

Métodos Iterativos de Decomposição Seqüencial com Mudança de Escala em Tomografia

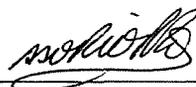
Michel Eduardo Beleza Yamagishi

200207473

Métodos Iterativos de Decomposição Sequencial com Mudança de Escala em Tomografia

Este exemplar corresponde à redação final da tese devidamente corrigida e defendida por Michel Eduardo Beleza Yamagishi e aprovada pela banca examinadora.

Campinas, 18 de dezembro de 2001



Prof. Dr. Álvaro Rodolfo De Pierro

Orientador

Banca Examinadora

Prof. Dr. Álvaro Rodolfo De Pierro

Prof. Dr. Sérgio Shiguemi Furuie

Prof. Dr. Nelson Delfino D'Ávila Mascarenhas

Profa. Dra. Cecil Chow Robilotta

Prof. Dr. José Mário Martinez Pérez

Tese apresentada ao Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas, como requisito parcial para obtenção do título de Doutor em Matemática Aplicada

UNICAMP
BIBLIOTECA CENTRAL
SEÇÃO CIRCULANTE

UNIDADE BC
Nº CHAMADA F/UNICAMP
Y14m
V _____ EX _____
TOLERA 47862
PREÇO 16-837102
C _____ D X
PREÇO R\$ 11,00
DATA 15-02-02
Nº CPD _____

CM0016378B-4

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

Yamagishi, Michel Eduardo Beleza

Y14m Métodos iterativos de decomposição sequencial com mudança de escala em tomografia / Michel Eduardo Beleza Yamagishi -- Campinas, [S.P. :s.n.], 2001.

Orientador : Álvaro Rodolfo De Pierro

Tese (doutorado) - Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

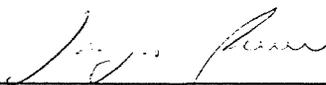
1. Métodos iterativos. 2. Tomografia computadorizada. 3. Otimização. I. De Pierro, Álvaro Rodolfo. II. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. III. Título.

Tese de Doutorado defendida em 18 de Dezembro de 2001 e aprovada

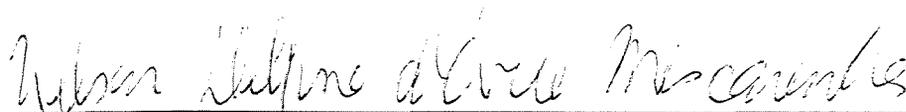
Pela Banca Examinadora composta pelos Profs. Drs.



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



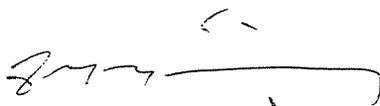
Prof (a). Dr (a). SÉRGIO SHIGUEMI FURUIE



Prof (a). Dr (a). NELSON DELFINO D'ÁVILA MASCARENHAS



Prof (a). Dr (a). CECIL CHOW ROBILOTTA



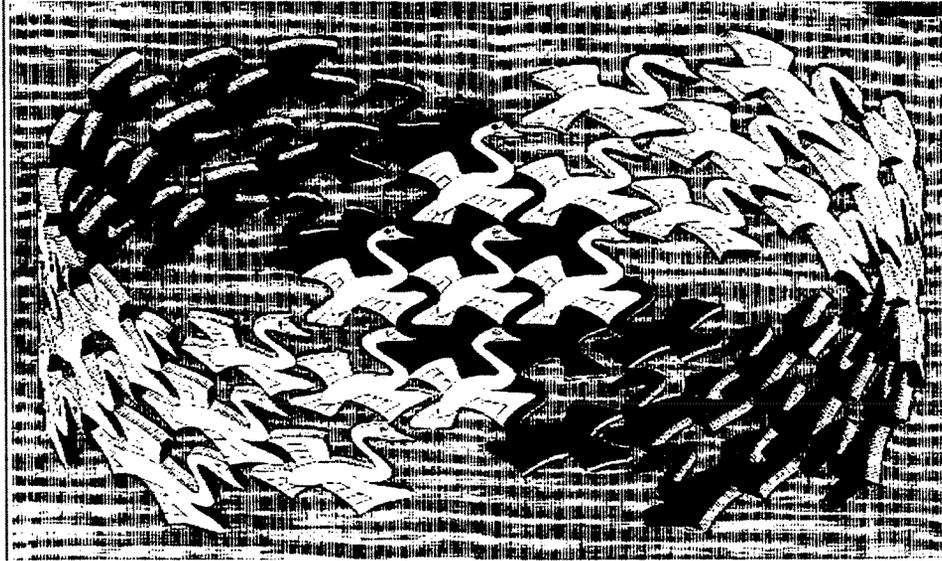
Prof (a). Dr (a). JOSÉ MÁRIO MARTINEZ PÉREZ

RESUMO

Apresentamos uma extensão do RAMLA para o problema de Máxima Verossimilhança Regularizado em Tomografia por Emissão. Provamos que, se a seqüência gerada por este método converge, então deve convergir para a solução. Propomos também novos métodos iterativos de reconstrução para os problemas de Máxima Verossimilhança e Máxima Verossimilhança Regularizado em Tomografia por Transmissão. Estes algoritmos são extensões em CT do RAMLA. Mostramos que os novos algoritmos produzem soluções similares ou melhores que o EM para CT e outros métodos de subconjuntos ordenados com a vantagem de serem mais rápidos e terem boas propriedades de convergência.

ABSTRACT

We present an extension of RAMLA for Regularized Maximum Likelihood (MAP) in Emission Tomography (ECT) reconstruction. We show that, if the sequence generated by this method converges, then it must converge to the true MAP solution. New iterative algorithms are presented for Maximum Likelihood (ML) and MAP reconstruction in Transmission Tomography (CT). The algorithms are natural extensions to CT of RAMLA. We show that the new algorithms for ML and MAP solutions produce similar, or even better results than the EM algorithm and other ordered subsets methods, but in much fewer iterations, and their convergence properties are better than other algorithms.



*"The gods did not reveal all things to men at the start;
but as time goes on, by searching, they discover more and
more"*

(Xenophanes)

Dedico este trabalho a minha esposa Mônica, a minha filha Laura, aos meus pais Marlene e Teruaki e aos meus avós Raymunda e William (*in memoriam*).

AGRADEÇO, EM PRIMEIRO LUGAR, A DEUS PELA CONCLUSÃO DESTA
TRABALHO. LOGO EM SEGUIDA, A MINHA FAMÍLIA PELO AMOR, APOIO
E ENCORAJAMENTO. AO MEU ORIENTADOR PROF. DR. ÁLVARO R. DE
PIERRO QUE, AO LONGO DESTES ANOS, SOUBE CONDUZIR A NOSSA PES-
QUISA COM MUITA SABEDORIA. À FAPESP PELO APOIO FINANCEIRO
QUE POSSIBILITOU A REALIZAÇÃO DESTA PESQUISA. À UNICAMP, MI-
NHA MÃE CIENTÍFICA. NELA ME GRADUEI, FIZ O MESTRADO E, AGORA,
CONCLUO O DOUTORADO COM MUITA GRATIDÃO. E A TODOS OS PRO-
FESSORES, AMIGOS, COLEGAS E FUNCIONÁRIOS DA UNICAMP E, EM
PARTICULAR, DO IMECC.

Sumário

1	Introdução	3
1.1	Matemática Aplicada	6
2	O Problema	8
2.1	Aspecto Físico	9
2.1.1	Fiat Lux	10
2.1.2	Raios-X	10
2.1.3	Tomografia Computadorizada	12
2.2	Aspecto Matemático	14
2.3	Imagens Tomográficas	15
2.3.1	Imagem Digital	16
2.3.2	Espaço de Frequência	16
2.4	Transformada de Radon	18
2.4.1	Propriedades da Transformada de Radon	18
2.4.2	Transformada de Radon Discreta	19
2.4.3	Interpolação pelo Inteiro Mais Próximo	20
2.4.4	Transformada Inversa de Radon	21
3	CT e ECT	24
4	Seguindo a História: Métodos Iterativos	26
4.1	Matriz de Projeção	26
4.2	ART	27
5	Máxima Verossimilhança: EM e RAMLA	30
5.1	Exemplos de Funções de Verossimilhança em Tomografia	31
5.1.1	Verossimilhança em ECT	31
5.1.2	Verossimilhança em CT	31
5.2	EM	32
5.3	OS-EM	33

<i>SUMÁRIO</i>	2
5.4 RAMLA	34
5.4.1 O Algoritmo	34
5.4.2 OS-EM versus RAMLA	36
6 O Modelo Regularizado e BSREM	38
6.1 Priors	39
6.2 OS-GP	40
6.3 BSREM	40
6.3.1 Primeira Versão	41
6.3.2 Versão Final	41
6.4 OS-GP versus BSREM	43
6.4.1 BSREM na Literatura	45
7 Novos Algoritmos para Transmissão: T-RAMLA e T-BSREM	46
7.1 Introdução	46
7.2 Novo Algoritmo : T-RAMLA	47
7.2.1 Convergência	48
7.3 Versão Regularizada	55
7.3.1 O Algoritmo T-BSREM	55
7.3.2 Resultado de Convergência	56
7.4 Simulações	57
7.4.1 T-RAMLA versus EM (Transmissão)	57
7.4.2 T-RAMLA versus OS-EM (Transmissão)	60
7.4.3 T-BSREM versus OSGA	61
8 Conclusões	64
8.1 Vantagens dos Novos Algoritmos	64
8.2 Pesquisas Futuras	65
9 Apêndice A	66
10 Apêndice B	68
11 Apêndice C	74
12 Apêndice D	88

Capítulo 1

Introdução

Os pontos de vista que aparecem nesta Tese refletem apenas a concepção do autor (MEBY)

“But it should always be required that a mathematical subject not be considered exhausted until it has become intuitively evident...” (Felix Klein)

A *Tomografia Computadorizada* é bastante abrangente. Podemos classificá-la como um assunto *Multidisciplinar*. Muitas pessoas, de diferentes formações, ajudaram na sua criação: físicos, engenheiros, matemáticos, médicos, etc. Confirmando a relevância da Tomografia, há três prêmios Nobel atribuídos a pesquisas relacionadas ao seu desenvolvimento.

Como dizia Isaac Newton (1642 - 1727): *“If I have seen a little farther than others it is because I have stood on the shoulders of giants”*¹. A ciência raramente dá saltos qualitativos. Em geral, seus avanços são contínuos². Ou seja, novas descobertas estão intimamente ligadas a resultados anteriores. Tendo isso em vista, resolvemos, na medida do possível, apresentar alguns dados históricos, procurando, dessa forma, dar ao leitor a sensação de *continuidade* entre os trabalhos anteriores e o nosso.

Entretanto, para evitar falsas expectativas, é preciso deixar claro dois pontos: (1) não foi nosso objetivo fazer um apanhado histórico exaustivo e detalhado. Buscamos indicar apenas aqueles trabalhos que, ao nosso ver, influenciaram direta ou indiretamente nossas pesquisas. Ou seja, este texto não pretende ser uma espécie de “review” sobre Tomografia. (2) Algumas partes importantes foram colocadas em segundo plano ou simplesmente ignoradas: a parte física, os métodos Diretos (ou de Transformação) de Reconstrução, etc. Isso para evitar fugir muito do tema principal.

Feitas essas considerações iniciais, vamos ao tema central: *“Métodos Iterativos de Decomposição Sequencial em Tomografia”*. Por se tratar de uma Tese sobre Métodos Iterativos, apenas descreveremos superficialmente os “Métodos Diretos”, em especial, o Método da Retroprojeção Filtrada ou, simplesmente, FBP

¹Uma tradução livre: “Se enxerguei um pouco mais que os outros, é porque estava sobre os ombros de gigantes”. Entre estes “gigantes” podemos citar Descartes, Kepler e Galileu. Veja [2].

²“Não devemos comparar a marcha da ciência com as transformações de uma cidade, onde os edifícios envelhecidos são impiedosamente demolidos para dar lugar às novas construções, mas sim com a evolução contínua dos tipos zoológicos que se desenvolvem sem cessar e acabam por se tornar irreconhecíveis aos olhares comuns, mas onde um olho experimentado reencontra sempre os vestígios do trabalho anterior dos séculos passados.” (Poincaré, [73], pp. 9)

(Filter Backprojection) e variantes. Tais métodos são muito importantes e, particularmente, eficientes quando a estatística de fótons é alta. Como será discutido oportunamente, tanto em Tomografia por Emissão quanto em Tomografia por Transmissão, quando a estatística de fótons é baixa, os Métodos Diretos deixam muito a desejar. É necessário modificar o modelo de forma a introduzir a natureza estatística do fenômeno para obter imagens hoje reconhecidamente superiores.

Um exemplo desses modelos "estatísticos" é baseado na Maximização da função de Verossimilhança. E para se maximizar essa função são empregados métodos iterativos, sendo, em geral, aceitos como melhores que o FBP, do ponto de vista da qualidade das imagens obtidas, o algoritmo conhecido como EM (de Expectation-Maximization). Entretanto, o EM é lento. São necessárias de 30 a 100 iterações (dependendo dos dados) para se obter imagens de boa qualidade, sendo que cada iteração do EM é essencialmente equivalente, em número de operações, à reconstrução usando o FBP, que é um método direto. Naturalmente, essa desvantagem poderia ser superada se de alguma forma o EM fosse acelerado.

Pesquisas independentes resultaram em vários métodos muito mais rápidos (aceleram o EM mais de uma ordem de grandeza, como por exemplo, o *Space-alternating generalized expectation maximization algorithm*-SAGE [21], Algoritmo do De Pierro [66], *Ordered Subsets EM* OS-EM de Hudson et al [39], Browne-De Pierro [5]) o que fez possível sua utilização na prática. Nesta Tese nos concentraremos numa família de métodos iterativos, que denominaremos de decomposição seqüencial com mudança de escala ('scaling') hoje usados nos 'scanners' mais modernos de *Positron Emission Tomography* - PET. Os dois métodos originais são: OS-EM [39] e o Row-Action Maximum Likelihood Algorithm - RAMLA [5]. Ambos baseados na idéia de Subconjuntos Ordenados. Contudo, somente o último possui um resultado de convergência para o máximo; enquanto o primeiro oscila, e, comprovadamente, não converge para o máximo da função de verossimilhança.

À medida que as iterações aumentam, devido ao fato de que este é um problema inverso mal condicionado, as imagens obtidas por esses dois algoritmos começam a apresentar o conhecido efeito "sal e pimenta", ou seja, o ruído começa a se sobrepor ao sinal. Uma possível solução é parar as iterações antes que o ruído seja maior que o sinal; outra forma é acrescentar informação 'a priori' usando um termo de regularização, modificando o modelo original. Tanto OS-EM como RAMLA consistem numa decomposição por blocos de projeções da função de verossimilhança mais a escolha de uma direção adequada de crescimento para cada bloco, que consiste numa mudança de escala diagonal aplicada ao gradiente do bloco. No caso de RAMLA, a subrelaxação produz a convergência para um ponto de máximo do ponto de vista teórico e evita as oscilações nas aplicações. Esta dissertação contém a extensão natural destas idéias em várias direções: para o modelo de máxima verossimilhança em Tomografia por transmissão, e para o modelo regularizado, tanto de emissão como de transmissão.

Os autores do OS-EM propuseram um método para o modelo regularizado: o OS-GP. Nossa contribuição começa exatamente nesse momento. Em primeiro lugar, suspeitamos que o OS-GP assim como o OS-EM tinha problemas de convergência, isto é, não convergia para o máximo da função de verossimilhança regularizada. Com um contra-exemplo simples demonstramos isso.

O passo seguinte foi criar o BSREM, que é a versão do RAMLA para o modelo regularizado. O principal

desafio foi como tratar de forma eficiente a parte de regularização. Esse desafio foi enfrentado de mais de uma forma diferente até a versão final do algoritmo. Parte dos resultados aparecem publicados em IEEE Transactions on Medical Imaging ([72] e Apêndice C).

A escolha da função de regularização ou, simplesmente, *priors* também foi objeto de estudo detalhado. Se o prior “penalizar” demais as descontinuidades, então as imagens podem ficar demasiadamente suavizadas (oversmoothing). Para isso não ocorrer, uma possibilidade é a escolha de *priors* mais complexos, que, em geral não são côncavos. Entretanto, *priors* não côncavos resultam em problemas de maximização global. Nesse caso, o BSREM, se convergir, o faz para um ponto crítico. Através de simulações, comparamos o BSREM a uma versão modificada do BSREM é capaz de lidar com os *priors* não convexos (côncavos), os resultados nos mostraram que o BSREM original obtém resultados equivalentes ao BSREM modificado, sugerindo que o BSREM original maximiza satisfatoriamente funções objetivo mais gerais (pelo menos tão bem quanto os algoritmos desenvolvidos para tais fins). Essas conclusões foram apresentadas e publicadas nos Proceedings da SPIE-99 ([71] e Apêndice B)

Finalmente, após tratar o caso de Tomografia por Emissão, perguntamos se seria possível estender satisfatoriamente este tipo de algoritmos rápidos a Tomografia por Transmissão. É sabido que, igualmente, o EM para Tomografia por Transmissão (TR-EM) é muito lento; e, após ler o trabalho sobre o TR-EM proposto por Lange e Carson, provamos que ele não é convergente.

Criamos um algoritmo rápido com os mesmo resultados de convergência do RAMLA. Esse novo algoritmo foi batizado de T-RAMLA. É claro que o modelo regularizado também foi abordado e como resultado criamos o T-BSREM assim como uma versão em blocos do método de Lange, Bahn e Little [48] para as comparações (o artigo correspondente submetido a publicação está no Apêndice D).

Essa é uma breve descrição do que será apresentado nos capítulos subsequentes. Nossa intenção foi sobrevoar rapidamente o texto, e dar ao leitor uma idéia do que já existia em termos de algoritmos iterativos e o que nós desenvolvemos. Após algumas considerações filosóficas sobre a matemática aplicada e os problemas inversos, no próximo capítulo apresentamos uma introdução ao problema da tomografia e no capítulo 3 uma descrição da Tomografia por emissão. Os capítulos 4 e 5 contêm uma descrição de alguns dos principais métodos iterativos em tomografia com ênfase nos métodos relacionados com o algoritmo EM, a decomposição seqüencial, e o ‘scaling’, que motivaram esta dissertação. No capítulo 6 discutimos a questão da regularização e apresentamos a extensão natural de RAMLA para o problema regularizado, que chamamos de BSREM. No capítulo 7 aplicamos as idéias dos capítulos anteriores no modelo de máxima verossimilhança em Tomografia por transmissão, tanto no caso em que regularizamos com um termo de penalização como quando apenas usamos um critério de parada experimental. Em todos os casos apresentamos simulações e comparações com outros métodos. Finalmente no capítulo 8 temos as conclusões do nosso trabalho e as possíveis direções de pesquisa originadas nesta dissertação. Nos Apêndices estão reproduzidos os artigos publicados e/ou submetidos como consequência deste trabalho de pesquisa. Apesar de conscientemente, deixamos muitos episódios sobre o desenvolvimento dos métodos nas entrelinhas, principalmente, os fracassos. Entretanto, sem eles, não teríamos apreendido tanto, e, tampouco, avançado no sentido correto. É uma injustiça deixá-los de fora. Mas, seja por costume seja por economia de letras, nas teses, nos livros, nos artigos e etc os fracassos são

omitidos e só aparecem os resultados positivos.

1.1 Matemática Aplicada

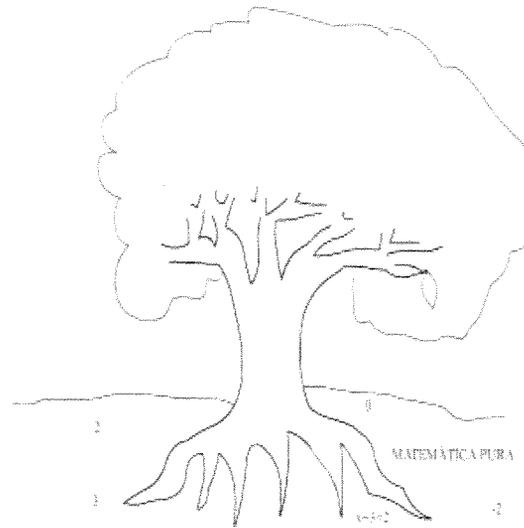
Vários autores (só para citar alguns, veja [18],[15],[37]) já tentaram definir os limites entre a Matemática Pura e Aplicada, ao nosso ver, sem muito sucesso. Não existe consenso entre os próprios matemáticos sobre o assunto. Entretanto, há uma afirmação aceita por “gregos e troianos”: não é possível fazer Matemática Aplicada sem a Matemática Pura.

A história nos mostra os gregos, verdadeiros pais da Matemática, procurando nela as leis pelas quais Deus havia criado o Universo.

“To understand why the Greeks created so much vital mathematics, one must investigate their objectives. It was the urgent and irrepressible desire of the Greeks to understand the physical world that impelled them to create and value mathematics. Mathematics was part and parcel of the investigation of nature and the key to comprehension of the universe, for mathematical laws are the essence of its design” ([43], pp. 145).

Já identificamos aí o desenvolvimento da Matemática Pura motivado por uma aplicação. E isso será regra por vários séculos. Por isso é tão difícil achar a fronteira entre a Matemática Pura e a Aplicada. Não é possível dizer, com precisão, onde uma termina e a outra começa.

De forma caricata, a Matemática Aplicada pode ser vista como uma árvore. Suas raízes estão enterradas na Matemática Pura e dela absorvem nutrientes. O tronco é formado por aqueles resultados comuns a todas as “grandes aplicações” da matemática. Os ramos ou galhos são as aplicações propriamente ditas: medicina, física, biologia, economia, otimização, problemas inversos, etc. Suas folhas são particularizações dessas aplicações. Por exemplo, no galho dos problemas inversos existe um ramo chamado “Reconstrução Tomográfica”. Nesse ramo há um sub-ramo chamado “Métodos Iterativos” e nesse sub-ramo há uma folha chamada “T-RAMLA” e outra folha chamada “T-BSREM”.



Essa é uma analogia simples e de forma alguma pretende ser “final”. Tampouco retrata adequadamente a complexidade do problema. A árvore, neste contexto, serve para ligar nosso tema “Reconstrução Tomográfica” à Matemática Aplicada.

Capítulo 2

O Problema

“...conduzir por ordem meus pensamentos, começando pelos objetos mais simples e mais fáceis de conhecer, para subir, pouco a pouco, como por degraus, até o conhecimento dos mais compostos...” (René Descartes, Discurso do Método)

Seguindo o conselho de René Descartes, tentaremos partir dos “objetos mais fáceis de conhecer” até o “conhecimento dos mais compostos”. Ou seja, descreveremos de forma simplificada a “física” do problema (objetos mais fáceis de conhecer) e, depois, apresentaremos um modelo matemático (conhecimento dos mais compostos) para o fenômeno físico.

A forte convicção dos antigos gregos de Deus ter criado o universo seguindo leis matemáticas influenciou muito o desenvolvimento da ciência. Durante séculos, as descobertas científicas pareciam confirmar cada vez mais essa crença.

A Matemática era por assim dizer a “rainha das ciências”, a “linguagem de Deus”, a forma mais segura de se conhecer a “verdade”. Matemática e Verdade eram palavras sinônimas.

Entretanto, a partir do século XIX, essa concepção foi seriamente abalada. A descoberta de geometrias não-euclidianas e estranhas álgebras forçaram os matemáticos a uma atitude mais humilde: os modelos matemáticos não necessariamente correspondiam à realidade. Os próprios fundamentos da matemática sofreram críticas[25] violentas (resultando no trabalho de G. J. Chaitin¹ que introduziu no seio da Matemática dois conceitos extremamente incômodos: “Randomness” [10] e “Unknowable” [11]).

Em 1921, Einstein sintetizou da seguinte forma a relação entre a “matemática” e a “realidade”:

“Insofar as the propositions of mathematics give an account of reality they are not certain; and insofar as they are certain they do not describe reality...But it is, on the other hand, certain

¹O trabalho de Chaitin é extremamente importante dentro dos “Fundamentos da Matemática”. Ele de certa forma estendeu os resultados de Gödel e Turing sobre a inconsistência/incompletude dos sistemas formais. O espantoso no seu trabalho foi mostrar que muitas vezes não conseguimos demonstrar certos teoremas não por incompetência, mas por falta de Estrutura (“Randomness”) e própria Matemática, e que há objetos matemáticos que são simplesmente “inconhecíveis” (“Unknowable”)! Para maiores informações: <http://www.cs.auckland.ac.nz/CDMTCS/chaitin>.

that mathematics in general and geometry in particular owe their existence to our need to learn something about the properties of real objects” ([44], pp. 97)

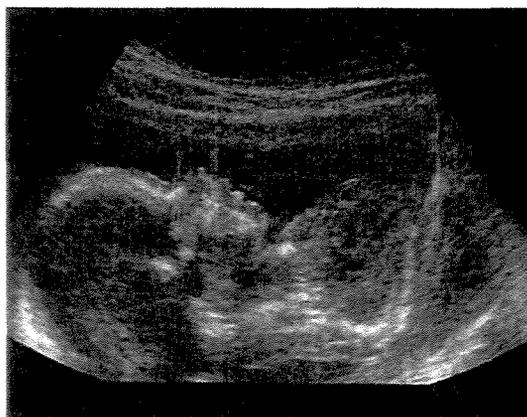
Assim, temos claro que o modelo matemático apenas *aproxima*² a realidade; e para enfatizar isso, os aspectos físicos e matemáticos do problema serão apresentados em seções separadas.

2.1 Aspecto Físico

Ter a visão de “raio-X” do Super-Homem sempre esteve entre os principais desejos da maioria dos homens (e mulheres). Ao lado da simples curiosidade, muitas outras justificativas podem ser enumeradas:

1. Diagnósticos médicos;
2. Controle de qualidade de peças;
3. Segurança;

Nos aeroportos internacionais, a visão de raio-X poderia ser utilizada para verificar se algum passageiro está transportando armas de fogo, bombas ou qualquer outro objeto (facas são agora especialmente visadas) que colocasse em risco a segurança dos demais passageiros. Nas grandes indústrias, serviria para procurar defeitos em peças metálicas sem destruí-las. Na medicina, a necessidade de diagnósticos e monitoramentos (veja figura abaixo) mais confiáveis e sem a necessidade de se fazer incisões nos pacientes. Só essas aplicações já justificariam plenamente o desejo dos seres humanos em “enxergar dentro ou através” dos objetos.



Ecografia de Monitoramento

²“Se considerarmos uma lei qualquer, de antemão podemos estar certos de que ela só pode ser aproximativa.” (Henri Poincaré). Veja Capítulo XI - A Ciência e a realidade - em [73].

2.1.1 Fiat Lux

Mas nossos olhos não vêem como os olhos do Super-Homem. São bem limitados. Só conseguem distinguir cores dentro de um comprimento de onda denominado “*espectro visível*” que vai, aproximadamente, de $380m\mu$ a $780m\mu$, onde $m\mu = \text{nanometro} = 10^{-9}m$. Na tabela abaixo, relacionamos os comprimentos de onda e as suas respectivas cores.

Cor	Faixa do Espectro ($m\mu$)
Violeta	380 – 440
Azul	440 – 490
Verde	490 – 565
Amarelo	565 – 590
Laranja	590 – 630
Vermelho	630 – 780

Para entender melhor como nossos olhos enxergam as cores, é necessário dizer algumas palavras sobre o modelo físico da *luz*.

A luz, segundo a física, tem uma natureza dual: é *partícula e onda* ao mesmo tempo. Um raio luminoso é formado por pequenas partículas chamadas *fótons*, e os fótons em movimento determinam uma onda cuja intensidade em cada ponto é igual à probabilidade de encontrar um foton nesse ponto. Esse modelo físico explica os fenômenos envolvendo a luz em que ora ela se comporta como partícula, ora como uma ondulagem eletromagnética.

Os fótons de um feixe de luz se deslocam a uma velocidade constante c que depende do meio; a ondulagem associada tem uma determinada frequência f . Quando os fótons encontram a retina, são gerados impulsos elétricos que, no cérebro, são traduzidos em percepção de cor³.

O comprimento de onda de um foton é determinado pela sua frequência f e pela sua velocidade segundo a fórmula $\lambda f = c$, onde λ representa o comprimento de onda.

2.1.2 Raios-X

Em 1901, o prêmio Nobel de Física foi concedido ao físico alemão Wilhelm Conrad Röntgen (1845 - 1923)

“in recognition of the extraordinary services he has rendered by the discovery of the remarkable rays subsequently named after him”.

³Para uma discussão mais aprofundada veja os capítulos 35 e 36 do volume I do Richard Feynman[22].



Röntgen

www.nobel.se/physics/laureates/1901/roentgen-bio.html

Embora só tenha recebido o prêmio Nobel em 1901, os primeiros experimentos feitos por Röntgen datam de 8 de novembro de 1895. Röntgen identificou um tipo “especial de luz⁴” capaz de atravessar objetos de espessuras modestas. Não demorou para ele perceber uma aplicação importantíssima.

Em 23 de janeiro de 1896, durante uma conferência, ele tirou uma “foto” da mão do biólogo alemão Rudolf Albert von Kölliker (1817 - 1905), usando uma fonte emissora de raios-X e um “filme” como anteparo (detector).



Mão de Von-Kölliker

www.vmas.kitasato-u.ac.jp/radiology/CALS/von-Kolliker.htm

Na “foto”, denominada na época de röntgenograma, os ossos da mão de Kölliker eram perfeitamente visíveis. Isso porque os raios-X não atravessavam os ossos da mão de Kölliker ou eram *bastante atenuados*, enquanto os tecidos moles (pele, gordura, sangue, etc) não ofereciam resistência a passagem dos raios-X.

Para alguns contemporâneos de Röntgen, seus experimentos foram responsáveis pela segunda grande revolução científica (sendo a primeira creditada aos experimentos de Galileu).

⁴Röntgen desconhecia completamente a natureza dos raios recém descobertos. Daí o nome “raios-X”. Só mais tarde *Max von Laue* mostrou que os raios-X tinham a mesma natureza eletromagnética da luz, diferindo apenas no seu comprimento de onda.

Röntgen morreu no dia 10 de fevereiro de 1923. Quase três décadas após seus primeiros experimentos. Causa da morte: carcinoma no intestino. Provavelmente, resultado de sua longa exposição aos raios-X.

Com a descoberta de Röntgen, a visão do “Super-Homem” ficou mais próxima.

2.1.3 Tomografia Computadorizada

A principal limitação dos raios-X é a sobreposição de estruturas. É fácil entender por que isso acontece. Quando realizamos uma tomada radiográfica, projetamos num plano um objeto, essencialmente, tridimensional. No momento da projeção, as estruturas são sobrepostas. Isso acarreta dificuldades de interpretação da imagem.

Quando a aplicação é médica, geralmente, o clínico solicita uma nova incidência radiográfica. A escolha desse ângulo é influenciada pelo conhecimento *a priori* que o médico tem da anatomia humana. Se ainda assim a dúvida persistir, outras incidências radiográficas são necessárias.

Se com imagens de raios-X em diferentes ângulos é possível dirimir dúvidas com relação às estruturas tridimensionais, então é natural perguntar se com um número grande de “imagens de raios-X” não seria possível reconstruir o interior do objeto, preservando as posições relativas entre as estruturas tridimensionais.

Cormack e Hounsfield

Os avanços teóricos, geralmente, precedem os avanços tecnológicos. A Tomografia Computadorizada não fugiu a regra.

Em 1963, Allan M. Cormack (1924 - 1998), físico, publicou um trabalho [12] no qual deu uma interpretação matemática para o problema de reconstruir a secção de um objeto através de suas projeções radiográficas. Com essa interpretação, Cormack deu um passo fundamental no desenvolvimento da tomografia.

Podemos imaginar um experimento mais simples para compreender o objetivo da “reconstrução tomográfica”. Para esse “experimento” precisaremos de um objeto qualquer e uma lanterna. Coloquemos esse objeto sobre uma mesa. Com a lanterna, iluminemos o objeto perfazendo 360 graus em torno da mesa, mantendo fixa nossa distância ao objeto. Para cada ângulo, se formará uma “sombra” (projeção) do objeto. A coleção dessas sombras são os nossos “dados”. Será possível “reconstruir a forma” tridimensional do objeto a partir de suas sombras?

No caso da tomografia, também temos uma espécie de “sombra” do objeto. Essa “sombra” é resultado da atenuação sofrida pelo raio-X ao atravessar o objeto. Dependendo das estruturas interceptadas, essa “sombra” pode ser mais ou menos intensa, e em alguns casos até mesmo ausente. O objetivo é, tendo como dados essas “sombras”, reconstruir uma secção do objeto, isto é, um corte do objeto.



