UNIVERSITÉ DE SHERBROOKE

TOPICS IN IMAGE RECONSTRUCTION FOR HIGH RESOLUTION POSITRON EMISSION TOMOGRAPHY

par

Vitali Selivanov

Département de Médecine nucléaire et de radiobiologie

Thèse présentée à la Faculté de médecine
en vue de l'obtention du grade de
philosophiae doctor (Ph.D.) en radiobiologie

Sherbrooke (Québec) Canada

2002

National Library of Canada
Acquisitions and Bibliographic Services
395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale du Canada
Acquisitions et services bibliographiques
395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-80550-6

Canada

Contents

ABBREVIATIONS ..... **iv**

LIST OF FIGURES ..... **vi**

LIST OF TABLES ..... **xi**

ABSTRACT ..... **xii**

RÉSUMÉ ..... **xiv**

INTRODUCTION **1**

CHAPTER 1. BACKGROUND **3**

1.1. PRINCIPLES OF PET 3

1.2. PET INVERSE PROBLEM 6

1.2.1. Radon transform model *7*

1.2.2. Functional semidiscrete model *8*

1.2.3. Stochastic extension of the functional model *9*

1.2.4. Stochastic model *10*

1.2.5. Poisson vs. Gaussian data *11*

1.2.6. Deterministic vs. stochastic modeling *12*

1.3. DISCRETE IMAGE AND QUADRATURE OPTIONS 13

1.4. IMAGE PROJECTION 18

1.5. BACKPROJECTION 20

1.6. SYSTEM MATRIX PROPERTIES AND INVERSE PROBLEM ANALYSIS 21

1.7. REGULARISATION 24

CHAPTER 2. IMAGE RECONSTRUCTION TECHNIQUES **26**

2.1. General Classification 26

2.2. Summation Method 28

2.3. Transform Methods 29

2.4. Series Expansion Methods 32

2.4.1. Orthogonal series 32

2.4.2. Direct matrix method 33

2.4.3. Algebraic methods 35

2.4.4. Statistical methods 36

2.4.5. Iterative reconstruction acceleration 42

2.5. 3-D image reconstruction issues 43

CHAPTER 3. RESEARCH QUESTIONS **46**

3.1. System matrix 46

3.2. Stopping rule 49

3.3. SVD potential 50

3.4. Accurate image reconstruction in real time 51

CHAPTER 4. RESULTS **52**

4.1. Selivanov V V, Picard Y, Cadorette J, Rodrigue S, and Lecomte R 2000 Detector Response Models for Statistical Iterative Image Reconstruction in High Resolution PET IEEE Trans. Nucl. Sci. ***47*** *1168-75* 53

4.2. Selivanov V V, Lapointe D, Bentourkia M, and Lecomte R 2001 Cross-Validation Stopping Rule for ML-EM Reconstruction of Dynamic PET Series: Effect on Image Quality and Quantitative Accuracy IEEE Trans. Nucl. Sci. ***48*** *883-9* 62

4.3. Selivanov V V and Lecomte R 2001 Fast PET Image Reconstruction Based on SVD Decomposition of the System Matrix IEEE Trans. Nucl. Sci. ***48*** *761-7* 70

4.4. Selivanov V V and Lecomte R List-Mode PET Image Reconstruction Based on Regularized Pseudo-Inverse of the System Matrix to be submitted 78

CHAPTER 5. DISCUSSION **111**

5.1. DETECTOR RESPONSE AS A SYSTEM MODEL 111

5.2. TO STOP OR NOT TO STOP... 113

5.3. REGULARISATION IS THE KEY 115

5.4. HOW PROHIBITIVE IS THE SVD COMPUTATION? 116

5.5. REAL TIME PET IMAGING PROSPECTS 118

CONCLUSION **121**

ACKNOWLEDGEMENTS **123**

APPENDIX A **125**

APPENDIX B **127**

REFERENCES **142**

INDEX **157**

Abbreviations

1-D	one-dimensional
2-D	two-dimensional
3-D	three-dimensional
ART	algebraic reconstruction technique
CLT	central limit theorem
CT	computed (or computerised) tomography
CV	cross-validation
DFM	direct Fourier method
DRF	detector response function
ECT	emission computed tomography
EM	expectation maximisation
FBP	filtered backprojection
FOV	field-of-view
KL	Karhunen-Loève (transform)
LOR	line-of-response
MAP	maximum *a posteriori*
ML	maximum likelihood
MLE	maximum likelihood estimate
ML-EM	maximum likelihood via expectation maximisation (algorithm)
OS-EM	(maximum likelihood via) expectation maximisation with ordered subsets (algorithm)
PET	positron emission tomography
PSF	point spread function
PWLS	penalised weighted least-squares
RAMLA	row-action maximum likelihood algorithm
ROI	region-of-interest
r.v.	random variable

SNR	signal-to-noise ratio
SPECT	single photon emission computed tomography
SVD	singular value decomposition
TAC	time-activity curve
TOR	tube-of-response
TSVD	truncated singular value decomposition
WLS	weighted least-squares

List of Figures

FIGURE 1. SIMPLE CASE OF RADIOTRACER DISTRIBUTION AND TWO IDEALISED CONTINUOUS PROJECTIONS AT SELECTED ANGLES. 6

FIGURE 2. FROM LEFT TO RIGHT: A) AN EXAMPLE OF THE SUPPORT REGION Ω, WHICH IS THE ELLIPSE INTERIOR; B) A SET OF CONVENTIONAL SQUARE PIXELS; C) A SET OF POLAR PIXELS. 15

FIGURE 3. TUBE-OF-RESPONSE FOR A GIVEN DETECTOR PAIR.. 16

FIGURE 4. NATURAL PIXEL DEFINED AS AN INTERSECTION OF THE TUBE-OF-RESPONSE SHOWN IN THE PREVIOUS FIGURE AND THE PET SCANNER FIELD-OF-VIEW. 17

FIGURE 5. THE SINOGRAM OF AN OFF-CENTRE "POINT" SOURCE. 18

FIGURE 6. THE DIAGRAM SHOWING TWO ALTERNATIVE ROUTES THAT COULD BE TAKEN FOR AN APPROXIMATE SOLUTION OF THE INVERSE PROBLEM. 27

FIGURE 7. SCHEMATICS OF THE ML-EM IMAGE RECONSTRUCTION AND AN EXAMPLE OF THE FEASIBILITY REGION.. 39

FIGURES APPEARING AS A PART OF THE FOLLOWING PUBLICATIONS (NOTE THAT THE REFERENCES TO OTHER FIGURES IN THE FOLLOWING CAPTIONS REFER TO SAME ARTICLE):

[4.1] SELIVANOV V V, PICARD Y, CADORETTE J, RODRIGUE S, AND LECOMTE R 2000 DETECTOR RESPONSE MODELS FOR STATISTICAL ITERATIVE IMAGE RECONSTRUCTION IN HIGH RESOLUTION PET *IEEE TRANS. NUCL. SCI.* **47** 1168-75

FIGURE 1. PHANTOMS USED TO ACQUIRE DATA WITH THE SHERBROOKE ANIMAL PET SCANNER.....

54

FIGURE 2. GEOMETRIC MODEL USED FOR THE CALCULATION OF THE ANALYTIC DRF. 55

FIGURE 3. TWO EXAMPLES OF SIX DIFFERENT DRF MODELS IN COMPARISON TO EXPERIMENTAL DRF..... 56

FIGURE 4. RESOLUTION PHANTOM RECONSTRUCTED USING SIX DIFFERENT DRF APPROXIMATIONS AND THE CROSS-VALIDATION PROCEDURE (FIRST RECONSTRUCTION SERIES) 58

FIGURE 5. PROFILES OF THE RESOLUTION PHANTOM (FIRST RECONSTRUCTION SERIES) 58

FIGURE 6. CONTRAST PHANTOM RECONSTRUCTED USING SIX DIFFERENT DRF APPROXIMATIONS AND THE CROSS-VALIDATION PROCEDURE (FIRST RECONSTRUCTION SERIES) 58

FIGURE 7. PROFILES OF THE CONTRAST PHANTOM ILLUSTRATING THE EDGE ARTEFACT (FIRST RECONSTRUCTION SERIES) 58

FIGURE 8. RELATIVE RECOVERY FACTOR AS A FUNCTION OF HOT SPOT SIZE, CALCULATED FROM THE CONTRAST PHANTOM IMAGES RECONSTRUCTED USING SIX DIFFERENT DRF MODELS (FIRST RECONSTRUCTION SERIES). 59

FIGURE 9. RELATIVE STANDARD DEVIATION AS A FUNCTION OF HOT SPOT SIZE CALCULATED FROM THE CONTRAST PHANTOM IMAGES RECONSTRUCTED USING SIX DIFFERENT DRF MODELS (FIRST RECONSTRUCTION SERIES). 59

FIGURE 10. RELATIVE STANDARD DEVIATION WITHIN THE LARGEST HOT REGION OF THE CONTRAST PHANTOM AS A FUNCTION OF ITERATION NUMBER (SECOND RECONSTRUCTION SERIES). 59

FIGURE 11. PROFILES OF THE CONTRAST PHANTOM ILLUSTRATING THE EDGE ARTEFACT (SECOND RECONSTRUCTION SERIES).....59

FIGURE 12. RELATIVE RECOVERY FACTOR AS A FUNCTION OF HOT SPOT SIZE CALCULATED FROM THE CONTRAST PHANTOM IMAGES RECONSTRUCTED USING SIX DIFFERENT DRF MODELS (SECOND RECONSTRUCTION SERIES). 60

FIGURE 13. RELATIVE STANDARD DEVIATION AS A FUNCTION OF HOT SPOT SIZE CALCULATED FROM THE CONTRAST PHANTOM IMAGES RECONSTRUCTED USING SIX DIFFERENT DRF MODELS (SECOND RECONSTRUCTION SERIES). 60

[4.2] SELIVANOV V V, LAPOINTE D, BENTOURKIA M, AND LECOMTE R 2001 CROSS-VALIDATION STOPPING RULE FOR ML-EM RECONSTRUCTION OF DYNAMIC PET SERIES: EFFECT ON IMAGE QUALITY AND QUANTITATIVE ACCURACY *IEEE TRANS. NUCL. SCI.* **48** 883-9

FIGURE 1. PHANTOM MODELING THE LEFT VENTRICLE OF A RAT HEART WAS USED TO ACQUIRE DATA FOR ASSESSING QUANTITATIVE ACCURACY OF DIFFERENT RECONSTRUCTION PROTOCOLS. 64

FIGURE 2. NUMBER OF ML-EM ITERATIONS, DEFINED BY THE AUTOMATIC CV PROCEDURE, AS A FUNCTION OF THE TOTAL COUNTS IN PHANTOM PROJECTIONS. 65

