	1.3975e+051	0.3549
<i>ER<sub>min</sub></i>	0.1197	0.1617	0.1155	1.1789e+049	0.1333
$\lambda$	0.00698	0.0103	0.00854	3.2660e-063	0.00424
$\lambda_{max}$	0.00757	0.01055	0.01054	3.2660e-063	0.00540
$\lambda_{min}$	0.00633	0.00998	0.00643	3.2660e-063	0.00281

Tabela 4.2: Comparação dos resultados dos métodos LSQR e RRGMRRES, para o problema *Heat*.

	LSQR			RRGMRES	
	MOR	DP	Min_Psi	MOR	DP
	NR = 0.1%				
$ER$	0.6642	0.0382	0.0362	0.8029	89.2476
$ER_o$	0.0268	0.0268	0.0268	1.0755	1.0755
$It$	2	17	18	1	99
$\hat{k}$	22	22	22	1	1
	NR = 1%				
$ER$	0.6642	0.1024	0.0817	0.8029	520.0721
$ER_o$	0.0759	0.0759	0.0759	1.0755	1.0755
$It$	2	11	15	1	99
$\hat{k}$	16	16	16	1	1
	NR = 5%				
$ER$	0.6642	0.2399	0.1995	0.8030	134.2660
$ER_o$	0.1632	0.1632	0.1632	1.0765	1.0765
$It$	2	7	9	1	79
$\hat{k}$	12	12	12	1	1

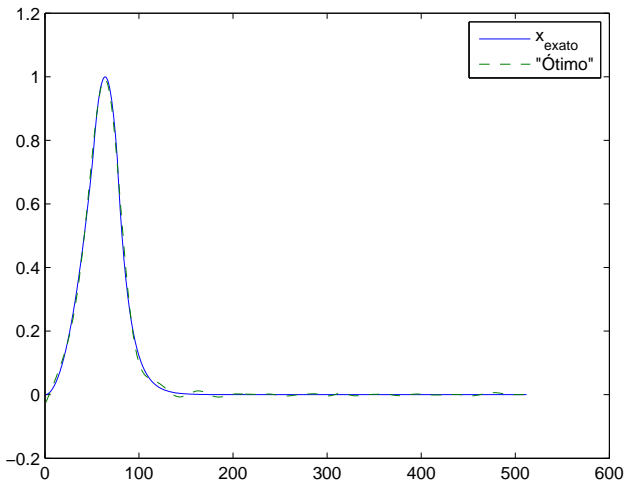


Figura 4.11: Solução exata,  $x_{exata}$ , e solução regularizada,  $x_{\lambda}$ , para o problema *Heat*, com erro relativo de 0.1% em  $b$ .

## 4.2.2 Baart

As comparações seguintes, referentes ao problema teste *Baart*, correspondem a uma matriz  $A \in \mathbb{R}^{512 \times 512}$  que possui um número de condição igual a  $3.2176e + 018$ , e os resultados se encontram nas tabelas 4.3 e 4.4.

Tabela 4.3: Comparação dos resultados dos métodos da Curva-L, Discrepância, Ponto fixo, Quase otimalidade e GCV, para o problema *Baart*.

	Curva-L	Discrep.	Ponto fixo	Quase otim	GCV
NR = 0.1%					
<i>ER</i>	0.1147	0.1172	0.1211	0.1234	6.6386e+003
<i>ER<sub>max</sub></i>	0.1449	0.1310	0.1464	0.1586	3.3085e+005
<i>ER<sub>min</sub></i>	0.1010	0.1079	0.0967	0.1053	0.0524
$\lambda$	0.0014	0.0023	0.0034	0.0037	0.0008
$\lambda_{max}$	0.0019	0.0023	0.0073	0.0148	0.0014
$\lambda_{min}$	0.0011	0.0023	0.0010	0.0004	2.1494e-010
NR = 1%					
<i>ER</i>	0.1595	0.1655	0.1583	0.1704	6.6379e+004
<i>ER<sub>max</sub></i>	0.1848	0.1812	0.1913	0.4097	3.3085e+006
<i>ER<sub>min</sub></i>	0.1311	0.1403	0.1111	0.1239	0.1102
$\lambda$	0.0137	0.0237	0.0191	0.0143	0.0061
$\lambda_{max}$	0.01976	0.02386	0.03605	0.02678	0.01019
$\lambda_{min}$	0.00847	0.02352	0.00599	0.00098	2.1494e-010
NR = 5%					
<i>ER</i>	0.2588	0.2844	0.2003	0.1922	3.3191e+005
<i>ER<sub>max</sub></i>	0.3012	0.3230	0.3271	0.3088	1.6542e+007
<i>ER<sub>min</sub></i>	0.2404	0.2635	0.1329	0.1403	0.1305
$\lambda$	0.0975	0.1271	0.0546	0.0374	0.0130
$\lambda_{max}$	0.1076	0.1289	0.1393	0.1831	0.0388
$\lambda_{min}$	0.0712	0.1237	0.0156	0.0109	2.1494e-010



Tabela 4.4: Comparação dos resultados dos métodos LSQR e RRGMRRES, para o problema *Baart*.

	LSQR			RRGMRES	
	MOR	DP	Min_Psi	MOR	DP
	NR = 0.1%				
$ER$	0.1609	0.1659	0.1200	0.6880	0.1659
$ER_o$	0.1110	0.1110	0.1110	0.0342	0.0342
$It$	4	3	4	1	3
$\widehat{k}$	5	5	5	4	4
	NR = 1%				
$ER$	0.2158	0.1667	0.1667	0.6880	0.1667
$ER_o$	0.1532	0.1532	0.1532	0.0379	0.0379
$It$	3	3	3	1	3
$\widehat{k}$	4	4	4	4	4
	NR = 5%				
$ER$	0.3114	0.3418	0.3418	0.6880	0.3169
$ER_o$	0.1792	0.1792	0.1792	0.0770	0.0770
$It$	3	2	2	1	3
$\widehat{k}$	4	4	4	3	3

Neste problema teste Baart, a relação entre parâmetros de regularização máximos e mínimos obtidos pelo método da GCV, para quaisquer nível de erros  $NR$  no vetor  $b$ , foi da ordem de  $1.0e + 010$ , fazendo, assim, com que as soluções obtidas tivessem erros relativos médios  $ER$  muito grandes, ocasionando soluções inapropriadas. Todos os outros métodos diretos geraram soluções com erro relativo médio  $ER$  baixo. Analisando os métodos iterativos, todos os critérios de parada se comportaram do mesmo modo, exceto o critério de Morigi quando associado ao método RRGMRRES.