Allan M. Cormack

www.nobel.se/medicine/laureates/1979/cormack-autobio.html

Cormack observou ser necessário encontrar uma função bidimensional $f(x, y)$ que descrevesse como os raios-X são atenuados pelas diversas estruturas no interior dos objetos, tendo como dados os valores médios da absorção de cada raio (as famosas integrais de linha).

Após oito anos do primeiro trabalho de Cormack, Godfrey Newbold Hounsfield, engenheiro, trabalhando na Electric and Musical Industries (EMI) construiu o primeiro “scanner tomográfico”, o EMI scanner. Esse equipamento foi especialmente desenvolvido para fazer imagens do crânio.



Godfrey N. Hounsfield

www.nobel.se/medicine/laureates/1979/hounsfield-autobio.html

Em 1979, Cormack e Hounsfield dividem o prêmio Nobel de Medicina

“for development of computer assisted tomography”.

Assim, temos três cientistas agraciados com o prêmio Nobel devido ao desenvolvimento das “imagens tomográficas”. Isso testemunha a importância do assunto, seja do ponto de vista científico seja do ponto de vista da aplicação prática (médica).

É claro que há muitos outros “aspectos físicos” não tratados aqui, como por exemplo, a interação entre radiação e a matéria (atenuação, espalhamento, absorção), efeitos da exposição à radiação e cálculo de doses de radiação em tecidos vivos [14], etc. Todos esses tópicos são importantíssimos, mas fogem ao escopo de nosso trabalho. Um resumo muito bom pode ser encontrado em [1] nos apêndices C e D.

2.2 Aspecto Matemático

Depois da interpretação matemática dada por Cormack, os modelos evoluíram rapidamente. É interessante notar, entretanto, que a principal ferramenta matemática do modelo foi desenvolvida em um contexto totalmente alheio à Tomografia Computadorizada.



Johann Radon

www-groups.dcs.st-and.ac.uk/history/Mathematicians/Radon.html

Johann Radon (1887 -1956) nasceu em Tetschen (atual Decin na República Tcheca). Radon fez muitos trabalhos em geometria diferencial e cálculo variacional; porém, a razão de citá-lo aqui é outro. Trata-se de seu trabalho envolvendo “*Transformações Matemáticas*”, em especial, uma transformação que recebeu o seu nome, isto é, a Transformada de Radon (TR).

A Transformada de Radon pode ser definida de diversas formas. A mais fácil de se compreender é a seguinte. Seja $g(x, y)$ uma função real no plano. A Transformada de Radon da função $g(x, y)$ é dada por

$$R(g) = \bar{g}(p, \tau) = \int_{-\infty}^{\infty} g(x, px + \tau) dx \quad \forall p, \tau \quad (2.1)$$

onde p é a inclinação da reta e τ é a intersecção com o eixo y da reta em questão. Em outras palavras, $\tilde{g}(p, \tau)$ é a integral de linha sobre a reta $y = px + \tau$. Variando-se p e τ cobrimos todo o plano.

Radon, em 1917, publicou um trabalho⁵ [74] onde apresentou uma fórmula para inverter 2.1,

$$g(x, y) = \int_{-\infty}^{\infty} \frac{-1}{2\pi^2\tau^2} * \tilde{g}(p, \tau) dp \quad (2.2)$$

onde $*$ indica a operação de convolução. Tanto a TR quanto a fórmula para invertê-la serão estudadas em detalhes em um capítulo especial.

Por enquanto, nosso objetivo principal é estabelecer a conexão entre o problema físico e o modelo matemático. Como mencionamos acima, no caso de Tomografia por Transmissão, os dados podem ser interpretados como a somatória das atenuações encontrada no caminho entre a fonte de radiação e o detector.

Se a atenuação das estruturas internas do objeto for “modelada” como uma função $g(x, y)$, os dados serão dados por

$$\int_{-\infty}^{\infty} g(x, px + \tau) dx$$

ou seja, a Transformada de Radon da função $g(x, y)$. Portanto, para reconstruirmos o objeto desejado a partir dos dados, basta inverter a Transformada de Radon.

Esse modelo pode ser facilmente estendido para o caso de Tomografia por Emissão. Nesse caso, os dados serão reinterpretados como a “contagem” dos fótons que chegam ao detector. Além disso, a estatística de fótons é muito menor, o que se traduz em uma maior possibilidade de ruídos nos dados. Isso torna o problema mais difícil de se resolver (problema mal posto ou com mal condicionamento ou problema infactível).

Antes de estudar um pouco mais a TR, é necessário falar sobre como entendemos a palavra “imagem”.

2.3 Imagens Tomográficas

“Não é, com efeito, empresa fácil transmitir e explicar o que pretendemos, porque as coisas novas são sempre compreendidas por analogias com as antigas” (Francis Bacon, Novum Organum)

A palavra *imagem* é bem conhecida. Há, entretanto, mais de uma conotação possível ao empregá-la. Veja os seguintes exemplos:

- a *imagem* do político está desgastada;
- estou olhando sua *imagem* através do espelho;
- sem a *imagem* tomográfica seria difícil diagnosticar esta doença.

⁵Com a permissão da BSB B. G. Teubner Verlagsgesellschaft, a AMS publicou a versão original alemã: J. Radon; **Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten**; Proceedings of Simposia in Applied Mathematics, Vol 27, pp. 71-86, 1982.

Existem outros exemplos, mas esses já são suficientes. Em nosso contexto, a palavra *imagem* estará vinculada a um objeto matemático. Isto é, uma abstração (modelo matemático) das “imagens” do mundo real.

Utilizamos basicamente duas representações da imagem enquanto objeto matemático: **contínua**⁶ e **digital**.

A primeira consiste em uma função $im : U \rightarrow C$, onde $U \subset \mathbb{R}^3$ é uma superfície e C é um espaço de cor⁷. Ou seja, é uma função que associa a cada ponto $(x, y, z) \in U$ uma cor do espaço C . A imagem digital é a versão discretizada da contínua, isto é, tanto o conjunto domínio U quanto o seu contra domínio C são **discretos**. Na Computação Gráfica, a discretização de C recebe um nome especial: **quantização**. Esta representação é especialmente adequada para a manipulação computacional (processamento) das imagens. Portanto, esta será a mais utilizada.

2.3.1 Imagem Digital

Como as paisagens do mundo físico (3-D) são projetadas na retina de nossos olhos (2-D), podemos considerar $U \subset \mathbb{R}^2$; em outras palavras, para nosso sentido visual as imagens são, essencialmente, bidimensionais. Além disso, sem perda de generalidade, podemos ainda supor que U é uma região do plano retangular⁸.

Falta ainda dizer uma palavra sobre a discretização de U . Geralmente, esta é feita da forma tradicional, dividindo-se a região em pequenos retângulos de dimensões iguais. Cada retângulo recebe o nome de **pixel** (Picture Element). A função im , neste caso, associará a cada retângulo a cor resultante da média das cores dos pontos pertencentes ao referido retângulo.

Já o espaço de cor C será unidimensional, ou seja, as imagens serão monocromáticas. Além disso, há uma restrição a mais: só trabalharemos com os valores visíveis aos olhos humanos. Portanto, na verdade, C será um intervalo finito $C = [a, b]$. Para discretizar esse intervalo, basta tomar valores igualmente espaçados, por exemplo.

Podemos pensar a imagem digital como uma **matriz** A de dimensões $m \times n$. Onde m será o número de pixels no eixo-y e n será o número de pixels no eixo-x. Os pixels serão identificados por um par de coordenadas (i, j) . A cada pixel (i, j) será associado, através da função im , o valor a_{ij} .

Feitas essas considerações, já podemos pensar em “manipular” matematicamente as imagens, pois, agora, elas são “matrizes”, objeto matemáticos bem familiares.

2.3.2 Espaço de Freqüência

Dentre as principais operações com imagens está a *filtragem*. Esta geralmente se dá no espaço de Fourier ou espaço de freqüências.

⁶Contínua no sentido de não-discreta.

⁷Para uma descrição mais aprofundada, consultar [29] e [40]. Podemos pensar em C como um “Espaço Vetorial”. O exemplo mais conhecido é o RGB. Onde C é um Espaço Vetorial de dimensão 3, cujos elementos da base são as cores Vermelha (Red), Verde (Green) e Azul (Blue). Se o espaço C tiver dimensão 1, dizemos que a imagem é monocromática. É importante lembrar que apenas uma pequena região do espaço de cor é visível aos olhos humanos.

⁸No caso em que essa condição não seja verificada, podemos circunscrever a imagem com um retângulo de dimensões apropriadas.

A filtragem aparece nos mais diferentes contextos dentro do processamento de imagens. Seja em Compressão de Imagens, onde está associada a um filtro de passa-baixa, ou seja em Reconstrução Tomográfica onde aparece um filtro de passa-alta.

Para entendermos um pouco melhor a filtragem, vamos definir a *Transformada de Fourier Discreta (TFD) Bidimensional* que nos leva ao espaço de frequências.

A TFD, $U = (u_{kl})$, de uma imagem $A = (a_{mn})$ com N linhas e N colunas é dada por

$$u_{kl} = \frac{1}{N} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} a_{mn} \exp\left(\frac{-2\pi i km}{N}\right) \exp\left(\frac{-2\pi i ln}{N}\right), \quad 0 \leq k, l \leq N-1 \quad (2.3)$$

e sua transformada inversa será

$$a_{mn} = \frac{1}{N} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} u_{kl} \exp\left(\frac{2\pi i km}{N}\right) \exp\left(\frac{2\pi i ln}{N}\right), \quad 0 \leq m, n \leq N-1. \quad (2.4)$$

Em notação matricial, $U = F^*AF$. O número de operações para efetuar uma TFD usando a *FFT* (Fast Fourier Transform) ([40],[84]) é da ordem de $O(N^2 \log_2 N)$, que é bem melhor que usar simplesmente a equação 2.3.

Na imagem a seguir, realizamos a TDF. Na segunda imagem temos a representação de seu módulo.



Já na imagem a seguir utilizamos um filtro de passa-baixa. Veja o efeito sobre a imagem.



2.4 Transformada de Radon

“Trying to find the solution, we may repeatedly change our point of view, our way of looking at the problem.”
(G. Polya)

Como a Transformada de Radon (TR) é a principal ferramenta utilizada no modelo matemático, faremos um resumo das principais propriedades e teoremas relacionados a ela.

Aproveitamos para introduzir uma forma alternativa de escrever a TR

$$\tilde{g}(p, \tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \delta(y - px - \tau) dx dy, \quad (2.5)$$

onde $\delta(\cdot)$ é a função delta de Dirac. A fórmula 2.5 é exatamente igual a fórmula 2.1. A vantagem da representação acima está na utilização das propriedades da função distribuição $\delta(\cdot)$.

2.4.1 Propriedades da Transformada de Radon

Uma transformação matemática é tanto mais útil do ponto de vista prático quanto mais boas propriedades ela tiver. Ser linear e inversível são duas das principais propriedades desejáveis em uma transformação matemática. Veremos a seguir que a TR possui essas duas e algumas outras propriedades interessantes.

Linearidade

A Transformada de Radon é uma Transformação Linear. Assim, se tivermos n funções $g_i(x, y)$, $i = 1, \dots, n$, definimos uma nova função $h(x, y)$ como a soma ponderada das funções $g_i(x, y)$

$$h(x, y) = \sum_{i=1}^n w_i g_i(x, y),$$

onde $w_i \in \mathfrak{R}$.

A TR de $h(x, y)$ é

$$\tilde{h}(p, \tau) = \sum_{i=1}^n w_i \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_i(x, y) \delta(y - px - \tau) dx dy = \sum_{i=1}^n w_i \tilde{g}_i(p, \tau), \quad (2.6)$$

ou seja, a Transformada de Radon da soma ponderada de funções nada mais é que a soma ponderada das Transformadas de Radon das funções em questão. Essa é sem dúvida a propriedade mais importante da Transformada de Radon.

Deslocamento

Se definirmos a função $h(x, y)$ como um deslocamento da função $g(x, y)$, isto é, $h(x, y) = g(x + a, y + b)$, qual será a TR da nova função $h(x, y)$?

$$\tilde{h}(p, \tau) = \int_{-\infty}^{\infty} g(x - a, px + \tau - b) dx,$$

mundando-se de variável de integração, $z = x - a$, temos

$$\tilde{h}(p, \tau) = \int_{-\infty}^{\infty} g(z, p(z + a) + \tau - b) dz = \tilde{g}(p, \tau - b + pa).$$

Notamos que a inclinação da reta não muda. Geometricamente, isso era esperado.

Escalonamento

Vamos definir a função $h(x, y)$ da seguinte forma $h(x, y) = g\left(\frac{x}{a}, \frac{y}{b}\right)$, onde $a > 0$ e $b > 0$. A TR da função $h(x, y)$ assim definida será

$$\tilde{h}(p, \tau) = \tilde{h}(p, \tau) = \int_{-\infty}^{\infty} g\left(\frac{x}{a}, \frac{px + \tau}{b}\right) dx,$$

mudando-se a variável de integração, $z = \frac{x}{a}$, obtemos

$$\tilde{h}(p, \tau) = a \int_{-\infty}^{\infty} g\left(z, \frac{paz + \tau}{b}\right) dz = a \tilde{g}\left(\frac{pa}{b}, \frac{\tau}{b}\right).$$

2.4.2 Transformada de Radon Discreta

Como trabalharemos com imagens digitais, precisaremos discretizar a TR. E isso pode ser feito de diversas formas diferentes, dependendo da aplicação.

A seguir apresentaremos uma versão simplificada da TR discreta. Primeiramente, vamos discretizar as variáveis envolvidas na transformação

$$\begin{aligned}
x &= x_m = x_{min} + m\Delta x, \quad m = 0, 1, \dots, M-1 \\
y &= y_n = y_{min} + n\Delta y, \quad n = 0, 1, \dots, N-1 \\
p &= p_k = p_{min} + k\Delta p, \quad k = 0, 1, \dots, K-1 \\
\tau &= \tau_h = \tau_{min} + h\Delta\tau, \quad h = 0, 1, \dots, H-1
\end{aligned}$$

veja que a notação é a usual, por exemplo, x_{min} denota a primeira posição no eixo x , Δx é o incremento ou a distância entre dois pontos consecutivos no eixo x , e, finalmente, m é o índice de cada um dos M pontos no eixo x .

Feitas essas considerações, vejamos como fica a aproximação discreta da TR. Usando 2.1, temos

$$\tilde{g}(p_k, \tau_h) = \int_{-\infty}^{+\infty} g(x, p_k x + \tau_h) dx \approx \Delta x \sum_{m=0}^{M-1} g(x_m, p_k x_m + \tau_h). \quad (2.7)$$

A função $g(x, y)$ nos fornece uma imagem digital da seguinte forma

$$g(m, n) = g(x_m, y_n), \quad (2.8)$$

analogamente para a TR discreta

$$\tilde{g}(k, h) = \tilde{g}(p_k, \tau_h). \quad (2.9)$$

Entretanto, dada a imagem digital $g(m, n)$, surge um problema com relação a equação 2.7. Em geral, $p_k x_m + \tau_h$ não coincide com os pontos y_n . Ou seja, a equação 2.7 requer pontos que não são encontrados na imagem digital.

Há várias formas de contornar esse problema.

2.4.3 Interpolação pelo Inteiro Mais Próximo

Essa talvez seja a solução mais simples. Ela consiste em tomar o inteiro mais próximo do número

$$\frac{p_k x_m + \tau_h - y_{min}}{\Delta y},$$

vamos denotar esse inteiro por $i(m; k, h)$. Dessa forma, a TR discreta pode ser escrita como

$$\tilde{g}(k, h) = \Delta x \sum_{m=0}^{M-1} g(m, i(m; k, h)). \quad (2.10)$$

Dois observações importantes:

1. O ponto $(m, i(m; k, h))$ pode não estar contido na imagem $g(m, n)$. Se ele estiver fora da imagem, podemos igualar a zero o valor da imagem g nesse ponto. Isso significa que esse ponto não tem

CAPÍTULO 2. O PROBLEMA

contribuição alguma na TR discreta.

2. O incremento Δx pode ser desprezado, pois ele não acrescenta informação alguma a TR discreta. Ele só seria necessário se a TR discreta aproximasse quantitativamente a TR contínua.

Interpolação Linear

Uma melhor aproximação pode ser obtida utilizando-se a Interpolação Linear no eixo y da imagem. Vejamos como proceder.

Usaremos a notação $[x]$ para denotar a função que escolhe o maior inteiro n tal que $n \leq x$. Feita essa observação, a TR discreta usando interpolação linear pode ser escrita como

$$\tilde{g}(k, h) = \Delta x \sum_{m=0}^{M-1} (1 - \alpha) g(m, n) + \alpha g(m, n + 1), \quad (2.11)$$

onde $n = \left\lfloor \frac{pkx_m + Th - y_{min}}{\Delta y} \right\rfloor$ e $\alpha = \frac{pkx_m + Th - y_{min}}{\Delta y} - n$.

É fácil verificar que esta discretização da TR é computacionalmente mais cara que a anterior, mas, em geral, métodos mais sofisticados de interpolação exigem um esforço computacional maior. Esta é a escolha que deve ser feita na hora de escolher o algoritmo de interpolação.

2.4.4 Transformada Inversa de Radon

Há várias formas de se obter a Transformada Inversa de Radon (TIR). Entre elas podemos citar por razões teóricas a forma proposta originalmente por Radon 2.2 e aquela baseada no Teorema de Projeção de Fourier, ou “*Fourier Slice Theorem*”, por razões de implementação computacional.

Vamos apresentar esta última. Aproveitando a oportunidade, introduziremos uma terceira forma de representar a TR, a chamada forma *Normal da TR*.

Teorema de Projeção de Fourier

A forma Normal da TR recebe este nome devido a “equação normal” da reta (ρ, θ) sobre a qual será feita a integração, isto é,

$$\rho = x \cos(\theta) + y \sin(\theta).$$

Usando essa forma da equação da reta e o auxílio da função distribuição $\delta(\cdot)$, a TR pode ser escrita como

$$\tilde{g}(\rho, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \delta(\rho - x \cos(\theta) - y \sin(\theta)) dx dy.$$

Agora, façamos uma pequena revisão da Transformada de Fourier bidimensional.

A Transformada de Fourier bidimensional da função $g(x, y)$ é

$$G(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \exp(-2\pi i (ux + vy)) dx dy, \quad (2.12)$$

onde $i = \sqrt{-1}$. A Transformada inversa é

$$g(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G(u, v) \exp(2\pi i (ux + vy)) du dv. \quad (2.13)$$

Mudando para coordenadas polares no espaço de frequência,

$$\begin{aligned} u &= r \cos(\theta) \\ v &= r \sen(\theta) \end{aligned}$$

Podemos reescrever a equação 2.12 da seguinte forma

$$\begin{aligned} G(r \cos(\theta), r \sen(\theta)) = \\ \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \delta(\rho - x \cos(\theta) - y \sen(\theta)) \exp(-2\pi i r \rho) dx dy \right] d\rho \end{aligned}$$

ou ainda,

$$G(r \cos(\theta), r \sen(\theta)) = \quad (2.14)$$

$$\int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \delta(\rho - x \cos(\theta) - y \sen(\theta) dx dy \right] \exp(-2\pi i r \rho) d\rho$$

a parte entre colchetes nesta última equação é a TR, $\tilde{g}(\rho, \theta)$.

Portanto,

$$G(r \cos(\theta), r \sen(\theta)) = \int_{-\infty}^{\infty} \tilde{g}(\rho, \theta) \exp(-2\pi i v \tau) d\tau, \quad (2.15)$$

isto é, a Transformada de Fourier unidimensional da Transformada de Radon nos fornece a Transformada de Fourier bidimensional de $g(x, y)$.

Esse é essencialmente o *Teorema de Projeção de Fourier*. Facilita muito a inversão da TR, pois basta realizar a inversão da Transformada de Fourier bidimensional que, por sua vez, tem algoritmos bem eficientes (Fast Fourier Transform - FFT, por exemplo).

Desse teorema surgem os principais algoritmos diretos para reconstrução de imagens tomográficas. Sendo o mais importante o *Método da Retroprojeção Filtrada*, ou *Filtered Backprojection (FBP)*. Os Métodos Diretos [36] para reconstrução tomográfica por si só já mereciam um estudo detalhado. Entretanto, não serão aqui abordados, pois fugiria do nosso tema principal. Uma ótima introdução aos métodos diretos pode ser

encontrada em [53]. Nesse trabalho, além de uma excelente introdução ao FBP e ao método de Newton, há um enorme número de referências na bibliografia sobre os métodos diretos.

Método Direto: FBP

O FBP ainda é o método de reconstrução mais utilizado pelos scanners. Quando há uma estatística de fótons elevada e não há muito ruído nos dados, o FBP gera imagens de boa qualidade, o que justifica sua utilização em muitos casos. Entretanto, na ausência dessas premissas, o FBP é de longe superado pelos métodos iterativos que levam em conta a natureza estatística da coleta dos dados (o EM e similares, por exemplo).

Vamos ao FBP. Introduzindo coordenadas polares na equação 2.13, temos

$$g(x, y) = \int_0^{2\pi} \int_0^{\infty} r G(r \cos(\theta), r \sin(\theta)) \exp(2\pi i r (x \cos(\theta) + y \sin(\theta))) dr d\theta, \quad (2.16)$$

ou equivalentemente

$$g(x, y) = \int_0^{\pi} \int_{-\infty}^{\infty} |r| G(r \cos(\theta), r \sin(\theta)) \exp(2\pi i r (x \cos(\theta) + y \sin(\theta))) dr d\theta, \quad (2.17)$$

pelo *Slice Fourier Theorem*, podemos escrever

$$\begin{aligned} \hat{g}(\rho, \theta) &= \int_{-\infty}^{\infty} |r| G(r \cos(\theta), r \sin(\theta)) \exp(2\pi i r \rho) dr \\ &= \int_{-\infty}^{\infty} |r| \left(\int_{-\infty}^{\infty} \tilde{g}(\tilde{\rho}, \theta) \exp(-2\pi i r \tilde{\rho}) d\tilde{\rho} \right) \exp(2\pi i r \rho) dr, \end{aligned} \quad (2.18)$$

que não passa de uma filtragem com o filtro $|r|$ (highpass), depois substituindo-se em 2.17, chegamos a parte da retroprojeção

$$g(x, y) = \int_0^{\pi} \hat{g}(x \cos(\theta) + y \sin(\theta), \theta) d\theta. \quad (2.19)$$

Filtrar (2.18) e retroprojetar (2.19) constituem as duas etapas do FBP. Há muitas versões diferentes do FBP. A maior parte diferindo apenas no filtro usado em 2.18. A escolha do filtro é fundamental nos casos onde o ruído pode ser amplificado por filtros highpass. Veja o capítulo 10 do livro [40] para um estudo mais detalhado.

Capítulo 3

CT e ECT

Há uma divisão que precisa ser esclarecida antes de falarmos nos métodos de reconstrução. É a divisão em Tomografia por Transmissão ou, simplesmente, CT e Tomografia por Emissão ou ECT.

Do ponto de vista clínico, esta divisão diz respeito ao tipo de avaliação pretendida. Se o médico quer fazer um estudo anatômico do paciente, então a Tomografia por Transmissão é a indicada; se, por outro lado, procura-se observar o funcionamento de um determinado órgão, ou seja, a fisiologia, então deve-se preferir a Tomografia por Emissão.

Outra diferença importante é a forma como os dados são obtidos. Na Tomografia por Transmissão, o modelo é aquele já apresentado na introdução, ou seja, o paciente (ou objeto) é colocado entre uma fonte de radiação e um detector. Dessa forma, tanto a fonte quanto o detector estão *fora* do paciente. Em Emissão, aplica-se (seja por via endovenosa ou por via oral) uma dose de um traçador radioativo no paciente. Esse traçador depende, entre outros fatores, do órgão a ser examinado. Pois, como o estudo é fisiológico, é necessário escolher um traçador que seja metabolizado pelos tecidos do órgão em questão. Daí a diferença surge, porque a *fonte* de radiação está dentro do paciente.

Há dois tipos principais de Tomografia por Emissão: PET (Positron Emission Tomography) [82] e SPECT (Single-Photon Emission Computed Tomography). A primeira é a mais precisa (e mais cara). É baseada na utilização de radioisótopos que decaem emitindo positrons.

O positron viaja curtas distâncias (da ordem de milímetros no máximo) antes de aniquilar-se com um elétron dos tecidos do paciente. No momento dessa aniquilação são emitidos dois fótons, cuja energia é de 0.511 MeV, que se dispersam, aproximadamente, em direções contrárias; ou seja, na mesma linha, mas em sentidos opostos.

O trabalho do detector é incrementar o contador de atividade todas as vezes que perceber a chegada simultânea de dois fótons.

Já no segundo caso, isto é, em SPECT, utilizam-se radioisótopos que decaem emitindo raios gama. Ao contrário do caso anterior, esses raios gama são emitidos como um único fóton. O detector conta as chegadas desses fótons e incrementa o contador.

Uma diferença crucial é a estatística de fótons. Em CT, temos uma estatística de fótons alta; enquanto, em ECT, especialmente em SPECT, temos uma estatística de fótons baixa. Essa diferença é o ponto chave de nosso trabalho, pois os Métodos Diretos são muito eficientes quando a estatística de fótons é alta (pouco ruído). Contudo, quando a estatística de fótons é baixa, o ruído é consideravelmente maior e os Métodos Diretos já não são adequados, nascendo a necessidade de modelos estatísticos que são resolvidos via algoritmos iterativos.

Uma última diferença que desejamos mencionar é a natureza dos dados coletados. No caso da Tomografia por Transmissão os dados refletem a *atenuação* sofrida pela radiação no interior do organismo. Assim, quando vemos uma Tomografia por Transmissão, estamos na verdade vendo um mapa de atenuação. Já em Tomografia por Emissão, é a *atividade* do radioisótopo que é mapeada.

É claro que novas técnicas têm associado as imagens de ECT e CT com a finalidade de produzir diagnósticos mais precisos.

Capítulo 4

Seguindo a História: Métodos Iterativos

“Nature is not embarrassed by difficulties of analysis.” (Augustin Fresnel)

O primeiro método de reconstrução tomográfica implementado em um scanner não foi um Método Direto. A máquina de Cormack e Hounsfield (EMI scanner) utilizava um método iterativo para reconstrução tomográfica.

Cormack tratou o seu modelo [12] de uma forma ligeiramente diferente da proposta por Radon, provavelmente por não conhecer o trabalho deste. Cormack recuperava o mapa de atenuação f através da solução da “Equação de Abel” [13]. Apesar de afirmar que essa forma era “*quite amenable to numerical analysis*”, no momento de implementar no EMI Scanner, preferiu um método iterativo.

4.1 Matriz de Projeção

Em linguagem moderna, imaginemos um operador linear S aplicado sobre uma função $g(x, y)$,

$$\tilde{g}(x, y) = Sg(x, y),$$

onde o resultado, ou a imagem do operador S , denotamos por $\tilde{g}(x, y)$.

No caso da tomografia, o operador S pode ser aproximado por uma matriz $A \in \mathbb{R}^{m \times n}$. Esta matriz é denominada *Matriz do Sistema* ou *Matriz de Projeção*. Cada elemento $a_{ij} \in A$ pode ser aproximado [34] pelo comprimento linear da intersecção do i -ésimo raio com o j -ésimo pixel.

O problema então assume a forma de um *Sistema Linear*

$$Ax = b, \tag{4.1}$$

onde $x \approx g(x, y)$ e $b \approx \tilde{g}(x, y)$.

Vejamos algumas de suas principais características:

- em geral, não é quadrada ($m \neq n$).
- é normalmente muito grande, pois o número de linhas corresponde ao número total de raios que por sua vez é igual ao número de vistas (ângulos) vezes o número de raios em cada vista (por exemplo 180 vistas vezes 201 raios por vista = 36180 linhas), e o número de colunas corresponde ao número de pixels na imagens (por exemplo, uma imagem quadrada da ordem de 256 por 256 gera uma matriz com 65536 colunas).
- é (quase) singular, isto é, possui *valores singulares* ([38],[28]) bem próximos a zero. Esta é a principal característica de um problema “mal condicionado”.
- não possui uma estrutura simples, como por exemplo, uma matriz de banda ou triangular possuem.
- é esparsa.

4.2 ART

Com essas características pouco atraentes, resolver 4.1 não é uma tarefa simples. Em situações dessa natureza, pode-se aplicar o *ART* (Algebraic Reconstruction Technique) que foi publicado pela primeira vez na área biomédica em 1970 [30], segundo Yair Censor [7]. Foi esse o método utilizado por Cormack e Hounsfield, e trata-se de um método iterativo.

É um método conhecido pelos matemáticos desde 1937; embora com um nome diferente: algoritmo de Kaczmarz [41]. A idéia é aproximar a solução do sistema através de *Projeções Ortogonais*¹ sobre os hiperplanos gerados pelas equações do sistema linear.

O algoritmo começa com um $x^0 \in \mathbb{R}^n$ arbitrário (chute inicial: pode ser uma imagem uniformemente constante), e atualiza a imagem em cada iteração de acordo com a fórmula

$$x^{k+1} = x^k + \frac{b_i - \langle a^i, x^k \rangle}{\|a^i\|^2} a^i, \quad (4.2)$$

onde a^i denota a i -ésima linha da matriz A .

Se definirmos o conjunto

$$H_i = \{x \in \mathbb{R}^n / \langle a^i, x \rangle = b_i\},$$

então o ponto x^{k+1} não passa da projeção ortogonal do ponto x^k sobre o hiperplano H_i . Se o sistema for consistente, isto é, se o conjunto $S = \bigcap_i H_i$ não for vazio, então a seqüência gerada por 4.2 converge para $x^* \in S$, veja capítulo 2 do livro [80]. Essa mesma teoria garante a convergência para a versão relaxada do ART, ou seja,

¹Uma excelente introdução aos métodos iterativos baseados em Projeções Ortogonais é o capítulo 2 de [80]. Ai a teoria é revisada partindo do trabalho pioneiro de John Von Neumann até os resultados mais recentes.

$$x^{k+1} = x^k + \lambda_k \frac{b_i - \langle a^i, x^k \rangle}{\|a^i\|^2} a^i, \quad (4.3)$$

onde $\varepsilon_1 \leq \lambda_k \leq 2 - \varepsilon_2$ com $\varepsilon_1, \varepsilon_2 > 0$.

Caso o sistema $Ax = b$ seja inconsistente, 4.3 gerará m pontos fixos, onde m é o número de linhas da matriz A . Se fizermos $\lambda_k \rightarrow 0$, a seqüência convergirá ao ponto x^* solução do problema

$$\min \sum_{i=1}^m \left(\frac{b_i - \langle a^i, x \rangle}{\|a^i\|} \right)^2, \quad (4.4)$$

ou seja, x^* não é solução de $Ax = b$, mas uma solução de mínimos quadrados [9].

Essa observação é importantíssima, pois é necessário entender o significado da solução encontrada quando $\lambda_k \rightarrow 0$.

Além disso, o problema 4.4, nos permite reescrever a equação 4.3 numa forma ligeiramente diferente, mas que sugere uma nova interpretação

$$x^{k+1} = x^k + \lambda_k D \nabla L_i(x^k), \quad (4.5)$$

sendo

$$L_i(x) = \left(\frac{b_i - \langle a^i, x \rangle}{\|a^i\|} \right)^2$$

e

$$D = \text{diag}(a_{i1}, a_{i2}, \dots, a_{in}).$$

A equação 4.5 pode ser entendida não como “projeções relaxadas”, mas como um tipo de *Método do Gradiente*. E, essa foi, segundo De Pierro [5], a verdadeira inspiração para o RAMLA.

Há ainda uma generalização do ART, conhecida como Block-ART [8].

$$x^{k+1} = x^k + \lambda_k \sum_{i=1}^m w_i^k \left(\frac{b_i - \langle a^i, x^k \rangle}{\|a^i\|^2} \right) a^i,$$

onde $\varepsilon_1 \leq \lambda_k \leq 2 - \varepsilon_2$, $\varepsilon_1, \varepsilon_2 > 0$ e w^k são vetores cujos elementos são ou zero ou um. Os vetores w^k determinam quais hiperplanos entrarão na projeção ortogonal em cada iteração. Isso dá uma grande flexibilidade ao algoritmo, pois os “blocos” podem ser de tamanho fixo ou variável.

Para outras “versões” do ART, veja, por exemplo, [7]. Para implementação computacional, consulte [35].

A simplicidade da fórmula 4.2 é um dos pontos positivos do ART. Em cada nova iteração, projeta-se a imagem em um determinado hiperplano. Dessa forma, pouca memória (comparado a outros métodos) é necessária e o número de operações também não é elevado. O problema é que são necessárias muitas iterações para se obter uma imagem com boa qualidade.

