

Problemas Inversos: Métodos Iterativos, Regularização e Validação Cruzada Generalizada

Reginaldo J. Santos*

Departamento de Matemática Aplicada

Instituto de Matemática, Estatística e Ciência da Computação

Universidade Estadual de Campinas

IMECC – UNICAMP

CP 6065, 13081-970, Campinas - SP - Brasil

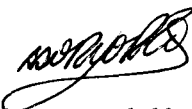
12 de janeiro de 1995

*Parcialmente financiado pela CAPES-PICD. Licenciado do Departamento de Matemática, Universidade Federal de Minas Gerais, ICEx - UFMG, CP 702, 30161-970, Belo Horizonte-MG-Brasil. e-mail: regi@mat.ufmg.br

Problemas Inversos: Métodos Iterativos, Regularização e Validação Cruzada Generalizada

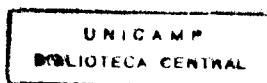
Este exemplar corresponde a redação final da tese devidamente corrigida e defendida pelo **Sr. Reginaldo de Jesus Santos** e aprovada pela Comissão Julgadora.

Campinas, 12 de janeiro de 1995.



Prof. Dr. Álvaro Rodolfo De Pierro

Tese apresentada ao Instituto de Matemática, Estatística e Ciência da Computação, UNICAMP, como requisito parcial para obtenção do Título de **DOCTOR** em Ciências em Matemática Aplicada.



Resumo

Estudamos aqui métodos numéricos para resolver problemas inversos. Provamos resultados sobre a consistência de métodos iterativos lineares estacionários convergentes para solução de quadrados mínimos de um sistema linear. Demonstramos a equivalência entre truncar um método iterativo linear estacionário e regularização de Tikhonov. Nossos resultados estendem, para o caso de posto incompleto, os de H. Fleming. Estendemos, para problemas não lineares, o método de escolha do parâmetro de regularização chamado Validação Cruzada Generalizada (GCV), introduzido por G. Whaba. Provamos resultados sobre o comportamento assintótico do parâmetro determinado por GCV para problemas não lineares que estendem os de G. Golub, M. Heath e G. Whaba. D. Girard introduziu uma variação do método GCV, que usa um método Monte-Carlo para o cálculo do traço de uma matriz simétrica ou simetrizável. Demonstramos resultados sobre o comportamento assintótico da estimativa do traço, para matrizes quaisquer, que generalizam resultados de D. Girard. Aplicamos os resultados anteriores em Tomografia Computadorizada como critério de parada de métodos iterativos.

Abstract

In this thesis we study numerical methods for solving inverse problems. We prove results on consistency of iterative linear stationary methods which converge to the least squares solution of a linear system of algebraic equations. We prove that solutions by direct regularization of linear systems are equivalent to truncated iterations of certain type of iterative methods. Our proofs extend previous results of H. Fleming to the rank-deficient case, giving a unified approach that includes the underdetermined and overdetermined problems.

We extend Generalized Cross-Validation (GCV) to the case in which the problem and the influence operator are nonlinear. From this extension we deduce stopping rules for general linear stationary methods and for the conjugate gradients (CG) method. We use a Monte-Carlo approach to compute the GCV functional. We prove results on the asymptotic optimality of our extension of GCV and on the Girard's Monte-Carlo method to estimate the trace of general matrices. Finally, we apply our results to the Positron Emission Tomography problem using the stationary method ART and CG.

Conteúdo

Introdução	1
1 Consistência de métodos lineares estacionários	9
1.1 Introdução	9
1.2 Resultados	10
2 Equivalência entre regularização e iteração truncada	15
2.1 Introdução	15
2.2 Resultados preliminares	17
2.3 Equivalência de soluções	21
3 Extensão de Validação Cruzada Generalizada (GCV)	27
3.1 Introdução	27
3.2 GCV Estendido para problemas não lineares	31
3.3 Monte-Carlo GCV	36
4 GCV como critério de parada	41
4.1 Introdução	41
4.2 Métodos estacionários	43
4.3 Aplicação na tomografia de emissão de pósitrons (PET)	46
Bibliografia	69

Lista de Figuras

4.1	Geometria de PET para o caso simplificado de oito detetores	49
4.2	Esquema da geometria divergente em SNARK93	50
4.3	Geometria divergente de SNARK93 usada para simular a geometria de PET mostrada na Fig. 4.1	51
4.4	Cinco estimativas Monte-Carlo de $(\frac{1}{30292}Tr(I_{30292} - A(k)))^2$ para o método ART com $\omega = 0.025$	52
4.5	Funções de perda típicas para o método ART com $\omega = 0.025$ no caso de $E(\epsilon) = 0$ e $E(\epsilon\epsilon^t) = 100I_{30292}$	53
4.6	Funções de perda típicas para o método ART com $\omega = 0.025$ no caso de erros típicos de PET	54
4.7	Um phantom de 95×95 pixels gerado aleatoriamente e suas reconstruções usando ART com $\omega = 0.025$	55
4.8	Comparação das densidades exatas com as das reconstruções de todos os pixels localizados na coluna 42, usando ART com $\omega = 0.025$	56
4.9	Cinco estimativas Monte-Carlo de $(\frac{1}{30292}Tr(I_{30292} - DA(k)(b)))^2$ para o método PCCGMR com $\omega = 0.0$	61
4.10	Cinco estimativas Monte-Carlo de $(\frac{1}{30292}Tr(I_{30292} - DA(k)(b)))^2$ para o método PCCGMR com $\omega = 0.025$	61
4.11	Funções de perda típicas para o método PCCGMR com $\omega = 0.0$	62
4.12	Funções de perda típicas para o método PCCGMR com $\omega = 0.025$	62
4.13	Um phantom de 95×95 pixels gerado aleatoriamente e suas reconstruções usando PCCGMR com $\omega = 0.0$	63

4.14	Comparação das densidades exatas com as das reconstruções de todos os pixels localizados na coluna 42, usando PCCGNR com $\omega = 0.0$	64
4.15	Funções de perda típicas para a primeira iteração do método ART variando o parâmetro de relaxação ω	65
4.16	Funções de perda típicas para a segunda iteração do método ART variando o parâmetro de relaxação ω	65
4.17	Funções de perda típicas para a terceira iteração do método ART variando o parâmetro de relaxação ω	66
4.18	Funções de perda típicas para o método ART tomando o parâmetro de relaxação ótimo a cada iteração	66
4.19	Um phantom de 95×95 pixels gerado aleatoriamente e suas reconstruções usando ART com ω ótimo em cada iteração	67
4.20	Comparação das densidades exatas com as das reconstruções de todos os pixels localizados na coluna 42, usando ART com ω ótimo em cada iteração	67

Introdução

Muitos problemas na indústria e nas ciências puras consistem na determinação de grandezas associadas à estrutura interna de sistemas físicos a partir do comportamento do sistema sob a ação de um agente externo. Estes problemas são denominados *problemas inversos*, em contraposição aos problemas diretos, onde a estrutura interna é conhecida e se quer prever o comportamento do sistema sob ação de um agente externo.

Os problemas inversos são, em geral, mal postos, no sentido de Hadamard [33], pois pequenas perturbações nos dados podem causar enormes variações na solução. Hadamard acreditava que o modelo por trás deste tipo de problema estava errado. Porém, muitas aplicações levaram a consideração de tais problemas mal postos, cuja solução tem um sentido físico bem definido. As bases teóricas para a solução de problemas inversos estão bem estabelecidas, pelo menos para o caso linear, e podem ser encontradas nos livros de Groetsch [32] e Louis [50]. Em [23], de Engl, está uma revisão sobre problemas inversos enfatizando suas origens industriais e o tratamento matemático do ponto de vista teórico. Revisões analisando os aspectos numéricos são [35], de Hanke e Hansen, [72] de Varah e [9] de Björck e Eldén. O último foi escrito para um simpósio internacional, mas infelizmente nunca foi publicado. Também em [3,7,52,55] podem ser encontrados métodos numéricos a disposição. Com relação a software, existem poucos pacotes destinados a solução de problemas inversos, [2,12,19,37,64].

O assunto principal desta tese é como encontrar f que satisfaz

$$Kf = g \quad (0.1)$$

onde $K : \mathcal{H}_1 \rightarrow \mathcal{H}_2$ é um operador entre espaços de Hilbert, injetivo, contínuo, mas cujo operador inverso $K^{-1} : \mathcal{R}(K) \rightarrow \mathcal{H}_1$ é não limitado (descontínuo). Um exemplo é a equação integral de Fredholm de primeira espécie

$$\int_{\Omega} k(s, t, f(t)) dt = g(s), s \in \Omega \subseteq \mathbb{R}^q. \quad (0.2)$$

Exemplos práticos incluem heliosismologia inversa (veja [58]), microscopia de scanamento confocal (veja [5,6]), espalhamento inverso (veja [47]), sensoriamento remoto (veja [68]), estimação de parâmetros em equações diferenciais (veja [17]). Nos experimentos numéricos desta tese consideramos apenas a Tomografia Computadorizada (veja [39,44,54]). Em [67] pode ser encontrada uma descrição resumida de uma ampla e variada coleção de exemplos.

Como, na prática, somente temos acesso a uma aproximação da função g em (0.1), que é contaminada com erros, temos $g = g^{\text{exato}} + \epsilon$, $g^{\text{exato}} \in \mathcal{R}(K)$ e $\|\epsilon\|/\|g^{\text{exato}}\|$ é pequeno. Não devemos esperar que g pertença a $\mathcal{R}(K)$, assim a tentativa de computar $K^{-1}g$ no lugar de $K^{-1}g^{\text{exato}}$ divergirá ou obteremos um resultado sem sentido, independente de quão pequena seja a perturbação ϵ .

Isto leva ao conceito de *regularização*, que segundo Tikhonov e Arsenin [66, seção II.1] significa, que se $g = g^{\text{exato}} + \epsilon$, então um método para “resolver” o problema deve determinar aproximações $f(\epsilon)$ tais que

$$\lim_{\epsilon \rightarrow 0} f(\epsilon) = f^{\text{exato}}. \quad (0.3)$$

O método de Tikhonov [66,32] consiste determinar estas aproximações $f(\epsilon)$ substituindo (0.1) por

$$\text{minimizar } \|Kf - g\|^2 + \alpha \|L(f - f_0)\|^2. \quad (0.4)$$

O segundo termo em (0.4) representa informação “a priori” sobre o problema físico. L é normalmente a identidade ou um operador diferencial, α é um parâmetro positivo

controlando a quantidade de regularização e f_0 é uma estimativa inicial da solução. Se α é muito pequeno o problema (0.4) é muito próximo do problema original e ocorrem instabilidades. Por outro lado, se α for grande demais, o problema a resolver tem pouca relação com o problema original. A escolha do parâmetro ótimo é essencial neste tipo de problemas.

Para se obter uma solução numérica para o problema (0.1), precisamos discretizá-lo de alguma forma. Há duas filosofias básicas a este respeito. Uma consiste em discretizar depois de regularizar (veja [32,25,24,46,56,57]). A outra sugere discretizar primeiro e então regularizar, ou seja, aplica-se algum método de discretização a equação (0.1) e então usa-se algum processo de regularização. Neste trabalho seguimos o segundo método. Após a discretização, o problema recai na resolução de um sistema de equações

$$F(x) = b, \quad (0.5)$$

onde F é uma função de \mathbb{R}^n em \mathbb{R}^m e b é um m -vetor. Como o problema (0.1) é mal-posto, se a discretização for fina o suficiente, então o sistema (0.5) será severamente mal condicionado, impondo também neste caso de dimensão finita alguma forma de regularização.

Existem dois procedimentos básicos para regularizar (0.5). O primeiro é regularizar (0.5) através do método de Tikhonov, ou seja, resolver o problema

$$\text{minimizar } \|F(x) - b\|^2 + \alpha \|B(x - x^0)\|^2, \quad (0.6)$$

onde B é a matriz identidade ou a discretização de um operador diferencial. Podemos resolver (0.6) através de um método direto, ou usar um método iterativo (para o caso linear veja [9,35]).

Uma outra forma, é aplicar um método iterativo convergente ao sistema (0.5), mesmo se ele for inconsistente, mas que convirja para uma solução de (0.5) se ele for consistente e usar um critério de parada para escolher um iterado antes que apareçam instabilidades. Por exemplo, um método que convirja para uma solução de quadrados mínimos, ou seja,

$$\text{minimizar } \|F(x) - b\|^2. \quad (0.7)$$

Este segundo procedimento, conhecido como *iteração truncada*, estabelece um equilíbrio entre precisão e regularização semelhante a aquele representado pelo primeiro e segundo termos em (0.6). Para certos métodos iterativos esta metodologia funciona muito bem, para outros simplesmente não funciona [18]. Pois a cada passo de um método iterativo estamos resolvendo um problema que é uma aproximação do problema original, mas que pode ser ou não bem condicionado.

Recentemente, Fleming [26] obteve uma equivalência entre os dois tipos de procedimentos, se $F(x) = Ax$, onde A é uma matriz de ordem $m \times n$ com posto completo. Foi provado em [26] que toda solução regularizada de (0.5) é igual a iteração truncada de um método iterativo da forma

$$x^{k+1} = x^k + MA^tW^tW(b - Ax^k), \quad k = 0, 1, 2, \dots \quad (0.8)$$

onde M e W são matrizes não singulares. Reciprocamente, cada iteração do método (0.8) é a solução de um problema da forma (0.6). Fleming considerou separadamente o caso sobredeterminado ($\text{posto}(A) = n < m$) e o caso subdeterminado ($\text{posto}(A) = n > m$). No Capítulo 2 estendemos estes resultados a matrizes de posto incompleto. No Capítulo 1 mostramos que (0.8) é a forma mais geral de um método iterativo linear estacionário convergente para soluções de

$$\text{minimizar } \|W(Ax - b)\|^2. \quad (0.9)$$

Como dissemos, a escolha do parâmetro ótimo de regularização, no caso da regularização de Tikhonov, ou a escolha do índice de parada, no caso de iteração truncada, é essencial para se encontrar uma solução aceitável para um problema mal-posto. Muitas vezes a escolha é feita simplesmente com base na experiência do usuário. Mas existem métodos que tentam encontrar este parâmetro ótimo supondo mais alguma informação sobre o problema. Alguns supõem que a norma do erro, $\|\epsilon\|$, é conhecida. O mais popular nesta linha é o baseado no princípio da discrepância de Morozov [53] (veja também [32]). Por este princípio, o parâmetro ótimo é tal que a norma do residuo, $\|F(x) - b\|$ é da ordem de grandeza da norma do erro, $\|\epsilon\|$. Pelo menos no caso linear, com hipóteses adequadas sobre a solução exata, esta escolha

leva a valores ótimos do parâmetro. Outros métodos que seguem esta linha podem ser encontrados em [35] e nas referências lá contidas.

Existem também os métodos que não supõem conhecimento da norma do erro, $\|\epsilon\|$. Um deles é o critério da L-curva: plota-se o termo da penalização, $\|B(x_\lambda - x_0)\|$ versus a norma do resíduo, $\|F(x_\lambda) - b\|$ em escala log-log. A curva correspondente tem a aparência de um L , onde se distingue uma parte “íngreme”, uma parte “plana” e um “canto” separando estas duas partes. O parâmetro ótimo é que corresponde ao canto da curva, que é o ponto onde a curvatura é máxima. Veja [36], para um “survey” a respeito deste método. Não existem resultados teóricos rigorosos que provem a eficiência desta metodologia, mas as experiências em [36] mostram a sua robustez, no caso linear (veja [35] e a bibliografia aí contida para outros métodos deste tipo).

Talvez o método mais popular quando não se supõe conhecida $\|\epsilon\|$ é *Validação Cruzada Generalizada (GCV)*, introduzido por Grace Wahba para o problema linear [76] (veja também [30,15]). Para $F(x) = Ax$, o parâmetro ótimo de GCV é a solução de

$$\text{minimizar } \frac{\frac{1}{m} \|b - Ax_\lambda\|^2}{\left[\frac{1}{m} \text{Tr}(I - A(\lambda))\right]^2}, \quad (0.10)$$

onde a matriz $A(\lambda)$ satisfaz $A(\lambda)b = Ax_\lambda$. Um aspecto particularmente atraente neste método é que existem resultados de convergência, para o caso em que a matriz $A(\lambda)$, para cada λ , é simétrica e positiva definida, supondo-se apenas que o vetor de erros ϵ é uma variável aleatória com média igual ao vetor nulo e a matriz de covariância é da forma $\sigma^2 I$, com σ desconhecido (veja [30,79]).

O’Sullivan e Wahba definem em [61,59] a seguinte extensão de GCV para o problema regularizado não linear (0.6)

$$\text{minimizar } \frac{\frac{1}{m} \|b - F(x_\lambda)\|^2}{\left[\frac{1}{m} \text{Tr}(I - \tilde{A}(\lambda))\right]^2}. \quad (0.11)$$

onde $\tilde{A}(\lambda) = J(x_\lambda)(J(x_\lambda)^t J(x_\lambda) + \lambda B)^{-1} J(x_\lambda)^t$ e $J(x)$ é o Jacobiano de $F(x)$. Eles não provam resultado algum de convergência, mas apresentam resultados numéricos, indicando que esta técnica é promissora também no caso não linear (veja [60]). Vogel em [75] também aplica este método.

No Capítulo 3 definimos a seguinte extensão do funcional de GCV para problemas não lineares e para uma família de problemas regularizados qualquer.

$$V(\lambda) = \frac{\frac{1}{m} \|b - F(x_\lambda)\|^2}{[\frac{1}{m} \text{Tr}(I - DA(\lambda)(b))]^2}, \quad (0.12)$$

onde o *operador de influência* é definido por $A(\lambda)(b) = F(x_\lambda)$ e $DA(\lambda)(b)$ é o Jacobiano da aplicação $A(\lambda)(\cdot)$ avaliado em b . E provamos um resultado básico de convergência, que é uma generalização daquele que foi provado em [30] para o caso linear. Um inconveniente para a aplicação de GCV está no cálculo do traço no denominador do funcional. Girard, em [28], encontrou uma forma engenhosa de se obter uma aproximação do traço de uma matriz simétrica $B \in \mathbb{R}^{n \times n}$, que é o seguinte:

(i) Gera-se um vetor pseudo-aleatório $w = (w_1, \dots, w_n)^t$ de uma distribuição normal, ou seja, $w \sim \mathcal{N}(0, I_{n \times n})$;

(ii) Toma-se

$$\frac{w^t B w}{w^t w} \quad (0.13)$$

como uma aproximação de

$$\frac{1}{n} \text{Tr}(B). \quad (0.14)$$

Ele provou a confiabilidade deste método para problemas de grande porte, se B for simétrica, ou pelo menos simetrizável. Também no Capítulo 3 estendemos estes resultados para uma matriz B qualquer. No Capítulo 4 aplicamos os resultados do Capítulo 3, para usar GCV como critério de parada de métodos iterativos lineares. Nosso problema modelo para as experiências numéricas foi a *Tomografia Computadorizada de Emissão de Pósitrons (PET)*.

A idéia da Tomografia é reconstruir uma função f com suporte compacto, $\Omega \subset \mathbb{R}^2$ por exemplo, conhecendo-se um conjunto de integrais de linha de f ao longo de retas, Γ_i que interceptam a região Ω . Ou seja,

$$\int_{\Gamma_i} f(r, \phi) ds = y_i. \quad (0.15)$$

O conjunto de retas Γ_i depende da geometria do “scanner”. Isto é uma semi-discretização da *Transformada de Radon* de f (veja [54,51]).

Cada tipo de Tomografia calcula grandezas físicas diferentes. Em *Tomografia de Transmissão de Raios X*, $f(r, \phi)$ é a atenuação em cada ponto de Ω . Medimos as atenuações sofridas pelos raios em cada reta que liga um emissor a um detetor, $y_i = -\log(E_{d_i}/E_{f_i})$, onde E_{f_i} e E_{d_i} são as intensidades do feixe de raios X no par emissor-detetor i , respectivamente. Em *Tomografia Sísmica*, o interesse está na estrutura de uma seção da crosta terrestre. Para obter os dados, ondas eletromagnéticas são enviadas através da terra e são medidos os tempos que gastam para chegar aos detetores, $y_i = \tau_i$. E $f(r, \phi) = 1/v(r, \phi)$, onde $v(r, \phi)$ é a velocidade de propagação da onda em cada ponto. Este modelo não inclui das difração e reflexões nas diferentes camadas (veja [4]). Em PET, injeta-se um isótopo radioativo num paciente, que se distribui no corpo de acordo com determinado processo fisiológico. Coloca-se um anel de detetores em volta do paciente, como na Fig. 4.1. Aqui, $f(r, \phi)$ é a distribuição da atividade destas emissões. Cada pósitron se aniquila com um elétron produzindo dois fótons emitidos em sentidos opostos. Assim, se um par i de detetores recebem dois fótons ao mesmo tempo, então uma emissão ocorreu na reta Γ_i . Neste caso, y_i é o número de fótons coletados simultaneamente pelo par de detetores i durante um certo intervalo de tempo (veja [73,65]).

Finalmente, agradeço a todos que tornaram possível a realização deste trabalho, principalmente Álvaro R. De Pierro, pela orientação e dedicação durante o doutorado e J. Mário Martínez, Walter Mascarenhas e Ronaldo Dias, pelas proveitosas discussões.

Capítulo 1

Consistência de métodos lineares estacionários

1.1 Introdução

Vamos considerar uma classe de métodos iterativos para encontrar x que resolva o problema

$$\text{minimize } \|W(Ax - b)\|^2, \quad (1.1)$$

onde A é uma matriz de ordem $m \times n$, b é um m -vetor e W é uma matriz com m colunas.