FIGURE 3. DECAY CORRECTED COUNT RATE c_t IN THE PHANTOM ROI AS A FUNCTION OF THE TOTAL COUNTS IN THE PROJECTION DATA. 66

FIGURE 4. RATIO OF c_t FOR ML-EM WITH THE CROSS-VALIDATION TO c_t FOR 200 ML-EM ITERATIONS. 66

FIGURE 5. PHANTOM IMAGES CONTRAST AS A FUNCTION OF THE TOTAL COUNTS IN THE PROJECTION DATA. 66

FIGURE 6. DEPENDENCE OF THE NUMBER OF ML-EM ITERATIONS, DEFINED BY THE AUTOMATIC CV PROCEDURE, ON THE TOTAL COUNTS FOR THE RAT SERIES (CV STOPPING POINT), FUNCTION FITTED TO THE CV STOPPING POINTS (CVF), AND THE FITTED FUNCTION DERIVED FOR THE PHANTOM (CVF FOR PHANTOM, GIVEN IN FIG.2 AS WELL). 66

FIGURE 7. SELECTED FRAMES FROM THE DYNAMIC RAT SCAN, RECONSTRUCTED WITH DIFFERENT TECHNIQUES. 67

FIGURE 8. EARLY PART OF THE BLOOD POOL TIME-ACTIVITY CURVES. 67

FIGURE 9. LATE PART OF THE BLOOD POOL TIME-ACTIVITY CURVES. 68

FIGURE 10. RMRGLC ESTIMATES COMPUTED FOR MYOCARDIAL ROIS USING TACS OF THE BLOOD POOL AS INPUT FUNCTIONS. 68

[4.3] SELIVANOV V V AND LECOMTE R 2001 FAST PET IMAGE RECONSTRUCTION BASED ON SVD DECOMPOSITION OF THE SYSTEM MATRIX *IEEE TRANS. NUCL. SCI.* **48** 761-7

FIGURE 1. SINGULAR VALUE SPECTRA OF THE SYSTEM MATRIX FOR THE SHERBROOKE ANIMAL PET SCANNER AND THREE DIFFERENT IMAGE GRIDS. 73

FIGURE 2. SAME AS FIG. 1 ON A LOG-LOG SCALE SHOWING THE EXISTENCE OF A "PLATEAU" OF SINGULAR VALUES AND AN ABRUPT DROP-OFF OF THE SPECTRUM TAIL RESPONSIBLE FOR THE SEVERE ILL CONDITIONING OF THE RECONSTRUCTION INVERSE PROBLEM. 73

FIGURE 3. RADIAL FWHM LOCAL RESOLUTION ESTIMATES OF A POINT SOURCE IN THE FOV CENTER, RECONSTRUCTED WITH TSVD WITH VARYING TRUNCATION INDEX T, 96×96-PIXEL IMAGE.....74

FIGURE 4. POINT SOURCE IN THE FOV CENTER RECONSTRUCTED WITH TSVD, 96×96 PIXEL IMAGE...74

FIGURE 5. FWHM RESOLUTION ESTIMATES (LOCAL AND GLOBAL), DERIVED USING POINT SOURCE IMAGES RECONSTRUCTED WITH TSVD, T=1217, 96×96 PIXEL IMAGE. 75

FIGURE 6. PHANTOM IMAGES RECONSTRUCTED WITH: A) FBP; B) ML-EM (100 ITERATIONS); C) TSVD, MATCHING FBP (GLOBAL) RESOLUTION (T=950); D) TSVD, MATCHING GLOBAL RECONSTRUCTED IMAGE RESOLUTION TO THE INTRINSIC SCANNER RESOLUTION IN THE FOV CENTER (T=1217). 75

FIGURE 7. IMAGE PROFILES THROUGH THE TWO LARGEST HOLES OF THE PHANTOM ALONG THE LINE SHOWN IN FIG. 6D. A) FBP; B) ML-EM, 100 ITERATIONS; C) TSVD, T=950; D) TSVD, T=1217. 76

FIGURE 8. RECOVERY FACTORS CALCULATED USING RECONSTRUCTED PHANTOM IMAGES SHOWN IN FIG. 6. 76

[4.4] SELIVANOV V V AND LECOMTE R LIST-MODE PET IMAGE RECONSTRUCTION BASED ON REGULARIZED PSEUDO-INVERSE OF THE SYSTEM MATRIX *TO BE SUBMITTED*

FIGURE 1. SCHEMATIC OF THE EVENT-BY-EVENT RECONSTRUCTION PROCESS USING THE REGULARIZED PSEUDO-INVERSE MATRIX. 105

FIGURE 2. SEQUENCE OF PHANTOM IMAGES DEMONSTRATING THE USE OF THE INCREMENTAL TSVD RECONSTRUCTION FOR A SYSTEM HAVING A COMPLETE RING OF DETECTORS. 106

FIGURE 3. SEQUENCE OF IMAGES DEMONSTRATING THE USE OF THE INCREMENTAL TSVD RECONSTRUCTION FOR ROTATING BANKS OF DETECTORS, AN INCOMPLETE RING, ACQUIRING DATA AT ONE PROJECTION ANGLE AT A TIME. 107

FIGURE 4. RADIAL PROFILES THROUGH THE EXPECTED STANDARD DEVIATION MAPS WITH THE TSVD RECONSTRUCTION FOR THREE DIFFERENT TRUNCATION LEVELS AND THE IMAGE OF 64×64 PIXELS. 108

FIGURE 5. RADIAL PROFILES THROUGH THE EXPECTED STANDARD DEVIATION MAPS WITH THE TSVD RECONSTRUCTION FOR THREE DIFFERENT TRUNCATION LEVELS AND THE IMAGE OF 96×96 PIXELS. 109

List of Tables

TABLES APPEARING AS A PART OF THE FOLLOWING PUBLICATIONS:

[4.1] SELIVANOV V V, PICARD Y, CADORETTE J, RODRIGUE S, AND LECOMTE R 2000 DETECTOR RESPONSE MODELS FOR STATISTICAL ITERATIVE IMAGE RECONSTRUCTION IN HIGH RESOLUTION PET *IEEE TRANS. NUCL. SCI.* **47** 1168-75

TABLE 1. NOTATION USED IN THE PAPER 55

TABLE 2. TRANSITION MATRIX SIZE (NON-ZERO ELEMENTS ONLY), MB 56

TABLE 3. NUMBER OF ITERATIONS SUGGESTED BY THE CROSS-VALIDATION STOPPING RULE 57

[4.2] SELIVANOV V V, LAPOINTE D, BENTOURKIA M, AND LECOMTE R 2001 CROSS-VALIDATION STOPPING RULE FOR ML-EM RECONSTRUCTION OF DYNAMIC PET SERIES: EFFECT ON IMAGE QUALITY AND QUANTITATIVE ACCURACY *IEEE TRANS. NUCL. SCI.* **48** 883-9

TABLE I. NOTATION USED 64

[4.3] SELIVANOV V V AND LECOMTE R 2001 FAST PET IMAGE RECONSTRUCTION BASED ON SVD DECOMPOSITION OF THE SYSTEM MATRIX *IEEE TRANS. NUCL. SCI.* **48** 761-7

TABLE I. TRUNCATION INDEX T BASED ON SPATIAL RESOLUTION ANALYSIS (USING GLOBAL RESOLUTION ESTIMATES) 74

[4.4] SELIVANOV V V AND LECOMTE R LIST-MODE PET IMAGE RECONSTRUCTION BASED ON REGULARIZED PSEUDO-INVERSE OF THE SYSTEM MATRIX *TO BE SUBMITTED*

TABLE I. EXPECTED SIZE OF THE (REGULARIZED) PSEUDO-INVERSE MATRIX FOR VARIOUS SINOGRAM SIZE 110

Abstract

Ill-posed problems are a topic of an interdisciplinary interest arising in remote sensing and non-invasive imaging. However, there are issues crucial for successful application of the theory to a given imaging modality. Positron emission tomography (PET) is a non-invasive imaging technique that allows assessing biochemical processes taking place in an organism *in vivo*. PET is a valuable tool in investigation of normal human or animal physiology, diagnosing and staging cancer, heart and brain disorders. PET is similar to other tomographic imaging techniques in many ways, but to reach its full potential and to extract maximum information from projection data, PET has to use accurate, yet practical, image reconstruction algorithms. Several topics related to PET image reconstruction have been explored in the present dissertation. The following contributions have been made:

- A system matrix model has been developed using an analytic detector response function based on linear attenuation of γ-rays in a detector array. It has been demonstrated that the use of an oversimplified system model for the computation of a system matrix results in image artefacts. (IEEE Trans. Nucl. Sci., 2000)
- The dependence on total counts modelled analytically was used to simplify utilisation of the cross-validation (CV) stopping rule and accelerate statistical iterative reconstruction. It can be utilised instead of the original CV procedure for high-count projection data, when the CV yields reasonably accurate images. (IEEE Trans. Nucl. Sci., 2001)

- A regularisation methodology employing singular value decomposition (SVD) of the system matrix was proposed based on the spatial resolution analysis. A characteristic property of the singular value spectrum shape was found that revealed a relationship between the optimal truncation level to be used with the truncated SVD reconstruction and the optimal reconstructed image resolution. (IEEE Trans. Nucl. Sci., 2001)

- A novel event-by-event linear image reconstruction technique based on a regularised pseudo-inverse of the system matrix was proposed. The algorithm provides a fast way to update an image potentially in real time and allows, in principle, for the instant visualisation of the radioactivity distribution while the object is still being scanned. The computed image estimate is the minimum-norm least-squares solution of the regularised inverse problem.

Résumé

Les problèmes mal posés représentent un sujet d'intérêt interdisciplinaire qui surgires dans la télédétection et des applications d'imagerie. Cependant, il subsiste des questions cruciales pour l'application réussie de la théorie à une modalité d'imagerie. La tomographie d'émission par positron (TEP) est une technique d'imagerie non-invasive qui permet d'évaluer des processus biochimiques se déroulant à l'intérieur d'organismes in vivo. La TEP est un outil avantageux pour la recherche sur la physiologie normale chez l'humain ou l'animal, pour le diagnostic et le suivi thérapeutique du cancer, et l'étude des pathologies dans le cœur et dans le cerveau. La TEP partage plusieurs similarités avec d'autres modalités d'imagerie tomographiques, mais pour exploiter pleinement sa capacité à extraire le maximum d'information à partir des projections, la TEP doit utiliser des algorithmes de reconstruction d'images à la fois sophistiquée et pratiques. Plusieurs aspects de la reconstruction d'images TEP ont été explorés dans le présent travail. Les contributions suivantes sont d'objet de ce travail:

- Un modèle viable de la matrice de transition du système a été élaboré, utilisant la fonction de réponse analytique des détecteurs basée sur l'atténuation linéaire des rayons γ dans un banc de détecteur. Nous avons aussi démontré que l'utilisation d'un modèle simplifié pour le calcul de la matrice du système conduit à des artefacts dans l'image. (IEEE Trans. Nucl. Sci., 2000)
- La modélisation analytique de la dépendance décrite à l'égard de la statistique des images a simplifié l'utilisation de la règle d'arrêt par contre-vérification (CV) et a permis d'accélérer la reconstruction statistique itérative. Cette règle

peut être utilisée au lieu du procédé CV original pour des projections aux taux de comptage élevés, lorsque la règle CV produit des images raisonnablement précises. (IEEE Trans. Nucl. Sci., 2001)

- Nous avons proposé une méthodologie de régularisation utilisant la décomposition en valeur propre (DVP) de la matrice du système basée sur l'analyse de la résolution spatiale. L'analyse des caractéristiques du spectre de valeurs propres nous a permis d'identifier la relation qui existe entre le niveau optimal de troncation du spectre pour la reconstruction DVP et la résolution optimale dans l'image reconstruite. (IEEE Trans. Nucl. Sci., 2001)

- Nous avons proposé une nouvelle technique linéaire de reconstruction d'image événement-par-événement basée sur la matrice pseudo-inverse régularisée du système. L'algorithme représente une façon rapide de mettre à jour une image, potentiellement en temps réel, et permet, en principe, la visualisation instantanée de distribution de la radioactivité durant l'acquisition des données tomographiques. L'image ainsi calculée est la solution minimisant les moindres carrés du problème inverse régularisé.