### 4.2.3 Deriv2

As comparações seguintes, referentes ao problema teste *Deriv2*, correspondem a uma matriz  $A \in \mathbb{R}^{512 \times 512}$  com um número de condição igual a  $7.9681e + 004$ , e os resultados estão apresentados na tabelas 4.5 e 4.6.

Para o problema teste Deriv2 todos os métodos, diretos e iterativos, forneceram soluções adequadas com erros relativos médios  $ER$  variando entre 15% e 37% em função do aumento do nível do erro  $NR$  no vetor de dados  $b$ .

Tabela 4.5: Comparação dos resultados dos métodos da Curva-L, Discrepância, Ponto fixo, Quase otimalidade e GCV, para o problema *Deriv2*.

	Curva-L	Discrep.	Ponto fixo	Quase otim	GCV
NR = 0.1%					
<i>ER</i>	0.1747	0.1890	0.1574	0.1696	0.1671
<i>ER<sub>max</sub></i>	0.1965	0.2153	0.1708	0.1814	0.2255
<i>ER<sub>min</sub></i>	0.1512	0.1592	0.1435	0.1566	0.1409
$\lambda$	0.0867e-03	0.0759e-03	0.2159e-03	0.3125e-03	0.1030e-03
$\lambda_{max}$	0.0908e-03	0.0779e-03	0.2727e-03	0.4460e-03	0.1248e-03
$\lambda_{min}$	0.0852e-03	0.0742e-03	0.1706e-03	0.2214e-03	0.0619e-03
NR = 1%					
<i>ER</i>	0.2240	0.2256	0.2302	0.2496	0.2521
<i>ER<sub>max</sub></i>	0.2647	0.2636	0.2758	0.2804	0.4401
<i>ER<sub>min</sub></i>	0.1934	0.1976	0.1969	0.2303	0.1970
$\lambda$	0.00073	0.00082	0.00099	0.00145	0.00048
$\lambda_{max}$	0.00079	0.00083	0.00159	0.00245	0.00065
$\lambda_{min}$	0.6965e-03	0.8055e-03	0.7298e-03	0.9909e-03	0.2514e-03
NR = 5%					
<i>ER</i>	0.3112	0.3291	0.2999	0.3291	0.3765
<i>ER<sub>max</sub></i>	0.3602	0.3590	0.3595	0.3612	1.6522
<i>ER<sub>min</sub></i>	0.2887	0.3126	0.2444	0.2905	0.2459
$\lambda$	0.0034	0.0045	0.0029	0.0044	0.0014
$\lambda_{max}$	0.0036	0.0045	0.0045	0.0065	0.0021
$\lambda_{min}$	0.0031	0.0043	0.0020	0.0027	0.0003

Tabela 4.6: Comparação dos resultados dos métodos LSQR e RRGMRRES, para o problema *Deriv2*.

	LSQR			RRGMRES	
	MOR	DP	Min_Psi	MOR	DP
	NR = 0.1%				
<i>ER</i>	0.1648	0.1720	0.1569	0.1854	0.1913
<i>ER<sub>o</sub></i>	0.1514	0.1514	0.1514	0.1517	0.1517
<i>It</i>	14	11	14	11	10
$\widehat{k}$	16	16	16	12	12
	NR = 1%				
<i>ER</i>	0.2503	0.2709	0.2312	0.2633	0.2783
<i>ER<sub>o</sub></i>	0.2213	0.2213	0.2213	0.2224	0.2224
<i>It</i>	9	6	7	7	5
$\widehat{k}$	9	9	9	8	8
	NR = 5%				
<i>ER</i>	0.3241	0.3703	0.3255	0.3399	0.3716
<i>ER<sub>o</sub></i>	0.2910	0.2910	0.2910	0.2922	0.2922
<i>It</i>	5	4	4	5	3
$\widehat{k}$	7	7	7	6	6

#### 4.2.4 Foxgood

As comparações seguintes, referentes ao problema teste *Foxgood*, correspondem a uma matriz  $A \in \mathbb{R}^{512 \times 512}$  de número de condição igual a  $3.2445e + 019$ , e os resultados se encontram nas tabelas 4.7 e 4.8.

Neste problema teste Foxgood, o método LSQR com o novo critério de parada forneceu a melhor solução para um nível de erro  $NR$  de 0.1%, com a solução possuindo um erro relativo médio  $ER$  praticamente igual ao erro relativo ótimo médio  $ER_o$ . Novamente o método da GCV foi o que apresentou os piores resultados, pois o erro relativo médio  $ER$  das soluções regularizadas chegou a  $6.69e + 003$ .

Tabela 4.7: Comparação dos resultados dos métodos da Curva-L, Discrepância, Ponto fixo, Quase otimalidade e GCV, para o problema *Fox-good*.

	Curva-L	Discrep.	Ponto fixo	Quase otim	GCV
NR = 0.1%					
<i>ER</i>	0.0223	0.0240	0.0106	0.0105	342.8316
<i>ER<sub>max</sub></i>	0.0652	0.0623	0.0232	0.0257	1.6877e+004
<i>ER<sub>min</sub></i>	0.0071	0.0069	0.0033	0.0026	0.0016
$\lambda$	0.00082	0.00077	0.00345	0.00300	0.00121
$\lambda_{max}$	0.0011	0.0008	0.0058	0.0118	0.0020
$\lambda_{min}$	0.00056	0.00076	0.00192	0.00150	2.9845e-009
NR = 1%					
<i>ER</i>	0.0286	0.0267	0.0253	0.0290	3.4180e+003
<i>ER<sub>max</sub></i>	0.0819	0.0802	0.0478	0.0466	1.6877e+005
<i>ER<sub>min</sub></i>	0.0041	0.0025	0.0045	0.0203	0.0048
$\lambda$	0.0061	0.0077	0.0119	0.0133	0.0046
$\lambda_{max}$	0.0075	0.0078	0.0174	0.0214	0.0073
$\lambda_{min}$	0.0046	0.0076	0.0057	0.0027	2.9845e-009
NR = 5%					
<i>ER</i>	0.0391	0.0546	0.0421	0.0395	212.9913
<i>ER<sub>max</sub></i>	0.0845	0.0856	0.0836	0.0824	6.6900e+003
<i>ER<sub>min</sub></i>	0.0132	0.0321	0.0109	0.0164	0.0093
$\lambda$	0.0213	0.0395	0.0270	0.0224	0.0095
$\lambda_{max}$	0.0280	0.0397	0.0388	0.0338	0.0149
$\lambda_{min}$	0.0159	0.0389	0.0125	0.0145	3.2207e-007

Tabela 4.8: Comparação dos resultados dos métodos LSQR e RRGMRRES, para o problema *Forgood*.