Seguindo a terminologia criada por Forsythe [27] os *métodos iterativos lineares estacionários* são aqueles que possuem a forma geral

$$x^{k+1} = Gx^k + h \quad k = 0, 1, 2, \dots \quad (1.2)$$

onde G é uma matriz de ordem $n \times n$ e h é um n -vetor.

Em [81] (veja também [80]) Young mostrou que todo método iterativo linear estacionário convergente para uma solução do sistema

$$Ax = b, \quad (1.3)$$

se $m = n$, pode ser escrito como

$$x^{k+1} = x^k + M(b - Ax^k), \quad k = 0, 1, 2, \dots \quad (1.4)$$

para uma matriz M inversível.

Para uma matriz A de ordem $m \times n$, Chen em [14] mostrou que se $\text{posto}(A) = n$, então qualquer método iterativo linear estacionário convergente cujos pontos limite são exatamente as soluções do problema de quadrados mínimos associado a matriz A (*completamente consistente com o problema de quadrados mínimos*) pode ser escrito como

$$x^{k+1} = x^k + MA^t(b - Ax^k), \quad k = 0, 1, 2, \dots \quad (1.5)$$

para uma matriz não singular M de ordem $n \times n$ (veja também [7, seção 20]). Ou seja, o método iterativo mais geral da forma (1.2) é equivalente ao método de Richardson de primeira ordem aplicado ao sistema de equações normais condicionado

$$MA^t(Ax - b) = 0 \quad (1.6)$$

Mostramos aqui que este resultado é verdadeiro, mesmo se $\text{posto}(A) < n$. Além disso, mostramos também qual é a forma mais geral de métodos lineares estacionários para problemas de quadrados mínimos generalizados (1.1).

1.2 Resultados

Os limites de seqüências geradas por um método iterativo do tipo (1.2), são as soluções do sistema linear

$$(I - G)x = h. \quad (1.7)$$

Suponhamos que o objetivo da aplicação do método iterativo seja resolver o sistema linear

$$Bx = c, \quad (1.8)$$

para uma dada matriz B de ordem $m \times n$ e um dado vetor c .

Neste ponto, uma questão que aparece é qual é a caracterização das matrizes G e dos vetores h tais que o conjunto solução de (1.7) é igual ao conjunto solução do sistema linear (1.8).

Teorema 1.1 *Considere os sistemas (1.8) e $Dx = f$, para matrizes B e D $m \times n$. Suponha que o sistema (1.8) seja solúvel.*

(i) *Temos que $\mathcal{S}(B,c) \subseteq \mathcal{S}(D,f)$ se, e somente se, existe uma matriz M tal que*

$$D = MB \text{ e } f = Mc; \quad (1.9)$$

(ii) *Além disso, $\mathcal{S}(B,c) = \mathcal{S}(D,f)$ se, e somente se, (1.9) se verifica para alguma matriz M inversível.*

Este Teorema e o Lema a seguir que é usado na sua demonstração são de Young [80,81]. A demonstração de Young faz uso da forma canônica de Jordan. Apresentamos aqui uma nova demonstração bem mais simples.

Lema 1.2 *Sejam B e D matrizes de ordem $m \times n$.*

(i) *Temos que $\mathcal{N}(B) \subseteq \mathcal{N}(D)$ se, e somente se, existe uma matriz M tal que*

$$D = MB \quad (1.10)$$

(ii) *Além disso, $\mathcal{N}(A) = \mathcal{N}(B)$ se, e somente se, (1.10) se verifica para alguma matriz M inversível.*

Dem. A existência de M satisfazendo as condições é claramente suficiente nos dois casos.

Suponha, agora, que $\mathcal{N}(B) \subseteq \mathcal{N}(D)$. Defina M como sendo a matriz da transformação linear que em $\mathcal{R}(B)$ é dada por $M(Bv) = Dv$ e em $\mathcal{R}(B)^\perp$ por qualquer aplicação linear de $\mathcal{R}(B)^\perp$ em $\mathcal{R}(D)^\perp$. A matriz M claramente satisfaz (1.10) e está bem definida pois, se $Bv = Bv'$, então, $v - v' \in \mathcal{N}(B)$ que por hipótese está contido em $\mathcal{N}(D)$, portanto $Dv = Dv'$.

Agora, se $\mathcal{N}(B) = \mathcal{N}(D)$, defina M da mesma forma que no caso anterior, mas em $\mathcal{R}(B)^\perp$ tome uma bijecção. Agora, $Dv = Dv'$ implica, pelo argumento aplicado acima, que $Bv = Bv'$, ou seja, M é inversível. ■

Dem. do Teorema 1.1. A existência de uma matriz M satisfazendo as condições é claramente suficiente nos dois casos.

Agora, suponha que $\emptyset \neq \mathcal{S}(B, c) \subseteq \mathcal{S}(D, f)$. Seja u uma solução de (1.8). Se $\mathcal{N}(B) \neq \{0\}$ e $v \in \mathcal{N}(B)$, então $u + v \in \mathcal{S}(B, c)$, logo da hipótese segue que $u + v$ e $u \in \mathcal{S}(D, f)$, o que implica que $v \in \mathcal{N}(D)$. Logo $\mathcal{N}(B) \subseteq \mathcal{N}(D)$, o que implica pelo lema 1.2 que existe uma matriz M tal que $MB = D$. Além disso, $Mc = MBx = Dx = f$, para $x \in \mathcal{S}(B, c)$. Isto prova (i). Com o mesmo argumento e usando a segunda parte do lema 2 se prova (ii). ■

Uma consequência do teorema 1.1 é que se a matriz B é quadrada e o sistema (1.8) é solúvel, então o conjunto solução $\mathcal{S}(B, c) = \mathcal{S}(I - G, h)$ se, e somente se,

$$G = I - MB = M(M^{-1} - B) \text{ e } h = Mc. \quad (1.11)$$

Como os métodos produzidos fazendo um “splitting”

$$B = Q - R, \quad (1.12)$$

são da forma

$$x^{k+1} = Q^{-1}Rx^k + Q^{-1}c \quad k = 0, 1, 2, \dots \quad (1.13)$$

segue de (1.11) que todo método cujos pontos limites são exatamente as soluções de (1.8) são obtidos fazendo o “splitting” (1.12) com

$$Q = M^{-1} \text{ e } R = M^{-1} - B. \quad (1.14)$$

Corolário 1.3 *Seja W uma matriz com m colunas. Todo método linear estacionário (1.14) cujos pontos limites são exatamente as soluções do problema de quadrados mínimos generalizado*

$$\min \|W(Ax - b)\|^2 \quad (1.15)$$

pode ser escrito como

$$x^{k+1} = x^k + MA^tW^tW(b - Ax^k), \quad k = 0, 1, 2, \dots \quad (1.16)$$

para alguma matriz M não singular.

Dem. O problema (1.15) é equivalente a resolver as equações normais

$$A^t W^t W A x = A^t W^t W b \quad (1.17)$$

que satisfaz as condições do teorema 1.1. Considere um método linear estacionário da forma (1.2). Como os pontos limites de (1.2) são as soluções do sistema (1.7), então os pontos limites de (1.2) coincidem com as soluções de (1.15) se, e somente se,

$$G = I - M A^t W^t W A \text{ e } h = M A^t W^t W b \quad (1.18)$$

para alguma matriz M inversível, de onde segue (1.16). ■

Capítulo 2

Equivalência entre regularização e iteração truncada

2.1 Introdução

Uma forma de se obter soluções estáveis de problemas inversos é substituir o problema

$$Ax = b \tag{2.1}$$

pela regularização de Tikhonov [66,32,9]. Isto é, a solução é obtida minimizando o funcional

$$F_\alpha(x) = \|Ax - b\|^2 + \alpha \|L(x - x_0)\|^2. \tag{2.2}$$

O segundo termo em (2.2) representa uma informação “a priori” sobre o problema. L é normalmente a discretização de um operador de derivadas, impondo algum suavizamento na solução, α é um parâmetro positivo controlando a quantidade de suavizamento sobre a solução e x_0 é uma estimativa da solução.

Uma outra forma de resolver (2.1) é aplicar um método iterativo às equações normais

$$A^t Ax = A^t b. \tag{2.3}$$

Um algoritmo usado em muitas aplicações é o método de Landweber generalizado [48,63], que é dado por

$$x^{k+1} = x^k + DA^t(b - Ax^k), \quad k = 0, 1, 2, \dots \quad (2.4)$$

onde $D = F(A^tA)$ e F é um polinômio ou uma função racional. No início do processo a precisão dos iterados cresce, mas depois de algum tempo um efeito de deteriorização aparece por causa do mal condicionamento. Uma solução estável pode ser obtida usando um critério de parada para escolher um iterado antes que esse efeito apareça. Este procedimento, conhecido como iteração truncada, estabelece um equilíbrio entre precisão e suavizamento semelhante a aquele representado pelo primeiro e segundo termos em (2.2).

Recentemente H. Fleming [26] obteve uma equivalência entre os dois tipos de métodos, se A tem posto completo. Foi provado em [26] que toda solução regularizada de (2.1) é igual a iteração truncada de um método iterativo da forma

$$x^{k+1} = x^k + MA^tP^{-1}(b - Ax^k), \quad k = 0, 1, 2, \dots \quad (2.5)$$

onde M é uma matriz não singular e P é uma matriz simétrica positiva definida. E reciprocamente, toda iteração truncada do método (2.5) é o mínimo do funcional (2.2). Fleming considerou separadamente o caso sobredeterminado ($\text{posto}(A) = n < m$) e o caso subdeterminado ($\text{posto}(A) = n > m$). Aqui, estendemos estes resultados a matrizes de posto incompleto.

É conhecido que ambos os métodos, regularização de Tikhonov e iteração truncada, pertencem a classe dos chamados “esquemas de aproximações espectrais”, isto é, as aproximações regularizadas podem ser expandidas usando o mesmo conjunto de auto-funções, diferindo somente na escolha dos chamados “filtros” (veja por exemplo [35,32,78]). Aqui, assim como no paper de Fleming [26], mostramos que a segunda forma de regularização é um caso particular da primeira.

2.2 Resultados preliminares

Como anteriormente, vamos considerar métodos iterativos da forma

$$x^{k+1} = Gx^k + f, \quad k = 0, 1, 2, \dots \quad (2.6)$$

onde G é uma matriz de ordem $n \times n$ e f é um vetor em \mathbb{R}^n . É claro que se a seqüência $\{x^k\}$ converge para x^* , então este ponto limite é solução do sistema

$$(I - G)x = f. \quad (2.7)$$

Dizemos que uma matriz é *convergente* se $\lim_{k \rightarrow \infty} G^k$ existe. Este limite existe se, e somente se, as seguintes condições são satisfeitas (veja [80]):

- (a) O raio espectral de G é menor ou igual a um.
- (b) Se λ é um auto-valor de G tal que $|\lambda| = 1$, então $\lambda = 1$ e todos os divisores elementares que correspondem a λ são lineares, isto é, λ não tem vetores principais.

Se G é uma matriz convergente, então $\text{ind}(I - G) \leq 1$, onde $\text{ind}(A)$ é o índice de A (isto é, o menor inteiro não negativo q tal que $\mathcal{R}(A^q) = \mathcal{R}(A^{q+1})$, conforme [13, Definição 7.2.1]). Se $\text{ind}(A) = q$, então $\mathbb{R}^n = \mathcal{N}(A^q) \oplus \mathcal{R}(A^q)$ (conforme [13, Lema 7.2.1]). Assim, se G é uma matriz convergente, então $\mathbb{R}^n = \mathcal{N}(I - G) \oplus \mathcal{R}(I - G)$.

O próximo Teorema descreve o iterado gerado por (2.6).

Teorema 2.1 *Seja G uma matriz de ordem $n \times n$ convergente. Então*

$$\mathbb{R}^n = \mathcal{N}(I - G) \oplus \mathcal{R}(I - G), \quad (2.8)$$

e a seguinte expressão é válida:

$$x^k = x_1^0 + kf_1 + G^k(x_2^0 - (I - G_2)^{-1}f_2) + (I - G_2)^{-1}f_2, \quad (2.9)$$

onde $f_1, x_1^0 \in \mathcal{N}(I - G)$ e $f_2, x_2 \in \mathcal{R}(I - G)$ são tais que $f = f_1 + f_2$, $x^0 = x_1^0 + x_2^0$, e $G_2 = G \Big|_{\mathcal{R}(I - G)}$.

Dem. Usando (2.6), x^k pode ser escrito como

$$x^k = G^k x^0 + \sum_{j=0}^{k-1} G^j f . \quad (2.10)$$

Seja W o subespaço gerado pelos vetores e auto-vetores associados com os auto-valores de G diferentes de um. Claramente

$$\mathbb{R}^n = \mathcal{N}(I - G) \oplus W . \quad (2.11)$$

Sejam $x_1^0, f_1 \in \mathcal{N}(I - G)$ e $x_2, f_2 \in W$ tais que $x^0 = x_1^0 + x_2^0$ e $f = f_1 + f_2$. Definindo $\hat{G}_2 = G \Big|_W$ e aplicando (2.10) obtemos

$$x^k = x_1^0 + k f_1 + \hat{G}_2^k x_2^0 + \sum_{j=0}^{k-1} \hat{G}_2^j f_2 . \quad (2.12)$$

Como \hat{G}_2 não possui um como auto-valor e W é $(I - G)$ -invariante, $I - \hat{G}_2$ tem inversa e

$$\sum_{j=0}^{k-1} \hat{G}_2^j = (I - \hat{G}_2^k)(I - \hat{G}_2)^{-1} . \quad (2.13)$$

Portanto, usando (2.12) e (2.13), temos que

$$x^k = x_1^0 + k f_1 + \hat{G}_2^k (x_2^0 - (I - \hat{G}_2)^{-1} f_2) + (I - \hat{G}_2)^{-1} f_2 . \quad (2.14)$$

Falta provar que $W = \mathcal{R}(I - G)$. Para isto, usamos a equação (2.14). Se $f \in W$, então $f_1 = 0$ e a seqüência $\{x^k\}$ é convergente, portanto (2.7) é solúvel e $f \in \mathcal{R}(I - G)$ (Isto é uma conseqüência de (2.14) e o fato de que os auto-valores de \hat{G}_2 são menores que um em módulo, mas pode ser deduzido de [16]). Por outro lado, se $f \in \mathcal{R}(I - G)$, então podemos tomar $x^0 = x^*$, uma solução de (2.7). A seqüência resultante é convergente porque G é uma matriz convergente e, pela equação (2.14), f_1 tem que ser zero; assim $f \in W$.

Concluimos que $\hat{G}_2 = G_2 = G \Big|_{\mathcal{R}(I-G)}$ e o resultado segue daí. ■

Dada a forma de Jordan de uma matriz A ,

$$A = P \begin{bmatrix} J_0 & 0 \\ 0 & J_1 \end{bmatrix} P^{-1}$$

(todos os blocos de Jordan correspondentes a $\lambda = 0$ de A são colecionados em J_0), a inversa de Drazin A^D de A é definida por

$$A^D = P \begin{bmatrix} 0 & 0 \\ 0 & J_1^{-1} \end{bmatrix} P^{-1}$$

(veja por exemplo [13, Definições 7.2.2, 7.2.3 e Teorema 7.2.1]). Assim, a equação (2.9) pode ser reescrita como

$$x^k = x_1^0 + kf_1 + G^k(x_2^0 - (I - G)^D f) + (I - G)^D f .$$

Considere, agora, o problema regularizado

$$\text{minimizar } \|Ax - b\|_P^2 + \|x - a\|_Q^2 , \quad (2.15)$$

onde $P \in \mathbb{R}^{m \times m}$ e $Q \in \mathbb{R}^{n \times n}$ são matrizes simétricas e positivas definidas, a é um vetor em \mathbb{R}^n e as normas são definidas por

$$\|z\|_P^2 = z^t P^{-1} z \quad (2.16)$$

(o mesmo para Q). Vamos também considerar um método iterativo da forma

$$x^{k+1} = x^k + MA^t P^{-1}(b - Ax^k) , \quad k = 0, 1, 2, \dots \quad (2.17)$$

onde M é uma matriz não singular. Usando a notação da seção anterior

$$G = I - MA^t P^{-1} A .$$

Lema 2.2 *A solução x^* do problema (2.15) sempre existe e pode ser escrita como*

$$x^* = (I + QA^t P^{-1} A)^{-1}(a - d) + d , \quad (2.18)$$

onde

$$d = (MA^t P^{-1} A)_2^{-1} MA^t P^{-1} b \quad e \quad (MA^t P^{-1} A)_2 = MA^t P^{-1} A \Big|_{\mathcal{R}(MA^t P^{-1} A)} .$$

Dem. É fácil ver que

$$x^* = (I + QA^tP^{-1}A)^{-1}(QA^tPb + a). \quad (2.19)$$

Pelo Teorema 2.1, $(MA^tP^{-1}A)_2$ tem uma inversa e, somando e subtraindo

$$(I + QA^tP^{-1}A)^{-1}(MA^tP^{-1}A)_2^{-1}MA^tP^{-1}b$$

em (2.19) obtemos

$$x^* = (I + QA^tP^{-1}A)^{-1}(a - d) + (I + QA^tP^{-1}A)^{-1}(QA^tP^{-1}b + d). \quad (2.20)$$

O sistema

$$A^tP^{-1}b = A^tP^{-1}Ax \quad (2.21)$$

tem solução . Aplicando M a ambos os lados de (2.21), deduzimos que

$$MA^tP^{-1}b \in \mathcal{R}(MA^tP^{-1}A).$$

Portanto,

$$MA^tP^{-1}b = MA^tP^{-1}Ad \quad (2.22)$$

Como M e Q são não singulares, podemos trocar M por Q em (2.22), obtendo

$$QA^tP^{-1}b = (QA^tP^{-1}A)(MA^tP^{-1}A)_2^{-1}MA^tP^{-1}b. \quad (2.23)$$

De (2.23) obtemos que

$$(I + QA^tP^{-1}A)^{-1}(QA^tP^{-1}b + d) = d. \quad (2.24)$$

O resultado segue de (2.20). ■

Apresentamos a seguir resultados relacionados a diagonalização simultânea de matrizes simétricas.

Teorema 2.3 *Suponha que A e B são matrizes simétricas de ordem $n \times n$ e defina $C(\mu)$ por*

$$C(\mu) = \mu A + (1 - \mu)B.$$

Se existe $\mu \in [0, 1]$ tal que $C(\mu)$ é não negativa definida e $\mathcal{N}[C(\mu)] = \mathcal{N}(A) \cap \mathcal{N}(B)$, então existe uma matriz não singular X tal que $X^t A X$ e $X^t B X$ são diagonais.

Dem. Ver [31, Cap. 8].

Corolário 2.4 *Se A é uma matriz simétrica e B é simétrica e positiva definida, então existe uma matriz não singular X tal que $X^t A X$ e $X^t B X$ são diagonais.*

Dem. O resultado segue do Teorema 2.3, fazendo $\mu = 0$. ■

2.3 Equivalência de soluções

Teorema 2.5 *Toda solução regularizada do sistema (2.1) é igual a iteração truncada da forma (2.17); isto é dadas as matrizes P e Q em (2.15) e um inteiro positivo k_0 existe uma matriz M tal que x^{k_0} dada por (2.17) é solução de (2.15).*

Dem. Como Q e $A^t P^{-1} A$ são simétricas e Q^{-1} é positiva definida, podemos diagonalizá-las simultaneamente. Assim, pelo Corolário 2.4, existe uma matriz não singular X tal que

$$X^t Q^{-1} X = \text{diag}\left(\frac{1}{q_1}, \dots, \frac{1}{q_n}\right) \quad (2.25)$$

e

$$X^t A^t P^{-1} A X = \text{diag}(p_1, \dots, p_n), \quad (2.26)$$

com $q_i > 0$ e $p_i \geq 0$ para $i = 1, \dots, n$. Conseqüentemente

$$X^{-1} Q A^t P^{-1} A X = X^{-1} Q X^{-t} (X^t A^t P^{-1} A X) = \text{diag}(p_1 q_1, \dots, p_n q_n). \quad (2.27)$$

Dado um índice de truncamento k_0 , seja

$$M = X \text{diag}(\lambda_1, \dots, \lambda_n) X^t = X D X^t \quad (2.28)$$

onde

$$\lambda_i = \begin{cases} \frac{1}{p_i}[1 - (1 + p_i q_i)^{-1/k_0}], & \text{se } p_i \neq 0, \\ 0, & \text{caso contrário.} \end{cases} \quad (2.29)$$

Usando (2.27), (2.28) e (2.29) obtemos

$$\begin{aligned} (I - MA^t P^{-1} A)^{k_0} &= (I - XDX^t A^t P^{-1} A)^{k_0} = \\ &= \{X(I - DX^t A^t P^{-1} AX)X^{-1}\}^{k_0} = \\ &= \{X \operatorname{diag}(1 - \lambda_i p_i) X^{-1}\}^{k_0} = \\ &= X \operatorname{diag}(1 - \lambda_i p_i)^{k_0} X^{-1} = \\ &= X \operatorname{diag}(1 + p_i q_i)^{-1} X^{-1} = \\ &= (I + QA^t P^{-1} A)^{-1}. \end{aligned} \quad (2.30)$$

Agora

$$I - MA^t P^{-1} A = X \operatorname{diag}(1 + p_i q_i)^{-1/k_0} X^{-1};$$

portanto, M dada por (2.28) define um método (2.17) que é convergente.