O ART foi o primeiro método iterativo aplicado a reconstrução tomográfica (CT), e Herman e Meyer [35] sugeriram sua aplicação em PET, mas nos últimos anos, surgiram novas idéias. Algumas ainda inspiradas no paradigma do ART; outras, entretanto, buscaram outros paradigmas, como por exemplo, o nosso próximo assunto: o EM (Expectation-Maximization).

Capítulo 5

Máxima Verossimilhança: EM e RAMLA

Grosso modo, Máxima Verossimilhança é uma técnica de Estimação, originariamente proposta por Fisher ([23],[24]).

Seja B uma variável aleatória cuja *função de densidade de probabilidade* (f.d.p) é $f(b; x_1, \dots, x_n)$, onde x_1, \dots, x_n são n parâmetros que precisam ser estimados. Dadas m amostras independentes da variável aleatória B , b_1, \dots, b_m , a Função de Verossimilhança é a probabilidade de obter b como dados se a imagem for x , ou seja,

$$P(b/x) = \prod_{i=1}^m f(b_i; x_1, \dots, x_n). \quad (5.1)$$

O objetivo é encontrar os parâmetros x_1, \dots, x_n de tal forma a maximizar a função $P(b/x)$. Para facilitar as contas maximiza-se o logaritmo, ou seja,

$$L(x) = \sum_{i=1}^m \ln(f(b_i; x)),$$

onde $x = (x_1, \dots, x_n)$.

A estimativa de máxima verossimilhança x^* pode ser obtida resolvendo-se as equações de verossimilhança abaixo

$$\frac{\partial L(x)}{\partial x} = 0,$$

que, na verdade, corresponde a um conjunto de n equações, em geral, não lineares.

Há alguns comentários importantes. O primeiro é que a função de verossimilhança pode conter muitos pontos estacionários; portanto, na melhor das hipóteses a solução da equação de verossimilhança nos fornece um máximo local. No caso particular em que $L(x)$ for côncava, então todo ponto estacionário é um máximo

global (veja o teorema 2.5 de [62]). Mesmo no caso em que $L(x)$ for côncava, a maximização pode ser uma tarefa delicada. Basta imaginar que o máximo esteja numa região quase plana (flat).

5.1 Exemplos de Funções de Verossimilhança em Tomografia

Como a emissão e a detecção de fótons são processos de Poisson, é razoável pensar em um tratamento estatístico para o problema.

Em 1976, Rockmore e Macovski [77], foram os primeiros a sugerir maximizar a função de verossimilhança para se obter uma estimativa da atividade de radioisótopos dentro do corpo de um paciente. O novo modelo foi o começo de uma notável melhora na qualidade das imagens em ECT. Lange e Carson [49] sugeriram o mesmo para CT de baixa estatística.

5.1.1 Verossimilhança em ECT

Vejam a função de verossimilhança para o caso da Tomografia por Emissão. Nesse caso, temos b_1, b_2, \dots, b_m amostras observadas da variável aleatória B cuja p.d.f é Poisson com média

$$\langle a^i, x \rangle = \sum_{j=1}^n a_{ij} x_j \quad i = 1, \dots, m,$$

onde a^i denota a i -ésima linha. Portanto, $f(b_i; x_1, \dots, x_n) = \frac{\langle a^i, x \rangle^{b_i}}{b_i!} \exp(-\langle a^i, x \rangle)$.

$$L(x) = \ln \left(\prod_{i=1}^m \frac{\langle a^i, x \rangle^{b_i}}{b_i!} \exp(-\langle a^i, x \rangle) \right).$$

Que desprezando-se as constantes, pode ser escrita como

$$L(x) = \sum_{i=1}^m b_i \ln \langle a^i, x \rangle - \langle a^i, x \rangle. \quad (5.2)$$

5.1.2 Verossimilhança em CT

Já em CT, a média do vetor aleatório B será

$$d_i \exp(-\langle a^i, x \rangle),$$

onde d_i são as medições do “blankscan”. Logo,

$$f(b_i; x_1, \dots, x_n) = \frac{(d_i \exp(-\langle a^i, x \rangle))^{b_i}}{b_i!} \exp(-d_i \exp(-\langle a^i, x \rangle)).$$

Daí,

$$L(x) = \ln \left(\prod_{i=1}^m \frac{(d_i \exp(-\langle a^i, x \rangle))^{b_i}}{b_i!} \exp(-d_i \exp(-\langle a^i, x \rangle)) \right),$$

analogamente, desprezando-se as constantes,

$$L(x) = \sum_{i=1}^m -d_i \exp(-\langle a^i, x \rangle) - b_i \langle a^i, x \rangle. \quad (5.3)$$

5.2 EM

Primeiramente é preciso deixar claro que o EM é, na verdade, uma *metodologia de otimização*. Em outras palavras, propõe apenas um esquema de otimização que é dividido em duas partes. A primeira parte diz respeito a uma fase de estimação (Expectation step), conhecida como fase *E*, e a segunda parte uma fase de otimização (Maximization step), denotada como fase *M*.

Entretanto, tanto na fase *E*, quanto na fase *M*, nenhum algoritmo em especial é especificado. É nesse sentido que dizemos que o EM é uma metodologia de otimização. Daí a quantidade de algoritmos EM aplicados nas mais diferentes áreas da ciência [57]. Antes mesmo de seu surgimento oficial¹ em 1977 no trabalho de Dempster, Laird e Rubin [17], o EM já era associado de forma informal a um número considerável de aplicações diferentes.

Em tomografia, Shepp e Vardi [78] em 1982, e Lange e Carson [49] em 1984, independentemente, propuseram o algoritmo EM para maximizar a função de verossimilhança em tomografia seja por Emissão seja por Transmissão.

Sem entrar em muitos detalhes, o EM em ECT (CT será discutido no Capítulo 7) pode ser assim descrito. Dada uma imagem inicial x^0 positiva, calcule x_j^{k+1} , usando a fórmula

$$x_j^{k+1} = \frac{x_j^k}{\sum_{i=1}^m a_{ij}} \sum_{i=1}^m \frac{b_i a_{ij}}{\langle a^i, x^k \rangle}, \quad (5.4)$$

onde $j = 1, \dots, n$ e n é o número de pixels na imagens. Desnecessário dizer que essa fórmula contém a fase *E* e a fase *M* simultaneamente.

A fórmula acima surge, segundo Meinel [58], pela primeira vez, nos trabalhos independentes de Richardson [76] e Lucy [55] sobre restauração de imagens. É interessante notar que a equação de Richardson-Lucy é um caso particular (quando $p = 1$) da fórmula

$$x_j^{k+1} = \frac{x_j^k}{\sum_{i=1}^m a_{ij}} \left[\sum_{i=1}^m \frac{b_i a_{ij}}{\langle a^i, x^k \rangle} \right]^p. \quad (5.5)$$

¹É preciso enfatizar que o “esquema” do EM já era empregado antes do trabalho de Dempster et. al. Contudo, só após esse trabalho, houve a unificação sob o nome de EM.

Segundo Meinel [58], para valores de $p > 1$, observa-se experimentalmente uma melhor taxa de convergência; embora a estabilidade seja prejudicada. Em 1987, Eiichi Tanaka [81] utilizou essa idéia para acelerar o EM. Segundo Tanaka, a convergência é melhorada desde que os parâmetros envolvidos sejam escolhidos de “forma adequada”. Entretanto, a aceleração obtida pelo algoritmo de Tanaka não chega a uma ordem de grandeza. Veremos mais tarde que existem outras formas de acelerar o EM em, pelo menos, uma ordem de grandeza.

A convergência do EM só foi definitivamente provada em 1984 no trabalho de Csiszar e Tusnady [16].

De Pierro, em [69], chega na fórmula 5.4 sem utilizar nenhum argumento estatístico. A idéia foi construir uma função $Q(x, x^k) \leq L(x)$ (sendo que a igualdade só é satisfeita quando $x = x^k$). Assim, De Pierro aproximou a máxima verossimilhança, maximizando $Q(x, x^k)$, e a fórmula para maximizar $Q(x, x^k)$ é igual a 5.4. Essa idéia inspirou e ainda inspira muitos trabalhos em Tomografia (ECT e CT), só para citar um dos mais recentes veja [19]. O segredo consiste em desacoplar as incógnitas, facilitando a maximização da função objetivo.

Já é sabido que o EM gera imagens com qualidade superior às imagens obtidas via FBP [79]. Entretanto, são necessárias muitas iterações para se obter tais imagens. Dessa forma, as pesquisas começaram a buscar formas de acelerar o EM.

5.3 OS-EM

O primeiro algoritmo que realmente obteve sucesso em acelerar² o EM, e que foi adotado pelos fabricantes de ‘scanners’, foi o OS-EM (Ordered Subsets - EM) [39].

Proposto por Hudson e Larkin, em 1994, esse algoritmo, segundo os próprios autores, segue a filosofia “*dividir para conquistar*”. Embora a analogia³, da forma como foi feita, com os métodos de Jacobi e Gauss-Seidel para resolver sistemas lineares tenha sido infeliz, o OS-EM, realmente, acelera o EM consideravelmente (em pelo menos uma ordem de grandeza), assim como o Gauss-Seidel acelera o método de Jacobi.

A parte do “dividir” se refere a uma partição no conjunto de equações que compõe o sistema linear $Ax = b$. Suponhamos que existam m equações. Tomemos uma partição $\pi = \{S_1, S_2, \dots, S_p\}$ composta por p subconjuntos do conjunto $M = \{1, 2, 3, \dots, m\}$.

Feito isso, dada uma imagem x^0 positiva, o algoritmo é assim definido

$$x_j^{(k,i+1)} = \frac{x_j^{(k,i)}}{\sum_{t \in S_i} a_{tj}} \sum_{t \in S_i} \frac{b_t a_{tj}}{\langle a^t, x^{(k,i)} \rangle}, \quad j = 1, \dots, n \text{ e } i = 0, \dots, p-1 \quad (5.6)$$

sendo que $x^{(k,p)} = x^{k+1}$ e $x^{(k,0)} = x^{k-1}$.

²Além do algoritmo sugerido por Tanaka, outras formas tradicionais de aceleração como “line search” foram propostas em [42] por Linda Kaufman e em [54] por Lewitt e Muehllehner, técnicas baseadas em “Vector Extrapolation” [75] ou gradientes conjugados [45, 46, 60, 59] e outras formas alternativas como o SAGE de Fessler e Hero [21] e “Coordinate Ascent approach” de Bouman e Sauer [3].

³“The difference between Vardi-Shepp and ordered subsets EM seems similar, conceptually, to the difference between Jacobi and Gauss-Seidel algorithms for solutions of linear equations”

A ordem em que os subconjuntos ou “blocos” S_i são tomados pode variar. Hudson e Larkin sugerem tomar subconjuntos de maneira que dois subconjuntos sucessivos possuam equações cujas direções dos respectivos raios (fótons) sejam aproximadamente ortogonais. Dada a importância da ordem escolhida esse tipo de algoritmo também é denominado de *algoritmos de subconjuntos ordenados*.

Em [39], os autores provaram um resultado de convergência para o caso consistente, já encontrado anteriormente em [67] que não é o caso de interesse (sem ruído).

De Pierro e Browne [5], no apêndice B, mostraram com um contra-exemplo simples que para o caso inconsistente (que é o mais interessante) o OS-EM não converge. Pior: gera p pontos estacionários e fica oscilando entre eles.

Ou seja, o OS-EM não encontra a solução de máxima verossimilhança. Geralmente, pára em valores bem inferiores ao máximo procurado. Matematicamente falando: não encontra a solução do problema. Como se trata de um problema mal condicionado, isto não é uma desvantagem muito grande, mas a alternância entre subconjuntos produz oscilações que são evitadas por RAMLA.

5.4 RAMLA

O RAMLA (Row-Action Maximum Likelihood Algorithm) [5] não foi inspirado no OS-EM como uma primeira leitura pode sugerir; embora utilize um princípio semelhante ao OS-EM, que é dividir as equações do sistema em blocos, ou seja, também é um algoritmo de subconjuntos ordenados.

O RAMLA nasce a partir de um paradigma encontrado no trabalho Herman e Meyer [35] sobre ART.

Nesse trabalho, Herman e Meyer compararam o desempenho do ART e do EM com relação a uma tarefa específica. Verificaram que uma escolha apropriada da seqüência de projeções ortogonais melhoravam o desempenho do ART, obtendo uma solução aceitável num tempo muito menor que o EM. Resumidamente, o que eles sugeriram foi escolher hiperplanos aproximadamente ortogonais em iterações sucessivas. Isso é bastante razoável. Pensemos num exemplo simples de um sistema 2×2 , onde as equações são retas perpendiculares. Usando ART, partindo de um ponto x^0 qualquer, em duas iterações chegamos à solução. É essa a idéia por detrás da sugestão de Herman e Meyer.

De Pierro e Browne se perguntaram se existiria um algoritmo, que tomando uma linha por vez (Row-Action⁴) a exemplo do ART, maximizasse a função de verossimilhança em ECT. A resposta foi o RAMLA.

5.4.1 O Algoritmo

Como tomar uma linha do sistema por vez, e procedendo dessa forma maximizar a função de verossimilhança? Consideremos as observações abaixo.

Primeira observação: na equação 4.5 vemos o ART como uma espécie de “Método do Gradiente” aplicado em forma seqüencial, com um ‘scaling’ adequado.

Segunda observação: o algoritmo EM é também um tipo de “Método do Gradiente”, com um ‘scaling’ diagonal muito particular, mas sem a decomposição em blocos. Vamos reescrever 5.4,

⁴Y. Censor em [6] “patentou” os métodos denominados “Row-action”.

$$x_j^{k+1} = x_j^k + \frac{x_j^k}{\sum_{i=1}^m a_{ij}} \sum_{i=1}^m a_{ij} \left(\frac{b_i}{\langle a^i, x^k \rangle} - 1 \right),$$

em forma vetorial

$$x^{k+1} = x^k + D(x^k) \nabla L(x^k),$$

onde

$$D(x) = \text{diag} \left(\frac{x_1}{\sum_{i=1}^m a_{i1}}, \dots, \frac{x_n}{\sum_{i=1}^m a_{in}} \right).$$

A partir dessas duas observações, o RAMLA foi definido de forma análoga como

$$x_j^{k+1} = x_j^k + \lambda_k x_j^k \left(\frac{b_i}{\langle a^i, x^k \rangle} - 1 \right) a_{ij}, \quad (5.7)$$

assim, dada x^0 uma imagem inicial positiva, tomando-se linha por linha, em uma ordem pre-estabelecida, e com as seguintes condições sobre $\lambda_k > 0$

$$\lambda_k a_{ij} \leq 1 \forall i, j, \lambda_k \rightarrow 0 \text{ e } \sum_{k=0}^{\infty} \lambda_k = \infty$$

o algoritmo gera a seqüência de imagens $\{x^k\}$ que converge para a máxima verossimilhança.

A fórmula 5.7, segundo De Pierro, aparece pela primeira vez em [83], mas sem relaxação e sem estabelecer nenhuma relação com a verossimilhança Poisson. Em [67] é tratada a convergência no caso consistente e apresentada uma generalização multiplicativa.

Veja que podemos escrever a equação 5.7 como

$$x^{k+1} = x^k + \lambda_k D(x^k) \nabla L_{i_k}(x^k),$$

sendo que

$$D(x) = \text{diag}(x_1, \dots, x_n)$$

e

$$L_{i_k}(x) = b_i \ln \langle a^{i_k}, x \rangle - \langle a^{i_k}, x \rangle.$$

i_k define a forma como é escolhida a ordem dos blocos. Note que a função de verossimilhança em ECT é

$$L(x) = \sum_{i=1}^m L_i(x).$$

Uma generalização imediata de 5.7 é tomar um bloco de linhas por vez ao invés de apenas uma única linha. Nesse caso, as modificações são

$$x_j^{(k,i)} = x_j^{(k,i-1)} + \lambda_k x_j^{(k,i-1)} \sum_{l \in S_{i-1}} a_{lj} \left(\frac{b_l}{\langle a^l, x^{(k,i-1)} \rangle} - 1 \right), \quad (5.8)$$

onde $j = 1, \dots, n$ e $i = 1, \dots, p$. Cada S_i é um subconjunto da partição de $I = \{1, 2, \dots, m\}$. As condições sobre λ_k serão

$$0 < \lambda_k \sum_{l \in S_i} a_{lj} \leq 1, \quad \lambda_k \rightarrow 0 \text{ e } \sum_{k=0}^{\infty} \lambda_k = \infty.$$

A fórmula 5.8 para o caso particular de um único bloco, ou seja, todas as linhas de uma única vez, aparece em [58].

A primeira observação importante sobre 5.8 é que o OS-EM é um caso particular do RAMLA, bastando tomar

$$\lambda_j^{(k,i)} = \frac{1}{\sum_{l \in S_i} a_{lj}}.$$

Outra observação interessante é que mesmo introduzindo relaxação no OS-EM, ainda assim não convergiríamos para a solução de máxima verossimilhança, mas para uma solução “ponderada” da verossimilhança [72].

De Pierro e Browne provaram a convergência do RAMLA com algumas hipóteses (sobre λ_k e sobre a sequência $L(x^k)$). Tais hipóteses, embora fortes, são necessárias devido ao problema relacionado à fronteira do domínio. Caso o máximo estivesse no interior, provamos que não são necessárias tais hipóteses.

5.4.2 OS-EM versus RAMLA

Os dois algoritmos são bem mais rápidos que o EM puro, portanto, de saída, já têm uma vantagem enorme sobre o EM. Também utilizam a mesma estratégia de dividir as equações em blocos ou “subconjuntos ordenados”. Entretanto, o RAMLA tem garantido um resultado de convergência; enquanto o OS-EM simplesmente não converge. Aí reside uma das principais diferenças entre os dois algoritmos.

Embora, por questões óbvias, o número de iterações seja limitado, na prática, a poucas iterações, a convergência não é somente uma simples propriedade matemática inútil. Essa diferença garante, por exemplo, que o RAMLA atinja valores de verossimilhança muito superiores aos atingidos pelo OS-EM já nas primeiras iterações.

Tanto no OS-EM quanto no RAMLA, à medida que as iterações vão aumentando, o ruído vai ficando

maior que o sinal⁵, a imagem vai ficando cheia de “*sal e pimenta*”. Surge um problema interessante: quando se deve parar o algoritmo? Qual é o momento em que o ruído começa a suplantar o sinal?

Há, pelo menos, duas formas diferentes de resolver esse problema. Ou se procura um critério de parada, e esse critério deve de alguma forma detectar o momento certo de parar, ou seja, parar antes que a qualidade da imagem seja comprometida pelo ruído; ou muda-se o modelo, incorporando informação adicional sobre o problema. Uma forma de mudar o modelo incorporando nova informação é regularizar a função objetivo. Isto coloca a questão de estender os algoritmos tipo RAMLA para situações mais gerais, que incluam a regularização. Nos próximos capítulos apresentaremos o modelo regularizado e a extensão dos algoritmos para este caso e para o problema de transmissão, assim como alguns resultados de convergência.

⁵O máximo da função de verossimilhança está numa região chata rodeada de “imagens” cheias de altas frequências. e onde a probabilidade de encontrar uma imagem suave é muito baixa devido à dimensão do problema

Capítulo 6

O Modelo Regularizado e BSREM

“ If Newton and Leibniz had thought that continuous functions do not necessarily have a derivative - and this is the general case - the differential calculus would never have been created.” (Emile Picard)

“ The series is divergent; therefore we may be able to do something with it.” (Oliver Heaviside)

Terminamos o Capítulo anterior deixando no ar a idéia de regularizar o modelo. É claro que a justificativa dada no final do Capítulo anterior é válida, mas não é a principal. A razão principal é incorporar informações ao modelo. É por isso que a regularização também é conhecida na literatura como *abordagem Bayesiana* [52] ou modelo MAP (Maximum A Posteriori).

Usar informações a priori contidas no termo de regularização é uma forma elegante de estabilizar um problema inverso. Em outras palavras, ao invés de maximizar $P(b/x)$, ou seja, a probabilidade de se obter o vetor b dado o vetor x , maximizamos $P(x/b)$ via fórmula de Bayes

$$P(x/b) = \frac{P(x \& b)}{P(b)} = \frac{P(b/x) P(x)}{P(b)}, \quad (6.1)$$

onde $P(x)$ e $P(b)$ são as probabilidades de x e b , respectivamente. Tomando-se o logaritmo e desprezando os termos que não dependem de x ,

$$G(x) = \ln(P(x/b)) = L(x) + \gamma R(x),$$

onde $L(x) = \ln(P(x/b))$ é a nossa função de verossimilhança e $\gamma R(x) = \ln(P(x))$ é o termo que incorpora as informações a priori sobre x , ou, simplesmente, o termo de regularização. γ é um parâmetro real positivo que mede a importância relativa de satisfazer a consistência com os dados versus a informação ‘a priori’. Portanto, nosso problema é maximizar a função $G(x)$, isto é,

$$\max_{x \geq 0} (L(x) + \gamma R(x)). \quad (6.2)$$

O termo de regularização $\gamma R(x)$, também conhecido como *prior*.

6.1 Priors

A escolha da função $R(x)$ não é arbitrária. Há vários fatores envolvidos nessa escolha. Desde as características da imagem (se é suave ou não) até as propriedades de $R(x)$, como por exemplo, continuidade, concavidade, positividade, etc.

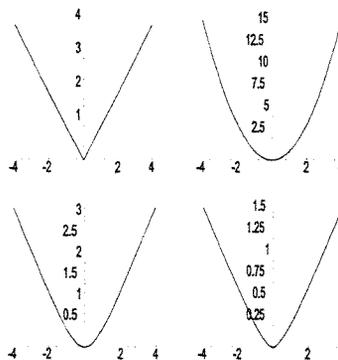
Existem diversas referências importantes sobre a escolha dos priors. Mas, nossos trabalhos foram influenciados principalmente por dois deles [47] e [33]. No primeiro, Lange faz um excelente apanhado dos priors mais utilizados, além de apresentar suas vantagens e, eventuais, desvantagens no caso da Tomografia Computadorizada (seja em emissão ou transmissão).

De uma forma geral, os Priors usados na prática (Priors de Gibbs) têm a seguinte forma

$$R(x) = - \sum_{j,k \in N} w_{jk} r(x_j - x_k),$$

onde γ e w_{jk} são constantes positivas, e $r(x)$ é denominada função potencial. Veja a seguir alguns exemplos de funções potencial

Função Potencial $r(x)$	Referência
$r(x) = x^2$	Hebert e Leahy [33]
$r(x) = \frac{x^2}{1+x^2}$	Geman e McClure [26],[27]
$r(x) = \ln(1+x^2)$	Hebert e Leahy [33]
$r(x) = \ln(\cosh(x))$	Green [31]
$r(x) = \frac{1}{2} \left(x + \frac{1}{1+ x } - 1 \right)$	Lange [47]
$r(x) = x - \ln(1+ x)$	Lange [47]
$r(x) = x + \exp(- x) - 1$	Lange [47]



Da esquerda para direita e de cima para baixo, temos $r(x) = |x|$, $r(x) = x^2$, $r(x) = \ln(\cosh(x))$ e $r(x) = \frac{1}{2} \left(|x| + \frac{1}{1+|x|} - 1 \right)$.

Vejamos alguns critérios que levamos em conta para a escolha da função potencial.

Em primeiro lugar, $r(x)$ não pode penalizar muito as altas frequências, pois suavizaríamos demais a imagem, causando o “*over smoothing*” ($r(x) = x^2$, por exemplo). No caso de funções $r(x)$ convexas o ideal seria utilizar $r(x) = |x|$; entretanto, essa função não é diferenciável (propriedade importante do ponto de vista da análise matemática).

Assim, entre as funções $r(x) = x^2$ e $r(x) = |x|$, em nossos trabalhos, tendo em conta que não era nosso objetivo comparar formas de regularização, mas algoritmos, optamos por uma boa aproximação diferenciável da segunda, que é a função $r(x) = \ln(\cosh(x))$, sugerida por P. Green [32]. Ela é diferenciável, convexa e não penaliza demasiadamente as altas frequências. Contudo, não pretendemos defender a função $r(x) = \ln(\cosh(x))$ como o “melhor” prior. É óbvio que dependendo da aplicação, outras funções podem ser mais indicadas. Há inclusive na literatura muitos defensores do uso de ‘priors’ mais gerais, (não convexas) [61].

Procuramos utilizar priors não convexas em nossos algoritmos. Entretanto, apesar dos bons resultados nas simulações, nos deparamos com uma grande dificuldade matemática: tínhamos um problema de otimização global para resolver. Inspirados em [61], fizemos um trabalho com priors não convexas em Tomografia por Emissão [71] com bons resultados comparados aos priors convexas.

Temos repetidas vezes falado em “suavização” da imagem nos referindo a regularização, entretanto qualquer informação sobre a imagem pode ser incorporada ao modelo em forma de priors.

Vejamos alguns algoritmos para o modelo regularizado.

6.2 OS-GP

Hudson e Larkin no mesmo trabalho [39] que apresentaram o OS-EM, também sugeriram um algoritmo denominado OS-GP (embora não mencionado no texto, acreditamos que venha de “Orderer Subsets-Gibbs Priors”, pois o algoritmo usa como termo regularizador os priors de Gibbs) para o modelo regularizado. Esse algoritmo é uma adaptação do OSL (One-Step Late) [32] proposto por Peter Green. Mais precisamente: consiste em aplicar a idéia de subconjuntos ordenados ao OSL. O algoritmo pode ser assim descrito.

Dada a imagem inicial x^0 positiva, e feita a divisão em p subconjuntos ordenados das equações,

$$x_j^{(k,i)} = \frac{x_j^{(k,i-1)}}{\sum_{l \in S_{i-1}} a_{lj} + \gamma \frac{\partial R(x^{(k,i-1)})}{\partial x_j}} \sum_{l \in S_{i-1}} \frac{b_l a_{lj}}{\langle a^l, x^{(k,i-1)} \rangle}, \quad (6.3)$$

onde $x^{(k,p)} = x^{k+1}$ e a função $R(x)$ sendo um Gibbs prior.

6.3 BSREM

Nossa primeira contribuição em termos de algoritmos rápidos em Tomografia Computadorizada foi pensando no modelo regularizado.

Apresentamos em [72] um novo algoritmo rápido para o modelo regularizado: o BSREM (Block Sequential Regularized Expectation Maximization). Esse algoritmo é uma extensão do RAMLA.

A principal dificuldade foi decidir qual a melhor forma de lidar com a parte de regularização.

6.3.1 Primeira Versão

A primeira abordagem foi tratar a regularização de forma separada, utilizando a idéia usada por De Pierro em [69] na parte de regularização. Isto é, desacoplar as variáveis na parte da regularização, utilizando uma função $W(x, x^k) \leq R(x)$. Essa idéia se mostrou infrutífera, seja do ponto de vista da implementação computacional (muitas contas necessárias para determinar a função $W(x, x^k)$) seja do ponto de vista matemático (não conseguimos nenhum resultado de convergência). Isso nos levou a procurar uma forma mais simples e eficiente de lidar com a regularização.

6.3.2 Versão Final

Após diversas versões, a opção mais simples, que é dar a regularização o mesmo tratamento dado à função objetivo, se mostrou a mais apropriada. Primeiro porque os resultados de convergência obtidos para o RAMLA foram naturalmente estendidos (exceto pela perda da positividade) para a versão regularizada. Segundo porque as simulações forneceram excelentes imagens, e a qualidade dessas imagens confirmadas por medidas de desempenho (valor da função objetivo e outras “figures of merit”).

O algoritmo pode ser assim descrito.

Dada a imagem x^0 positiva, calcule

$$x_j^{(k,i)} = x_j^{(k,i-1)} + \lambda_k x_j^{(k,i-1)} \sum_{l \in S_{i-1}} a_{lj} \left(\frac{b_l}{\langle a^l, x^{(k,i-1)} \rangle} - 1 \right), \quad (6.4)$$

seja $x^{k+\frac{1}{2}} = x^{(k,p)}$

$$x^{k+1} = x^{k+\frac{1}{2}} + \lambda_k \gamma D(x^{k+\frac{1}{2}}) \nabla R(x^{k+\frac{1}{2}}). \quad (6.5)$$

Em outras palavras, as fórmulas 6.4 e 6.5, na verdade, poderiam ser resumidas em uma única fórmula, sendo que o termo de regularização seria uma parte da função objetivo. Neste caso, nada impede que também na regularização se usem “blocos”.

É claro que quando $\gamma = 0$ o algoritmo resultante é RAMLA e a positividade está garantida. Quando $\gamma \neq 0$ algum tipo de controle é necessário para que x^{k+1} seja estritamente positivo em 6.5. Na prática isto não acontece (ou seja, para valores razoáveis de γ as coordenadas do novo iterado permanecem positivas), e, de qualquer maneira, sempre é possível manter os iterados positivos acrescentando um termo logarítmico do tipo $-\epsilon \sum_{j=1}^n x_j$ com ϵ muito pequeno.

Priors Não-Côncavos

O BSREM tal como foi descrito acima é suficientemente flexível no que diz respeito a escolha da função $R(x)$. O que aconteceria se escolhêssemos priors não côncavos? Em primeiro lugar, como já mencionamos

anteriormente, esbarraríamos em um problema de Otimização Global, que todos sabem ser extremamente difícil de resolver (salvo em casos muito particulares).

Nesse caso, teríamos que nos contentar com a convergência para um ponto crítico (que pode ou não ser o máximo global). Para verificar se esse ponto crítico é razoável, implementamos duas versões do BSREM. A primeira versão foi o BSREM com o prior não côncavo. A outra versão foi uma modificação do BSREM, denominado BSREM-GNC, inspirada no método conhecido como *Graduated Non Convexity (GNC)* [61] utilizado por Nikolova et. al.

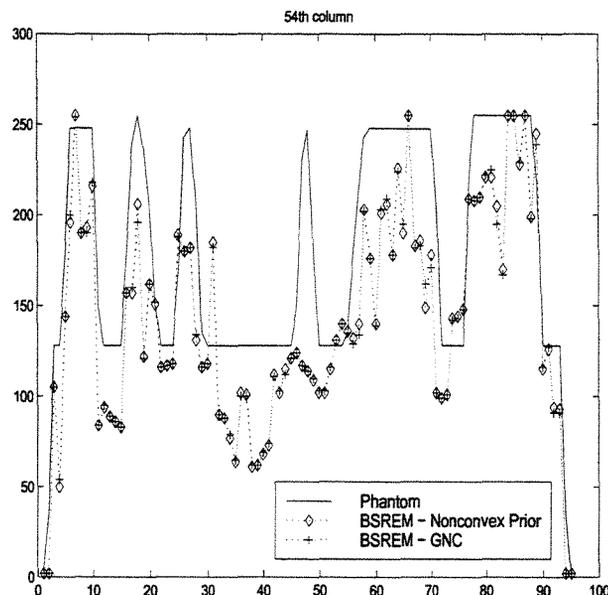
O GNC tem como idéia principal “convexificar” o prior não convexo (côncavo) somando um termo convexo (côncavo). Ou seja, seja $R(x)$ o prior não convexo (côncavo), o GNC adiciona um termo convexo (côncavo) a $R(x)$, e conforme as iterações vão passando esse termo vai tendendo a zero.

$$R_\alpha(x) = R(x) + \alpha x^2,$$

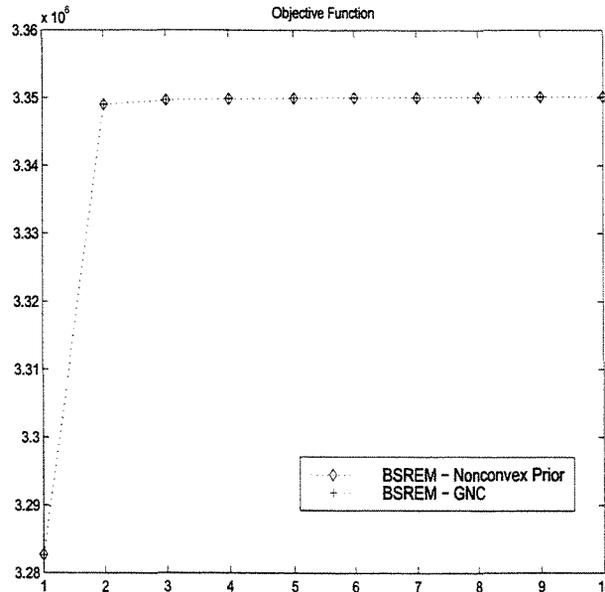
onde $\alpha \rightarrow 0$ com o passar das iterações.

Nosso objetivo era comparar o BSREM sem modificação com o BSREM-GNC. Os resultados [71] nos mostraram que o BSREM sem modificação se comportou de forma satisfatória comparado ao BSREM modificado, atingindo valores de verossimilhança equivalentes e gerando imagens de boa qualidade.