	LSQR			RRGMRES	
	MOR	DP	Min_Psi	MOR	DP
	NR = 0.1%				
<i>ER</i>	0.0209	0.0299	0.0083	0.0213	0.0299
<i>ER<sub>o</sub></i>	0.0082	0.0082	0.0082	0.0082	0.0082
<i>It</i>	3	3	3	3	3
$\widehat{k}$	4	4	4	4	4
	NR = 1%				
<i>ER</i>	0.0430	0.0311	0.0311	0.0409	0.0311
<i>ER<sub>o</sub></i>	0.0252	0.0252	0.0252	0.0253	0.0253
<i>It</i>	3	2	2	3	2
$\widehat{k}$	3	3	3	3	3
	NR = 5%				
<i>ER</i>	0.1338	0.0360	0.0360	0.1338	0.0360
<i>ER<sub>o</sub></i>	0.0358	0.0358	0.0358	0.0359	0.0359
<i>It</i>	2	2	2	2	2
$\widehat{k}$	3	3	3	3	3

## 4.2.5 Gravity

Para este problema teste foi utilizada uma matriz  $A \in \mathbb{R}^{512 \times 512}$ , com um número de condição  $2.2471e+019$ . Os resultados obtidos se encontram nas tabelas 4.9 e 4.10. Na figura 4.12 encontra-se a solução regularizada,  $x_\lambda$ , obtida pelo método GCV.

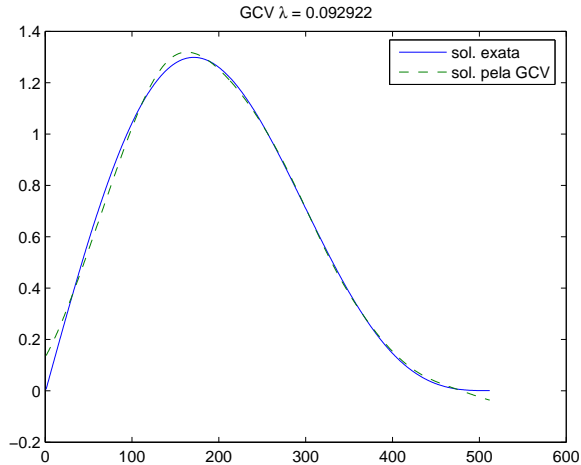


Figura 4.12: Solução exata,  $x_{exata}$ , e solução regularizada,  $x_\lambda$ , para o problema *Gravity*, com erro relativo de 1% em  $b$ .

Para o problema teste Gravity, nos três níveis de erro  $NR$  aplicados no vetor de dados, os métodos diretos e os iterativos originaram soluções regularizadas adequadas e, novamente, a exceção ficou para a GCV que com o aumento do nível de erro  $NR$  não forneceu soluções adequadas.



Tabela 4.9: Comparação dos resultados dos métodos da Curva-L, Discrepância, Ponto fixo, Quase otimalidade e GCV, para o problema *Gravity*.

	Curva-L	Discrep.	Ponto fixo	Quase otim	GCV
NR = 0.1%					
<i>ER</i>	0.0304	0.0366	0.0136	0.0131	0.2919
<i>ER<sub>max</sub></i>	0.0598	0.0677	0.0178	0.0184	12.0712
<i>ER<sub>min</sub></i>	0.0117	0.0156	0.0074	0.0082	0.0061
$\lambda$	0.0070	0.0058	0.0465	0.0350	0.0174
$\lambda_{max}$	0.0084	0.0059	0.0630	0.0692	0.0247
$\lambda_{min}$	0.0056	0.0057	0.0275	0.0164	3.7236e-005
NR = 1%					
<i>ER</i>	0.0404	0.0390	0.0281	0.0295	2.4794
<i>ER<sub>max</sub></i>	0.0684	0.0661	0.0497	0.0519	120.7126
<i>ER<sub>min</sub></i>	0.0193	0.0178	0.0151	0.0201	0.0166
$\lambda$	0.0551	0.0586	0.1619	0.1372	0.0706
$\lambda_{max}$	0.0634	0.0591	0.2236	0.2222	0.0929
$\lambda_{min}$	0.0451	0.0573	0.1011	0.0528	3.7236e-005
NR = 5%					
<i>ER</i>	0.0570	0.0520	0.0506	0.0530	12.2119
<i>ER<sub>max</sub></i>	0.1239	0.1120	0.0870	0.0845	603.5632
<i>ER<sub>min</sub></i>	0.0251	0.0287	0.0287	0.0274	0.0275
$\lambda$	0.2184	0.2958	0.3727	0.3560	0.1681
$\lambda_{max}$	0.2613	0.2983	0.5179	0.5843	0.2265
$\lambda_{min}$	0.1823	0.2882	0.2331	0.1843	3.7236e-005

Tabela 4.10: Comparação dos resultados dos métodos LSQR e RRMRES, para o problema *Gravity*.