Falta provar que $x^{k_0} = x^*$ é a solução do problema (2.15). Pelo Lema 2.2, a expressão (2.18) é válida. Se fazemos $x^0 = a$ e aplicamos (2.30), segue que

$$x^* = (I + QA^t P^{-1} A)^{-1} x_1^0 + (I - MA^t P^{-1} A)^k (x_2^0 - d) + d, \quad (2.31)$$

onde $x_1^0 \in \mathcal{N}(MA^t P^{-1} A)$ e $x_2^0 \in \mathcal{R}(MA^t P^{-1} A)$ são tais que $x^0 = x_1^0 + x_2^0$. Mas,

$$(I + QA^t P^{-1} A)^{-1} x_1^0 = x_1^0, \quad (2.32)$$

porque $x_1^0 \in \mathcal{N}(MA^t P^{-1} A) = \mathcal{N}(A)$. Assim, pelo Teorema 2.1, $x^* = x^k$. ■

Agora vamos estabelecer a recíproca do Teorema 2.5.

Teorema 2.6 *Toda iteração truncada de um método da forma (2.17), onde M é uma matriz simétrica e positiva definida, é a solução do problema regularizado da forma (2.15); isto é, para todo k e matrizes M e P , existe uma matriz Q tal que x^k dada por (2.17) é solução de (2.15).*

Dem. Como M^{-1} e $A^t P^{-1} A$ são simétricas e M^{-1} é positiva definida, pelo Corolário 2.4, podemos diagonalizá-las simultaneamente. Assim, existe uma matriz não singular Y tal que

$$Y^t M^{-1} Y = \text{diag}\left(\frac{1}{m_1}, \dots, \frac{1}{m_n}\right), \quad (2.33)$$

e

$$Y^t (A^t P^{-1} A) Y = \text{diag}(a_1, \dots, a_n), \quad (2.34)$$

com $a_i \geq 0$ e $m_i > 0$ para $i = 1, \dots, n$.

Defina

$$Q = Y \text{diag}(\mu_1, \dots, \mu_n) Y^t, \quad (2.35)$$

onde

$$\mu_i = \begin{cases} \frac{1}{a_i} [(1 - a_i m_i)^{-k} - 1], & \text{se } a_i \neq 0, \\ 1, & \text{caso contrário.} \end{cases} \quad (2.36)$$

Usando (2.33), (2.34), (2.35) e (2.36) obtemos que

$$\begin{aligned} (I + Q A^t P^{-1} A)^{-1} &= Y \text{diag}(1 + \mu_i a_i)^{-1} Y^{-1} \\ &= Y \text{diag}(1 - a_i m_i)^k Y^{-1} = \\ &= (I - M A^t P^{-1} A)^k. \end{aligned} \quad (2.37)$$

O método (2.17) é convergente, então, temos que $1 - a_i m_i < 1$, se $a_i \neq 0$, para $i = 1, \dots, n$, implicando que $\mu_i > 0$. Assim, Q é positiva definida. Podemos aplicar o Teorema 2.1 e (2.37) para obtermos

$$x^k = x_1^0 + (I + Q A^t P^{-1} A)^{-1} (x_2^0 - d) + d, \quad (2.38)$$

Mas $x_1^0 \in \mathcal{N}(A)$, então

$$x_1^0 = (I + QA^tP^{-1}A)^{-1}x_1^0. \quad (2.39)$$

Se fazemos $a = x^0$, e usando o Lema 2.2, concluímos que $x^k = x^*$. ■

Exemplo 2.1. Considere o método de Landweber [48]

$$x^{k+1} = x^k + \omega A^t(b - Ax^k), \quad k = 0, 1, 2, \dots \quad (2.40)$$

onde ω é um número real positivo. Se P é uma matriz ortogonal, tal que

$$P^t A^t A P = \text{diag}(a_1, \dots, a_n), \quad (2.41)$$

então, usando a demonstração do Teorema 2.6, x^k é a solução do problema

$$\text{minimizar } \|Ax - b\|_2^2 + \|D^{-\frac{1}{2}}P^t x\|_2^2, \quad (2.42)$$

onde

$$D = \text{diag}(\mu_1, \dots, \mu_n) \quad (2.43)$$

e

$$\mu_i = \begin{cases} \frac{1}{a_i}[(1 - \omega a_i)^{-k} - 1], & \text{se } a_i \neq 0, \\ 1, & \text{caso contrário;} \end{cases} \quad (2.44)$$

para $i = 1, \dots, n$.

Exemplo 2.2. Se quisermos resolver o problema

$$\text{minimizar } \|Ax - b\|^2 + \alpha^2 \|Bx\|^2, \quad (2.45)$$

para uma matriz $B \in \mathbb{R}^{p \times n}$, usando um método iterativo, Hanke e Hansen sugerem em [35] que os métodos devam ser aplicados às equações normais preconditionadas

$$B_A^\dagger (B_A^\dagger)^t A^t A x = B_A^\dagger (B_A^\dagger)^t A^t b, \quad (2.46)$$

onde B_A^\dagger é a inversa de B pesada por A (veja [22]), definida por

$$B_A^\dagger = W \begin{pmatrix} \text{diag}(\tau_i^{-1}) \\ 0 \end{pmatrix} V^t, \quad (2.47)$$

se a decomposição GSVD de (A, B) (veja [31]) for dada por

$$A = U \begin{pmatrix} \text{diag}(\rho_i) & 0 \\ 0 & I_{n-p} \\ 0 & 0 \end{pmatrix} W^{-1}, \quad B = V \begin{pmatrix} \text{diag}(\tau_i) & 0 \end{pmatrix} W^{-1}, \quad (2.48)$$

onde $U = [u_1, \dots, u_m] \in \mathbb{R}^{m \times m}$, $V = [v_1, \dots, v_p] \in \mathbb{R}^{p \times p}$ são ortogonais e $W = [w_1, \dots, w_n] \in \mathbb{R}^{n \times n}$ é inversível.

Vamos mostrar a seguir que nenhum iterado do método de Landweber aplicado a (2.46) corresponde a solução do problema (2.45), se $p = n$ e B for inversível.

Em termos da notação apresentada neste Capítulo, (2.45) pode ser escrito como

$$\text{minimizar } \|Ax - b\|^2 + \|x\|_Q^2, \quad (2.49)$$

onde $Q^{-1} = \sqrt{\alpha} B^t B$. De (2.48) obtemos

$$W^t Q^{-1} W = \text{diag}(\sqrt{\alpha} \tau_i) \quad (2.50)$$

e

$$W^t A^t A W = \text{diag}(\rho_i). \quad (2.51)$$

Assim, a matriz W diagonaliza simultaneamente Q^{-1} e $A^t A$ e usando a demonstração do Teorema 2.5 obtemos que a matriz

$$M = W \text{diag} \left(\frac{1}{\rho_i^2} \left[1 - \left(1 + \frac{\rho_i^2}{\sqrt{\alpha} \tau_i} \right)^{-\frac{1}{k}} \right] \right) W^t \quad (2.52)$$

é tal que

$$x^k = x^{k-1} + M A^t (b - A x^{k-1}) \quad (2.53)$$

é solução de (2.45). Por outro lado, na estratégia de Hanke e Hansen

$$x^k = x^{k-1} + \tilde{M} A^t (b - A x^{k-1}), \quad (2.54)$$

onde

$$\tilde{M} = \omega B_A^t (B_A^t)^t = W \text{diag}(\omega \tau_i^2) W^t, \quad (2.55)$$

ou seja, com esta estratégia não se consegue o objetivo pretendido.

Capítulo 3

Extensão de Validação Cruzada Generalizada (GCV)

3.1 Introdução

Consideremos o problema de encontrar uma solução x de

$$f(x) + \epsilon = b, \quad (3.1)$$

onde f é uma aplicação de \mathbb{R}^n em \mathbb{R}^m , b é um m -vetor de observações e ϵ um m -vetor de erros. Supõe-se que as componentes de ϵ são variáveis aleatórias de média zero não correlacionadas e com variância σ^2 (desconhecida), isto é,

$$E\{\epsilon\} = 0, \quad E\{\epsilon\epsilon^t\} = \sigma^2 I, \quad (3.2)$$

onde I é a matriz identidade.

No caso em que $f(x) = Ax$, onde A é uma matriz de posto completo de ordem $m \times n$, o Teorema de Gauss-Markov garante que a solução de quadrados mínimos de (3.1), $\tilde{x} = (A^t A)^{-1} A^t b$, é o melhor estimador linear não viciado de x , no sentido de que tem a menor variância, $E(\|\tilde{x} - x\|^2)$ (veja por exemplo [62]). Mas, se A é mal condicionada esta menor variância ainda é muito alta. Sabe-se que permitir que o estimador seja viciado pode reduzir tremendamente a variância (veja, por exemplo, [69,70]).

Suponhamos que temos uma família $\{x_\lambda\}$ de estimadores de soluções de (3.1-3.2). Um exemplo típico é a família de soluções do problema regularizado

$$\text{minimizar } \|f(x) - b\|^2 + \lambda x^t B x, \quad (3.3)$$

para $\lambda > 0$, onde B é uma matriz de ordem $n \times n$, que introduz informação *a priori* ao problema, no sentido de que devemos esperar da solução que $x^t B x$ seja pequeno. Por exemplo $B = L^t L$, onde L é a discretização de um operador de derivadas.

Coloca-se a questão de encontrar um valor de λ que corresponda a menor perda de informação possível. Uma possibilidade poderia ser encontrar um valor de λ que minimizasse a variância $E(\|x_\lambda - x\|^2)$. Entretanto, tal valor de λ depende de σ^2 e do valor de x , que é desconhecido. Uma outra possibilidade poderia ser obter um valor de λ que em média produzisse o menor erro quadrático médio em estimar as componentes de $f(x)$, isto é, um valor de λ que

$$\text{minimize } E\{T(\lambda)\}, \quad (3.4)$$

onde

$$T(\lambda) = \frac{1}{m} \|f(x) - f(x_\lambda)\|^2 \quad (3.5)$$

e x é uma solução do sistema de equações $f(x) = b - \epsilon$.

Assim, temos o mesmo problema de não conhecermos o valor de x e de σ^2 , mas neste caso existem métodos que encontram aproximações para a solução de (3.4-3.5). Um deles que é muito popular é o de *Validação Cruzada Generalizada* (GCV).

Para $f(x) = Ax$, onde A é uma matriz de ordem $m \times n$ e x_λ uma função linear dos dados, ou seja, de b , define-se a *matriz de influência* como a matriz $A(\lambda)$ tal que

$$A(\lambda)b = Ax_\lambda. \quad (3.6)$$

Para $A(\lambda)$ simétrica e semi positiva definida, Wahba [30,15] definiu o funcional de GCV, $V(\lambda)$ por

$$V(\lambda) = \frac{\frac{1}{m} \|b - Ax_\lambda\|^2}{\left[\frac{1}{m} \text{Tr}(I - A(\lambda))\right]^2}. \quad (3.7)$$

Usando GCV, o parâmetro ótimo é determinado como um minimizador de $V(\lambda)$. Esta estimativa é uma versão invariante por rotação do funcional definido por Allen

em [1], chamado de PRESS ou *Validação Cruzada*. Uma discussão sobre a origem de GCV se encontra em [30]. Em [30,15] foi provado que, para $A(\lambda)$ uma matriz simétrica e semi positiva definida, se λ_0 é o minimizador de $E\{T(\lambda)\}$ e $\tilde{\lambda}$ é o minimizador de $E\{V(\lambda)\}$, então

$$\frac{E\{T(\tilde{\lambda})\}}{E\{T(\lambda_0)\}} \leq \frac{1 + h(\lambda_0)}{1 - h(\tilde{\lambda})}, \quad (3.8)$$

onde $h(\lambda) = \left(2\mu_1(\lambda) + \frac{\mu_1(\lambda)^2}{\mu_2(\lambda)}\right) \frac{1}{(1 - \mu_1(\lambda))^2}$, $\mu_1(\lambda) = (1/m)TrA(\lambda)$ e $\mu_2(\lambda) = (1/m)TrA^2(\lambda)$. Assim, se $h(\lambda_0)$ e $h(\tilde{\lambda})$ são pequenos, o erro quadrático médio para λ igual ao minimizador de $E\{V(\lambda)\}$ não é muito maior que o menor erro quadrático médio possível $\min_{\lambda} E\{T(\lambda)\}$. Também em [30] foi mostrado que para a família de soluções do problema de regularização de Tikhonov (3.3), $x_{\lambda} = (A^t A + \lambda B)^{-1} A^t b$, se

(i) O número de equações m tende a infinito, enquanto n permanece fixo, ou

(ii) Os auto-valores de AA^t são da forma $mk^{-\alpha}$, para algum $\alpha > 1$ e $\frac{1}{m}Tr(AA^t)$ é finito, quando m tende a infinito

então, $\frac{ET(\tilde{\lambda})}{ET(\lambda_0)} \downarrow 1$.

Em [61,59] é definida a seguinte extensão do funcional de GCV para o problema regularizado não linear (3.3)

$$\tilde{V}(\lambda) = \frac{\frac{1}{m} \|b - f(x_{\lambda})\|^2}{\left[\frac{1}{m} Tr(I - \tilde{A}(\lambda))\right]^2}. \quad (3.9)$$

onde $\tilde{A}(\lambda) = J(x_{\lambda})(J(x_{\lambda})^t J(x_{\lambda}) + \lambda B)^{-1} J(x_{\lambda})^t$ e $J(x)$ é o Jacobiano de $f(x)$. Os autores não provam nenhum resultado que indique que o mínimo de $E\tilde{V}(\lambda)$ esteja próximo do mínimo de $ET(\lambda)$, mas apresentam resultados numéricos que indicam ser esta uma técnica promissora também no caso não linear.

Definimos a seguinte extensão do funcional de GCV para problemas não lineares, mas para uma família de estimadores quaisquer.

$$V(\lambda) = \frac{\frac{1}{m} \|b - f(x_{\lambda})\|^2}{\left[\frac{1}{m} Tr(I - DA(\lambda)(b))\right]^2}, \quad (3.10)$$

onde o *operador de influência* é definido por $A(\lambda)(b) = f(x_\lambda)$ e $DA(\lambda)(b)$ é o Jacobiano da aplicação $A(\lambda)(\cdot)$ avaliado em b . Provamos que também vale (3.8) no caso geral.

Uma dificuldade na aplicação de GCV como critério para escolha do parâmetro λ está no cálculo do traço no denominador de (3.7) para diversos valores de λ . Para a regularização de Tikhonov, o método standard é usar a decomposição em valores singulares de A (veja [30,77]). Em [21,20] Eldén, ao invés disso, usa a bidiagonalização de A . Mas, para problemas de grande porte em que a matriz A é não estruturada, torna-se difícil a aplicação de ambos os métodos. Para superar esta dificuldade existe, no caso linear, uma variante de GCV devida a Girard [28] chamada Monte-Carlo GCV, que pode ser implementada a baixo custo computacional. O método consiste em calcular uma estimativa confiável do traço de $I - A(\lambda)$. Este algoritmo é simplesmente o seguinte:

(i) Gera-se um vetor pseudo-aleatório $w = (w_1, \dots, w_m)^t$ de uma distribuição normal;

(ii) Toma-se

$$\frac{w^t(w - A(\lambda)w)}{w^tw} \quad (3.11)$$

como uma aproximação de

$$\frac{1}{m} \text{Tr}(I - A(\lambda)). \quad (3.12)$$

Girard provou a confiabilidade deste método para problemas de grande porte, apenas se $A(\lambda)$ for simétrica, ou pelo menos simetrizável; em [28] provou que (3.11) é uma estimativa confiável do traço (3.12), no sentido de que

$$E\left(\frac{w^t(w - A(\lambda)w)}{w^tw}\right) = \frac{1}{m} \text{Tr}(I - A(\lambda)) \quad (3.13)$$

e

$$\sigma\left(\frac{w^t(w - A(\lambda)w)}{w^tw}\right) = \sqrt{\frac{2}{m+2}} d, \quad (3.14)$$

onde $\sigma(\cdot)$ é o desvio padrão e d é o desvio padrão dos auto-valores de $I - A(\lambda)$ ou da simetrização de $I - A(\lambda)$.

Mostramos aqui uma generalização deste resultado, para matrizes quaisquer.

3.2 GCV Estendido para problemas não lineares

Gostaríamos de obter um bom estimador para o erro quadrático médio em estimar $f(x)$, ou seja, para

$$T(\lambda) = \frac{1}{m} \|f(x) - f(x_\lambda)\|^2, \quad (3.15)$$

onde x é uma solução do sistema linear $f(x) = b - \epsilon$. O ponto de partida para aproximar $T(\lambda)$ é

$$U(\lambda) = \frac{1}{m} \|b - f(x_\lambda)\|^2 \quad (3.16)$$

como mostram os resultados seguintes.

Lema 3.1 *Sejam $F(\lambda)$, $g(\lambda)$, $G(\lambda)$, $r(\lambda)$ e $H(\lambda)$ funções reais de variáveis reais e α um parâmetro real. Se*

$$G(\lambda) = F(\lambda) + \alpha(1 - 2g(\lambda)) + r(\lambda) \quad (3.17)$$

e

$$H(\lambda) = \frac{G(\lambda)}{(1 - g(\lambda))^2}, \quad (3.18)$$

então é válida a seguinte expressão

$$\frac{H(\lambda) - F(\lambda) - \alpha}{F(\lambda)} = \frac{1}{(1 - g(\lambda))^2} \left(-\alpha \frac{g(\lambda)^2 + r(\lambda)}{F(\lambda)} + 2g(\lambda) - g(\lambda)^2 \right) \quad (3.19)$$

Dem. De (3.17) e (3.18) temos que

$$\begin{aligned} \frac{H(\lambda) - F(\lambda) - \alpha}{F(\lambda)} &= \frac{1}{(1 - g(\lambda))^2} \left(\frac{F(\lambda) + \alpha(1 - 2g(\lambda)) + r(\lambda)}{F(\lambda)} - \frac{(1 - g(\lambda))^2(F(\lambda) + \alpha)}{F(\lambda)} \right) \\ &= \frac{1}{(1 - g(\lambda))^2} \left(1 - \alpha \frac{g(\lambda)^2}{F(\lambda)} - (1 - g(\lambda))^2 + \frac{r(\lambda)}{F(\lambda)} \right), \end{aligned} \quad (3.20)$$

de onde segue (3.19). ■

Lema 3.2 *Sejam $f(\lambda)$, $g(\lambda)$, $h(\lambda)$ funções reais de variável real, com $f(\lambda) > 0$ e α um parâmetro real. Se $f(\lambda)$ tem um mínimo global λ_0 , $g(\lambda)$ tem um mínimo global $\tilde{\lambda}$ e vale que*

$$\frac{|f(\lambda) - g(\lambda) - \alpha|}{f(\lambda)} < h(\lambda) \quad (3.21)$$

então

$$\frac{f(\tilde{\lambda})}{f(\lambda_0)} < \frac{1 + h(\lambda_0)}{1 - h(\tilde{\lambda})}. \quad (3.22)$$

Dem. De (3.21) segue que para todo λ vale

$$f(\lambda)(1 - h(\lambda)) < g(\lambda) + \alpha < f(\lambda)(1 + h(\lambda)). \quad (3.23)$$

Assim de (3.23) temos que

$$f(\tilde{\lambda})(1 - h(\tilde{\lambda})) < g(\tilde{\lambda}) + \alpha \leq g(\lambda_0) + \alpha < f(\lambda_0)(1 + h(\lambda_0)). \quad (3.24)$$

De onde segue o resultado. ■

Proposição 3.3 *Seja $\{x_\lambda\}$ uma família de estimadores de solução do problema (3.1), para o qual vale (3.2). Se para cada λ , $f(x_\lambda) = A(\lambda)(b)$, onde $A(\lambda)(\cdot)$ é uma aplicação continuamente diferenciável, então*

$$EU(\lambda) = ET(\lambda) + \sigma^2 \left(1 - \frac{2}{m} \text{Tr}(DA(\lambda)(b)) \right) + E(\epsilon^t O_\lambda(\epsilon^t \epsilon)), \quad (3.25)$$

onde $U(\lambda)$ é dada por (3.16), $T(\lambda)$ por (3.15), $DA(\lambda)(b)$ é o Jacobiano da aplicação $A(\lambda)(\cdot)$ em relação a b e a função $O_\lambda(\epsilon^t \epsilon)$ é tal que $\|O_\lambda(\epsilon^t \epsilon)\| \leq M \epsilon^t \epsilon$, para algum $M > 0$.

