

Teoria e Experimentação com Métodos
Incrementais Relaxados em Tomografia por
Emissão

Elias Salomão Helou Neto

09 de maio de 2005

**FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO IMECC DA UNICAMP**

Bibliotecário: Maria Júlia Milani Rodrigues – CRB8a / 2116

Helou Neto, Elias Salomão

H369t Teoria e experimentação com métodos incrementais relaxados em tomografia por emissão / Elias Salomão Helou Neto -- Campinas, [S.P. :s.n.], 2005.

Orientador : Álvaro Rodolfo De Pierro

Dissertação (mestrado) - Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Tomografia. 2. Tomografia – Emissão. 3. Tomografia computadorizada. I. De Pierro, Álvaro Rodolfo. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.

Título em inglês: Theory and experimentation with relaxed incremental methods in emission tomography

Palavras-chave em inglês (Keywords): 1. Tomography. 2. Tomography – Emission. 3. Computerized tomography.

Área de concentração: Tomografia computadorizada

Titulação: Mestre em Matemática Aplicada

Banca examinadora: Prof. Dr. Álvaro Rodolfo De Pierro (UNICAMP)
Prof. Dra. Joyce da Silva Bevilacqua (USP)
Prof. Dr. Roberto Andreani (UNICAMP)

Data da defesa: 12/04/2005

Teoria e Experimentação com Métodos Incrementais Relaxados em Tomografia por Emissão

Este exemplar corresponde à redação final da dissertação devidamente corrigida e defendida por Elias Salomão Helou Neto e aprovada pela comissão julgadora.

Campinas, 9 de maio de 2005



Prof. Dr. Álvaro Rodolfo De Pierro
Orientador

Banca Examinadora:

Álvaro Rodolfo De Pierro
Joyce da Silva Bevilacqua
Roberto Andreani

Dissertação apresentada ao Instituto de Matemática, Estatística e Computação Científica, UNICAMP como requisito para obtenção do título de Mestre em Matemática Aplicada.

Dissertação de Mestrado defendida em 12 de Abril de 2005 e aprovada pela Banca Examinadora composta pelos Profs. Drs.



Prof (a). Dr (a). ALVARO RODOLFO DE PIERRO



Prof (a). Dr (a). JOYCE DA SILVA BEVILACQUA



Prof (a). Dr (a). ROBERTO ANDREANI

Dedico esta dissertação a meus pais e à Aline, que esteve ao meu lado desde o princípio até o fim do trabalho, sempre me apoiando.

Sumário

Resumo	8
Abstract	9
Introdução	10
Tomografia	10
Tomografia por Transmissão	10
Tomografia por Emissão	12
Modelagem Estatística em Tomografia por Emissão de Pósitrons . . .	14
Algoritmos para o Modelo Estatístico	16
Objetivos	17
1 Tomografia por Emissão de Pósitrons	19
1.1 Emissão e Desintegração de Pósitrons	19
1.2 Atenuação	20
1.3 Espalhamento	21
1.4 Outros Efeitos	22
2 Métodos Analíticos	24
2.1 Teorema da Projeção	24
2.2 Retroprojeção Filtrada	26
2.3 Outros Filtros	29
3 Reconstrução Estatística de Imagens em Tomografia por Emissão de Pósitrons	31
3.1 Expectation Maximization Clássico	32

	7
3.2	Gradientes Incrementais 35
3.3	Ordered Subsets - Expectation Maximization 36
3.4	Row Action Maximum Likelihood Algorithm 39
3.5	Dynamic RAMLA 40
4	Simulações Numéricas 44
4.1	Simulando a Leitura de Tomografia por Emissão de Pósitrons . . 44
4.2	Avaliando as Reconstruções 45
4.3	Composição e Ordenação dos Subconjuntos 45
4.4	Parâmetros de Relaxação 52
4.5	Interpretação dos Resultados 53
5	Demonstrações 61
5.1	Regularidade 61
5.2	Convergência 66
5.3	Conclusões e Trabalho Futuro 71
	Referências Bibliográficas 73
	Lista de Figuras 75
A	Implementação 77
A.1	Projeção Ay 77
A.2	Retroprojeção Filtrada 78
A.3	Ordered Subsets - Expectation Maximization 78
A.4	Row Action Maximum Likelihood Algorithm 80

Resumo

O principal assunto desta dissertação é a teoria e prática em métodos iterativos incrementais relaxados para reconstrução estatística de imagens em Tomografia por Emissão de Póstrons — PET. Especial atenção é dedicada ao estudo teórico dos algoritmos, com a apresentação de resultados novos sobre convergência com o parâmetro de relaxação variando dentre as subiterações e sobre a regularidade da sequência gerada por tais algoritmos no caso específico da função de log-verossimilhança que surge em PET. Além disso, uma comparação experimental entre a qualidade da imagem obtida com métodos relaxados e não relaxados é efetuada.

Abstract

The main subject of this dissertation is the theory and practice in iterative incremental relaxed methods for statistical image reconstruction in Positron Emission Tomography — PET. Special attention is drawn to the theoretical study of the algorithms, with the presentation of new results concerning convergence with a relaxation parameter which is allowed to vary within the subiterations and about the regularity of the sequence generated by such algorithms in the specific case of the log-likelihood function that arises in PET. Furthermore, an experimental comparison is made to test the quality of the image obtained with relaxed methods against those generated by non-relaxed ones.

Introdução

Tomografia

O objetivo da tomografia é obter a imagem de uma secção transversal, ou talvez uma imagem tridimensional, do interior de um objeto de estudo de maneira não invasiva. Como os objetos de estudo podem diferir muito em sua natureza, as características que se desejam revelar nestas imagens variam grandemente, assim como as técnicas usadas para obtê-las.

Desnecessário dizer, um vasto campo de aplicação de técnicas de tomografia é a medicina. Neste contexto, as imagens devem revelar características úteis para o diagnóstico ou avaliação do estado clínico do paciente (o objeto de estudo). Detalhes anatômicos podem ser fundamentais em certos casos, enquanto que em outros o metabolismo e/ou o equilíbrio bioquímico podem ser mais importantes.

Tomografia por Transmissão

Historicamente, a primeira técnica tomográfica utilizada para obter imagens médicas foi a baseada em raios X. Como a atenuação desta radiação varia conforme a densidade do meio percorrido, a obtenção de dados sobre a anatomia interna é possível posicionando uma fonte de raios X de um lado do paciente e um aparelho receptor do outro lado. Variando o ângulo deste aparato, podemos medir a atenuação da radiação como se fossem projeções (integrais de linha ao longo de retas) e, desta forma, utilizar versões discretas da fórmula de inversão da transformada de Radon para obter as imagens desejadas com um esforço computacional reduzido.

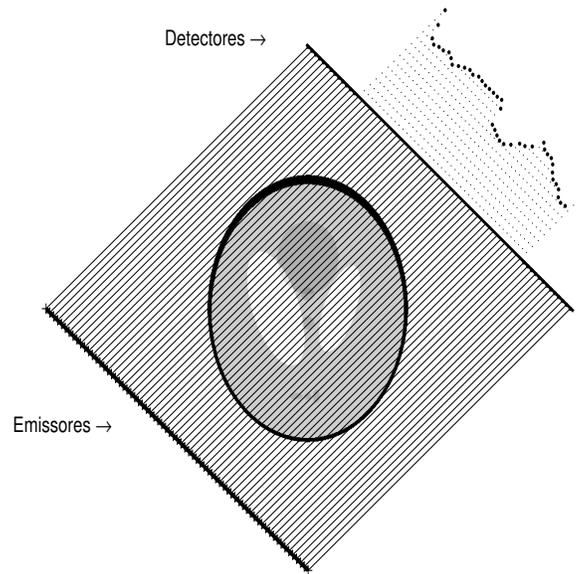


Figura 1: Tomada de dados em tomografia por transmissão.

A Figura 1 exibe uma representação esquemática da aquisição de dados em tomografia por raios X. Pode ser visto o conjunto de emissores e detectores durante um momento da leitura. Também é mostrado um gráfico das atenuações medidas, que serão utilizadas como projeções na reconstrução da imagem. Apesar da representação ser bidimensional, a extensão para três dimensões é facilmente obtida “empilhando” vários detectores e emissores como os mostrados.

Pode-se também iluminar o objeto com um emissor pontual que emita radiação em formato de leque, este tipo de geometria é conhecido como *feixe divergente*, em contraposição ao *feixe paralelo* exibido na figura.

A mesma técnica de coleta de dados pode ser aplicada para qualquer sinal transmitido através do objeto de estudo. Quando utilizamos estes dados para obtenção de imagens tomográficas temos a *tomografia por transmissão*.

Tomografia por Emissão

Se o interesse for por um processo fisiológico, podemos utilizar elementos químicos radioativos chamados de *marcadores* para rastrear metabólitos específicos¹ que conhecidamente façam parte do processo particular em estudo. Assim, se for possível medir de forma apropriada, no exterior do corpo do paciente, a radiação emitida, seremos capazes de obter uma imagem da intensidade de determinado processo fisiológico ou da concentração de uma substância específica no interior do organismo do paciente. Este é o princípio da *tomografia por emissão*.

Dentre estas técnicas, destacam-se PET (Positron Emission Tomography — Tomografia por Emissão de Pósitrons) e SPECT (Single Photon Emission Tomography — Tomografia por Emissão de Fóton Único). No primeiro caso, o isótopo utilizado como marcador decai emitindo um pósitron. Tal partícula percorre uma curta trajetória, até que se aniquila ao se chocar com um elétron emitindo dois fótons em trajetórias opostas. Estes fótons emitidos são coletados em um cilindro formado por detectores em torno do paciente e as detecções simultâneas são contadas separadamente para cada par de detectores. Isto fornece dados sobre as intensidades de emissão ao longo de várias direções (aquelas que unem os pares de detectores), de forma que a obtenção de uma imagem é possível a partir destas contagens. Em SPECT, a substância radioativa emite um único fóton. A detecção pura e simples deste fóton não é suficiente e o uso de um colimador é necessário para que os dados possam ser usados para tomografia. Nesta dissertação, serão abordados algoritmos para a obtenção de imagens em tomografia por emissão. Estes métodos podem ser utilizados de forma semelhante para PET ou SPECT, apesar de que apenas o caso de PET é discutido explicitamente.

Na Figura 2, pode ser visto um desenho esquemático simplificado de uma leitura PET de curta duração. Os \times indicam os locais onde pósitrons foram emitidos e as linhas pontilhadas representam as trajetórias dos fótons emitidos

¹Como um exemplo pode-se utilizar ^{18}F para obter FDG (Flúor-DeoxiGlicose) que é bioquimicamente semelhante à glicose e toma parte em diversos processos metabólicos.

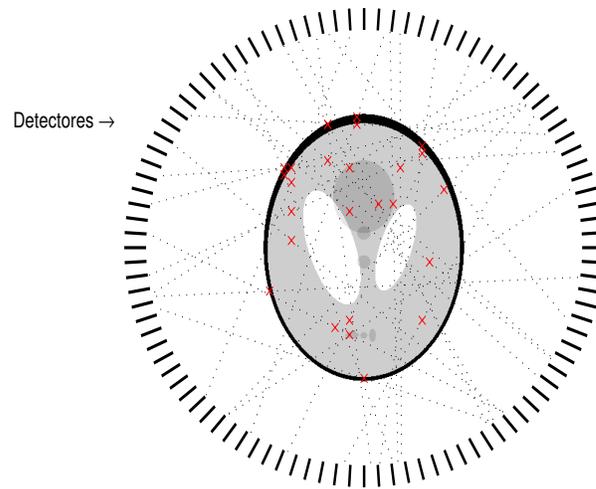


Figura 2: Exemplo de leitura PET.

após as interações dos pósitrons com elétrons vizinhos. Apesar de não ser mostrado, cada pósitron percorre um curto espaço antes de se aniquilar com um elétron e causar a emissão dos fótons.

Cada detector é representado, na figura, pelo espaço entre dois traços adjacentes no anel em torno do objeto de estudo. Estes detectores são capazes de acusar o contato do fóton com sua superfície. Se dois fótons foram emitidos simultaneamente, suas detecções devem ocorrer dentro de um intervalo muito pequeno, o que permite emparelhar as detecções causadas por uma aniquilação pósitron-elétron. Os dados resultantes da leitura são os números de detecções (quase) simultâneas para cada par de detectores.

Em PET, não temos controle sobre a emissão, que ocorre dentro do objeto de estudo. Entretanto, a observação mostra que a radiação é emitida uniforme-

mente em todas as direções e que a taxa de emissão de pósitrons é proporcional à quantidade de marcador. Isto implica em que as contagens de fótons podem funcionar como projeções da imagem desejada² sobre as linhas que ligam os detectores de cada par. Desta maneira, pode-se utilizar inversão discreta da transformada de Radon para obter imagens em PET.

Por outro lado, a emissão de radiação é, na realidade, um fenômeno aleatório. O comportamento descrito no parágrafo anterior deve valer quando um alto número de coincidências é contado, o que nem sempre é verdade. Além da natureza probabilística da emissão de pósitrons, há outras fontes de ruído em PET que podem ser modeladas de forma simples como processos estatísticos, mas apresentam dificuldades quando procura-se incorporá-las à transformada de Radon. Os melhores exemplos são a distância percorrida pelo pósitron antes de interagir com um elétron e o fato de que os fótons não são emitidos em direções exatamente opostas após a aniquilação pósitron-elétron. Além destes exemplos, alguns tomógrafos medem o tempo entre as duas detecções em cada par de detectores e esta informação pode ser incorporada à reconstrução da imagem facilmente no modelo estatístico (veja [VSK85] para detalhes).

Modelagem Estatística em Tomografia por Emissão de Pósitrons

O problema da baixa relação sinal/ruído exige uma abordagem que leve em conta a natureza aleatória dos fenômenos envolvidos numa leitura PET. Assim, considerando que a emissão de pósitrons é um fenômeno de Poisson, se a imagem (tridimensional e dinâmica) $f : \mathbb{R}^3 \times \mathbb{R}_+ \mapsto \mathbb{R}_+$ desejada é proporcional à distribuição espacial da intensidade de emissão, então o número de pósitrons emitidos durante a leitura na região $R \in \mathbb{R}^3$ é uma variável de Poisson de parâmetro:

$$\int_0^T \int_R f(\mathbf{x}, t) d\mathbf{x} dt \quad (1)$$

²No caso, a imagem almejada é um mapa da concentração de marcador na secção transversal de interesse.

Apesar do t3pico de imagens dinâmicas ser bastante pesquisado, não é o intuito desta dissertação se aprofundar no assunto e, daqui por diante, as imagens serão consideradas estáticas, portanto constantes no tempo. Desta forma, definindo $f(\mathbf{x}) := Tf(\mathbf{x}, t)$, a integral acima se torna:

$$\int_R f(\mathbf{x})d\mathbf{x}. \quad (2)$$

Seja $a_i(\mathbf{x})$ a probabilidade de que um pósitron emitido no ponto \mathbf{x} gere fótons que serão detectados simultaneamente no par i de detectores e $V_i := \{\mathbf{x} \in \mathbb{R}^3 | a_i(\mathbf{x}) > 0\}$. O número de detecções simultâneas no par i de detectores é, com esta notação, uma V.A. Poisson de parâmetro:

$$\int_{V_i} a_i(\mathbf{x})f(\mathbf{x})d\mathbf{x}. \quad (3)$$

Como os dados são finitos, apenas um número finito de parâmetros pode ser estimado. Este argumento, aliado a limitações computacionais, exige uma discretização da imagem. Uma forma geral de proceder é escolher um conjunto de funções $\{f_1, \dots, f_n\}$ e procurar uma estimativa da imagem no espaço gerado por esta base.

Dentre estas, uma escolha razoável é a imagem que maximiza a probabilidade de se obter os dados, ou seja, a solução do seguinte problema:

$$\begin{aligned} & \max P(\mathbf{b}|f) \\ & \text{s.a. :} \\ & f(\mathbf{x}) \geq 0; \\ & f(\mathbf{x}) = \sum_{j=1}^n y_j f_j(\mathbf{x}). \end{aligned} \quad (4)$$

Onde $\mathbf{b} \in \mathbb{R}^m$ é um vetor contendo os dados do tomógrafo. Isto é, cada componente b_i de \mathbf{b} contém o número de coincidências contadas no par i de detectores. Utilizando a restrição de discretização:

$$\int_{V_i} a_i(\mathbf{x})f(\mathbf{x})d\mathbf{x} = \int_{V_i} a_i(\mathbf{x}) \sum_{j=1}^n y_j f_j(\mathbf{x})d\mathbf{x} = \sum_{j=1}^n y_j \int_{V_i} a_i(\mathbf{x})f_j(\mathbf{x})d\mathbf{x}. \quad (5)$$

Definindo $a_{ij} := \int_{V_i} a_i(\mathbf{x}) f_j(\mathbf{x}) d\mathbf{x}$, $A := (a_{ij})$ e $\mathbf{y} := (y_j)$:

$$P(\mathbf{b}|f) = P(\mathbf{b}|\mathbf{y}) = \prod_{i=1}^m e^{-(A\mathbf{y})_i} \frac{(A\mathbf{y})_i^{b_i}}{b_i!}. \quad (6)$$

Aplicando logaritmos e eliminando termos constantes, resta o problema de otimização:

$$\begin{aligned} \max L(\mathbf{y}) &:= \sum_{i=1}^m \{b_i \log(A\mathbf{y})_i - (A\mathbf{y})_i\} \\ \text{s.a. :} & \\ &\sum_{j=1}^n y_j f_j(\mathbf{x}) \geq 0. \end{aligned} \quad (7)$$

Apesar da função $L(\mathbf{x})$ do problema acima ser côncava, ele ainda não é tratável devido à forma da restrição de não-negatividade. Uma solução é discretizar em uma grade regular de *voxels* cúbicos [VSK85]. Isto permite substituir (7) por:

$$\begin{aligned} \max L(\mathbf{y}) &= \sum_{i=1}^m \{b_i \log(A\mathbf{y})_i - (A\mathbf{y})_i\} \\ \text{s.a. :} & \\ &y_j \geq 0 \quad j = 1, \dots, n. \end{aligned} \quad (8)$$

Algoritmos para o Modelo Estatístico

Qualquer método iterativo utilizado para resolver (8) envolve, grosso modo, o equivalente ao cálculo de $\nabla L(\mathbf{y})$ a cada iteração. Nos problemas reais, o número de incógnitas pode facilmente passar de $2,6 * 10^5$ (por exemplo uma imagem de $128 \times 128 \times 16$ voxels) e, com um número de equações semelhante, alta velocidade de convergência é necessária.

O algoritmo EM (Expectation Maximization) pode ser obtido com simples manipulação das condições KKT (Karush Kuhn Tucker) para o problema (8). Demonstrações de convergência são conhecidas há algum tempo. Uma referência útil no assunto é o artigo de Vardi, Shepp e Kaufman [VSK85], onde também podem ser encontrados uma bibliografia interessante e um panorama histórico

do algoritmo. Apesar de possuir o desenvolvimento teórico concluído há algum tempo, a aplicação prática do algoritmo EM sempre foi restrita devido à sua lenta convergência e, conseqüentemente, ao seu esforço computacional elevado.

Graças à qualidade superior, à modelagem flexível, ao fato de assegurar automaticamente a não negatividade da imagem, entre outras razões, as imagens estatísticas continuaram atraentes e estimularam a procura de métodos mais rápidos para resolver (8). Dentre as técnicas utilizadas com maior sucesso, ainda que com implementação muito simples, estão as baseadas em uma atualização incremental da solução. Em todos estes métodos, as iterações são compostas de diversas subiteraões, nas quais apenas um subconjunto dos dados é considerado de cada vez. Espera-se que, em cada uma destas subiteraões, a melhora na solução seja equivalente a uma iteração completa de um algoritmo não incremental, porém com o esforço computacional reduzido proporcionalmente ao tamanho do subconjunto considerado. Esta abordagem tem diversas características que são convenientes ao caso de tomografia por emissão e merece um estudo mais aprofundado.

Objetivos

A intenção desta monografia é apresentar e discutir métodos incrementais cuja a convergência é garantida pela diminuição controlada do tamanho de passo (também conhecido como parâmetro de relaxação) de tal forma que o erro na aproximação do gradiente ao fim de uma iteração completa torne-se menor com o passar das iteraões.

Há dois principais objetivos na presente dissertação. O primeiro é fornecer demonstrações de convergência para tais algoritmos. Recentemente, Ahn e Fessler [AhF03] apresentaram resultados interessantes para o caso mais geral em que um termo *a priori* pode ser incorporado à função objetivo. Seus métodos envolvem, porém, uma alteração da função de verossimilhança a fim de garantir a limitação e continuidade de Lipschitz do gradiente na região factível. As demonstrações que serão aqui apresentadas excluem a necessidade desta alteração

(que tem o custo de uma iteração), ao menos para o caso fundamental sem regularização (8), e permitem que o tamanho de passo possa variar de acordo com as subiterações do método. O segundo objetivo é estudar como a forma de diminuição do tamanho de passo pode influenciar a convergência do algoritmo e a qualidade da imagem reconstruída quando comparadas a resultados obtidos com métodos sem relaxação.

Incidentalmente, é dada uma visão intuitiva do algoritmo de Retroprojeção Filtrada com a intenção de estimular o interesse pela maior flexibilidade da modelagem estatística em relação ao modelo das projeções.

Capítulo 1

Tomografia por Emissão de Póstrons

O presente capítulo destina-se a dar uma descrição mais minuciosa dos efeitos presentes em uma leitura PET sem se deter em aspectos quantitativos. O entendimento destes fenômenos abre espaço para modelagens mais sofisticadas, que tendem a diminuir a influência do ruído na imagem reconstruída porque levam em conta várias das possíveis fontes da flutuação nos dados, indo além da simples consideração do fato de que a emissão de póstrons é um fenômeno de Poisson.

1.1 Emissão e Desintegração de Póstrons

A emissão dos póstrons pelo isótopo radioativo é um fenômeno intrinsecamente aleatório regido por uma distribuição de Poisson, ou seja, o número de póstrons emitidos por unidade de volume em um dado espaço de tempo é uma variável aleatória de Poisson com parâmetro proporcional à concentração média do marcador naquele volume e espaço de tempo. Nas situações em que uma leitura PET é executada, esta concentração varia com o tempo. Entretanto, em grande parte dos casos, o interesse clínico estará em elementos estáticos do paciente e a variação da concentração do marcador ao longo do tempo não será fundamental na identificação destes elementos, pois ela ocorrerá devido ao decaimento do

marcador, que ocorre de forma proporcional à concentração do mesmo, fazendo com que a leitura seja equivalente a uma tomografia de uma imagem estática cujo valor é proporcional à concentração inicial da imagem. Mais precisamente, teremos $f(\mathbf{x}, t) = e^{-\lambda t} f(\mathbf{x}, 0)$, onde λ é uma constante característica do marcador. Logo

$$\int_0^T \int_R f(\mathbf{x}, t) d\mathbf{x} dt = \int_0^T \int_R e^{-\lambda t} f(\mathbf{x}, 0) d\mathbf{x} dt = \int_R \frac{1 - e^{-\lambda T}}{\lambda} f(\mathbf{x}, 0) d\mathbf{x}. \quad (1.1)$$

Assim, definindo $f(\mathbf{x}) := \frac{1 - e^{-\lambda T}}{\lambda} f(\mathbf{x}, 0)$, podemos ver que, nesses casos, uma imagem pode ser obtida utilizando o problema (8) apresentado na introdução.