	LSQR			RRGMRES	
	MOR	DP	Min_Psi	MOR	DP
	NR = 0.1%				
<i>ER</i>	0.0159	0.0224	0.0161	0.0158	0.0224
<i>ER<sub>o</sub></i>	0.0120	0.0120	0.0120	0.0121	0.0121
<i>It</i>	9	7	8	9	7
$\widehat{k}$	10	10	10	10	10
	NR = 1%				
<i>ER</i>	0.0337	0.0472	0.0318	0.0356	0.0472
<i>ER<sub>o</sub></i>	0.0274	0.0274	0.0274	0.0279	0.0279
<i>It</i>	7	5	8	7	5
$\widehat{k}$	8	8	8	8	8
	NR = 5%				
<i>ER</i>	0.0626	0.0710	0.0531	0.0650	0.0923
<i>ER<sub>o</sub></i>	0.0499	0.0499	0.0499	0.0498	0.0498
<i>It</i>	6	4	5	6	4
$\widehat{k}$	6	6	6	6	6

## 4.2.6 Llaplace

Para as comparações referentes ao problema teste *Llaplace*, utilizou-se uma matriz  $A \in \mathbb{R}^{512 \times 512}$ , com número de condição  $7.6160e + 032$ . Nas tabelas 4.11 e 4.12 encontram-se os resultados obtidos.

Dentre os métodos diretos, o método do ponto fixo foi o que alcançou melhores resultados. Para a matriz  $A \in \mathbb{R}^{512 \times 512}$  os métodos da discrepância e da quase otimalidade não originaram soluções regularizadas, pelo fato de terem ocorrido divisões por zero durante a compilação, mas para matrizes  $A \in \mathbb{R}^{n \times n}$ , com  $n$  inferior a 380, esses métodos se comportaram de modo análogo aos outros métodos diretos. Para os métodos iterativos, o LSQR com o novo critério de parada gerou soluções com erro relativo médio  $ER$  muito próximo ao ótimo  $ER_o$ , neste problema teste Llaplace.

Tabela 4.11: Comparação dos resultados dos métodos da Curva-L, Discrepância, Ponto fixo, Quase otimalidade e GCV, para o problema *Llaplace*.

	Curva-L	Discrep.	Ponto fixo	Quase otim	GCV
NR = 0.1%					
<i>ER</i>	0.8038	NaN	0.1829	NaN	1.2627
<i>ER<sub>max</sub></i>	0.8039	NaN	0.1905	NaN	36.3440
<i>ER<sub>min</sub></i>	0.8038	NaN	0.1743	NaN	0.1654
$\lambda$	4.8721	NaN	0.0086	NaN	0.0023
$\lambda_{max}$	4.8721	NaN	0.0133	NaN	0.0039
$\lambda_{min}$	4.8721	NaN	0.0052	NaN	4.9341e-006
NR = 1%					
<i>ER</i>	0.8038	NaN	0.2067	NaN	0.8405
<i>ER<sub>max</sub></i>	0.8040	NaN	0.2166	NaN	16.2112
<i>ER<sub>min</sub></i>	0.8036	NaN	0.1950	NaN	0.1855
$\lambda$	4.8721	NaN	0.0535	NaN	0.0173
$\lambda_{max}$	4.8721	NaN	0.0745	NaN	0.0310
$\lambda_{min}$	4.8721	NaN	0.0304	NaN	0.0001
NR = 5%					
<i>ER</i>	0.8039	NaN	0.2280	NaN	0.3677
<i>ER<sub>max</sub></i>	0.8049	NaN	0.2466	NaN	5.0344
<i>ER<sub>min</sub></i>	0.8026	NaN	0.2076	NaN	0.1963
$\lambda$	4.8721	NaN	0.1548	NaN	0.0611
$\lambda_{max}$	4.8721	NaN	0.2266	NaN	0.0899
$\lambda_{min}$	4.8721	NaN	0.0831	NaN	0.0017

Tabela 4.12: Comparação dos resultados dos métodos LSQR e RRGMRRES, para o problema *Llaplace*.

	LSQR			RRGMRES	
	MOR	DP	Min_Psi	MOR	DP
	NR = 0.1%				
<i>ER</i>	0.2637	0.1913	0.1881	0.7519	0.1949
<i>ER<sub>o</sub></i>	0.1752	0.1752	0.1752	0.4042	0.4042
<i>It</i>	4	10	10	1	11
$\widehat{k}$	13	13	13	3	3
	NR = 1%				
<i>ER</i>	0.2636	0.2166	0.2061	0.7519	0.2231
<i>ER<sub>o</sub></i>	0.1992	0.1992	0.1992	0.4063	0.4063
<i>It</i>	4	7	8	1	8
$\widehat{k}$	11	11	11	9	9
	NR = 5%				
<i>ER</i>	0.2594	0.2631	0.2304	0.7521	0.2406
<i>ER<sub>o</sub></i>	0.2196	0.2196	0.2196	0.4026	0.4026
<i>It</i>	7	5	6	1	6
$\widehat{k}$	9	9	9	8	8

## 4.2.7 Phillips

Na resolução deste problema teste utilizou-se uma matriz  $A \in \mathbb{R}^{512 \times 512}$  cujo número de condição é  $1.1354e + 008$ . Os resultados se encontram nas tabelas 4.13 e 4.14.

Tabela 4.13: Comparação dos resultados dos métodos da Curva-L, Discrepância, Ponto fixo, Quase otimalidade e GCV, para o problema *Phillips*.

	Curva-L	Discrep.	Ponto fixo	Quase otim	GCV
NR = 0.1%					
$ER$	0.0662	0.0915	0.0091	0.0094	0.0199
$ER_{max}$	0.0915	0.1233	0.0130	0.0135	0.2720
$ER_{min}$	0.0500	0.0714	0.0057	0.0055	0.0078
$\lambda$	0.0065	0.0049	0.0515	0.0586	0.0250
$\lambda_{max}$	0.0072	0.0050	0.0670	0.0867	0.0302
$\lambda_{min}$	0.0057	0.0048	0.0397	0.0444	0.0025
NR = 1%					
$ER$	0.0501	0.0604	0.0219	0.0228	0.0525
$ER_{max}$	0.0744	0.0910	0.0310	0.0293	0.2459
$ER_{min}$	0.0300	0.0341	0.0136	0.0171	0.0169
$\lambda$	0.0581	0.0502	0.1629	0.2117	0.0736
$\lambda_{max}$	0.0625	0.0507	0.2215	0.2655	0.0949
$\lambda_{min}$	0.0524	0.0490	0.1103	0.1761	0.0193
NR = 5%					
$ER$	0.0529	0.0438	0.0377	0.0384	0.1445
$ER_{max}$	0.0824	0.0711	0.0637	0.0629	1.2169
$ER_{min}$	0.0238	0.0158	0.0173	0.0181	0.0257
$\lambda$	0.2028	0.2539	0.3592	0.3471	0.158
$\lambda_{max}$	0.2331	0.2561	0.4993	0.4523	0.1984
$\lambda_{min}$	0.1830	0.2480	0.2331	0.2889	0.0195

Tabela 4.14: Comparação dos resultados dos métodos LSQR e RRGMRRES, para o problema *Phillips*.