A figura abaixo compara pixel a pixel a 54ª coluna do phantom e das imagens reconstruídas pelos BSREM e BSREM-GNC.



Em relação a função de verossimilhança, os dois algoritmos têm desempenhos equivalentes. Veja figura a seguir.



A seguir as imagens obtidas com os dois algoritmos.



Esses resultados demonstraram a grande flexibilidade do BSREM. Mesmo utilizando priors não convexos (côncavos), ele gera imagens de boa qualidade e atinge altos níveis de verossimilhança.

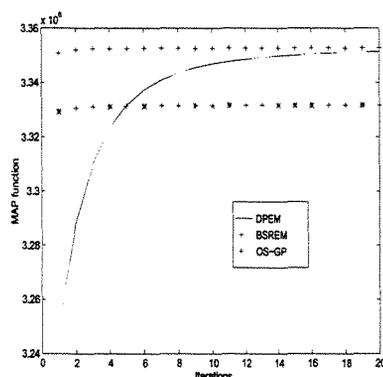
6.4 OS-GP versus BSREM

No Apêndice C encontramos uma cópia de [72] onde fazemos uma comparação de desempenho detalhada entre o BSREM e o OS-GP. Entretanto, resumidamente, podemos mencionar as seguintes diferenças entre os dois algoritmos.

- A convergência do BSREM é provada sob hipóteses ligeiramente diferentes daquelas necessárias para o RAMLA (a principal diferença é que não é possível assegurar a positividade, embora nos experimentos nunca tenha ocorrido um termo não positivo, tal como mencionado anteriormente). Já o OS-GP simplesmente não converge como foi demonstrado com um contra-exemplo em [72]. Com pequenas modificações em γ (veja Apêndice III em [72]) o OS-GP gera iterandos negativos, mostrando sua

sensibilidade com relação a pequenas variações em γ .

- Como o OS-GP não converge, o BSREM alcança valores maiores de verossimilhança. Para verificar experimentalmente a convergência do BSREM, escolhemos o DPEM [69] que, apesar de mais lento, tem uma prova de convergência completa para comparar os valores de verossimilhança atingidos. Veja o gráfico abaixo. Nele representamos o desempenho do BSREM, do OS-GP e do DPEM. Claramente o BSREM é mais rápido que os outros dois. O OS-GP atinge um valor de verossimilhança inferior tanto ao BSREM quanto ao DPEM. O DPEM assintoticamente aproxima o valor de verossimilhança atingido pelo BSREM.



- O OS-GP em cada subiteração “regulariza” a imagem, enquanto o BSREM “regulariza” apenas uma única vez por iteração. Isso significa que, além de mais operações aritméticas [20], o OS-GP tende a suavizar demais a imagem (over smoothing). O gráfico abaixo ilustra essa afirmação. Compare as imagens obtidas pelo BSREM e pelo OS-GP na vigésima iteração. Esta última é visivelmente mais suave que a primeira.

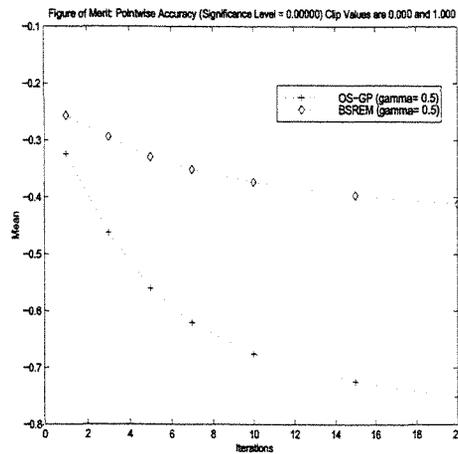


- O BSREM supera o OS-GP com relação a várias FOM (Figure Of Merit). O gráfico abaixo ilustra o desempenho com relação a Pointwise Accuracy, cuja definição é o oposto da “distância normalizada”

entre a imagem reconstruída, r , e o phantom, p , isto é,

$$-\left(\frac{\sum_{j=1}^n (p_j - r_j)^2}{\sum_{j=1}^n (p_j - \hat{p})^2}\right)^{\frac{1}{2}},$$

onde \hat{p} é a média dos pixels do phantom p .



6.4.1 BSREM na Literatura

Mesmo antes de sua publicação na IEEE Trans. Med. Imaging, vários autores já citavam o BSREM como uma boa alternativa em termos de algoritmos iterativos para o modelo regularizado em ECT. Um exemplo disso é o editorial “*Recent Development in Iterative Image Reconstruction for PET and SPECT*” escrito por R. Leahy e C. Byrne na IEEE Trans. Med. Imaging [51].

Outro artigo que cita o BSREM foi escrito por H. Erdogan e J. A. Fessler, “*Ordered subsets algorithms for transmission tomography*” [20]. Erdogan e Fessler afirmam que a nossa técnica de tratar a regularização como parte da função objetivo é uma forma eficiente de diminuir o tempo e o custo computacional por iteração nos algoritmos de subconjuntos ordenados.

Capítulo 7

Novos Algoritmos para Transmissão: T-RAMLA e T-BSREM

7.1 Introdução

Vimos nos capítulos anteriores, como os problemas de reconstrução em Tomografia por Emissão são resolvidos através de métodos iterativos, mencionamos o ART, o EM e algoritmos para acelerar o EM, como por exemplo, o RAMLA.

Neste capítulo tratamos da extensão de RAMLA e BSREM para o problema de transmissão.

Em 1984, Lange e Carson [49] propuseram um algoritmo baseado no EM para o caso do problema de ML em Tomografia por Transmissão.

Entretanto, o passo de Maximização era extremamente difícil, pois envolvia equações transcendentais. Lange e Carson sugeriram algumas aproximações (séries truncadas) para contornar esse problema.

Dez anos depois, Ollinger [65] através de experimentos notou que o algoritmo proposto por Lange e Carson gerava imagens mais “suaves” que a imagem original. Isso levou Ollinger a suspeitar que a aproximação feita por Lange e Carson na fase de maximização gerava um algoritmo que não convergia para o máximo da função de verossimilhança. Por isso, ele propôs que a fase maximização fosse realizada utilizando o método de Newton Relaxado (a relaxação foi introduzida para estabilizar o algoritmo). Dessa forma, o algoritmo de Ollinger, chamado de GEM (generalized EM) garantia o aumento da função de verossimilhança em cada iteração. Porém, tanto a resolução via método de Newton quanto a aproximação por séries truncadas, além de lentas, aumentavam consideravelmente o custo computacional por iteração, como observado em [64]. Dessa forma, outras alternativas começaram a surgir para otimizar a função de ML em Tomografia por Transmissão.

Lange et al. [48], em 1987, sugeriram um algoritmo para maximizar a função de ML que utilizava como direção de subida o gradiente com mudança de escala. Essa idéia encontrou eco em outros trabalhos como por exemplo em [64].

Em 1995, Lange e Fessler [50], dentro do contexto de problemas regularizados, baseado na idéia de De

Pierro [68], apresentaram o *Convex Algorithm*. Segundo os autores, este novo algoritmo, além de computacionalmente superior ao EM, possuía boas propriedades de convergência local e global. Contudo, a principal dificuldade destes algoritmos é a lentidão no aumento da função de verossimilhança e, conseqüentemente, em obter imagens de boa qualidade.

Em 1999, Erdogan e Fessler [20], utilizaram a idéia de subconjuntos ordenados, e propuseram um algoritmo mais rápido, porém não convergente (sem subrelaxação nem o ‘scaling’ apropriado)

Nossa contribuição vai nessa linha de aprimoramento dos algoritmos para maximizar a função de verossimilhança (regularizada) no caso da Tomografia por Transmissão. A idéia é essencialmente a mesma utilizada para o RAMLA, ou seja, utilizando subconjuntos ordenados, em cada iteração, atualizamos a imagem indo na direção relaxada e escalada do gradiente.

7.2 Novo Algoritmo : T-RAMLA

Para descrevermos o novo algoritmo, é necessário recordar as fórmulas das funções de verossimilhança para o problema de Transmissão.

Em Transmissão [50], a função de verossimilhança¹ é

$$L(x) = \sum_i^m -d_i \exp(-\langle a^i, x \rangle) - y_i \langle a^i, x \rangle \quad (7.1)$$

,onde d é o vetor contendo as medições de controle². É bom lembrar que esta função é côncava.

Primeiramente, tomamos p subconjuntos disjuntos S_i , $i = 1, \dots, p$, formados a partir de uma partição do intervalo de números inteiros $M = [1, 2, 3, \dots, m]$. A função de verossimilhança pode dessa forma ser reescrita como

$$L(x) = \sum_{i=1}^p L_i(x),$$

onde

$$L_i(x) = \sum_{k \in S_i} -d_k \exp(-\langle a^k, x \rangle) - y_k \langle a^k, x \rangle.$$

Vamos calcular o gradiente das funções $L_i(x)$

$$(\nabla L_i(x))_j = \frac{\partial L_i(x)}{\partial x_j} = \sum_{k \in S_i} a_{kj} d_k \exp(-\langle a^k, x \rangle) - a_{kj} y_k \quad (7.2)$$

Com essas informações, denotando a imagem inicial por $x^{(0)}$ e sendo k e i os índices para uma iteração completa e para as subiterações, respectivamente, podemos definir o novo algoritmo como

¹Neste modelo, são ignorados os “background events” [20].

²Na literatura, estas medições são conhecidas como “blank scan”. Assim, d_i é o número esperado de fótons emitidos pela fonte ao longo da direção i [50]. Na prática, para se obter o vetor b é necessário uma seção de transmissão com o scanner sem nenhum objeto no seu interior, daí o nome “blank scan”.

$$x_j^{(k,i+1)} = x_j^{(k,i)} + \lambda_k x^{(k,i)} \sum_{l \in S_i} a_{lj} d_l \exp\left(-\langle a^l, x^{(k,i)} \rangle\right) - a_{lj} y_l \quad (7.3)$$

onde $j = 1, \dots, n$, $i = 1, \dots, p$ (número de pixels e de sub-blocos, respectivamente) e $x^{(k,0)} = x^{k-1}$ e $x^{(k,p)} = x^{k+1}$. Ou seja, mudança no segundo índice equivale a mudança na escolha do subconjunto de equações; enquanto mudança no primeiro índice equivale a um ciclo completo.

Chamaremos esse novo algoritmo de *T-RAMLA*, *Transmission - Row Action Maximum Likelihood Algorithm*. O nome deixa claro a relação com o RAMLA [5].

7.2.1 Convergência

Provaremos para o T-RAMLA o mesmo resultado de convergência obtido em [5] para o RAMLA. Na verdade, salvo onde a função de verossimilhança aparece explicitamente, a demonstração é exatamente igual, sendo em alguns trechos uma tradução do original.

Positividade

Para obtermos uma sequência de pontos positivos, isto é, $x_j > 0 \forall j$, partimos de um ponto inicial positivo e tomamos λ_k de tal forma a satisfazer a condição abaixo em todas as iterações

$$0 < \lambda_k c_{ij} \leq 1 \text{ (se } c_{ij} \neq 0), \quad (7.4)$$

onde

$$c_{ij} = \sum_{l \in S_i} a_{lj} y_l.$$

Com esta condição satisfeita, podemos reescrever 7.3 da seguinte forma

$$x_j^{(k,i+1)} = x_j^{(k,i)} (1 - \lambda_k c_{ij}) + x_j^{(k,i)} \lambda_k \sum_{l \in S_i} a_{lj} d_l \exp\left(-\langle a^l, x^{(k,i)} \rangle\right). \quad (7.5)$$

O primeiro termo do lado direito da equação é positivo devido a 7.4; o termo seguinte é claramente positivo.

Essa é uma das vantagens do T-RAMLA. Uma escolha apropriada de λ_k resulta em iterandos positivos. Isso já não é possível em alguns outros algoritmos propostos, como por exemplo aquele sugerido em [64], onde os autores apontam a possibilidade de perder a positividade como um dos pontos fracos do seu algoritmo (“*A potential weakness is that it accepts negative values in the solution*”), e no EM para transmissão.

Limitação

Quando se pretende provar qualquer resultado de convergência, o primeiro passo é estudar a limitação da sequência gerada. Aqui surge a primeira necessidade de adaptação da demonstração feita em [5]. Para provar

a limitação da sequência gerada pelo T-RAMLA, aparecerá o gradiente da função de verossimilhança do problema de transmissão, que precisará ser devidamente majorado.

Proposição 1: A sequência gerada por 7.3 é limitada.

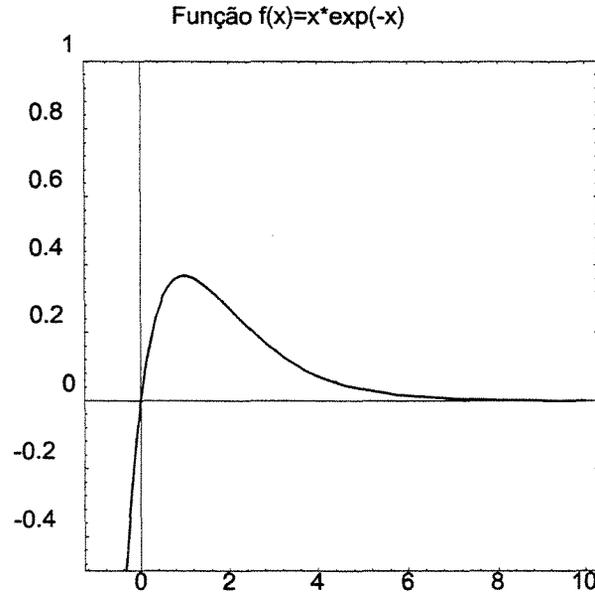
Demonstração: Majorando o segundo termo do lado direito de 7.5, temos

$$x_j^{(k,i)} \sum_{l \in S_i} a_{lj} d_l \exp \left(- \langle a^l, x^{(k,i)} \rangle \right) \leq \sum_{j=1}^n x_j^{(k,i)} \sum_{l \in S_i} a_{lj} d_l \exp \left(- \langle a^l, x^{(k,i)} \rangle \right)$$

ou seja,

$$x_j^{(k,i)} \sum_{l \in S_i} a_{lj} d_l \exp \left(- \langle a^l, x^{(k,i)} \rangle \right) \leq \sum_{l \in S_i} d_l \langle a^l, x^{(k,i)} \rangle \exp \left(- \langle a^l, x^{(k,i)} \rangle \right).$$

Um estudo simples da função $f(x) = x \exp(-x)$ (veja gráfico abaixo) mostra que seu máximo é atingido quando $x = 1$. Dessa forma, $f(x) \leq \exp(-1) < 1$.



Usando essa informação, obtemos

$$x_j^{(k,i)} \sum_{l \in S_i} a_{lj} d_l \exp \left(- \langle a^l, x^{(k,i)} \rangle \right) < \sum_{l \in S_i} d_l = q_i. \tag{7.6}$$

Seja

$$q = \max_i q_i,$$

para λ fixo, majoramos 7.5 da seguinte forma

$$x_j^{(k,i+1)} \leq x_j^{(k,i)} \delta + \eta, \quad (7.7)$$

onde

$$\delta = \max_{c_{ij} \neq 0} (1 - \lambda c_{ij})$$

e

$$\eta = \lambda q.$$

A partir daí os argumentos são idênticos aos de [5], que para *efeito de completitude* serão a seguir reproduzidos.

Para simplificar a notação, substituiremos o duplo índice (k, i) por um único índice k .

Usando 7.7 iterativamente, obtemos

$$\begin{aligned} x_j^{k+1} &\leq x_j^k \delta + \eta \leq x_j^0 \delta^k + \eta \sum_{l=0}^k \delta^l \\ &\leq x_j^0 + g(\lambda), \end{aligned} \quad (7.8)$$

onde

$$g(\lambda) = \eta \frac{1}{1 - \delta},$$

levando em conta a positividade dos termos em questão. $g(\lambda)$ tende a zero quando λ tende a zero. Portanto, para $\lambda \in [0, c)$, onde $c = \max \frac{1}{c_{ij}}$, a função $g(\lambda)$ é limitada. Dessa forma, para todo par (k, λ) , $x_j^k(\lambda)$ é limitado. Segue o resultado desejado. ◊

Teorema 1: Se L é estritamente côncava e $\{L(x^k)\}$ é uma seqüência convergente (monótona, por exemplo), então, se $\lambda_k \rightarrow 0$ e $\sum \lambda_k = +\infty$, a seqüência gerada por 7.3 converge para o máximo de L .

Uma vez provada a limitação e positividade da seqüência $\{x^k\}$, a demonstração desse Teorema é a mesma do Teorema 1 encontrado em [5]. Para evitar remeter o leitor ao referido artigo, segue a demonstração tal qual feita por De Pierro.

Antes de demonstrar este teorema, serão necessários diversas proposições auxiliares. Seguindo [5], faremos as demonstrações dessas proposições com a finalidade de obter a demonstração do teorema 1 como resultado final.

Proposição 2: $D\nabla L_i(x)$ e $D\nabla L(x)$ são limitadas e contínuas em C .

Primeiramente, vamos esclarecer a notação: $D = \text{diag}(x_1, x_2, \dots, x_n)$ e $C =$ fecho do envelope convexo de $x^{(k,i)}$, pela proposição 1, C é limitado.

Vamos a demonstração da proposição 2.

$$\begin{aligned} x_j \frac{\partial L_i(x)}{\partial x_j} &= x_j \sum_{l \in S_i} a_{lj} d_l \exp(-\langle a^l, x \rangle) - a_{lj} y_l \leq \sum_{l \in S_i} d_l \langle a^l, x \rangle \exp(-\langle a^l, x \rangle) - a_{lj} x_j y_l \\ &\leq \sum_{l \in S_i} d_l - a_{lj} x_j y_l, \end{aligned} \quad (7.9)$$

de 7.9 e do fato de $x \in C$, que é limitado, concluímos que $D\nabla L_i(x)$ e $D\nabla L_i(x)$ são limitadas em C . A continuidade em C é resultado da continuidade em \mathbb{R}^n . \diamond

Pela proposição 2, o fator que multiplica λ_k em 7.3 é limitado, então como $\lambda_k \rightarrow 0$ obtemos que

$$x^{(k,i)} - x^{(k,i-1)} \xrightarrow{k \rightarrow \infty} 0. \quad (7.10)$$

Usando 7.3, é fácil mostrar que

$$x^{(k,i)} - x^{(k,p)} = \sum_{l=i+1}^p (x^{(k,l-i)} - x^{(k,l)}), \quad (7.11)$$

usando 7.10 e $x^{(k,p)} = x^{k+1}$ na equação acima,

$$x^{(k,i)} - x^{k+1} \xrightarrow{k \rightarrow \infty} 0, \quad i = 1, \dots, p. \quad (7.12)$$

Seja

$$h^k = \sum_{i=1}^p D^{(k,i-1)} \nabla L_i(x^{(k,i-1)}). \quad (7.13)$$

Pela proposição 2 e 7.12,

$$h^k - D^k \nabla L(x^k) \xrightarrow{k \rightarrow \infty} 0. \quad (7.14)$$

Proposição 3: Se $L(x^k)$ converge para algum limite $L^* < \infty$, L estritamente côncava, e x^* é um ponto de acumulação 7.3 tal que $x_j^* \nabla L(x^*)_j = 0$ $j = 1, \dots, n$, então x^* é o máximo de L .

Demonstração: Vamos tomar uma subsequência $\{x^{l_k}\}$ convergente a x^* . Pela própria construção da sequência $x_j^* \geq 0$. Tentaremos provar que $\nabla L(x^*)_j \leq 0$ para todo j , isto é, não há direção de subida a partir de x^* .

Consideremos os seguintes conjuntos

$$J = \{j \in [1, n] \mid \nabla L(x^*)_j > 0\} \quad (7.15)$$

e

$$S = \{x \mid \langle a^l, x \rangle \neq 0 \forall l\}. \quad (7.16)$$

O conjunto S é claramente não vazio, e como $L^* < \infty$ temos que $x^* \in S$. Nosso objetivo é provar que $J = \emptyset$.

Usando a continuidade do gradiente de L em $\mathfrak{R}^n \cap S$, sabemos que, para $j \in J$, existe uma esfera B centrada em x^* com raio $r > 0$, tal que no interior de B , $\nabla L(x)_j$ é maior que uma constante positiva ψ .

Dessa forma,

$$\sum_{i=1}^p \nabla L_i(x^{(k,i-1)})_j = \left(\sum_{i=1}^p [\nabla L_i(x^{(k,i-1)})_j - \nabla L_i(x^k)_j] \right) + \nabla L(x^k)_j, \quad (7.17)$$

o termo entre parênteses no lado direito de 7.17 tende a zero (pela continuidade e 7.12), então para $j \in J$ e l_k suficientemente grande

$$\sum_{i=1}^p \nabla L_i(x^{(k,i-1)})_j \geq \frac{\psi}{2} > 0. \quad (7.18)$$

Por outro lado,

$$\begin{aligned} x_j^{k+1} &= x_j^k \prod_{i=1}^p \left(1 + \lambda_k \nabla L_i(x^{(k,i-1)})_j \right) \\ &= x_j^k \left[1 + \lambda_k \sum_{i=1}^p \nabla L_i(x^{(k,i-1)})_j + O(\lambda_k^2) \right], \end{aligned} \quad (7.19)$$

e, usando 7.18 e o fato que as derivadas são limitadas, concluímos que, para l_k suficientemente grande, o fator que multiplica x_j^k em 7.19 é maior que um para $j \in J$ e conseqüentemente

$$x_j^{l_k+1} > x_j^{l_k}. \quad (7.20)$$

Tomemos k suficientemente grande para que $x^{l_k} \in B$. Seja $t_k = \max \{q < l_k \mid x^q \in B\}$. Faça $t_k = 0$ se $x^q \in B$ para todo $q < l_k$. Então $x^q \in B$ para $t_k + 1 \leq q \leq l_k$ e, para $j \in J$, temos que

$$x_j^{l_k} \geq x_j^{l_k+1} \geq x_j^q \geq x_j^{t_k+1}. \quad (7.21)$$

Se existe um t tal que $t_k = t$ para um k suficientemente grande, isto é, se a seqüência permanece em B para k suficientemente grande, então temos

$$0 = x_j^* = \lim_{k \rightarrow \infty} x_j^{l_k} \geq x_j^{t_k+1} > 0.$$

Concluimos que

$$\lim_{k \rightarrow \infty} t_k = \infty,$$

e segue de 7.21 que $\lim_{k \rightarrow \infty} x_j^{t_{k+1}} = 0$ para todo $j \in J$, e isso implica que $\lim_{k \rightarrow \infty} x_j^{t_k} = 0$, porque $x^{k+1} - x^k \rightarrow 0$. Contudo, $x^{t_k} \notin B$, então existe uma subsequência

$$x_j^{t_k} \rightarrow x_j^{**},$$

tal que $x_j^{**} = 0$ para todo $j \in J$. $x^{**} \in B$ significa que $\|x^* - x^{**}\| \geq r > 0$; ainda mais, $L(x^*) = L(x^{**})$, pois tínhamos assumido que $L(x^k)$ convergia a $L(x^*)$. Portanto, temos dois pontos diferentes satisfazendo a condição de máximo em $R_J = \{x \in \mathfrak{R}_+^n \mid x_j = 0 \text{ para } j \in J\}$. Isso é uma contradição, pois L não pode ter dois máximos em R_J porque L é estritamente côncava. Logo isso implica que $J = \emptyset$. \diamond

Nas proposições 4, 5 e 6, assumimos que $D^k \nabla L(x^k)$ possui a seguinte propriedade: existe $\delta > 0$ tal que para todo k ,

$$\|D^k \nabla L(x^k)\| \geq \delta > 0, \quad (7.22)$$

onde $\|\cdot\|$ denota a norma euclidiana usual.

Proposição 4: Para todo $\eta_1 > 0$, existe $K_1 \in \mathbb{N}$ tal que para $k \geq K_1$

$$\|x^{k+1} - x^k - \lambda_k D^k \nabla L(x^k)\| < \eta_1 \lambda_k \|D^k \nabla L(x^k)\|. \quad (7.23)$$

Demonstração: De 7.3 e 7.13, temos

$$x^{k+1} - x^k - \lambda_k D^k \nabla L(x^k) = \lambda_k (h^k - D^k \nabla L(x^k)). \quad (7.24)$$

As equações 7.13 e 7.14 implicam 7.23. \diamond

A equação 7.23 possui algumas conseqüências de caráter geométrico. Vejamos nas duas proposições seguintes algumas delas.

Proposição 5: i) existe $\delta_1 > 0$ e $\delta_2 > 0$ tais que $\nabla L(x^k)^t D^k \nabla L(x^k) > \delta_1$ e $\|\nabla L(x^k)\| > \delta_2$. ii) Existe $K_2 \in \mathbb{N}$ e $\delta_3 > 0$ tais que para $k \geq K_2$, $(\bar{h}^k)^t \nabla L(x^k) \geq \delta_3$, onde $\bar{h}^k = \frac{x^{k+1} - x^k}{\|x^{k+1} - x^k\|}$.

Demonstração: i) Se existe uma seqüência $\{x^{k_i}\}$ tal que $\nabla L(x^{k_i})^t D^k \nabla L(x^{k_i}) \rightarrow 0$, então $D^k \nabla L(x^{k_i}) \rightarrow 0$, o que contradiz 7.22. O mesmo resultado segue se $\nabla L(x^{k_i}) \rightarrow 0$. ii) Da primeira parte da demonstração, o ângulo θ_k entre $D^k \nabla L(x^{k_i})$ e $\nabla L(x^{k_i})$ deve estar no intervalo abaixo

$$0 \leq \theta_k \leq C_1 < \frac{\pi}{2}, \quad (7.25)$$

onde C_1 é uma constante. Seja ϕ_k o ângulo entre \bar{h}^k e $D^k \nabla L(x^{k_i})$. De 7.23, sabemos que

$$\text{sen}(\phi_k) \leq \eta_1, \quad (7.26)$$

tomando-se η_1 suficientemente pequeno, o resultado segue. \diamond

Proposição 6: Existe K_3 tal que para $k \geq K_3$, teremos

$$L(x^{k+1}) \geq L(x^k). \quad (7.27)$$

Demonstração: As fórmulas 7.14 e 7.22 implicam que existe $\delta_4 > 0$ tal que

$$\|h^k\| \geq \delta_4. \quad (7.28)$$

Consideremos agora,

$$\bar{x}^k = x^k + \tau_k h^k, \quad (7.29)$$

onde τ_k é tal que $\bar{x}^k = \operatorname{argmax}_{\tau \geq 0}(x^k + \tau h^k)$, isto é, \bar{x}^k é o ponto de máximo na direção h^k . Seja $\tilde{\tau} = \inf(\tau_k)$. Se $\tilde{\tau} = 0$, existem subsequências tal que $\tilde{\tau}_{k_i} \rightarrow 0$, $x^{k_i} \rightarrow \bar{x}$, $h^{k_i} \rightarrow h$ e $\nabla L(\bar{x})^t h = 0$; mas isso contradiz a segunda parte da Proposição 5 (estamos assumindo continuidade de ∇L em \bar{x} , mas a argumentação é a mesma se $\nabla L(x^{k_i})$ não for limitada.). Portanto, concluímos que $\tilde{\tau} > 0$ e existe K_3 tal que para $k \geq K_3$, $\lambda_k \leq \tilde{\tau}$ e segue o resultado. \diamond

O corolário abaixo é consequencia imediata da Proposição 6.

Corolário 1: L e L_i são continuamente diferenciáveis em C .

Demonstração: Usando o fato que L é crescente deduzimos que $\langle a^i, x^k \rangle$ é diferente de zero para todo i , e isso implica no resultado. \diamond

Proposição 7: Existem $\xi > 0$ e $K_4 \in \mathbb{N}$ tais que para $k \geq K_4$

$$L(x^{k+1}) - L(x^k) \geq \lambda_k \xi. \quad (7.30)$$

Demonstração: Pela segunda parte da Proposição 5, a derivada direcional de L em x^k na direção do vetor $x^{k+1} - x^k$ é maior que um número positivo μ . Tomemos $\varepsilon > 0$ tal que se $u, v \in C$ e $\|u - v\| < \varepsilon$, então a derivada direcional de L nos pontos u e v , em qualquer direção dada, difere no máximo por $\frac{\mu}{2}$, usando a equicontinuidade deduzida no Corolário 1 (Do Corolário 1, ∇L e ∇L_i são contínuas no conjunto compacto C , logo são equicontínuas.). Tomemos K_4 tal que $\|x^{k+1} - x^k\| < \varepsilon$. Então para $k \geq K_4$, a derivada direcional ao longo do seguimento unindo x^{k+1} e x^k é maior que $\frac{\mu}{2}$. Mas isso implica que

$$L(x^{k+1}) - L(x^k) \geq \frac{\mu}{2} \|x^{k+1} - x^k\| > (1 - \eta_1) \lambda_k \|\nabla L(x^k)\| \quad (7.31)$$

$$> (1 - \eta_1) \lambda_k \delta_2 \frac{\mu}{2}, \quad (7.32)$$

onde utilizamos as Proposições 4 e a primeira parte da Proposição 5 em 7.31 e 7.32, respectivamente. Tomando $\xi = (1 - \eta_1) \delta_2 \frac{\mu}{2}$ em 7.30, o resultado segue. \diamond

Finalmente, estamos prontos para a demonstração do Teorema 1.

Demonstração Teorema 1: Seja

$$M = \sup_{x \in \mathbb{R}_+^n} (L(x)) \quad (7.33)$$

e suponhamos que exista N tal que

$$L(x^k) \leq N < M. \quad (7.34)$$

Pela Proposição 3, os pontos limites que satisfazem

$$x_j^* \nabla L(x^*)_j = 0 \quad j = 1, \dots, n \quad (7.35)$$

são os pontos de máximo de L , logo 7.34 implica que 7.22 é satisfeita; portanto, usando a Proposição 7 recursivamente temos

$$L(x^k) - L(x^0) \geq \xi \sum_{i=0}^{k-1} \lambda_i. \quad (7.36)$$

Tomando o limite nos dois lados da equação acima, e usando que $\sum \lambda_k = \infty$, temos que $L(x^k) \rightarrow \infty$, o que é uma contradição visto que a seqüência $\{L(x^k)\}$ é limitada.. Assim, 7.34 não é verdade, e existe uma subsequência $\{x^{k_i}\}$ convergindo para x^* e $L(x^*) = M$. Isso implica que toda subsequência converge para x^* . \diamond

7.3 Versão Regularizada

Em [20], discutindo Tomografia por Transmissão, Erdogan e Fessler imaginavam ser possível derivar um algoritmo convergente para o problema regularizado seguindo as mesmas idéias contidas em [5] e [72]. Vejamos como isso pode ser feito.

Primeiramente, vamos descrever a função de verossimilhança penalizada.

$$G(x) = L(x) + \gamma R(x), \quad (7.37)$$

onde $L(x) = \sum_i^m -d_i \exp(-\langle a^i, x \rangle) - y_i \langle a^i, x \rangle$ e $\gamma R(x)$, que é a regularização, é a parte onde, na medida do possível, são incorporadas as informações sobre a imagem x . Em [72], damos uma interpretação Bayesiana para 7.37, daí chamarmos esse modelo de “*Maximum A Posteriori Poisson Model*”, ou simplesmente, MAP.

7.3.1 O Algoritmo T-BSREM

Usando a mesma notação do T-RAMLA, definimos o T-BSREM da seguinte forma: dado um ponto x^0 inicial positivo, e seguindo a convenção

$$x^{(k,0)} = x^{k-1}, \quad x^{(k,p)} = x^{k+1}, \quad (7.38)$$

o algoritmo é

$$x_j^{(k,i)} = x_j^{(k,i-1)} + \lambda_k x_j^{(k,i-1)} \sum_{l \in S_i} a_{lj} d_l \exp\left(-\langle a^l, x^{(k,i-1)} \rangle\right) - a_{lj} y_l \quad (7.39)$$

para $j = 1, \dots, n$ e $i = 1, \dots, p$.

Fazendo $x^{k+\frac{1}{2}} = x^{(k,p)}$, definimos

$$x^{k+1} = x^{k+\frac{1}{2}} + \lambda_k \gamma D^{k+\frac{1}{2}} \nabla R\left(x^{k+\frac{1}{2}}\right), \quad (7.40)$$

onde $D^k = \text{diag}(x_1^k, \dots, x_n^k)$.

Lembrando sempre as condições sobre λ_k , isto é,

$$\lambda_k \rightarrow 0 \quad (7.41)$$

$$\sum_{k=0}^{\infty} \lambda_k = \infty. \quad (7.42)$$

Dada uma função $G(x)$ estritamente côncava, continuamente diferenciável, e feita a decomposição dessa função da seguinte forma

$$G(x) = \sum_{i=1}^m G_i(x), \quad (7.43)$$

assumindo que as funções vetoriais $D(x) \nabla G_i(x)$ para todo i sejam contínuas, definimos um algoritmo mais geral como segue

$$x^{k+1} = x^k + \lambda_k \sum_{i=1}^m D^{(k,i-1)} \nabla G_i(x^{(k,i-1)}). \quad (7.44)$$

Veja que a função de verossimilhança e qualquer prior côncavo podem ser escritos de forma unificada como sugere a equação 7.43.