Uma vez que um pósitron é emitido, ele percorre um pequeno caminho até que se aniquila em um encontro com um elétron, gerando dois fótons com energia de 511keV que viajam em direções quase opostas. O fato da aniquilação não ocorrer no local exato da emissão do pósitron pode ser levado em consideração nas funções $a_i(\mathbf{x})$ que representam a probabilidade de um pósitron emitido no ponto \mathbf{x} ser detectado no par i de detectores. Da mesma forma, pode-se modelar o fato da emissão dos fótons gerados pela aniquilação não ocorrer em direções exatamente opostas. Seguindo os passos apresentados na introdução, podemos ver que estes ajustes apenas seriam notados nos coeficientes da matriz A na função objetivo do problema (8).

Veja a Figura 1.1 para uma ilustração destes fenômenos. O desenho mostra o ponto onde o pósitron foi emitido (simbolizado por \times), o ponto onde ocorreu a aniquilação (ponto comum entre os círculos com $+$ e $-$ dentro), a trajetória dos fótons caso a emissão fosse exatamente em direções opostas (linha tracejada) e a emissão que realmente ocorreu (linha sólida).

1.2 Atenuação

Outro fenômeno a ser considerado é a atenuação de fótons. Isto é, a possibilidade de um dos fótons emitidos ser absorvido pelo objeto de estudo antes de alcançar o anel detector fazendo com que apenas um dos fótons do par seja detectado e, conseqüentemente, inutilizando a detecção. A fim de levar em consideração

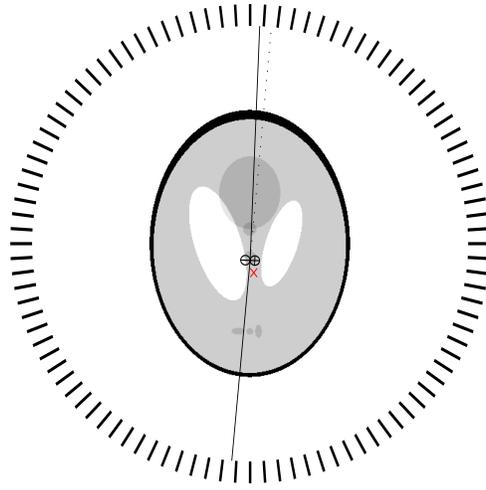


Figura 1.1: Emissão e desintegração de pósitron.

este fenômeno, a solução usual é efetuar uma tomografia prévia por transmissão utilizando uma fonte de radiação com energia semelhante à dos fótons emitidos na aniquilação pósitron-elétron (511keV). O resultado desta tomografia permite que os dados sejam corrigidos para levar em conta a atenuação ou que se possa alterar a matriz A considerando o *mapa de atenuação*, ou seja, a imagem obtida pela tomografia por transmissão.

1.3 Espalhamento

Pode ocorrer que um dentre os fótons emitidos desvie sua trajetória por espalhamento. Uma vez que a energia dos fótons emitidos é bem conhecida, a perda de energia correspondente ao desvio da trajetória permite eliminar a detecção

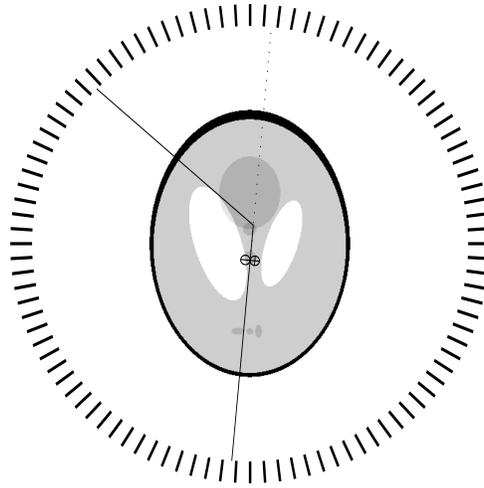


Figura 1.2: Exemplo do fenômeno de espalhamento.

de fótons espalhados. Este efeito acabaria sendo levado em conta na tomografia prévia por transmissão e incluído como parte da atenuação. Modelos mais sofisticados podem ser pensados para considerar o fato de que há uma janela em torno de 511keV onde uma detecção é considerada válida. A Figura 1.2 mostra uma representação esquemática do desvio que pode ser causado e como o mesmo pode gerar uma leitura errônea na detecção simultânea.

1.4 Outros Efeitos

Sem pretensões de fornecer uma lista completa dos fenômenos presentes em uma tomografia por emissão de pósitrons, foram identificados, nas seções anteriores deste capítulo, os efeitos que são dominantes em uma leitura PET. Existem

ainda outros fenômenos secundários que podem se tornar importantes, como a possibilidade de mais de uma aniquilação ocorrer num espaço muito curto de tempo, tornando impossível determinar quais detecções de fótons devem ser emparelhadas, ou mesmo uma eventual movimentação do paciente durante a leitura, a qual poderia, em casos extremos, inutilizar o exame. Uma vez que não está entre os objetivos desta dissertação dissecar a física da Tomografia por Emissão de Pósitrons, nem outros efeitos são fundamentais na obtenção da imagem, não há razão para prosseguir com considerações acerca de detalhes menos importantes, tendo sido as idéias básicas expostas com clareza.

Capítulo 2

Métodos Analíticos

A intenção deste capítulo não é fornecer uma fundamentação teórica rígida, mas apenas apresentar uma intuição sobre as idéias por trás dos métodos analíticos que são comumente utilizados na reconstrução de imagens em tomografia computadorizada. Para a teoria omitida, as referências centrais são os livros de Kak e Slaney [KaS88] e de F. Natterer [Nat01], especialmente o segundo para a validação rigorosa dos resultados esboçados aqui. Os algoritmos de real interesse (métodos incrementais para recuperação estatística de imagens em tomografia por emissão) na presente dissertação serão discutidos com clareza de detalhes em capítulos posteriores.

Vale a pena mencionar que as projeções de uma imagem, conforme a definição que será dada logo abaixo, equivalem, no caso bidimensional, à transformada de Radon. Em dimensões maiores, a última considera integrais sobre hiperplanos ao invés de integrais de linha. Extensões dos resultados apresentados abaixo para o caso da transformada de Radon são possíveis, o leitor interessado deve procurar as referências acima.

2.1 Teorema da Projeção

Podemos definir uma imagem (em tons de cinza) n -dimensional como uma função $f : \mathbb{R}^n \mapsto \mathbb{R}$. O resultado por trás dos métodos analíticos para obtenção de imagens em tomografia relaciona a transformada de Fourier das projeções

com a da imagem. Como somente amostras das projeções são conhecidas podemos conhecer apenas componentes abaixo de um limite de frequência no espaço de Fourier. Isto torna necessária a hipótese de que a imagem seja *banda-limitada*, ao menos aproximadamente, para que possamos reconstruí-la com algum grau de precisão. Uma discussão relacionando a amostragem das projeções com a resolução esperada para a imagem obtida está fora do escopo da presente dissertação e pode ser encontrada em [Nat01].

Uma *projeção* é a integral de linha ao longo de uma reta. De forma precisa, seja $f : \mathbb{R}^n \mapsto \mathbb{R}$ a nossa imagem n -dimensional. Cada projeção é calculada como $\mathcal{P}[f(\cdot)](\mathbf{d}, \mathbf{x}) = \int_{\mathbb{R}} f(\mathbf{x} + t\mathbf{d})dt$, onde $\mathbf{x}, \mathbf{d} \in \mathbb{R}^n$ e $\|\mathbf{d}\|_2 = 1$. Por esta expressão pode-se ver que as projeções não se alteram por translações de \mathbf{x} na direção de \mathbf{d} , de forma que geralmente restringe-se \mathbf{x} ao conjunto $\mathbf{d}^\perp := \{\mathbf{x} | \mathbf{x} \cdot \mathbf{d} = 0\}$. Isto permite que $\mathcal{P}[f(\cdot)]$ seja escrita (para \mathbf{d} fixo) como uma função de $\mathbb{R}^{n-1} \mapsto \mathbb{R}$. Seja $B^\perp \in \mathbb{R}^{n \times (n-1)}$ uma matriz cujas colunas formam uma base ortonormal para \mathbf{d}^\perp . Assim, para $\mathbf{s} \in \mathbb{R}^{n-1}$:

$$\mathcal{P}[f(\cdot)](\mathbf{d}, \mathbf{s}) := \int_{\mathbb{R}} f(B^\perp \mathbf{s} + t\mathbf{d})dt. \quad (2.1)$$

A *transformada de Fourier* de uma função $f : \mathbb{R}^n \mapsto \mathbb{C}$ é definida, para $\mathbf{w} \in \mathbb{R}^n$, como:

$$\mathcal{F}[f(\cdot)](\mathbf{w}) := \int_{\mathbb{R}^n} f(\mathbf{t})e^{-i\mathbf{w} \cdot \mathbf{t}}d\mathbf{t}. \quad (2.2)$$

Define-se também a *transformada inversa de Fourier* de $f : \mathbb{R}^n \mapsto \mathbb{C}$, para $\mathbf{t} \in \mathbb{R}^n$:

$$\mathcal{F}^{-1}[f(\cdot)](\mathbf{t}) := \frac{1}{(2\pi)^n} \int_{\mathbb{R}^n} f(\mathbf{w})e^{i\mathbf{t} \cdot \mathbf{w}}d\mathbf{w}. \quad (2.3)$$

Assumimos, por hipótese, que as transformadas de Fourier da imagem e das projeções existam e que a relação de inversão abaixo vale:

$$\mathcal{F}^{-1}[\mathcal{F}[f(\cdot)](\cdot)](\mathbf{t}) = \mathcal{F}[\mathcal{F}^{-1}[f(\cdot)](\cdot)](\mathbf{t}) = f(\mathbf{t}). \quad (2.4)$$

O teorema fundamental para reconstrução de imagens pela suas projeções, conhecido como *teorema da fatia de Fourier* ou *teorema da projeção*, pode ser

obtido de maneira muito simples. Avaliemos $\mathcal{F}[\mathcal{P}[f(\cdot)](\mathbf{d}, \cdot)](\mathbf{v})$ usando a definição (2.1), para \mathbf{d} , $\|\mathbf{d}\|_2 = 1$ fixo:

$$\begin{aligned}\mathcal{F}[\mathcal{P}[f(\cdot)](\mathbf{d}, \cdot)](\mathbf{v}) &= \int_{\mathbb{R}^{n-1}} \mathcal{P}[f(\cdot)](\mathbf{d}, \mathbf{s}) e^{-i\mathbf{v}\cdot\mathbf{s}} d\mathbf{s} \\ &= \int_{\mathbb{R}^{n-1}} \int_{\mathbb{R}} f(B^\perp \mathbf{s} + \mathbf{s}\mathbf{d}) d\mathbf{s} e^{-i\mathbf{v}\cdot\mathbf{s}} d\mathbf{s}.\end{aligned}\quad (2.5)$$

Agora, suponhamos que $\mathbf{w} \in \mathbf{d}^\perp$, isto significa que $\mathbf{w} = B^\perp \mathbf{v}$ para algum $\mathbf{v} \in \mathbb{R}^{n-1}$, mais precisamente $\mathbf{v} = B^{\perp T} \mathbf{w}$. Notando que $(B^{\perp T} \mathbf{w}) \cdot \mathbf{s} = \mathbf{w} \cdot (B^\perp \mathbf{s} + \mathbf{s}\mathbf{d})$:

$$\begin{aligned}\mathcal{F}[\mathcal{P}[f(\cdot)](\mathbf{d}, \cdot)](B^{\perp T} \mathbf{w}) &= \int_{\mathbb{R}^{n-1}} \int_{\mathbb{R}} f(B^\perp \mathbf{s} + \mathbf{s}\mathbf{d}) e^{-i\mathbf{w}\cdot(B^\perp \mathbf{s} + \mathbf{s}\mathbf{d})} d\mathbf{s} d\mathbf{s} \\ &= \int_{\mathbb{R}^n} f(B\mathbf{t}) e^{-i\mathbf{w}\cdot(B\mathbf{t})} d\mathbf{t}.\end{aligned}\quad (2.6)$$

Onde $B = \begin{pmatrix} \mathbf{d} & | & B^\perp \end{pmatrix}$ e $\mathbf{t} = \begin{pmatrix} s \\ \mathbf{s} \end{pmatrix}$. Como $B^T B = I$, $\det(B) = \pm 1$. Isto implica que sob a transformação de variáveis $\mathbf{x} = B\mathbf{t}$ teremos:

$$\mathcal{F}[\mathcal{P}[f(\cdot)](\mathbf{d}, \cdot)](B^{\perp T} \mathbf{w}) = \mathcal{F}[f(\cdot)](\mathbf{w}). \quad (2.7)$$

Há uma contrapartida óbvia desta fórmula:

$$\mathcal{F}[\mathcal{P}[f(\cdot)](\mathbf{d}, \cdot)](\mathbf{v}) = \mathcal{F}[f(\cdot)](B^\perp \mathbf{v}). \quad (2.8)$$

O resultado acima demonstra que um hiperplano perpendicular a uma dada direção da transformada de Fourier da imagem pode ser obtido pela transformada $n-1$ dimensional das projeções da imagem naquela direção. Isto permite, em princípio, que a imagem seja reconstruída com o auxílio de (2.4) se projeções suficientes forem conhecidas.

2.2 Retroprojeção Filtrada

Se $\mathcal{F}[f(\cdot)](\mathbf{w})$ for escrito em coordenadas cilíndricas e a fórmula de inversão da transformada de Fourier aplicada, teremos $\mathbf{w} = (\omega \cos \theta \quad \omega \sin \theta \quad w_3 \quad \dots \quad w_n)^T$ e a integral se torna:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^n} \int_0^\pi \int_{-\infty}^\infty \dots \int_{-\infty}^\infty \mathcal{F}[f(\cdot)](\mathbf{w}) e^{i\mathbf{w}\cdot\mathbf{x}} |\omega| d\omega dw_3 \dots dw_n d\theta. \quad (2.9)$$

Denotando $\mathbf{v} = (\omega \ w_3 \ \dots \ w_n)^T$, $B_\theta^\perp = \begin{pmatrix} \cos \theta & & \\ \sin \theta & & \\ & & I_{n-2} \end{pmatrix}$ e $\mathbf{x}_\theta = (x_1 \cos \theta + x_2 \sin \theta \ x_3 \ \dots \ x_n)$ resulta:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^n} \int_0^\pi \int_{\mathbb{R}^{n-1}} \mathcal{F}[f(\cdot)](B_\theta^\perp \mathbf{v}) e^{i\mathbf{v} \cdot \mathbf{x}_\theta} |\omega| d\mathbf{v} d\theta. \quad (2.10)$$

Aplicando a igualdade (2.8) temos, para¹ $\mathbf{d}_\theta = \begin{pmatrix} -\sin \theta \\ \cos \theta \\ \mathbf{0} \end{pmatrix}$:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^n} \int_0^\pi \int_{\mathbb{R}^{n-1}} \mathcal{F}[\mathcal{P}[f(\cdot)](\mathbf{d}_\theta, \cdot)](\mathbf{v}) |\omega| e^{i\mathbf{v} \cdot \mathbf{x}_\theta} d\mathbf{v} d\theta. \quad (2.11)$$

Agora, defina $\Omega(\mathbf{x}) = |x_1|$. A forma final da fórmula de inversão fica:

$$f(\mathbf{x}) = \frac{1}{2\pi} \int_0^\pi \mathcal{F}^{-1} \left[\mathcal{F}[\mathcal{P}[f(\cdot)](\mathbf{d}_\theta, \cdot)](\cdot) \Omega(\cdot) \right] (\mathbf{x}_\theta) d\theta. \quad (2.12)$$

Supondo que θ tenha sido amostrado a intervalos regulares, $\Delta\theta = \frac{\pi}{m}$:

$$\begin{aligned} f(\mathbf{x}) &\approx \frac{1}{2\pi} \sum_{k=0}^{m-1} \mathcal{F}^{-1} \left[\mathcal{F}[\mathcal{P}[f(\cdot)](\mathbf{d}_{k\Delta\theta}, \cdot)](\cdot) \Omega(\cdot) \right] (\mathbf{x}_{k\Delta\theta}) \Delta\theta \\ &= \frac{1}{2m} \sum_{k=0}^{m-1} \mathcal{F}^{-1} \left[\mathcal{F}[\mathcal{P}[f(\cdot)](\mathbf{d}_{k\Delta\theta}, \cdot)](\cdot) \Omega(\cdot) \right] (\mathbf{x}_{k\Delta\theta}). \end{aligned} \quad (2.13)$$

Esta fórmula, que envolve uma operação de *filtragem* (a multiplicação no espaço de Fourier), seguida de uma *retroprojeção* (a soma sob todos os ângulos) dos dados, é a base para o algoritmo de FBP (do inglês para Retroprojeção Filtrada, Filtered Backprojection). Na prática, as transformadas de Fourier são efetuadas via FFT e interpolação pode ser necessária para efetuar a retroprojeção.

Deve-se notar que a fórmula acima pode ser obtida por outras transformações de variáveis, levando a resultados semelhantes. A melhor forma de girar a direção das projeções durante a retroprojeção vai depender da geometria do tomógrafo. Numa leitura PET normal, o detector consiste de um cilindro em torno da região de interesse. A escolha óbvia é, portanto, girar em torno do eixo do cilindro detector. Uma discretização deste eixo seria equivalente a uma seqüência de operações bidimensionais ao longo de fatias transversais, tornando este o caso especial de maior interesse.

¹Neste caso, $\mathbf{0}$ representa o vetor coluna de zeros com $n - 2$ componentes.

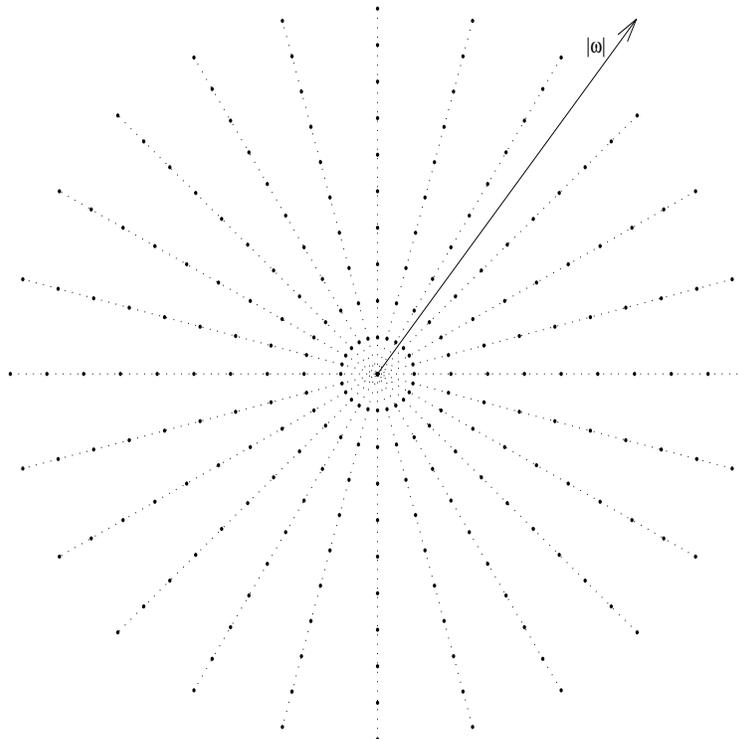


Figura 2.1: Amostras das projeções vistas no espaço de Fourier.

Observando a figura 2.1 pode-se ter uma visão mais intuitiva de cada um dos passos envolvidos em (2.13). Cada uma das linhas tracejadas é obtida pela transformada de Fourier de uma das projeções, conforme (2.8). Notando que, conforme a frequência aumenta, a densidade dos pontos no espaço de Fourier diminui, justifica-se o passo da filtragem em que é dado o devido peso às componentes de cada projeção no espaço de Fourier. De fato, a filtragem seguida da retroprojeção equivalem, aproximadamente, à inversão de Fourier da função radial por partes obtida completando as fatias $k\Delta\theta \leq \theta < (k+1)\Delta\theta$, $k = 0, \dots, m-1$ com valores constantes iguais ao das transformadas das projeções amostradas $\mathcal{F}[\mathcal{P}[f(\cdot)](\mathbf{d}_{k\Delta\theta}, \cdot)](\mathbf{v})$.

2.3 Outros Filtros

Desconsiderando efeitos de espalhamento, atenuação e a largura não nula dos detectores, se a contagem de fótons de uma leitura PET for alta o suficiente, de acordo com a lei dos grandes números, os dados podem funcionar como projeções da distribuição do marcador na região de interesse e, desta forma, as idéias acima poderiam ser utilizadas para a obtenção da imagem desejada.

A experiência mostra, entretanto, que a flutuação estatística em PET é grande demais para ser desconsiderada e a utilização do filtro $|\omega|$ pode amplificar demais as componentes de alta frequência do ruído e, por isso, melhores resultados podem ser obtidos multiplicando este filtro, no espaço de Fourier, por algo que tire a ênfase das frequências mais altas.

Vejamos uma ilustração desta idéia com um exemplo numérico simples, utilizando o algoritmo de FBP descrito acima. Primeiro na versão original e depois multiplicando o filtro $|\omega|$ por uma função sinc² (filtro de Shepp-Logan). Na Figura 2.2 podemos ver, na imagem, o efeito benéfico da desenfaturização das altas frequências, que torna a imagem mais suave. A filtragem pode, porém, suavizar demais, tornando as bordas menos definidas.

A imagem utilizada foi o conhecido *espectro de Shepp-Logan*³, visto na Figura 4.1 (definições detalhadas serão dadas no capítulo 4). Os dados foram adquiridos com 120 vistas igualmente espaçadas e o número de coincidências contadas foi 715,863. Cada projeção foi calculada e depois ruído de Poisson foi aplicado. A largura não nula do detector foi ignorada, assim como todos outros efeitos presentes numa leitura PET.