	LSQR			RRGMRES	
	MOR	DP	Min_Psi	MOR	DP
	NR = 0.1%				
<i>ER</i>	0.0243	0.0103	0.0243	0.0243	0.0231
<i>ER<sub>o</sub></i>	0.0082	0.0082	0.0082	0.0083	0.0083
<i>It</i>	6	9	5	4	7
$\widehat{k}$	10	10	10	10	10
	NR = 1%				
<i>ER</i>	0.0246	0.0248	0.0246	0.0260	0.0272
<i>ER<sub>o</sub></i>	0.0224	0.0224	0.0224	0.0207	0.0207
<i>It</i>	6	5	6	5	4
$\widehat{k}$	9	9	9	9	9
	NR = 5%				
<i>ER</i>	0.0710	0.0883	0.0365	0.0441	0.0913
<i>ER<sub>o</sub></i>	0.0293	0.0293	0.0293	0.0299	0.0299
<i>It</i>	6	4	8	6	3
$\widehat{k}$	7	7	7	7	7

No problema teste Phillips, o método do ponto fixo obteve melhores resultados, dentre os métodos diretos. Já entre os métodos iterativos, o método LSQR com o novo critério de parada se comportou melhor com o aumento dos níveis de erro  $NR$  no vetor de dados.

## 4.2.8 Shaw

As soluções deste problema teste *Shaw* se encontram nas tabelas 4.15 e 4.16 e foram obtidas a partir de uma matriz  $A \in \mathbb{R}^{512 \times 512}$  com número de condição  $4.3053e + 019$ .

Tabela 4.15: Comparação dos resultados dos métodos da Curva-L, Discrepância, Ponto fixo, Quase otimalidade e GCV, para o problema *Shaw*.

	Curva-L	Discrep.	Ponto fixo	Quase otim	GCV
NR = 0.1%					
<i>ER</i>	0.0492	0.0489	0.0483	0.0499	4.5571e+06
<i>ER<sub>max</sub></i>	0.0751	0.0720	0.0564	0.0764	2.2782e+08
<i>ER<sub>min</sub></i>	0.0321	0.0332	0.0387	0.0426	0.0370
$\lambda$	0.0021	0.0023	0.0074	0.0042	0.0025
$\lambda_{max}$	0.0028	0.0023	0.0110	0.0074	0.0035
$\lambda_{min}$	0.0013	0.0023	0.0047	0.0006	6.7715e-13
NR = 1%					
<i>ER</i>	0.0761	0.0834	0.0958	0.1256	4.5571e+07
<i>ER<sub>max</sub></i>	0.1431	0.1432	0.1571	0.1599	2.2782e+09
<i>ER<sub>min</sub></i>	0.0520	0.0533	0.0552	0.0493	0.0513
$\lambda$	0.0178	0.0234	0.0331	0.0712	0.0100
$\lambda_{max}$	0.0218	0.0236	0.0700	0.1224	0.0201
$\lambda_{min}$	0.0153	0.0231	0.0175	0.0063	0.0000
NR = 5%					
<i>ER</i>	0.1512	0.1584	0.1503	0.1622	2.2786e+08
<i>ER<sub>max</sub></i>	0.2147	0.2009	0.1918	0.3380	1.1391e+10
<i>ER<sub>min</sub></i>	0.1219	0.1338	0.0800	0.0948	0.0673
$\lambda$	0.0846	0.1198	0.1046	0.1178	0.0400
$\lambda_{max}$	0.1014	0.1209	0.1533	0.1688	0.0664
$\lambda_{min}$	0.0641	0.1174	0.0447	0.0086	6.7715e-13



Tabela 4.16: Comparação dos resultados dos métodos LSQR e RRGMRRES, para o problema *Shaw*.

	LSQR			RRGMRES	
	MOR	DP	Min_Psi	MOR	DP
	NR = 0.1%				
<i>ER</i>	0.0516	0.0477	0.0477	0.5880	0.0592
<i>ER<sub>o</sub></i>	0.0445	0.0445	0.0445	0.0445	0.0445
<i>It</i>	8	7	7	1	6
$\hat{k}$	9	9	9	9	9
	NR = 1%				
<i>ER</i>	0.0819	0.1642	0.1099	0.5880	0.1642
<i>ER<sub>o</sub></i>	0.0644	0.0644	0.0644	0.0665	0.0665
<i>It</i>	7	5	6	1	5
$\hat{k}$	8	8	8	8	8
	NR = 5%				
<i>ER</i>	0.2033	0.1684	0.1684	0.5880	0.1684
<i>ER<sub>o</sub></i>	0.1419	0.1419	0.1419	0.1420	0.1420
<i>It</i>	6	4	4	1	4
$\hat{k}$	7	7	7	7	7

Neste problema teste Shaw, o método da GCV gerou soluções inadequadas, enquanto que os outros métodos de regularização utilizados geraram soluções bem semelhantes entre si.

### 4.2.9 Tomo

Para a resolução deste problema teste foi tomado  $n = 25$  em *Regularization Tools* [15], gerando assim, uma matriz  $A \in \mathbb{R}^{625 \times 625}$ , cujo número de condição é igual a  $4.3053e + 019$ . As tabelas 4.17 e 4.18 permitem a comparação dos métodos de resolução empregados.