Introduction

In 1956, Allan M. Cormack, a nuclear physicist involved with radiotherapy treatment planning hypothesised that X-ray beams projected through the human body at different angles but along a single plane, would provide a better view of the body's internal structure than the procedures utilised at the time (Cormack, 1980). At that time a state-of-the-art diagnostic X-ray examination implied the transmission of X-rays through tissue resulting in a planar projection image on a film. During the next several years, he had been intermittently working on a reconstruction method that could convert the data collected at several angles into an image representing the cross-section of a body. In 1963, Cormack performed testing on a simple head phantom applying the results of his studies in reconstruction theory. Two papers on the subject were published in the Journal of Applied Physics in 1963 and 1964 (Cormack, 1963; Cormack, 1964), but received almost no attention. Ironically, it was not until 1970 that Cormack learned that the problem of determining a function from its line integrals was first solved by Johann Radon in 1917 (Cormack, 1980; Cormack, 1992).

In 1967, unaware of the work by Cormack, Godfrey N. Hounsfield also took on a goal of utilising measurements of X-ray transmission, taken from all possible directions through a body, for revealing its internal structure. After a series of computer simulations and experimental efforts with an improvised scanner, encouraging results had been obtained (Hounsfield, 1976; Hounsfield, 1980). Unlike Cormack, Hounsfield was successful in generating interest in his research. In 1971, the first clinical prototype brain

scanner was installed at the Atkinson Morley's Hospital, Wimbledon, England (Hounsfield, 1973).

For their pioneering efforts, Hounsfield and Cormack were awarded the 1979 Nobel Prize in Medicine for the "development of computer assisted tomography". Several novel tomographic imaging modalities have emerged as a result of intensive research during the past four decades. Positron emission tomography (PET), which has set a new standard in functional medical imaging, was one of them.

The present work makes a contribution on several topics of PET image reconstruction. Accommodation of an adequate PET system response model and the issue of a stopping rule for iterative reconstruction termination are considered. A reconstruction approach using the theory of pseudo-inverse matrices, which was considered at the early stages of theoretical developments but was left out of the mainstream for more than two decades, is shown to be feasible today. A regularisation technique based on the systematic spatial resolution analysis and singular value spectrum truncation is developed. Finally, a list-mode image reconstruction algorithm based on the latter approach is proposed.

The manuscript is organised as follows. A brief introduction into PET principles and the corresponding inverse problem are given in Chapter 1. Existing image reconstruction approaches are put into perspective in Chapter 2. Limitations of the existing techniques are discussed and issues of research interest are pointed out in Chapter 3. Our results are summarised in Chapter 4 in the form of four articles. Chapter 5 contains discussion and possible directions for future work.

Chapter 1. Background

PET is a non-invasive imaging technique that allows assessing biochemical processes taking place in a living organism *in vivo*. PET is a valuable tool in investigation of normal human or animal physiology, diagnosing and staging cancer, heart and brain disorders. A brief overview of PET can be found in (Mandelkern, 1995; Raichle, 1998).

1.1. Principles of PET

A *radiopharmaceutical* is a substance containing very small amounts of radioactive nuclei, called *radiotracer* or simply *tracer*, used to follow the course of a chemical or physiological process without perturbing it significantly. In PET, a subject is administered a radiopharmaceutical which is used either for perfusion, metabolic or receptor-binding imaging (Saha *et al.*, 1992).

The radiopharmaceuticals used in PET are labelled with radionuclides that decay by emitting a positron, which subsequently annihilates with an electron of the surrounding body tissue. Two photons of 511 keV each are produced as a result and propagate in approximately opposite directions. This physical process is described by the following symbolic sequence:

$$p^+ \rightarrow n + \beta^+ + \nu \Rightarrow \beta^+ + e^- \rightarrow 2\gamma, \qquad (1.1.1)$$

where p^+ - proton, n - neutron, β^+ - positron, ν - neutrino, e^- - electron, and γ - annihilation photons.

A pair of detectors placed along the photons' path outside the subject being imaged may register the annihilation photons. Usually the subject is positioned within a ring of detectors. Any two opposite detectors in the ring are electronically collimated to signal a virtually simultaneous detection, which is called a *coincidence event*. The two signalling detectors define a *line-of-response* (LOR). In contemporary PET scanners, there are tens of thousands LORs ready to register photons emitted at different angles along a single plane, i.e. in a two-dimensional (2-D) mode. Multiple rings of detectors may be stacked together as a hollow cylinder to allow three-dimensional (3-D) sets of LORs to be acquired. Coincidence events collected with a given set of LORs are called *projection data*.

There are several phenomena related to PET data acquisition that must be taken into account. They may be divided into two categories: those related to the basic physics of positron emission and high energy photon interaction in matter, and those introduced purely by the presence of an imperfect (though state-of-the-art) detection instrument. The first group includes positron range after emission (Levin and Hoffman, 1999; Cho *et al.*, 1975), non-collinearity of the two annihilation photons (Muehllehner, 1976), photoelectric absorption of photons in matter, Compton scattering of photons, and gradual decrease of the radioisotope concentration over time due to decay of the positron emitter (Sorenson and Phelps, 1987). The second group includes random coincidences (Hoffman *et al.*, 1981), non-uniform efficiency of different detectors (Casey and Hoffman, 1986), photon scatter in detectors (Msaki *et al.*, 1996), spatial response non-uniformity due to detector arrangement (Hoffman *et al.*, 1982), and system dead time due

to the inability to handle high rates of incident photons (Budinger, 1998; Germano and Hoffman, 1991; Daube-Witherspoon and Carson, 1991).

As a result of all mentioned factors, PET data represent corrupted information on the actual physical phenomenon of the radiotracer decay within a subject. Much research effort has been devoted to overcoming the limitations of PET by searching for better scintillation crystals (Melcher, 2000), improving scanner electronics and design (Derenzo *et al.*, 1993; Links, 1998) and taking into account relevant phenomena during *image reconstruction*, which is an estimation of the unobserved radioactive decay density [Bq/cc] given the observed projection data.

The maps of radioactive decay density are analysed either by a human observer visually or with the help of a computer using dedicated mathematical models in order to find anomalies or yet unknown patterns of radiotracer distribution. These valuable diagnostic or research data shed light on organism functioning (Hoh *et al.*, 1997; Saha *et al.*, 1992).

Unlike X-ray computed tomography (CT), which probes tissue density and yields anatomical information, PET provides data characteristic to molecular function and makes monitoring of the physiological processes possible *in vivo*. The ability to detect very low concentrations of the tracer (on the order of 10^{-12} M) is a clear advantage of PET (Volkow *et al.*, 1997). The feasibility of accurate correction for photon attenuation in tissue and higher spatial and temporal resolution render PET superior to Single Photon Emission Computed Tomography (SPECT), an imaging technique utilising radiotracers decaying by single γ-ray emission (Budinger *et al.*, 1979). Since PET and SPECT are based on related principles, sometimes they are collectively referred to as *emission*

computed tomography (ECT) as opposed to transmission X-ray CT. Additionally, ECT involves the problem of determination of the radioisotope distribution and the distribution of attenuation coefficient that is necessary for accurate quantification whereas X-ray CT is concerned solely with the distribution of attenuation coefficient.

1.2. PET inverse problem

Tomographic data are object "views" or "projections" at different angles. Given PET projection data, one has to obtain the underlying emission density that could have been the source of acquired projections (see figure 1). This would be an approximation to the actual tracer distribution in the subject. The emission density map is also called an *image* since it is common to have it visualized after reconstruction.

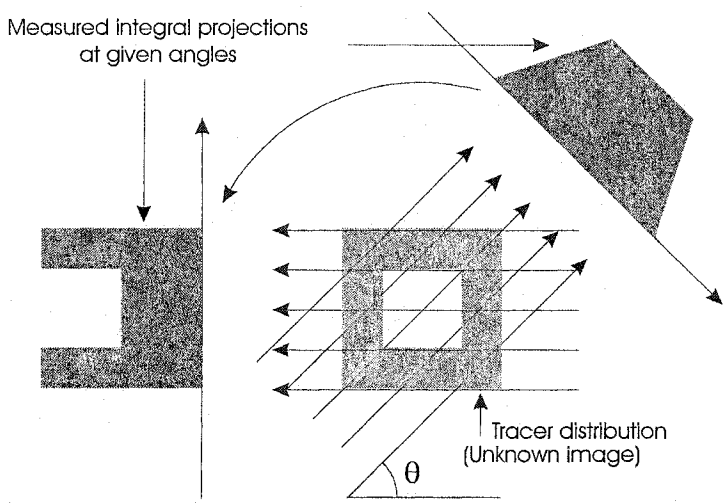

Figure 1. Simple case of radiotracer distribution and two idealized continuous projections at selected angles. Projections are shown as if they were not corrupted by the factors inherent to PET data acquisition. Note that projections at angles θ and $\theta \pm 180^{\circ}$ are identical.

In a broad sense, an *inverse problem* arises when making inference about a quantity that cannot be observed directly. In tomography, acquired data are used to determine characteristics of a physical phenomenon that cannot be followed otherwise. PET scanners provide projections of the tracer distribution given sufficient time to collect enough coincidence events, which are also referred to as *counts*. The most likely tracer distribution within the subject being imaged must be derived with dedicated mathematical methods using acquired projection data.

Solution of a tomographic inverse problem is the subject of *image reconstruction* and provides an *image estimate*. There are a number of different image reconstruction approaches that have been developed to date including many minor variations. This will be clarified further in Chapter 2, but first let us briefly outline the apparatus of image reconstruction, starting with the PET inverse problem, i.e. the mathematical problem that has to be solved.

1.2.1. Radon transform model

The first published work elaborating on the inverse problem similar to that of CT was by Radon in 1917 (Radon, 1986). He found a solution to the problem of determining a function of two variables from a set of its straight-line integral values.