A Figura 2.3 ilustra o efeito do ruído na transformada de uma projeção. De cima para baixo temos: a transformada de Fourier (em valor absoluto) de uma das projeções; o mesmo para o caso com ruído, evidenciando as componentes de alta frequência introduzidas; por fim, o efeito do filtro sinc.

²A função sinc é definida como $\text{sinc}(x) = \frac{\text{sen}(\pi x)}{\pi x}$, se $x \neq 0$ e $\text{sinc}(0) = 1$.

³A palavra “espectro” parece ser a mais apropriada tradução do termo inglês “phantom” geralmente utilizado em processamento de imagens. O vocábulo “fantasma” guardaria maior equivalência com “ghost”, ambos associados mais fortemente ao sobrenatural.

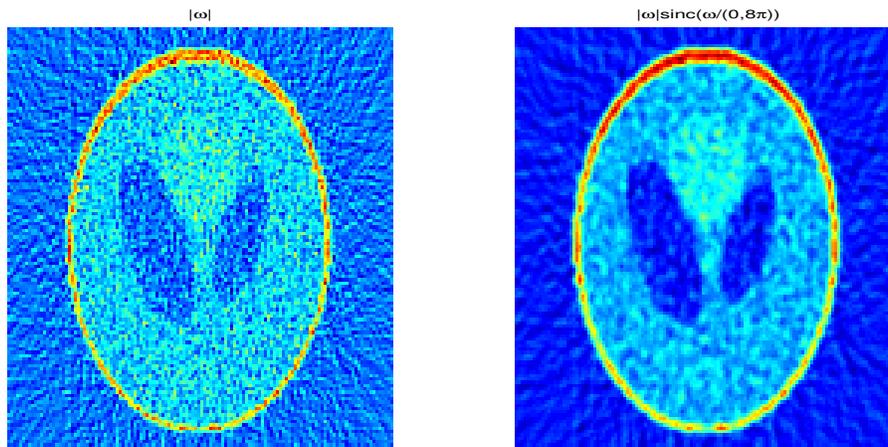


Figura 2.2: Comparação entre filtros em FBP.

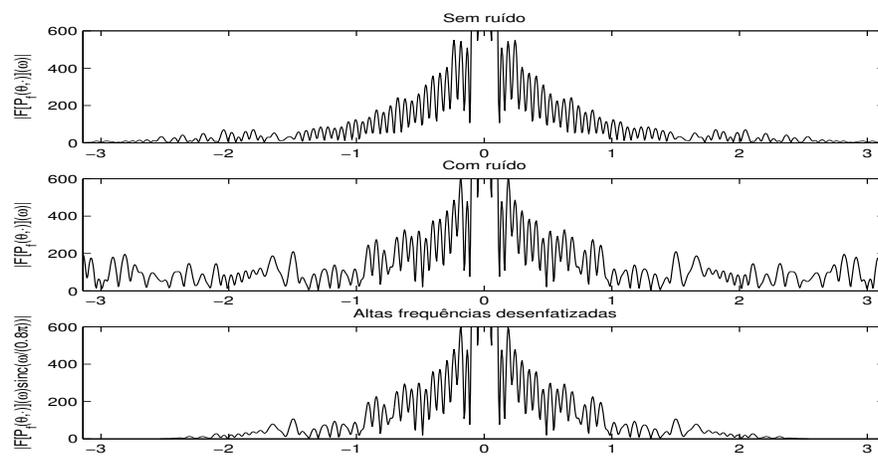


Figura 2.3: Efeito do ruído na transformada de uma projeção.

Capítulo 3

Reconstrução Estatística de Imagens em Tomografia por Emissão de Pósitrons

No capítulo anterior, foram apresentados métodos analíticos para a reconstrução de imagens por projeções que podem ser utilizados em PET. As hipóteses para a utilização destes métodos são, entretanto, muito simplificadoras. Melhores resultados podem ser obtidos com a modelagem estatística apresentada na introdução, já que muitos dos efeitos presentes na leitura podem ser incorporados ao problema de forma natural. A modelagem mais sofisticada apresenta, porém, a contrapartida de ser um problema impossível de ser resolvido analiticamente, tornando necessário o uso de métodos iterativos para aproximarmos a solução.

A abordagem estatística leva a um programa não linear convexo. Como argumentado na introdução, praticamente qualquer método padrão proposto para a sua solução envolve o cálculo do gradiente da função objetivo em cada iteração. Devido às dimensões do problema, o custo computacional desta operação é bastante elevado, tornando necessária uma alta velocidade de convergência. Neste capítulo serão apresentados métodos para a solução deste problema, começando com o algoritmo EM e prosseguindo com alguma de suas versões incrementais. Aspectos práticos de escolha de parâmetros serão abordados no capítulo 4, deixando os estudos teóricos de convergência para o capítulo 5.

3.1 Expectation Maximization Clássico

As condições KKT para (8) são:

$$y_j^* \geq 0 \quad j = 1, \dots, n; \quad (3.1a)$$

$$\mu_j^* y_j^* = 0 \quad j = 1, \dots, n; \quad (3.1b)$$

$$\mu_j^* \geq 0 \quad j = 1, \dots, n; \quad (3.1c)$$

$$\sum_{i=1}^m a_{ij} \left(\frac{b_i}{(A\mathbf{y}^*)_i} - 1 \right) + \mu_j^* = 0 \quad j = 1, \dots, n. \quad (3.1d)$$

Multiplicando cada uma das equações (3.1d) por y_j^* e utilizando (3.1b) temos:

$$y_j^* = \frac{y_j^*}{\sum_{i=1}^m a_{ij}} \sum_{i=1}^m \frac{a_{ij} b_i}{(A\mathbf{y}^*)_i} \quad j = 1, \dots, n. \quad (3.2)$$

Partindo de qualquer ponto inicial $\mathbf{y}^{(0)} \in \mathbb{R}_{++}^n$, o algoritmo EM (Expectation Maximization) é definido como a iteração de ponto fixo implícita na condição necessária para otimalidade acima:

$$y_j^{(k+1)} = \frac{y_j^{(k)}}{\sum_{i=1}^m a_{ij}} \sum_{i=1}^m \frac{a_{ij} b_i}{(A\mathbf{y}^{(k)})_i} \quad j = 1, \dots, n. \quad (3.3)$$

É fácil ver que estas iterações mantêm a positividade da solução inicial. Outra característica interessante pode ser obtida somando $s_j y_j^{(k+1)}$ na definição do algoritmo, onde $s_j = \sum_{i=1}^m a_{ij}$:

$$\sum_{j=1}^n s_j y_j^{(k+1)} = \sum_{i=1}^m b_i \quad \forall \mathbf{y}^{(k)} \in \mathbb{R}^n. \quad (3.4)$$

Isto mostra que qualquer que seja a solução inicial, a iteração seguinte terá uma contagem de fótons compatível com os dados. Outro ponto de interesse no algoritmo EM é que não apenas ele converge globalmente para uma solução de máxima verossimilhança, como converge monotonicamente. A implicação prática é que podemos partir de qualquer estimativa e melhorar a verossimilhança com uma iteração do algoritmo. Para demonstração destes resultados veja [VSK85] e as referências contidas no artigo.

Apesar de todas as características atraentes, o algoritmo EM apresenta uma razão de convergência proibitiva, podendo levar várias centenas de iterações

para atingir convergência numérica. Um outro ponto negativo é que a imagem de máxima verossimilhança pode apresentar um aspecto “granuloso”, efeito de uma excessiva adequação da solução aos dados. Se, partindo de uma solução inicial homogênea, as iterações do algoritmo forem interrompidas precocemente pode-se evitar (ou ao menos controlar) a amplificação excessiva do ruído. Mesmo nestes casos, ainda é necessário iterar algumas dezenas de vezes para a obtenção da imagem.

A Figura 3.1 mostra uma comparação simples entre FBP e EM efetuada aplicando os dois algoritmos ao mesmo conjunto de dados. Podemos ver que, apesar de ambas reconstruções serem altamente afetadas pelo ruído, a solução de máxima verossimilhança não apresenta os elementos “radiais” apresentados pela imagem analítica. Apesar da imagem de FBP exibir um melhor contraste, a imagem de máxima verossimilhança pode ser facilmente corrigida aplicando uma estratégia de truncamento. Em contrapartida, na Figura 2.2, vemos que um filtro apropriado em FBP pode gerar um resultado mais suave do que o obtido pelo algoritmo EM. O número de iterações utilizados para o algoritmo EM foi da ordem de grandeza do que é recomendado em [VSK85], valor obtido experimentalmente.

Na Figura 3.2 as imagens obtidas por FBP com o filtro sinc e a de máxima verossimilhança são exibidas após normalizadas: as componentes abaixo de zero são zeradas e os valores acima de um limite superior y_{max} são truncados e exibidos como se valessem y_{max} ¹. Nota-se que, por inspeção visual, é difícil determinar qual é melhor reconstrução.

Vale lembrar que os testes não simularam várias das fontes de ruído que podem ser modeladas no método estatístico e são ignoradas por FBP. Além disso, o algoritmo EM não requer a escolha de nenhum parâmetro, enquanto tivemos que escolher, por inspeção, o fator $\frac{1}{0,8\pi}$ usado no algoritmo de FBP. Resultados como estes estimularam a busca de alternativas para a solução de (8) sem o peso computacional imposto pelo algoritmo EM.

¹Foi utilizado y_{max} igual a 95% do maior valor obtido na imagem de FBP.

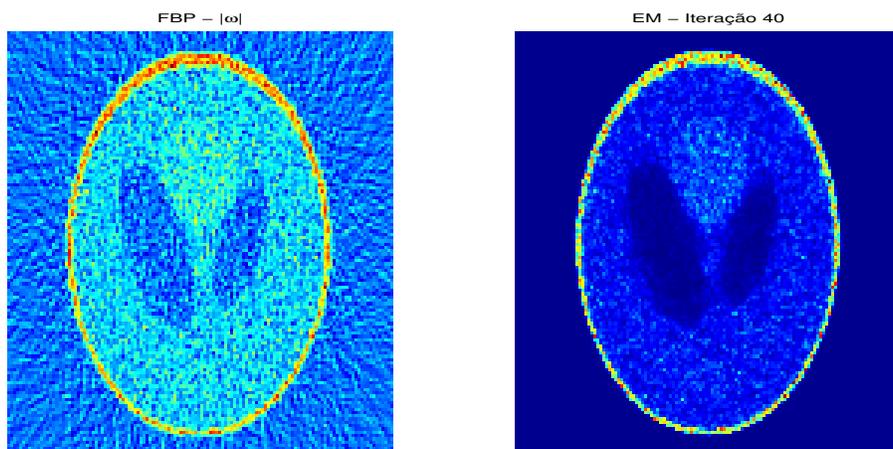


Figura 3.1: Comparação FBP×EM.

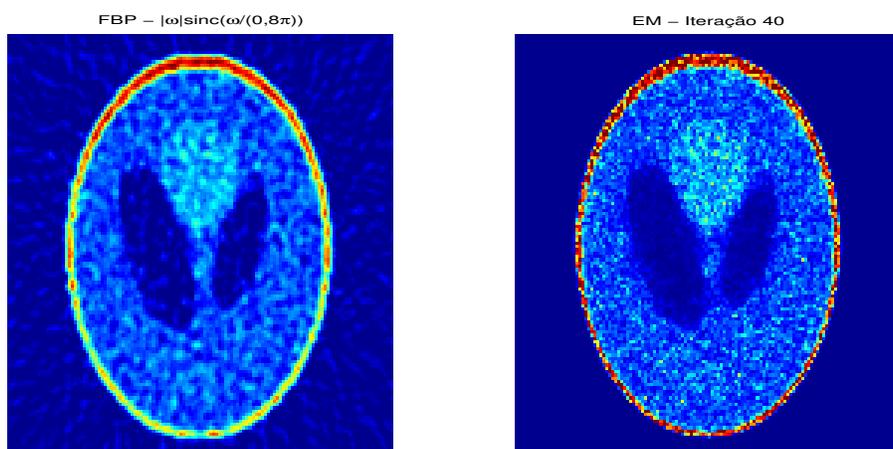


Figura 3.2: Comparação FBP×EM, imagens normalizadas.

3.2 Gradientes Incrementais

Seja o problema

$$\begin{aligned} \min c(\mathbf{y}) &= \sum_{l=1}^N c_l(\mathbf{y}) \\ \text{s.a. : } \mathbf{y} &\in \mathbb{X}. \end{aligned} \quad (3.5)$$

Onde $c_l(\mathbf{y}) : \mathbb{R}^n \mapsto \mathbb{R}$ são funções convexas e \mathbb{X} é um subconjunto não vazio, fechado e convexo de \mathbb{R}^n . O método dos gradientes incrementais é definido como:

$$\begin{aligned} \mathbf{y}^{(k,0)} &= \mathbf{y}^{(k)} \\ \mathbf{y}^{(k,l)} &= \mathcal{P}_{\mathbb{X}} \left(\mathbf{y}^{(k,l-1)} - \lambda_k \nabla c_l(\mathbf{y}^{(k,l-1)}) \right) \\ \mathbf{y}^{(k+1)} &= \mathbf{y}^{(k,N)}. \end{aligned} \quad (3.6)$$

Onde $\mathcal{P}_{\mathbb{X}}(\mathbf{y}) := \arg \min_{\mathbf{x} \in \mathbb{X}} \|\mathbf{x} - \mathbf{y}\|$. Se os gradientes ∇c_l são limitados em $\overline{\text{conv}}\{\mathbf{y}^{(k,l)}\}$ e o conjunto das soluções ótimas é não vazio e limitado, as seguintes condições para o tamanho passo são suficientes para a convergência do algoritmo para uma solução ótima:

$$\lambda_k > 0, \quad \sum_{k=0}^{\infty} \lambda_k = \infty, \quad \sum_{k=0}^{\infty} \lambda_k^2 < \infty. \quad (3.7)$$

Veja o artigo de Nedić e Bertsekas [NeB01] para a demonstração deste resultado.

A aplicação deste método ao caso de tomografia por emissão esbarra na possível não limitação do gradiente. Não é difícil ver que podemos substituir a função objetivo por uma aproximação linear ou quadrática nas vizinhanças de descontinuidades, de forma a contornar este problema sem alterar a solução de máxima verossimilhança. O algoritmo OS-SPS (Ordered Subsets - Separable Paraboloidal Surrogates) [AhF03] precisa de uma modificação deste tipo e inclui uma pré-multiplicação do gradiente por uma matriz constante diagonal definida positiva.

A grande vantagem dos gradientes incrementais é vista quando os gradientes $\nabla c_l(\mathbf{y})$ se assemelham, em direção, a $\nabla c(\mathbf{y})$. Neste caso, cada subiteração é praticamente equivalente a uma iteração do método não incremental, com o custo computacional reduzido proporcionalmente ao número de subiterações.

3.3 Ordered Subsets - Expectation Maximization

O primeiro algoritmo que pode ser reconhecido como uma aplicação de métodos incrementais para reconstrução de imagens em tomografia é ART (Algebraic Reconstruction Technique), veja o trabalho de Herman e Meyer [HeM93]. ART não procura resolver o problema de máxima verossimilhança, mas sim um problema de quadrados mínimos ponderado.

Para resolver (8), entretanto, o primeiro algoritmo incremental² amplamente reconhecido foi OS-EM (Ordered Subsets - Expectation Maximization) de Hudson e Larkin [HuL94]. A idéia é aplicar, sucessivamente, iterações EM a subconjuntos dos dados. Precisamos escolher N subconjuntos \mathcal{N}_l de índices tais que:

$$\bigcup_{l=1}^N \mathcal{N}_l = \{1, \dots, m\} \text{ e } \mathcal{N}_i \cap \mathcal{N}_j = \emptyset, \quad i \neq j. \quad (3.8)$$

A definição de OS-EM é, então:

$$\begin{aligned} \mathbf{y}^{(k,0)} &= \mathbf{y}^{(k)} \\ y_j^{(k,l)} &= \frac{y_j^{(k,l-1)}}{\sum_{i \in \mathcal{N}_l} a_{ij}} \sum_{i \in \mathcal{N}_l} \frac{b_i a_{ij}}{(A\mathbf{y}^{(k,l-1)})_i}, \quad l = 1, \dots, N \\ \mathbf{y}^{(k+1)} &= \mathbf{y}^{(k,N)}. \end{aligned} \quad (3.9)$$

Estas subiterações podem ser escritas como:

$$\mathbf{y}^{(k,l)} = \mathbf{y}^{(k,l-1)} + D_l(\mathbf{y}^{(k,l-1)}) \nabla L_l(\mathbf{y}^{(k,l-1)}). \quad (3.10)$$

Onde $D_l(\mathbf{y}) := \text{diag} \left\{ \frac{y_j}{\sum_{i \in \mathcal{N}_l} a_{ij}} \right\}$ e $L_l(\mathbf{y}) := \sum_{i \in \mathcal{N}_l} \{b_i \log(A\mathbf{y})_i - (A\mathbf{y})_i\}$. Esta expressão nos remete a (3.6), mas alguns detalhes saltam à vista: o algoritmo dispensa a operação $\mathcal{P}_{\mathbb{R}_+}$ porque é multiplicativo e, portanto, garante a positividade de cada iteração; não há uma estratégia de diminuição do tamanho de passo, o que impede a convergência no caso em que o sistema $A\mathbf{y} = \mathbf{b}$ é inconsistente. Nestes casos, o algoritmo alcança um ciclo limite dentre as subiterações. Finalmente, o algoritmo deixa em aberto o número e a escolha dos subconjuntos.

²Na comunidade de tomografia, algoritmos incrementais são conhecidos como algoritmos de “subconjuntos ordenados”. Em inglês, *ordered subsets* ou, sucintamente, OS.

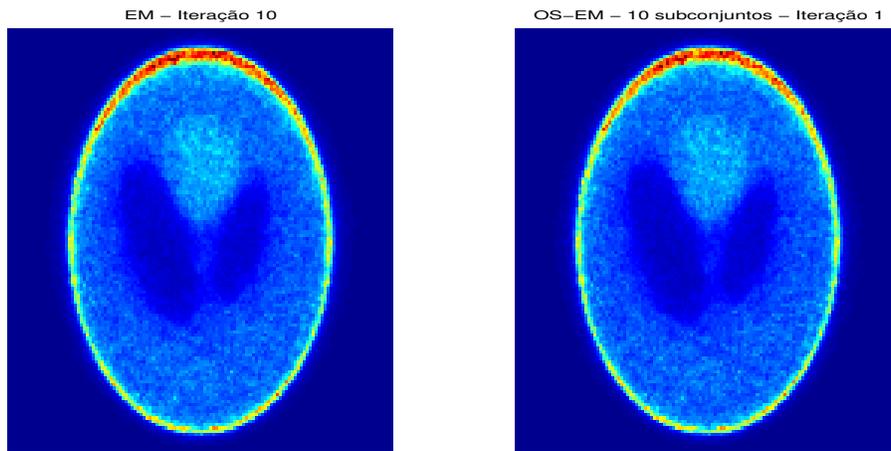


Figura 3.3: Comparação EM×OS-EM (10 subconjuntos).

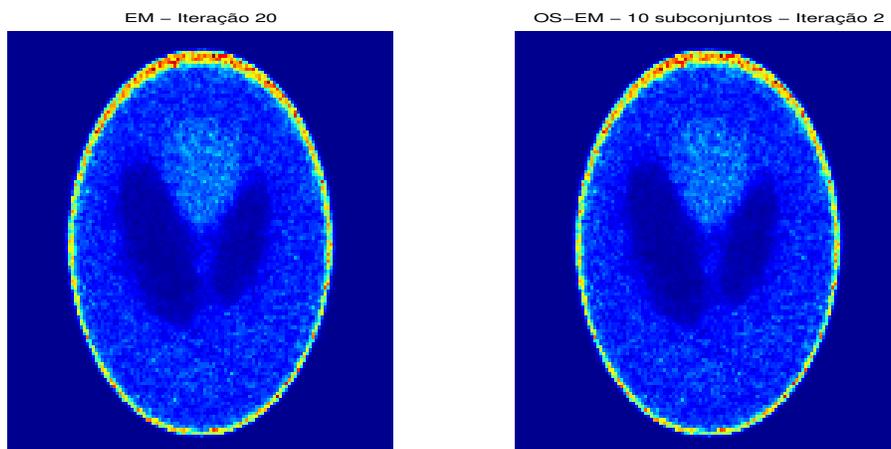


Figura 3.4: Comparação EM×OS-EM (10 subconjuntos).

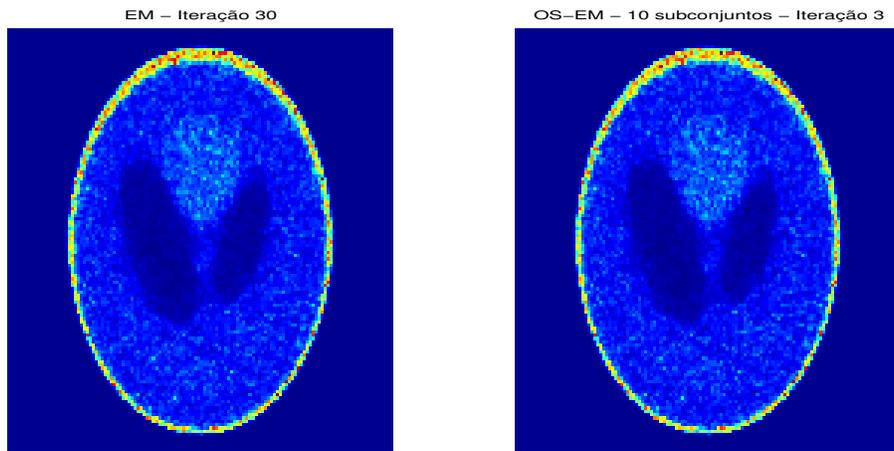


Figura 3.5: Comparação EM×OS-EM (10 subconjuntos).

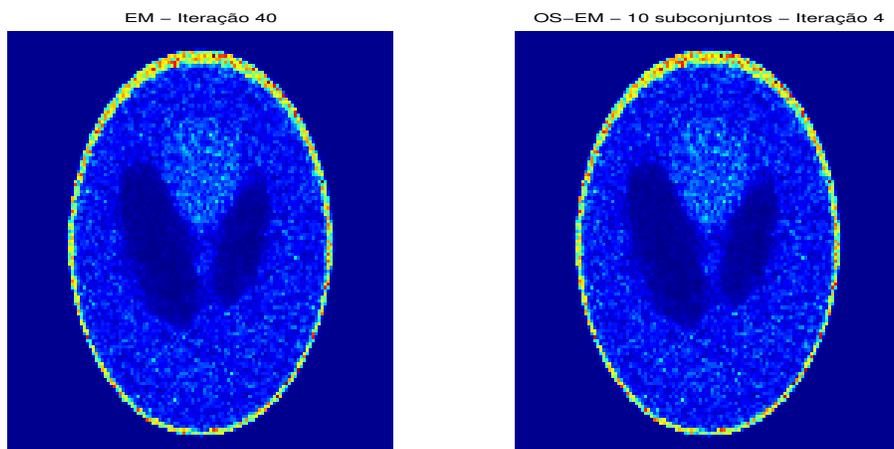


Figura 3.6: Comparação EM×OS-EM (10 subconjuntos).