7.3.2 Resultado de Convergência

Para o algoritmo 7.44, provamos o seguinte resultado de convergência.

Teorema 2: Se a seqüência $\{G(x^k)\}$ tem um limite finito único (se for monótona, por exemplo), então a seqüência $\{x^k\}$ gerada por 7.44, se limitada e positiva, converge para o máximo da função $G(x)$, desde que as condições 7.41 e 7.42 sejam satisfeitas.

A demonstração desse Teorema é encontrado em [72]. Algumas observações se fazem necessárias. A primeira é que o Teorema 2 é essencialmente igual ao Teorema 1. A principal diferença reside no fato de se supor limitação e positividade.

Isso se faz necessário devido a parte de regularização (priors). Entretanto, a hipótese sobre o comportamento de $\{G(x^k)\}$ implica na limitação da seqüência $\{x^k\}$: se a seqüência $\{G(x^k)\}$ é convergente, então é limitada. Considere $G^* = \inf_k(G(x^k))$, e o conjunto $B = \{x/G(x) \geq G^*\}$, então, pela própria definição do conjunto B , $\{G(x^k)\} \in B$. Como a função G é estritamente côncava e além disso tende a $-\infty$ quando $x \rightarrow \infty$ (propriedade compartilhada tanto pela função de verossimilhança quanto pelos priors), segue que as suas curvas de nível são limitadas (em particular em B). Portanto, a seqüência $\{x^k\}$ é limitada. Já a positividade da seqüência $\{x^k\}$ é um problema mais delicado. Se x^k é positivo, então $x^{k+\frac{1}{2}}$ também o será. Isso foi provado no Teorema 1. Entretanto, nada nos garante que x^{k+1} será positivo. Nos nossos experimentos, programamos o algoritmo de forma a garantir a positividade da seqüência, usando um *threshold*. Contudo, em nossas simulações, esse recurso nunca foi necessário.

A segunda observação é que a hipótese forte sobre o comportamento da seqüência $\{G(x^k)\}$ é necessária devido à dificuldade em tratar a fronteira. Como já observamos em [72], essa também era a dificuldade ao estudar a convergência do EM até que Csiszar e Tusnady [16] provaram, em 1984, o resultado definitivo de convergência.

De Pierro, em [70], vem tratando a questão da convergência de métodos iterativos em Tomografia por emissão. Seu objetivo é justamente remover essa condição forte sobre a seqüência $\{G(x^k)\}$.

7.4 Simulações

"Assume what is required to be done as already done." (Pappus)

Nosso objetivo é verificar através de simulações duas características básicas dos novos algoritmos: (i) é mais rápido que o EM (transmissão) e (ii) atinge valores de verossimilhança equivalentes ou superiores ao EM e sua versão acelerada OS-EM.

Todas as simulações foram realizadas utilizando o SNARK93 [4]. Este é um pacote desenvolvido na Universidade da Pennsylvania para simulações envolvendo algoritmos de reconstrução tomográfica. A principal vantagem do SNARK93 é a disponibilidade de diversas rotinas escritas em FORTRAN que facilitam a implementação dos algoritmos iterativos.

7.4.1 T-RAMLA versus EM (Transmissão)

Já sabemos que o EM para o caso de transmissão, assim como em ECT, é lento [20]. São necessárias muitas iterações para se obter um nível alto de verossimilhança, e uma imagem de boa qualidade (por volta de 30 a

80 iterações [56]).

O EM no caso da Tomografia por Transmissão [49] tem a seguinte forma

$$x_j^{k+1} = \frac{\sum_{i=1}^m (M_{ij} + N_{ij})}{\frac{1}{2} \sum_{i=1}^m (M_{ij} - N_{ij}) a_{ij}}, \quad (7.45)$$

onde

$$M_{ij} = b_i + d_i \left[\exp \left(- \sum_{l \in S_{ij}} a_{il} x_l^k \right) - \exp (- \langle a^i, x^k \rangle) \right] \quad (7.46)$$

e

$$N_{ij} = b_i + d_i \left[\exp \left(- \sum_{l \in S_{ij} \cup \{j\}} a_{il} x_l^k \right) - \exp (- \langle a^i, x^k \rangle) \right], \quad (7.47)$$

o conjunto S_{ij} denota os pixels entre a fonte e o pixel j (o pixel j não pertence a S_{ij}) ao longo da projeção i .

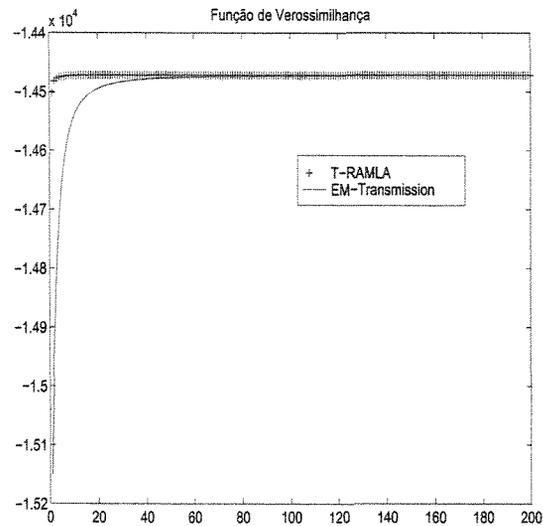
A equação 7.45 é uma aproximação (série truncada) para a solução da equação transcendental

$$\sum_{i=1}^m -N_{ij} a_{ij} + \sum_{i=1}^m \frac{(M_{ij} - N_{ij}) a_{ij}}{\exp(a_{ij} x_j) - 1} = 0. \quad (7.48)$$

Ollinger [65] argumenta que é mais seguro resolver a equação acima utilizando o método de Newton ou métodos equivalentes.

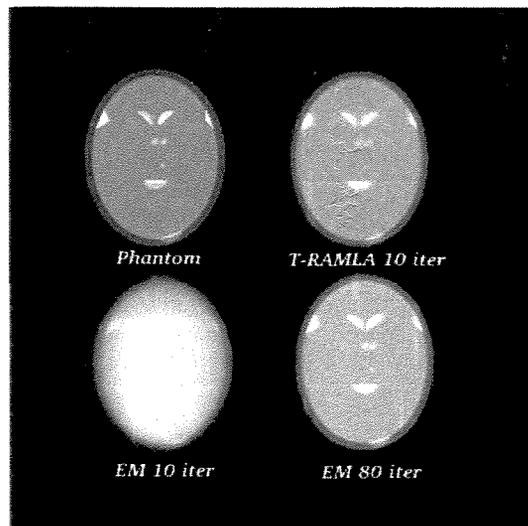
Para verificar se o T-RAMLA é mais rápido que o EM, fizemos o seguinte experimento: utilizamos 120 vistas (ângulos de coleta de dados) com 231 raios em cada vista, com uma média de 10000000 fótons, e espalhamento (SCATTER 0.24 0.445) e com apenas 10% dos fótons emitidos sendo detectados. Utilizamos 16 blocos no T-RAMLA, $\lambda_0 = 1.5$ e $\lambda_k = 1.5/k^{1/4}$. A escolha dos λ 's foi experimental. A seqüência λ_k assim gerada satisfaz as condições de convergência, mas permite passos suficientemente compridos no início. Nos experimentos com T-BSREM usamos como função potencial o $\text{Log}(\cosh)$.

Vejamos o gráfico da função de verossimilhança em função do número de iterações.



Pelo gráfico é possível perceber que o T-RAMLA já nas 5 primeiras iterações atinge níveis altos de verossimilhança, enquanto o EM só atingirá valores equivalentes a partir da 60 iteração, confirmando que o T-RAMLA, de fato, é mais rápido.

Fato também confirmado pela qualidade das imagens obtidas. Veja alguns exemplos

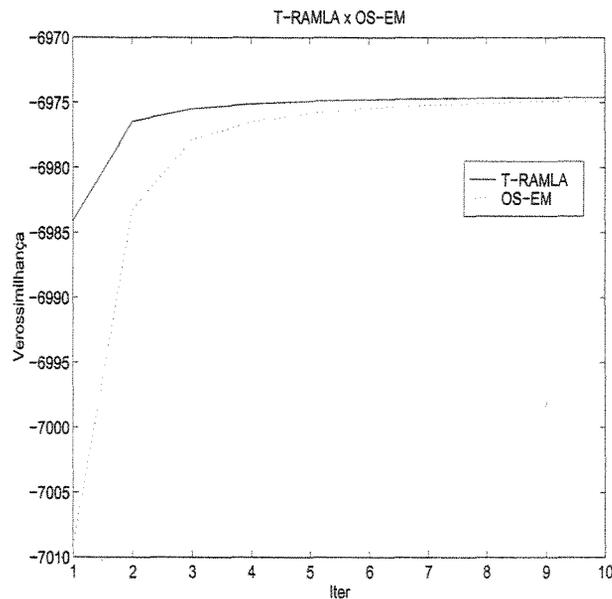


É visível a diferença entre a décima iteração do EM e a décima iteração do T-RAMLA. O EM só atingirá uma imagem com qualidade semelhante por volta da octagésima iteração.

7.4.2 T-RAMLA versus OS-EM (Transmissão)

Podemos “acelerar” o EM introduzindo os subconjuntos ordenados, a exemplo do que foi feito por Mangic et. al. [56]. Bons resultados são obtidos com essa estratégia, seja do ponto de vista da aceleração, seja do ponto de vista da qualidade das imagens.

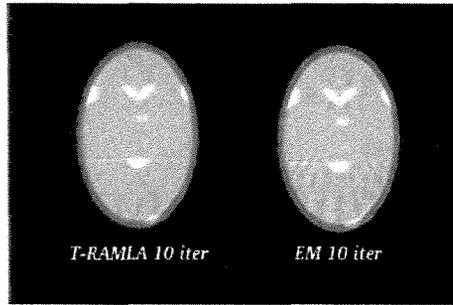
Entretanto, ao realizar um experimento onde o ruído é maior, ou seja, quando diminuimos a estatística de fótons (média de 10000) e o número de vistas (60 vistas), o T-RAMLA atinge valores maiores de verossimilhança, deixando o concorrente sempre por baixo nas 10 primeiras iterações. Comportamento semelhante verificado no caso da Tomografia por emissão, onde o OS-EM, por não convergir, fica com valores da função de verossimilhança abaixo do RAMLA.



Outro experimento interessante é reduzir drasticamente o número de vistas (ângulos) na obtenção dos dados. Do ponto de vista clínico, isso diminui consideravelmente o tempo dos exames, além de submeter o paciente a menos radiação.

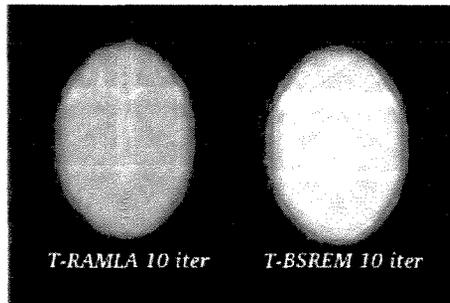
Nesses experimentos encontramos aquela que talvez seja a principal deficiência do EM. Em alguns casos chegando ao extremo de não gerar imagem alguma; enquanto, o T-RAMLA continua obtendo imagens com qualidade razoável, apesar de alguns artefatos.

Nos gráficos abaixo, reduzimos o número de vistas a 24, isto é, uma vista a cada 15 graus. Veja que EM-Transmissão gera uma imagem com muito mais artefatos que a imagem gerada pelo T-RAMLA.



Se tomarmos uma vista a cada 30 graus, o T-RAMLA ainda é capaz de mostrar uma imagem que, apesar dos artefatos, ainda lembra o phantom; enquanto o EM não gera imagem alguma, ou seja, começam a surgir valores negativos na imagem, demonstrando que o algoritmo não está bem definido.

Na figura abaixo, mostramos a imagem obtida via T-RAMLA após 10 iterações, e ao lado uma imagem obtida com T-BSREM. Veja que a qualidade da imagem regularizada é superior, indicando que nestes casos extremos a regularização pode melhorar a qualidade da imagem. Vejamos na secção seguinte alguns experimentos com o T-BSREM.



7.4.3 T-BSREM versus OSGA

Para avaliar o desempenho do T-BSREM, tivemos que criar uma versão baseada em Subconjuntos Ordenados do "Gradient Algorithm" de Lange e Fessler [50]. O Gradient Algorithm original tem a seguinte fórmula de iteração

$$x_j^{k+1} = x_j^k + \frac{x_j^k}{\sum_i b_i a_{ij} + x_j^k \frac{\partial^2 R(x^k)}{\partial x_j^2}} \frac{\partial \Delta(x^k)}{\partial x_j}, \quad (7.49)$$

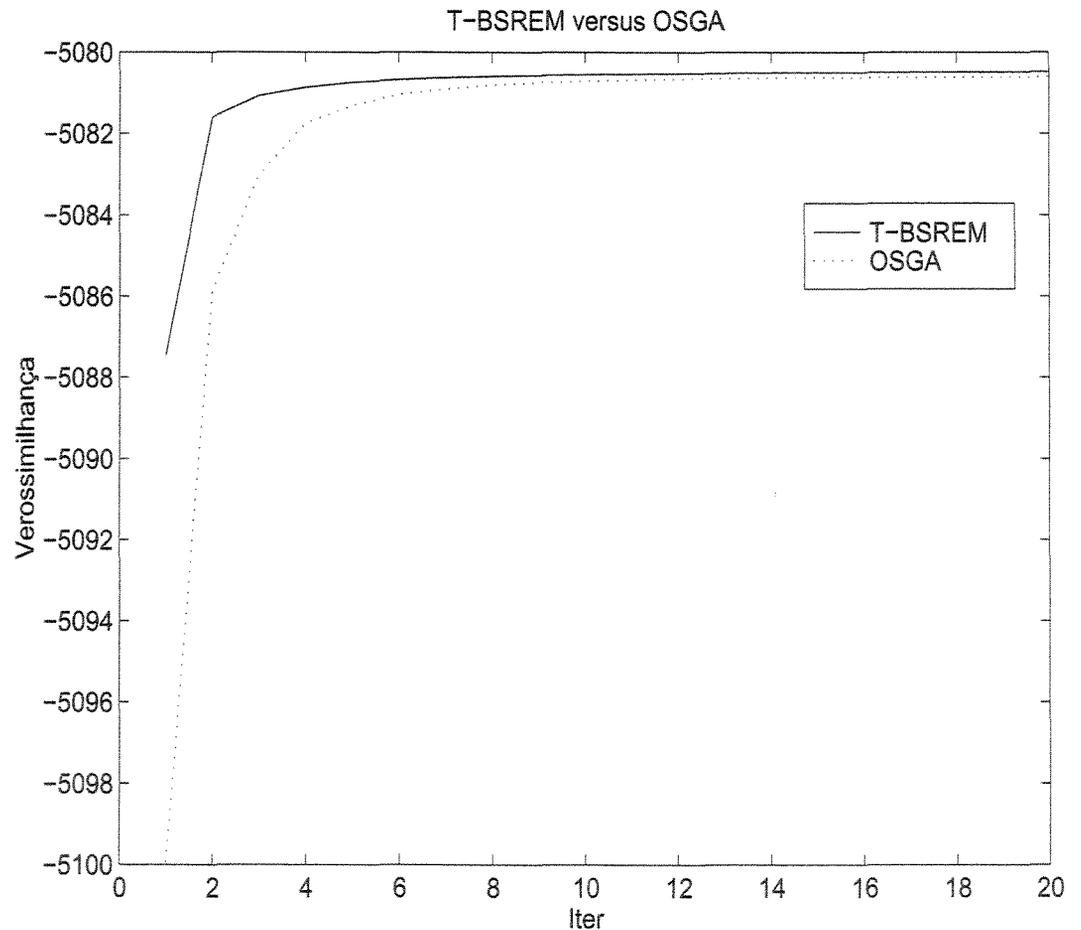
onde $\Delta(x) = L(x) - \gamma R(x)$. $L(x)$ é a função de verossimilhança para Transmissão e $R(x)$ é o prior. Em nossos experimentos, adotamos $\ln(\cosh(x))$ como a função potencial de R .

A versão que utiliza Subconjuntos Ordenados desenvolvida por nós, de agora em diante denominada OSGA³, é facilmente obtida a partir da fórmula 7.49. Essa versão é bem mais rápida que o algoritmo original

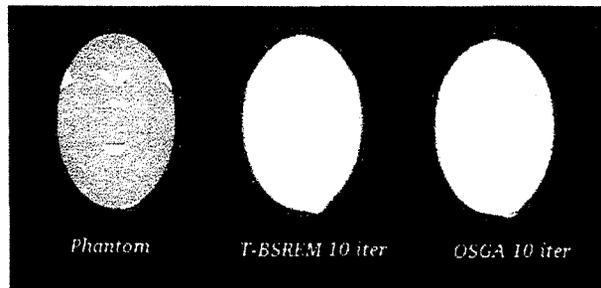
³O "Gradient Algorithm", [50], tem boas propriedades de convergência. Entretanto, ao introduzir a metodologia de Subconjuntos

proposto em [50] e gera imagens de excelente qualidade quando o ruído não é muito grande. Entretanto assim como as versões não regularizadas, o OSGA sob condições desfavoráveis de ruído, é superado em desempenho pelo T-BSREM. Em casos extremos chega a não gerar imagem alguma (iterandos negativos). A seguir apresentaremos os resultados típicos de experimentos onde há baixa estatística de fótons. Nessa simulação, utilizamos 16 blocos, com baixa estatística de fótons (média de 10000), 45 vistas (uma coleta cada 8 graus) com 231 raios ou projeções. E a seqüência de $\lambda_k = \frac{\lambda_0}{k}$ e $\gamma = 0.1$.

A seguir o gráfico do valor da função de verossimilhança.



É importante observar que o T-BSREM atinge valores maiores de verossimilhança que o OSGA. Entretanto o OSGA, em geral, fornece imagens de maior qualidade já nas primeiras iterações. Veja as imagens abaixo Ordenadas, o OSGA não compartilha de tais propriedades.



Capítulo 8

Conclusões

8.1 Vantagens dos Novos Algoritmos

Neste trabalho apresentamos três novos algoritmos rápidos em Tomografia Computadorizada: um para Tomografia por Emissão (PET) e os outros dois para Tomografia por Transmissão.

O primeiro algoritmo denominamos de Block Sequential Regularized Expectation Maximization -BSREM. Este surgiu como uma extensão do RAMLA para o caso regularizado, tornando-se uma alternativa aos outros algoritmos, em especial, ao OS-GP.

O modelo regularizado é importante por incorporar informações conhecidas “a priori” sobre a imagem. Além disso, pode ser utilizado para evitar que o ruído (“sal e pimenta”) se sobreponha ao sinal.

Entre as vantagens do BSREM podemos citar o resultado de convergência que assegura a obtenção de valores de verossimilhança superiores aos obtidos pelo OS-GP (já nas primeiras iterações) que, por sua vez, se traduz em imagens de boa qualidade. Além disso, a forma como o termo de regularização é tratado facilita significativamente os resultados teóricos e a implementação computacional além de evitar possíveis efeitos indesejáveis sobre a imagem como o “oversmoothing”.

O BSREM vem sendo bem aceito mesmo antes de sua publicação em [72].

Os outros dois algoritmos dizem respeito à Tomografia por Transmissão. Vale ressaltar que a utilização dos métodos iterativos em Tomografia por transmissão só é justificada se a estatística de fótons for baixa (como por exemplo, no caso dos equipamentos de PET que colhem dados de transmissão para a correção por atenuação). Caso a estatística de fótons seja alta, os métodos diretos são excelentes.

Propomos dois algoritmos para a Tomografia por Transmissão: o T-RAMLA e o T-BSREM. O primeiro para o modelo não regularizado e o segundo para o modelo regularizado.

No caso particular da Tomografia por Transmissão, o EM, como proposto em [49], é lento e não converge. O T-RAMLA surge como um algoritmo rápido e com resultado de convergência. Uma característica especialmente importante é que o T-RAMLA gera iterandos positivos ao contrário do EM que pode gerar iterandos negativos.

Os experimentos confirmaram o bom desempenho do T-RAMLA comparado ao EM.

Comparado ao OSGA, o T-BSREM se mostrou superior com relação aos valores da função de verossimilhança; contudo, não obteve imagens de qualidade significativamente superior ao OSGA, exceto nos casos onde este último não gerava imagem alguma. Portanto, o T-BSREM é uma boa alternativa nos casos onde o OSGA falha (baixa estatística de fótons).

8.2 Pesquisas Futuras

No decorrer de nossas pesquisas, várias perguntas interessantes foram surgindo. Entre elas, o comportamento do T-BSREM se utilizarmos priors não convexos (côncavos). Nossa expectativa é que o T-BSREM tenha a mesma flexibilidade que o BSREM, ou seja, maximize satisfatoriamente a função objetivo, mas isso deve ser comprovado experimentalmente.

Um outro problema que estamos particularmente interessados é a recuperação simultânea de atenuação e atividade a partir de dados de emissão. Essa é uma forma “elegante” de corrigir emissão por atenuação sem a necessidade de uma coleta de dados de transmissão em separado. Um algoritmo recente foi proposto por Nuyts et. al. [63]. A idéia é maximizar a parte referente à atividade (emissão) utilizando o EM, e maximizar a parte da atenuação usando um algoritmo proposto por Nuyts et. al. em [64]. Entretanto, esse algoritmo tem duas grandes desvantagens: (i) o algoritmo que maximiza a parte de atenuação não tem positividade garantida. (ii) o número médio de iterações para se obter boas imagens é da ordem de 1000. Esse número de iterações é proibitivo se levarmos em conta que estamos trabalhando com o dobro de pixels (ou voxels), ou seja, metade para a atividade e o restante para a atenuação. Nossa idéia é utilizar o BSREM para maximizar a parte da atividade (emissão) e o T-BSREM para a parte de atenuação. Além da aceleração, esperamos obter valores de verossimilhança superiores pelas boas propriedades de convergência.

Capítulo 9

Apêndice A

Mostraremos através de dois contra-exemplos simples que o algoritmo EM para transmissão proposto por Lange e Carson [49] e sua versão acelerada proposta por Manglos et. al. [56] não são convergentes.

Primeiramente se o sistema for convergente. Consideremos o seguinte sistema linear

$$\begin{cases} \frac{1}{2}x_1 + x_2 = \frac{1}{2} \\ \frac{1}{2}x_1 + \frac{1}{4}x_2 = \frac{1}{4} \end{cases}.$$

A função de verossimilhança será

$$L(x_1, x_2) = -\exp\left(-\frac{1}{2}x_1 - x_2\right) - \exp\left(-\frac{1}{2}x_1 - \frac{1}{4}x_2\right) - \frac{3x_1}{8} - \frac{9x_2}{9},$$

que tem como máximo o ponto $(3 \ln(2), 2 \ln(2))$.

Aplicando-se o EM, partindo-se do ponto $(\frac{1}{4}, \frac{3}{8})$ que satisfaz a primeira equação, chega-se ao ponto $(15.9562, 0)$ que não é o ponto de máximo.

Consideremos, agora, o seguinte sistema inconsistente

$$\begin{cases} \frac{1}{2}x_1 + x_2 = \frac{1}{2} \\ \frac{1}{2}x_1 + x_2 = \frac{1}{4} \end{cases}. \quad (9.1)$$

Para esse caso, a função de verossimilhança será

$$L(x_1, x_2) = -2 \exp\left(-\frac{1}{2}x_1 - x_2\right) - \frac{3}{4} \left(\frac{1}{2}x_1 + x_2\right). \quad (9.2)$$

Para maximizar a função acima, basta igualar o gradiente a zero, ou seja,

$$\nabla L = 0.$$

Dessa forma, encontramos que a solução de máxima verossimilhança x^* satisfaz a equação

$$\frac{1}{2}x_1 + x_2 = -\ln\left(\frac{3}{8}\right), \quad (9.3)$$

onde $-\ln\left(\frac{3}{8}\right) \cong 0.9808$.

Utilizando-se o EM, partindo do ponto inicial $(\frac{1}{4}, \frac{3}{8})$, chegamos ao ponto (com cinco casas decimais) $(-4.37571, 2.46522)$ que não satisfaz a equação 9.3, e possui a primeira coordenada negativa, violando a positividade. Isso demonstra que o algoritmo não é convergente, e tampouco gera uma seqüência positiva.

Já a versão com subconjuntos ordenados, partindo do mesmo ponto inicial, fica oscilando entre os pontos $(-2.1571, 1.8186)$ e $(-2.36532, 1.6625)$. Comportamento semelhante ao encontrado no OSEM.

Por outro lado, com o T-RAMLA, partindo-se do mesmo ponto inicial e tomando-se $\lambda_k = 1/k$, após 80 iterações, chegamos ao ponto $(0.365515, 0.778139)$ que satisfaz a equação 9.3 até a quarta casa decimal. Partindo-se de diferentes pontos iniciais e/ou utilizando outras seqüências de λ_k (sempre respeitando as condições), obtemos outros pontos que satisfazem a equação 9.3.

Capítulo 10

Apêndice B

Fast Iterative Methods Applied to Tomography Models with General Gibbs Priors

Alvaro R. De Pierro* and Michel E. B. Yamagishi**

*A. R. De Pierro is with State University of Campinas, e-mail: alvaro@ime.unicamp.br

**M. E. Belez Yamagishi is with State University of Campinas, e-mail: michel@ime.unicamp.br

ABSTRACT

Recently, we presented a very fast method (BSREM) for solving regularized problems in emission tomography Refs. 1, 2, that is convergent to Maximum a Posteriori (MAP) solutions for convex priors. The method generalizes in a natural way the Expectation Maximization (EM) algorithm and some of its extensions to the MAP case. It consists of decomposing the likelihood function in blocks containing sets of projections plus a block that corresponds to the prior function, and iterating for each block using scaled gradient directions, resembling a smoothed iteration algorithm. In the general nonconvex case (nonconvex Gibbs priors), it can be proven that the algorithm converges to a critical point, instead of the sought global maximum. In spite of this, BSREM is so fast and flexible, that it can be used to search for global (or better) optima when the model is not convex. In this article, we present an implementation of BSREM, that works as the local optimization method for each step of a Graduated Convexity (GNC) approach as compared with BSREM itself without any convexification. We illustrate the behavior of the method with applications to emission tomography.

Keywords: Maximum Likelihood, Bayesian Estimation, Positron Emission Tomography

1. INTRODUCTION

For a given inverse problem with data b , the Bayesian approach to its solution consists of maximizing the “A posteriori” Probability (MAP), that is, the probability of obtaining a solution x , given the data b , $P(x|b)$, (see Refs. 3, 4). Via the Bayes formula, the mathematical model is given by

$$\max_{x \geq 0} P(x|b) = \frac{P(b/x)P(x)}{P(b)} = \frac{P_L(b/x) P(x)}{P(b)}, \quad (1)$$

where $P(x)$ and $P(b)$ are the probabilities of x and b respectively. Taking logarithms and eliminating terms not depending on x , we obtain

$$\max_{x \geq 0} G(x) = L(x) + \gamma F(x), \quad (2)$$

where $\gamma F(x)$ represents the information we have about the distribution of x , or, in other words, the “approximation” of $\log P(x)$.

In the particular case of positron emission tomography (PET) Ref. 5 coincidences are counted along m lines (determined by pair of detectors) and b is the m -dimensional vector whose i th element, b_i ($b_i \geq 0$), is the number of coincidences for the i th line ($i = 1, \dots, m$) during the data collection period; we shall refer to b as the measurement vector. If a_{ij} ($a_{ij} \geq 0$) denotes the probability that a positron emitted from pixel j results in a coincidence at the i -th line of response (in our experiments this probability is approximated by the length of the intersection), then b_i is a sample of a Poisson distribution whose expected value is

$$\langle a_i, x \rangle = \sum_{j=1}^m a_{ij} x_j \quad (3)$$

Send correspondence to A. R. De Pierro: State University of Campinas, Department of Applied Mathematics, CP 6065, CEP 13081-970, Campinas, SP, Brazil

where \langle , \rangle is the standard inner product and a^i is the i th column of the transpose A^T of the $m \times n$ projection matrix $A = (a_{ij})$ Ref. 6.

The probability of obtaining the measurement vector b if the image vector is x (i.e., the likelihood function) is

$$P_L(b|x) = \prod_{i=1}^m \left[\frac{\langle a_i, x \rangle^{b_i}}{b_i!} \exp(-\langle a_i, x \rangle) \right]. \quad (4)$$

So, in this case, the log-likelihood function is

$$L(x) = \sum_{i=1}^m b_i \log \langle a_i, x \rangle - \langle a_i, x \rangle. \quad (5)$$

We will assume that $F(x)$ is a function of the form

$$F(x) = - \sum_{l=1}^t f_l(\langle s_l, x \rangle), \quad (6)$$

$f_l, l = 1, \dots, t$, are single variable real functions, $s_l = \{s_{lj}\} (l = 1, \dots, t)$ real vectors and $\gamma > 0$

In recent years, algorithms that take into account the particular geometry of the scanners have been proposed, allowing very fast reconstructions when applied to the maximum likelihood (ML) model (Refs. 7–9); in Ref.1 and Ref.2, the algorithm in Ref. 7 was extended to one solving the more general MAP model, and called BSREM. Ref.2 presents a set of statistically based experiments, showing the high performance of BSREM, using convex priors. In the next Section, we briefly describe BSREM and give the justification for its application to nonconvex models. In Section 4 we give a brief description of the Graduated Non Convexity (GNC) approach. In Section 5 we present PET simulations comparing BSREM and GNC using BSREM.

2. A FAST ALGORITHM FOR THE MAP MODEL

Let us define a partition of the integer interval $M = [1, p]$ into a family of p disjoint subsets S_i , such that

$$M = \cup_{i=1}^p S_i. \quad (7)$$

Consider now a decomposition of L using the partition defined above, that is,

$$L(x) = \sum_{i=1}^p L_i(x), \quad (8)$$

where, for $i = 1, \dots, p$,

$$L_i(x) = \sum_{t \in S_i} y_t \ln \langle a_t, x \rangle - \langle a_t, x \rangle. \quad (9)$$

Then, using the same notation of Ref. 7, we define the BSREM (for Block Sequential Regularized Expectation Maximization) algorithm as follows. Given a sequence of positive real numbers $\{\lambda_k\}$ satisfying,

$$\lambda_k \xrightarrow{k \rightarrow \infty} 0, \quad (10)$$

and

$$\sum_{k=0}^{\infty} \lambda_k = +\infty, \quad (11)$$

Given a positive starting point x^0 , $k = 0, 1, 2, \dots$ and $x^{(k,0)} = x^k$, for $i = 1, \dots, p$,

$$x^{(k,i)} = x^{(k,i-1)} + \lambda_k D^{(k,i-1)} \nabla L_i(x^{(k,i-1)}). \tag{12}$$

Set $x^{k+\frac{1}{2}} = x^{(k,p)}$ and define

$$x^{k+1} = x^{k+\frac{1}{2}} + \lambda_k \gamma D^k \nabla F(x^{k+\frac{1}{2}}), \tag{13}$$

where D^k is the diagonal matrix with the variables in the diagonal as in Ref. 7.

In order to keep positivity, in some cases, it is necessary to enforce it using some threshold after iterating the regularization part. In Ref.⁷ we present some convergence properties of the algorithm assuming strong hypotheses, but suggesting, for the more general case, full convergence to the MAP estimate.

3. THE GRADUAL CONVEXIFICATION APPROACH

As pointed out before, BSREM was proven to converge to a maximum in the concave case, and a slight modification of the proofs can show convergence to a critical point in the general nonconvex (nonconcave) case. It is well known that finding a global maximum of a nonconcave function is a very high cost computational task in the case of huge problems like those arising in tomography. However, several methods exist that, although not guaranteeing a global optimum, they attain at least local optima with some properties. One of these methods is the extension of Graduated Non Convexity (GNC) recently proposed by Nikolova et al. in Ref.10 (we will call it EGNC). In Ref.10, the f_i 's in (6) are piecewise Gaussian functions depending on two parameters, c and a defined as

$$f_i(t) = \Phi_c(t) + at^2, \tag{14}$$

where

$$\Phi_c(t) = \begin{cases} (\lambda t)^2, & |t| < q_c, \\ \alpha - \frac{1}{2}c(|t| - r_c)^2, & q_c \leq |t| < r_c, \\ \alpha, & |t| \geq r_c, \end{cases} \tag{15}$$

and $q_c = T(1 + 2\lambda^2/c)^{-1/2}$, $r_c = T(1 + 2\lambda^2/c)^{1/2}$. When using these functions, small values of c convexify the function and a stabilizes the problem for ill-conditioning and should tend to zero. T represents the interval where the function is quadratic (big values of T mean smoother images).