Let $x=(x_1,x_2)\in \mathbf{R}^2$ be a point, line $L_{p,\theta}$ be given by equation $x_1\cos\theta + x_2\sin\theta = p$, and the line integral of a real function $f(x_1,x_2)$ along $L_{p,\theta}$ be given by

$$R_\theta(p)=\int_{-\infty}^{\infty} f(p\cos\theta - s\sin\theta, p\sin\theta + s\cos\theta)ds\,. \qquad (1.2.1)$$

With some assumptions Radon proved that the function f could be uniquely determined from the complete set of integral projections, i.e. when $R_\theta(p)$ is known for all possible lines $L_{p,\theta}$, i.e. with infinitely fine sampling at all possible angles. The problem posed in the Radon's paper is not applicable to PET directly, since a complete (infinite) set of projections is not available in practice and infinitely thin LORs are not best suited for robust modeling of system response, but his approach became the basis for a number of practical image reconstruction techniques. Integral functional (1.2.1) is known as the *Radon transform* of the function f. When $x \in \mathbf{R}^n$ and L is any line in $\mathbf{R}^n$, the integral

$$R_L = \int_L f(x)dx, \tag{1.2.2}$$

of which (1.2.1) is a special case, is known as the *X-ray transform* (Louis and Natterer, 1983).

1.2.2. Functional semidiscrete model

Though an idealized problem formulation based on the Radon (and the X-ray) transform was adopted in PET first, better modeling of underlying physical processes requires replacing it with a more sophisticated photon detection model. The following continuous-to-discrete mapping involving a set of simultaneous integral equations can be appropriate (e.g. Baker *et al.*, 1992):

$$n(d) \cong \int_\Omega r(d,x) f(x)dx, \quad d = \overline{1,N}. \tag{1.2.3}$$

where $x \in \Omega \subset \mathbf{R}^n$ is a point of an Euclidean n-dimensional space ($n = 2$ or $n = 3$), the support region Ω, associated with the PET scanner field-of-view (FOV), is bounded and

compact, $f(x)$ is the unknown emission density, $r(d,x)$ is the kernel specified by a model of PET detection process – a spatially varying system response to a source located at point x for the detector pair having index d, $n(d)$ is the number of counts acquired with the detector pair having index d, and N is the total number of active detector pairs. We assume that all N detector pairs are ordered so that a unique pair is referred to by specifying a unique index d. This is in contrast to a more conventional way of having projection angle θ and bin position p in a projection $n(p,\theta)$ as two separate parameters. Single equations in (1.2.3) are varieties of the *first-kind Fredholm integral equation*, if strict equalities are assumed (Hansen, 1992).

Quite generally an image may be viewed as a vector in the Hilbert space* $\mathbf{L}_2(\Omega)$, where $\mathbf{L}_2(\Omega)$ denotes the space of square-integrable functions** defined over Ω (Louis and Natterer, 1983). $\mathbf{L}_2(\Omega)$ would be the *image space* in this case.

The dynamic aspect of PET sometimes requires time as the fourth image dimension, i.e. $f(x,t)$, where $t \in [a,b]$ and $0 \le a < b < \infty$.

1.2.3. Stochastic extension of the functional model

It is intuitively clear that the true continuous emission map is impossible to recover using the finite number of projections, and one faces a problem that is characterized as ill posed. "An ill-posed experimental problem is one for which there is not as much information in your experimental data as you really need to find out what you want to

---

* Hilbert space is a complete vector space in which a norm and a scalar product are defined.

** Function $f(x)$ is square-integrable over Ω if $\int_{\Omega} |f(x)|^2 dx < \infty$

know" (Wahba, 1987). By a more rigorous definition of Hadamard, a problem is *well posed* if the solution exists, is unique, and depends continuously on the data, otherwise, a problem is *ill posed* (Franklin, 1970).

True projections are not available due to the physical nature of emission scanning. Many factors contribute to the corruption of projections as mentioned in section 1.1. Thus, having equality in (1.2.3) would be next to impossible. An additional zero-mean error vector $\varepsilon(d),\ d=\overline{1,N}$ is introduced into the deterministic model (1.2.3) to bring it to equality:

$$n(d)=\int_{\Omega} r(d,x)f(x)dx+\varepsilon(d),\ d=\overline{1,N}. \tag{1.2.4}$$

It is also commonly assumed that the noise $\varepsilon(d),\ d=\overline{1,N}$ can be represented with uncorrelated stochastic variables having some (probably unknown) probability distribution.

1.2.4. Stochastic model

The process of measuring tomographic "projections" in ECT is stochastic in nature. It is widely accepted that the Poisson distribution* describes the counting statistics for large quantities of radioactive nuclei. The importance of taking into account the stochastic data errors during image reconstruction was pointed out as early as in mid-1970s (Rockmore and Macovski, 1976). Thus, a *statistical model* of the PET inverse problem has been proposed (Shepp and Vardi, 1982; Vardi *et al.*, 1985):

---

* Poisson distribution gives the probability of observing x events given the average number of events μ per time interval according to $P(x;\mu)=\frac{\mu^x}{x!}e^{-\mu}$.

$$E[\eta(d)] = \int_{\Omega} p(d|x)\lambda(x)dx, \quad d = \overline{1,N}, \tag{1.2.5}$$

where $\lambda(x)$ is the unknown emission density at point x and is assumed to be the mean of an independent random variable (r.v.) having the Poisson distribution, $p(d|x)$ is the probability of an emission at point x being registered by the detector pair d, $E[\eta(d)]$ is the expected value of r.v. $\eta(d)$. The number of counts acquired with the detector pair d gives one sample $n(d)$ of $\eta(d)$, and $\lambda(x)$ has to be estimated. $\eta(d), d = \overline{1,N}$ are independent Poisson r.v. as well.

1.2.5. Poisson vs. Gaussian data

In practice one records data that are limited by the properties of detection system as well as contaminated by the non-negligible effects of the associated physical phenomena mentioned in section 1.1. A practical conclusion is that PET data are not expected to be purely Poisson. Indeed, it was shown that data precorrected for random coincidences may be approximated by the Gaussian distribution* (Fessler, 1994). Moreover, it is known that the Poisson distribution could be hard to tell from the Gaussian distribution for large values of the mean. This "large" mean is not that large in practice, e.g. (Bevington and Robinson, 1992) advises that "for values of the mean greater than about 10, the Gaussian distribution closely approximates the shape of the Poisson distribution." This provides an option of assuming the normal data error distribution, at least under the

* Gaussian (or normal) distribution gives the probability of obtaining value x according to the following law: $P(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-\mu)^2}{2\sigma^2}}$, where μ is the mean and σ is the standard deviation.

mentioned conditions. Therefore, the left-hand side of (1.2.5) could be changed to accommodate these two cases:

$$\omega(d) = \int_{\Omega} p(d|x)\lambda(x)dx, \quad d = \overline{1,N}. \tag{1.2.6}$$

The expectation $\omega(d)$ could correspond to either Poisson or Gaussian distribution and this choice may be made based on the analysis of the data sample at hand. The applicability range could be derived and the validity of the resulting solution would depend on the fulfilment of the underlying assumption on the data nature. It should be noted that a Poisson distribution is easier to describe as it is fully defined by its mean, which is equal to its variance. However, the Gaussian distribution is somewhat more convenient and a considerable number of mathematical methods developed for other applications have been (or could be) transferred to tomography. See Appendix A for an additional explanation.

1.2.6. Deterministic vs. stochastic modeling

Equations (1.2.3) and (1.2.6) have seemingly similar appearance, but they are distinct conceptually and the major difference is in the assumptions on the measured data vector. The former is considered *deterministic*, whereas the latter represents the *stochastic model* of PET. Equation (1.2.4) is the bridge connecting the two. Note also that $\lambda(x) \in \mathbf{L}_2(\Omega)$ since the support is bounded and the r.v. mean is finite in any practical situation. The system response $r(d,x)$ may be easily translated to probabilities $p(d|x)$ with an appropriate normalization.

A model of PET data acquisition must be set forth as it entails the logic of getting a reasonable solution to the inverse problem. Based on the adopted formulation of the PET inverse problem, (approximate) solution of the set of simultaneous equations (1.2.3) or (1.2.4) or (1.2.6) is the subject of *PET image reconstruction*. The symbols λ and f will be used hereafter to denote an image to emphasise its statistical or functional properties, respectively.

Image reconstruction is just an intermediate step in deriving definitive knowledge from PET data. Therefore, it is pointless to argue which model is the best one when detached from the context. The assessment could be performed by comparing reconstruction results on the basis of objective (quantitative) image estimation parameters as well as based on the subjective suitability for providing reliable and conclusive results. Minor objective improvements in image quality do not always lead to improved subjective interpretation of reconstructed images. It is known that the change of inverse problem formulation accompanied by the appropriate change of the image reconstruction method does not necessarily yield clinically important changes in the image estimate. Therefore, either model may be justified if respective assumptions on the data nature are closely satisfied.

1.3. Discrete image and quadrature options

The space $\mathbf{L}_2(\Omega)$ is separable, i.e. any object $g(x) \in \mathbf{L}_2(\Omega)$ can be represented as an infinite series (Barrett, 1999):

$$g(x) = \sum_{i=1}^{\infty} g_i \psi_i(x), \tag{1.3.1}$$

where $\{\psi_i(x)\}$ is the *orthonormal basis* for $\mathbf{L}_2(\Omega)$, i.e.

$$(\psi_n, \psi_m) = \delta_{nm}, \tag{1.3.2}$$

where δ_{nm} is the Kronecker delta function:

$$\delta_{nm} = \begin{cases} 1, & n = m \\ 0, & n \neq m \end{cases} \tag{1.3.3}$$

The coefficients in the series are given by the scalar product

$$g_i = (\psi_i, g). \tag{1.3.4}$$

A number of image reconstruction techniques use discrete image representation right from the start, i.e. with the inverse problem formulation. Remembering that an image can be represented as an infinite series (1.3.1), one may choose to approximate it with a finite number of basis vectors for simplicity. It is possible to use a non-orthogonal basis as well, e.g. functions $\phi_i(x)$, $i = \overline{1, M}$ spanning some space $\mathbf{\Phi}(\Omega)$. It is preferable to have $\mathbf{\Phi}(\Omega) \subset \mathbf{L}_2(\Omega)$, which is automatic if $\phi_i(x) \in \mathbf{L}_2(\Omega)$ for all i. $\mathbf{\Phi}(\Omega)$ is the *image representation space*. The linear approximation to the image in this case would be

$$f(x) \cong \sum_{i=1}^{M} f_i \phi_i(x), \tag{1.3.5}$$

where $\phi_i(x)$ are not necessarily normalized and orthogonal, M is the total number of these functions. The functions $\phi_i(x)$ are called *expansion functions* to distinguish them from the basis in $\mathbf{L}_2(\Omega)$ featured in (1.3.1).