Mesmo sem ser globalmente convergente, OS-EM apresenta uma aceleração notável nas primeiras iterações e foi adotado em alguns tomógrafos comerciais de PET. As figuras 3.3 até 3.6 exibem uma comparação entre as iterações do algoritmo EM e as de OS-EM com dez subconjuntos. Como podemos ver, a equivalência entre uma iteração completa do algoritmo EM e uma subiteração OS-EM é muito próxima, ao menos nestas primeiras iterações.

Os dados utilizados foram os mesmos dos exemplos anteriores. Cada subconjunto foi formado por todas as projeções de 12 ângulos, igualmente espaçados. Os ângulos amostrados foram $0^\circ; 1,5^\circ; \dots; 178,5^\circ$, o primeiro subconjunto compunha-se das projeções nos ângulos $0^\circ; 15^\circ; \dots; 165^\circ$, o segundo $1,5^\circ; 16,5^\circ; \dots; 166,5^\circ$, e assim sucessivamente.

3.4 Row Action Maximum Likelihood Algorithm

O algoritmo OS-EM, apresentado na seção anterior, não converge, em geral, porque é um algoritmo incremental com tamanho de passo fixo. Pode-se tornar OS-EM convergente diminuindo gradualmente o número de subconjuntos, caso em que o algoritmo se reduz ao EM clássico nas últimas iterações. Outra opção é adotar uma estratégia de diminuição do tamanho de passo. A aplicação deste tipo de estratégia em OS-EM não garante a convergência para a solução de máxima verossimilhança porque as funções $D_l(\mathbf{y})$ não são iguais para todas as subiterações [AhF03], [BrP96]. A solução é utilizar a mesma função em todas as subiterações. O algoritmo resultante é RAMLA (Row Action Maximum Likelihood Algorithm), proposto por Browne e De Pierro [BrP96]:

$$\begin{aligned} \mathbf{y}^{(k,0)} &= \mathbf{y}^{(k)} \\ \mathbf{y}^{(k,l)} &= \mathbf{y}^{(k,l-1)} + \lambda_k D(\mathbf{y}^{(k,l-1)}) \nabla L_l(\mathbf{y}^{(k,l-1)}), \quad l = 1, \dots, N \\ \mathbf{y}^{(k+1)} &= \mathbf{y}^{(k,N)}. \end{aligned} \quad (3.11)$$

Onde $D(\mathbf{y}) = \text{diag} \left\{ \frac{y_j}{p_j} \right\}$, $p_j > 0$. Para efeitos teóricos de convergência, a escolha dos fatores p_j é indiferente e nem é um objetivo desta dissertação discutir os efeitos práticos de tal escolha. Duas sugestões interessantes são

$p_j = \max_{l=1, \dots, N} \{ \sum_{i \in \mathcal{N}_l} a_{ij} \}$ e $p_j = \frac{\sum_{i=1}^m a_{ij}}{N}$, já que se λ_k for mantido constante igual a 1 e $N = 1$ (um único subconjunto contendo todos os dados), RAMLA se reduz ao EM clássico.

Além do número e da composição dos subconjuntos, RAMLA exige a escolha dos parâmetros λ_k . Se $L(\mathbf{y})$ é estritamente côncava, as condições (3.7) para o tamanho de passo garantem a convergência da sequência $\mathbf{y}^{(k)}$ gerada pelo algoritmo³ para a solução de máxima verossimilhança. Estas condições são, entretanto, assintóticas e não determinam completamente o conjunto de parâmetros, permitindo que os primeiros sejam determinados por meio de “treinamento” (veja a seção III de [BrP96]) e apenas depois, estratégias de diminuição do tamanho de passo sejam adotadas. Na literatura, por exemplo em [AhF03] e [TaK03], os tamanhos de passo adotados são⁴, normalmente, $O(\frac{1}{k})$. As condições (3.7) permitem, entretanto, que regras com comportamento assintótico igual a $O(\frac{1}{k^{\frac{1}{2} + \delta}})$, $\delta \in (0, \frac{1}{2}]$ sejam adotadas. De Pierro relatou (via comunicação oral) resultados melhores (em termos de maximização da verossimilhança) com regras $O(\frac{1}{\sqrt{k}})$. Apesar destas regras não se encaixarem nas demonstrações de convergência a seguir, com a adoção de δ suficientemente pequeno podemos aproximar o comportamento delas.

3.5 Dynamic RAMLA

Se o tamanho de passo puder variar dentre as subiterações de RAMLA, teremos:

$$\begin{aligned} \mathbf{y}^{(k,0)} &= \mathbf{y}^{(k)} \\ \mathbf{y}^{(k,l)} &= \mathbf{y}^{(k,l-1)} + \lambda_{kl} D(\mathbf{y}^{(k,l-1)}) \nabla L_l(\mathbf{y}^{(k,l-1)}), \quad l = 1, \dots, N \\ \mathbf{y}^{(k+1)} &= \mathbf{y}^{(k,N)}. \end{aligned} \quad (3.12)$$

Esta flexibilidade pode permitir escolhas para o tamanho de passo que favoreçam a qualidade da imagem. Uma forma interessante de variar estes parâmetros, proposta em [TaK03], procura fazer com que a propagação do erro estatístico

³É necessária uma condição do tipo $\lambda_k \leq \lambda$, para algum $\lambda > 0$ que garanta a positividade das iterações. Detalhes sobre a existência de tal λ , sobre condições para a convexidade estrita de $L(\mathbf{y})$ e sobre as demonstrações de convergência serão dados no capítulo 5.

⁴Neste parágrafo, por uma questão de concisão, $O(f(k))$ significa também $o(f(k))$

de cada projeção seja equilibrada ao fim de cada iteração. Com o tamanho de passo constante durante uma iteração completa, as chances de que o ruído das projeções processadas nas últimas subiterações tenha uma influência maior na imagem são grandes.

Num primeiro momento, imaginemos que os subconjuntos que vamos utilizar sejam compostos de uma projeção apenas cada um. Por uma questão de simplicidade, consideraremos que o dado b_i é acessado na i -ésima iteração. É claro que esta segunda hipótese não representa perda alguma de generalidade, já que podemos denotar por b_i qualquer uma das contagens obtidas e reordenar de forma adequada as linhas de A . A argumentação contida em [TaK03] começa reescrevendo (3.12):

$$\begin{aligned} y_j^{(k,l)} &= y_j^{(k,l-1)} + \lambda_{kl} \frac{y_j^{(k,l-1)}}{p_j} a_{lj} \left(\frac{b_l}{(\mathbf{A}\mathbf{y}^{(k,l-1)})_l} - 1 \right) \\ &= y_j^{(k,l-1)} + \lambda_{kl} \frac{y_j^{(k,l-1)}}{p_j} a_{lj} \left(\frac{b_l - (\mathbf{A}\mathbf{y}^{(k,l-1)})_l}{(\mathbf{A}\mathbf{y}^{(k,l-1)})_l} \right). \end{aligned} \quad (3.13)$$

Denotemos por \mathbf{y} a imagem real desejada. A igualdade acima pode ser aproximada por:

$$y_j^{(k,l)} \approx y_j^{(k,l-1)} + \lambda_{kl} \frac{y_j}{p_j} a_{lj} \left(\frac{b_l - (\mathbf{A}\mathbf{y}^{(k,l-1)})_l}{(\mathbf{A}\mathbf{y})_l} \right). \quad (3.14)$$

Se $\mathbf{y}^{(k,l-1)}$ é igual à imagem real e b_l for escrita como $(\mathbf{A}\mathbf{y})_l + \epsilon_l$, sendo ϵ_l o desvio estatístico, teremos, por (3.13):

$$y_j^{(k,l)} - y_j = \frac{\lambda_{kl} y_j a_{lj} \epsilon}{p_j (\mathbf{A}\mathbf{y})_l}. \quad (3.15)$$

Agora, utilizando (3.14) e supondo que $b_i = (\mathbf{A}\mathbf{y})_i$, $i \neq l$:

$$\begin{aligned} y_j^{(k,l+1)} - y_j &\approx \frac{\lambda_{kl} y_j a_{lj} \epsilon}{p_j (\mathbf{A}\mathbf{y})_l} + \frac{\lambda_{kl+1} y_j a_{l+1j}}{p_j (\mathbf{A}\mathbf{y})_{l+1}} \left(A(\mathbf{y} - \mathbf{y}^{(k,l)}) \right)_{l+1} \\ &= \frac{\lambda_{kl} y_j a_{lj} \epsilon}{p_j (\mathbf{A}\mathbf{y})_l} - \frac{\lambda_{kl+1} y_j a_{l+1j}}{p_j (\mathbf{A}\mathbf{y})_{l+1}} \sum_{j=1}^n a_{l+1j} \frac{\lambda_{kl} y_j a_{lj} \epsilon}{p_j (\mathbf{A}\mathbf{y})_l} \\ &= \frac{\lambda_{kl} y_j \epsilon}{p_j (\mathbf{A}\mathbf{y})_l} \left\{ a_{lj} - \frac{\lambda_{kl+1} a_{l+1j} \bar{g}_{ll+1}}{(\mathbf{A}\mathbf{y})_{l+1}} \right\}. \end{aligned} \quad (3.16)$$

Onde $\bar{g}_{mn} = \sum_{j=1}^n \frac{a_{mj} a_{nj} y_j}{p_j}$. Então, seja ϵ^* a componente do erro na projeção

$(\mathbf{A}\mathbf{y}^{(k,l+1)})_l$:

$$\begin{aligned}\epsilon^* &= (\mathbf{A}\mathbf{y}^{(k,l+1)})_l - (\mathbf{A}\mathbf{y})_l = \left(A(\mathbf{y}^{(k,l+1)} - \mathbf{y}) \right)_l \\ &\approx \sum_{j=1}^n a_{lj} \frac{\lambda_{kl} y_j \epsilon}{p_j (\mathbf{A}\mathbf{y})_l} \left\{ a_{lj} - \frac{\lambda_{kl+1} a_{l+1j} \bar{g}_{l+1}}{(\mathbf{A}\mathbf{y})_{l+1}} \right\} \\ &= \frac{\lambda_{kl} \epsilon}{(\mathbf{A}\mathbf{y})_l} \left\{ \bar{g}_{ll} - \frac{\lambda_{kl+1} \bar{g}_{l+1}^2}{(\mathbf{A}\mathbf{y})_{l+1}} \right\}.\end{aligned}\quad (3.17)$$

Assim, podemos escrever

$$\frac{\epsilon^*}{\epsilon} \approx \lambda_{kl} \frac{\bar{g}_{ll}}{(\mathbf{A}\mathbf{y})_l} \left\{ 1 - \lambda_{kl+1} \frac{\bar{g}_{l+1} \bar{g}_{l+1}}{(\mathbf{A}\mathbf{y})_{l+1}} \frac{\bar{g}_{l+1} \bar{g}_{l+1}}{\bar{g}_{ll} \bar{g}_{l+1}} \right\}.\quad (3.18)$$

Assumindo que⁵ $\bar{g}_{ll} \approx (\mathbf{A}\mathbf{y})_l$, $\bar{g}_{l+1} \approx (\mathbf{A}\mathbf{y})_{l+1}$ e $\frac{y_j}{p_j} \approx c$, c constante então encontramos

$$\frac{\epsilon^*}{\epsilon} \approx \lambda_{kl} (1 - \lambda_{kl+1} g_{l+1}^2).\quad (3.19)$$

Onde $g_{li} = \frac{\sum_{j=1}^n a_{lj} a_{ij}}{\sqrt{\sum_{j=1}^n a_{lj}^2 \sum_{j=1}^n a_{ij}^2}}$. Repetindo este raciocínio sucessivas vezes podemos definir uma aproximação $\epsilon_l^{(k)}$ para a razão $\frac{(\mathbf{A}\mathbf{y}^{(k+1)})_l - (\mathbf{A}\mathbf{y})_l}{(\mathbf{A}\mathbf{y}^{(k,l-1)})_l - (\mathbf{A}\mathbf{y})_l}$, que seria uma medida da eficiência com que o ruído se propaga a partir da l -ésima projeção:

$$\epsilon_l^{(k)} = \lambda_{kl} \prod_{i=l+1}^m \{1 - \lambda_{ki} g_{li}^2\}.\quad (3.20)$$

O interesse agora é escolher valores de λ_{kl} tais que $\epsilon_l^{(k)}$ seja constante em cada iteração. Os valores apropriados poderiam ser determinados por algum método iterativo. Porém, se substituirmos g_{li}^2 por uma constante denotada por $\frac{1}{\beta_0}$ (por exemplo a média entre os g_{li}^2), utilizando $\lambda_{kl} = \frac{\beta_0}{\beta_0 + l + \gamma k}$ teremos:

$$\begin{aligned}\epsilon_l^{(k)} &\approx \frac{\beta_0}{\beta_0 + l + \gamma k} \prod_{i=l+1}^m \left\{ 1 - \frac{\beta_0}{\beta_0 + i + \gamma k} \frac{1}{\beta_0} \right\} \\ &= \frac{\beta_0}{\beta_0 + l + \gamma k} \prod_{i=l+1}^m \left\{ \frac{\beta_0 + (i-1) + \gamma k}{\beta_0 + i + \gamma k} \right\} \\ &= \frac{\beta_0}{\beta_0 + l + \gamma k} \frac{\beta_0 + l + \gamma k}{\beta_0 + m + \gamma k} = \frac{\beta_0}{\beta_0 + m + \gamma k}.\end{aligned}\quad (3.21)$$

Que não depende de l , como desejado.

⁵De acordo com [TaK03], estas hipóteses são aproximadamente verdadeiras quando a imagem é um disco uniforme, as linhas da matriz A estão normalizadas e $p_j = \max_{i=1, \dots, m} a_{ij}$.

Agora vamos a fatos simples acerca desta sequência que fazem com que ela se adeque às hipótese das demonstrações de convergência do capítulo 5. Primeiro, é óbvio que $\lim_{k \rightarrow \infty} \frac{\lambda_{k1}}{\lambda_{kl}} = 1$, $l = 1, \dots, m$ e que a razão $\frac{\lambda_k}{\lambda_{k+1}}$ é limitada. Além disso:

$$\begin{aligned} |\lambda_{k1} - \lambda_{kl}| &= \frac{\beta_0}{\beta_0 + 1 + \gamma k} - \frac{\beta_0}{\beta_0 + l + \gamma k} \\ &= \beta_0 \left(\frac{\beta_0 + l + \gamma k - (\beta_0 + 1 + \gamma k)}{(\beta_0 + 1 + \gamma k)(\beta_0 + l + \gamma k)} \right) \leq \frac{\beta_0(l-1)}{\gamma^2 k^2}. \end{aligned} \quad (3.22)$$

Isto permite mostrar que $|\lambda_{k1} - \lambda_{kl}| = O(\lambda_k^{-2})$, $l = 1, \dots, m$, o que vai ao encontro das condições utilizadas no capítulo 5 para garantir a convergência do algoritmo. Se utilizarmos valores apropriados para p_j (por exemplo $p_j = \max_{l=1, \dots, N} \{ \sum_{i \in \mathcal{N}_l} a_{ij} \}$) podemos garantir a positividade das iterações e assegurar que DRAMA converge se $L(\mathbf{y})$ for estritamente côncava. Isto demonstra a conjectura aventada por Tanaka e Kudo em [TaK03] sobre a convergência do método. Para detalhes sobre a adaptação deste raciocínio para subconjuntos compostos por mais de uma projeção, o leitor pode procurar o referido artigo.

Capítulo 4

Simulações Numéricas

Neste capítulo, procura-se dar uma amostra dos resultados esperados com o uso de alguns dos algoritmos expostos no capítulo anterior. Nenhum experimento com DRAMA é realizado, o leitor deve se reportar a [TaK03] para resultados obtidos com o método. O objetivo principal, que guia a organização dos testes, é compreender que tipo de sequência de parâmetros de relaxação pode ser utilizada em RAMLA a fim de torná-lo competitivo (seja com relação à convergência ou à qualidade da imagem) com OS-EM. O interesse é, portanto, comparar o desempenho dos algoritmos e de forma alguma procura-se fornecer uma validação dos algoritmos; uma vez que ambos vêm sendo empregados largamente isto é desnecessário.

4.1 Simulando a Leitura de Tomografia por Emissão de Pósitrons

Nesta dissertação, um *espectro* é uma imagem discretizada utilizada para gerar os dados que simularão a tomografia¹. Em todos os experimentos realizados, a leitura foi simulada com uma projeção do espectro seguida pela aplicação de ruído de Poisson. A projeção é calculada como $A\mathbf{x}$ onde cada coeficiente a_{ij} é igual ao comprimento da intersecção entre a linha i ligando o centro de cada um dos detectores do par i e o pixel j . Não foram simulados efeitos como o

¹Não há risco de confusão com o termo homônimo, utilizado com o significado da representação da imagem no espaço de Foutier.

espalhamento e a atenuação dos fótons ou a distância percorrida pelo fóton antes de se aniquilar com um elétron. A Figura 4.1 exibe a discretização do *espectro de Shepp-Logan*, cuja definição pode ser encontrada no capítulo 3 de [Ka88]. Nas figuras 4.2 e 4.3 temos, respectivamente, as projeções da imagem e o resultado após a simulação do ruído de Poisson. A contagem aproximada de fótons desejada pode ser obtida por uma normalização apropriada da imagem antes do ruído ser gerado.

4.2 Avaliando as Reconstruções

É necessária a adoção de um critério preciso para a avaliação da qualidade da imagem. Uma das formas mais simples de quantificar a qualidade da reconstrução é a distância euclidiana entre a imagem reconstruída e o espectro utilizado. Uma vez que diferentes imagens podem ter intensidades totais diferentes é preciso que um fator de escala que leve isto em consideração seja introduzido. A *precisão pontual* é definida como:

$$-\sqrt{\frac{\sum_{j=1}^n (y_j - z_j)^2}{\sum_{j=1}^n (z_j - \bar{z})^2}}. \quad (4.1)$$

Onde \mathbf{z} é o espectro utilizado, \mathbf{y} é a imagem obtida pelo método sendo avaliado e $\bar{z} := \frac{1}{n} \sum_{j=1}^n z_j$.

4.3 Composição e Ordenação dos Subconjuntos

Os parâmetros livres dos algoritmos apresentados no capítulo anterior são o número de subconjuntos, a composição destes subconjuntos, a ordem em que eles serão processados e a sequência de parâmetros de relaxação. Em experimentos preliminares com OS-EM, foram testadas, para diversos graus de incrementalidade, quatro diferentes formas de escolher e ordenar os subconjuntos. A forma dos subconjuntos utilizados, a maneira como eles foram ordenados e os resultados dos testes são o assunto desta seção.

Quanto aos subconjuntos propriamente ditos, os testes utilizaram duas possíveis escolhas. Em ambas, todos os subconjuntos continham um

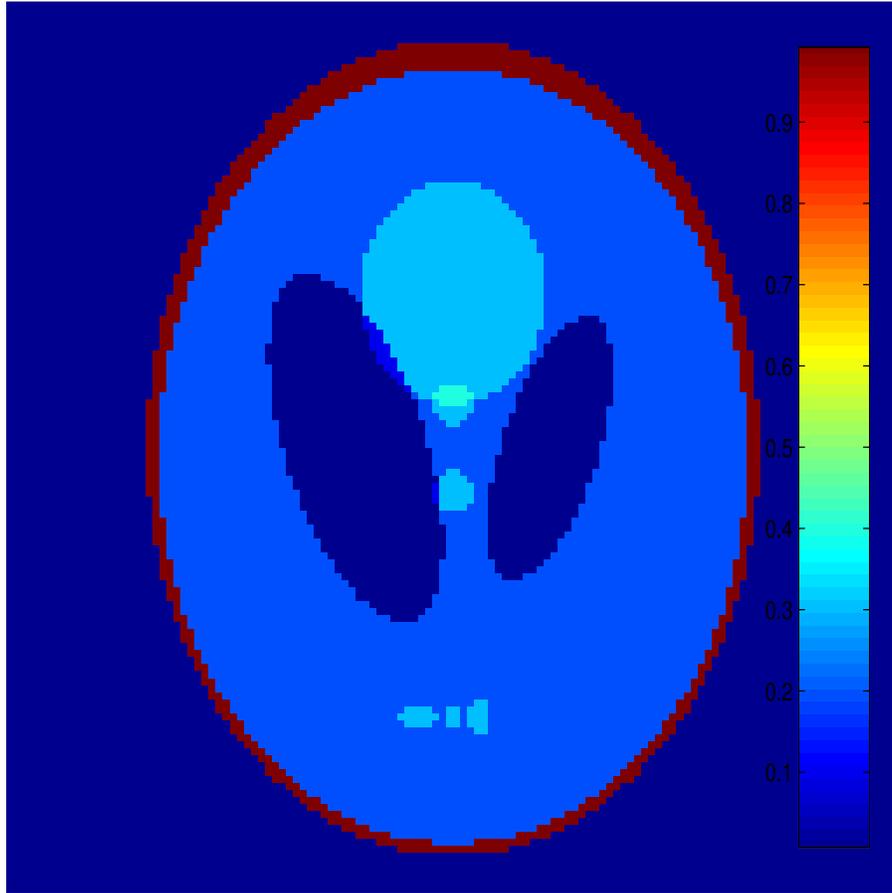


Figura 4.1: Espectro de Shepp-Logan.