4. SIMULATIONS

We have performed simulations with PET data comparing the behavior of pure BSREM against EGNC where the intermediate maximization (for each value of c and a) is performed by using a few iterations of BSREM with the following modification: the iteration formulas (12) and (13) were changed to

$$x^{(k,i)} = x^{(k,i-1)} + \lambda_k D^k \left(\nabla L_i(x^{(k,i-1)}) + \frac{\gamma}{p} \nabla F(x^{(k,i-1)}) \right), \tag{16}$$

set $x^{k+1} = x^{(k,p)}$.

Results of those experiments are illustrated by figures 1-4, showing typical results. The number of blocks, S_i , was 16, λ_k for BSREM was taken $.2/k$ for each iteration. For EGNC a was taken $.5/k^2$ and c was increased linearly to infinity. Figure 1 shows a PET phantom and corresponding reconstructions by BSREM and EGNC combined with BSREM. Figure 2 shows the 54th line of both reconstruction. Figure 3 shows the behavior of the values of the objective function with the iteration number. Finally, Figure 4 represents the quadratic difference between the values of the objective function.



Figure 1: PET phantom and reconstructions

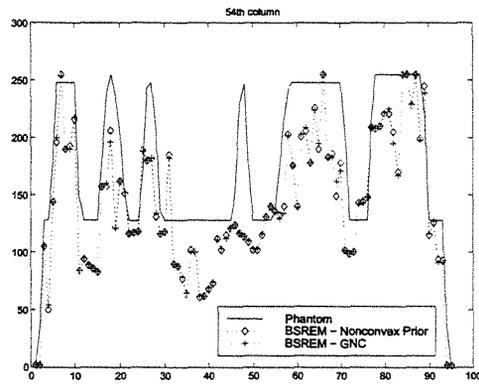


Figure 2: 54th column

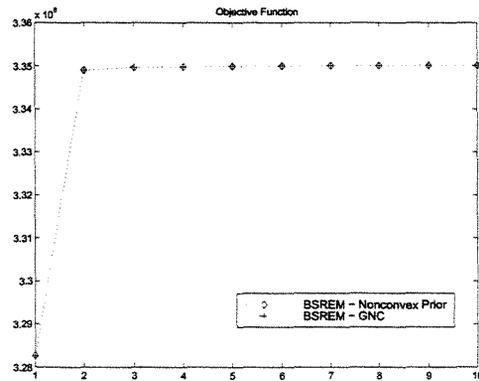


Figure 3: Objective function values

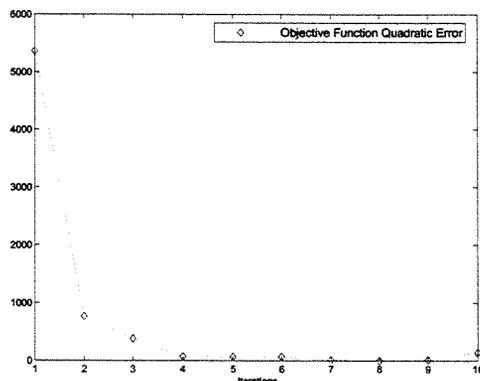


Figure 4: Square difference

5. CONCLUSIONS

The goal of this work was to show experimentally that BSREM is a very flexible algorithm and behaves very well when used in its pure form as compared to a convexification-regularization algorithm like EGNC. Of course these results are preliminary because of the large number of parameters involved in the implementations; further experimental work is needed in order to validate them confirming the capacity of BSREM to achieve high values of the MAP function in few iterations.

Acknowledgements

Work by the first author was partially supported by CNPq grant No. 301699/81 and FAPESP grant No. 98/00618-2. The second author was supported by FAPESP grant No. 96/09186-2

REFERENCES

1. A. R. DePierro, "Fast bayesian estimation methods in emission tomography," in *Bayesian Inference for Inverse Problems*, A. M. Djafari, ed., *Proc. SPIE* **3459**, pp. 100–105, 1998.
2. A. R. DePierro and M. B. Yamagishi, "Fast em-like methods for maximum 'a posteriori' estimates in emission tomography," *submitted*.
3. E. Levitan and G. T. Herman, "A maximum a posteriori probability expectation maximization algorithm for image reconstruction in emission tomography," *IEEE Trans. Med. Imaging* **6**, pp. 185–192, 1987.
4. S. Geman and D. McClure, "Statistical methods for tomographic image reconstruction," *Bull. Intl. Stat. Inst.* **LII-4**, pp. 5–21, 1987.
5. M. M. Ter-Pogossian, M. Raichle, and B. E. Sobel, "Positron emission tomography," *Scientif. Amer.* **243**, pp. 170–181, 1980.
6. Y. Vardi, L. A. Shepp, and L. Kaufman, "A statistical model for positron emission tomography," *J. Amer. Statist. Assoc.* **80**, pp. 8–20, 1985.
7. J. Browne and A. R. DePierro, "A row-action alternative to the em algorithm for maximizing likelihoods in emission tomography," *IEEE Trans. Med. Imaging* **15**, pp. 687–700, 1996.
8. H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans. Med. Imaging* **13**, pp. 601–609, 1994.
9. H. Hudson, B. Hutton, and R. Larkin, "Accelerated em reconstruction using ordered subsets," *J. Nucl. Med.* **33**, p. 960, 1992.
10. M. Nikolova and J. I. A. Mohammad-Djafari, "Inversion of large support ill-posed linear operators using a piecewise gaussian mrf," *IEEE Trans. Image Proc.* **7**, pp. 571–585, 1998.

Capítulo 11

Apêndice C

Fast EM-like Methods for Maximum ‘A Posteriori’ Estimates in Emission Tomography

Alvaro R. De Pierro, Michel Eduardo Beleza Yamagishi

Abstract— The Maximum Likelihood (ML) approach in emission tomography provides images with superior noise characteristics compared to conventional filtered backprojection (FBP) algorithms. The Expectation Maximization (EM) algorithm is an iterative algorithm for maximizing the Poisson likelihood in ECT that became very popular for solving the ML problem because of its attractive theoretical and practical properties. Recently ([3], [22]) block sequential versions of the EM algorithm that take advantage of the scanner’s geometry have been proposed in order to accelerate its convergence. In [22], the OS-EM (Ordered Subsets EM) method was applied to the ML problem and a modification (OS-GP) to the MAP (Maximum A Posteriori) regularized approach without showing convergence. In [3], we presented a relaxed version of OS-EM (RAMLA) that converges to an ML solution. In this article we present an extension of RAMLA for MAP reconstruction. We show that, if the sequence generated by this method converges, then it must converge to the true MAP solution. Experimental evidence of this convergence is also shown. To illustrate this behavior we apply the algorithm to positron emission tomography simulated data comparing its performance to OS-GP.

Keywords— Expectation maximization algorithm, ordered-subsets maximum likelihood algorithm, positron emission tomography, regularization.

I. INTRODUCTION

EMISSION computed tomography (ECT) [6] is the quantitative determination of the changes in the chemistry and flow physiology of radioactive labeled compounds inside the body. Mathematically, the problem con-

sists of reconstructing a function representing the distribution of radioactivity from measured data that are the total activity along lines of known location. Measurements are much more noisy than in X-ray Computed Tomography (CT), so, it is desirable to have a reconstruction approach incorporating an estimation procedure that takes into account the statistical nature of the noise.

In the particular case of positron emission tomography (PET) [41], the isotope used emits positrons which annihilate with nearby electrons, thereby generating two photons traveling away from each other in (approximately) opposite directions. The collected data is the number of photon pairs detected in time coincidence for each line determined by a pair of detectors. Since the emission and detection of photons are Poisson processes [16], it was first suggested by Rockmore and Macovski [36] that improvements in image reconstruction from ECT data may be achieved by estimating the radioactive distribution using maximum likelihood (ML) estimation. Shepp and Vardi [38], and independently Lange and Carson [25], proposed the use of the Expectation Maximization (EM) [9] algorithm to maximize the Poisson likelihood in PET. Since then, the ML-EM approach has become a very popular method, providing very good images compared to reconstructions produced with the conventional filtered backprojection (FBP) algorithm [39], not only in PET, but also in Single Photon Emission Computed Tomography (SPECT). The advantages of ML-EM, and, in general, of iterative methods, over FBP has already been recognized by the research community (see [37]); in spite of this, it took a long time for these algorithms to be adopted by commercial ECT systems. One of the main reasons for this is that the EM algorithm has an extremely

A. R. De Pierro is with State University of Campinas, Department of Applied Mathematics (e-mail: alvaro@ime.unicamp.br). Work partially supported by CNPq grant n^o 301699/81 and FAPESP grant n^o 98/00618-2

M. E. Beleza Yamagishi is with State University of Campinas, Department of Applied Mathematics (e-mail: michel@ime.unicamp.br) Work supported by FAPESP grant n^o 96/09186-2

slow rate of convergence and thus a large number of iterations may be required to achieve an acceptable image. Several modifications have been suggested to accelerate the convergence of the algorithm using line search [30], [24] and vector extrapolation [35] techniques. Along the 90's some faster alternatives were proposed to speed up convergence; among them, it is worth mentioning the Space Alternating Generalized Expectation Maximization (SAGE) algorithms by Fessler and Hero [13] and the Coordinate Ascent approach of Bouman and Sauer [2], achieving an order of magnitude acceleration. Others have considered preconditioned conjugate gradient techniques [26], [27], [32], [33]. In two papers published in 1992 [21] and 1994 [22] Hudson, Hutton and Larkin applied the EM algorithm sequentially to blocks of projections obtaining very good images in few iterations, but without showing convergence proof in the general inconsistent case. They called their algorithm Ordered-Subsets Expectation Maximization (OS-EM). Subsequently, Browne and De Pierro [3] in a parallel research, inspired by work of Herman and Meyer on the use of ART in PET reconstruction [18] introduced relaxation in the 1-block version of OS-EM, and called the resulting algorithm RAMLA, proving that relaxing in the direction defined by the algorithm applied to the blocks is the key to obtain convergence to an ML solution. Since its appearance, as far as we know, OS-EM (and RAMLA) has been the only fast iterative algorithm adopted by commercial scanners (PET or PET-SPECT scanners). This fact and the high quality of the resulting images (see [31]) motivates this article, that deals with the natural extension of OS-EM for solving the Maximum A Posteriori (MAP) model for ECT. An interesting historical observation at this point is that the 'subsets' or 'blocking' idea for acceleration purposes, previously appeared in two papers by Holte et al. [20] and Hebert et al. [15] but, went apparently unnoticed by the medical imaging algorithmic community.

As it is well known, ECT is an ill-posed problem; that is, small variations in the data produce large variations in the solution. When using an iterative algorithm for the ML

problem, at some point, as the likelihood increases, the image starts deteriorating; the algorithm enforces consistency with data, and, because of the size of the problem, more and more high frequencies related to noise show up (see Figure 4 (ii)). So, iterations should be stopped before this happens, by using a suitable stopping criterion. A better approach to overcome the ill-posedness effects is to consider prior information through regularization terms, allowing more sophisticated and accurate models. In this case, asymptotic convergence to the right limit (the MAP solution) becomes definitely more important. In [22], an extension of OS-EM, called OS-GP (we assume that this comes from Ordered Subsets Gibbs Priors, although this does not appear explicitly in the article), to the regularized case is presented, together with some numerical experiments. OS-GP was derived from the application of Green's One Step Late (OSL) [14] to OS-EM. However, unlike OSL, that can be proved to be, at least, locally convergent, this new version does not converge to the MAP solution (see Appendix III). Also, it can be easily shown that not even by introducing relaxation as in [3], it is possible to enforce convergence of the algorithm to a MAP solution. It is worth noting that some of the accelerated algorithms mentioned before, have also extensions to the MAP problem, for example, [13] and [2]. In a recent article [23], it was proposed the use of classical interior point methods for the MAP problem in PET. Also, in [23], many additional references on the subject can be found.

In this article we present a new method, that is a natural extension of OS-EM and RAMLA for the MAP estimation problem. In the Appendix we present a proof of the fact that when the full sequence converges, the limit point is a solution (Appendix I) and a simple example illustrating the local behavior of the algorithm (Appendix II). Also, a counterexample for the non-convergence of OS-GP to a MAP solution is shown (Appendix III). In the next two sections we describe the ML and MAP models, the previous OS-type algorithms for solving them and the new method. In section IV we present numerical experiments using PET

simulated data. In Section V we discuss the results and, finally, Section VI includes some concluding remarks and research directions.

II. ORDERED SUBSETS-TYPE METHODS FOR ML SOLUTIONS

We consider now the discretized two-dimensional PET model. In what follows, x_j denotes the expected number of coincidences per unit area in the j th pixel ($j = 1, \dots, n$) (we are assuming here that the “tubes” are of constant cross-section and that dividing the radioactivity concentration per unit volume by the area of this constant cross section gives us the “number of coincidences per unit length”); x represents the n -dimensional vector whose j th component is x_j and will be referred to as the image vector.

Suppose we count coincidences along m lines. We denote by b the m -dimensional vector whose i th element, b_i ($b_i \geq 0$), is the number of coincidences which are counted for the i th line ($i = 1, \dots, m$) during the data collection period; we shall refer to b as the measurement vector. If a_{ij} ($a_{ij} \geq 0$) denotes the probability that a positron emitted from pixel j results in a coincidence at the i -th line of response (in this paper this probability is approximated by the weighted length of the intersection), then b_i is a sample of a Poisson distribution whose expected value is

$$\langle a^i, x \rangle = \sum_{j=1}^n a_{ij} x_j \quad (1)$$

where $\langle \cdot, \cdot \rangle$ is the standard inner product and a^i is the i th column of the transpose A^T of the $m \times n$ projection matrix $A = (a_{ij})$ [42].

The probability of obtaining the measurement vector b if the image vector is x (i.e., the likelihood function) is

$$P_L(b|x) = \prod_{i=1}^m \left[\frac{\langle a^i, x \rangle^{b_i}}{b_i!} \exp(-\langle a^i, x \rangle) \right]. \quad (2)$$

The reconstruction problem is to estimate the image vector x given the data measurements b .

As mentioned before one approach to this problem, first

suggested in [36], is the ML method which estimates the x that maximizes $P_L(b|x)$ subject to nonnegative constraints on x , or, equivalently, finds the $x \geq 0$ which maximizes

$$L(x) = \sum_{i=1}^m b_i \log \langle a^i, x \rangle - \langle a^i, x \rangle. \quad (3)$$

In [39] and [25], the EM algorithm [9] was proposed for solving (3). This results in an iterative algorithm, which, starting with a strictly positive vector $x^{(0)}$, successively estimates the image vector by

$$x_j^{(k+1)} = \frac{x_j^{(k)}}{\sum_{i=1}^m a_{ij}} \sum_{i=1}^m \frac{a_{ij} b_i}{\langle a^i, x \rangle}, \quad j = 1, \dots, n \quad (4)$$

The EM algorithm’s convergence rate is extremely slow and thus a significant amount of computation is often needed to achieve an acceptable image, (essentially, applying filtered backprojection costs one iteration of the EM, and, in general, at least fifty iterations are required). The Ordered Subsets method, derived from the EM, consists of applying the EM algorithm to subsets of projections ordered in some special manner [22], [21]. In [3] we showed that OS-EM does not converge to an ML solution and that this can be achieved by applying a special relaxation in the direction defined by OS-EM, obtaining the algorithm described below.

Define a partition of the integer interval $M = [1, m]$ into a family of p disjoint subsets S_i , such that

$$M = \cup_{i=1}^p S_i. \quad (5)$$

Starting with a positive $x^{(0)}$, if k denotes the index for a complete cycle and i the index for a sub-iteration, we first define

$$x^{(k,0)} = x^{k-1}, \quad x^{(k,p)} = x^{k+1}, \quad (6)$$

(i.e., a change of the second index indicates that we are choosing the next subset of equations and the first index is up-dated when the algorithm has gone through all the subsets and a new cycle begins) and we have the following relaxed OS-EM,

$$x_j^{(k,i)} = x_j^{(k,i-1)} + \lambda_k x_j^{(k,i-1)} \sum_{l \in S_i} a_{lj} \left(\frac{b_l}{\langle a^l, x^{(k,i-1)} \rangle} - 1 \right), \quad (7)$$

for $j = 1, \dots, n$, $i = 1, \dots, p$, where λ_k is a sequence of positive relaxation parameters such that

$$\lambda_k \xrightarrow[k \rightarrow \infty]{} 0, \quad (8)$$

and

$$\sum_{k=0}^{\infty} \lambda_k = +\infty. \quad (9)$$

Remark 1. It is worth noting that the structure of the above algorithm is defined by: i) decomposing the likelihood function as a sum of functions defined by the sets of projections, ii) computing the gradient of each of these functions, iii) premultiplying this gradient by a diagonal matrix with the current coordinates in the diagonal and iv) performing a relaxation step in the direction defined by this scaled gradient.

Remark 2. Another important observation (apparently not commented in the literature) is that (7) with $\lambda_k = 1$ is not exactly OS-EM. OS-EM is retrieved from (7) taking $\lambda_{k,j} = \frac{1}{\sum_{l \in S_i} a_{lj}}$ for every k . When relaxing OS-EM appropriately, using a relaxation parameter with properties (8-9), it will converge to a weighted maximum Poisson likelihood, with weights given by $\sum_{l \in S_i} a_{lj}$, in the same way that ART approaches a weighted least squares solution (see [7]). That is because the EM is a scaling independent algorithm defined by Kullback-Leibler projections ([8]), that are geometric procedures.

III. A FAST METHOD FOR THE MAXIMUM A POSTERIORI POISSON MODEL IN ECT

As quoted in the Introduction, the Bayesian approach [29], represents a more complete and elegant way to stabilize an inverse problem by using prior information contained in a regularization term. That is, instead of max-

imizing the likelihood (3), $P(x|b)$, is maximized, i.e., via the Bayes formula, the new model becomes

$$\max_{x \geq 0} P(x|b) = \frac{P(x \& b)}{P(b)} = \frac{P_L(b/x) P(x)}{P(b)}, \quad (10)$$

where $P(x)$ and $P(b)$ are the probabilities of x and b respectively. Taking logarithms and eliminating terms not depending on x , we obtain

$$\max_{x \geq 0} G(x) = L(x) + \gamma F(x), \quad (11)$$

where $\gamma F(x)$ represents the information we have about the distribution of x , or, in other words, the approximation of $P(x)$, being γ a positive parameter.

In [14], P. Green presented an algorithm for computing the solution of (11) based on Green's algorithm, that he called One Step Late Expectation Maximization (OSL-EM). It consists of approximating $F(x)$ by its first order Taylor expansion, and then applying the EM algorithm to the resulting function; the main disadvantage of this approach is that the iterations are not necessarily well defined and no global convergence proof was provided so far. In [22], the authors tried to generalize OS-EM to the regularized case, obtaining the following algorithm, that they called OS-GP.

Given x^0 , define for $k = 0, \dots, \dots$, $x^k = x^{k,0}$, and for $i = 1, \dots, p$

$$x_j^{(k,i)} = \frac{x_j^{(k,i-1)} \sum_{l \in S_i} \frac{a_{lj} b_l}{\langle a^l, x^{(k,i-1)} \rangle}}{\sum_{l \in S_i} a_{lj} + \gamma \frac{\partial F(x^{(k,i-1)})}{\partial x_j}}, \quad j = 1, \dots, n, \quad (12)$$

and $x^{k+1} = x^{k,p}$.

Now, it is not only the fact that the OS-GP may not be well defined as OSL, but, it doesn't converge to the MAP solution at all. An example for this behavior is shown in Appendix III. It is also very easy to prove that even relaxing in the directions defined by the algorithm, following the approach in [3] MAP solutions cannot be attained.

Considering the previous observations on OS-GP and the Remarks in Section II on RAMLA's generalization, we are able now to define a natural extension of the algorithm

in [3] to the regularized case.

We define the Block Sequential Regularized Expectation Maximization (from now on BSREM) Algorithm as follows.

Once again, given x^0 , define for $k = 0, \dots, \dots$, $x^k = x^{(k,0)}$, and for $i = 1, \dots, p$

$$x_j^{(k,i)} = x_j^{(k,i-1)} + \lambda_k x_j^{(k,i-1)} \sum_{l \in S_i} a_{lj} \left(\frac{b_l}{\langle a^l, x^{(k,i-1)} \rangle} - 1 \right), \quad (13)$$

for $j = 1, \dots, n$, $i = 1, \dots, p$, Set $x^{k+\frac{1}{2}} = x^{(k,p)}$ and define

$$x^{k+1} = x^{k+\frac{1}{2}} + \lambda_k \gamma D^k \nabla F(x^{k+\frac{1}{2}}), \quad (14)$$

where D^k is the diagonal matrix with the corresponding variables in the diagonal as in [3], that is, $D^k = \text{diag}(x_1^{k+\frac{1}{2}}, \dots, x_n^{k+\frac{1}{2}})$. Positivity is necessary for the iteration to be well-defined, so in the algorithm implementation if any of the elements of x^{k+1} fall below a small threshold, we reset those elements to the threshold value before proceeding to the next iteration. As pointed out by a reviewer, the D^k factor appears quite naturally when one considers cross entropy.

In Appendix I it is proven about the sequence generated by the above algorithm, that, when convergent, its limit point is necessarily a solution.

In [3] (Appendix I), a convergence result was proven for a general form of the algorithm. That is, given a continuously differentiable strictly concave real valued function $G(x)$ and a decomposition of G of the form

$$G(x) = \sum_{i=1}^m G_i(x), \quad (15)$$

(we also assume that the vector functions $D(x) \nabla G_i(x)$ are continuous) then, as before we define the general algorithm as

$$x^{k+1} = x^k + \lambda_k \sum_{i=1}^m D^{(k,i-1)} \nabla G_i(x^{(k,i-1)}). \quad (16)$$

For algorithm (16) the following result was proven in [3]:

Theorem 1 If the sequence $\{G(x^k)\}$ has a unique (finite) limit point (if it is monotone for example), then, the sequence $\{x^k\}$ generated by (16), if bounded and positive, converges to the point maximizing G , provided that (8) and (9) hold.

Remark 1. The above result, proven in [3], requires further explanation. The hypothesis on the convergence of the sequence $\{G(x^k)\}$ does not mean convergence to the maximum. It is just assuming that the sequence is well behaved, with only one (finite) limit point. This is in particular true when the sequence is monotone (a behaviour that has been observed in all the experiments). The theorem states that, with this assumption, convergence of the images to the one of maximum regularized likelihood is guaranteed. As a matter of fact, for a fixed λ there is a neighborhood of the maximum, where the function decreases but increases outside. By decreasing the relaxation parameter, the size of that neighborhood goes to zero. In [3], Theorem 1 was complemented by proving that, when G is just the likelihood, then, the sequence is always bounded. In our penalized case, this result cannot be proven in its general form, but, boundedness is implied by the assumption on the behavior of $\{G(x^k)\}$, as follows. If $\{G(x^k)\}$ is convergent, it is bounded. Consider $G^* = \inf_k G(x^k)$, and the set $B = \{x / G(x) \geq G^*\}$, then $\{G(x^k)\} \in B$. But, G is not only strictly concave, but tends to $-\infty$ when x tends to ∞ (a property of the likelihood function and concave priors as well), so, its level sets are bounded (in particular B). Therefore, the sequence $\{x^k\}$ is bounded. In Appendix I, we present a weaker and shorter result, to illustrate, in an easier way, that relaxation with properties (8) and (9) is necessary, in order to get a MAP solution when using ordered subsets-type methods.

Remark 2. The necessity of the strong hypothesis on the behavior of $\{G(x^k)\}$ comes from the difficulties to deal with the boundary. This was the same drawback as with the EM

algorithm for several years, until Csiszar and Tusnady [8] proved full convergence in 1984. Shepp and Vardi's proof in their well known paper of 1982 [38] assumes implicitly that the limit points belong to the interior of the positive orthant (if not, Zangwill's result cannot be applied). For our algorithm, it is easy to check in the proofs in [3], that, if the limit points are not in the boundary, the strong hypothesis on $\{G(x^k)\}$ is no longer needed.

Remark 3. Another important observation is that $x^{k+\frac{1}{2}}$ is always strictly positive (this is proven in [3]) whenever x^k is positive. So, if $\gamma = 0$ (no regularization, that is the case in [3]) the next iterate is always well defined (strictly positive). In the general case, we have to assume that x^{k+1} is strictly positive. In our experiments, we kept a safety threshold to make sure that this happens and the iterations are well defined. In spite of this control in our experiments, nonpositive variables were never detected.

IV. SIMULATION EXPERIMENTS

We present in this Section the results of two sets of experiments (I and II). The first set aims at showing that BSREM fulfils the goal of maximizing the MAP function whereas OS-GP does not, that is the main purpose of this article. The second set of experiments is part of a statistical study illustrating the behavior of OS-GP and BSREM when applied to simulated PET data of the brain. This study helped us to better understand the behavior of BSREM with respect to its parameters. Our choice of comparing to OS-GP in both cases was because the latter is the proposed extension of OS-EM to the MAP problem, in the same way as BSREM is the natural extension of RAMLA.

A. Experiments I

In order to support experimentally our claim that BSREM converges to the maximum of the MAP function (11) whereas OS-GP does not, we ran numerous examples, testing BSREM and OS-GP against De Pierro's modified EM method (DPEM) [10], much slower, but with a known

complete convergence proof. For these experiments, we used a quadratic prior, as in [17] with regularization parameter $\gamma = 0.3$. Simulations were done using brain phantoms as described in the second set of experiments. The choice of a quadratic prior for this set of experiments was motivated by the fact that it has very large derivatives (larger than any other reasonable prior suggested in the literature), the best test for possible divergence (see Appendix II). In spite of that, the algorithm always appeared to converge (and very fast), even when using different starting images. Better priors (models) tend to have smaller derivatives, in order to penalize less the discontinuities, that is the case of *logcosh* (almost constant derivatives), or piecewise Gaussian priors (derivative zero outside a small neighborhood) [11]. Results for a typical run are shown in Figure 1 for the MAP function values. It can be seen that, after two iterations, OS-GP values remain constant, far below those attained by BSREM and DPEM. On the other hand, BSREM attains very high MAP values in only two iterations, with DPEM attaining similar values after more than 20 iterations, remaining almost the same for at least 200 iterations. In all cases (200 to 1000 iterations of DPEM), the square distances between the images obtained by DPEM and BSREM are much smaller than those obtained by DPEM and OS-GP, meaning that BSREM is approaching the same limit as the convergent algorithm DPEM. The number of blocks was chosen to be 16, because it was the one giving faster and better results for OS-GP.

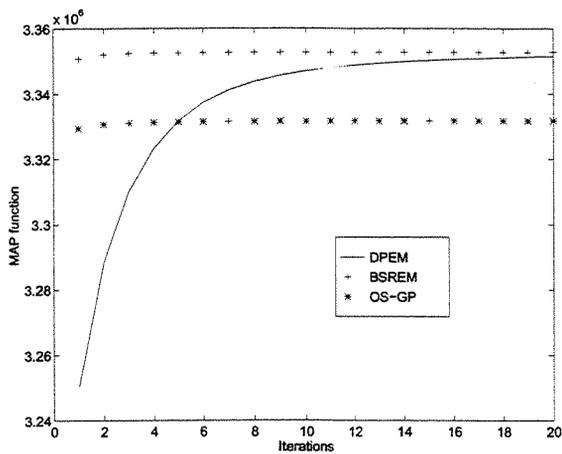


Fig. 1. Regularized likelihood vs number of iterations for DPEM, BSREM and OS-GP

B. Experiments II

For the second set of experiments we have used SuperSnark, that is part of a larger software package, SNARK93 [5], a programming system developed and maintained by the Medical Image Processing Group at the University of Pennsylvania. SNARK93 was designed to help researchers in developing, testing and evaluating algorithms for 2-D tomographic reconstruction.

The experiments aim at comparing OS-GP and BSREM from the point of view of performing a specific imaging task using the Figure of Merit [FOM] that measures the estimation of radionuclide uptake at individual points within neurological structure of interest (PA FOM). This FOM is a weighted distance between the phantom and the reconstruction. The goal of this set of experiments was to show that, for a reasonable prior, BSREM keeps the distance to the phantom much closer than OSGP.

A description of the FOM, the way it was computed and the statistical methodology can be found in Appendix IV and more details on the evaluation methodologies using FOM's and SuperSnark in [4].

The prior chosen for these experiments was the *logcosh* suggested by Green [14] (we used an 8-pixels neighborhood with weights one for pixels with a common edge and $\frac{1}{\sqrt{2}}$ for pixels with a common vertex) and there were two free parameters for BSREM (regularization and initial relaxation)

and one for OS-GP (just regularization). The regularization parameters were chosen according to the results in a large sets of experiments (*training experimentation* [4]). For γ in the range $[.1, .8]$ the best results for both algorithms were obtained, and the relative behaviors were similar; for greater values, images become oversmoothed (especially those generated by OS-GP).

For BSREM, the relaxation sequence was not optimized for each iteration as in [3], but, our experiments indicated that the right strategy is to take $\lambda_k = \frac{\epsilon}{k}$, that satisfies the convergence conditions, where ϵ is the only new (positive) parameter to be chosen. Also according to our experimental results, a reasonable value for ϵ with brain data was 0.6. As before the number of blocks was chosen to be 16, because it was the one giving faster and better results for OS-GP.

Poisson noise was introduced into measurements and the total number of counts was 1,187,722.

Figure 2 shows the results for the FOM with noise as described above. The significance level for the statistical study appears on the top.

Figure 3 shows the behavior of the objective function for a typical run of both algorithms for this set of experiments, OS-GP and BSREM. Figure 4 shows several images obtained during the experiments. Comparing the reconstructed images in the first row, (iii) looks smoother than (v); (ii) shows the effect of noise after 20 iterations of OS-EM and, (iv) and (vi) the regularized results after the same number of iterations. The number of blocks used in the experiments was 16 and no special ordering as in [21] or [18] was used. This number (16) was experimentally the best (faster to achieve high MAP values) for OS-GP, that critically depends on it. For BSREM, larger number of sets can be used, but with smaller relaxation parameters. Figure 5 shows the images obtained after 500 iterations of DPEM, and 20 iterations of BSREM and OSGP using a quadratic prior, $\gamma = 20.0$; for BSREM, λ_0 was chosen 0.015 and $\lambda_k = \frac{\lambda_0}{k^3}$. The phantom is the same of Figure 4. So,

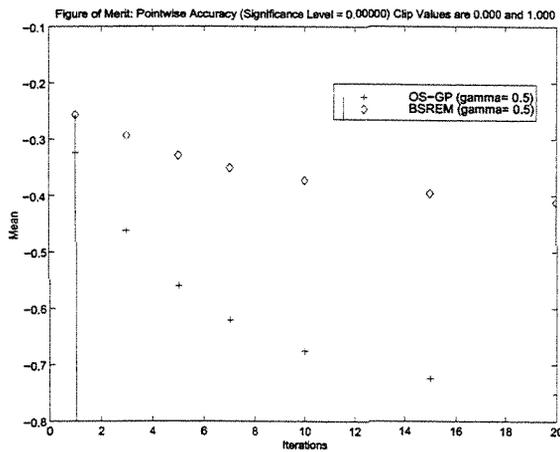
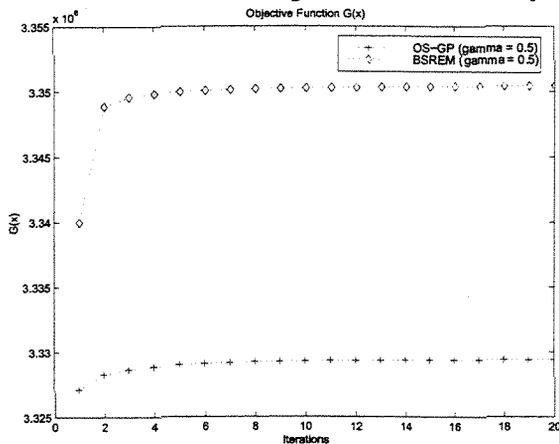


Fig. 2. Pointwise Accuracy

Fig. 3. Objective function $G(x)$

even for a quadratic prior and slowly decreasing values of the relaxation parameters, no divergence was observed and the images obtained are very similar to those generated by DPEM.