A common practical method of obtaining a set of expansion functions involves representing the support Ω with a set of smaller compact regions Ω_i, called *pixels* in 2-D and *voxels* in 3-D case. The unity $\bigcup$ of all pixels* must contain Ω:

$$\Omega \subset \bigcup_i \Omega_i \tag{1.3.6}$$

A pixel Ω_i may serve as the support for an expansion function $\phi_i(x)$ in the simplest case, and the value of $\phi_i(x)$ is assumed constant over Ω_i. This simplistic approach lacks theoretic justification and poses the problem of pixel grid optimization, however, the set of expansion functions would be automatically orthogonal if defined on non-overlapping pixels. These can be conventional square pixels of the same size, or a more complicated set of *polar pixels* may be used (Kearfott, 1985; Kaufman, 1987), see figure 2 below for an illustration.

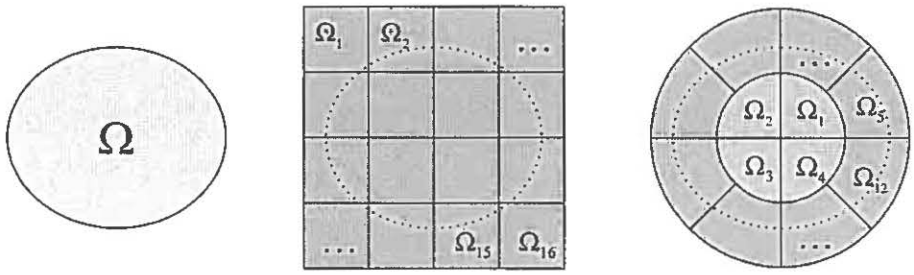

Figure 2. From left to right: a) an example of the support region Ω, which is the ellipse interior; b) a set of conventional square pixels; c) a set of polar pixels.

A grid of square pixels is very suitable for subsequent image visualisation with digital displays, therefore it is most widely utilised in practice. Another attractive alternative is

* The description given hereafter will be limited to the 2-D case to simplify examples but without loosing generality.

the grid of natural pixels (Buonocore *et al.*, 1981) that is easy to appreciate from the following explanation. Any two opposite detectors in the ring can register annihilation photons originating only from a compact region which is usually significantly smaller than the support Ω but large enough to be poorly represented by a single infinitely thin line connecting two detectors as the Radon transform model implies. A compact region that actually represents the FOV of two given detectors is called a *tube-of-response* (TOR). An example of a strip TOR model is shown in figure 3.

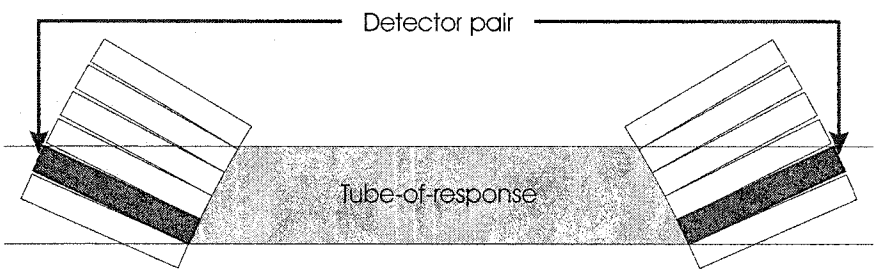

Figure 3. Tube-of-response for a given detector pair. It is assumed that the two detectors in the ring are able to register coincident annihilation photons originating from any point of the tube-of-response. The model may be applied in 2-D image reconstruction.

TOR is a more realistic detection model for PET as compared to LOR. The inverse problem can be written using the TORs. The model (1.2.6) becomes, for instance:

$$\omega(d) = \int_{S_d} p(d|x)\lambda(x)dx, \quad d = 1,...,N, \tag{1.3.7}$$

where S_d is the TOR corresponding to the detector pair d. The intersection of a TOR with the PET scanner FOV yields a *natural pixel* and is shown in figure 4. An orthonormal pixel basis based on natural pixels has been proposed in (Baker *et al.*, 1992).

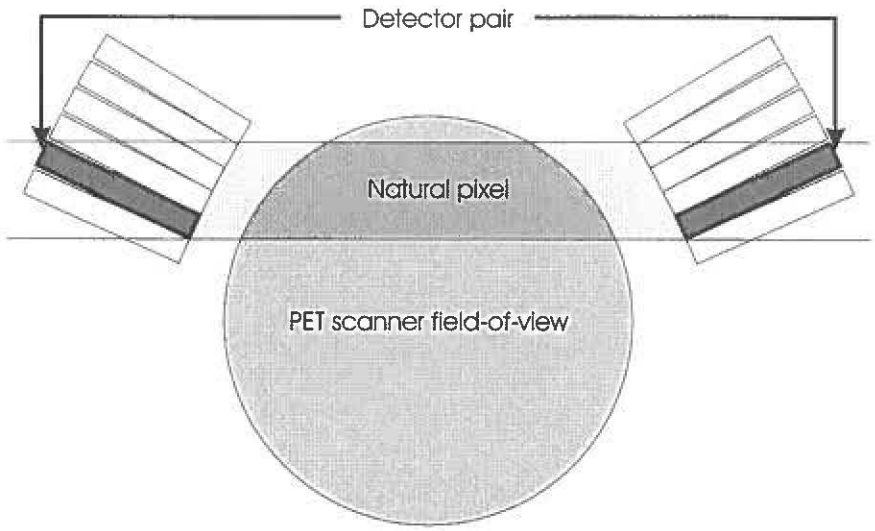

Figure 4. Natural pixel defined as an intersection of the tube-of-response shown in the previous figure and the PET scanner field-of-view.

It is possible, however, to define a convenient set of expansion functions to facilitate the problem solution and expand the solution in another set afterwards if need arises. A multitude of candidates for expansion functions are feasible, e.g. a set of spherically symmetric volume elements (blobs) was proposed (Lewitt, 1992; Matej and Lewitt, 1996) to simplify calculation of image projections and to have control over image smoothness properties.

The choice of a pixel grid is an important one, especially when a system with unconventional geometry is used for data acquisition. An improper choice would result in unnecessary complications aggravating the issues of solution stability and uniqueness. Irrespective of the choice of expansion functions, however, the final image is remapped to conventional square pixels to facilitate image display and sharing in digital form.

1.4. Image projection

An idealized illustration of continuous object projections in 2-D was given already in figure 1. Actual raw PET data are a finite set of numbers representing the number of counts acquired in a given TOR over time. This set of values, arranged in a 2-D matrix with one matrix dimension representing projection angles and another one bins in a projection (within a tomographic plane), is usually called a *sinogram* due to the sinusoidal trace that an off-centred point source leaves in this matrix (see figure 5). A sinogram row usually represents a single *object projection* at a given angle. Then, elements of the row represent different *projection bins* that this discrete object projection is comprised of.

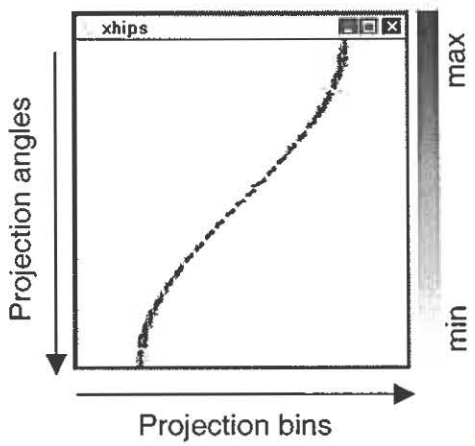

Figure 5. The sinogram of an off-centre "point" source (scanned through a line source extending axially).

PET imaging may be quite generally modelled with the continuous-to-discrete mapping $\mathbf{P}: \mathbf{L}_2(\Omega) \to \mathbf{R}^N$ as discussed in section 1.2. The linear operator $\mathbf{P}$ is bounded, i.e. there exists a positive number c such that $\|\mathbf{P}f\| < c\|f\|$ for all f. However, its inverse is unbounded and $\mathbf{P}$ is often substituted with its discrete counterpart $\mathbf{P}_\Phi : \Phi(\Omega) \to \mathbf{R}^N$ for practical purposes. When computed for a given PET system, $\mathbf{P}_\Phi$ is represented with a $N \times M$ matrix $P = \left\{ p_{ij};\ i = \overline{1,N},\ j = \overline{1,M} \right\}$. Then, the systems of

integral equations (1.2.3) and (1.2.4) can be reduced to the following matrix forms, respectively:

$$n \cong Pf\,. \tag{1.4.1}$$

and

$$n = Pf + \varepsilon\,. \tag{1.4.2}$$

where n is an N-dimensional vector of projection data; the matrix P is comprised of probabilities, i.e.

$$\sum_{i=1}^{N} p_{ij} = 1,\ \ j = \overline{1,M}\,, \tag{1.4.3}$$

which are obtained by the normalization of the system response model, and $f = \left\{ f_i : i = \overline{1,M} \right\}$ is the set of expansion coefficients characteristic for the image approximation expressed as the finite series (1.3.5).

The statistical model (1.2.6) can be represented as the system of linear equations just the same way:

$$\omega = P\lambda\,, \tag{1.4.4}$$

where ω is an N-dimensional vector of expectation of a Poisson or Gaussian r.v. as discussed in section 1.2.5, and $\lambda = \left\{ \lambda_i : i = \overline{1,M} \right\}$ is the set of expansion coefficients representing the image.

One should keep in mind that the actual incarnation of operator $\mathbf{P}$ is determined by the physical modeling for a particular scanner and the chosen image representation. Different photon detection models as well as other relevant physical phenomena

associated with the system response model will, generally, yield distinct matrices P. The modeling of the system response is called the *direct problem*, and is an essential part of solving the inverse problem. The matrix P, approximating operator $\mathbf{P}$, is called either the *system matrix* or the *transition matrix*[*]. The system matrix is usually very sparse, i.e. has many zero entries. This is mainly due to the modeling that assumes that TOR covers a finite space region, which is much smaller than the FOV (see figure 4); hence, a given TOR might see only a fraction of the set of local expansion functions. That happens if expansion functions are localized. Image projection becomes as simple as mapping $\mathbf{R}^M$ into $\mathbf{R}^N$ with the help of the system matrix if $\mathbf{P}_\Phi$ is set.

1.5. Backprojection

Another essential operation has yet to be introduced, which is referred to as the *backprojection*. It is the adjoint of image projection described in the previous section and maps from the finite dimensional projection space to the infinite dimensional image space, i.e. $\mathbf{P}^\dagger : \mathbf{R}^N \to \mathbf{L}_2(\Omega)$ (Natterer, 1980):

$$\left(\mathbf{P}^\dagger \mathrm{g}\right)(x) = \sum_{i=1}^{N} g_i \chi_i \,, \tag{1.5.1}$$

where g_i are the weights and

$$\chi_i(x) = \begin{cases} 1, & x \in S_d \\ 0, & \text{otherwise} \end{cases} \tag{1.5.2}$$

[*] The term "projection matrix" can be met as well. However, one should not think of the system matrix as a projector, in terms of the theory of matrix operators. A real valued projector must be symmetric and idempotent, i.e. $P = P^{\mathrm{T}}$ and $P^2 = P$, respectively, and the system matrix generally does not qualify.