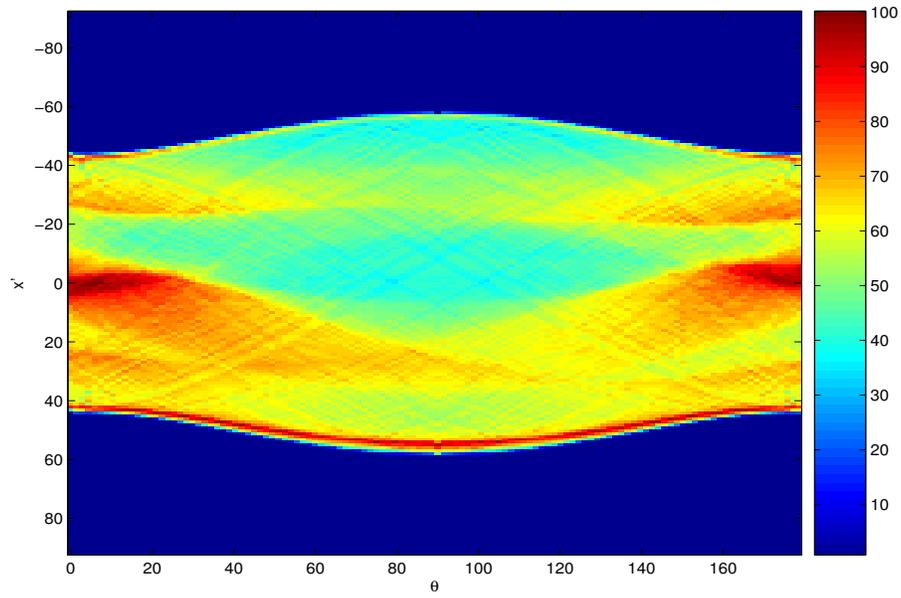
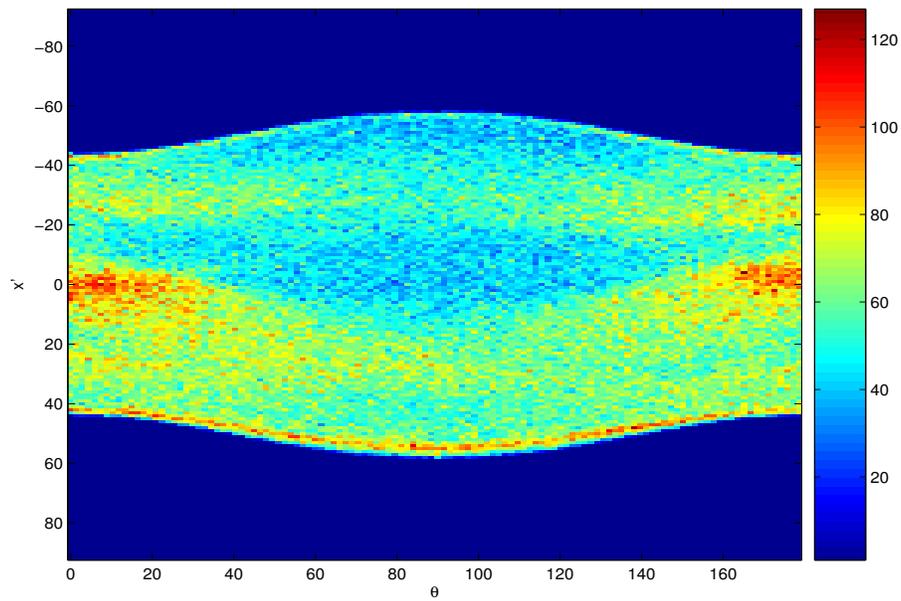


Figura 4.2: Transformada sem Ruído.

Figura 4.3: Transformada com Ruído — $\sum_{i=1}^m b_i = 715.863$.

número igual de vistas². No primeiro caso, os subconjuntos, que foram chamados de “subconjuntos sequenciais”, compunham-se das vistas $\{\frac{m}{N}(l-1)+1, \frac{m}{N}(l-1)+2, \dots, \frac{m}{N}l\}$, onde l é o número do subconjunto. A segunda forma de compor os subconjuntos, chamada de “subconjuntos equilibrados”, tinha cada subconjunto formado pelas vistas $\{l, l+N, l+2N, \dots, l+N(\frac{m}{N}-1)\}$.

Com relação à ordem de processamento dos subconjuntos, as duas opções utilizadas foram a “sequencial natural” $\mathcal{N}_1, \mathcal{N}_2, \dots, \mathcal{N}_N$ e a “sequência perpendicular”, descrita no trabalho clássico de Herman e Meyer [HeM93]. Combinadas com as duas formas de escolher os subconjuntos, estas ordenações nos dão quatro possibilidades para testarmos o comportamento de OS-EM. A fim de dar uma noção mais pictória do significado destas escolhas, as figuras 4.4 até 4.7 mostram a duas primeiras subiterações de OS-EM para cada uma das possíveis combinações de ordenação/composição dos subconjuntos. Foram utilizados 24 subconjuntos, cada um composto de 5 vistas. A imagem utilizada para iniciar o algoritmo era uniforme, o que permite depreender, nestas primeiras subiterações, uma boa noção da formação dos subconjuntos.

É importante comparar o desempenho do algoritmo utilizando cada uma destas possibilidades. As figuras de 4.8 até 4.10 mostram a evolução da qualidade da imagem obtida com as diferentes escolhas para os parâmetros em diversos graus de incrementalidade. Podemos notar que quanto maior o número de subconjuntos, mais crítica é a importância da ordenação e que os subconjuntos sequenciais sofrem uma influência grande do modo como são ordenados. Para ambas composições dos subconjuntos, a sequência perpendicular sai em vantagem. Além disso, as melhores imagens foram obtidas utilizando os subconjuntos sequenciais, indicando que esta escolha seria a mais apropriada, quando combinada com a sequência perpendicular.

²Uma vista é a coleção de todas as projeções paralelas entre si e cada uma delas pode, portanto, ser determinada pelo ângulo entre uma das projeções que a compõem e um eixo de referência. Em PET, o ângulo entre vistas consecutivas é constante, de forma que, se temos q vistas e considerarmos que o eixo de referência pertence à primeira vista, a l -ésima vista seria aquela com ângulo de $\frac{l-1}{q}180^\circ$ em relação a tal eixo.

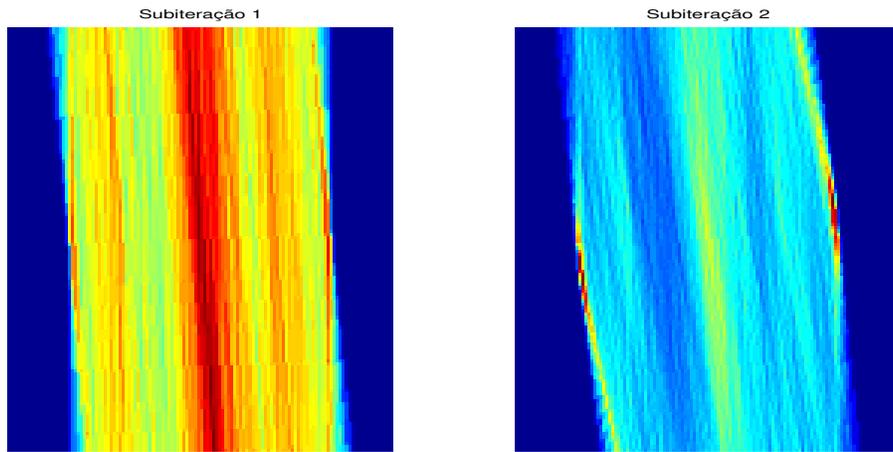


Figura 4.4: OS-EM — Subconjuntos Sequenciais — Sequência Natural.

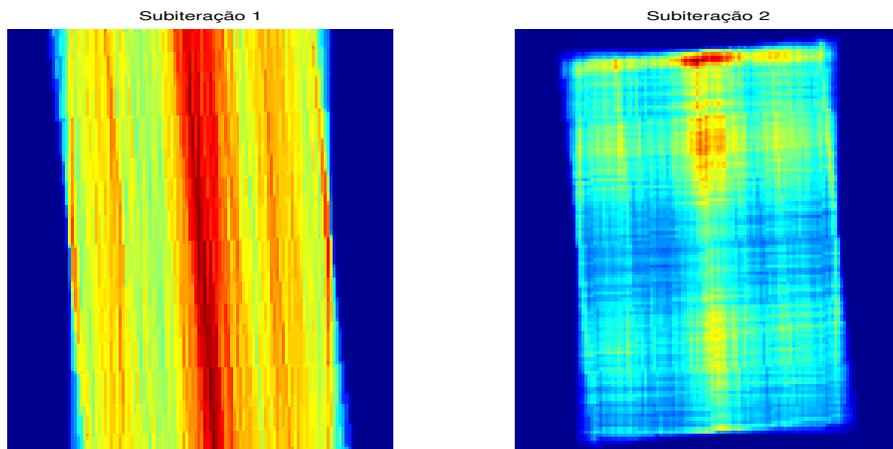


Figura 4.5: OS-EM — Subconjuntos Sequenciais — Sequência Perpendicular.

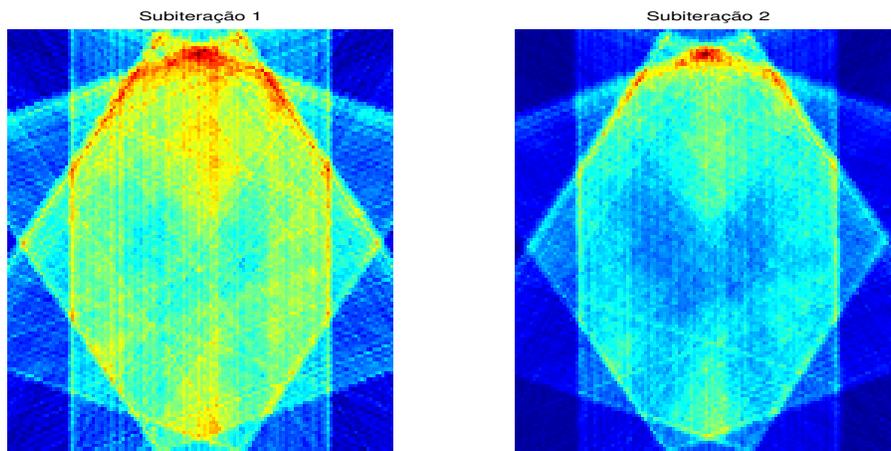


Figura 4.6: OS-EM — Subconjuntos Equilibrados — Sequência Natural.

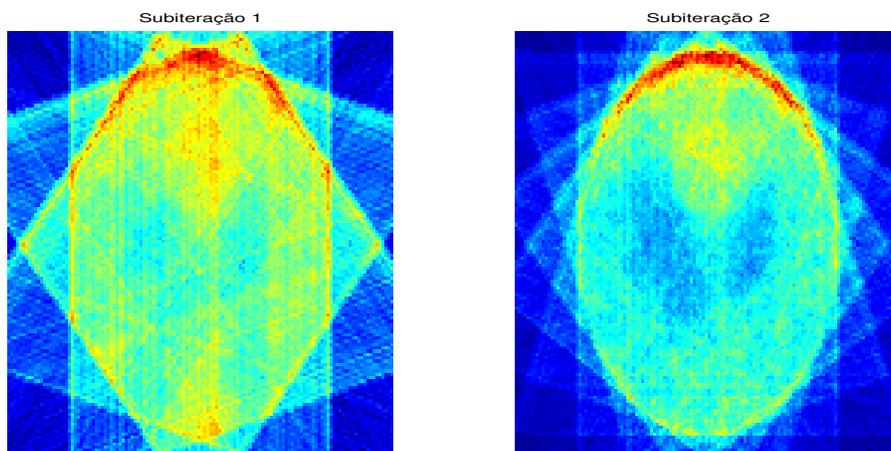


Figura 4.7: OS-EM — Subconjuntos Equilibrados — Sequência Perpendicular.

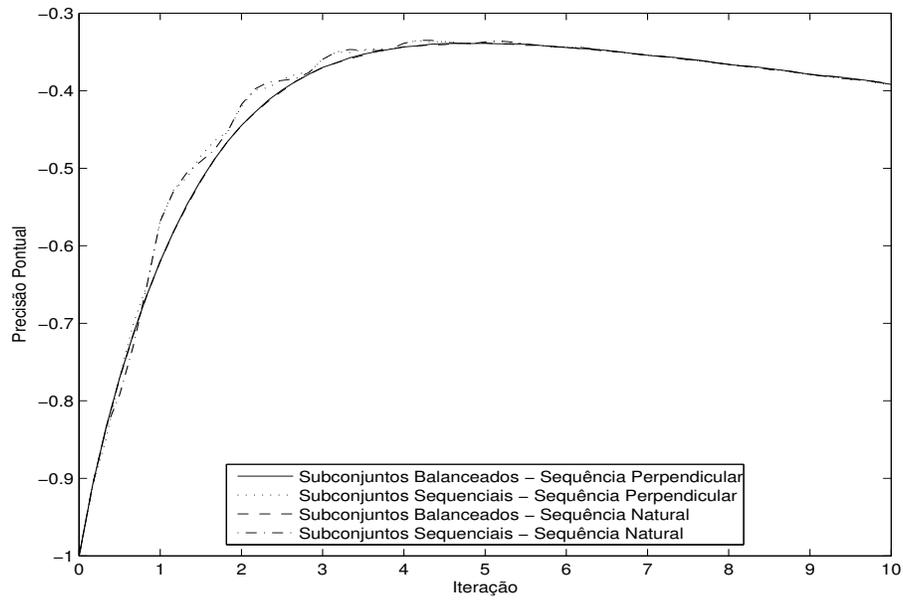


Figura 4.8: OS-EM — 6 Subconjuntos.

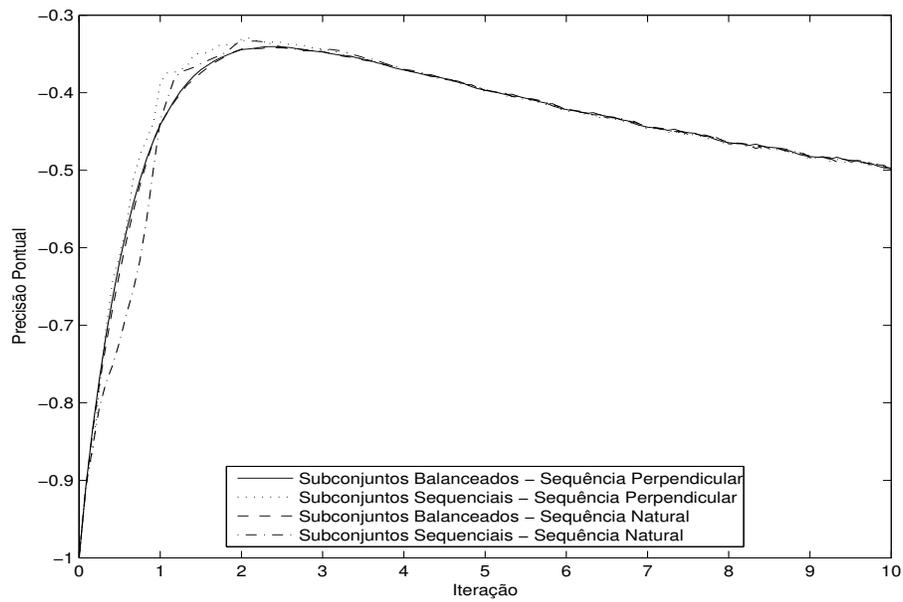


Figura 4.9: OS-EM — 12 Subconjuntos.

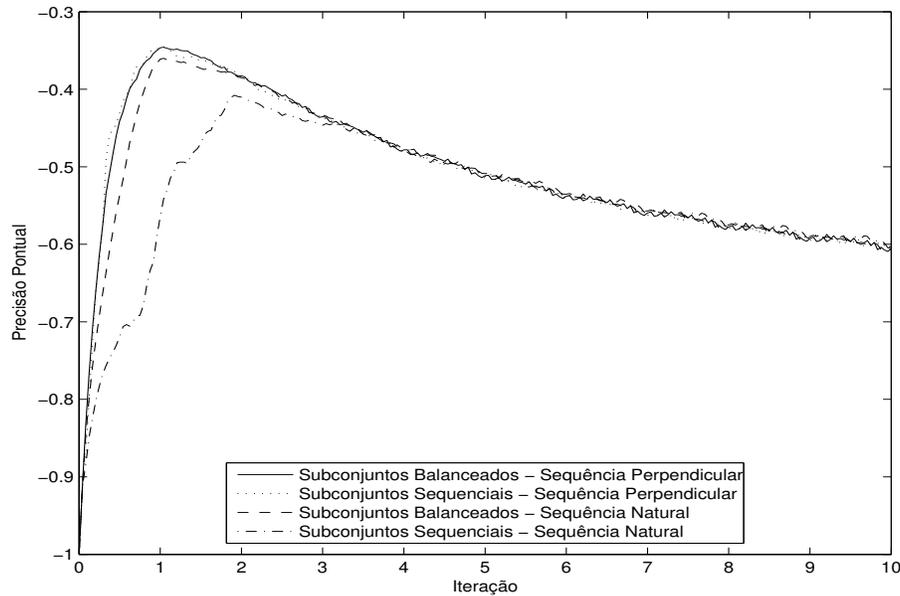


Figura 4.10: OS-EM — 24 Subconjuntos.

4.4 Parâmetros de Relaxação

Uma regra simples para os parâmetros de relaxação que atende as condições assintóticas para convergência utilizadas nas demonstrações do próximo capítulo é $\lambda_k = \frac{\lambda_0}{\alpha k + 1}$, onde α é uma constante positiva. Aplicamos esta regra com $\alpha = 1$ e $\lambda_0 = 1$ a RAMLA com $p_j = \frac{\sum_{i=1}^m a_{ij}}{N}$ e a sequência perpendicular de subconjuntos sequenciais. Nos experimentos com RAMLA, não basta informar que usamos uma solução inicial positiva uniforme, porque RAMLA não possui a propriedade de auto-ajustar a imagem à contagem de fótons que OS-EM possui. A fim de não prejudicar o desempenho do algoritmo, a solução inicial utilizada foi a imagem homogênea cuja transformada de Radon tem uma contagem compatível com as dos dados. Outro problema que pode surgir é o aparecimento de números não positivos nas primeiras iterações. Nestes casos, basta retornar as componentes não positivas a um valor positivo pequeno. A proposição 1 do próximo capítulo garante que para k grande o suficiente este expediente deixará

de ser necessário e as demonstrações de convergência se aplicam.

Como podemos ver pelo resultado exibido na Figura 4.11, utilizando esta regra, RAMLA perde em velocidade para OS-EM quando o número de subconjuntos é pequeno, porém, no caso $N=24$, temos uma clara vantagem a favor de RAMLA. Assim, fica claro que um melhor compromisso pode ser encontrado entre o grau de incrementalidade do algoritmo e a velocidade de diminuição do tamanho de passo. Se quisermos $\alpha(N)$ uma função linear tal que $\alpha(1) = 0$ (EM Clássico) e $\alpha(24) = 1$ (situação a partir da qual RAMLA levou vantagem) teremos $\alpha(N) = \frac{N-1}{23}$. A idéia é que, ao invés de utilizarmos $\lambda_k = \frac{\lambda_0}{\alpha k + 1}$, utilizemos $\lambda_k = \frac{\lambda_0}{\alpha(N)k + 1}$, que no nosso caso se torna:

$$\lambda_k = \frac{\lambda_0}{\frac{N-1}{23}k + 1}. \quad (4.2)$$

Os resultados com esta fórmula estão exibidos nas figuras 4.12 até 4.14. Entre um e outro teste, alguns parâmetros da simulação dos dados (contagem de fótons e número de vistas) foram alterados. Como tivemos um desempenho excessivamente lento quando poucos subconjuntos foram utilizados, pareceu natural testar a alternativa $\lambda_k = \frac{\lambda_0}{\frac{N-1}{d}k + 1}$. Os resultados podem ser vistos nas figuras 4.15 até 4.17. Deve-se frisar que mesmo que alguns parâmetros tenham sido modificados, a imagem continuou com a dimensão de 128 por 128 pixels constante durante os testes. Os resultados obtidos podem não se repetir no caso de imagens de tamanho muito diferente deste testado. A idéia básica de utilizar uma fórmula do tipo $\lambda_k = \frac{\lambda_0}{\frac{N-1}{d}k + 1}$ pode, entretanto, ser mantida se o parâmetro d for corretamente dimensionado.

4.5 Interpretação dos Resultados

Como as figuras de 4.12 até 4.17 mostram, não é difícil encontrar uma sequência de parâmetros de relaxação que faça RAMLA atingir valores superiores para a precisão pontual do que os obtidos por OS-EM na maioria das iterações e esta escolha é razoavelmente robusta com relação à variação de parâmetros da simulação tais como nível de ruído e geometria do tomógrafo. Este resultado

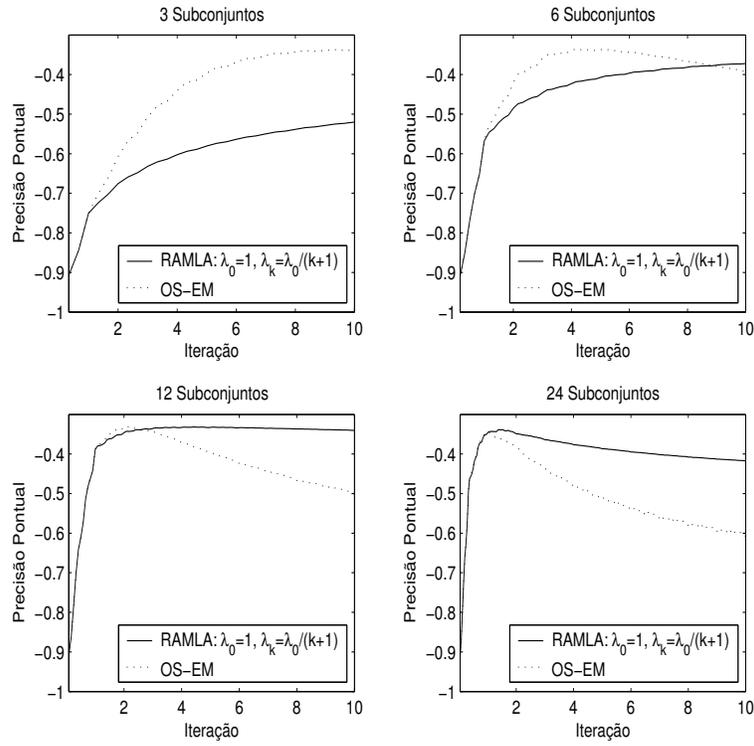


Figura 4.11: OS-EM×RAMLA — $\sum_{i=1}^m b_i = 715,813$ — 120 vistas.

é interessante por que mostra que a adoção de critérios de parada tem mais chance de obter boas imagens com RAMLA do que com OS-EM. Além disso, estas seqüências utilizadas encaixam-se com perfeição aos resultados do próximo capítulo que garantem a convergência de RAMLA.