V. RESULTS AND DISCUSSION

As pointed out at the beginning, our main goal in this work was to present a natural extension of OS-EM-RAMLA type algorithms to the MAP problem, comparing it with the previous attempt to do this, represented by OS-GP. The mathematical proofs in [3] and the Appendices, as well as the numerical evidences, show that BSREM converges to the maximum value and Figures 1 and 3 illustrate the fact that it attains very high values of the objective function in few iterations. On the other hand, for OS-GP,

the values of the objective function remain almost fixed and much lower. These asymptotic properties (convergence to a maximum value) are not so important when dealing with the the nonregularized problem, where the maximum likelihood is not needed, but, just high likelihood values are necessary, which are very rapidly attained by OS-EM. However, when trying to add prior information in the model, attaining high values of the MAP function (and asymptotically the maximum) becomes crucial in order to obtain a better fitting of the model to this information. So, our first set of experiments (Figure 1) shows that BSREM is an appropriate reconstruction algorithm for a given prior and that it converges to a MAP solution, even for the quadratic one.

The second set of experiments, with a better prior, shows that the weighted distance to the phantom is kept bounded by BSREM, and relatively constant with respect to the first iteration, also smaller than the distance to OS-GP iterates, that tends to diverge also faster (Figure 2).

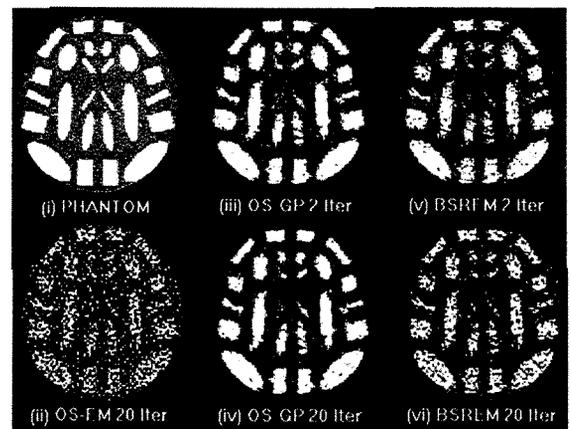


Fig. 4. PET Images. Logcosh prior



Fig. 5. PET images. Quadratic prior

VI. CONCLUDING REMARKS

We have presented a very fast algorithm for computing MAP solutions in Emission Tomography. The new algorithm attains high values of the MAP function at a very low number of iterations, providing high quality PET images, better than those generated by the OS-GP method, its closer relative in the family of the ordered subsets EM-like methods. It is worth noting that (13-14) is not the only possible algorithm derived using the same idea and many different variations and modifications can give rise to different, and maybe still better, algorithms. One possibility is choosing different ways of decomposing the prior function; for example, after iterating through one block of projections, iterate that part of the smoothing function that is related only to the pixels that are intersected by the rays in that block. Our choice of iterating using a unique block for the prior was motivated by its simplicity and by the fact that it works like filtering in between steps of the main iteration method, but, with the advantage, that it converges to a solution of the chosen MAP model (with a known meaning) and this cannot be proven for EMS-like algorithms (see [40], [28], [34], [12]). Another advantage of iterating separately the ML and the prior, that we have seen, and can be exploited, is the fact that this may help to take the iterations out of local minima when dealing with nonconvex priors. Of course, this departs from the observation that the algorithm converges to local minima or critical points (see [11], for the general nonconvex case (convexity is not needed in Appendix I) and modifications of the convergence proofs in [3] can be made in that direction.

Taking into account the previous remarks, future directions we are considering for our research are: (1) a statistical comparison of BSREM, with other well known fast methods like SAGE [13] and coordinate ascent [2]; (2) application of BSREM to SPECT problems (our choice for PET in this paper comes our previous experience on the problem and its simplicity when we are dealing with speed comparison); (3) experimenting with different ways to take

advantage of the decomposition of the prior as mentioned before; (4) the use of BSREM as a secondary algorithm to deal with more complex models with nonconvex priors. Also, a crucial issue, from the point of view of the application of BSREM (as well as RAMLA) is the choice of the relaxation parameters. Preliminary experiments seem to indicate that sequences with slower convergence rates than $\frac{1}{k}$ can work better from the point of view of achieving higher MAP values.

From the theoretical point of view, we are now also working in the direction of improving our convergence results.

ACKNOWLEDGEMENTS

We are grateful to Richard Leahy for helpful comments regarding an early version of this article and to the anonymous reviewers for their valuable comments and suggestions. Many improvements are due to the valuable criticisms of Jeff Fessler.

APPENDIX I

In [3], a proof of convergence of the algorithm (16) was given, assuming a reasonable behavior of the objective function. Completing those results, and in order to show a simpler evidence of the behavior of the algorithm, we give a brief and simple proof of the fact that the sequence generated by (13-14) necessarily converges to the MAP solution (when convergent). For the sake of simplicity, we consider the general algorithm (16), and, of course we assume continuous differentiability of the functions.

Suppose now that x^k tends to x^* , then $x^{(k,i)}$ tend to x^* for every i , because $x^{(k,i-1)} - x^{(k,i)}$ tends to zero for every i (and this is a consequence of the fact that λ_k tends to zero).

Let us now expand the algorithm's series from the starting point to k :

$$x^{k+1} = x^0 + \sum_{l=0}^k \lambda_l \sum_{i=1}^m D^{(l,i-1)} \nabla G_i(x^{(l,i-1)}). \quad (17)$$

For the sake of simplicity, let us define $a_l = \sum_{i=1}^m D^{(l,i-1)} \nabla G_i(x^{(l,i-1)})$. So, for k tending to infinity we get

$$x^* = x^0 + \sum_{l=0}^{\infty} \lambda_l a_l, \quad (18)$$

Using the fact that x^k tends to some finite limit x^* , and the continuity of the vector functions $D(x) \nabla G_i(x)$, we deduce that a^l converges to some limit vector a . For a given j , suppose that $a_j > 0$. Therefore, there exists an integer s such that, for $l > s$, a_j^l is greater than a positive number \bar{a}_j . Then

$$\sum_{l=s}^{\infty} \lambda_l a_j^l > \bar{a}_j \sum_{l=s}^{\infty} \lambda_l, \quad (19)$$

and the left hand side tends to infinity, because of (9); a contradiction. The same argument applies assuming that $a_j < 0$. So a should be zero.

From the arguments above, we deduce that, for $j = 1, \dots, n$,

$$x_j^* \nabla_j G(x^*) = 0. \quad (20)$$

If $x_j^* > 0$, then, clearly, from (20), $\nabla_j G(x^*) = 0$. Now, suppose that $x_j^* = 0$ for some j and that $\nabla_j G(x^*) > 0$; by continuity arguments we know that $\sum_{i=1}^m \nabla G_i(x^{(k,i)})_j - \nabla G(x^k)_j$ tends to zero. So, for k large enough,

$$\sum_{i=1}^m \nabla_j G_i(x^{(k,i)}) > \eta > 0. \quad (21)$$

On the other hand, algorithm (16) can be expressed as

$$x_j^{k+1} = x_j^k \prod_{i=1}^m (1 + \nabla_j G_i(x^{(k,i-1)})), \quad (22)$$

or

$$x_j^{k+1} = x_j^k (1 + \lambda_k \sum_{i=1}^m \nabla_j G_i(x^{(k,i-1)}) + o(\lambda_k^2)). \quad (23)$$

That is, using (23) and boundedness, there exists k_0 , such that

$$x_j^{k+1} > x_j^k \quad \forall k > k_0, \quad (24)$$

therefore, x_j^k cannot tend to zero, a contradiction that implies that $\nabla_j G(x^*) \leq 0$. So, x^* satisfies the Kuhn-Tucker

conditions (see [1]), and it is a maximizer if the problem is concave, or at least a critical point in the general case.

APPENDIX II

In the following we analyse, by means of a simple example, the boundedness issue. Let us consider the function

$$f(x) = \ln x - x - \gamma x^2. \quad (25)$$

The function above is just the Poisson likelihood in one variable penalized by a quadratic term and applying BSREM to it gives:

$$x^{k+\frac{1}{2}} = (1 - \lambda_k)x^k + \lambda_k, \quad (26)$$

$$x^{k+1} = x^{k+\frac{1}{2}} - 2\lambda_k \gamma (x^{k+\frac{1}{2}})^2. \quad (27)$$

Substituting (24) in (25) for a fixed λ we obtain the following iteration function

$$\phi(x) = (1 - \lambda)x + \lambda - 2\lambda\gamma((1 - \lambda)x + \lambda)^2. \quad (28)$$

Computing the derivative of $\phi(x)$, we get the sufficient condition for ϕ to be a contraction

$$|\phi'(x)| = |(1 - \lambda) - 4\lambda\gamma(1 - \lambda)((1 - \lambda)x + \lambda)| \leq 1, \quad (29)$$

that is; for

$$\frac{\lambda}{1 - \lambda} - \frac{1}{4(1 - \lambda)^2 \gamma} \leq x \leq \frac{2 - \lambda}{4(1 - \lambda)^2 \lambda \gamma} - \frac{\lambda}{1 - \lambda}, \quad (30)$$

ϕ is a contraction, and we expect for the sequence starting in this region to be bounded. We perform several experiments for this small example, confirming our predictions. As a matter of fact, when preventing the algorithm of generating nonnegative iterates (setting the variable to a small number if this is the case), it is possible to obtain wider convergence regions. As an illustration, for $\lambda = 0.5$ and $\gamma = 1.$, convergence to the right solution ($x = 0.5$) occurs for every starting point because the iterate is set immediately to a small number if the starting point is large. When applied to the above function, starting with a reasonable point like $x = 1.$, that is the solution of problem for $\gamma = 0$ (no regularization), the sequence converges fast

at the beginning and then it slows down. This is approximately the behavior that we have observed in realistic PET simulations. As a matter of fact, the first part of the first iteration obtains a reasonable approximation for the image, avoiding any possible divergence or nonboundedness.

APPENDIX III

Next, we present a counterexample to the convergence of OS-GP to a MAP solution. Let us consider the following inconsistent system of equations.

$$\begin{cases} \frac{1}{2}x_1 + x_2 = 1 \\ \frac{1}{2}x_1 + x_2 = 2 \end{cases} \quad (31)$$

and the potential function $V(x) = \frac{1}{2}(x_1 - x_2)^2$. Then, the MAP problem in this case consists of maximizing the function

$$G(x_1, x_2) = 3\ln\left(\frac{x_1}{2} + x_2\right) - 2\left(\frac{x_1}{2} + x_2\right) - \frac{\beta}{2}(x_1 - x_2)^2. \quad (32)$$

The solution for the problem above is $x_1^* = x_2^* = 1$ for every positive β . If we apply OS-GP for $\beta = 0.5$ and the starting point $x^0 = (\frac{1}{3}, \frac{11}{6})$, the sequence oscillates between the points $\hat{x} = (0.5986, 0.7535)$ and $\tilde{x} = (1.3456, 0.6643)$. Changing β to other values the sequence converges to different pairs of limit points. For example: for $\beta = .6$ (a small change of the ‘smoothing’ parameter) and the same starting point as before, the sequence strongly oscillates producing large and small values (even negative ones) successively. As a conclusion, OS-GP is not an adequate procedure for solving the MAP optimization problem.

VII. APPENDIX IV

A ‘basic phantom’ was randomly selected from a list of phantom files, each of which containing the description of a specified basic brain phantom. The *Pointwise Accuracy* of a reconstructed image is defined as the negative of the normalized root mean squared distance between the

reconstructed image and the digitized phantom ; that is,

$$- \left(\frac{\sum_{j=1}^n (p_j - r_j)^2}{\sum_{j=1}^n (p_j - \hat{p})^2} \right)^{\frac{1}{2}}. \quad (33)$$

For the simulations 26 geometrical phantoms were selected from an atlas containing typical brain structures; from this set 50 phantoms were randomly generated using typical PET values [4]. For each phantom, projection data were generated assuming that the scanner consists of a ring of 300 detectors with each detector in coincidence with 101 detectors opposite to it. Phantoms were digitized using a 95x95 pixels grid.

To assess the statistical significance of our experiments we followed the methodology in [5].

REFERENCES

- [1] M. Avriel, *Nonlinear Programming: Analysis and Methods*, Prentice Hall, Englewood Cliffs, NJ, 1976.
- [2] C.A. Bouman and K. Sauer, ‘‘A unified approach to statistical tomography using coordinate descent optimization’’, *IEEE Trans. Image Proc.*, vol. 5, 3, pp. 480-492, 1996.
- [3] J.A. Browne and A.R. De Pierro, ‘‘A row-action alternative to the EM algorithm for maximizing likelihoods in emission tomography’’, *IEEE Transactions on Medical Imaging*, V15, 4, 687-699, October 1996.
- [4] J.A. Browne and G.T. Herman, ‘‘Computerized evaluation of image reconstruction algorithms’’, *International Journal of Imaging Systems and Technology*, vol.7, pp. 256-267, 1996.
- [5] J.A. Browne, G.T. Herman and D. Odhner, ‘‘SNARK93- A programming system for image reconstruction from projections’’, Technical Report No. MIPG198, Department of Radiology, University of Pennsylvania, 1993.
- [6] T.F. Budinger, G.T. Gullberg and R.H. Huesman, ‘‘Emission computed tomography’’, in *Image Reconstruction from Projections: Implementation and Applications*, G. T. Herman (ed), Chapter 5, pp. 147-246, Springer Verlag, Berlin, Heidelberg, 1979.
- [7] Y. Censor, P.P.B. Eggermont and D. Gordon, ‘‘Strong underrelaxation in Kaczmarz’s method for inconsistent systems’’, *Numer. Math.* 41, pp. 83-92, 1983.
- [8] I. Csisz ar and G. Tusn ady, ‘‘Information geometry and alternating minimization procedures’’, in *Statistics & Decisions*, Supplement Issue No. 1, R. Oldenbourg Verlag, pp. 205-237, 1984.
- [9] A.P. Dempster, N.M. Laird and D.D. Rubin, ‘‘Maximum likelihood from incomplete data via the EM algorithm’’, *J. Roy. Stat. Soc., Series B*, vol. 39, pp. 1-38, 1977.
- [10] A.R. De Pierro, ‘‘A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography’’, *IEEE Trans. Med. Imaging*, vol. 14, 1, pp. 132-137, 1995.

- [11] A.R. De Pierro and M.E. Belezá Yamagishi, "Fast iterative methods applied to tomography models with general Gibbs priors", Conference on Mathematical Modeling, Bayesian Estimation and Inverse Problems, Proceedings of SPIE, Vol. 3816, pp. 134-138, 1999.
- [12] P.P.B. Eggermont and V.N. LaRiccia, "Maximum smoothed likelihood density estimation for inverse problems", *Annals of Statistics*, 23,, 1, pp. 199-220, 1995.
- [13] J.A. Fessler and A.O. Hero, "Space-alternating generalized expectation maximization algorithm", *IEEE Trans. on Signal Proc.*, vol. 42, pp. 2664-2676, 1994.
- [14] P.J. Green, 'Bayesian reconstructions from emission tomography data using a modified EM algorithm', *IEEE Trans. Medical Imaging*, 9, 1, pp. 84-93, 1990.
- [15] T.J. Hebert, R. Leahy and M. Singh, "Three-dimensional maximum-likelihood reconstruction for an electronically collimated single-photon-emission imaging system", *J. Optical Society of America*, 7, 7, pp. 1305-1313, 1990.
- [16] G.T. Herman, *Image Reconstruction from Projections: The Fundamentals of Computerized Tomography*, Academic Press, New York, NY, 1980.
- [17] G.T. Herman, A.R. De Pierro and N. Gai, "On methods for maximum a posteriori image reconstruction with a normal prior", *Journal of Visual Communication and Image Representation*, V3, 4, 316-324, 1992.
- [18] G.T. Herman and L.B. Meyer, "Algebraic reconstruction techniques can be made computationally efficient", *IEEE Trans. Med. Imaging*, vol. 12, 3, pp. 600-609, 1993.
- [19] G.T. Herman and D. Odhner, "Performance evaluation of an iterative image reconstruction algorithm for positron emission tomography", *IEEE Trans. Med. Imaging*, vol. 10, 3, pp. 336-346, 1991.
- [20] S. Holte, P. Schmidlin, A. Linden, G. Rosenqvist and L. Eriksson, "Iterative image reconstruction for positron emission tomography: a study of convergence and quantitation problems", *IEEE Trans. Nuclear Science*, 37, 2, pp. 629-635, 1990.
- [21] H.M. Hudson, B.F. Hutton and R. Larkin, "Accelerated EM reconstruction using ordered subsets", *J. Nucl. Med.*, vol. 33, p. 960, (abstract), 1992.
- [22] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data", *IEEE Trans. Med. Imaging*, vol. 13, n^o 4, pp. 601-609, 1994.
- [23] C.A. Johnson, J. Seidel and A. Sofer, "Interior-point methodology for 3-D PET reconstruction", preprint, Center for Information Technology, National Institutes of Health, Bethesda, MD, 1999.
- [24] L. Kaufman, "Implementing and accelerating the EM algorithm for positron emission tomography", *IEEE Trans. Med. Imaging*, vol. 6, pp. 37-51, 1987.
- [25] K. Lange and R. Carson, "EM reconstruction algorithms for emission and transmission tomography", *J. Comput. Assisted Tomog.*, vol. 8, pp. 306-316, 1987.
- [26] D.S. Lalush and B.M.W. Tsui, "A fast and stable maximum a posteriori conjugate gradient reconstruction algorithm", *Med. Phys.*, vol. 22, 8, pp. 1273-1284, 1995.
- [27] D.S. Lalush and B.M.W. Tsui, "The importance of preconditioners in fast Poisson-based iterative reconstruction algorithms for SPECT", in **1995 IEEE Nuclear Science Symposium Conference Record**, pp. 1326-1330, 1996.
- [28] G.A. Latham and R.S. Anderssen, "Assesing quantification for the EMS algorithm", *Linear Alg. Applics.*, 210, pp. 89-122, 1994.
- [29] E. Levitan and G.T. Herman, "A maximum a posteriori probability expectation maximization algorithm for image reconstruction in emission tomography", *IEEE Trans. Med. Imaging*, vol. 6, n^o 3, pp. 185-192, 1987.
- [30] R.M. Lewitt and G. Muehlethner, "Accelerated iterative reconstruction for positron emission tomography based on the EM algorithm for maximum likelihood estimation" *IEEE Trans. Med. Imaging*, vol. 5, pp. 16-22, 1986.
- [31] S. Matej and J. A. Browne, "Performance of a fast maximum likelihood algorithm for fully 3D PET reconstruction", *Proc. Internat. Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, Aix-les-Bains, France, pp. 247-251, 1995.
- [32] E.Ü. Mumcuoglu, R.M. Leahy, S.R. Cherry and Z. Zhou, "Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET images", *IEEE Trans. Med. Imag.*, vol. 13, 4, pp. 687-701, 1994.
- [33] E.Ü. Mumcuoglu, R.M. Leahy and S.R. Cherry, "Bayesian reconstruction of PET images: methodology and performance analysis", *Phys. Med. Biol.*, vol 41, pp 1777-1807, 1996.
- [34] D. Nychka, "Some properties of adding a smoothing step to the EM algorithm", *Statist. Probab. Lett.* 9, pp. 187-193, 1990.
- [35] N. Rajeevan, K. Rajgopal and G. Khishna, "Vector-extrapolated fast maximum likelihood estimation algorithms for emission tomography", *IEEE Trans. Med. Imaging*, vol. 11, n^o 3, pp. 9-20, 1992.
- [36] A.J. Rockmore and A. Macovski, "A maximum likelihood approach to emission image reconstruction from projections", *IEEE Trans. Nucl. Sci.*, vol. 23, pp. 1428-1432, 1976.
- [37] M.S. Rosenthal, J. Cullom, W. Hawkins, S.C. Moore, B.M.W. Tsui and M. Yester, "Quantitative SPECT imaging: a review and recommendations by the focus committee of the Society of Nuclear Medicine Computer and Instrumentation Council", *J. Nucl. Med.*, 36, pp. 1489-1513, 1995.
- [38] L.A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography", *IEEE Trans. Med. Imaging*, vol. 1, pp. 113-121, 1982.
- [39] L.A. Shepp, Y. Vardi, J.B. Ra, S.K. Hilal and Z.H. Cho, "Maximum likelihood PET with real data", *IEEE Trans. Nucl. Sci.*, vol. 31, pp. 910-913, 1984.
- [40] B.W. Silverman, M.C. Jones, J.D. Wilson and D.W. Nychka, "A smoothed EM approach to indirect estimation problems, with par-

tical reference to stereology and emission tomography", J.R. Statist. Soc. B, 52, 2, pp. 271-324, 1990.

- [41] M.M. Ter-Pogossian, M. Raichle and B.E. Sobel, "Positron emission tomography", Scientific Amer., vol. 243, n^o 4, pp. 170-181, 1980.
- [42] Y. Vardi, L.A. Shepp and L. Kaufman, "A statistical model for positron emission tomography", J. Amer. Statist. Assoc., vol. 80, pp. 8-20, 1985.

Capítulo 12

Apêndice D

Fast scaled gradient decomposition methods for maximum likelihood transmission tomography

Michel E Beleza Yamagishi† and Alvaro R De Pierro†§

†University of Campinas, Department of Applied Mathematics, CP 6065, CEP 13081-970, Campinas, SP, Brazil

E-mail: alvaro@ime.unicamp.br

Abstract. New iterative algorithms are presented for Maximum Likelihood (ML) and Regularized Maximum Likelihood (MAP) reconstruction in Transmission Tomography (CT). The algorithms are natural extensions to CT of RAMLA, a well known method for ML reconstruction in Emission Computed Tomography (ECT). We show that the new algorithm for ML solutions produces similar, or even better results than the EM algorithm, but in much fewer iterations. Also, its convergence properties are better than other ordered subsets methods.

PACS number: 42.30, 87.5

Submitted to: *PMB*

§ To whom correspondence should be addressed

1. Introduction

Lange and Carson [10] were the first to propose the use of a ML approach for the low statistics CT reconstruction problem and the EM algorithm for computing the solution. The fact that the maximization step involves the solution of single variable transcendental equations led to several approximations improving the original version ([2], [14]), In [14] Ollinger, based on experimental evidence, suggested that the approximated version of the EM of Lange and Carson was not convergent and proposed the use of a relaxed Newton method for the line search. In spite of the improvement, the computational cost of each iteration, as observed in [13], and the slow increase of the likelihood function were important drawbacks.

In 1987 Lange et al. [9] proposed an algorithm based on a scaled gradient direction and a line search, avoiding the truncation problems of the approximation in the EM (see also [13]), but still computationally expensive and slow. In [11] an algorithm based on the minorizing functions approach of [4] was proposed for the regularized problem, at a reasonable cost for each iteration, convergent, but also slow.

Manglos et al [12] used an ordered subsets approach [8] to obtain an accelerated method for ML solutions. Erdogan and Fessler [7] combined a modification of the minorizing functions approach with ordered subsets to obtain regularized ML solutions. But, as shown in [5], unless appropriate scaling and relaxation is used, these methods do not converge to a solution of the proposed model.

Our contribution in this article is a new algorithm for ML reconstruction in CT, that is fast because it uses ordered subsets as [12], it has guaranteed positivity, and shares the convergence properties of RAMLA. Also, we adapt the new algorithm to the regularized problem. In section 2 we formulate the problem and present the

main algorithm. In section 3 we describe the convergence results that are complemented in Appendix I. section 4 is devoted to the extension of the main algorithm to the MAP problem. In section 5 we describe the results of simulations comparing the performance of our algorithms with the EM algorithm and we also introduce a preliminary version of a new and promising algorithm for the MAP problem as an alternative for T-BSREM. Section 6 is dedicated to some concluding remarks.

2. The New Algorithm: T-RAMLA

We aim at finding a solution of the transmission tomography ML problem. As in [10], the ML approach looks for solutions maximizing the function

$$L(x) = \sum_i^m -d_i \exp(-\langle a^i, x \rangle) - y_i \langle a^i, x \rangle, \quad (1)$$

where d_i is the number of photons emitted along the i -th line (blank scan) and y_i the number of detected photons. a^i $i = 1, \dots, m$ are the rows of the projection matrix and x the image vector (n pixels). \langle, \rangle denotes the standard inner product. It is easy to prove that L is a concave function.

Consider now the partition of the integer interval $M = [1, 2, 3, \dots, m]$ as a union of p disjoint subsets S_i , $i = 1, \dots, p$.

Let us now decompose L as

$$L(x) = \sum_{i=1}^p L_i(x), \quad (2)$$

where

$$L_i(x) = \sum_{k \in S_i} -d_k \exp(-\langle a^k, x \rangle) - y_k \langle a^k, x \rangle. \quad (3)$$

Calculating the gradient of the functions $L_i(x)$ gives

$$(\nabla L_i(x))_j = \frac{\partial L_i(x)}{\partial x_j} = \sum_{k \in S_i} a_{kj} d_k \exp(-\langle a^k, x \rangle) - a_{kj} y_k \quad (4)$$

Using the previous definitions and denoting the starting image as $x^{(0)}$, the new algorithm is defined as follows. k and i are the indices for the iterations and subiterations respectively.

$$x_j^{(k,i+1)} = x_j^{(k,i)} + \lambda_k x^{(k,i)} \sum_{l \in S_i} a_{lj} d_l \exp(-\langle a^l, x^{(k,i)} \rangle) - a_{lj} y_l \quad (5)$$

where $j = 1, \dots, n$, $i = 1, \dots, p$ (pixels and subsets respectively) and $x^{(k,p)} = x^{k+1}$. That is, the second index goes through the subsets of equations (projections) and the first one is associated to a complete cycle. λ_k is a sequence of positive relaxation parameters such that

$$\lambda_k \xrightarrow[k \rightarrow \infty]{} 0, \quad (6)$$

and

$$\sum_{k=0}^{\infty} \lambda_k = +\infty. \quad (7)$$

We will call the new algorithm *T-RAMLA* because it is a natural extension of RAMLA (from Row Action Maximum Likelihood Algorithm) [1] to transmission tomography (in spite of the fact that we are allowing the possibility of several rows instead of a single one).

3. Convergence Results

In [1] (Appendix I), a convergence result was proven for a general form of the previous algorithm. That is, given a continuously differentiable strictly concave real

valued function $G(x)$ and a decomposition of G of the form

$$G(x) = \sum_{i=1}^m G_i(x), \quad (8)$$

then, a general algorithm can be defined as follows

$$x^{k+1} = x^k + \lambda_k \sum_{i=1}^m D^{(k,i-1)} \nabla G_i(x^{(k,i-1)}). \quad (9)$$

where $D^{(k,i-1)}$ is the diagonal matrix with its diagonal elements being the coordinates of $x^{(k,i-1)}$ (in general $D(x)$ will be the diagonal matrix of the coordinates of x). And the result is the following.

Theorem 1 Assuming that the vector functions $D(x) \nabla G_i(x)$ are continuous, if the sequence $\{G(x^k)\}$ has a unique (finite) limit point (if it is monotone for example), then, the sequence $\{x^k\}$ generated by 9, if bounded and positive, converges to the point maximizing G , provided that 6 and 7 hold.

In our case $G_i(x)$ is defined by (2) and we need to prove that algorithm 5 satisfies the hypotheses of the Theorem, that is, that the generated sequence remains positive and bounded, as well as the continuity property. This is shown in Appendix I. In Appendix II we present some simple examples illustrating the limiting behavior of the methods mentioned in this article as compared to previous ones.

4. The Bayesian Approach: T-BSREM.

We consider in this section the extension of our algorithm for the computation of Maximum ‘A Posteriori’ (MAP) solutions, the Bayesian approach. Instead of maximizing the likelihood, we aim at maximizing the penalized likelihood function,

$$G(x) = L(x) + \gamma R(x), \quad (10)$$

where $L(x)$ is the likelihood function, $R(x)$ is a function containing ‘prior’ information and γ a real positive parameter allowing a trade-off between the ‘prior’ information and the consistency with the data.

Using the same notation of T-RAMLA and following [5] we define T-BSREM. Given a positive starting point x^0 and considering that

$$x^{(k,0)} = x^k, \quad x^{(k,p)} = x^{k+\frac{1}{2}}, \quad (11)$$

the sequence is defined as

$$x_j^{(k,i)} = x_j^{(k,i-1)} + \lambda_k x_j^{(k,i-1)} \sum_{l \in S_i} a_{lj} d_l \exp\left(-\langle a^l, x^{(k,i-1)} \rangle\right) - a_{lj} y_l \quad (12)$$

for $j = 1, \dots, n$ e $i = 1, \dots, p$. Then we define

$$x^{k+1} = x^{k+\frac{1}{2}} + \lambda_k \gamma D^{k+\frac{1}{2}} \nabla R\left(x^{k+\frac{1}{2}}\right), \quad (13)$$

where $D^k = \text{diag}(x_1^k, \dots, x_n^k)$. The sequence of relaxation parameters satisfies conditions 6 and 7.

5. Simulations

We performed several experiments comparing T-RAMLA against T-EM (the EM algorithm for transmission tomography ML), as well as its ordered subsets (OS) version. Our goal is to show that our algorithm and its extension, respectively for ML and MAP solutions is faster than the EM, attaining higher values of the objective function, even if compared to the OS version.

All the simulations were performed using SNARK93 [3], a reconstruction software developed by the Medical Image Processing Group of the University of Penn-

sylvania.

T-RAMLA versus T-EM

It is well known that the EM for the transmission tomography ML problem is very slow [7], and many iterations are necessary to obtain high likelihood levels and image quality.

The EM algorithm for CT (T-EM) [10] has the following form.

$$x_j^{k+1} = \frac{\sum_{i=1}^m (M_{ij} + N_{ij})}{\frac{1}{2} \sum_{i=1}^m (M_{ij} - N_{ij}) a_{ij}}, \quad (14)$$

where

$$M_{ij} = b_i + d_i \left[\exp \left(- \sum_{l \in S_{ij}} a_{il} x_l^k \right) - \exp \left(- \langle a^i, x^k \rangle \right) \right] \quad (15)$$

and

$$N_{ij} = b_i + d_i \left[\exp \left(- \sum_{l \in S_{ij} \cup \{j\}} a_{il} x_l^k \right) - \exp \left(- \langle a^i, x^k \rangle \right) \right], \quad (16)$$

the set S_{ij} denotes the pixels between the source and pixel j (the j -th pixel does not belong to S_{ij}) along projection i .

Equation 14 is an approximation to the solution of the transcendental equation

$$\sum_{i=1}^m -N_{ij} a_{ij} + \sum_{i=1}^m \frac{(M_{ij} - N_{ij}) a_{ij}}{\exp(a_{ij} x_j) - 1} = 0. \quad (17)$$

Ollinger [14] argues that is safer to solve the equation above using Newton's method. We performed several experiments comparing T-RAMLA with the above version of the EM. We used 120 views with 231 rays per view. The number of blocks of projections (subsets) was 16 for T-RAMLA. The sequence of relaxation parameters of our choice was defined by $\lambda_0 = 1.5$ and $\lambda_k = 1.5/k^{1/4}$. This sequence was found

after some experimentation; it satisfies the convergence conditions and it tends to zero very slowly, but allows relatively large steps at the beginning. The mean number of photons was 10000000, but only 10 % went detected. Figure 1 shows the behavior of the likelihood function for a typical case

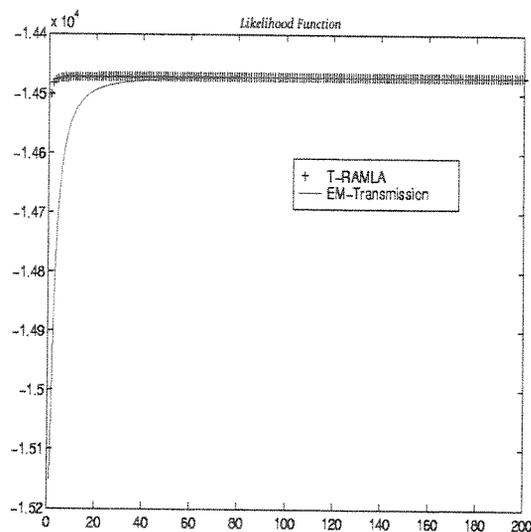


Figure 1: Likelihood Function for T-RAMLA and T-EM.

T-RAMLA attains very high likelihood levels in the first 5 iterations, but the EM needs at least 60 iterations for the same likelihood values. Figure 2 shows the phantom (used for all the experiments) and the images corresponding to 10 iterations of T-RAMLA, and 10 and 80 iterations of the EM. The EM attains the same quality of 10 iterations of T-RAMLA only after 80 iterations.

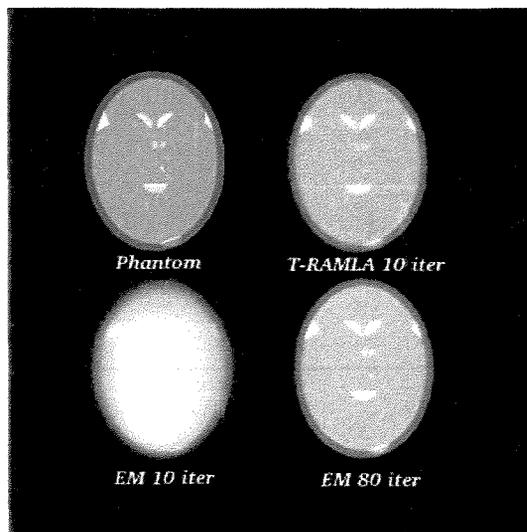


Figure 2: Clockwise from upper left: phantom, T-RAMLA 10 iterations, EM 10 and 80 iterations.