Dem. Como vale (3.1), temos que

$$\begin{aligned} \|b - f(x_\lambda)\|^2 - \|f(x) - f(x_\lambda)\|^2 &= \|f(x) + \epsilon - f(x_\lambda)\|^2 - \|f(x) - f(x_\lambda)\|^2 = \\ &= \|\epsilon\|^2 + 2\epsilon^t(f(x) - f(x_\lambda)). \end{aligned} \quad (3.26)$$

Aqui, desenvolvemos os quadrados para obter (3.26). Além disso como $A(\lambda)(\cdot)$ é continuamente diferenciável, temos que

$$A(\lambda)(f(x)) = A(\lambda)(f(x) + \epsilon) - DA(\lambda)(f(x) + \epsilon)\epsilon + O_\lambda(\epsilon^t \epsilon) \quad (3.27)$$

onde $DA(\lambda)(f(x) + \epsilon)$ é o Jacobiano da aplicação $A(\lambda)(\cdot)$ em relação a b avaliada em $f(x) + \epsilon$ e a função $O_\lambda(\epsilon^t \epsilon)$ é tal que $\|O_\lambda(\epsilon^t \epsilon)\| \leq M \epsilon^t \epsilon$, para algum $M > 0$. Assim, usando (3.1) e a equação acima obtemos

$$f(x_\lambda) = A(\lambda)(b) = A(\lambda)(f(x) + \epsilon) = A(\lambda)(f(x)) + DA(\lambda)(b)\epsilon + O_\lambda(\epsilon^t \epsilon). \quad (3.28)$$

Dividindo (3.26) por m e calculando a esperança obtemos

$$\begin{aligned} E(U(\lambda)) &= E(T(\lambda)) + \frac{1}{m} E(\|\epsilon\|^2) + \frac{2}{m} [E(\epsilon^t f(x)) - E(\epsilon^t f(x_\lambda))] = \\ &= E(T(\lambda)) + \sigma^2 - \frac{2}{m} [E(\epsilon^t A(\lambda)(f(x))) + E(\epsilon^t DA(\lambda)(b)\epsilon) + E(\epsilon^t O_\lambda(\epsilon^t \epsilon))] = \\ &= E(T(\lambda)) + \sigma^2 [1 - \frac{2}{m} Tr(DA(\lambda)(b))] + E(\epsilon^t O_\lambda(\epsilon^t \epsilon)), \end{aligned} \quad (3.29)$$

aqui usamos (3.2), (3.28), a linearidade da esperança e os fatos de que

$$E(\epsilon^t f(x)) = E(\epsilon^t) f(x) = 0$$

e que para qualquer matriz B de ordem $m \times m$, em vista de (3.2) vale

$$E(\epsilon^t B \epsilon) = \sum_{ij} B_{ij} E(\epsilon_i \epsilon_j) = \sigma^2 Tr(B).$$

Fica assim demonstrado o resultado. ■

Teorema 3.4 *Seja $\{x_\lambda\}$ uma família de estimadores de solução do problema (3.1), para o qual vale (3.2). Se para cada λ , $f(x_\lambda) = A(\lambda)(b)$, onde $A(\lambda)$ é uma aplicação continuamente diferenciável, então para o funcional de Validação Cruzada Generalizada (GCV) definido por*

$$V(\lambda) = \frac{\frac{1}{m} \|b - f(x_\lambda)\|^2}{[\frac{1}{m} Tr(I - DA(\lambda)(b))]^2}, \quad (3.30)$$

vale a seguinte expressão:

$$\frac{EV(\lambda) - ET(\lambda) - \sigma^2}{ET(\lambda)} = \frac{1}{(1 - \mu_1(\lambda))^2} \left(\frac{-\sigma^2 \mu_1(\lambda)^2 + r(\lambda)}{ET(\lambda)} + 2\mu_1(\lambda) - \mu_1(\lambda)^2 \right), \quad (3.31)$$

onde $\mu_1(\lambda) = \frac{1}{m} Tr(DA(\lambda)(b))$ e $r(\lambda) = E(\epsilon^t O_\lambda(\epsilon^t \epsilon))$.

Dem. Pela Proposição 3.3 vale a expressão (3.25). Aplicando o Lema 3.1 obtemos (3.31). ■

Corolário 3.5 *Com as mesmas hipóteses do Teorema 3.4 vale a seguinte desigualdade*

$$\frac{ET(\tilde{\lambda}_1)}{ET(\lambda_0)} \leq \frac{1 + h(\lambda_0)}{1 - h(\tilde{\lambda}_1)} \quad (3.32)$$

onde λ_0 é o mínimo global de $ET(\lambda)$, $\tilde{\lambda}_1$ é o mínimo global de $EV(\lambda)$;

$$h(\lambda) = \frac{1}{(1 - \mu_1(\lambda))^2} \left(\frac{\sigma^2 \mu_1(\lambda)^2 + |r(\lambda)|}{ET(\lambda)} + 2|\mu_1(\lambda)| + \mu_1(\lambda)^2 \right), \quad (3.33)$$

e $\mu_1(\lambda)$, $r(\lambda)$ são definidos no Teorema 3.4.

Dem. Este corolário decorre simplesmente do Teorema 3.4 e do Lema 3.2. ■

Assim, se $h(\lambda_0)$ e $h(\tilde{\lambda})$ são pequenos, o erro quadrático médio para λ igual ao minimizador de $E\{V(\lambda)\}$ não é muito maior que o menor erro quadrático médio possível $\min_{\lambda} E\{T(\lambda)\}$.

Um problema, assim, é provar que sob condições satisfeitas na prática $h(\tilde{\lambda})$ e $h(\lambda_0)$ tendem a zero quando a dimensão do problema tende a infinito. Em [77,30] existem resultados deste tipo, para o caso em que $f(x) = Ax$, onde A é uma matriz de ordem $m \times n$ e $\{x_{\lambda}\}$ é a família de soluções do problema de Tikhonov (3.3), ou seja, $A(\lambda) = A(A^t A + \lambda B)^{-1} A^t$, como dissemos na introdução deste capítulo.

Corolário 3.6 *Com as mesmas hipóteses do Teorema 3.4, se $f(x) = Ax$, onde A é uma matriz de ordem $m \times n$ e para cada λ , $A(\lambda)$ é uma aplicação afim, ou seja, da forma*

$$A(\lambda)(b) = A_0(\lambda)b + b_0(\lambda), \quad (3.34)$$

então

$$\frac{ET(\tilde{\lambda})}{ET(\lambda_0)} \leq \frac{1 + h(\lambda_0)}{1 - h(\tilde{\lambda})}, \quad (3.35)$$

onde $\tilde{\lambda}$ é o mínimo global de $EV(\lambda)$,

$$h(\lambda) = \frac{1}{(1 - \mu_1(\lambda))^2} \left(\frac{\mu_1(\lambda)^2}{\mu_2(\lambda)} + 2|\mu_1(\lambda)| + \mu_1(\lambda)^2 \right) \quad (3.36)$$

e onde $\mu_1(\lambda) = \frac{1}{m} \text{Tr}(A_0(\lambda))$ e $\mu_2(\lambda) = \frac{1}{m} \text{Tr}(A_0(\lambda)^t A_0(\lambda))$.

Dem. Em vista do Corolário 3.5, basta mostrarmos que

$$\frac{\sigma^2 \mu_1(\lambda)^2}{ET(\lambda)} < \frac{\mu_1(\lambda)^2}{\mu_2(\lambda)} \quad (3.37)$$

Agora,

$$\begin{aligned} \|Ax - Ax_\lambda\|^2 &= \|Ax - A_0(\lambda)b - b_0(\lambda)\|^2 = \\ &= \|Ax - A_0(\lambda)Ax - b_0(\lambda) - A_0(\lambda)\epsilon\|^2 = \\ &= \|(I - A_0(\lambda))Ax - b_0(\lambda)\|^2 + \|A_0(\lambda)\epsilon\|^2 + \\ &\quad - 2\epsilon^t A_0(\lambda)^t [(I - A_0(\lambda))Ax - b_0(\lambda)]. \end{aligned} \quad (3.38)$$

Calculando a esperança da expressão acima, obtemos

$$ET(\lambda) = \|(I - A_0(\lambda))Ax - b_0(\lambda)\|^2 + \sigma^2 \mu_2(\lambda), \quad (3.39)$$

de onde decorre (3.37) e o resultado. ■

O resultado acima, por si, já é uma generalização de (3.8) que foi demonstrado em [30,15].

3.3 Monte-Carlo GCV

Como dissemos na Introdução deste Capítulo, uma dificuldade na aplicação de GCV como critério para escolha do parâmetro λ está no cálculo do traço no denominador de (3.7) para diversos valores de λ . Uma forma de superar esta dificuldade é usar um método do tipo Monte-Carlo para obter uma estimativa confiável do traço. Apresentamos a seguir resultados nesta direção que são uma generalização para matrizes quaisquer daqueles que Girard provou em [28].

Teorema 3.7 *Seja $w = (w_1, \dots, w_n)$ um vetor de componentes aleatórias com distribuição normal, isto é, $w \sim \mathcal{N}(0, I_{n \times n})$. Seja B uma matriz de ordem $n \times n$, e $T_B(w)$ a variável aleatória*

$$T_B(w) = w^t B w. \quad (3.40)$$

Então $T_B(w)$ é um estimador não viciado de $Tr(B)$ com desvio padrão $(Tr(BB^t) + Tr(B^2))^{1/2}$, isto é,

$$ET_B(w) = Tr(B) \quad e \quad (3.41)$$

$$\sigma(T_B(w)) = (Tr(BB^t) + Tr(B^2))^{1/2}. \quad (3.42)$$

Dem. Como $E(w_i^2) = 1$ e $E(w_i w_j) = 0$, se $i \neq j$, então

$$ET_B(w) = E\left(\sum_{i,j} B_{ij} w_i w_j\right) = \sum_{i,j} B_{ij} E(w_i w_j) = Tr(B). \quad (3.43)$$

Agora, como $E(w_i^4) = 3$, $E(w_i w_j w_k w_l) = 0$, se $i \neq j, k, l$ e $E(w_i^2 w_j^2) = 1$, se $i \neq j$, então

$$\begin{aligned} E((T_B(w))^2) &= E\left(\left(\sum_{i,j} B_{ij} w_i w_j\right)^2\right) = \\ &= E\left(\sum_{i,j,k,l} B_{ij} B_{kl} w_i w_j w_k w_l\right) = \\ &= \sum_{i,j,k,l} B_{ij} B_{kl} E(w_i w_j w_k w_l) = \end{aligned}$$

$$\begin{aligned}
&= 3 \sum_i B_{ij}^2 + \sum_{i \neq j} B_{ii} B_{jj} + \sum_{i \neq j} B_{ij} B_{ij} + \sum_{i \neq j} B_{ij} B_{ji} = \\
&= \text{Tr}(B)^2 + \text{Tr}(BB^t) + \text{Tr}(B^2). \tag{3.44}
\end{aligned}$$

Assim,

$$\begin{aligned}
\sigma^2(T_B(w)) &= E((T_B(w))^2) - E(T_B(w))^2 = \\
&= \text{Tr}(B)^2 + \text{Tr}(BB^t) + \text{Tr}(B^2) - \text{Tr}(B)^2. \tag{3.45}
\end{aligned}$$

De onde segue (3.42). ■

Corolário 3.8 *Sejam w e B como no teorema anterior. Seja $T_B^*(w)$ a variável aleatória definida por*

$$T_B^*(w) = \frac{w^t B w}{w^t w}. \tag{3.46}$$

Então $T_B^(w)$ é um estimador não viciado de $\frac{1}{n} \text{Tr}(B)$ com desvio padrão $\sqrt{\frac{\frac{\text{Tr}(BB^t) + \text{Tr}(B^2)}{n} - 2(\frac{\text{Tr}(B)}{n})^2}{n+2}}$, isto é,*

$$ET_B^*(w) = \frac{1}{n} \text{Tr}(B) \quad e \tag{3.47}$$

$$\sigma(T_B^*(w)) = \sqrt{\frac{\frac{\text{Tr}(BB^t) + \text{Tr}(B^2)}{n} - 2(\frac{\text{Tr}(B)}{n})^2}{n+2}}. \tag{3.48}$$

Dem. Uma maneira de encontrar os momentos de $T_B^*(w)$ é observar que $(w^t B w)/(w^t w)$ e $w^t w$ são independentemente distribuídos. Então é fácil deduzir que os momentos de $T_B^*(w)$ são iguais aos momentos de $w^t B w$ dividido pelos momentos de $w^t w$, assim como fez Girard em [28]. Assim,

$$E\left(\frac{w^t B w}{w^t w}\right) = \frac{E(w^t B w)}{E(w^t w)} = \frac{\text{Tr}(B)}{n} \tag{3.49}$$

e

$$E\left(\left(\frac{w^t B w}{w^t w}\right)^2\right) = \frac{\text{Tr}(B)^2 + \text{Tr}(BB^t) + \text{Tr}(B^2)}{n^2 + 2n}. \tag{3.50}$$

Agora,

$$\begin{aligned}
\sigma^2(T_B^*(w)) &= E((T_B^*(w))^2) - E(T_B^*(w))^2 = \\
&= \frac{n(\text{Tr}(BB^t) + \text{Tr}(B^2)) - 2\text{Tr}(B)^2}{(n^2 + 2n)n} = \\
&= \frac{\frac{\text{Tr}(BB^t) + \text{Tr}(B^2)}{n} - 2\left(\frac{\text{Tr}(B)}{n}\right)^2}{n + 2}.
\end{aligned} \tag{3.51}$$

O que demonstra o resultado. ■

Observemos que se B é simétrica, então o Corolário 3.8 recai no Teorema 2.2 de [28], que para $B = I - A(\lambda)$ corresponde a (3.13) e (3.14).

Se $\|B\|_1$, $\|B\|_2$ ou $\|B\|_\infty$, permanece limitada quando n tende a infinito, então $\sigma(T_B^*(w))$ tende a zero com a velocidade de $1/\sqrt{n+2}$ como mostramos abaixo.

Corolário 3.9 *Sejam w e B como no teorema anterior. Se $\|B\|_i \leq M$, para $i = 1, 2$ ou ∞ e algum $M > 0$, então*

$$\sigma(T_B^*(w)) \leq \frac{2M}{\sqrt{n+2}}. \tag{3.52}$$

Dem. Como para uma matriz $A \in \mathbb{R}^{n \times n}$ são válidas as desigualdades (veja, por exemplo [31, Cap. 2])

$$\max |A_{ij}| \leq \|A\|_i \tag{3.53}$$

e

$$\|A\|_F \leq \sqrt{n} \|A\|_i; \tag{3.54}$$

para $i = 1, 2$ e ∞ , então, obtemos as seguintes relações

$$|\text{Tr}(B)| \leq \sum_i |B_{ii}| = n \max |B_{ii}| \leq n \|B\|_i \leq n M, \tag{3.55}$$

que leva a

$$|\text{Tr}(B^2)| \leq n \|B^2\|_i \leq n \|B\|_i^2 \leq n M^2; \tag{3.56}$$

e

$$\text{Tr}(BB^t) = \sum_{ij} B_{ij}^2 = \|B\|_F^2 \leq n \|B\|_i^2 \leq n M^2, \quad (3.57)$$

para $i = 1, 2$ e ∞ . O resultado segue facilmente de (3.48), usando as desigualdades acima. ■

A desigualdade (3.52) está longe de ser a melhor possível. Para casos não tão gerais, Girard em [28] provou que a velocidade de decrescimento da dispersão $\sigma(T_B^*(w))$ pode ser da ordem de $1/n$, quando n tende a infinito.

Capítulo 4

GCV como critério de parada

4.1 Introdução

Nosso objetivo neste capítulo é aplicar os resultados do capítulo anterior para determinar um critério de parada de um método iterativo. Agora, cada iterado x^k é um estimador da solução de (3.1) e o índice k cumpre o papel do parâmetro λ do capítulo anterior.

Os inconvenientes principais para empregar GCV como critério de parada de um método iterativo eram essencialmente dois. Um deles era o fato de que o funcional de GCV não estava definido para estimadores não lineares quaisquer. Por exemplo, mesmo para um método linear estacionário, o operador de influência é não linear. Os resultados do Capítulo 3 possibilitam superarmos esta dificuldade. O outro inconveniente é o cálculo do traço no denominador do funcional de GCV em cada iteração k .

Com relação ao cálculo do traço no denominador do funcional de GCV existe o seguinte resultado. Para um método iterativo convergente da forma

$$x^{k+1} = x^k + MA^t(b - Ax^k), \quad k = 0, 1, 2, \dots \quad (4.1)$$

onde M é uma matriz de ordem $n \times n$, positiva definida que possua raiz quadrada simétrica (ou seja, M é simétrica e positiva definida) Wahba [78] mostrou que, para

$x_0 = 0$, a matriz de influência é dada por

$$A(k) = I - (I - AMA^t)^k. \quad (4.2)$$

Na seção 4.2 apresentamos uma fórmula mais geral para a matriz de influência de qualquer método estacionário convergente, cujos pontos limite são exatamente as soluções de quadrados mínimos associado a matriz A .

Mesmo neste caso é difícil o cálculo do traço no denominador de (3.7). Em [78], Wahba sugere o uso da decomposição em valores singulares de $AM^{\frac{1}{2}}$. Mas, em muitos problemas aplicam-se métodos iterativos para não se ter que fazer transformações numa matriz como A .

Nossa alternativa para calcular o traço do funcional de GCV é o método Monte-Carlo, tal como sugerido por Girard em [28]. Os Corolários 3.8 e 3.9 estendem a confiabilidade do método para matrizes não necessariamente simétricas. Devemos observar que, para métodos lineares estacionários, não é necessário usar o operador de influência para aplicar Monte-Carlo GCV, já que este é uma aplicação afim da forma $A(k)(y) = A_0(k)y + b_0$ e $A_0(k)w$ é simplesmente a aplicação de A no k -ésimo iterado gerado pelo método, com o vetor $b = w$ e o ponto inicial igual ao vetor nulo. Apesar disso, o conhecimento da expressão do operador de influência possibilita uma diminuição considerável do custo computacional, como mostramos na Seção 4.2.

Nas simulações de PET descritas na Seção 4.3 aplicamos Monte-Carlo GCV como critério de parada de dois métodos iterativos lineares. O primeiro método é ART [39], originalmente inventado por Kaczmarz [43], que é um método linear estacionário cujo operador de influência é uma aplicação afim (se o ponto inicial é não nulo) com a parte linear não simétrica. O segundo método é Gradientes Conjugados aplicado às equações normais preconditionado com SSOR [10]. Neste caso o operador de influência é não linear. Finalmente usamos Monte-Carlo GCV para determinar o parâmetro de relaxação em cada iteração de ART.

Devemos salientar que o método GCV já foi utilizado, antes, por Girard em [29] num problema de Tomografia Computadorizada, mas para o caso em que a matriz tem uma estrutura que possibilita a aplicação de um método direto ao problema de regularização de Tikhonov (3.3) correspondente e também ao cálculo do traço do

denominador do funcional de GCV. No nosso caso a matriz não é estruturada, e o tamanho do problema nos leva ao emprego de métodos iterativos. Monte-Carlo GCV já foi empregado como critério de parada de métodos iterativos, tomando o ponto inicial igual a zero. Em [35], GCV é usado como critério de parada de dois métodos iterativos: ν -métodos e CGNR. No caso de CGNR (gradientes conjugados aplicados às equações normais), o operador de influência é não linear mas os autores substituem o denominador do funcional de GCV, por uma função afim $1 - \frac{1}{m}(n - p) - \frac{1}{m}k$ (onde p é o posto da matriz A), que não é o que encontramos na prática, mesmo se o método for estacionário, como constatamos nos gráficos mostrados nas Figuras 4.4, 4.9 e 4.10. Naturalmente os resultados deixam a desejar. Os ν -métodos como definido por Brakhage em [11] constituem, na verdade, uma família de métodos iterativos (ν é um parâmetro a ser escolhido) obtida como uma modificação de CGNR, em que tanto o parâmetro que otimiza a busca unidimensional, quanto aquele que determina a ortogonalidade dos vetores de busca, são escolhidos independentemente do vetor b (dependendo apenas de ν). Neste caso o operador de influência é linear, simétrico e semi-definido positivo, desde que o ponto inicial do método iterativo seja o vetor nulo.

4.2 Métodos estacionários

Teorema 4.1 *Considere um método convergente da forma*

$$x^{k+1} = x^k + M(b - Ax^k), \quad k = 0, 1, 2, \dots \quad (4.3)$$

onde M é uma matriz $n \times m$. Então o operador de influência, $A(k)$, para este método é dada por

$$A(k)(b) = [I - (I - AM)^k]b + Ax_1^0 + A(I - MA)^k x_2^0. \quad (4.4)$$

Dem. Pelo Teorema 2.1, temos que

$$\mathbb{R}^n = \mathcal{N}(MA) \oplus \mathcal{R}(MA) \quad (4.5)$$

e

$$x^k = x_1^0 + (I - MA)^k x_2^0 + [I - (I - MA)^k](MA)_2^{-1}(Mb)_2 + k(Mb)_1 \quad (4.6)$$

onde $x_1^0, (Mb)_1 \in \mathcal{N}(MA)$ e $x_2^0, (Mb)_2 \in \mathcal{R}(MA)$, são tais que

$$Mb = (Mb)_1 + (Mb)_2, \quad x^0 = x_1^0 + x_2^0, \quad e \quad (MA)_2 = MA \Big|_{\mathcal{R}(MA)}.$$

Sendo o método convergente para todo x_0 , tem que acontecer que $(Mb)_1 = 0$, ou seja, $Mb \in \mathcal{R}(MA)$, para todo m -vetor b . Portanto, por (3.6), temos

$$A(k)(b) = A[I - (I - MA)^k](MA)_2^{-1}Mb + Ax_1^0 + A(I - MA)^k x_2^0. \quad (4.7)$$

Agora,

$$\begin{aligned} A[I - (I - MA)^k](MA)_2^{-1}M &= -A \sum_{j=1}^k \binom{k}{j} (-MA)^j (MA)_2^{-1}M = \\ &= -\sum_{j=1}^k \binom{k}{j} (-AM)^j A(MA)_2^{-1}M = \\ &= -\sum_{j=1}^k \binom{k}{j} (-AM)^{j-1} [-A(MA(MA)_2^{-1})M] = \\ &= -\sum_{j=1}^k \binom{k}{j} (-AM)^{j-1} (-AM) = \\ &= I - (I - AM)^k \end{aligned} \quad (4.8)$$

O que demonstra o resultado. ■

Observemos que apesar do método ser linear e estacionário, para um ponto inicial x_0 não nulo, o operador de influência não é linear e muitas vezes é conveniente escolher x_0 diferente de zero (veja [45,39]).