The discrete counterpart of the backprojection operator $\mathbf{P}_{\Phi}^{\dagger}: \mathbf{R}^N \to \Phi(\Omega)$ is employed if one has chosen to use practical image approximations. Therefore, assuming that the same image representation is used for backprojection as for projection, the backprojection operator will be approximated by the transpose of the system matrix P^{T}, since for a real matrix the adjoint and transpose are synonymous. Thus, with the image approximation given in $\Phi(\Omega)$, image backprojection is as simple as mapping $\mathbf{R}^N$ back into $\mathbf{R}^M$.

1.6. System matrix properties and inverse problem analysis

The challenges of the inverse problem solution can be directly associated with the properties of the respective projection operator $\mathbf{P}$. Getting rid of the problem ill-posedness could be a sufficient reason for representing a continuous image as a finite series, however this approximation does not necessarily convert the inverse problem into a well posed one in the sense of Hadamard. Some insight can be gained by examining the projection operator or its discrete approximation - matrix *P*.

Several important definitions have to be recalled first. The set of all vectors that the operator $\mathbf{P}$ act upon is called the *domain* of $\mathbf{P}$. The *range of operator* $\mathbf{P}$ is defined as the set of all vectors *g* that can be reached by applying $\mathbf{P}$ to the members of a given functional space, i.e. $\mathscr{R}(\mathbf{P}) = \{g : g = \mathbf{P}f, \text{ for some } f \text{ from the domain of } \mathbf{P}\}$. The *nullspace of operator* $\mathbf{P}$ is defined as the set of all functions that are mapped to zero, i.e. $\mathscr{N}(\mathbf{P}) = \{f : \mathbf{P}f = 0\}$. The same definitions are applied to matrices that are considered in the case of discrete-to-discrete mapping. The operator $\mathbf{P}$ is *injective*, or one-to-one, if each vector of the original functional space is mapped to a single vector in $\mathscr{R}(\mathbf{P})$. The

operator $\mathbf{P}$ would be *surjective*, or onto, if all vectors of the target space would be in the range $\mathscr{R}(\mathbf{P})$. The operator $\mathbf{P}$ is *invertible* if it is injective and surjective. The nullspace of an invertible operator is trivial, i.e. $\mathscr{N}(\mathbf{P}) = \{0\}$. The operator that is not invertible is said to be *singular*. The real-world imaging operators are always singular (Barrett, 1999).

A powerful approach to inverse problem analysis involves the *singular value decomposition* (SVD) of the projection operator. Here we will describe the discrete case only since it is directly relevant for the following presentation. The application of SVD to the analysis of continuous-to-continuous mapping may be found in (Davison, 1983; Louis, 1986; Caponnetto and Bertero, 1997).

It is well known that any $N \times M$ matrix P can be decomposed into a product of three special matrices (Strang, 1980):

$$P = UDV^T, \tag{1.6.1}$$

where $D = diag(\mu_1, \mu_2, ..., \mu_M)$ is a diagonal matrix of *singular values*, $U = (u_{ij})_{\overline{i=1,N};\ \overline{j=1,M}}$ is an $N \times M$ matrix with orthonormal columns, that are referred to as the *left singular vectors* $u_j \equiv u_{\bullet j} = (u_{ij})_{\overline{i=1,N}}$, $V = (v_{ij})_{\overline{i=1,M};\ \overline{j=1,M}}$ is an $M \times M$ matrix with orthonormal columns, that are referred to as the *right singular vectors* $v_j \equiv v_{\bullet j} = (v_{ij})_{\overline{i=1,M}}$. The factored representation (1.6.1) is known as the matrix SVD. The left singular vectors u_j corresponding to the non-zero singular values form an orthonormal basis spanning the range of P. The right singular vectors v_j corresponding to the zero singular values form an orthonormal basis for the nullspace of P. The left and right singular vectors are

related via

$$Pv_j = \mu_j u_j \tag{1.6.2}$$

and

$$P^{\mathrm{T}} u_j = \mu_j v_j . \tag{1.6.3}$$

The set of singular values is called the *singular value spectrum.* The singular values are usually ordered so that

$$\mu_1 \geq \mu_2 \geq \ldots \geq \mu_M \geq 0 . \tag{1.6.4}$$

The decay rate of the singular values is an important indicator of information content for the inverse problem (Gilliam *et al.*, 1990; Wahba, 1980). In case of continuous mapping, the problem is *mildly ill posed* if the singular values decay slowly and there is a good chance of finding a stable (approximate) solution. The problem is *severely ill posed* and calls for a dedicated solution technique if the singular value spectrum drops towards zero rapidly. The ratio

$$C_P = \frac{\mu_1}{\mu_M} \tag{1.6.5}$$

is referred to as the *matrix condition number* and is a simple measure of the degree of inverse problem ill-posedness. A matrix is *singular* if $C_P = \infty$. If the matrix is not singular but the condition number is significantly large, the matrix is said to be *ill conditioned.* The inverse matrix P^{-1} would have a very large norm in that case, which would result in huge amplification of minor perturbations. The singular values are also directly related to the eigenvalues* σ_i of the self-adjoint matrix $P^{\mathrm{T}}P$ (hence, PP^{T} as

* The eigenvalues λ and eigenvectors x corresponding to a given matrix A satisfy the equation $Ax = \lambda x$.

well), which appears in the normal equations*, and

$$\sigma_i = \mu_i^2 . \tag{1.6.6}$$

The problem (1.4.1) could have an exact solution if $n \in \mathscr{R}(P)$. Unfortunately, that is usually not the case and one has to settle with an approximate (generalized) solution f^+, which can be chosen according to some optimality criterion.

1.7. Regularisation

The presence of noise in projection data is not the only challenge inherent to the PET inverse problem as we have seen already. Special methods have to be employed when handling severely ill posed (or ill conditioned) problems that involve replacing the initial problem with another one having more favourable properties. This technique is called *regularisation*. Following (Bertero *et al.*, 1988), we define a regulariser for our inverse problem.

Let $\mathbf{P}^+$ be an operator that gives a generalized solution to (1.2.3) or (1.2.4):

$$f^+ = \mathbf{P}^+ n . \tag{1.7.1}$$

A family of linear operators $\{\mathbf{P}_\omega^+\}_{\omega>0}$, $\mathbf{P}_\omega^+ : \mathbf{R}^N \to \mathbf{L}_2(\Omega)$ constitute a regulariser for an operator $\mathbf{P}^+$ if the following conditions are satisfied:

(i) for any $\omega > 0$, $\mathscr{R}(\mathbf{P}_\omega^+) \subset \mathbf{\Phi}(\Omega)$;

(ii) for any $\omega > 0$, $\|\mathbf{P}_\omega^+\| \le \|\mathbf{P}^+\|$;

---

* Normal equations arise with the least-squares approach to approximate solution of an optimization problem.

(iii) $\lim_{\omega \to 0} \mathbf{P}_{\omega}^{+} = \mathbf{P}^{+}$.

Thus $\mathbf{P}_{\omega}^{+}$ is a regularised approximation of $\mathbf{P}^{+}$ and the variable ω is called the regularisation parameter. Choosing the most appropriate value of ω is one of the main problems of regularisation theory. A useful overview of the image reconstruction problem and regularisation can be found in (Demoment, 1989).

Chapter 2. Image Reconstruction Techniques

Various inverse problem models could describe PET imaging as has been discussed in section 1.2. A variety of established mathematical tools and numerical algorithms can be utilised to derive a reasonable solution – a feasible image approximation, even if the inverse problem formulation stays the same. Thus, a great number of image reconstruction algorithms have been conceived. An overview given in the present chapter will clarify current state-of-the-art. The reviews given in (Natterer, 1999) and (Leahy and Qi, 2000) cover somewhat different perspectives and may be a brief introduction to the subject as well.

2.1. General Classification

The primary differences among reconstruction techniques stem from the adopted formulation of the PET inverse problem. This leads to two broad classes of algorithms, namely *deterministic* and *statistical*, based on the respective inverse problem models discussed in section 1.2. The choice of discrete or continuous image model differentiates *series-expansion* from *transform* methods, respectively. Yet another fundamental distinction among algorithms exists: they may be classified in *one-step* and *iterative* groups. One-step reconstruction methods aim at solving the inverse problem at a single (though complex) step, thus providing a "final" image estimate, whereas iterative methods attempt to reach a solution by successive improvement of an image estimate starting with some initial guess.

This classification clarifies underlying concepts, but image reconstruction algorithms might fall into several categories at the same time as the corresponding

classification reflects different and non-exclusive properties. Moreover, the boundaries between the opposite groups are sometimes vague. A deterministic model having no solution in the sense of Hadamard, for instance, may be solved approximately in the least-squares sense* and may involve statistical interpretation of data errors. A solution expanded into the set of eigenvectors of the system matrix can be related to conventional Fourier techniques (Llacer, 1979). A similar relationship has been studied in (Anastasio *et al.*, 2001) for the continuous case.

Another perspective onto the image reconstruction problem is presented in figure 6. Two alternative routes that employ regularisation at one stage or another may be taken to derive a "solution". Either the operator $\mathbf{P}$ is replaced by its discrete counterpart P which, in turn, is replaced by a regularised version P_ω, or the operator $\mathbf{P}$ is idealized (the Radon's approach is an example) and its regularised version $\mathbf{P}_\omega$ is studied completely in functional spaces and a numerical approximation is developed that involves discretisation P_ω of the regularised operator after that.

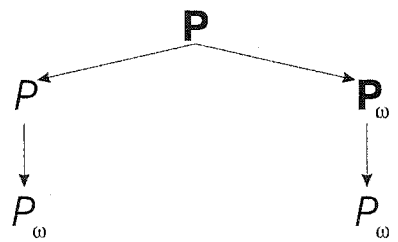

Figure 6. Diagram showing two alternative routes that could be taken when solving an inverse problem approximately.

* Function f is a least-squares solution of $Pf = n$ if $\inf\{\|Pu - n\| : u \in X\} = \|Pf - n\|$. The solution depends upon the choice of the norm.

It should be mentioned that the idealisation of $\mathbf{P}$ with the help of the Radon transform was studied first and the most. The powerful set of transform methods is the widely used result of developments with the Radon (or X-ray) transform. However, the series expansion approach employing early image discretisation is used more often in current research.

2.2. Summation Method

The easiest tomographic reconstruction method is the one called *summation* or back projection, due to the simple superposition of projections by spreading them back across the reconstruction plane. This would involve a single operation of backprojection introduced in section 1.5. The summation method yields image estimates that are blurred and the degree of blurring in the case of an infinite number of angles and the line integral measurement model is proportional to $1/r$, where r is the distance from the point source of radioactivity. For more details see e.g. the review in (Gordon and Herman, 1974).