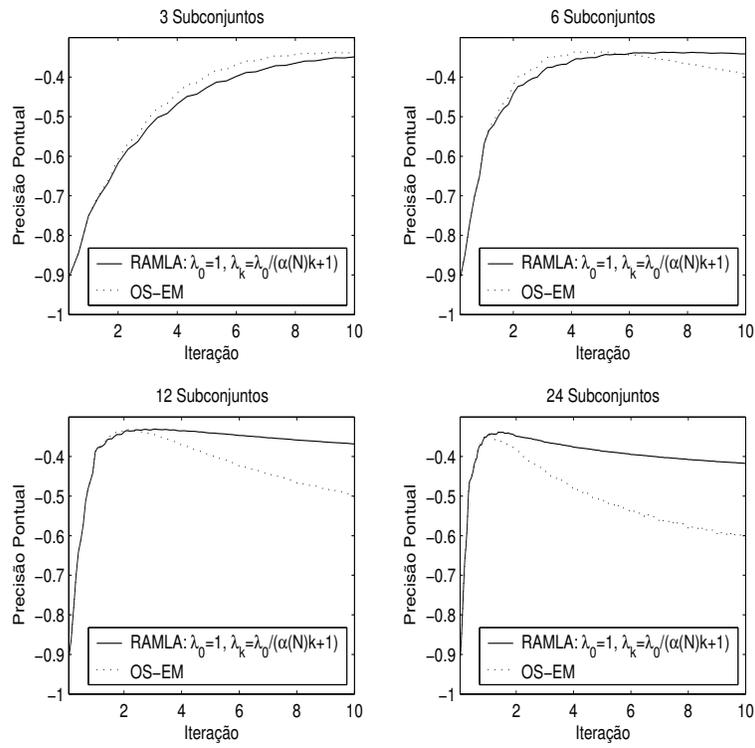


Figura 4.12: OS-EM×RAMLA — $\alpha(N) = \frac{N-1}{23} - \sum_{i=1}^m b_i = 715,813 - 120$ vistas.

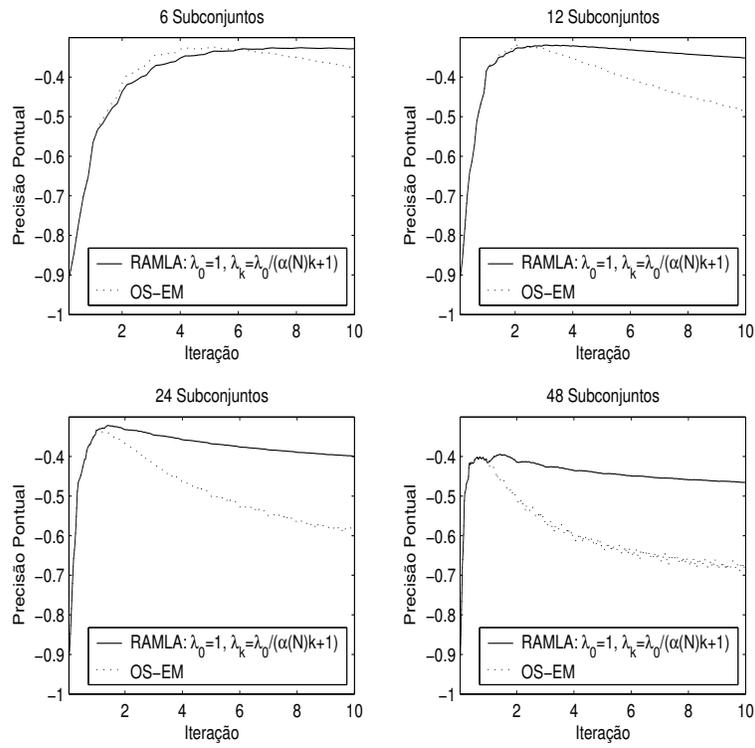


Figura 4.13: OS-EM×RAMLA — $\alpha(N) = \frac{N-1}{23} - \sum_{i=1}^m b_i = 764,713 - 384$ vistas.

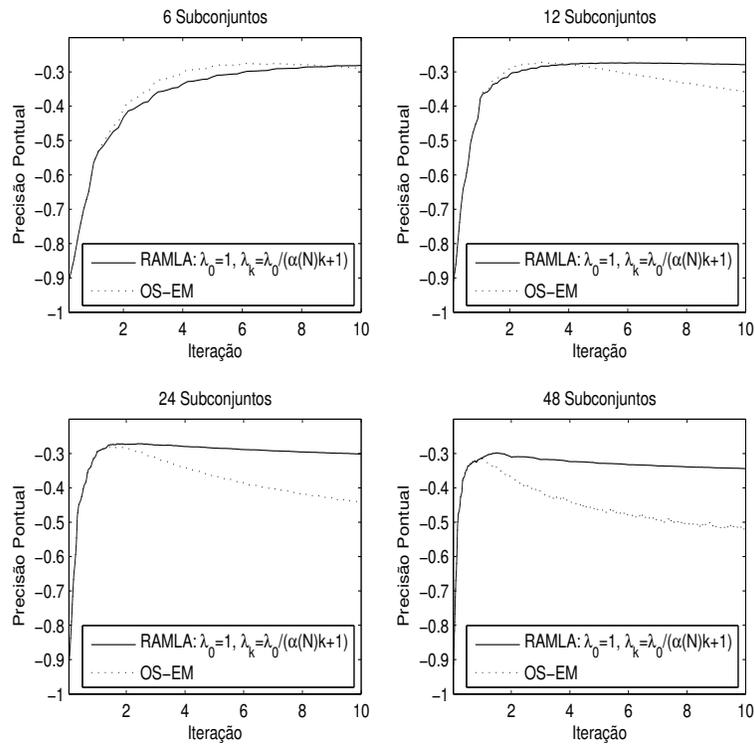


Figura 4.14: OS-EM×RAMLA — $\alpha(N) = \frac{N-1}{23} - \sum_{i=1}^m b_i = 1,528,687 - 384$ vistas.

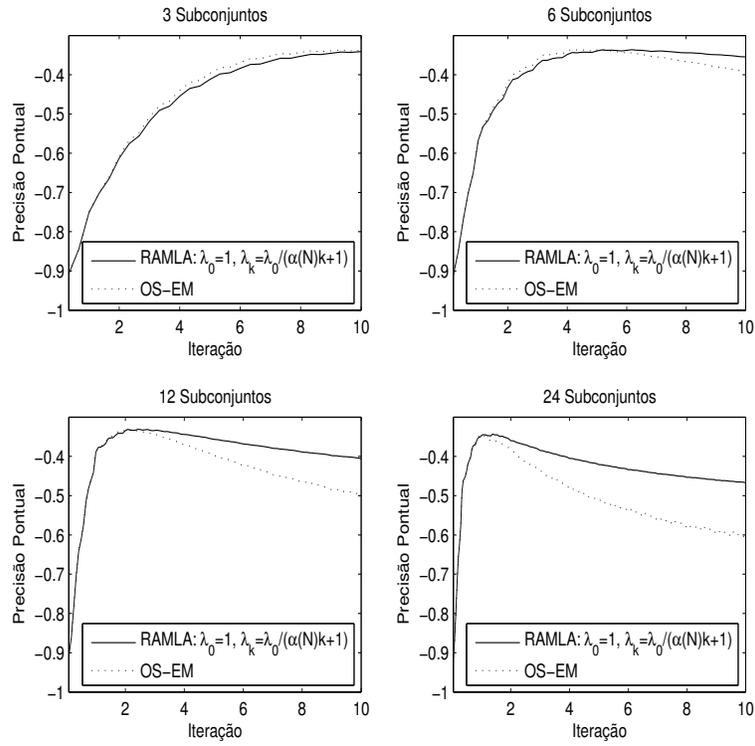


Figura 4.15: OS-EM×RAMLA — $\alpha(N) = \frac{N-1}{47} = \sum_{i=1}^m b_i = 715,813 = 120$ vistas.

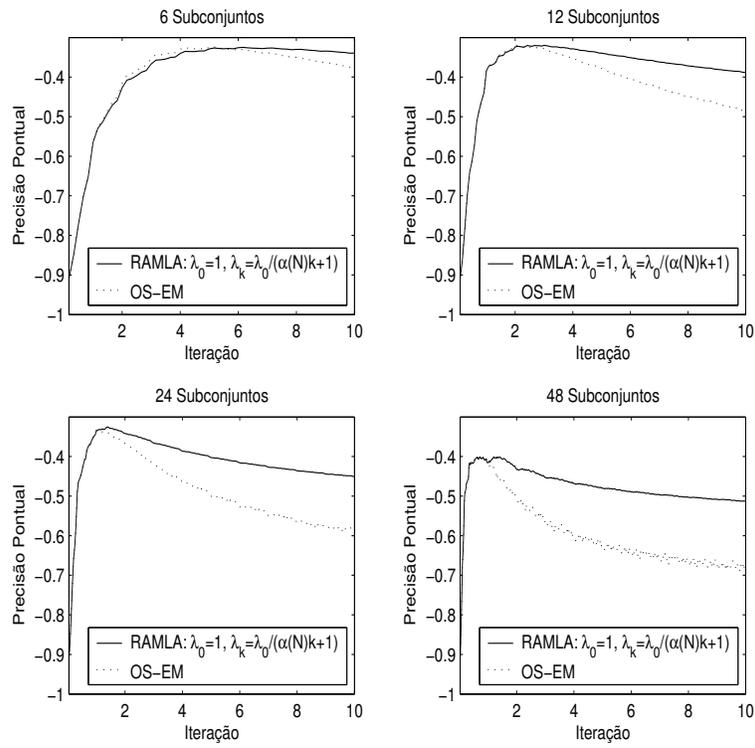


Figura 4.16: OS-EM×RAMLA — $\alpha(N) = \frac{N-1}{47} = \sum_{i=1}^m b_i = 764, 713 = 384$ vistas.

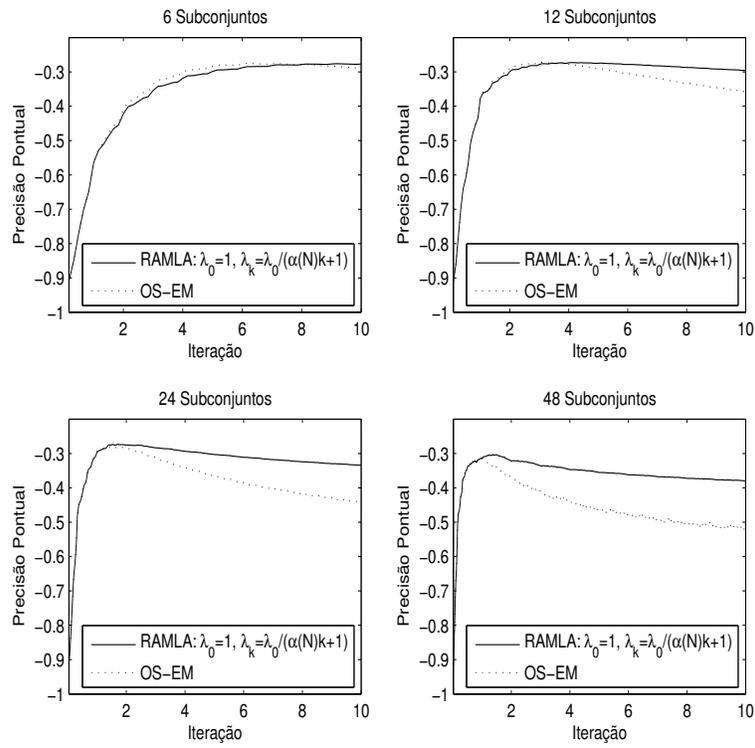


Figura 4.17: OS-EM×RAMLA — $\alpha(N) = \frac{N-1}{47} - \sum_{i=1}^m b_i = 1,528,687 - 384$ vistas.

Capítulo 5

Demonstrações

Neste capítulo, os detalhes sobre a demonstração de convergência para RAMLA (que permite que o tamanho de passo seja variável dentre as subiterações e inclui DRAMA) são explicitados. As hipóteses utilizadas são as seguintes:

- $\mathbf{y}^{(0)} \in \mathbb{R}_{++}^n$, $A \in \mathbb{R}_+^{m \times n}$, $p_j > 0$;
- A função $L(\mathbf{y})$ é estritamente côncava. Note que no caso de PET teremos $\nabla^2 L(\mathbf{y}) = -A^T W(\mathbf{y}) A$, onde $W(\mathbf{y}) = \text{diag} \left\{ \frac{b_i}{(A\mathbf{y})_i^2} \right\}$. Desta forma, vemos que $\mathbf{x}^T \nabla^2 L(\mathbf{y}) \mathbf{x} = - \left\| W(\mathbf{y})^{\frac{1}{2}} A \mathbf{x} \right\|_2^2$. Isto implica que $L(\mathbf{y})$ é estritamente côncava (para $m > n$) se e somente se $\text{posto}(W(\mathbf{y})^{\frac{1}{2}} A) = n$.
- $0 < \lambda_{kl} \leq \lambda$, onde λ foi escolhido de acordo com a proposição 1 a seguir e, denotando $\lambda_k = \lambda_{k1}$:

$$\sum_{k=0}^{\infty} \lambda_k = \infty, \quad \sum_{k=0}^{\infty} \lambda_k^2 < \infty, \quad |\lambda_k - \lambda_{kl}| = O(\lambda_k^2); \quad (5.1a)$$

$$\lim_{k \rightarrow \infty} \frac{\lambda_{kl}}{\lambda_k} = 1, \quad \frac{\lambda_k}{\lambda_{k+1}} < M. \quad (5.1b)$$

5.1 Regularidade

Nesta seção, são fornecidos alguns resultados acerca da regularidade da sequência gerada pelo algoritmo (3.12) quando aplicado à função objetivo $L(\mathbf{y}) = \sum_{i=1}^m \{b_i \log(A\mathbf{x})_i - (A\mathbf{x})_i\}$. O leitor deverá notar que as hipóteses

(5.1a) não são utilizadas — esta é a razão pela qual a primeira das condições em (5.1b) é considerada independente daquelas em (5.1a), mesmo sendo implicada por elas —, porém a estrutura algébrica da função objetivo é sistematicamente explorada. A hipótese $A \in \mathbb{R}_+^{m \times n}$ é utilizada, sem aviso, em quase todas as demonstrações desta seção.

Primeiro, precisamos nos preocupar com a factibilidade da sequência $\mathbf{y}^{(k)}$ gerada por (3.12). A seguinte proposição é útil:

Proposição 1 *Existe $\lambda > 0$ tal que se $0 < \lambda_{kl} \leq \lambda$ e $\mathbf{y}^{(0)} \in \mathbb{R}_{++}^n$ então $y_j^{(k,l)} > 0, \forall j, k, l$.*

Demonstração: Suponha, por indução, que $\mathbf{y}^{(k,l-1)} \in \mathbb{R}_{++}^n$:

$$\begin{aligned} y_j^{(k,l)} &= y_j^{(k,l-1)} \left(1 + \frac{\lambda_{kl}}{p_j} \sum_{i \in \mathcal{N}_i} a_{ij} \left(\frac{b_i}{(A\mathbf{y}^{(k,l-1)})_i} - 1 \right) \right) \\ &\geq y_j^{(k,l-1)} \left(1 - \frac{\bar{\lambda}_k}{p_j} \sum_{i \in \mathcal{N}_i} a_{ij} \right). \end{aligned} \quad (5.2)$$

Onde $\bar{\lambda}_k := \max_l \{\lambda_{kl}\}$. Isto significa que para obtermos $y_j^{(k,l)} > 0$ é suficiente que:

$$\bar{\lambda}_k \frac{\sum_{i \in \mathcal{N}_i} a_{ij}}{p_j} < 1, \quad \forall j, k, l \quad (5.3)$$

Assim, se $\lambda_{kl} \leq \lambda$, a desigualdade acima é satisfeita se:

$$\lambda < \min_{j,l} \frac{p_j}{\sum_{i \in \mathcal{N}_i} a_{ij}}. \quad (5.4)$$

De acordo com a hipótese $p_j > 0$ para todo j , sempre poderemos obter um valor de λ positivo que respeite a desigualdade acima. Uma vez que $\mathbf{y}^{(0)} \in \mathbb{R}_{++}^n$, a demonstração está completa. \square

Proposição 2 *Se, para dados C e I , $\eta > 0$ é um limitante inferior da sequência $(A\mathbf{y}^{(k,C)})_I$, gerada por (3.12) com λ escolhido de acordo com a proposição 1, $0 < \lambda_{kl} \leq \lambda$ e $\mathbf{y}^{(0)} \in \mathbb{R}_{++}^n$, então $(A\mathbf{y}^{(k,l)})_I, l = 1, \dots, N, k = 1, \dots$ também é limitada inferiormente por um número positivo.*

Demonstração: De fato, de acordo com as hipóteses, se fizermos:

$$\gamma_j^{(l)} = 1 - \lambda \frac{\sum_{i \in \mathcal{N}_l} a_{ij}}{p_j}, \quad \gamma = \min_{j,l} \gamma_j^{(l)}. \quad (5.5)$$

Então, pela maneira como λ foi construído, temos $\gamma > 0$. Além disso, aplicando (5.2) temos $\mathbf{y}^{(k,l+1)} \geq \gamma \mathbf{y}^{(k,l)}$. Isto leva ao conjunto de desigualdades:

$$\begin{aligned} \mathbf{y}^{(k,C+i)} &\geq \gamma^i \mathbf{y}^{(k,C)}, \quad i = 1, \dots, N - C; \\ \mathbf{y}^{(k+1,i)} &\geq \gamma^{N-C+i} \mathbf{y}^{(k,C)}, \quad i = 1, \dots, C - 1. \end{aligned} \quad (5.6)$$

Seja $\mu = \gamma^{N-1}$. Então:

$$\begin{aligned} \mathbf{y}^{(k,C+i)} &\geq \mu \mathbf{y}^{(k,C)}, \quad i = 1, \dots, N - C; \\ \mathbf{y}^{(k+1,i)} &\geq \mu \mathbf{y}^{(k,C)}, \quad i = 1, \dots, C - 1. \end{aligned} \quad (5.7)$$

O que implica:

$$\begin{aligned} (A\mathbf{y}^{(k,C+i)})_I &\geq \mu (A\mathbf{y}^{(k,C)})_I \geq \mu \eta > 0, \quad i = 1, \dots, N - C; \\ (A\mathbf{y}^{(k+1,i)})_I &\geq \mu (A\mathbf{y}^{(k,C)})_I \geq \mu \eta > 0, \quad i = 1, \dots, C - 1. \end{aligned} \quad (5.8)$$

Isso finaliza a demonstração porque as desigualdades acima valem para todo k . \square

Proposição 3 *Com λ conforme encontrado na proposição 1, $0 < \lambda_{kl} \leq \lambda$, $\frac{\lambda_k}{\lambda_{k+1}} < M$, $\lim_{k \rightarrow \infty} \frac{\lambda_{kl}}{\lambda_k} = 1$ e $\mathbf{y}^{(0)} \in \mathbb{R}_{++}^n$, a sequência gerada por (3.12) é tal que $(A\mathbf{y}^{(k,l)})_I$ é limitada inferiormente por um valor maior que zero sempre que $b_I > 0$.*

Demonstração: Suponha que $b_I > 0$ e seja \mathcal{N}_C o subconjunto que contenha I . Podemos descrever a iteração (3.12) como:

$$\begin{aligned} y_j^{(k,C)} &= y_j^{(k-1,C)} + \frac{1}{p_j} \sum_{l=C+1}^N \left\{ \lambda_{k-1l} y_j^{(k-1,l-1)} \sum_{i \in \mathcal{N}_l} a_{ij} \left(\frac{b_i}{(A\mathbf{y}^{(k-1,l-1)})_i} - 1 \right) \right\} \\ &\quad + \frac{1}{p_j} \sum_{l=1}^C \left\{ \lambda_{kl} y_j^{(k,l-1)} \sum_{i \in \mathcal{N}_l} a_{ij} \left(\frac{b_i}{(A\mathbf{y}^{(k,l-1)})_i} - 1 \right) \right\}. \end{aligned} \quad (5.9)$$

De onde teremos:

$$y_j^{(k,C)} \geq y_j^{(k-1,C)} + \frac{\lambda_k}{p_j} \left\{ -\frac{\lambda_{k-1}}{\lambda_k} \sum_{l=C+1}^N \frac{\lambda_{k-1l}}{\lambda_{k-1}} y_j^{(k-1,l-1)} \sum_{i \in \mathcal{N}_l} a_{ij} \right. \\ \left. - \sum_{l=1}^C \frac{\lambda_{kl}}{\lambda_k} y_j^{(k,l-1)} \sum_{\substack{i \in \mathcal{N}_l \\ i \neq I}} a_{ij} + \frac{\lambda_{kC}}{\lambda_k} y_j^{(k,C-1)} a_{Ij} \left(\frac{b_I}{(\mathbf{A}\mathbf{y}^{(k,C-1)})_I} - 1 \right) \right\}. \quad (5.10)$$

Como $\lim_{k \rightarrow \infty} \frac{\lambda_{kl}}{\lambda_k} = 1$, podemos escolher P tal que $k \geq P - 1$ implica $\frac{\lambda_{kl}}{\lambda_k} > \frac{1}{2}$ e $\frac{\lambda_{kl}}{\lambda_k} < \frac{3}{2}$. Agora, seja γ como definido em (5.5), então $\mathbf{y}^{(k,l)} \geq \gamma \mathbf{y}^{(k,l-1)} \Rightarrow \mathbf{y}^{(k,l-1)} \leq \frac{\mathbf{y}^{(k,l)}}{\gamma}$. Assim, como $\frac{\lambda_k}{\lambda_{k+1}} < M$, se $k \geq P$ e $\left(\frac{b_I}{(\mathbf{A}\mathbf{y}^{(k,C-1)})_I} - 1 \right) > 0$ e, se denotarmos $B = \max_{l,j} \sum_{i \in \mathcal{N}_l} a_{ij}$:

$$y_j^{(k,C)} \geq y_j^{(k-1,C)} + \frac{\lambda_k}{p_j} \left\{ -K y_j^{(k,C-1)} + \frac{1}{2} y_j^{(k,C-1)} a_{Ij} \left(\frac{b_I}{(\mathbf{A}\mathbf{y}^{(k,C-1)})_I} - 1 \right) \right\}. \quad (5.11)$$

Onde $K = \frac{3}{2} M B \sum_{l=1}^{N-1} \frac{1}{\gamma^l}$. Neste ponto, restringimos a nossa atenção para o conjunto de índices $J_I = \{j | a_{Ij} > 0\}$. Seja ϵ_I tal que $a_{Ij} \left(\frac{b_I}{\epsilon_I} - 1 \right) > 2K$, $\forall j \in J_I$. Agora analisamos os possíveis casos:

1. Caso $(\mathbf{A}\mathbf{y}^{(k,C-1)})_I \geq \epsilon_I$ então $(\mathbf{A}\mathbf{y}^{(k,C)})_I \geq \gamma \epsilon_I$;
2. A única situação restante é $(\mathbf{A}\mathbf{y}^{(k,C-1)})_I < \epsilon_I$. Entretanto, isto implica $y_j^{(k,C)} > y_j^{(k-1,C)}$, $\forall j \in J_I$ devido a (5.11) e a maneira como ϵ_I foi construído. Isto claramente significa $(\mathbf{A}\mathbf{y}^{(k,C)})_I > (\mathbf{A}\mathbf{y}^{(k-1,C)})_I$. Este limitante pode ser, então, utilizado para a aplicação recursiva desta análise: se não encontrarmos limite inferior caindo no caso 1 teremos $(\mathbf{A}\mathbf{y}^{(k,C)})_I > (\mathbf{A}\mathbf{y}^{(P,C)})_I \geq \gamma^C (\mathbf{A}\mathbf{y}^{(P)})_I$.