Seja $A_0(k)$ a parte linear de $A(k)$. Consideremos a aplicação de Monte-Carlo GCV como critério de parada de um método da forma (4.3). Como dissemos na introdução deste Capítulo, não é necessário usar o operador de influência para aplicar (3.11), pois $A_0(k)w$ é simplesmente o produto da matriz A pelo resultado da iteração k do método (4.3), com o vetor $b = w$ e o ponto inicial igual ao vetor nulo. Mas, usando o Teorema 4.1 obtemos o resultado seguinte, que fornece uma forma mais econômica

de calcular uma aproximação para o $Tr(I - A_0(k))$ no funcional de Monte-Carlo GCV (veja Teorema 3.4).

Proposição 4.2 *Para um método da forma (4.3), vale a seguinte igualdade*

$$Tr(I_m - A_0(k)) = \frac{1}{n} Tr[(m - n)I_n + n(I_n - MA)^k], \quad (4.9)$$

onde $A_0(k)$ é a parte linear do operador de influência $A(k)$.

Dem. Pelo Teorema 4.1, vale a seguinte igualdade

$$A_0(k) = I_m - (I_m - AM)^k = - \sum_{j=1}^k \binom{k}{j} (-AM)^j. \quad (4.10)$$

Assim, usando o fato de que $Tr(AB) = Tr(BA)$ e (4.10) deduzimos que

$$Tr(A_0(k)) = Tr\left[- \sum_{j=1}^k \binom{k}{j} (-MA)^j\right] = Tr[I_n - (I_n - MA)^k]. \quad (4.11)$$

Assim, usando (4.11), obtemos

$$Tr(I_m - A_0(k)) = m - Tr(A(k)) = m - n + Tr[(I_n - MA)^k]. \quad (4.12)$$

De onde segue facilmente (4.9). ■

Assim, em vista da proposição anterior temos o seguinte algoritmo para calcular para cada k o funcional de Monte-Carlo GCV.

Algoritmo 4.1.

- (i) Gera-se um vetor pseudo-aleatório $w = (w_1, \dots, w_n)^t \in \mathbb{R}^n$ de uma distribuição normal com desvio padrão igual a 1, ou seja, $w \sim \mathcal{N}(0, I_{n \times n})$;
- (ii) Toma-se

$$\Phi(k) = \left(\frac{w^t((m - n)w - n(I_n - MA)^k w)}{m w^t w} \right)^2, \quad (4.13)$$

onde o produto $(I_n - MA)^k w$ é igual a iteração k do método (4.3), com $x_0 = w$ e $b = 0$ (veja (2.10)), como uma aproximação de $(\frac{1}{m} Tr[I_m - A_0(k)(b)])^2$;

(iii) Finalmente, calcula-se

$$\frac{\frac{1}{m} \|b - Ax^k\|^2}{\Phi(k)} \quad (4.14)$$

como uma aproximação de $V(k)$.

Desta forma, economiza-se o produto da matriz A por um vetor.

4.3 Aplicação na tomografia de emissão de pósitrons (PET)

Fizemos experiências numéricas com o problema de reconstrução de imagens da Tomografia de Emissão de Pósitrons (PET).

Na Tomografia, o objetivo é reconstruir uma função f com suporte compacto, $\Omega \subset \mathbb{R}^2$ por exemplo, conhecendo-se um conjunto de integrais de linha de f sobre retas que interceptam a região Ω . Ou seja,

$$\int_{\Gamma_i} f(r, \phi) ds = y_i, \quad (4.15)$$

onde Γ_i , para cada i , é uma reta que intercepta a região Ω . O conjunto de retas Γ_i depende da geometria do “scanner”. A Fig. 4.1 mostra o esquema de um “scanner” de PET.

Descrevemos agora o método de discretização usado. Assumimos que a função f a ser reconstruída pode ser aproximada por uma combinação linear de funções de uma base,

$$f(r, \phi) \approx \sum_{j=1}^n x_j b_j(r, \phi). \quad (4.16)$$

Assumiremos que Ω é um quadrado. Substituindo a expressão acima na equação (4.15), obtemos

$$\sum_{j=1}^n l_{ij} x_j = y_i, \quad (4.17)$$

onde $l_{ij} = \int_{\Gamma_i} b_j(r, \phi) ds$. A maneira mais simples de escolher um base é subdividir a região Ω , em *pixels* e tomamos funções da base cujo valor é 1 dentro de um pixel específico e zero fora dele. Em [49] são utilizadas funções de base com simetria esférica,

mas aqui nos restringiremos à base formada pelas funções características dos pixels como base. Neste caso l_{ij} é simplesmente o comprimento da interseção da reta Γ_i com o pixel j . Em tal situação somente um número pequeno de l_{ij} são diferentes de zero, ou seja, a matriz do sistema linear é esparsa, mas em compensação, é não estruturada.

Utilizamos o sistema de programação SNARK93 [12] para simular os dados coletados por um tomógrafo de PET e para aplicar algoritmos de reconstrução, que é um programa escrito para auxiliar pesquisadores no desenvolvimento e avaliação de algoritmos de reconstrução de imagens em Tomografia. O SNARK93 é um sistema aberto, no sentido de que estão disponíveis as fontes escritas em FORTRAN 77 (algumas ferramentas são escritas em C). Neste sistema, os algoritmos são implementados em rotinas escritas em FORTRAN 77, que são então compiladas e linkadas ao resto das rotinas do programa.

Algumas informações são passadas ao programa através de um arquivo de entrada, como descrevemos resumidamente a seguir (veja [12] para mais detalhes).

- Informações Geométricas: NELEM, se a região é subdividida em NELEM x NELEM quadrados iguais (pixels); o comprimento do lado de cada pixel (PIXSIZ); a geometria do “scanner”, são possíveis a geometria paralela e a geometria divergente; no caso da geometria divergente, se ARC ou TANGENT; o número de projeções (número de FONTES); a distância da fonte à origem (RAIO) e a distância da fonte aos detectores (Fon-Det); os ângulos de cada projeção (THETA) (veja a Fig. 4.2); o ambiente para a coleta de dados: introdução ou não de erros, são disponíveis QUÂNTICO, ESPALHAMENTO (para Tomografia de Transmissão de raios X), ADITIVO, MULTIPLICATIVO com distribuição normal e POISSON (para PET).
- Imagem teste ou dados experimentais: no caso de imagem teste, podem ser dados objetos geométricos e suas localizações para a composição da imagem.

SNARK93 gera, então, os dados y_i , de acordo com as informações fornecidas acima. Observemos que a expressão FONTE não está de acordo com a Tomografia

de emissão de pósitrons. É que SNARK93 foi escrito inicialmente para simular a Tomografia de Transmissão de raios X. Para simular PET, SNARK93 utiliza muitas rotinas já escritas para a Tomografia de Transmissão. A Fig. 4.3 mostra um esquema de como ele usa a geometria DIVERGENTE do tipo ARC, para simular a coleta de dados de um “scanner” de PET.

A imagem a ser reconstruída (designada por *phantom*) foi obtida de um atlas computadorizado baseado na anatomia média de um cérebro. Ela foi discretizada com $n = 95 \times 95$ pixels. A geometria dos raios é a geometria divergente como mostrado na Fig. 4.2, simulando a geometria típica de PET, como mostrado nas Figs. 4.1 e 4.3. As equações (ou os raios) são tomados na seguinte ordem: Para cada conjunto de raios associado a uma fonte (veja Fig. 4.2), que chamamos de *vista*, os raios são acessados seqüencialmente na ordem anti-horária e depois variamos as vistas, também seqüencialmente, na ordem anti-horária. A ordem com que são tomados os raios pode implicar numa melhora da performance de certos algoritmos, como este que consideramos aqui nestes experimentos (veja, por exemplo, [71,40]). Mas, como o nosso objetivo era apenas verificar a validade de GCV, como critério de parada, tomamos a ordem trivial, descrita acima. A geometria divergente tinha 101 raios \times 300 vistas, sendo os raios igualmente espaçados, produzindo um total de $m = 30292$ equações (8 raios não interceptam a região e as equações correspondentes não são utilizadas).

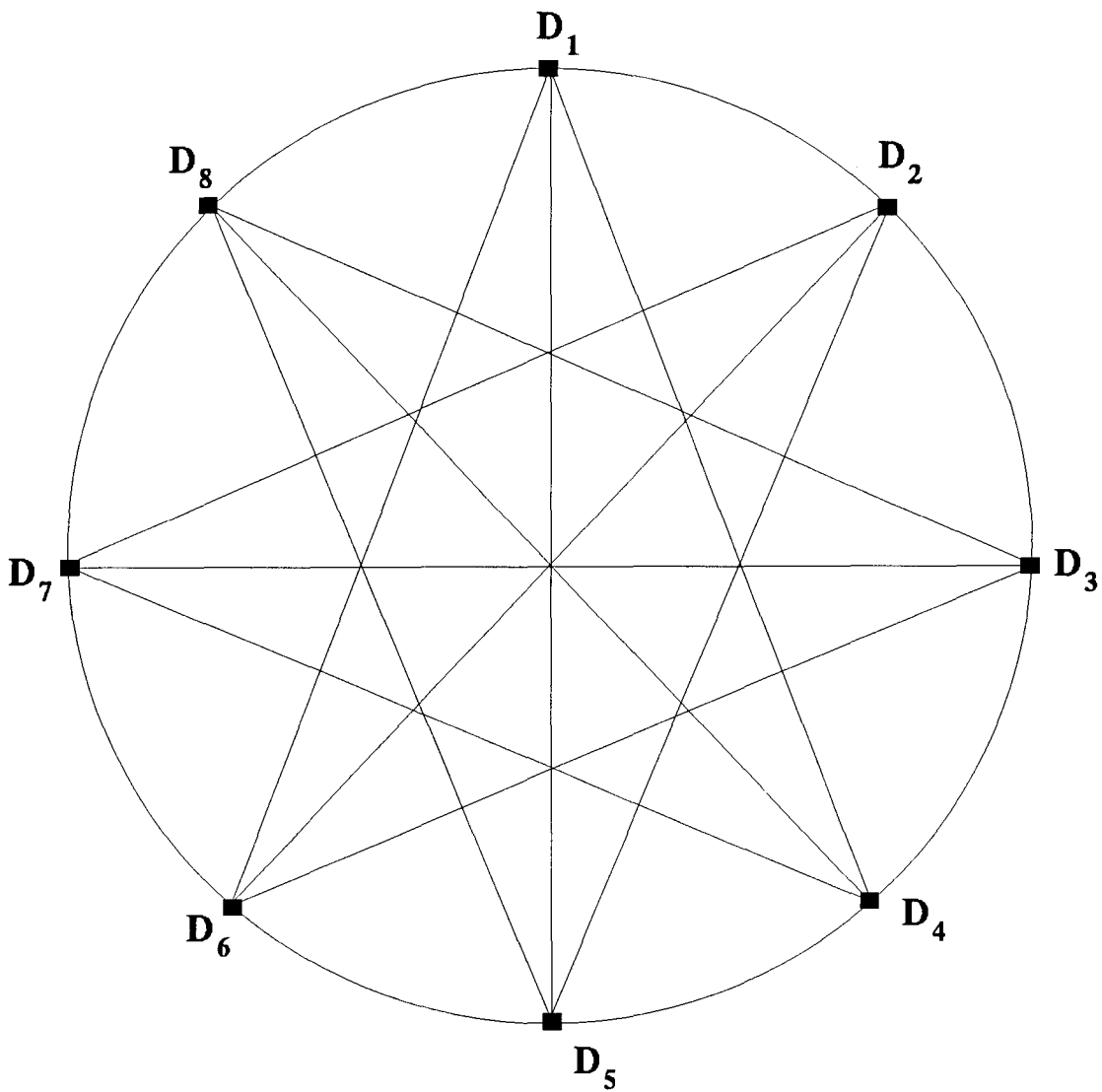


Figura 4.1: Geometria de PET para o caso simplificado de oito detectores

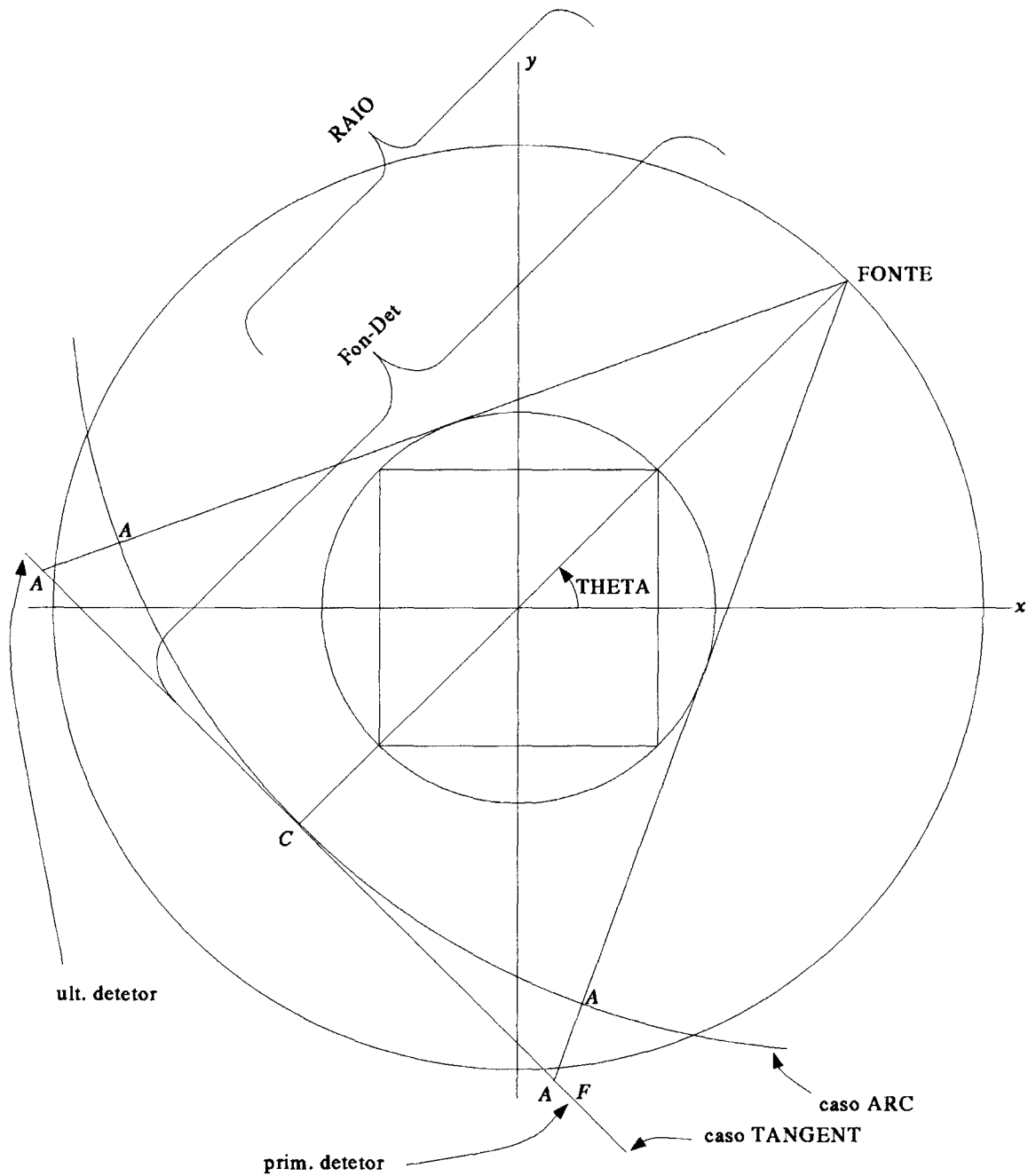


Figura 4.2: Esquema da geometria divergente em SNARK93

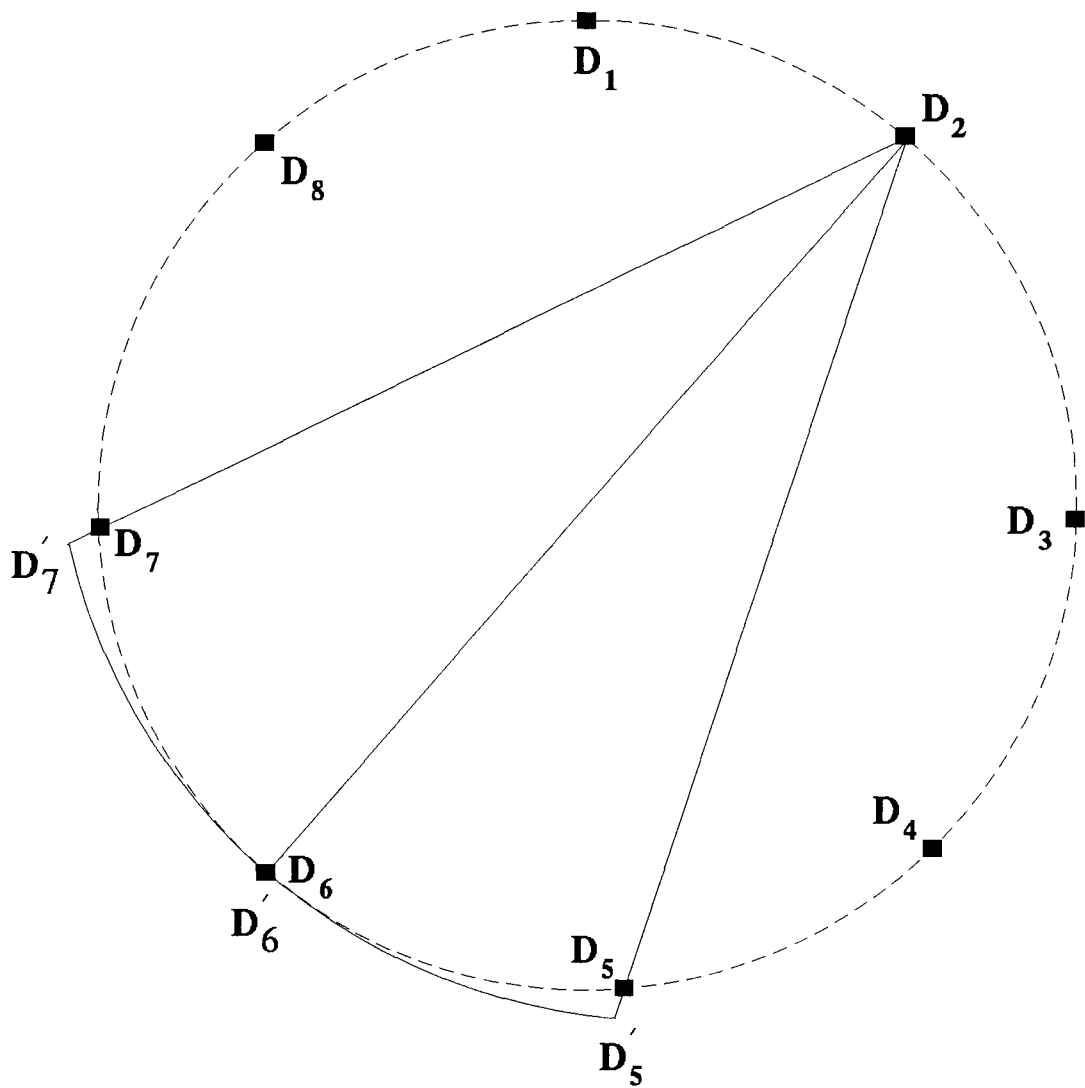


Figura 4.3: Geometria divergente de SNARK93 usada para simular a geometria de PET mostrada na Fig. 4.1

Consideramos a aplicação do algoritmo conhecido na literatura de Reconstrução de Imagens por ART (Técnicas Algébricas de Reconstrução) (veja por exemplo [39]), que é o algoritmo de Kaczmarz [43], definido por

$$\begin{cases} x^{(k+1,1)} = x^k \\ x^{(k+1,i+1)} = x^{(k+1,i)} + \omega_{k+1}(b_i - a_i^t x^{(k+1,i)})a_i \\ x^{k+1} = x^{(k+1,m+1)}, \end{cases} \quad (4.18)$$

para $i = 1, \dots, m$; onde $\omega_k = 0.025$, $k = 1, 2, \dots$; $A = [a_1, \dots, a_m]^t$, $x^0 = (a, \dots, a)^t$ e a é uma aproximação da densidade média do phantom dada por

$$a = \frac{\sum_{i=1}^m b_i}{\sum_{i=1}^m \sum_{j=1}^n a_{ij}}. \quad (4.19)$$

A escolha do ponto inicial uniforme é sugerida por Kaufmann (veja [45]) como uma escolha razoável de ponto inicial no problema de Reconstrução de Imagens em Tomografia, porque de outra forma as variações que aparecem na imagem inicial permanecem por algum tempo enquanto as iterações se desenvolvem.