Though it is hardly used as a standalone reconstruction method in practice, the backprojection operation is an essential part of most practical image reconstruction techniques. Efforts have been made to improve backprojection algorithms (Peters, 1981; Cho *et al.*, 1990; Egger *et al.*, 1998) along with designing special hardware for backprojection (Thompson and Peters, 1981; Hartz *et al.*, 1985; Jones *et al.*, 1990) to accelerate reconstruction.

2.3. *Transform Methods*

A distinct set of image reconstruction techniques is based on the *Fourier transform*[*], which in 2-D is defined as:

$$\mathcal{F}_2 f \equiv [\mathcal{F}_2 f](u,v) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) e^{-i(ux+vy)} dxdy, \tag{2.3.1}$$

where $f(x,y)$ is the image function, and $i^2 = -1$. A single projection at angle θ, when subject to the one-dimensional (1-D) Fourier transform

$$\mathcal{F}_1 R_\theta \equiv [\mathcal{F}_1 R_\theta](w) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} R_\theta(p) e^{-iwp} dp, \tag{2.3.2}$$

yields a central section of the 2-D Fourier transformed image at angle θ:

$$[\mathcal{F}_2 f](w\cos\theta, w\sin\theta) = [\mathcal{F}_1 R_\theta](w), \tag{2.3.3}$$

This relationship is referred to as the *projection slice theorem*, or the *central slice theorem*, or the *Fourier slice theorem* (Kak and Slaney, 1988).

A number of techniques employing Fourier transforms have been described. Reconstruction may be accomplished by transferring projections into the Fourier space, interpolating in the frequency domain to get the sampling on a square grid, and then transferring the resulting frequency estimate back into the image space with the *inverse Fourier transform*, which in 2-D will yield:

---

[*] The Fourier transform is defined as $\tilde{f}(\xi) \equiv [\mathcal{F}_N f](\xi) = (2\pi)^{-N/2} \int_{R^N} f(x) e^{-i\xi x} dx$; the inverse Fourier transform is defined as $[\mathcal{F}_N^{-1} \tilde{f}](x) = (2\pi)^{-N/2} \int_{R^N} \tilde{f}(\xi) e^{i\xi x} d\xi$.

$$f(x,y) = \mathcal{F}_2^{-1} \int_{2\pi} \mathcal{F}_1 R_\theta \, d\theta \,. \tag{2.3.4}$$

This approach is referred to as *direct Fourier method* (DFM). The major problem with DFM is the robust interpolation in the Fourier space to resample the points from polar to Cartesian grid, which results in image artefacts if a simplistic interpolation strategy is employed. For the latest developments with DFM see (Waldén, 2000; Gottlieb *et al.*, 2000).

Another approach is termed the *convolution method* and is based on the fact that the Fourier transform of the convolution of two functions* equals the product of their individual Fourier transforms

$$\mathcal{F}(\phi * \varphi) = \mathcal{F}\phi \times \mathcal{F}\varphi \,. \tag{2.3.5}$$

This leads to practical reconstruction techniques, since the measured emission data represent "ideal" projections convolved with the system point spread function (PSF)**. One popular algorithm starts with transferring each measured projection into the Fourier space, applying an appropriate filter to de-convolve the effect of image blurring (observed with the summation method), then transfers the result back into the projection space, and finally, backprojects the filtered projections onto the image grid. This technique is called *filtered backprojection* (FBP). The following equation can be an illustration:

$$\hat{f}_{FBP}(x,y) = \sum_{\theta_i} Backproject\left\{\mathcal{F}_1^{-1}\left[\tilde{c}(w) \times \mathcal{F}_1 \tilde{R}_{\theta_i}\right]\right\}. \tag{2.3.6}$$

---

* Convolution of 1-D functions $\phi(v)$ and $\varphi(v)$ is defined as $[\phi * \varphi](v) = \int_{-\infty}^{\infty} \phi(u)\varphi(v-u)du$.

** Not to be confused with PSF estimated from a reconstructed image.

It is customary to put a hat over f, to recognise the fact that this is an approximate solution, i.e. an *image estimate*. In (2.3.6), $\tilde{R}_{\theta_i}$ is the projection at angle θ_i sampled at a finite number of points, i.e. $\tilde{R}_{\theta_i}$ is a single row of the sinogram (as discussed in section 1.4), and $\tilde{c}(w)$ is a 1-D function, a sampled version of a window of the *ramp filter* $c(w)$ which arises as a correction for non-uniform sampling in the Fourier domain and is given by

$$c(w) = \begin{cases} |w|, & |w| \le w_{\max} \\ 0, & |w| > w_{\max} \end{cases}. \tag{2.3.7}$$

Here $w_{\max}$ is the cut-off frequency, which is determined from the given sampling with the Radon transform. This relationship stems from a result of the sampling theory that is known as the *Nyquist sampling rate*, which states that a signal must be sampled at least twice during each cycle of the highest frequency of the signal.

Extensive treatment of FBP theory and implementation is given in (Rowland, 1979; Budinger *et al.*, 1979; Kak and Slaney, 1988). FBP is currently the algorithm of choice in most practical applications due to its overall acceptable performance and relative computational simplicity, due to availability of efficient implementations of the fast Fourier transform and 1-D operations involved. Another ordering of backprojection, filtering and the Fourier transform is possible and would result in a different algorithm:

$$\hat{f}(x,y) = \mathscr{F}_2^{-1}\left\{ \tilde{c}(u) \times \mathscr{F}_2\left[\sum_{\theta_i} Backproject\left(\tilde{R}_{\theta_i}\right) \right] \right\}. \tag{2.3.8}$$

See (Rowland, 1979) or (Budinger *et al.*, 1979) for more detail.

Transform algorithms are computationally inexpensive but need a set of projections with angular sampling over 2π as well as high total counts in order to produce a high quality image. They are deterministic in nature; improvements in image quality have been attained by designing data dependent (adaptive) convolution filters to smooth the effect of measurement noise (Rowland, 1979; Budinger *et al.*, 1979). This intermediate filtering is equivalent to regularisation of the original problem. Pre-processing data to account for the system response has been explored as an additional option (Karuta and Lecomte, 1992; Liang, 1994). For more details on the mathematics of transform methods see (Lewitt, 1983; Herman, 1980).

2.4. Series Expansion Methods

Another class of methods, collectively called the series expansion methods, is based on the parametric image representation with expansion functions (1.3.5) right from the outset. An early overview of series expansion methods can be found in (Censor, 1983; Herman, 1980). Series expansion approach includes techniques developed for the deterministic as well as statistical PET models. Sometimes the representation of an image with (conventional) pixels (voxels) is understood as the essence of the approach, but this is an oversimplification since an image could be expanded in a variety of different orthogonal or non-orthogonal series as we mentioned already in section 1.3.

2.4.1. Orthogonal series

The tomographic image reconstruction based on orthogonal series expansion was among the first theoretically solid results. The structure of the Radon transform acting upon $\mathbf{L}_2(\Omega)$ can be established with the help of a basis in this space. It has been known

that a complete set of orthogonal functions in $L_2(\Omega)$ can be formed with Zernicke polynomials (Cormack, 1964; Louis, 1986). Marr reported that the Radon transform of this basis is closely related to the Chebyshev polynomials of the second kind and showed a procedure to construct the solution as a polynomial, such that its Radon transform fits the data in the least-squares sense (Marr, 1974). Recently, a practical and computationally efficient algorithm for 2-D image reconstruction based on the Chebyshev polynomials of the second kind have been shown in (Bortfeld and Oelfke, 1999). The Chebyshev polynomials of the second kind coupled with the natural pixel image decomposition have been used in (Kazantsev *et al.*, 2000) to derive an image estimate as a function of continuous variable, which allows, at least in principle, image visualization on an arbitrary grid.

2.4.2. Direct matrix method

Direct matrix inversion was considered early for image restoration in radionuclide imaging (Cahill and Blau, 1970). This method yielded compensation for image blur inherent in observations with real systems having finite spatial resolution (Iinuma and Nagai, 1967). A direct matrix approach has been described as an image reconstruction method in nuclear medicine as well and employed matrix pseudo-inverse P^+ but was intractable at the time (Budinger and Gullberg, 1974). Later, a maximum likelihood approach has been applied in emission imaging, taking into account the stochastic data nature and assuming that the measurements can be described by the Poisson process (Rockmore and Macovski, 1976). The latter approach involved pseudo-inverse matrix but stopped short of taking into account severe ill conditioning of the image reconstruction problem.

The minimum-norm least-squares solution of (1.4.2) may be found with the *pseudo-inverse* P^+ based on the SVD of the system matrix P:

$$P^+ = VD^+U^{\mathrm{T}}, \tag{2.4.1}$$

where D^+ is the diagonal matrix containing the reciprocals of the respective positive singular values, i.e.

$$D^+ = diag\left(\mu_1^{-1}, \mu_2^{-1}, ..., \mu_z^{-1}\right),\ z \le M, \tag{2.4.2}$$

and z is the total number of non-zero singular values.

The use of SVD enables explicit regularisation by spectral filtering which potentially improves the inverse problem conditioning. A simple regularisation approach employs truncation of the singular value spectrum at some non-zero value having index T (also referred to as the *truncation level*) and yields a "truncated" SVD (*TSVD*) solution:

$$\hat{f} = \left\{ \hat{f}_k : \hat{f}_k = \sum_{i=1}^{T} v_{ki} \mu_i^{-1} \sum_{j=1}^{N} u_{ji} b_j;\ k = \overline{1, M};\ T \le M \right\}. \tag{2.4.3}$$

An "optimal" truncation level can be derived with a number of approaches. The generalized *cross-validation*[*] was used in (Vogel, 1986). An application of the discrepancy principle[**] was studied and an alternative criterion was suggested in (Defrise and De Mol, 1987). The L-curve[***] approach was shown to assist in choosing the

---

[*] If the variance of the noise is unknown, one can use a technique called generalised cross-validation to get an approximate value of the regularisation parameter. See (Golub and von Matt, 1997) for the form of the function that must be minimised in this case.

[**] The discrepancy principle states that solution should not fit noisy data with the precision greater than that of the measurement, i.e. the function f is a solution if $\|Pf - n\| = \varepsilon$, where ε is the given noise level.

[***] L-curve is a plot, for all valid regularisation parameters, of the size (in certain metric) of the regularised solution versus the size of the corresponding residual.

truncation level (Hansen and O'Leary, 1993; Hansen, 1992). Several optimality criteria for SVD truncation have been studied in (Xu, 1998) and the minimisation of the mean squared error was recommended.

In PET, image reconstruction based on SVD has been attempted using special block circulant structure of matrix PP^T arising for rotationally invariant systems; an approximate SVD was computed with a supercomputer using these symmetries (Baker *et al.*, 1992). Image reconstruction based on matrix SVD has been applied in SPECT as well (Smith *et al.*, 1992; Gullberg *et al.*, 1996; Smith, 1996; Hsieh *et al.*, 1996). Despite these several attempts, the SVD approach has never been in the mainstream of medical image reconstruction due to the size of the system matrix and the time necessary to calculate the SVD.