A argumentação contida nos itens acima mostra que se denotarmos $\mu < \min_{k=1, \dots, P} \{\gamma \epsilon_I, \gamma^C (\mathbf{A}\mathbf{y}^{(k)})_I\}$, então $(\mathbf{A}\mathbf{y}^{(k,C)})_I > \mu$, $\forall k$. Uma vez que $\epsilon_I > 0$ e $(\mathbf{A}\mathbf{y}^{(k)})_I > 0$ para $k = 1, \dots, P$, então tal μ pode ser escolhido maior que zero e, assim, utilizando a proposição 2 o resultado está demonstrado. \square

Corolário 1 *Sob as hipóteses da proposição 3, $\exists \delta > 0$ tal que se $\mathbf{y} \in F_\delta$, onde $F_\delta := \{\mathbf{y} | \min_{\mathbf{x} \in \overline{\text{conv}}\{\mathbf{y}^{(k,l)}\}} \|\mathbf{y} - \mathbf{x}\| < \delta\}$, então as derivadas de qualquer ordem*

de $L_i(\mathbf{y})$, $i = 1, \dots, N$ existem e são limitadas em F_δ (como consequência todas são Lipschitz no conjunto F_δ).

Demonstração: Suponha que $\mathbf{y} \in \overline{\text{conv}}\{\mathbf{y}^{(k,l)}\}$. Isto significa que ou $\mathbf{y} \in \text{conv}\{\mathbf{y}^{(k,l)}\}$ ou \mathbf{y} é o limite de uma sequência de pontos de $\text{conv}\{\mathbf{y}^{(k,l)}\}$. Na primeira hipótese, sabemos pela proposição 3, que se $b_I > 0$ então existe $\xi_I > 0$ tal que:

$$(\mathbf{A}\mathbf{y})_I = \alpha(\mathbf{A}\mathbf{y}^{(k,l)})_I + (1 - \alpha)(\mathbf{A}\mathbf{y}^{(s,l)})_I > \xi_I.$$

Isto também implica que se \mathbf{y} for um ponto de acumulação de $\text{conv}\{\mathbf{y}^{(k,l)}\}$ e $b_I > 0$, então $(\mathbf{A}\mathbf{y})_I \geq \xi_I$. Logo, se $b_I > 0$ e $\mathbf{y} \in \overline{\text{conv}}\{\mathbf{y}^{(k,l)}\}$ então $(\mathbf{A}\mathbf{y})_I \geq \xi_I$. Agora, seja $B_\delta(\mathbf{y}) := \{\mathbf{x} \mid \|\mathbf{x} - \mathbf{y}\| < \delta\}$ e faça $\delta < \frac{\xi_I}{2\|\mathbf{A}\|}$ então $\mathbf{x} \in B_\delta(\mathbf{y}) \Rightarrow \|\mathbf{y} - \mathbf{x}\| < \frac{\xi_I}{2\|\mathbf{A}\|}$. Por outro lado, $|(A\mathbf{x})_I - (\mathbf{A}\mathbf{y})_I| \leq \|\mathbf{A}(\mathbf{x} - \mathbf{y})\| \leq \|\mathbf{A}\|\|\mathbf{y} - \mathbf{x}\| < \frac{\xi_I}{2}$. Logo, $\mathbf{x} \in B_\delta(\mathbf{y}) \Rightarrow (A\mathbf{x})_I > \frac{\xi_I}{2}$. Isto finaliza a demonstração porque as derivadas de $L_l(\mathbf{y})$ são da forma $\sum_{i \in \mathcal{N}_l | b_i > 0} \beta_i (-1)^{n+1} \frac{b_i}{(\mathbf{A}\mathbf{y})_i^n}$, para algum $\beta_i \geq 0$ (exceto pelas de primeira ordem, que tem adição de um termo constante, o que é indiferente).□

Nota: Este corolário tem, como consequência óbvia, a continuidade de Lipschitz de $D(\mathbf{y})\nabla L_l(\mathbf{y})$ em F_δ .

Proposição 4 A sequência $\mathbf{y}^{(k)}$, gerada por (3.12), sob as hipóteses da proposição 3 é limitada.

Demonstração: É suficiente mostrar a limitação para cada uma das sequências $y_J^{(k,l)}$. Primeiro, notamos que, como consequência da primeira igualdade em (5.2) e da proposição 3 acima, $\exists \Gamma > 0$ tal que $y_J^{(k,l)} \leq \Gamma y_J^{(k,l-1)} \forall k, l$. Seja agora $M > 0$ tal que $\frac{b_i}{a_{iJ}M} < 1 \forall i | a_{iJ} > 0$. Observando novamente a primeira igualdade em (5.2) vemos que se $y_J^{(k,l-1)} \geq M$ então $y_J^{(k,l)} < y_J^{(k,l-1)}$. Unindo os dois raciocínios podemos concluir que um limitante superior para a sequência $y_J^{(k,l)}$ é $\max\{\Gamma M, y_J^{(0)}\}$.□

5.2 Convergência

Esta seção utiliza algumas condições de regularidade sobre a sequência $\mathbf{y}^{(k,l)}$ para fornecer demonstrações de convergência para (3.12) sem especificar a função $L(\mathbf{y})$, exceto pela sua convexidade estrita.

Proposição 5 *Se $\{\mathbf{y}^{(k,l)}\} \subset \mathbb{R}_{++}^n$ é uma sequência limitada gerada por (3.12), $L_i(\mathbf{y})$, $\nabla L_i(\mathbf{y})$, $D(\mathbf{y})\nabla L_i(\mathbf{y})$, $i = 1, \dots, N$ e $\nabla^2 L(\mathbf{y})$ são limitados e Lipschitz em uma vizinhança F em torno de $\overline{\text{conv}}\{\mathbf{y}^{(k,l)}\}$ e valem as condições (5.1a), então $L(\mathbf{y}^{(k)})$ converge.*

Demonstração: Começamos avaliando a magnitude da seguinte diferença:

$$\epsilon^{(k)} := \sum_{l=1}^q \lambda_{kl} D(\mathbf{y}^{(k,l-1)}) \nabla L_l(\mathbf{y}^{(k,l-1)}) - \lambda_k D(\mathbf{y}^{(k)}) \nabla L(\mathbf{y}^{(k)}). \quad (5.12)$$

$$\begin{aligned} \|\epsilon^{(k)}\| &= \left\| \lambda_k D(\mathbf{y}^{(k)}) \nabla L(\mathbf{y}^{(k)}) - \sum_{l=1}^N \lambda_{kl} D(\mathbf{y}^{(k,l-1)}) \nabla L_l(\mathbf{y}^{(k,l-1)}) \right\| \\ &\leq \lambda_k \left\| \sum_{l=1}^N \left\{ D(\mathbf{y}^{(k)}) \nabla L_l(\mathbf{y}^{(k)}) - D(\mathbf{y}^{(k,l-1)}) \nabla L_l(\mathbf{y}^{(k,l-1)}) \right\} \right\| \\ &\quad + \left\| \sum_{l=1}^N D(\mathbf{y}^{(k,l-1)}) \nabla L_l(\mathbf{y}^{(k,l-1)}) (\lambda_k - \lambda_{kl}) \right\| \\ &\leq \lambda_k M_1 \sum_{l=1}^N \|\mathbf{y}^{(k)} - \mathbf{y}^{(k,l-1)}\| + M_2 \sum_{l=1}^N |\lambda_k - \lambda_{kl}|. \end{aligned} \quad (5.13)$$

Onde M_1 é a constante de Lipschitz e M_2 é um limitante superior para $\|D(\mathbf{y}^{(k,l)})\nabla L(\mathbf{y}^{(k,l)})\|$. Agora note que $|\lambda_{kl} - \lambda_k| = O(\lambda_k^2) \Rightarrow \left| \frac{\lambda_{kl}}{\lambda_k} - 1 \right| = O(\lambda_k) \Rightarrow \lim_{k \rightarrow \infty} \frac{\lambda_{kl}}{\lambda_k} = 1$. Assim, se k for grande tal que $\frac{\lambda_{kl}}{\lambda_k} < \frac{4}{3}$ temos $\lambda_{kl} < \frac{4}{3}\lambda_k$ e assim:

$$\begin{aligned} \|\mathbf{y}^{(k)} - \mathbf{y}^{(k,l-1)}\| &= \left\| \sum_{c=1}^{l-1} \mathbf{y}^{(k,c-1)} - \mathbf{y}^{(k,c)} \right\| \leq \sum_{c=1}^{l-1} \|\mathbf{y}^{(k,c-1)} - \mathbf{y}^{(k,c)}\| \\ &\leq M_2 \sum_{c=1}^{l-1} \lambda_{kc} \leq \frac{4}{3}(l-1)M_2\lambda_k \leq \frac{4}{3}NM_2\lambda_k. \end{aligned} \quad (5.14)$$

Unindo isto com (5.13), podemos sintetizar:

$$\|\epsilon^{(k)}\| \leq M \left\{ \lambda_k^2 + \sum_{l=1}^N |\lambda_k - \lambda_{kl}| \right\}. \quad (5.15)$$

Iterando na fórmula (3.12), teremos:

$$\begin{aligned} \mathbf{y}^{(k+1)} &= \mathbf{y}^{(k)} + \sum_{l=1}^N \lambda_{kl} D(\mathbf{y}^{(k,l-1)}) \nabla L_l(\mathbf{y}^{(k,l-1)}) \\ &= \mathbf{y}^{(k)} + \lambda_k D(\mathbf{y}^{(k)}) \nabla L(\mathbf{y}^{(k)}) + \epsilon^{(k)}. \end{aligned} \quad (5.16)$$

Com isto, como $\mathbf{y}^{(k+1)}$ e $\mathbf{y}^{(k)}$ pertencem a F , poderemos utilizar a fórmula de Taylor de segunda ordem (veja, por exemplo, o teorema 9.4 em [Apo69]):

$$\begin{aligned} L(\mathbf{y}^{(k+1)}) &= L(\mathbf{y}^{(k)}) + \nabla L(\mathbf{y}^{(k)})^T (\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)}) \\ &\quad + \frac{1}{2} (\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)})^T \nabla^2 L(\mathbf{y}^{(k)} + \xi(\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)})) (\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)}). \end{aligned} \quad (5.17)$$

Onde $\xi \in [0, 1]$. Como $\mathbf{y}^{(k)} + \xi(\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)}) \in F$, temos, graças à hipótese que assegura a limitação de $\nabla^2 L(\mathbf{y})$ em F :

$$|L(\mathbf{y}^{(k+1)}) - L(\mathbf{y}^{(k)})| \leq |\nabla L(\mathbf{y}^{(k)})^T (\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)})| + K \|\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)}\|^2. \quad (5.18)$$

Note que (5.14) implica que $\|\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)}\| \leq V \lambda_k$ para algum $V > 0$ grande o suficiente. Logo, teremos, com a ajuda de (5.16):

$$\begin{aligned} |L(\mathbf{y}^{(k+1)}) - L(\mathbf{y}^{(k)})| &\leq |\nabla L(\mathbf{y}^{(k)})^T (\mathbf{y}^{(k+1)} - \mathbf{y}^{(k)})| + KV^2 \lambda_k^2 \\ &= |\nabla L(\mathbf{y}^{(k)})^T (\lambda_k D(\mathbf{y}^{(k)}) \nabla L(\mathbf{y}^{(k)}) + \epsilon^{(k)})| + KV^2 \lambda_k^2 \\ &\leq \lambda_k \nabla L(\mathbf{y}^{(k)})^T D(\mathbf{y}^{(k)}) \nabla L(\mathbf{y}^{(k)}) \\ &\quad + U \left\{ \lambda_k^2 + \sum_{l=1}^q |\lambda_k - \lambda_{kl}| \right\}. \end{aligned} \quad (5.19)$$

A última desigualdade foi obtida com o auxílio de (5.15) e a hipótese de que $\nabla L(\mathbf{y}^{(k)})$ é limitado. Continuemos com o cálculo:

$$\begin{aligned}
|L(\mathbf{y}^{(k+I)}) - L(\mathbf{y}^{(k)})| &= \left| \sum_{j=0}^{I-1} \{L(\mathbf{y}^{(k+j+1)}) - L(\mathbf{y}^{(k+j)})\} \right| \\
&\leq \sum_{j=0}^{I-1} |L(\mathbf{y}^{(k+j+1)}) - L(\mathbf{y}^{(k+j)})| \\
&\leq \sum_{j=0}^{I-1} \lambda_{k+j} \nabla L(\mathbf{y}^{(k+j)})^T D(\mathbf{y}^{(k+j)}) \nabla L(\mathbf{y}^{(k+j)}) \\
&\quad + U \left\{ \sum_{j=0}^{I-1} \lambda_{k+j}^2 + \sum_{j=0}^{I-1} \sum_{l=1}^N |\lambda_{k+j} - \lambda_{k+j+l}| \right\}.
\end{aligned} \tag{5.20}$$

O leitor pode notar que, em vista das hipóteses (5.1a), se pudéssemos mostrar que $\sum_{k=0}^{\infty} \lambda_k \nabla L(\mathbf{y}^{(k)})^T D(\mathbf{y}^{(k)}) \nabla L(\mathbf{y}^{(k)}) < \infty$, não seria difícil ver que $L(\mathbf{y}^{(k)})$ seria uma sequência de Cauchy e, portanto, convergente. Supondo que isto não seja verdade e utilizando um raciocínio semelhante ao usado acima, obtemos uma contradição:

$$\begin{aligned}
L(\mathbf{y}^{(k+I)}) - L(\mathbf{y}^{(k)}) &= \sum_{j=0}^{I-1} \{L(\mathbf{y}^{(k+j+1)}) - L(\mathbf{y}^{(k+j)})\} \\
&\geq \sum_{j=0}^{I-1} \lambda_{k+j} \nabla L(\mathbf{y}^{(k+j)})^T D(\mathbf{y}^{(k+j)}) \nabla L(\mathbf{y}^{(k+j)}) \\
&\quad - U \left\{ \sum_{j=0}^{I-1} \lambda_{k+j}^2 + \sum_{j=0}^{I-1} \sum_{l=1}^N |\lambda_{k+j} - \lambda_{k+j+l}| \right\}.
\end{aligned} \tag{5.21}$$

Assim, se $\sum_{k=0}^{\infty} \lambda_k \nabla L(\mathbf{y}^{(k)})^T D(\mathbf{y}^{(k)}) \nabla L(\mathbf{y}^{(k)}) = \infty$ teríamos $L(\mathbf{y}^{(k)}) \rightarrow \infty$, tendo em vista (5.1a). Isto é uma contradição com a limitação de $L_i(\mathbf{y})$. \square

Proposição 6 *Sob as mesmas hipóteses da proposição 5, existe um ponto de acumulação $\mathbf{y}^{(h_k)} \rightarrow \mathbf{y}^*$ tal que $\|D(\mathbf{y}^*) \nabla L(\mathbf{y}^*)\| = 0$.*

Demonstração: Suponhamos que não existe uma subsequência $\mathbf{y}^{(h_k)}$ tal que $\lim_{k \rightarrow \infty} \nabla L(\mathbf{y}^{(h_k)})^T D(\mathbf{y}^{(h_k)}) \nabla L(\mathbf{y}^{(h_k)}) = 0$. Logo, existe $\epsilon > 0$ tal que

$\nabla L(\mathbf{y}^{(k)})^T D(\mathbf{y}^{(k)}) \nabla L(\mathbf{y}^{(k)}) > \epsilon$. Lembrando das hipóteses sobre o tamanho de passo, (5.21) implicaria que $L(\mathbf{y}^{(k)}) \rightarrow \infty$, contradizendo a proposição 5. Assim, temos uma subsequência $\mathbf{y}^{(h_k)} \rightarrow \mathbf{y}^*$, onde $\nabla L(\mathbf{y}^*)^T D(\mathbf{y}^*) \nabla L(\mathbf{y}^*) = 0$. Porém, $\nabla L(\mathbf{y}^*)^T D(\mathbf{y}^*) \nabla L(\mathbf{y}^*) = 0 \Rightarrow \|D(\mathbf{y}^*) \nabla L(\mathbf{y}^*)\| = 0$. \square

Teorema 1 *Se $L(\mathbf{y})$ é estritamente côncava e valem as hipóteses da proposição 5, então $\mathbf{y}^{(k)} \rightarrow \mathbf{y}^*$, onde \mathbf{y}^* é o único maximizador global de $L(\mathbf{y})$.*

Demonstração: Note que podemos escrever (3.12) como

$$\begin{aligned} y_j^{(k,l)} &= y_j^{(k,l-1)} \left(1 + \frac{\lambda_{kl}}{p_j} \nabla_j L_l(\mathbf{y}^{(k,l-1)}) \right) \\ &= y_j^{(k,l-1)} \left(1 + \frac{\lambda_k}{p_j} \nabla_j L_l(\mathbf{y}^{(k,l-1)}) + \frac{\lambda_{kl} - \lambda_k}{p_j} \nabla_j L_l(\mathbf{y}^{(k,l-1)}) \right). \end{aligned} \quad (5.22)$$

Portanto

$$\begin{aligned} y_j^{(k+1)} &= y_j^{(k)} \prod_{l=1}^N \left(1 + \frac{\lambda_k}{p_j} \nabla_j L_l(\mathbf{y}^{(k,l-1)}) + O(\lambda_k^2) \right) \\ &= y_j^{(k)} \left(1 + \frac{\lambda_k}{p_j} \sum_{l=1}^N \nabla_j L_l(\mathbf{y}^{(k,l-1)}) + O(\lambda_k^2) \right). \end{aligned} \quad (5.23)$$

Neste ponto, a idéia é mostrar que existe uma subsequência tal que $\mathbf{y}^{(h_k)} \rightarrow \mathbf{y}^*$. Caso exista, como, pela proposição 5, $L(\mathbf{y}^{(k)})$ converge, temos $L(\mathbf{y}^{(k)}) \rightarrow L(\mathbf{y}^*)$. Por outro lado, podemos ver que isto implica em $\mathbf{y}^{(k)} \rightarrow \mathbf{y}^*$, já que \mathbf{y}^* é o único maximizador global de $L(\mathbf{y})$ e $L(\mathbf{y})$ é contínua em F .

De acordo com a proposição 6, existe uma subsequência $\mathbf{y}^{(h_k)} \rightarrow \mathbf{y}^{**}$, onde $D(\mathbf{y}^{**}) \nabla L(\mathbf{y}^{**}) = \mathbf{0}$. Suponhamos que \mathbf{y}^{**} não é o otimizador procurado. Neste caso, o conjunto $J := \{j | y_j^{**} = 0, \nabla_j L(\mathbf{y}^{**}) > 0\}$ seria não vazio. Para $j \in J$, pela continuidade de $\nabla L(\mathbf{y})$ em F , sabemos que existe uma vizinhança $B_\delta(\mathbf{y}^{**}) := \{\mathbf{y} | \|\mathbf{y} - \mathbf{y}^{**}\| < \delta\}$ em torno de \mathbf{y}^{**} , tal que $\mathbf{y} \in B_\delta(\mathbf{y}^{**}) \Rightarrow \nabla_j L(\mathbf{y}) > \Psi$, para algum $\Psi > 0$. Por outro lado, a limitação de $D(\mathbf{y}) \nabla L_l(\mathbf{y})$ em F , mais $\lambda_{kl} \rightarrow 0$ (consequência de (5.1a)), implicam que $\|\mathbf{y}^{(k,l)} - \mathbf{y}^{(k)}\| \rightarrow 0$. Então, pela continuidade de Lipschitz das derivadas $\nabla L_l(\mathbf{y})$ em F , $\sum_{l=1}^N \nabla L_l(\mathbf{y}^{(k,l-1)}) - \nabla L(\mathbf{y}^{(k)}) \rightarrow 0$ e, portanto, para k grande o suficiente teremos $\sum_{l=1}^N \nabla_j L_l(\mathbf{y}^{(h_k,l-1)}) > \frac{\Psi}{2}$. Agora, utilizando (5.23), vemos que,

para h_k grande o suficiente e $j \in J$, teremos:

$$y_j^{(h_k+1)} > y_j^{(h_k)}. \quad (5.24)$$

Suponha que h_k é grande o bastante, de forma que $\mathbf{y}^{(h_k)} \in B_\delta(\mathbf{y}^{**})$. Seja $t_k = \max\{q < h_k \mid \mathbf{y}^{(q)} \notin B_\delta(\mathbf{y}^{**})\}$ (faça $t_k = 0$ se $\mathbf{y}^{(q)} \in B_\delta(\mathbf{y}^{**})$ para todo $q < h_k$). Então, $\mathbf{y}^{(q)} \in B_\delta(\mathbf{y}^{**})$, $q = t_k + 1, \dots, h_k$ e, para $j \in J$ teremos:

$$y_j^{(h_k)} > y_j^{(h_k-1)} > \dots > y_j^{(t_k+1)}. \quad (5.25)$$

Agora, veja que, se $\lim_{k \rightarrow \infty} t_k = t$ para algum t finito, então teremos $0 = y_j^{**} = \lim_{k \rightarrow \infty} y_j^{(h_k)} \geq y_j^{(t+1)} > 0$. Esta contradição implica que $\lim_{k \rightarrow \infty} t_k = \infty$ e daí, segue de (5.25), que $\lim_{k \rightarrow \infty} y_j^{(t_k+1)} = 0$. Uma vez que $\lambda_{kl} \rightarrow 0$ e $D(\mathbf{y}^{(k,i)}) \nabla L_l(\mathbf{y}^{(k,i)})$ é limitada, temos $\lim_{k \rightarrow \infty} |y_j^{(k+1)} - y_j^{(k)}| = 0$ e assim, $\lim_{k \rightarrow \infty} y_j^{(t_k)} = 0$, $j \in J$. Já que a sequência $\mathbf{y}^{(k,l)}$ é limitada, podemos garantir que temos uma subsequência $\mathbf{y}^{(s_{t_k})} \rightarrow \mathbf{y}^{***}$ convergente. Uma vez que $\mathbf{y}^{(t_k)} \notin B_\delta(\mathbf{y}^{**})$, sabemos que $\|\mathbf{y}^{***} - \mathbf{y}^{**}\| \geq \delta$. Por outro lado, podemos notar que \mathbf{y}^{**} cumpre as condições suficientes para ser um maximizador de $L(\mathbf{y})$ sujeito a $\mathbf{y} \in R := \{\mathbf{y} \mid y_j \geq 0, j = 1, \dots, n, y_j = 0, j \in J\}$. Uma vez que $L(\mathbf{y}^{(k)})$ converge, teremos $L(\mathbf{y}^{***}) = L(\mathbf{y}^{**})$. Além disso, $\mathbf{y}^{***} \in R$ e teríamos dois pontos satisfazendo a condição para o máximo em R . Em resumo, como isto contradiz a hipótese de concavidade estrita, mostramos que temos uma subsequência $\mathbf{y}^{(h_k)}$ que converge para o otimizador global. Uma vez que a sequência $L(\mathbf{y}^{(k)})$ converge, como $L(\mathbf{y})$ é estritamente côncava e contínua em F e a sequência $\mathbf{y}^{(k)}$ é limitada, é fácil mostrar que $\mathbf{y}^{(k)} \rightarrow \mathbf{y}^*$. \square

Nota: Apesar da motivação inicial do trabalho ter induzido ao uso da hipótese $|\lambda_{kl} - \lambda_k| = O(\lambda_k^2)$ em (5.1a), as hipóteses menos restritivas $\sum_{k=0}^{\infty} |\lambda_{kl} - \lambda_k| < \infty$ e $\frac{|\lambda_{kl} - \lambda_k|}{\lambda_k} \rightarrow 0$ (que é equivalente a $\frac{\lambda_{kl}}{\lambda_k} = 1$ em (5.1b)) podem ser utilizadas. Todas as demonstrações seriam as mesmas, exceto esta última que precisaria de alterações mínimas.