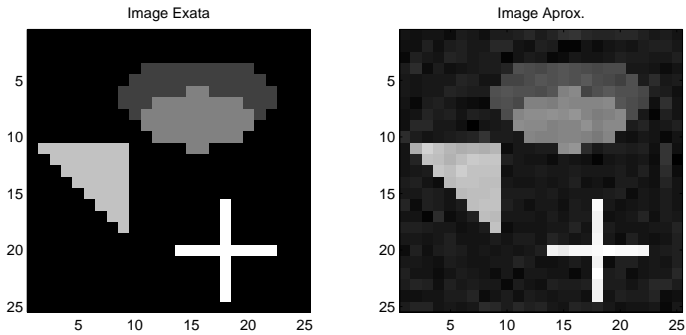


Figura 4.13: Solução  $x_{\text{exata}}$  para o problema *tomo* de dimensão 625.

Tabela 4.17: Comparação dos resultados dos métodos da Curva-L, Discrepância, Ponto fixo, Quase otimalidade e GCV, para o problema *Tomo*.

	Curva-L	Discrep.	Ponto fixo	Quase otim	GCV
NR = 0.1%					
<i>ER</i>	0.2189	0.2116	0.1139	0.2204	0.1055
<i>ER<sub>max</sub></i>	0.5429	0.5353	0.1175	0.5429	0.3353
<i>ER<sub>min</sub></i>	0.1004	0.0991	0.1100	0.1005	0.0696
$\lambda$	0.0002	0.0008	0.1601	2.3396e-06	0.0234
$\lambda_{max}$	0.0010	0.0022	0.1639	4.3362e-06	0.0448
$\lambda_{min}$	7.9471e-06	0.0001	0.1557	1.3393e-06	0.0007
NR = 1%					
<i>ER</i>	1.3844	1.0524	0.3026	1.8490	0.3278
<i>ER<sub>max</sub></i>	4.7636	3.5868	0.3082	4.7636	2.5902
<i>ER<sub>min</sub></i>	0.4248	0.4365	0.2978	0.7756	0.2555
$\lambda$	0.0063	0.0090	0.8429	0.0000	0.2598
$\lambda_{max}$	0.0313	0.0296	0.8754	5.5122e-006	0.3802
$\lambda_{min}$	0.000002	0.00085	0.8213	0.000002	0.00079
NR = 5%					
<i>ER</i>	4.2941	1.5182	0.4235	6.1402	0.3900
<i>ER<sub>max</sub></i>	10.7620	7.9868	0.4325	10.7620	0.4093
<i>ER<sub>min</sub></i>	0.5708	0.6471	0.4154	3.1849	0.3746
$\lambda$	0.0471	0.0783	2.2221	4.5541e-06	0.9905
$\lambda_{max}$	0.2343	0.1606	2.3123	7.3384e-06	1.2104
$\lambda_{min}$	2.9066e-06	0.0033	2.1699	2.6415e-06	0.7400

Para o problema teste Tomo, o método do ponto fixo e o da GCV forneceram soluções regularizadas melhores que as dos outros métodos diretos.

Tabela 4.18: Comparação dos resultados dos métodos LSQR e RRGMRRES, para o problema *Tomó*.

	LSQR			RRGMRES	
	MOR	DP	Min.Psi	MOR	DP
	NR = 0.1%				
<i>ER</i>	0.3079	0.1291	0.1253	0.7930	0.1122
<i>ER<sub>o</sub></i>	0.1122	0.1122	0.1122	0.9156	0.9156
<i>It</i>	8	63	64	2	99
$\widehat{k}$	99	99	99	2	2
	NR = 1%				
<i>ER</i>	0.2667	0.2331	0.1969	0.5981	0.2371
<i>ER<sub>o</sub></i>	0.1827	0.1827	0.1827	0.9132	0.9132
<i>It</i>	15	19	87	1	99
$\widehat{k}$	73	73	73	2	2
	NR = 5%				
<i>ER</i>	0.3697	0.3600	0.9275	0.5383	1.0097
<i>ER<sub>o</sub></i>	0.3264	0.3264	0.3264	0.9130	0.9130
<i>It</i>	5	7	99	2	99
$\widehat{k}$	15	15	15	1	1

## 4.2.10 Wing

Foi escolhida para a resolução deste problema teste uma matriz  $A \in \mathbb{R}^{256 \times 256}$ , cujo número de condição é igual a  $1.2052e + 020$ . Os resultados estão nas tabelas 4.19 e 4.20.

Para o problema teste *Wing*, nenhum dos métodos testados chegou a uma solução próxima da solução exata, independente do nível de erro *NR* no vetor de dados, conforme pode ser visto, tanto pelo erro relativo médio *ER*, na ordem de 60%, como pela solução gerada, por exemplo, pelo

Tabela 4.19: Comparação dos resultados dos métodos da Curva-L, Discrepância, Ponto fixo, Quase otimalidade e GCV, para o problema *Wing*.

	Curva-L	Discrep.	Ponto fixo	Quase otim	GCV
NR = 0.1%					
<i>ER</i>	0.6029	0.6032	0.5890	0.6040	1.3373e+04
<i>ER<sub>max</sub></i>	0.6097	0.6102	0.6047	0.6055	6.3943e+05
<i>ER<sub>min</sub></i>	0.6002	0.5999	0.4484	0.6033	0.4519
$\lambda$	0.2570e-03	0.3165e-03	0.8431e-03	0.5778e-03	0.2556e-03
$\lambda_{max}$	0.0003	0.0003	0.0015	0.0021	0.0005
$\lambda_{min}$	0.1205	0.3127	0.0354	0.1722	1.5245e-011
NR = 1%					
<i>ER</i>	0.6048	0.6045	0.6045	0.6063	1.3247e+005
<i>ER<sub>max</sub></i>	0.6065	0.6055	0.6051	0.6403	6.3943e+006
<i>ER<sub>min</sub></i>	0.6029	0.6041	0.6030	0.6031	0.6035
$\lambda$	0.0019	0.0032	0.0030	0.0026	0.0011
$\lambda_{max}$	0.0027	0.0032	0.0044	0.0051	0.0017
$\lambda_{min}$	0.0009	0.0032	0.0015	0.0002	1.5245e-011
NR = 5%					
<i>ER</i>	0.6050	0.6193	0.6053	0.6051	6.6233e+005
<i>ER<sub>max</sub></i>	0.6078	0.6289	0.6112	0.6076	3.1972e+007
<i>ER<sub>min</sub></i>	0.6036	0.6127	0.6037	0.6032	0.6032
$\lambda$	0.0056	0.0190	0.0066	0.0053	0.0020
$\lambda_{max}$	0.0077	0.0194	0.0106	0.0080	0.0033
$\lambda_{min}$	0.0030	0.0184	0.0033	0.0022	1.5245e-011

método do ponto fixo, figura 4.15. O fato de que, somente neste problema teste, nenhum dos métodos utilizados gerou uma solução adequada, nos levou a fazer uma verificação da condição discreta de Picard. A figura 4.14 mostra que os coeficientes de Fourier, representados por  $(x)$ , a partir do

Tabela 4.20: Comparação dos resultados dos métodos LSQR e RRGMRRES, para o problema *Wing*.