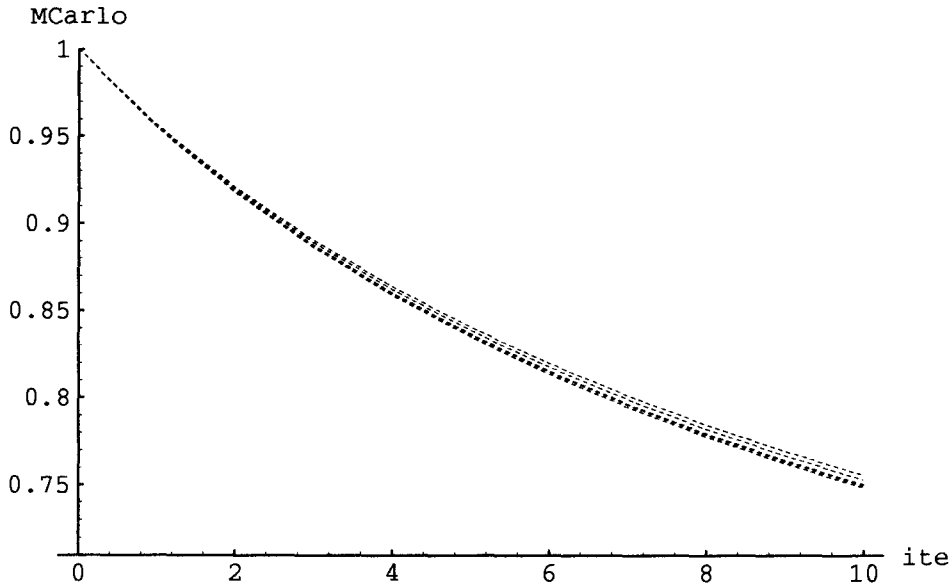


Figura 4.4: Cinco estimativas Monte-Carlo de $[\frac{1}{30292}Tr(I_{30292} - A(k))]^2$ para o método ART com $\omega = 0.025$

Fizemos experiências aplicando o Algoritmo 4.1 para estimar o denominador do funcional de MCarlo. Plotamos no gráfico da Fig. 4.4 os resultados de 5 estimativas

para o denominador do funcional de GCV. A proximidade das curvas indica que há uma pequena dispersão nas estimativas. A suavidade e não interseção das curvas favorece o não aparecimento de mínimos locais e uma precisão maior no valor do mínimo global do funcional correspondente, pois o erro que se comete na estimação do traço em cada iteração é mais ou menos uniforme.

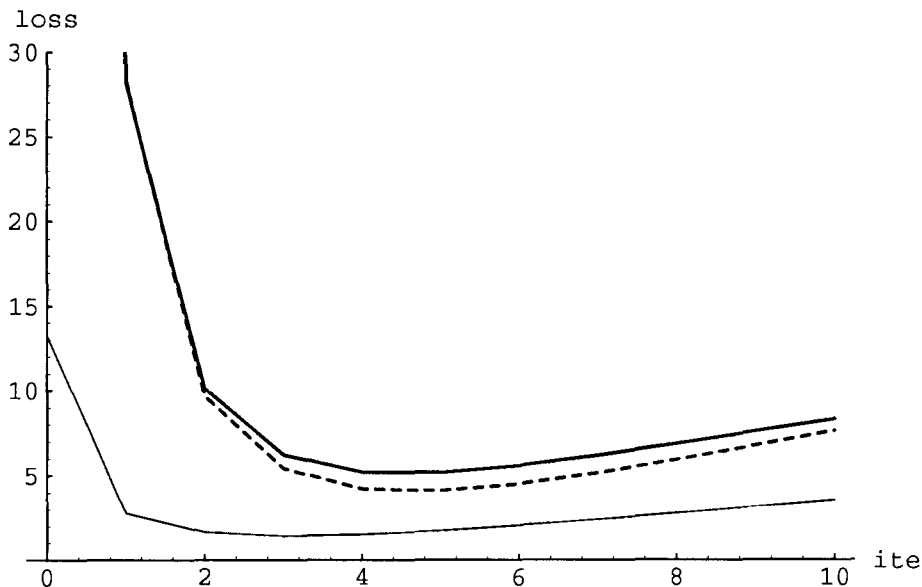


Figura 4.5: Funções de perda típicas para o método ART com $\omega = 0.025$ no caso de $E(\epsilon) = 0$ e $E(\epsilon\epsilon^t) = 100I_{30292}$. A linha fina corresponde a $100\frac{1}{95^2}\|x^k - x\|^2$; a grossa, a $\frac{1}{30292}\|Ax^k - Ax\|^2$; a tracejada, a MCarlo GCV - 100

No gráfico da Fig. 4.5 plotamos diferentes funções de perda para o método ART (4.18), considerando $b = Ax + \epsilon$, com ϵ sendo o chamado *ruído branco* com desvio padrão igual 10, ou seja, o único erro presente é aquele em que as suas componentes são números pseudo-aleatórios de uma distribuição normal com média igual a zero e desvio padrão igual 10, ou ainda, $\epsilon \sim \mathcal{N}(0, 100I_{30292})$.

No gráfico da Fig. 4.6 plotamos novamente diferentes funções de perda para o método ART (4.18), apenas que neste caso, o vetor b contém os erros da discretização, além dos erros que simulam aqueles, típicos de PET, que são obtidos de um gerador de números pseudo-aleatórios com uma distribuição de Poisson (veja [12]), sendo que o número total de fótons contados é igual 2.022.085.

Os gráficos mostram que tanto para o caso em que os erros satisfazem as hipóteses da teoria, quanto para o caso em que os erros são típicos de PET não somente o mínimo de MCarlo GCV coincide com o de $\|Ax^k - Ax\|^2$, mas as curvas são muito semelhantes.

Na Fig. 4.7 estão as imagens que correspondem ao caso em que os erros nos dados são aqueles, típicos de PET. Na Fig. 4.8 aparecem os cortes verticais com as comparações entre as reconstruções em cada iteração e o phantom, também para erros típicos de PET. Nas Figs. 4.7 e 4.8 podemos constatar que a distância medida no contra-domínio de A está mais em acordo com a visualização dos resultados, do que a distância medida no domínio.

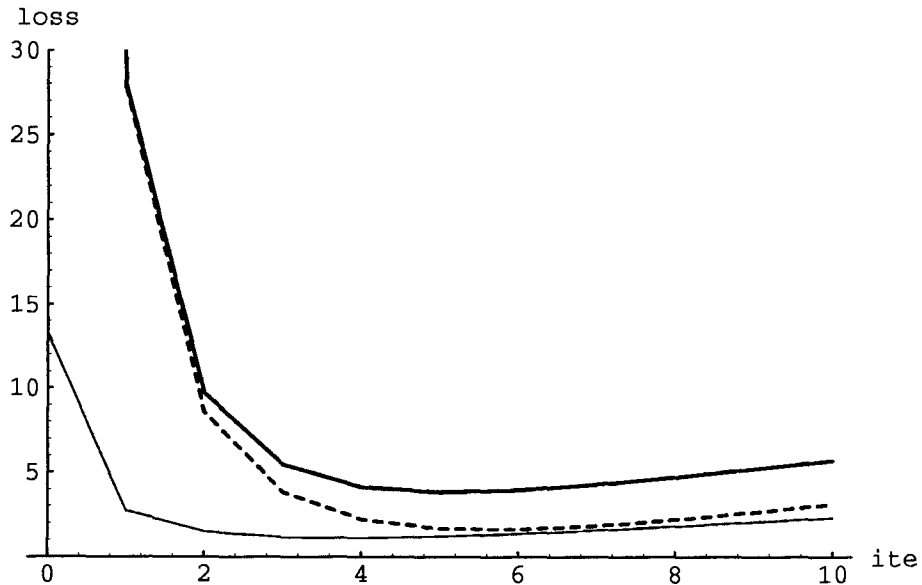


Figura 4.6: Funções de perda típicas para o método ART com $\omega = 0.025$ no caso de erros típicos de PET. A linha fina corresponde a $100 \frac{1}{95^2} \|x^k - x\|^2$; a grossa, a $\frac{1}{30292} \|Ax^k - Ax\|^2$, a tracejada, a MCarlo GCV - 70

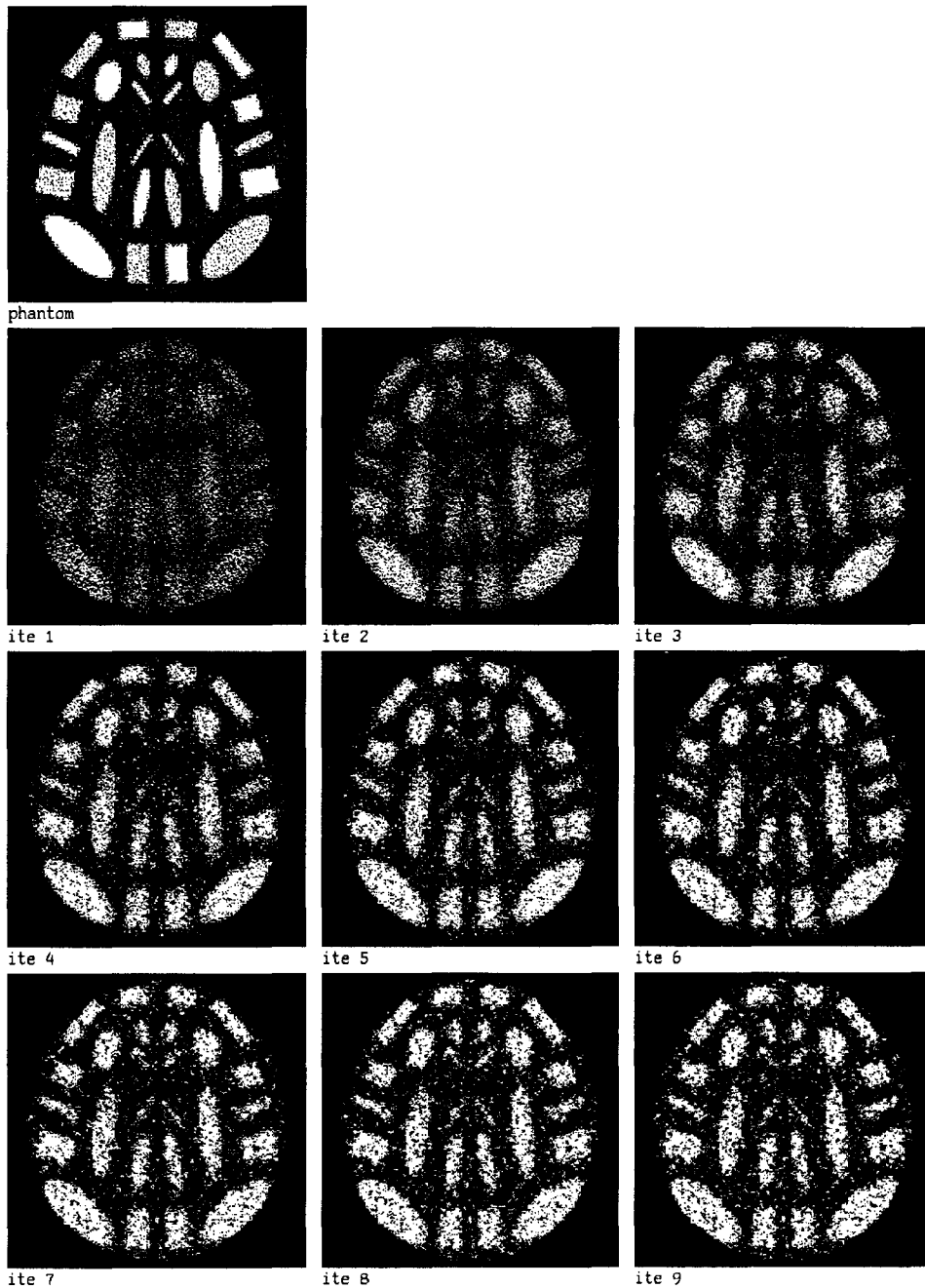


Figura 4.7: Um phantom de 95×95 pixels gerado aleatoriamente e suas reconstruções usando ART com $\omega = 0.025$

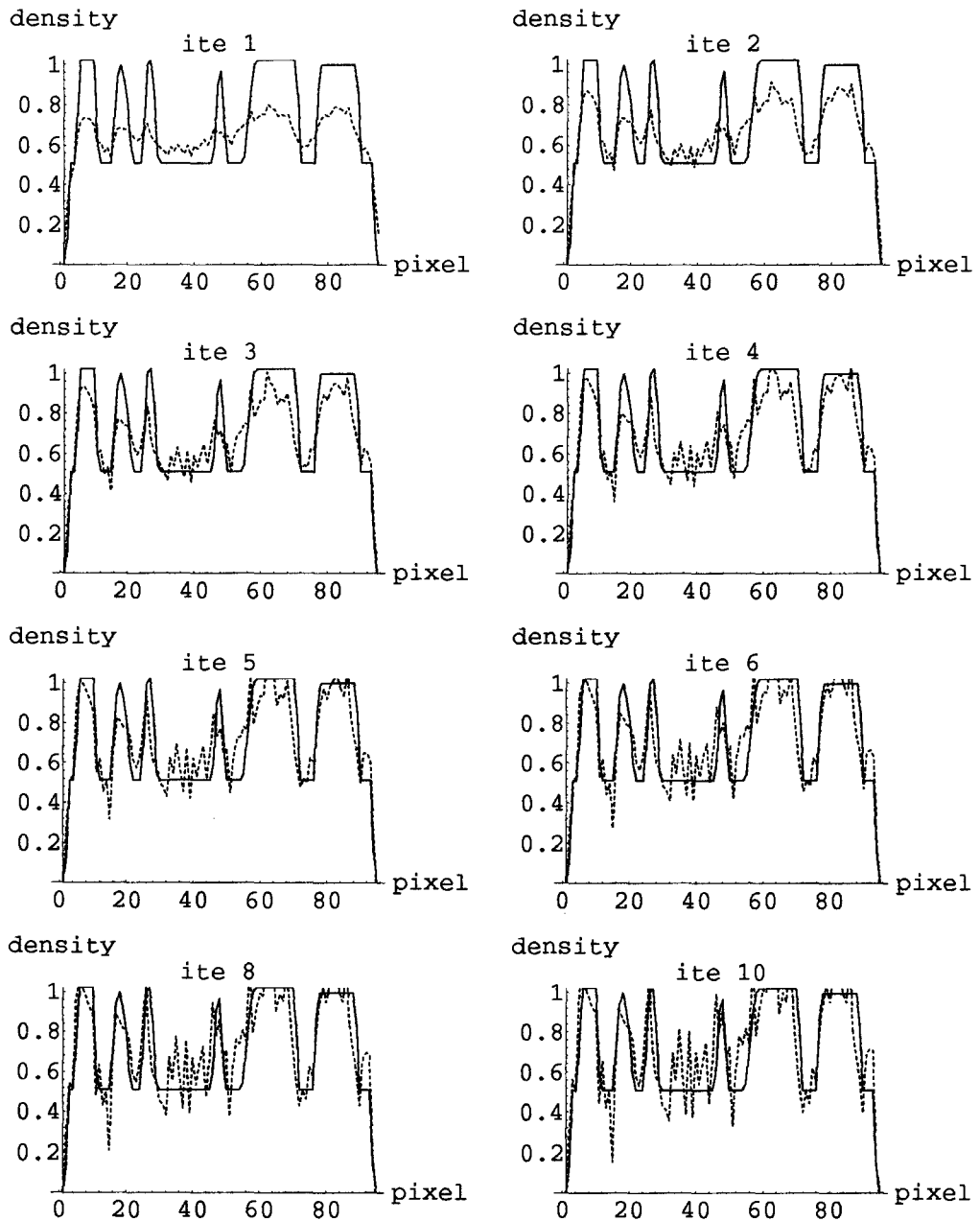


Figura 4.8: Comparação das densidades exatas com as das reconstruções de todos os pixels localizados na coluna 42, usando ART com $\omega = 0.025$

Fizemos experiências, também, usando um método iterativo não estacionário: Gradientes Conjugados aplicados às equações normais pré-condicionados com o método SSOR (PCCGNR), como explicaremos abaixo.

O método Gradientes Conjugados de Hestenes e Stiefel [42,41] (veja também [8,38]) aplicados às equações normais pode ser escrito como:

$$r^0 = b - Ax^0, s^0 = A^t r^0, w^0 = s^0$$

para $k = 0, 1, \dots$

$$p^k = Aw^k$$

$$\alpha^k = \frac{\|s^k\|^2}{\|p^k\|^2}$$

$$x^{k+1} = x^k + \alpha^k w^k$$

$$r^{k+1} = r^k - \alpha^k p^k$$

$$s^{k+1} = A^t r^{k+1}$$

$$\beta^k = \frac{\|s^{k+1}\|^2}{\|s^k\|^2}$$

$$w^{k+1} = s^{k+1} + \beta^k w^k$$

Agora, seja $AA^t = L + D + L^t$, onde L é a parte estritamente triangular inferior e D a diagonal de AA^t respectivamente. É bem conhecido (veja, por exemplo, [80,74,34]) que a matriz da iteração para o método SSOR (Sobre-Relaxação Sucessiva Simétrica) aplicado a $AA^t y = b$ é

$$Q_\omega = I - \omega(2 - \omega)(D + \omega L)^{-t} D (D + \omega L)^{-1} AA^t. \quad (4.20)$$

Assim, após dividirmos pela constante, que é um fator de escala, $\omega(2 - \omega)$ obtemos que o sistema de equações relacionado ao método SSOR aplicado a $AA^t y = b$ pode ser escrito como

$$C_\omega^{-t} C_\omega^{-1} AA^t y = C_\omega^{-t} C_\omega^{-1} b, \quad (4.21)$$

onde

$$C_\omega = (D + \omega L)D^{-\frac{1}{2}}. \quad (4.22)$$

Aplicamos, então, o método Gradientes Conjugados ao sistema

$$A^t C_\omega^{-t} C_\omega^{-1} A x = A^t C_\omega^{-t} C_\omega^{-1} b, \quad (4.23)$$

que pode, então ser escrito como:

$$r^0 = C_\omega^{-1}(b - Ax^0), \quad s^0 = A^t C_\omega^{-t} r^0, \quad w^0 = s^0$$

para $k = 0, 1, \dots$

$$p^k = C_\omega^{-1} A w^k$$

$$\alpha^k = \frac{\|s^k\|^2}{\|p^k\|^2}$$

$$x^{k+1} = x^k + \alpha^k w^k$$

$$r^{k+1} = r^k - \alpha^k p^k$$

$$s^{k+1} = A^t C_\omega^{-t} r^{k+1}$$

$$\beta^k = \frac{\|s^{k+1}\|^2}{\|s^k\|^2}$$

$$w^{k+1} = s^{k+1} + \beta^k w^k$$

Como os elementos de L não são explicitamente conhecidos, falta mostrar uma forma eficiente de computar os produtos $A^t C_\omega^{-t} r$ e $C_\omega^{-1} A w$. Em [10] foi mostrado isto, que vamos descrever abaixo.

Define-se $h_m = 0$. Para $i = m, m-1, \dots, 1$ computa-se o número

$$s_i = d_i^{-\frac{1}{2}} r_i - \omega d_i^{-1} a_i^t h_i \quad (4.24)$$

e o vetor

$$h_{i-1} = h_i + s_i a_i, \quad (4.25)$$

obtendo $h_0 = A^t C_\omega^{-t} r$, onde $D = \text{diag}(d_1, \dots, d_m)$.

De forma análoga, define-se $g_1 = w$. Para $i = 1, 2, \dots, m$ computa-se a componente

$$t_i = d_i^{-\frac{1}{2}} a_i^t g_i, \quad (4.26)$$

e o vetor

$$g_{i+1} = g_i - (\omega d_i^{-\frac{1}{2}} t_i) a_i, \quad (4.27)$$

obtendo $t = C_\omega^{-1} A w$, onde $D = \text{diag}(d_1, \dots, d_m)$.

Apesar deste método ser linear, x^k não é uma aplicação linear de b . O que implica que o operador de influência é não linear. Usamos o seguinte algoritmo para calcular a versão Monte-Carlo da extensão, que definimos no Capítulo anterior, do funcional de GCV para cada iteração k .

Algoritmo 4.2.

- (i) Gera-se um vetor pseudo-aleatório $w = (w_1, \dots, w_m)^t \in \mathbb{R}^m$ de uma distribuição normal com desvio padrão igual a 1, ou seja, $w \sim \mathcal{N}(0, I_{m \times m})$;
- (ii) Para cada k calculamos as iterações correspondentes a b , $b - \delta w$ e $b + \delta w$, onde $\delta = 10^{-4}$. Sejam x_1^k e x_2^k os iterados correspondentes a $b + \delta w$ e $b - \delta w$ respectivamente. Tomamos

$$\Phi(k) = \left(\frac{w^t [w - A(x_1^k - x_2^k)/2\delta]}{w^t w} \right)^2, \quad (4.28)$$

como uma aproximação de $(\frac{1}{m} \text{Tr}[I_m - DA(k)(b)])^2$;

- (iii) Para cada k tomamos

$$\frac{\frac{1}{m} \|b - Ax^k\|^2}{\Phi(k)} \quad (4.29)$$

como uma aproximação de $V(k)$.

Para a aplicação do método de PCCGNR, tomamos o vetor b com os erros da discretização, além dos erros que simulam aqueles, típicos de PET, que são obtidos de um gerador de números pseudo-aleatórios com uma distribuição de Poisson (veja [12]), sendo que o número total de fótons contados é, novamente, igual 2.022.085.

Na Fig. 4.9 estão os resultados de 5 estimativas para o denominador do funcional de GCV, com $\omega = 0.0$, que corresponde a aplicar CGNR ao sistema $Ax = b$ depois de escalar as linhas de A de forma que fiquem com norma igual a um. Na Fig. 4.10 estão os resultados de 5 estimativas para o denominador do funcional de GCV, com $\omega = 0.025$. Em ambos os casos temos uma pequena dispersão nas estimativas. A suavidade e não interseção das curvas favorece o não aparecimento de mínimos locais e uma precisão maior no valor do mínimo global do funcional correspondente, pois o erro que se comete na estimação do traço em cada iteração é mais ou menos uniforme.