T-RAMLA versus OS-EM

We also accelerated the EM using ordered subsets as in [12] and the results were similar to those obtained with T-BSREM for the same noise level as before. However, when the photon statistics was drastically decreased (mean 10000), as well as the number of views (60), T-RAMLA attains higher likelihood values. This is shown in Figure 3. A similar behavior was observed for RAMLA, when compared with OS-EM in emission tomography.

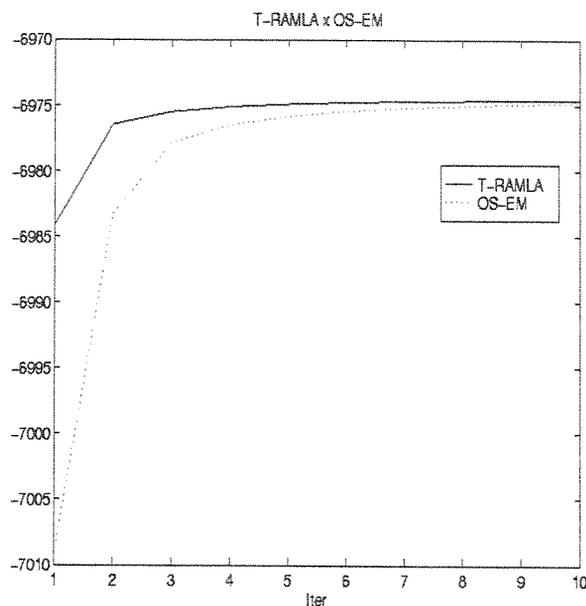


Figure 3: Likelihood Function for T-RAMLA and T-EM (very low statistics).

By further reducing the number of views, the images obtained by the EM fastly deteriorate, but the ones obtained by T-RAMLA are still acceptable. This is crucial in PET or SPECT transmission studies, because it can decrease considerably the exposure for the transmission scan. Figure 4 shows the result of reducing the number of views to 24, that is, one view each 15 degrees. The EM generates an image with much more artifacts than the one generated by T-RAMLA.

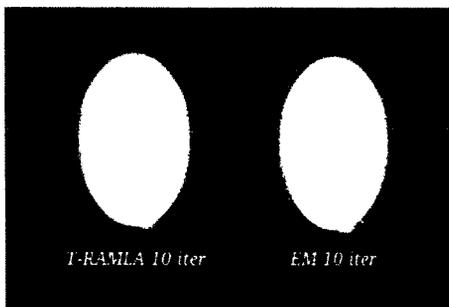


Figure 4: From left to right: T-RAMLA 10 iterations, EM 10 iterations (very low statistics and few views).

When we take one view each 30 degrees, after 10 iterations, T-RAMLA still shows an image in spite of the artifacts, while the EM does not generate any image at all, and negative values show up. The same behavior is observed when using OS-EM instead of the EM. Figure 5 shows an image obtained after 10 iterations of T-RAMLA and the other one after the same number of iterations of T-BSREM ($\gamma = 0.1$).

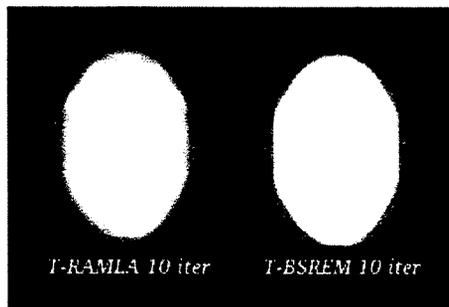


Figure 5: From left to right: T-RAMLA 10 iterations, T-BSREM 10 iterations (very low statistics and few views).

T-BSREM versus OS-GA

BSREM [5] in ECT is a natural extension of RAMLA [1] for MAP solutions;

in [8] it was presented an extension of OS-EM to the MAP model in ECT, called OS-GP. In this section, we introduce a similar version for CT derived from the scaled Gradient algorithm of Lange and Fessler [11]. The resulting iteration is

$$x_j^{k+1} = x_j^k + \frac{x_j^k}{\sum_i b_i a_{ij} + x_j^k \frac{\partial^2 R(x^k)}{\partial x_j^2}} \frac{\partial \Delta(x^k)}{\partial x_j}, \quad (18)$$

where $\Delta(x) = L(x) + \gamma R(x)$. In our experiments $-\ln(\cosh(x))$ was the potential function for R . We developed an ordered subsets version of the scaled gradient algorithm above, that we called OS-GA, easily obtained from (18).

OS-GA generates high quality images with low noise levels; however, for very high noise levels, its performance is worse than T-BSREM, and, in some cases does not generate any image. Results of a typical experiment with very low statistics are shown in Figure 6. For this simulation, 16 blocks of projections were used, 45 views with 231 rays each and the photon count was 10000. The relaxation sequence was $\lambda_k = \frac{\lambda_0}{k}$ and $\gamma = 0.1$ (experiments were performed using several values of γ , 0.1 gave the best results for the algorithms). Figure 6 shows the behavior of the MAP function for both algorithms.

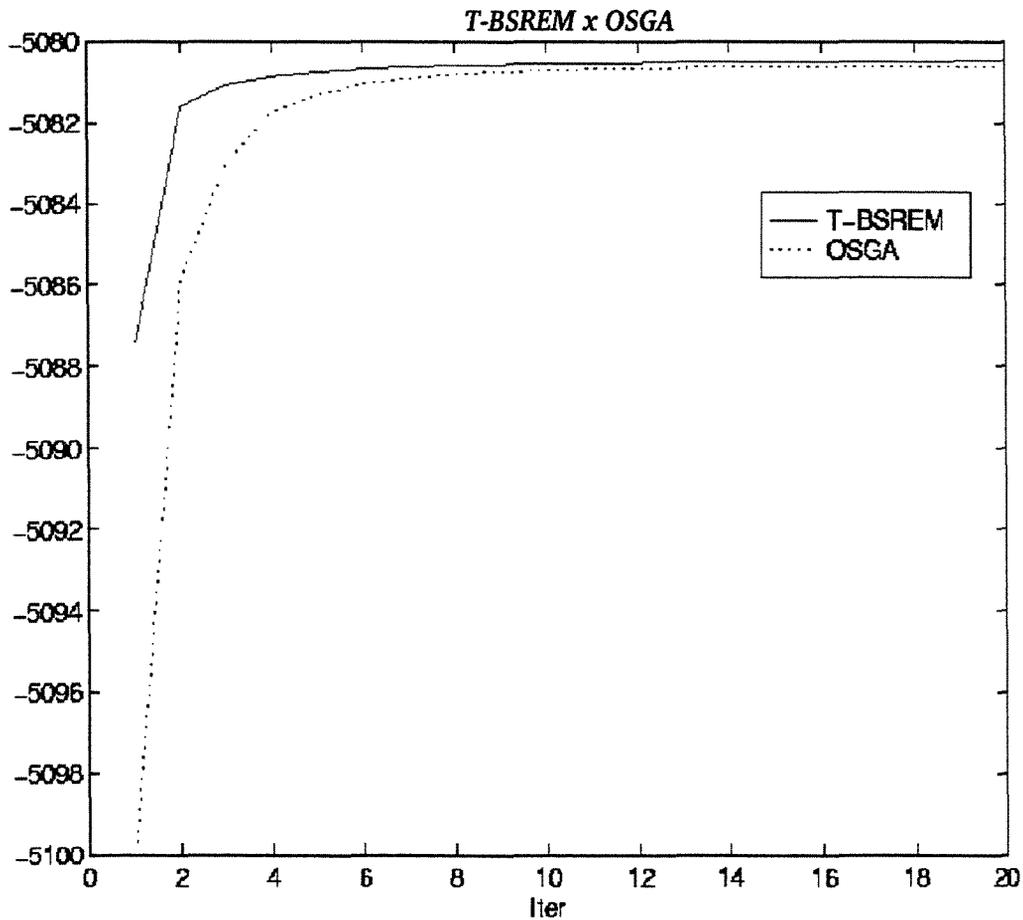


Figure 6: MAP Function for T-BSREM and OS-GA (very low statistics).

T-BSREM attains higher values of the likelihood, nevertheless OS-GA provides better images in the first few iterations (Figure 7).

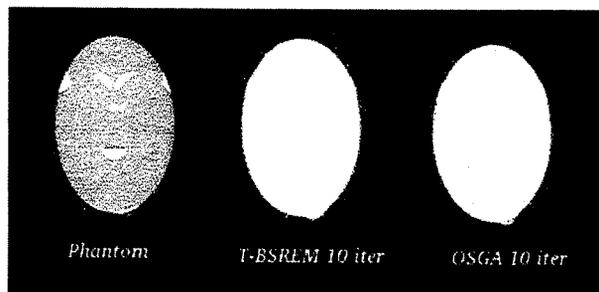


Figure 7: From left to right: Phantom, T-BSREM 10 iterations, OS-GA 10 iterations (very low statistics and few views).

6. Concluding Remarks

In this article we have introduced new scaled iterative decomposition methods for ML reconstruction in CT. The new methods are faster than the EM algorithm for CT, they have better convergence properties (increasing the likelihood, preserving positivity, ensuring boundedness in the case of T-RAMLA) than its ordered subsets versions and they produce less artifacts when applied to problems with small number of views.

We are now working on further experiments to optimize the choice of the relaxation parameters and a comparison with other methods. From the theoretical point of view we are trying to improve the convergence results in [1], that are applicable to T-RAMLA and T-BSREM and analyzing the convergence of OS-GA, that cannot be included in the general family defined by (9), but produces very good experimental results.

Acknowledgments

Work by M. E. Belezá Yamagishi was supported by FAPESP grant No 96/09186-2 and work by A. R. De Pierro by CNPq grants No 301699/81 and 201487/92-6, Brasil, and FAPESP grant No 98/00618-2.

Appendix I

Positivity

In order to satisfy the positivity condition, given an initial positive starting point we take λ_k satisfying

$$0 < \lambda_k c_{ij} \leq 1 \text{ (se } c_{ij} \neq 0), \quad (\text{A.1})$$

where

$$c_{ij} = \sum_{l \in S_i} a_{lj} y_l.$$

With the above condition we can rewrite 5 as

$$x_j^{(k,i+1)} = x_j^{(k,i)} (1 - \lambda_k c_{ij}) + x_j^{(k,i)} \lambda_k \sum_{l \in S_i} a_{lj} d_l \exp(-\langle a^l, x^{(k,i)} \rangle). \quad (\text{A.2})$$

The first term on the right side of the above equation is positive because of A.1, and the second is clearly positive. Taking into account that λ_k tends to zero, condition A.1 will be easily satisfied and the algorithm will be a well defined interior point method. This is one of the advantages that the T-RAMLA has over other ones, where positivity should be checked as in [13] or even the EM for transmission. \diamond

Boundedness

Our proof of boundedness follows closely that in [1]. Majorizing the second

term of the left hand side of A.2, we have

$$x_j^{(k,i)} \sum_{l \in S_i} a_{lj} d_l \exp(-\langle a^l, x^{(k,i)} \rangle) \leq \sum_{j=1}^n x_j^{(k,i)} \sum_{l \in S_i} a_{lj} d_l \exp(-\langle a^l, x^{(k,i)} \rangle)$$

that is

$$x_j^{(k,i)} \sum_{l \in S_i} a_{lj} d_l \exp(-\langle a^l, x^{(k,i)} \rangle) \leq \sum_{l \in S_i} d_l \langle a^l, x^{(k,i)} \rangle \exp(-\langle a^l, x^{(k,i)} \rangle).$$

A simple analysis of the function $f(x) = x \exp(-x)$ shows that the maximum is attained when $x = 1$. So, we have that

$$x_j^{(k,i)} \sum_{l \in S_i} a_{lj} d_l \exp(-\langle a^l, x^{(k,i)} \rangle) < \sum_{l \in S_i} d_l = q_i. \quad (\text{A.3})$$

Let

$$q = \max_i q_i,$$

for a fixed λ we majorize A.2 as follows

$$x_j^{(k,i+1)} \leq x_j^{(k,i)} \delta + \eta, \quad (\text{A.4})$$

where

$$\delta = \max_{c_{ij} \neq 0} (1 - \lambda c_{ij})$$

and

$$\eta = \lambda q.$$

Now, in order to simplify the notation we substitute the double index (k, i) by the single index k . Using A.4 iteratively we obtain

$$x_j^{k+1} \leq x_j^k \delta + \eta \leq x_j^0 \delta^k + \eta \sum_{l=0}^k \delta^l$$

where

$$g(\lambda) = \eta \frac{1}{1 - \delta},$$

taking into account the positivity of both terms. $g(\lambda)$ tends to zero when λ tends to zero. Therefore, if $c = \max \frac{1}{c_{ij}}$, $\lambda \in [0, c)$, the function $g(\lambda)$ is bounded. And for every (k, λ) , $x_j^k(\lambda)$ is bounded and the result follows. \diamond

Continuity

Now, assuming the boundedness (proven above) of the sequence, let us consider the set (closed and bounded of course) $C =$ closure of the convex hull of the sequence $x^{(k,i)}$.

$$\begin{aligned} x_j \frac{\partial L_i(x)}{\partial x_j} &= x_j \sum_{l \in S_i} a_{lj} d_l \exp(-\langle a^l, x \rangle) - a_{lj} y_l \leq \sum_{l \in S_i} d_l \langle a^l, x \rangle \exp(-\langle a^l, x \rangle) - a_{lj} x_j y_l \\ &\leq \sum_{l \in S_i} d_l - a_{lj} x_j y_l, \end{aligned} \tag{A.5}$$

From A.5 and the fact that $x \in C$, that is bounded, we conclude that $D\nabla L_i(x)$ and $D\nabla L_i(x) C$. The continuity in C follows from the continuity in \mathbb{R}^n . \diamond

Appendix I

Two simple examples showing that the EM algorithm proposed by Lange and Carson [10] as well as its ordered subsets version [12] do not converge to a ML solution. Consider the linear system

$$\begin{cases} \frac{1}{2}x_1 + x_2 = \frac{1}{2} \\ \frac{1}{2}x_1 + \frac{1}{4}x_2 = \frac{1}{4} \end{cases}.$$

The likelihood function is

$$L(x_1, x_2) = -\exp\left(-\frac{1}{2}x_1 - x_2\right) - \exp\left(-\frac{1}{2}x_1 - \frac{1}{4}x_2\right) - \frac{3x_1}{8} - \frac{9x_2}{9},$$

that has its maximum at $\left(\frac{1}{4}, \frac{3}{8}\right)$.

Applying the EM (Lange and Carson's version), the limit point is (15.9562, 0), that is not a maximum.

Let us consider now the following inconsistent linear system

$$\begin{cases} \frac{1}{2}x_1 + x_2 = \frac{1}{2} \\ \frac{1}{2}x_1 + x_2 = \frac{1}{4} \end{cases}. \quad (\text{A.1})$$

In this case the likelihood function is

$$L(x_1, x_2) = -2 \exp\left(-\frac{1}{2}x_1 - x_2\right) - \frac{3}{4}\left(\frac{1}{2}x_1 + x_2\right). \quad (\text{A.2})$$

Making the gradient equal to zero we find a ML solution x^* satisfying the equation

$$\frac{1}{2}x_1 + x_2 = -\ln\left(\frac{3}{8}\right), \quad (\text{A.3})$$

where $-\ln\left(\frac{3}{8}\right) \cong 0.9808$. Using as the starting point $\left(\frac{1}{4}, \frac{3}{8}\right)$ for the EM (LC), the limit is $(-4.37571, 2.46522)$, that has a negative coordinate and is not a solution. The ordered subset version, from the same initial point oscillates between $(-2.1571, 1.8186)$ and $(-2.36532, 1.6625)$. On the other hand T-RAMLA, from the same starting point and using the relaxation $\lambda_k = 1/k$ gets to the point $(0.365515, 0.778139)$ after 80 iterations, satisfying the equation A.3 up to 4 digits.

References

- [1] Browne J A and De Pierro A R 1996 A row-action alternative to EM algorithm for maximizing likelihoods in emission tomography *IEEE Trans. Med. Imaging* **15** 687-699.

- [2] Browne J A and Holmes T J 1992 Developments with maximum likelihood x-ray computed tomography *IEEE Trans. Med. Imaging* **11** 40-52.
- [3] Browne J A Herman G T and Odhner D 1993 SNARK93 - A programming system for image reconstruction from projections *Technical Report No. MIPG198, Department of Radiology, University of Pennsylvania*.
- [4] De Pierro A R 1993 On the relation between the ISRA and the EM algorithm for positron emission tomography *IEEE Trans. Med. Imaging* **12** 328-333.
- [5] De Pierro A R and Yamagishi M E B 2001 Fast EM-like Methods for Maximum "A Posteriori" Estimates in Emission Tomography *IEEE Trans. Med. Imaging* **20** 280-287.
- [6] De Pierro A R and Yamagishi M E B 1999 Fast Iterative Methods Applied to Tomography Models with General Gibbs Priors *Conference on Mathematical Modeling, Bayesian Estimation and Inverse Problems, Proc. of SPIE* **3816**134-138.
- [7] Erdogan H and Fessler J A 1999 Ordered subsets algorithms for transmission tomography *Phys. Med. Biol.* **44** 2835-2851.
- [8] Hudson H M and Larkin R S 1994 Accelerated image reconstruction using ordered subsets of projection data *IEEE Trans. Med. Imaging* **13** 601-609.
- [9] Lange K Bahn M and Little R 1987 A theoretical study of some maximum likelihood algorithms for emission and transmission tomography *IEEE Trans. Med. Imaging* **6** 106-114.
- [10] Lange K and Carson R 1984 EM reconstruction algorithm for emission and transmission tomography *J. Comput. Assist. Tomogr.* **8** 306-316.
- [11] Lange K. and Fessler J A 1995 Globally Convergent Algorithms for Maximum a Posteriori Transmission Tomography *IEEE Trans. Med. Imaging* **4** 1430-1438. 1995.
- [12] Manglos S H Gagne G M Krol A Thomas F D and Narayanaswamy R 1995 Transmission maximum-likelihood reconstruction with ordered subsets for cone beam CT *Phys. Med. Biol.* **40**1 225-1241.
- [13] Nuyts J De Man B Dupont P Defrise M Suetens P and Mortelmans L 1998 Iterative reconstruction for helical CT: a simulation study *Phys. Med. Biol.* **43** 729-737.
- [14] Ollinger J 1994 Maximum-likelihood reconstruction of transmission images in emission computed tomography via the EM algorithm *IEEE Trans. Med. Imaging* **13**, pp. 89-101

Referências Bibliográficas

- [1] H. H. Barrett and W. Swindell. *Radiological Imaging: the theory of image formation, detection, and processing*, volume 1. Academic Press, 1981.
- [2] E. T. Bell. *Men of Mathematics - The Lives and Achievements of the Great Mathematicians from Zeno to Poincaré*. Touchstone, 1937.
- [3] C. A. Bouman and K. Sauer. A unified approach to statistical tomography using coordinate descent optimization. *IEEE Trans. Image Proc.*, 5(3):480–492, 1996.
- [4] J. A. Browne, G. T. Herman, and D. Odhner. Snark93 - a programming system for image reconstruction from projections. *Technical Report*, (MIPG198, Department of Radiology, University of Pennsylvania), 1993.
- [5] J. A. Browne and A. R. De Pierro. A row-action alternative to em algorithm for maximizing likelihoods in emission tomography. *IEEE Trans. Med. Imaging*, 15(4):687–699, Oct 1996.
- [6] Y. Censor. Row-action methods for huge and sparse system and their application. *SIAM Rev.*, 23:444–466, 1981.
- [7] Y. Censor. Finite series-expansion reconstruction methods. *Proceedings of the IEEE*, 71(3):409–419, 1983.
- [8] Y. Censor. On variable block algebraic reconstruction techniques. In A. K. Louis G. T. Herman and F. Natterer, editors, *Mathematical Methods in Tomography*. Springer-Verlag, 1990.
- [9] Y. Censor, P. P. B. Eggermont, and D. Gordon. Strong underrelaxation in kaczmaz’s method for inconsistent systems. *Numer. Math.*, 41:83–92, 1983.
- [10] G. J. Chaitin. Randomness and mathematical proof. *Scientific American*, 232 (5):47–52, May 1975.
- [11] G. J. Chaitin. *The Unknowable*. Springer-Verlag, 1999.
- [12] A. M. Cormack. Representation of a function by its line integrals with some radiological applications. *J. Appl. Phys.*, 34:2722–2727, 1963.

- [13] A. M. Cormack. Computed tomography: some history and recent developments. *Proceedings of Symposia in Applied Mathematics*, 27:35–42, 1982.
- [14] A. M. Cormack and E. T. Quinto. The mathematics and physics of radiation dose planning using x-rays. *Contemporary Mathematics*, 113:41–55, 1990.
- [15] R. Courant and H. Robbins. *What is Mathematics?* Oxford, 1996.
- [16] I. Csiszar and G. Tusnady. Information geometry and alternating minimization procedures statistic and decisions. *Supplement Issue R. Oldenbourg Verlag*, 1:205–237, 1984.
- [17] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm (with discussion). *Journal of the Royal Statistic Society B*, 39:1–38, 1977.
- [18] J. Dieudonné. *Mathematics - The Music of Reason*. Springer, 1992.
- [19] H. Erdogan and J. A. Fessler. Monotonic algorithms for transmission tomography. *IEEE Trans. Med. Imaging*, 18(9):801–814, 1999.
- [20] H. Erdogan and J. A. Fessler. Ordered subsets algorithms for transmission tomography. *Phys. Med. Biol.*, 44:2835–2851, 1999.
- [21] J. A. Fessler and A. O. Hero. Space-alternating generalized expectation maximization algorithm. *IEEE Trans. on Signal Proc.*, 42:2664–2676, 1994.
- [22] R. Feynman, R. B. Leighton, and M. Sands. *The Feynman Lectures on Physics*, volume 1. Eddison-Wesley, 1977.
- [23] R. A. Fisher. Theory of statistical estimation. *Proceedings of Cambridge Philosophical Society*, 22:700–725, 1925.
- [24] R. A. Fisher. Two new properties of maximum likelihood. *Proceedings of the Royal Society of London A*, 144:285–307, 1934.
- [25] K. Gödel. *On Formally Undecidable Propositions of Principia Mathematica and Related Systems*. Dover, 1992.
- [26] S. Geman and D. McClure. Bayesian image analysis: An application to single photon emission tomography. *Proc. Statist. Comput. Sect., Amer. Statist. Assoc., Washington D. C.*, pages 12–18, 1985.
- [27] S. Geman and D. McClure. Statistical methods for tomographic image reconstruction. *ISI Tokyo session, Bull. Int. Statist. Inst.*, LII-4:5–21, 1987.
- [28] G. H. Golub and C. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, 2 edition, 1993.

- [29] J. Gomes and L. Velho. *Computação Gráfica: Imagem*. IMPA, 1994. Série de Computação e Matemática.
- [30] R. Gordon, R. Bender, and G. T. Herman. Algebraic reconstruction techniques (art) for three-dimensional electron microscopy and x-ray photography. *J. Theor. Biol.*, 29:471–481, 1970.
- [31] P. J. Green. Bayesian reconstruction from emission tomography data using a modified em algorithm. *IEEE Trans. Med. Imaging*, 9(1):84–93, 1990.
- [32] P. J. Green. Bayesian reconstructions from emission tomography data using a modified em algorithm. *IEEE Trans. Med. Imaging*, 9(1):84–93, 1990.
- [33] T. Hebert and R. Leary. A generalized em algorithm for 3-d bayesian reconstruction from poisson data using gibbs priors. *IEEE Trans. Med. Imaging*, 8:194–202, 1989.
- [34] G. T. Herman. *Image Reconstruction from Projections: The fundamentals of Computerized Tomography*. Academic Press, 1980.
- [35] G. T. Herman and L. B. Meyer. Algebraic reconstruction techniques can be made computationally efficient. *IEEE Trans. Med. Imaging*, 12(3):600–609, 1993.
- [36] G. T. Herman and A. Naparstek. Fast image reconstruction based on a radon inversion formula appropriate for rapidly collected data. *SIAM J. Appl. Math*, 33(3):511–533, Nov 1977.
- [37] R. Hersh. *What is Mathematics, Really?* Oxford, 1997.
- [38] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1994.
- [39] H. M. Hudson and R. S. Larkin. Accelerated image reconstruction using ordered subsets of projection data. *IEEE Trans. Med. Imaging*, 13(4):601–609, 1994.
- [40] A. K. Jain. *Fundamentals of Digital Image Processing*. Prentice Hall, 1989.
- [41] S. Kaczmarz. Angenäherte auflösung von systemen linearer gleichungen. *Bull. Acad. Polon. Sci. Lett., A.*, 35:355–357, 1937.
- [42] L. Kaufman. Implementing and accelerating the em algorithm for positron emission tomography. *IEEE Trans. Med. Imaging*, MI-6:37–51, 1987.
- [43] M. Kline. *Mathematical Thought from Ancient to Modern Times*, volume 1. Oxford, 1972.
- [44] M. Kline. *Mathematics: The Loss of Certainty*. Oxford, 1980.
- [45] D. S. Lalush and B. M. W. Tsui. A fast and stable maximum a posteriori conjugate gradient reconstruction algorithm. *Med. Phys.*, 22(8):1273–1284, 1995.

- [46] D. S. Lalush and B. M. W. Tsui. The importance of preconditioners in fast poisson-based iterative reconstruction algorithms for spect. *1995 IEEE Nuclera Science Symposium Conference Record*, pages 1326–1330, 1996.
- [47] K. Lange. Convergence of em image reconstruction algorithms with gibbs smoothing. *IEEE Trans. Med. Imaging*, 9(4):439–446, 1990.
- [48] K. Lange, M. Bahn, and R. Little. A theoretical study of some maximum likelihood algorithms for emission and transmission tomography. *IEEE Trans. Med. Imaging*, MI-6:106–114, 1987.
- [49] K. Lange and R. Carson. Em reconstruction algorithm for emission and transmission tomography. *J. Comput. Assist. Tomogr.*, 8:306–316, 1984.
- [50] K. Lange and J. A. Fessler. Globally convergent algorithms for maximum a posteriori transmission tomography. *IEEE Trans. on Image Processing*, 4(10):1430–1438, Oct. 1995.
- [51] R. Leahy and C. Byrne. Recent development in iterative image reconstruction for pet and spect. *IEEE Trans. Med. Imaging*, 19(4):257–260, Apr 2000.
- [52] E. Levitan and G. T. Herman. A maximum a posteriori probability expectation maximization algorithm for image reconstruction in emission tomography. *IEEE Trans. Med. Imaging*, 6(3):185–192, 1987.
- [53] R. M. Lewitt. Reconstruction algorithms: Transform methods. *Proceedings of the IEEE*, 71(3):390–408, 1983.
- [54] R. M. Lewitt and G. Muehlelehner. Accelerated iterative reconstruction for positron emission tomography based on the em algorithm for maximum likelihood estimation. *IEEE Trans. Med. Imaging*, 5:16–22, 1986.
- [55] L. B. Lucy. An iterative technique for rectification of observed distributions. *Astron. J.*, 79:745–754, 1974.
- [56] S. H. Manglos, G. M. Gagne, A. Krol, F. D. Thomas, and R. Narayanaswamy. Transmission maximum-likelihood reconstruction with ordered subsets for cone beam ct. *Phys. Med. Biol.*, 40:1225–1241, 1995.
- [57] G. J. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. Wiley Series in Probability and Statistics. John Wiley and Sons, 1997.
- [58] E. S. Meinel. Origins of linear and nonlinear recursive restoration algorithms. *J. Opt. Soc. Am. A.*, 3(6):787–799, 1986.
- [59] E. Ü. Mumcuoglu, R. M. Leahy, and S. R. Cherry. Bayesian reconstruction of pet images: methodology and performance analysis. *Phys. Med. Biol.*, 41:1777–1807, 1996.
- [60] E. Ü. Mumcuoglu, R. M. Leahy, S. R. Cherry, and Z. Zhou. Fast gradient-based methods for bayesian reconstruction of transmission and emission pet images. *IEEE Trans. Med. Imag.*, 13(4):687–701, 1994.

- [61] M. Nikolova, J. Idier, and A. Mohammad-Dajafari. Inversion of large-support ill-posed linear operators using a piecewise gaussian mrf. *IEEE Trans. Image Processing*, 7(4):571–585, 1998.
- [62] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer Series in Operation Research. Springer, 1999.
- [63] J. Nuyts, P. Dupont, S. Stroobants, R. Benninck, L. Mortelmans, and P. Suetens. Simultaneous maximum a posteriori reconstruction of attenuation and activity distributions from emission sinograms. *IEEE Trans. on Medical Imaging*, 18(5):393–403, May 1999.
- [64] J. Nuyts, B. De Man, P. Dupont, M. Defrise, P. Suetens, and L. Mortelmans. Iterative reconstruction for helical ct: a simulation study. *Phys. Med. Biol.*, 43:729–737, 1998.
- [65] J. Ollinger. Maximum-likelihood reconstruction of transmission images in emission computed tomography via the em algorithm. *IEEE Trans. Med. Imaging*, 13:89–101, 1994.
- [66] A. R. De Pierro. A generalization of the em algorithm for maximum likelihood estimates from incomplete data. *Technical Report No MIPG119, Department of Radiology, University of Pennsylvania*, May 1987.
- [67] A. R. De Pierro. On some nonlinear iterative relaxation methods in remote sensing. *Matemática Aplicada e Computacional*, 8(1):1–14, 1989.
- [68] A. R. De Pierro. On the relation between the isra and the em algorithm for positron emission tomography. *IEEE Trans. Med. Imaging*, 12:328–333, 1993.
- [69] A. R. De Pierro. A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography. *IEEE Trans. Med. Imaging*, 14(1):132–137, Mar 1995.
- [70] A. R. De Pierro. Convergence behavior of some fast em-like methods for ml and map estimates in emission tomography. *Preprint*, 2001.
- [71] A. R. De Pierro and M. E. B. Yamagishi. Fast iterative methods applied to tomography models with general gibbs priors. In *Conference on Mathematical Modeling, Bayesian Estimation and Inverse Problems, Proceedings of SPIE, Vol. 3816*, pp. 134–138, 1999.
- [72] A. R. De Pierro and M. E. B. Yamagishi. Fast em-like methods for maximum “a posteriori” estimates in emission tomography. *IEEE Trans. Med. Imaging*, 20(4):280–287, Apr 2001.
- [73] H. Poincaré. *O Valor da Ciência*. Contraponto, 1995.
- [74] J. Radon. Über die bestimmung von funktionen durch ihre integralwerte längs gewisser mannigfaltigkeiten. *Ber. Ver. Sächs. Akad. Wiss. Leipzig Math-Phys. Kl.*, 69:262–277, Apr 1917.
- [75] N. Rajeevan, K. Rajgopal, and G. Khishna. Vector-extrapolated fast maximum likelihood estimation algorithms for emission tomography. *IEEE Trans. Med. Imaging*, 11(3):9–20, 1992.

- [76] W. H. Richardson. Bayesian-based iterative method of image restoration. *J. Opt. Soc. Am.*, 62:55–59, 1972.
- [77] A. J. Rockmore and A. Macovski. A maximum likelihood approach to emission image reconstruction from projections. *IEEE Trans. Nucl. Sci.*, 23:1428–1432, 1976.
- [78] L. A. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Trans. Med. Imaging*, 1:113–121, 1982.
- [79] L. A. Shepp, Y. Vardi, J. B. Ra, S. K. Hilal, and Z. H. Cho. Maximum likelihood pet with real data. *IEEE Trans. Nucl. Sci.*, 31:910–913, 1984.
- [80] H. Stark. *Image Recovery: Theory and Application*. Academic Press, 1987.
- [81] E. Tanaka. A fast reconstruction algorithm for stationary positron emission tomography based on a modified em algorithm. *IEEE Trans. Med. Imaging*, MI-6(2):98–105, Jun 1987.
- [82] M. M. Ter-Pogossian, M. Raichle, and B. E. Sobel. Positron emission tomography. *Sci. Amer.*, 243(4):170–181, 1980.
- [83] S. Twomey. Comparison of constrained linear inversion and an iterative nonlinear algorithm applied to the indirect estimation of particle size distributions. *J. Comput. Phys.*, 18:188–200, 1975.
- [84] R. E. Woods and R. C. Gonzales. *Digital Image Processing*. Addison Wesley, 1992.