5.3 Conclusões e Trabalho Futuro

Como pudemos ver pelo corolário 1, na seção 5.1, as hipóteses listadas no começo deste capítulo são suficientes para garantir, no caso da função de log-verossimilhança de PET, as condições de regularidade exigidas no teorema 1, que garante a convergência do algoritmo (3.12). Estes resultados podem ser aplicados na situação encontrada em [TaK03] a qual, até agora, não possuía uma demonstração deste tipo. É interessante notar que os resultados condensados no corolário 1 não existiam na literatura até o momento nem mesmo para RAMLA em sua versão original, tendo sido ou admitidos *a priori* (como em [BrP96], implicitamente com a hipótese de que $L(\mathbf{y}^{(k)})$ convergiria), ou obtidos modificando a função objetivo de forma a evitar a descontinuidade na fronteira do conjunto factível (como em [AhF03]). Neste sentido, a contribuição desta dissertação é dupla pois mostra como pode ser realizada a variação no tamanho de passo dentre as iterações de forma a garantir a convergência para a solução de máxima verossimilhança e como manter o bom comportamento da sequência gerada pelo algoritmo de forma não restritiva.

Além dos resultados teóricos de convergência, os experimentos realizados indicam que métodos relaxados podem oferecer imagens com uma qualidade superior à obtida com métodos não relaxados, mesmo quando utilizados com um critério de parada, já que degradam menos as reconstruções nas últimas iterações e, quando utilizados com uma estratégia razoável de diminuição do tamanho de passo, não desaceleram a convergência nas primeiras iterações.

Trabalho futuro pode consistir na utilização de RAMLA ou DRAMA na reconstrução simultânea de emissão e atenuação com dados de emissão, em aplicações na reconstrução tridimensional das imagens (em contraponto ao método de “Fourier Rebinning” ou à simples redução a fatias bidimensionais), ou em outras aplicações promissoras em que alta velocidade de convergência do algoritmo seja fundamental para viabilizar a reconstrução. Testes com outras formas de avaliar a imagem e com simulações mais realistas teriam importância para comparar o desempenho dos métodos relaxados contra os não relaxados,

já que OS-EM foi adotado em tomógrafos comerciais e, na presente dissertação, surgiram indícios fortes de que RAMLA pode superar este algoritmo. Além disso, as idéias mostradas na seção anterior poderiam ser aplicadas de forma semelhante em demonstrações da convergência de métodos para o problema regularizado cujo parâmetro de relaxação varie dentre as subiterações, se as linhas principais de [AhF03] forem seguidas no desenvolvimento do algoritmo para assegurar as condições de regularidade necessárias.

Referências Bibliográficas

- [AhF03] S. Ahn; J. A. Fessler, *Globally Convergent Image Reconstruction for Emission Tomography Using Relaxed Ordered Subsets Algorithms*, IEEE Trans. Med. Imag., 22(2003), pp. 613-626;
- [Apo69] T. M. Apostol, *Calculus*, Vol II, 2nd edition, Wiley International Edition;
- [BrP96] J. A. Browne; A. R. De Pierro, *A Row-action Alternative to the EM Algorithm for Maximizing Likelihoods in Emission Tomography*, IEEE Trans. Med. Imag., 15(1996), pp. 687-699;
- [HeM93] G. T. Herman; L. B. Meyer, *Algebraic Reconstruction Techniques Can Be Made Computationally Efficient*, IEEE Trans. Med. Imag., 12(1993), pp. 600-609;
- [HuL94] H. M. Hudson; R. S. Larkin, *Accelerated Image Reconstruction Using Ordered Subsets of Projection Data*, IEEE Trans. Med. Imag., 13(1994), pp. 601-609;
- [KaS88] A. C. Kak; M. Slaney, *Principles of Computerized Tomographic Imaging*, IEEE Press, 1999 (Electronic Copy), N. York.
- [Nat01] F. Natterer, *The Mathematics of Computerized Tomography*, SIAM (reimpressão, publicado por Wiley em 1986), 2001, Philadelphia.
- [NeB01] A. Nedić; D.P. Bertsekas, *Incremental Subgradient Methods for Non-differentiable Optimization*, SIAM J. Optim., 12(2001), pp. 109-138;

- [PiY01] A. R. De Pierro; M. E. B. Yamagishi, *Fast EM-like Methods for Maximum 'a posteriori' Estimates in Emission Tomography*, IEEE Trans. Med. Imag., 20(2001), pp. 280-288;
- [TaK03] E. Tanaka; H. Kudo, *Subset-Dependent Relaxation in Block-Iterative Algorithms for Image Reconstruction in Emission Tomography*, Institute of Physics Publishing Physics in Medicine and Biology, 48(2003), pp. 1405-1422;
- [VSK85] Y. Vardi; L. A. Shepp; L. Kaufman, *A Statistical Model for Positron Emission Tomography*, Journal of the American Statistical Association, Vol. 80, No. 389(Mar., 1985), pp. 8-20;

Lista de Figuras

1	Tomada de dados em tomografia por transmissão.	11
2	Exemplo de leitura PET.	13
1.1	Emissão e desintegração de pósitron.	21
1.2	Exemplo do fenômeno de espalhamento.	22
2.1	Amostras das projeções vistas no espaço de Fourier.	28
2.2	Comparação entre filtros em FBP.	30
2.3	Efeito do ruído na transformada de uma projeção.	30
3.1	Comparação FBP×EM.	34
3.2	Comparação FBP×EM, imagens normalizadas.	34
3.3	Comparação EM×OS-EM (10 subconjuntos).	37
3.4	Comparação EM×OS-EM (10 subconjuntos).	37
3.5	Comparação EM×OS-EM (10 subconjuntos).	38
3.6	Comparação EM×OS-EM (10 subconjuntos).	38
4.1	Espectro de Shepp-Logan.	46
4.2	Transformada sem Ruído.	47
4.3	Transformada com Ruído — $\sum_{i=1}^m b_i = 715.863$	47
4.4	OS-EM — Subconjuntos Sequenciais — Sequência Natural.	49
4.5	OS-EM — Subconjuntos Sequenciais — Sequência Perpendicular.	49
4.6	OS-EM — Subconjuntos Equilibrados — Sequência Natural.	50
4.7	OS-EM — Subconjuntos Equilibrados — Sequência Perpendicular.	50
4.8	OS-EM — 6 Subconjuntos.	51

4.9 OS-EM — 12 Subconjuntos.	51
4.10 OS-EM — 24 Subconjuntos.	52
4.11 OS-EM×RAMLA — $\sum_{i=1}^m b_i = 715, 813$ — 120 vistas.	54
4.12 OS-EM×RAMLA — $\alpha(N) = \frac{N-1}{23}$ — $\sum_{i=1}^m b_i = 715, 813$ — 120 vistas.	55
4.13 OS-EM×RAMLA — $\alpha(N) = \frac{N-1}{23}$ — $\sum_{i=1}^m b_i = 764, 713$ — 384 vistas.	56
4.14 OS-EM×RAMLA — $\alpha(N) = \frac{N-1}{23}$ — $\sum_{i=1}^m b_i = 1, 528, 687$ — 384 vistas.	57
4.15 OS-EM×RAMLA — $\alpha(N) = \frac{N-1}{47}$ — $\sum_{i=1}^m b_i = 715, 813$ — 120 vistas.	58
4.16 OS-EM×RAMLA — $\alpha(N) = \frac{N-1}{47}$ — $\sum_{i=1}^m b_i = 764, 713$ — 384 vistas.	59
4.17 OS-EM×RAMLA — $\alpha(N) = \frac{N-1}{47}$ — $\sum_{i=1}^m b_i = 1, 528, 687$ — 384 vistas.	60

Apêndice A

Implementação

A.1 Projeção A_y

A função a seguir calcula a projeção A_y . A matriz A é especificada pelo vetor θ , que indica quais vistas teriam sido amostradas na tomografia. I é a matriz com a imagem propriamente dita.

```
function [r,xp]=radon2(I,theta);
lt=length(theta); m=max(size(I));
lx=ceil(sqrt(2)*m/2);
xp=-(lx+1):(lx+1); r=zeros(2*lx+3,lt);
x0=floor((m+1)/2); y0=x0; wbar=waitbar(0,'Transformada de Radon');
for sc=1:lt
    for tc=1:2*lx+3
        t=tan(theta(sc)*pi/180); s=sin(theta(sc)*pi/180); c=cos(theta(sc)*pi/180);
        %*****
        %*****
        if abs(s)>eps % Evita divisao por zero
            for j=1:m % Percorre todas as colunas
                xj=j-x0-1;
                iy0=ceil(xp(tc)/s-xj/t)+1;
                if (i<=m) & (i>0)
                    if c==0 xl=Inf; else xl=xp(tc)/c+t*(i-y0-(t<0)); end;
                    r(tc,sc)=r(tc,sc)+min(abs([1 j-x0-1-xl]/s))*I(i,j);
                end;
            end;
        end;
        if abs(c)>eps % Evita divisao por zero
            for i=1:m % Percorre todas as linhas
                yi=y0-i+(t>0);
                j=x0+floor(xp(tc)/c-yi*t)+1;
                if (j<=m) & (j>0)
                    if s==0 yl=Inf; else yl=xp(tc)/s+(x0-j)/t; end;
                    r(tc,sc)=r(tc,sc)+min(abs([1 y0-i+(t>0)-yl]/c))*I(i,j);
                end;
            end;
        end;
    end; waitbar(sc/lt);
end; close(wbar);
```

A.2 Retroprojeção Filtrada

Para utilizar estes comandos, chame a sua matriz com os dados de **r** e o vetor com as vistas, **theta**. A imagem ficará na matriz **img**.

```
% Dados: r, theta.
m=size(r,2); lp=size(r,1); n=128;
p=zeros(1024,m); p(1:lp,:)=r;

%Filtragem
f=fft(p); w=pi*(0:512)/512;
w=[w';w(end-1:-1:2)'];
for i=1:size(f,2);
    %Note o filtro alterado
    fp(:,i)=f(:,i).*w.*sinc(w/(0.8*pi));
end;
fp=real(ifft(fp)); fp((lp+1):end,:)=[];

%Retroprojeção
img=zeros(n); x=(1-n/2):(n/2); x=repmat(x,n,1);
y=rot90(x); ctheta=cos(theta*pi/180); stheta=sin(theta*pi/180);
for i=1:m;
    t=x*ctheta(i)+y*stheta(i); a=floor(t); proj=fp(:,i);
    img=img+(t-a).*proj(a+ceil(lp/2)+1)+(a+1-t).*proj(a+ceil(lp/2));
end;
img=img/(2*m);
```

A.3 Ordered Subsets - Expectation Maximization

niter: Número de iterações a executar;

r: Matriz com os dados;

theta: Vetor com os ângulos (em graus) em que as projeções foram tomadas.;

xp: Vetor indicando a que distância do centro passa cada uma das projeções que forma uma vista;

N: Matriz em que cada linha é um vetor que indica quais as vistas que fazem parte de cada subconjunto;

x: Solução inicial a ser utilizada na reconstrução;

str1: Nome a ser utilizado nos arquivos que conterão os resultados de cada iteração.

```
function x=osem(niter, r, theta, xp, N, x, str1);
[lx,lt]=size(r); m=max(size(x)); [ns, nel]=size(N);
x0=floor((m+1)/2); y0=x0;
A=zeros(m,m,ns);
%Calculo de \Sigma_{i \in N_l}^m a_{ij}*****
for k=1:ns
    for sc=N(k,:)
        for tc=1:lx
            t=tan(theta(sc)*pi/180); s=sin(theta(sc)*pi/180); c=cos(theta(sc)*pi/180);
            if abs(s)>eps % Evita divisao por zero
                for j=1:m % Percorre todas as colunas
                    xj=j-x0-1;
                    i=y0-ceil(xp(tc)/s-xj/t)+1;
                    if (i<=m) & (i>0)
                        if c==0 xl=Inf; else xl=xp(tc)/c+t*(i-y0-(t<0)); end;
                        A(i,j,k)=A(i,j,k)+min(abs([1 j-x0-1-xl]/s));
                    end;
                end;
            end;
        end;
    end;
    if abs(c)>eps % Evita divisao por zero
        for i=1:m % Percorre todas as linhas
```

```

        yi=y0-i+(t>0);
        j=x0+floor(xp(tc)/c-yi*t)+1;
        if (j<=m) & (j>0)
            if s==0 yl=Inf; else yl=xp(tc)/s+(x0-j)/t; end;
            A(i,j,k)=A(i,j,k)+min(abs([1 y0-i+(t>0)-yl]/c));
        end;
    end;
end;
end;
end;
end;
end;
%*****
%Iteração OS-EM
for v=1:niter;
    %Subiteração
    for subiter=1:ns;
        ip=zeros(size(r));
        %Calculo de (Ax)_i*****
        for sc=N(subiter,:);
            for tc=1:lx;
                t=tan(theta(sc)*pi/180); s=sin(theta(sc)*pi/180);
                c=cos(theta(sc)*pi/180);
                if abs(s)>eps % Evita divisao por zero
                    for j=1:m % Percorre todas as colunas
                        xj=j-x0-1;
                        i=y0-ceil(xp(tc)/s-xj/t)+1;
                        if (i<=m) && (i>0)
                            if c==0 xl=Inf; else xl=xp(tc)/c+t*(i-y0-(t<0)); end;
                            ip(tc,sc)=ip(tc,sc)+min(abs([1 j-x0-1-xl]/s))*x(i,j);
                        end;
                    end;
                end;
                if abs(c)>eps % Evita divisao por zero
                    for i=1:m % Percorre todas as linhas
                        yi=y0-i+(t>0);
                        j=x0+floor(xp(tc)/c-yi*t)+1;
                        if (j<=m) && (j>0)
                            if s==0 yl=Inf; else yl=xp(tc)/s+(x0-j)/t; end;
                            ip(tc,sc)=ip(tc,sc)+min(abs([1 y0-i+(t>0)-yl]/c))*x(i,j);
                        end;
                    end;
                end;
            end;
        end;
    end;
end;
%*****
soma=zeros(m);
%Calculo de \sum_{i\in\mathcal{N}_l}\left\{(b_i/(Ax_i)-1)a_{ij}\right\}
for sc=N(subiter,:);
    for tc=1:lx;
        t=tan(theta(sc)*pi/180); s=sin(theta(sc)*pi/180);
        c=cos(theta(sc)*pi/180);
        if abs(s)>eps % Evita divisao por zero
            for j=1:m % Percorre todas as colunas
                xj=j-x0-1;
                i=y0-ceil(xp(tc)/s-xj/t)+1;
                if (i<=m) && (i>0)
                    if c==0 xl=Inf; else xl=xp(tc)/c+t*(i-y0-(t<0)); end;
                    if ip(tc,sc)~=0 | r(tc,sc)~=0
                        soma(i,j)=soma(i,j)+(r(tc,sc)/ip(tc,sc))*...
                            min(abs([1 j-x0-1-xl]/s)); end;
                end;
            end;
        end;
    end;
end;
if abs(c)>eps % Evita divisao por zero
    for i=1:m % Percorre todas as linhas
        yi=y0-i+(t>0);
        j=x0+floor(xp(tc)/c-yi*t)+1;
        if (j<=m) && (j>0)

```

```

        if s==0 yl=Inf; else yl=xp(tc)/s+(x0-j)/t; end;
        if ip(tc,sc)~=0 | r(tc,sc)~=0
            soma(i,j)=soma(i,j)+(r(tc,sc)/ip(tc,sc))*...
                min(abs([1 y0-i+(t>0)-yl]/c)); end;
        end;
    end;
end;
end;
end;
end;
%*****
x=x./A(:, :, subiter).*soma;
s1=num2str(v); s2=num2str(subiter);
fid=fopen(['c:\dados\' str1 '-' num2str(v) '-' num2str(subiter)], 'wb');
fwrite(fid,x,'double'); fclose(fid);
end;
end;

```

A.4 Row Action Maximum Likelihood Algorithm

Esta função possui os mesmos argumentos de entrada que a função `osem` e mais:
str2: String que indica que regra utilizar para o parâmetro de relaxação;
10: Parâmetro de relaxação inicial;

A: Parâmetro opcional, se fornecido evita o recálculo dos valores $\sum_{i=1}^m a_{ij}$.

```

function [x,A]=ramla(niter, r, theta, xp, N, x, str1, str2, 10, A);
[lx,lt]=size(r); m=max(size(x));
[ns, nel]=size(N);
x0=floor((m+1)/2); y0=x0; ct=1;
%Calculo de \sum_{i=1}^m a_{ij} *****
if nargin==9
    for sc=1:lt
        for tc=1:lx
            t=tan(theta(sc)*pi/180); s=sin(theta(sc)*pi/180); c=cos(theta(sc)*pi/180);
            if abs(s)>eps % Evita divisao por zero
                for j=1:m % Percorre todas as colunas
                    xj=j-x0-1;
                    i=y0-ceil(xp(tc)/s-xj/t)+1;
                    if (i<=m) & (i>0)
                        if c==0 xl=Inf; else xl=xp(tc)/c+t*(i-y0-(t<0)); end;
                        A(i,j)=A(i,j)+min(abs([1 j-x0-1-xl]/s));
                    end;
                end;
            end;
            if abs(c)>eps % Evita divisao por zero
                for i=1:m % Percorre todas as linhas
                    yi=y0-i+(t>0);
                    j=x0+floor(xp(tc)/c-yi*t)+1;
                    if (j<=m) & (j>0)
                        if s==0 yl=Inf; else yl=xp(tc)/s+(x0-j)/t; end;
                        A(i,j)=A(i,j)+min(abs([1 y0-i+(t>0)-yl]/c));
                    end;
                end;
            end;
        end;
    end;
end;
end;
%Iteração RAMLA
for v=1:niter;
    eval(['lambda=' str2 '']);
    %Subiteração
    for subiter=1:ns;
        ip=zeros(size(r));
        %Calculo de (Ax)_i*****
    end;
end;

```

```

for sc=N(subiter,:);
for tc=1:lx;
t=tan(theta(sc)*pi/180); s=sin(theta(sc)*pi/180);
c=cos(theta(sc)*pi/180);
if abs(s)>eps % Evita divisao por zero
for j=1:m % Percorre todas as colunas
xj=j-x0-1;
iy0=ceil(xp(tc)/s-xj/t)+1;
if (i<=m) && (i>0)
if c==0 xl=Inf; else xl=xp(tc)/c+t*(i-y0-(t<0)); end;
ip(tc,sc)=ip(tc,sc)+min(abs([1 j-x0-1-xl]/s))*x(i,j);
end;
end;
end;
if abs(c)>eps % Evita divisao por zero
for i=1:m % Percorre todas as linhas
yi=y0-i+(t>0);
j=x0+floor(xp(tc)/c-yi*t)+1;
if (j<=m) && (j>0)
if s==0 yl=Inf; else yl=xp(tc)/s+(x0-j)/t; end;
ip(tc,sc)=ip(tc,sc)+min(abs([1 y0-i+(t>0)-yl]/c))*x(i,j);
end;
end;
end;
end;
end;
%*****
soma=zeros(m);
%Calculo de \sum_{i\in\mathcal{N}_1}\left\{(b_i/(Ax_i-1)a_{ij})\right\}
for sc=N(subiter,:);
for tc=1:lx;
t=tan(theta(sc)*pi/180); s=sin(theta(sc)*pi/180);
c=cos(theta(sc)*pi/180);
if abs(s)>eps % Evita divisao por zero
for j=1:m % Percorre todas as colunas
xj=j-x0-1;
iy0=ceil(xp(tc)/s-xj/t)+1;
if (i<=m) && (i>0)
if c==0 xl=Inf; else xl=xp(tc)/c+t*(i-y0-(t<0)); end;
if ip(tc,sc)~=0 | r(tc,sc)~=0
soma(i,j)=soma(i,j)+(r(tc,sc)/ip(tc,sc)-1)*...
min(abs([1 j-x0-1-xl]/s)); end;
end;
end;
end;
if abs(c)>eps % Evita divisao por zero
for i=1:m % Percorre todas as linhas
yi=y0-i+(t>0);
j=x0+floor(xp(tc)/c-yi*t)+1;
if (j<=m) && (j>0)
if s==0 yl=Inf; else yl=xp(tc)/s+(x0-j)/t; end;
if ip(tc,sc)~=0 | r(tc,sc)~=0
soma(i,j)=soma(i,j)+(r(tc,sc)/ip(tc,sc)-1)*...
min(abs([1 y0-i+(t>0)-yl]/c)); end;
end;
end;
end;
end;
end;
%*****
x=x+lambda*ns*x.*soma./A;
if min(min(x))<0 disp(['Iteração ' num2str(v) ' - Subiteração ' num2str(subiter)]); end;
clear I; I=find(x<=0); x(I)=eps;
fid=fopen(['c:\dados\' str1 '-' num2str(v) '-' num2str(subiter)],'wb');
fwrite(fid,x,'double'); fclose(fid);
end;
end;

```