Nas Figs. 4.11 e 4.12 estão plotados diferentes funções de perda para o método PCCGMR com $\omega = 0.0$ e $\omega = 0.025$ respectivamente.

Os gráficos mostram que não somente o mínimo de MCarlo GCV coincide com o de $\|Ax^k - Ax\|^2$, mas também as curvas são muito semelhantes.

Na Fig. 4.13 estão as imagens que correspondem a $\omega = 0.0$. Na Fig. 4.14 estão os cortes verticais com as comparações entre as reconstruções e o phantom em cada iteração, também para $\omega = 0.0$.

Finalmente, fizemos experiências usando GCV para escolher os parâmetros de relaxação ω_k no método ART (4.18). Para as primeiras três iterações plotamos nas Figs. 4.15, 4.16 e 4.17 diferentes funções de perda, variando apenas o parâmetro de relaxação ω , além de uma interpolação cúbica de valores do funcional de GCV. Na prática, pode-se a partir de uma tal interpolação determinar o valor de ω que minimiza o funcional de GCV. Com o valor de ω assim determinado para uma iteração, repete-se o mesmo procedimento para a iteração seguinte, até que não haja diminuição significativa da interpolação do funcional de GCV. Na Fig. 4.18 estão plotadas diferentes funções de perda para cada iteração, sendo que os valores dos parâmetros de relaxação ω_k foi escolhido em cada iteração usando o procedimento descrito acima. Nas Figs. 4.19 e 4.20 estão as imagens correspondentes e os cortes verticais com as comparações entre as reconstruções e o phantom em cada iteração, respectivamente.

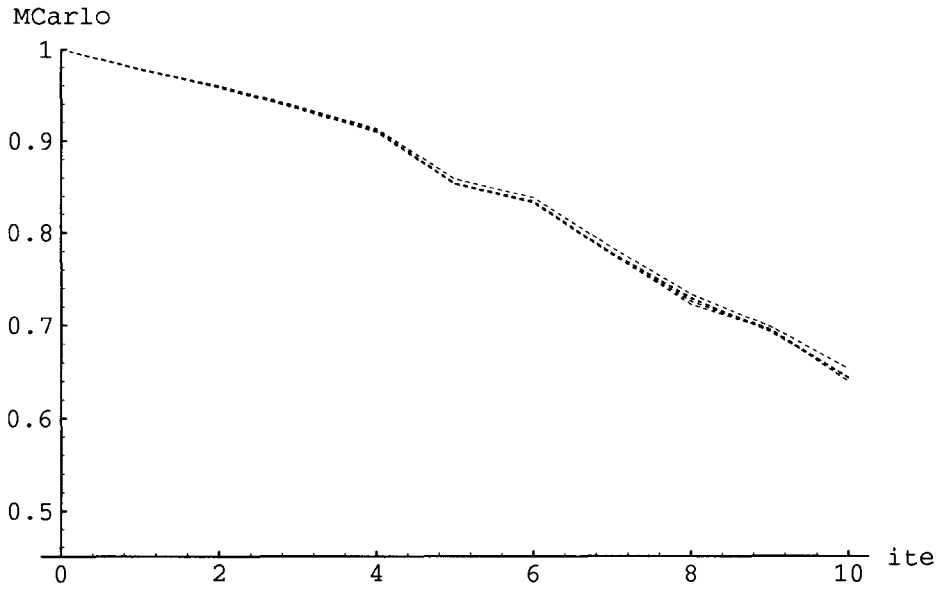


Figura 4.9: Cinco estimativas Monte-Carlo de $[\frac{1}{30292}Tr(I_{30292} - DA(k)(b))]^2$ para o método PCCGNR com $\omega = 0.0$

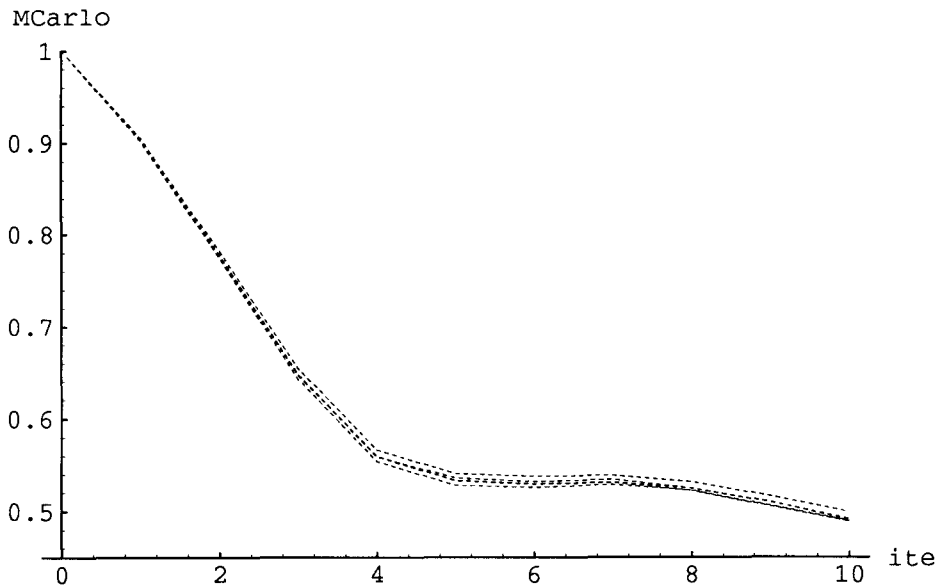


Figura 4.10: Cinco estimativas Monte-Carlo de $[\frac{1}{30292}Tr(I_{30292} - DA(k)(b))]^2$ para o método PCCGNR com $\omega = 0.025$

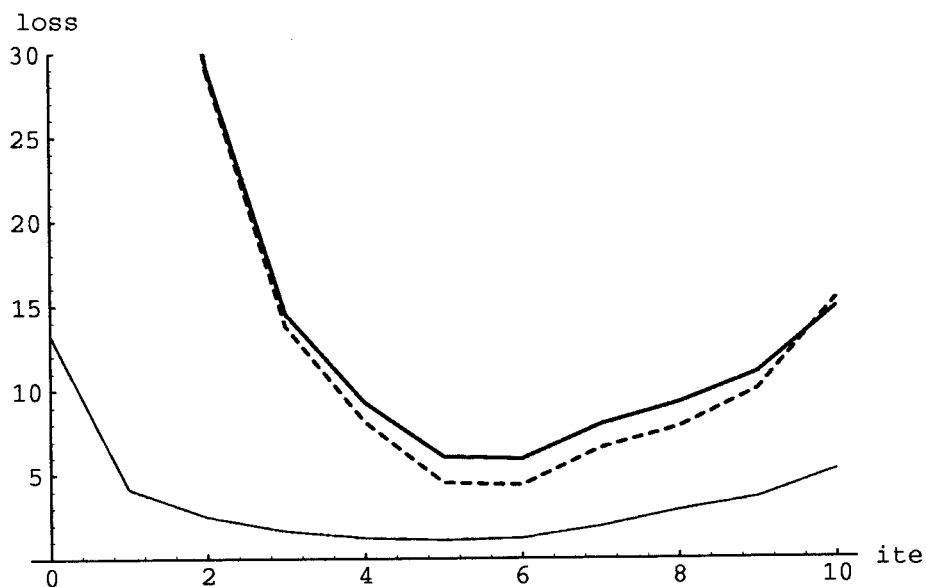


Figura 4.11: Funções de perda típicas para o método PCCGNR com $\omega = 0.0$. A linha fina corresponde a $100\frac{1}{95^2}\|x^k - x\|^2$; a grossa, a $\frac{1}{30292}\|Ax^k - Ax\|^2$; a tracejada, a MCarlo GCV - 70

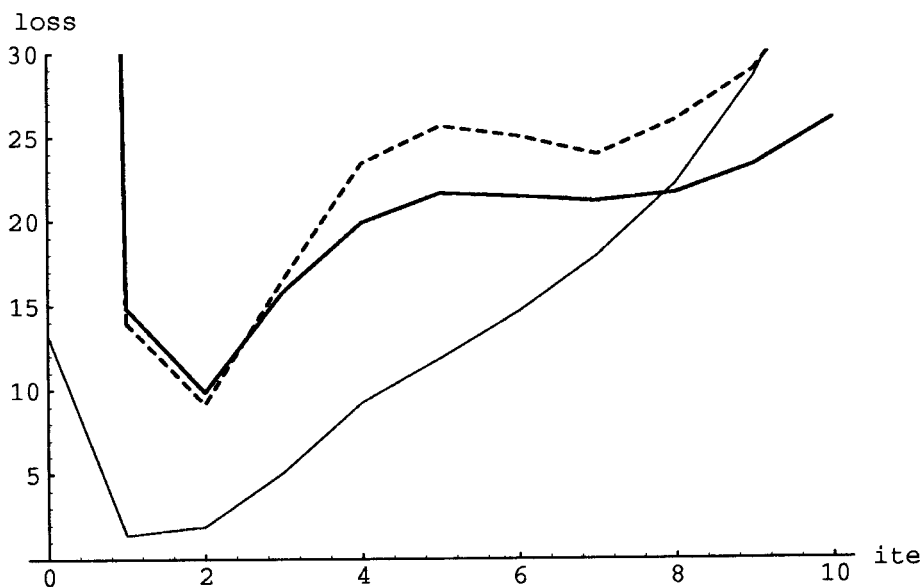


Figura 4.12: Funções de perda típicas para o método PCCGNR com $\omega = 0.025$. A linha fina corresponde a $100\frac{1}{95^2}\|x^k - x\|^2$; a grossa, a $\frac{1}{30292}\|Ax^k - Ax\|^2$; a tracejada, a MCarlo GCV - 70

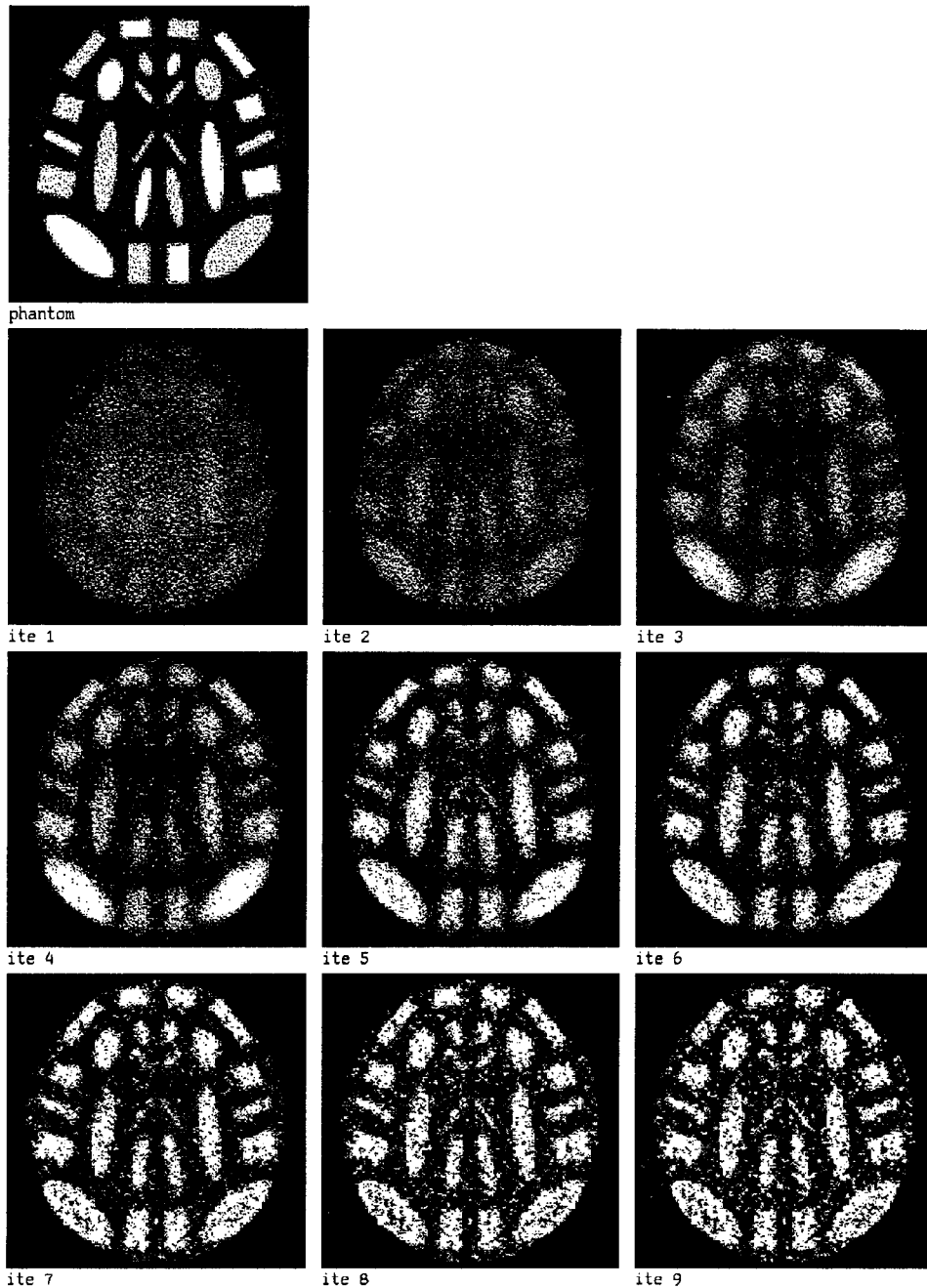


Figura 4.13: Um phantom de 95×95 pixels gerado aleatoriamente e suas reconstruções usando PCCGNR com $\omega = 0.0$

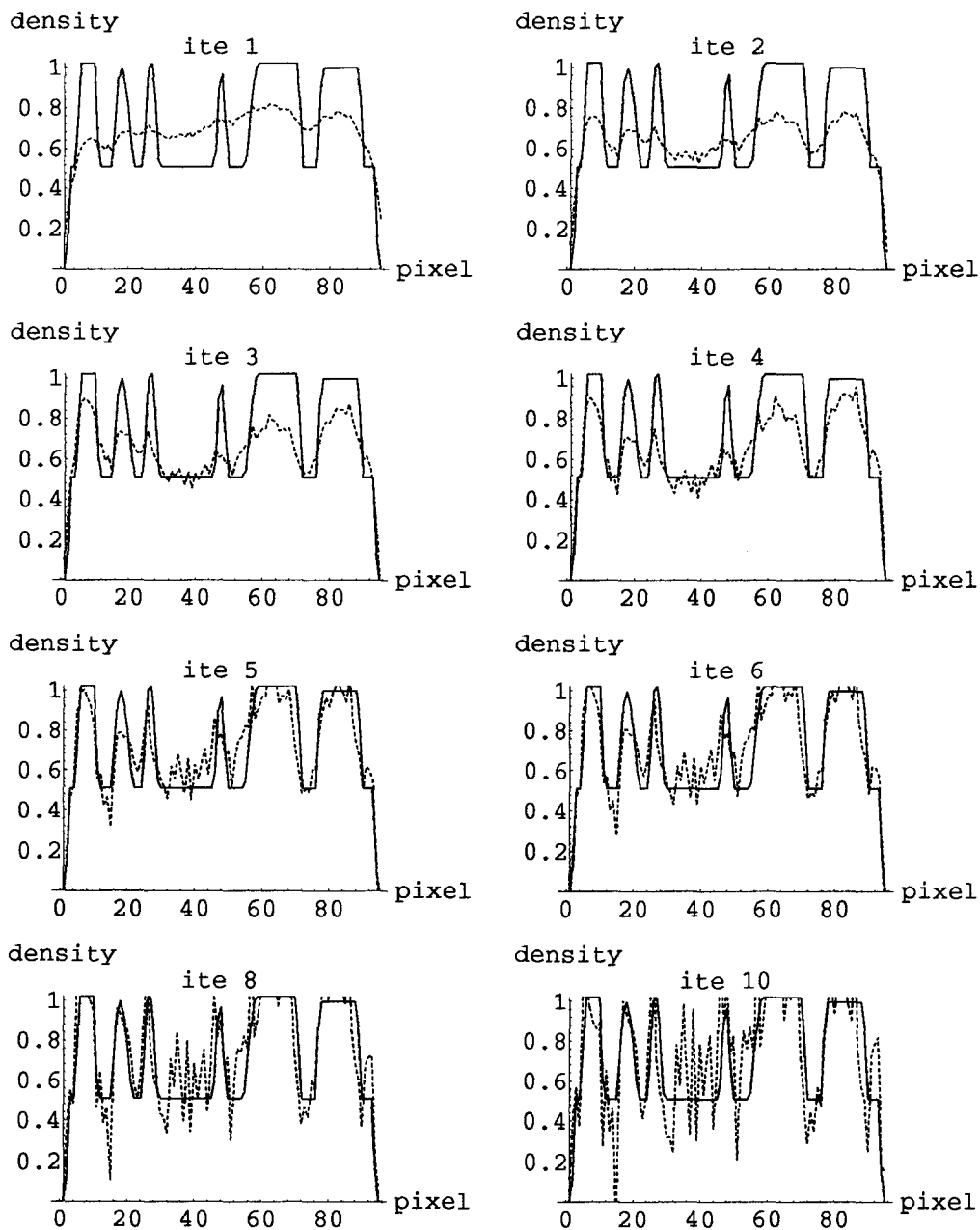


Figura 4.14: Comparação das densidades exatas com as das reconstruções de todos os pixels localizados na coluna 42, usando PCCGMR com $\omega = 0.0$

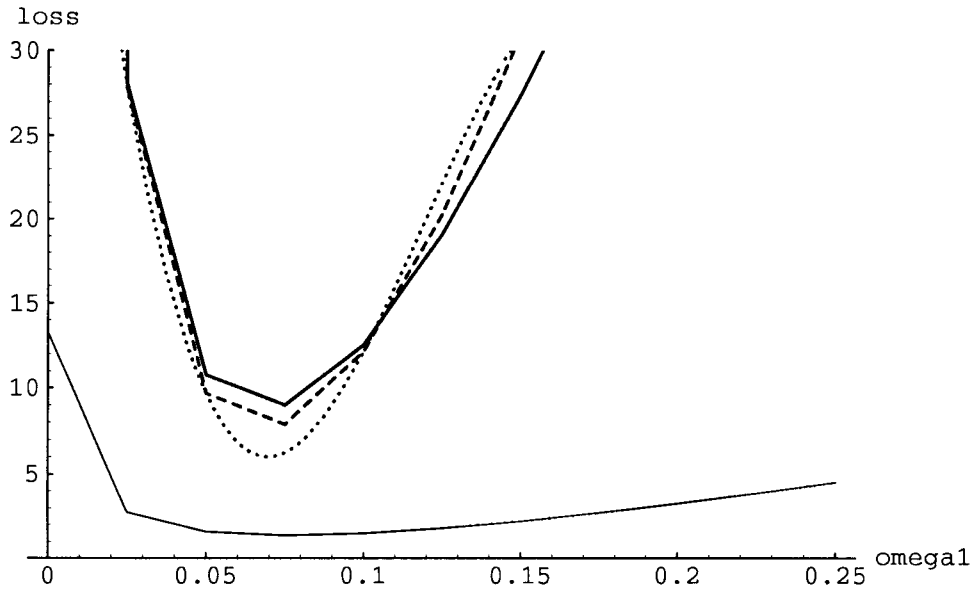


Figura 4.15: Funções de perda típicas para a primeira iteração do método ART variando o parâmetro de relaxação ω . A linha fina corresponde a $100\frac{1}{95^2}\|x^k - x\|^2$; a grossa, a $\frac{1}{30292}\|Ax^k - Ax\|^2$; a tracejada, a MCarlo GCV - 70; a pontilhada, a uma interpolação cúbica

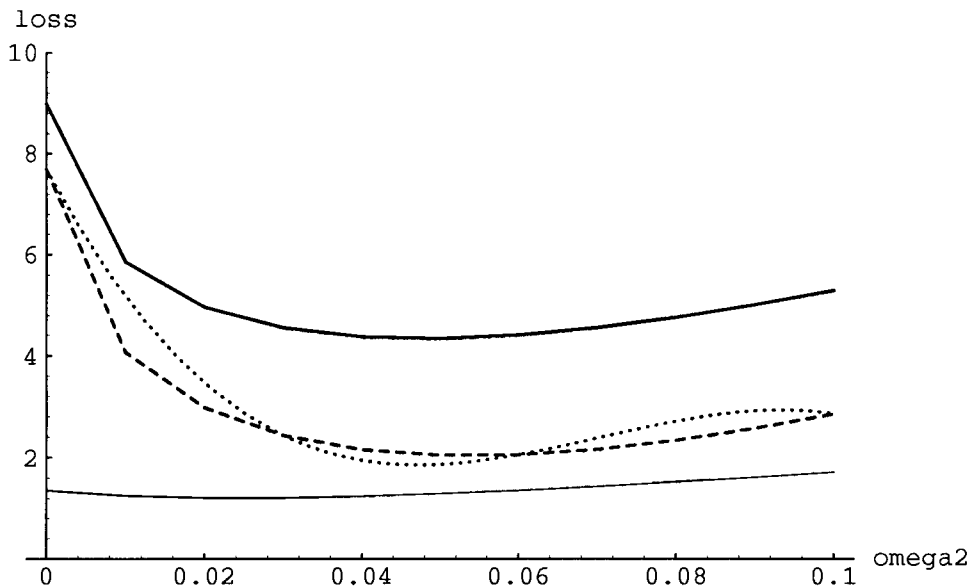


Figura 4.16: Funções de perda típicas para a segunda iteração do método ART variando o parâmetro de relaxação ω . A linha fina corresponde a $100\frac{1}{95^2}\|x^k - x\|^2$; a grossa, a $\frac{1}{30292}\|Ax^k - Ax\|^2$; a tracejada, a MCarlo GCV - 70; a pontilhada, a uma interpolação cúbica