	LSQR			RRGMRES	
	MOR	DP	Min_Psi	MOR	DP
	NR = 0.1%				
<i>ER</i>	0.6044	0.6044	0.6044	0.8237	0.6043
<i>ER<sub>o</sub></i>	0.5475	0.5475	0.5475	0.5963	0.5963
<i>It</i>	2	2	3	1	3
$\widehat{k}$	5	5	5	3	3
	NR = 1%				
<i>ER</i>	0.6044	0.6044	0.6044	0.7755	0.6044
<i>ER<sub>o</sub></i>	0.5944	0.5944	0.5944	0.6401	0.6401
<i>It</i>	2	2	2	2	2
$\widehat{k}$	4	4	4	3	3
	NR = 5%				
<i>ER</i>	0.6661	0.6047	0.6047	0.7887	0.6047
<i>ER<sub>o</sub></i>	0.6047	0.6047	0.6047	0.7689	0.7689
<i>It</i>	2	2	2	2	2
$\widehat{k}$	3	3	3	3	3

coeficiente de índice  $n = 4$ , estão sobre os valores singulares, representados por ( $\cdot$ ), fazendo com que a razão entre eles, representada por ( $o$ ), seja muito grande à medida que  $n$  cresce. Assim, este problema teste não possui uma solução de quadrado integrável. Por conseguinte, não satisfaz a condição discreta de Picard, não sendo possível apresentar uma solução adequada.

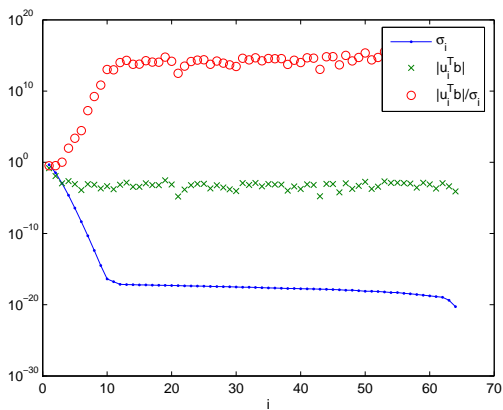


Figura 4.14: Valores singulares para a matriz  $A$  do problema *Wing*, coeficientes de Fourier e a razão entre eles, com  $n=256$  e erro relativo=5% no vetor  $b$ .

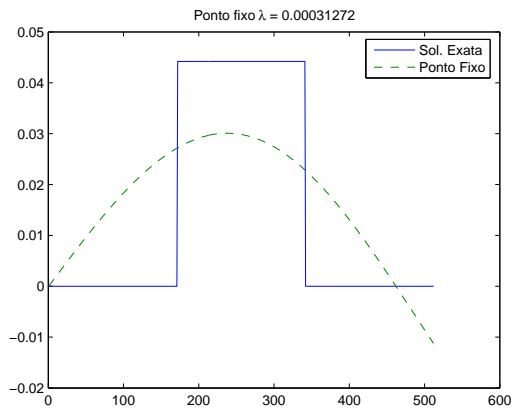


Figura 4.15: Solução exata,  $x_{exata}$ , e solução regularizada,  $x_\lambda$ , para o problema *Wing*, com erro relativo de 0.1% em  $b$ .



# Capítulo 5

## Conclusão

Nas simulações verificadas no capítulo anterior, pode-se observar que dentre os métodos diretos o que se comportou melhor, na maioria dos casos, 20 dos 30 casos estudados, foi o método do ponto fixo e aquele com os piores resultados foi o da GCV.

Em termos gerais e lembrando que os métodos iterativos possuem um esforço computacional menor do que os métodos diretos, pelo fato, de não precisarem calcular a SVD da matriz representativa do problema discretizado, o novo critério de parada foi mais eficiente em 10 casos, conforme pode ser verificado pelo quadro resumo que se encontra na tabela 5.1.

Agora, se compararmos somente os três critérios de parada empregados no método LSQR , cujo resumo que se encontra na tabela 5.2, podemos observar a nítida vantagem do novo critério de parada, apresentado neste trabalho, em relação aos demais critérios testados, estes melhores resul-

Tabela 5.1: Resumo dos métodos de regularização mais eficientes por problema teste.

	NR = 0.1%	NR = 1%	NR = 5%
Heat	ponto fixo	LSQR com novo crit.	curva-L
Baart	curva-L	ponto fixo	quase otimalidade
Deriv2	LSQR com novo crit.	curva-L	ponto fixo
Foxgood	LSQR com novo crit.	LSQR com novo crit.	LSQR com novo crit.
Gravity	quase otimalidade	ponto fixo	ponto fixo
Llaplace	ponto fixo	LSQR com novo crit.	ponto fixo
Phillips	ponto fixo	ponto fixo	LSQR com novo crit.
Shaw	LSQR com novo crit.	curva-L	LSQR com novo crit.
Tomo	quase otimalidade	LSQR com novo crit.	LSQR com disc

tados ocorreram em 81,5% dos testes realizados. Se a comparação for feita exclusivamente com o critério de parada da discrepância, vê-se que o novo critério gerou soluções com um erro relativo médio maior que o da discrepância somente em dois testes, ou seja em 7,4% dos casos. Cabe salientar que o novo critério de parada não utiliza a informação adicional do nível de ruído, ao contrário do critério da discrepância, que necessita deste ítem.

O estudo dos problemas testes avaliados nesta dissertação, realmente confirmam o fato de que a escolha de um parâmetro de regularização adequado, nos métodos diretos, conduz a solução procurada adequada e, nos métodos iterativos, se não for introduzido um critério de parada, a solução

Tabela 5.2: Comparação dos critérios de parada utilizados no método LSQR por problema teste, com base no erro relativo médio.