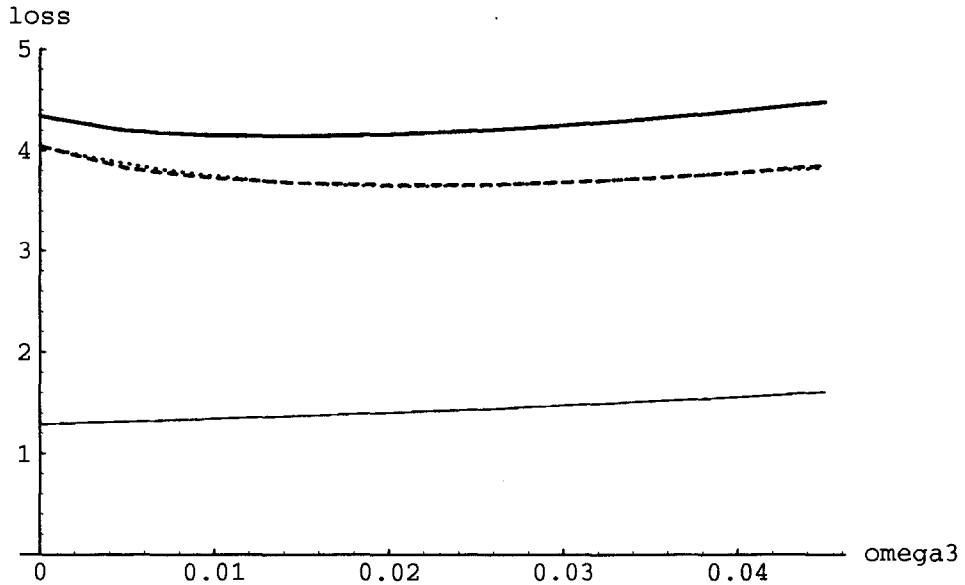


Figura 4.17: Funções de perda típicas para a terceira iteração do método ART variando o parâmetro de relaxação ω . A linha fina corresponde a $100\frac{1}{95^2}\|x^k - x\|^2$; a grossa, a $\frac{1}{30292}\|Ax^k - Ax\|^2$; a tracejada, a MCarlo GCV - 68; a pontilhada, a uma interpolação cúbica

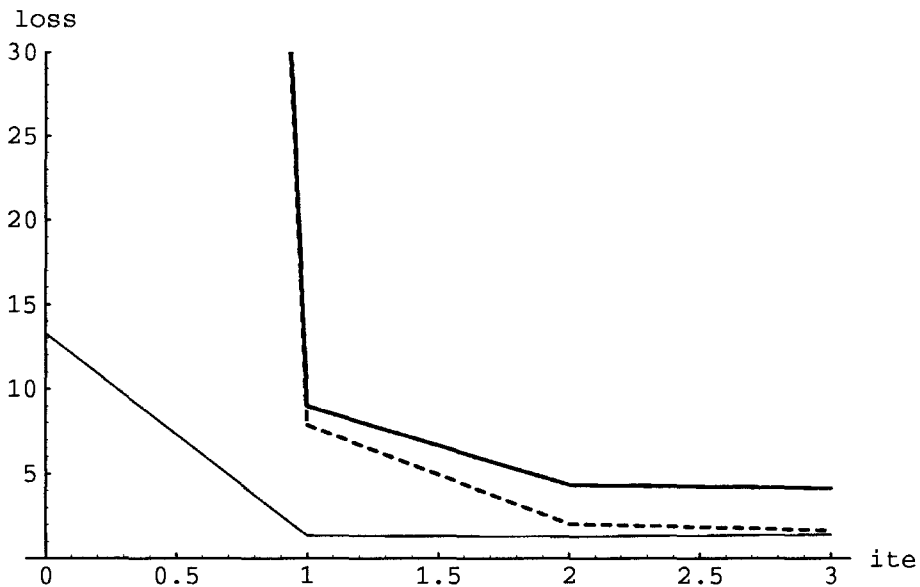


Figura 4.18: Funções de perda típicas para o método ART tomando o parâmetro de relaxação ótimo a cada iteração. A linha fina corresponde a $100\frac{1}{95^2}\|x^k - x\|^2$; a grossa, a $\frac{1}{30292}\|Ax^k - Ax\|^2$; a tracejada, a MCarlo GCV - 70

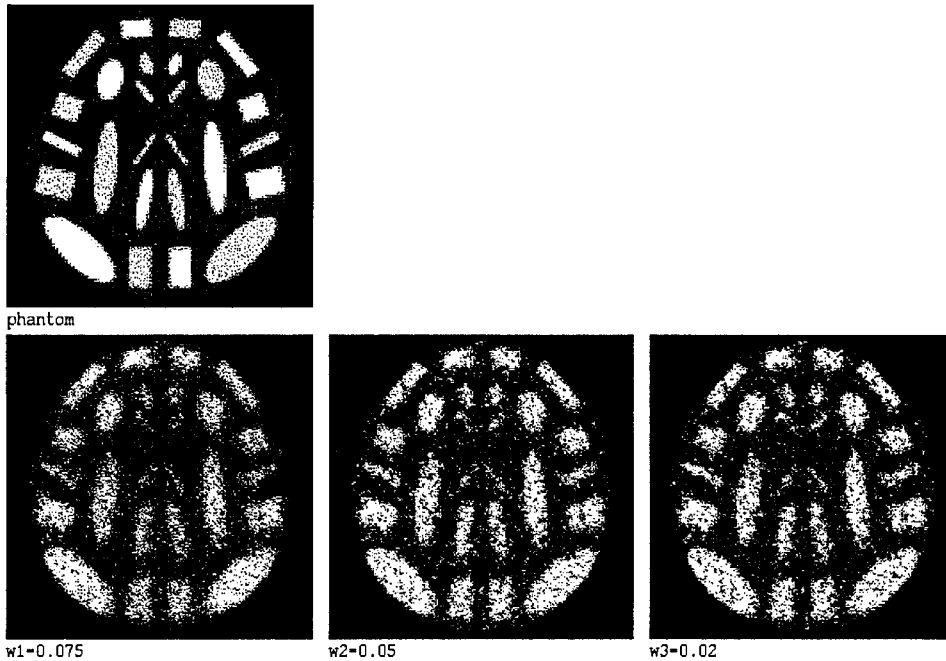


Figura 4.19: Um phantom de 95×95 pixels gerado aleatoriamente e suas reconstruções usando ART com ω ótimo em cada iteração

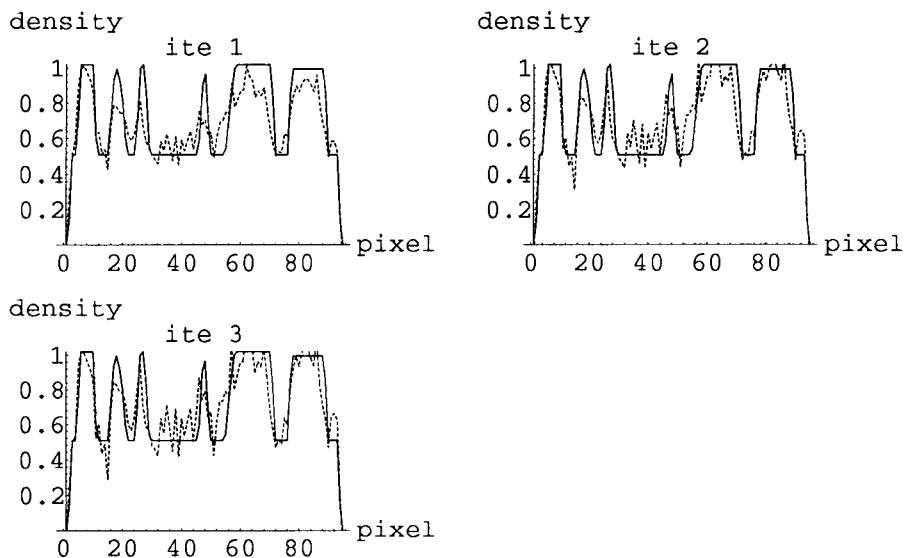


Figura 4.20: Comparação das densidades exatas com as das reconstruções de todos os pixels localizados na coluna 42, usando ART com ω ótimo em cada iteração

Bibliografia

- [1] D. M. Allen. The relationship between variable selection and data augmentation and a method for prediction. *Technometrics*, 16(1):125–127, 1974.
- [2] D. M. Bates, M. J. Lindstrom, G. Whaba, and B. S. Yandell. GCVPACK-routines for generalized cross validation. *Commun. Statist. Simul.*, 16:263–297, 1987.
- [3] J. Baumeister. *Stable Solution of Inverse Problems*. Friedr. Vieweg & Sohn, Braunschweig, 1987.
- [4] J. B. Bednar, L. R. Lines, R. H. Stolt, and A. B. Weglein. *Geophysical inversion*. SIAM, Philadelphia, 1992.
- [5] M. Bertero, P. Brianzi, and E. R. Pike. Super-resolution in confocal scanning microscopy. *Inverse Problems*, 3:195–212, 1987.
- [6] M. Bertero, C. De Mol, and E. R. Pike. Applied inverse problems in optics. In H. Engl and C. Groetsch, editors, *Inverse and Ill-Posed Problems*, Academic Press, New York, 1987.
- [7] A. Björck. Least squares methods. In P. G. Ciarlet and J. L. Lions, editors, *Handbook of Numerical Analysis, vol. I: Finite Difference Methods. Solutions of Equations in \mathbb{R}^n* , Elsevier North-Holland, Amsterdam, 1990.
- [8] A. Björck. Methods for sparse least squares problems. In J. R. Bunch and D. J. Rose, editors, *Sparse Matrix Computations*, Academic Press, New York, 1976.

- [9] A. Björck and L. Eldén. *Methods in Numerical Algebra for Ill-Posed Problems*. Technical Report LiTH-MAT-R-33-1979, Linköping Univ., Linköping, Sweden, 1979.
- [10] A. Björck and T. Elfving. Accelerated projection methods for computing pseudoinverse solutions of systems of linear equations. *BIT*, 19:145–163, 1979.
- [11] H. Brakhage. On ill-posed problems and the method of conjugate gradients. In H. Engl and C. Groetsch, editors, *Inverse and Ill-Posed Problems*, Academic Press, New York, 1987.
- [12] J. A. Browne, G. T. Herman, and D. Odhner. *SNARK93 a programming system for image reconstruction from projections*. Technical Report MIPG198, Department of Radiology, University of Pennsylvania, 1993.
- [13] S. L. Campbell and C. D. Meyer Jr. *Generalized Inverses of Linear Transformations*. Pitman, London San Francisco Melbourne, 1979.
- [14] Y. T. Chen. *Iterative methods for linear least squares problems*. Technical Report CS-75-04, University of Waterloo, Canada, 1975.
- [15] P. Craven and G. Wahba. Smoothing noisy data with spline functions. *Numer. Math.*, 31:377–403, 1979.
- [16] A. Dax. The convergence of linear stationary iterative processes for solving singular unstructured system of linear equations. *SIAM Review*, 32(4):611–635, 1990.
- [17] P. Deuffhard and E. Hairer, editors. *Numerical treatment of inverse problems in differential and integral equations*. Birkhäuser, Boston, 1983.
- [18] M. A. Diniz-Ehrhardt, J. M. Martínez, and S. A. Santos. Parallel projection methods and the resolution of ill-posed problems. *Computers Math. Applic.*, 27(1):11–24, 1994.

-
- [19] J. B. Drake. *ARIES: a computer program for the solution of first kind integral equations with noisy data*. Technical Report K/CSD/TM-43, Dept. of Computer Science, Oak Ridge National Laboratory, 1983.
- [20] L. Eldén. Algorithms for the regularization of ill-conditioned least squares problems. *BIT*, 17:134–145, 1977.
- [21] L. Eldén. A note on the computation of the generalized cross-validation function for ill-conditioned least squares problems. *BIT*, 24:467–472, 1984.
- [22] L. Eldén. A weighted pseudoinverse, generalized singular values, and constrained least squares problems. *BIT*, 22:487–502, 1982.
- [23] H. W. Engl. Regularization methods for the stable solution of inverse problems. *Surv. Math. Ind.*, 3:71–143, 1993.
- [24] H. W. Engl, K. Kunisch, and A. Neubauer. Convergence rates for tikhonov regularization of non-linear ill-posed problems. *Inverse Problems*, 5:523–540, 1989.
- [25] H. W. Engl and A. Neubauer. Convergence rates for Tikhonov regularization in finite-dimensional subspaces of Hilbert scales. *Proc. Amer. Math. Soc.*, 102:587–592, 1988.
- [26] H. E. Fleming. Equivalence of regularization and truncated iteration in the solution of ill-posed image reconstruction problems. *Linear Algebra and its Appl.*, 130:133–150, 1990.
- [27] G. E. Forsythe. Solving linear algebraic equations can be interesting. *Bull. Amer. Math. Soc.*, 59:299–329, 1953.
- [28] D. A. Girard. A fast ‘Monte-Carlo cross-validation’ procedure for large least squares problems with noisy data. *Numer. Math.*, 56:1–23, 1989.
- [29] D. A. Girard. Optimal regularized reconstruction in computerized tomography. *SIAM J. Sci. Stat. Comput.*, 8(6):934–950, 1987.

- [30] G. H. Golub, M. T. Heath, and G. Wahba. Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21:215–223, 1979.
- [31] G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins U.P., Baltimore, 2nd edition, 1989.
- [32] C. W. Groetsch. *The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind*. Pitman, Boston, 1984.
- [33] J. Hadamard. *Lectures on Cauchy's Problem in Linear Partial Differential Equations*. Yale University Press, New Haven, 1923.
- [34] L. A. Hageman and D. M. Young. *Applied Iterative Methods*. Academic Press, New York, 1981.
- [35] M. Hanke and P. C. Hansen. Regularization methods for large-scale problems. *Surv. Math. Ind.*, 3:253–315, 1993.
- [36] P. C. Hansen. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Review*, 34:561–580, 1992.
- [37] P. C. Hansen. Regularization tools, a MATLAB package for analysis and analysis and solution of discrete ill-posed problems. *Numer. Algorithms*, 6:1–35, 1994.
- [38] M. T. Heath. Numerical methods for large sparse linear least squares problems. *SIAM J. Sci. Stat. Comput.*, 5:497–513, 1984.
- [39] G. T. Herman. *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography*. Academic Press, New York, 1980.
- [40] G. T. Herman and L. B. Meyer. *Algebraic reconstruction techniques can be made computationally efficient*. Technical Report MIPG189, Department of Radiology, University of Pennsylvania, Philadelphia, 1992.
- [41] M. R. Hestenes. Pseudoinverses and conjugate gradients. *Comm. of the ACM*, 18(1):40–43, 1975.

- [42] M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *J. Res. Nat. Bur. Stds.*, 49(6):409–436, 1952.
- [43] S. Kaczmarz. Angenährte Auflösung von Systemen linearer Gleichungen. *Bull. Acad. Polon. Sci. Lett. A*, 35:355–357, 1937.
- [44] A. C. Kak and M. Slaney. *Principles of Computerized Tomographic Imaging*. IEEE Press, New York, 1988.
- [45] L. Kaufmann. Implementing and accelerating the EM algorithm for positron emission tomography. *IEEE Trans. Med. Imag.*, 6:37–51, 1987.
- [46] J. T. King and A. Neubauer. A variant of finite-dimensional Tikhonov regularization with a-posteriori parameter choice. *Computing*, 40:91–109, 1988.
- [47] G. Kristensson and C. R. Vogel. Inverse, scattering for acoustic waves using the penalized likelihood method. *Inverse Problems*, 2:461–479, 1986.
- [48] L. Landweber. An iteration formula for Fredholm integral equations of the first kind. *Amer. J. Math.*, 73:615–624, 1951.
- [49] R. Lewitt. Alternatives to voxels for image representation in iterative reconstruction algorithms. *Phys. Med. Biol.*, 37:705–716, 1992.
- [50] A. K. Louis. *Inverse und schlecht gestellte Probleme*. B. G. Teubner, Stuttgart, 1989.
- [51] A. K. Louis and F. Natterer. Mathematical problems of computerized tomography. *Proc. IEEE*, 71(3):379–389, 1983.
- [52] V. A. Morozov. *Methods for solving incorrectly posed problems*. Springer, New York Berlin Heidelberg, 1984.
- [53] V. A. Morozov. On the solution of functional equations by the method of regularization. *Soviet Math. Dokl.*, 7:52–74, 1970.

-
- [54] F. Natterer. *The Mathematics of Computerized Tomography*. J. Wiley & Sons and G. Teubner, Stuttgart, 1986.
- [55] F. Natterer. Numerical treatment of ill-posed problems. In G. Talenti, editor, *Inverse Problems*, Springer, New York Berlin Heidelberg Tokyo, 1987.
- [56] A. Neubauer. Tikhonov regularization for non-linear ill-posed problems: optimal convergence rates and finite-dimensional approximation. *Inverse Problems*, 5:541–557, 1989.
- [57] A. Neubauer and O. Scherzer. Finite-dimensional approximation of tikhonov regularized solutions of non-linear ill-posed problems. *Numer. Funct. Anal. Optim.*, 11:85–99, 1990.
- [58] Y. Osaki and H. Shibahashi, editors. *Progress of seismology of the sun and stars*. Springer, New York Berlin Heidelberg Tokyo, 1990.
- [59] F. O’Sullivan. Sensitivity analysis for regularized estimation in some system identification problems. *SIAM J. Sci. Stat. Comput.*, 12(6):1266–1283, 1991.
- [60] F. O’Sullivan. A statistical perspective on ill-posed inverse problems. *Statist. Sci.*, 1:502–527, 1986.
- [61] F. O’Sullivan and G. Wahba. A cross validated bayesian retrieval algorithm for nonlinear remote sensing experiments. *J. Comp. Phys.*, 59:441–455, 1985.
- [62] S. D. Silvey. *Statistical Inference*. Penguin, Harmondsworth, 1970.
- [63] O. N. Strand. Theory and methods related to the singular-function expansion and Landweber’s iteration for integral equations of the first kind. *SIAM J. Numer. Anal.*, 11:798–825, 1974.
- [64] J. J. te Riele. A program for solving first kind Fredholm integral equations by means of regularization. *Comput. Phys. Comm.*, 36:423–432, 1985.
- [65] M. M. Ter-Pogossian et al. Positron emission tomography. *Scientific American*, October:170–181, 1980.

- [66] A. N. Tikhonov and V. Y. Arsenin. *Solutions of Ill-Posed Problems*. Wiley, New York, 1977.
- [67] A. N. Tikhonov and A. V. Goncharsky. *Ill-posed Problems in the Natural Sciences*. MIR, Moscow, 1987.
- [68] S. Twomey. *Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurements*. Elsevier, New York, 1977.
- [69] A. van der Sluis and H. A. van der Vorst. Numerical solution of large, sparse linear algebraic systems arising from tomographic problems. In G. Nolet, editor, *Seismic Tomography*, D. Reidel Pub. Comp., Dordrecht, The Netherlands, 1987.
- [70] A. van der Sluis and H. A. van der Vorst. SIRT and CG type methods for the iterative solution of sparse linear least squares problems. *Linear Algebra and its Appl.*, 130:257–303, 1990.
- [71] M. van Dijke. *Iterative Methods in Image Reconstruction*. PhD thesis, Rijksuniversiteit Utrecht, Utrecht, The Netherlands, 1992.
- [72] J. M. Varah. A practical examination of some numerical methods for linear discrete ill-posed problems. *SIAM Review*, 21:100–111, 1979.
- [73] Y. Vardi, L. A. Shepp, and L. Kaufman. A statistical model for positron emission tomography. *J. Amer. Stat. Assoc.*, 80(389):8–37, 1985.
- [74] R. S. Varga. *Matrix Iterative Analysis*. Prentice-Hall, New Jersey, 1962.
- [75] C. R. Vogel. A constrained least squares regularization method for nonlinear ill-posed problems. *SIAM J. Control and Optim.*, 28(1):34–49, 1990.
- [76] G. Wahba. Practical approximate solutions to linear operator equations when the data are noisy. *SIAM J. Numer. Anal.*, 14:651–667, 1977.
- [77] G. Wahba. *Spline Models for Observational Data*. SIAM, Philadelphia, 1991.

-
- [78] G. Wahba. Three topics in ill-posed problems. In H. Engl and C. Groetsch, editors, *Inverse and Ill-Posed Problems*, Academic Press, New York, 1987.
- [79] G. Wahba and Y. Wang. When is the the optimal regularization parameter insensitiveto the choice of the loss function? *Commun. Statist. - Theory Meth.*, 19(5):1685–1700, 1990.
- [80] D. M. Young. *Iterative Solutions of Large Linear Systems*. Academic Press, New York, 1971.
- [81] D. M. Young. On the consistency of linear stationary iterative methods. *SIAM J. Numer. Anal.*, 9(1):89–96, 1972.