	NR = 0.1%	NR = 1%	NR = 5%
Heat	Novo	Novo	Novo
Baart	Novo	Novo e discrepância	Morigi
Deriv2	Novo	Novo	Morigi
Foxgood	Novo	Novo e discrepância	Novo e discrepância
Gravity	Morigi	Novo	Novo
Llaplace	Novo	Novo	Novo
Phillips	Discrepância	Novo e Morigi	Novo
Shaw	Novo e discrepância	Novo	Novo e discrepância
Tomo	Novo	Novo	Discrepância

gerada começa a se afastar da solução exata a partir de determinada iteração.

Para estudos futuros, no caso do problema *Wing* em que nenhum dos métodos aqui apresentados originou soluções adequadas, sugere-se um novo estudo utilizando uma norma diferente da norma 2 como, por exemplo, a norma 1. Sugere-se, ainda, um estudo aprofundado que fundamente o novo critério de parada aqui proposto.



# Referências

# Bibliográficas

- [1] BAUER, F; REISS, M. *Regularization independent of the noise level: an analysis of quasi-optimality*. Inverse Problems, vol. 24, 055009, 2008.
- [2] BAZÁN, F. S. V.; FRANCISCO, J. B. *An improved fixed-point algorithm for determining the Tikhonov regularization parameter* . Inverse Problems, 2009.
- [3] BAZÁN, F. S. V. *Fixed-point iterations in determining the Tikhonov regularization parameter* . Inverse Problems, vol. 24, 2008.
- [4] BAZÁN, F. S. V.; BORGES, L. *GKB-FP: an algorithm for large-scale discrete ill-posed problems*. Aceito para publicação em BIT.
- [5] BERTERO, M.; BOCCACCI, P. *Introduction to inverse problems in imaging*. IOP Publishing, London, 1998.

- [6] CALVETTI, D.; REICHEL, L; LEWIS, B. *On the regularizing properties of the GMRES method*. Numer. Math. 2002, pp. 605-625.
- [7] CALVETTI, D.; LEWIS, B.; REICHEL, L. *GMRES, L-Curves and discret ill- posed problems*. BIT, V. xx, 2002, pp. 44-65.
- [8] CHUNG, J.; NAGY, J. G.; O'LEARY, D. P. *Weighted-GCV method for Lanczos-Hybrid regularization*. Transactions on Numerical Analysis, Vol. 28, pp. 149-167, 2008.
- [9] DEMMEL, J. *Applied numerical linear algebra*. Siam, 1996.
- [10] FERNANDES, E. M. G. P. *Computação numérica*. ISBN: 972-96944-1-9, (411 páginas), Universidade do Minho, Braga, 1998.
- [11] GOLUB, G. H.; LOAN, C. F. V. *Matrix computations*. The Johns Hopkins University Press, London, 1996.
- [12] GROETSCH, C. W. *Elements of applicable functional analysis*. Dekker, New York, 1980.
- [13] HANSEN, P. C. *Rank-deficient and discret ill-posed problems: numerical aspects of linear inversion*. Society for Industrial and Applied Mathematics, Philadelphia, 1997.
- [14] HANSEN, P. C. *The L-curve and its use in the numerical treatment of inverse problems*. Computational Inverse Problems in Electrocardiography, WIT Press, Southampton, pp. 119-142, 2001.
- [15] HANSEN, P. C. *Regularization tools version 4.0 for Matlab 7.3*. Numerical Algorithms, 2007, pp. 189-194.
- [16] HILGENDORF, A. C. *Linear and nonlinear models for inversion of electrical conductivity profiles in field soils from EM-38 measurements*.

Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Mathematics with Operations Research and Statistics Option, New Mexico, 1997.

- [17] JACOBSEN, M. *Two-grid iterative methods for ill-posed problems*. Technical University of Denmark, Lyngby, 2000.
- [18] JIANG, M.; XIA, L.; SHOU, G.; TANG, M. *Combination of the LSQR method and a genetic algorithm for solving the electrocardiography inverse problem*. IOP Publishing, London, 2007.
- [19] MATHEWS, J. H.; FINK, K. D. *Numerical methods using MATLAB*. Prentice Hall, 1999.
- [20] MEYER, C. D. *Matrix analysis and applied linear algebra*. SIAM, Philadelphia, 2000.
- [21] MORIGI, S.; REICHEL, L.; SGALLARI, F.; ZAMA, F. *Iterative methods for ill-posed problems and semiconvergent sequences*. J. Comput. Appl. Math., vol. 193, pp. 157-167, 2006.
- [22] PAIGE, C. C.; SAUNDERS, M. A. *LSQR: An algorithm for sparse linear equations and sparse least squares*. ACM Transactions on Mathematical Software, Vol 8, No. 1, 1982, Pages 43-71.
- [23] PEDERSEN, J. *Modular algorithms for large-scale total variation image deblurring*. IMM-Thesis-2005-06, Lyngby, 2005.
- [24] RAMM, A. G. *Inverse problems - mathematical and analytical techniques with applications to engineering*. Springer, Boston, 2005.
- [25] REGIŃSKA, T. *A regularization parameter in discrete ill-posed problems*. SIAM, vol 17, pp. 740-749, 1996.

- [26] SANTOS, M. I. C. R. *Métodos numéricos para equações integrais de volterra com núcleo singular*. Universidade Técnica de Lisboa - Instituto Superior Técnico, 1996.
- [27] SILVA NETO, A. J.; MOURA NETO, F. D. *Problemas inversos. conceitos fundamentais e aplicações*. Editora da Universidade do Estado do Rio de Janeiro, 2005.
- [28] TREFETHEN, L. N.; BAU III, D. *Numerical linear algebra*. Society for Industrial and Applied Mathematics, Philadelphia, 1997.
- [29] VELHO, H. F. C. *Problemas inversos - conceitos básicos e aplicações*. IV Encontro de Modelagem Computacional, Nova Friburgo, IPRJ-UERJ, Rio de Janeiro, 2001, pp. 63-79.
- [30] WANG, Y. *A restarted conjugate gradient method for ill-posed problems*. Acta Mathematicae Applicatae Sinica, English Series Vol. 19, 2003.