2.4.3. Algebraic methods

The algorithm used by Hounsfield for image reconstruction in his prototype scanner was algebraic in nature and has been successfully applied in early practical implementations. A reconstruction method using the set of linear algebraic equations (1.4.1) and pursuing an iterative solution to the inverse problem is called an *algebraic reconstruction technique* (ART). Several ART variations were proposed that update the current pixel values in a different fashion (Gordon, 1974; Herman, 1980). Special data-access ordering was also investigated (Herman and Meyer, 1993). The algorithm convergence to a unique approximate solution is not guaranteed in general, since the exact solution does not exist in the presence of statistical errors in data. However, there is a version of ART that is claimed to converge to a regularised least-squares solution, see discussion in (Herman and Meyer, 1993). ART is not the only algorithm capable of

"solving" (1.4.1). To name a few, the steepest descent, Landweber iteration and the conjugate gradient method (Kawata and Nalcioglu, 1985) can be employed in iterative solution. Xu *et al.* have shown the generalized form of the iterative algebraic reconstruction, which includes several popular algorithms as special cases (Xu *et al.*, 1993):

$$\hat{f}^{(k+1)} = \hat{f}^{(k)} + r\mathbf{P}_{\Phi}^{\dagger}\left(n - \mathbf{P}_{\Phi}\hat{f}^{(k)}\right). \tag{2.4.4}$$

where $\hat{f}^{(k)}$ is the image estimate after iteration k, $\mathbf{P}_{\Phi}^{\dagger}$ is a backprojection operator specific to a given algorithm, and r is a positive relaxation parameter.

If the method is convergent and the inverse problem solution is known to exist, making an infinite number of iterations would ensure reaching the solution. However, iterating to full convergence is certainly not practical and is often meaningless, therefore, a reasonable rule to stop iterations has to be employed. Various criteria for estimating the solution feasibility, the so-called *stopping rules*, have been devised (Gordon, 1974; Defrise and De Mol, 1987).

2.4.4. Statistical methods

An image reconstruction approach involving the principle of maximum likelihood estimation and based on stochastic assumptions of the measurement process was proposed for ECT in mid-1970s (Rockmore and Macovski, 1976). The analytic maximum likelihood approach derived by Rockmore and Macovski was guaranteed to yield an unbiased and efficient* image estimate with Poisson data, but it did not take into

* Efficient estimate is the one that has the minimum variance.

account the severe ill conditioning of practical problems. More important, it was hardly tractable for the reconstruction problems of practical interest using that day's computers. Therefore, an iterative solution taking into account statistical data errors (Shepp and Vardi, 1982) got the most attention.

The likelihood of the observed data is estimated as

$$L(\lambda) \equiv P(n|\lambda) = \prod_{i=1}^{N} e^{-\lambda_i} \frac{\lambda_i^{n_i}}{n_i!}. \tag{2.4.5}$$

The log-likelihood

$$l(\lambda) \equiv \ln L(\lambda) \tag{2.4.6}$$

is considered afterwards for convenience and its global maximum is sought to solve the problem. Shepp and Vardi used an established mathematical technique for computing a *maximum likelihood estimate* (MLE) from incomplete data* known as the expectation maximization (EM) algorithm (Dempster *et al.*, 1977; Wu, 1983). The maximum likelihood (ML) estimation via EM, the *ML-EM* algorithm, is simple and elegant:

$$\hat{\lambda}_j^{(k+1)} = \hat{\lambda}_j^{(k)} \times \sum_{i=1}^{N} \frac{n_i p_{ij}}{\sum_{l=1}^{M} \hat{\lambda}_l^{(k)} p_{il}}, \quad j = \overline{1, M}. \tag{2.4.7}$$

The particularly attractive properties of ML-EM include natural incorporation of the image non-negativity and proven algorithm convergence (Lange and Carson, 1984; Vardi *et al.*, 1985). EM is but one algorithm to compute a maximum likelihood estimate. As noted in (Vardi *et al.*, 1985), EM is a gradient-type algorithm and (2.4.7) can be

* Incomplete data is understood in the sense that the mapping from $\Phi(\Omega)$ to $\mathcal{R}(P)$ is not injective, i.e. it is many-to-one.

written as

$$\hat{\lambda}_j^{(k+1)} = \hat{\lambda}_j^{(k)} + \hat{\lambda}_j^{(k)} \times \frac{\partial l\left(\hat{\lambda}^{(k)}\right)}{\partial \lambda_j}, \quad j = \overline{1, M}. \tag{2.4.8}$$

Based on this fact, it was proposed to accelerate ML-EM convergence by introducing an overrelaxation parameter into the iterative update equation, which yields additional multiplicative change of the correction item in the sum above (Lewitt and Muehllehner, 1986). However, this could potentially inhibit convergence and an occasional decrease in likelihood was indeed mentioned by Lewitt and Muehllehner for relaxation parameter values larger than 4 in their numerical experiments. ML-EM is a non-linear algorithm and the rate of recovery of smaller image detail is lower than that of larger structures (Liow and Strother, 1993). In an effort to make the convergence more uniform for an arbitrary image pattern, a way to improve algorithm's frequency response has been suggested in (Tanaka, 1987).

Despite its useful properties, the widespread application of the ML-EM algorithm was hindered by its high computational requirements, slow rate of convergence and, especially, by the *"noise" artefact*, i.e. increasingly speckled image appearance, which develops with the higher iteration numbers and destroys the appealing estimates attained with the previous iteration. Many efforts have been devoted to overcoming this artefact. Snyder and Miller applied the Grenander's method of sieves to suppress the artefact (Snyder and Miller, 1985). Terminating iterations early with a *stopping rule* has been studied as another way of avoiding the artefact (Llacer and Veklerov, 1989; Hebert, 1990; Coakley and Llacer, 1992). Initialising estimate $\hat{\lambda}^{(0)}$ with a smooth function is also important in avoiding early artefact development, since all the discontinuities present in

the initial estimate usually stay with the subsequent iteration (Kaufman, 1987). The "mechanics" of statistical iterative image reconstruction is better understood using the idea of feasibility applied to the ECT imaging (Llacer and Veklerov, 1989; Veklerov and Llacer, 1990; Llacer, 1990). In short, an image is presumed to be *feasible* if it "could have caused or produced the observed data by the known statistical process that governs the measurement." See figure 7 for a graphic illustration. Being improved iteratively, the image estimate usually passes through the *feasibility region*, which contains a set of "desirable" images. As the image is moved towards the MLE point with additional iterations, it becomes excessively noisy.

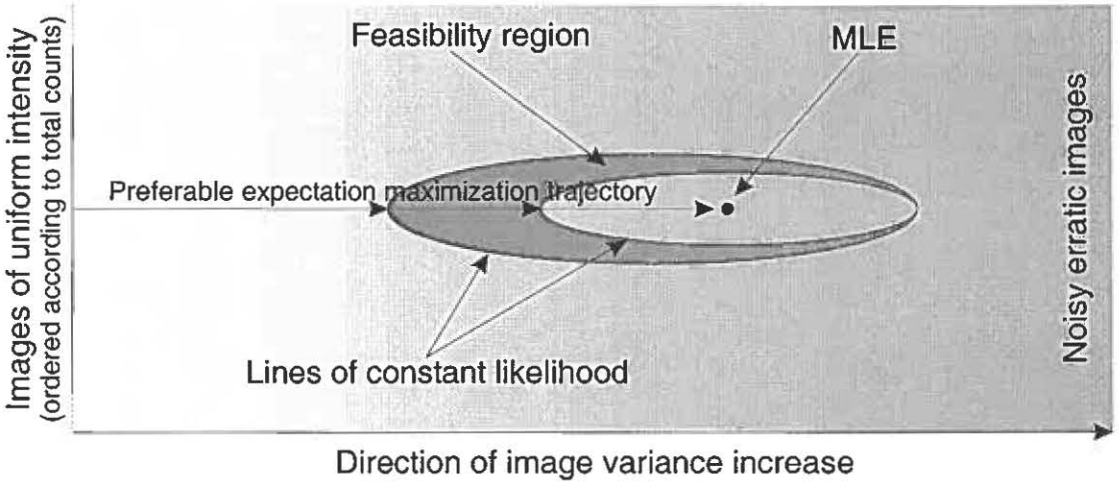

Figure 7. The schematics of the ML-EM image reconstruction and an example of the feasibility region are shown. The algorithm takes an image of uniform intensity as an initial estimate. ML-EM aims at the estimate having the maximum likelihood given measured data, which is theoretically reachable with an infinite number of iterations.

Regularisation techniques were applied to stabilize iterative image recovery. It was shown that utilisation of *a priori* information helps achieving more appealing images

(Lange *et al.*, 1987; Geman and McClure, 1987; Levitan and Herman, 1987). This so-called *Bayesian approach* allowed algorithmic enforcement of spatial correlation between pixels in the image estimate using the *Bayes' theorem*:

$$p(\lambda|n)=\frac{p(n|\lambda)p(\lambda)}{p(n)}. \tag{2.4.9}$$

The *a posteriori* probability $p(\lambda|n)$ explicitly takes into account a probabilistic model of one's expectations on image properties or structure, incorporated as the *a priori* probability distribution $p(\lambda)$, and leads to a maximum *a posteriori* (MAP) problem formulation. This is equivalent to penalized maximum likelihood estimation (Geman and McClure, 1987) that combines the log-likelihood $l(\lambda)$ with the penalty term $U(\lambda)$ into a joint objective function:

$$\mathrm{Z}_{U,\beta}(\lambda)=l(\lambda)-\beta U(\lambda). \tag{2.4.10}$$

The factor β sets a balance of sensitivity between the data and the prior information. This objective function has to be maximized in order to solve the regularised problem:

$$\hat{\lambda}=\arg\max_{\lambda}\mathrm{Z}_{U,\beta}(\lambda). \tag{2.4.11}$$

The optimization problem (2.4.11) is usually solved with numerical iterative techniques, but a non-iterative Bayesian solution has been considered as well in (Phillips, 1989). Various priors have been investigated, e.g. the use of a multivariate Gaussian probability distribution with the mean equal to a smoothed FBP reconstruction was proposed in (Levitan and Herman, 1987), the use of priors based on Markov random field model of the form

$$p(\lambda) = a^{-1} e^{-\beta U(\lambda)}, \tag{2.4.12}$$

where a is the normalization constant, was suggested in (Geman and McClure, 1987), a statistical model of the *a priori* probability expressed as a multinomial law was developed in (Liang *et al.*, 1989), the entropy prior has been employed in (Nunez and Llacer, 1990), even anatomical information obtained with other imaging modalities has been suggested to be used as a prior, see e.g. (Sastry and Carson, 1997). However, the choice of a prior and its optimization remain crucial for maintaining quantitative accuracy of the reconstructed images (Mumcuoğlu *et al.*, 1996).

The log-likelihood for the Poisson case is not the only measure appropriate for solution derivation. As discussed in section 1.2.5, the normal distribution could be an adequate model for PET data. In this case, a *weighted least-squares* (WLS) objective function could be employed (Kaufman, 1993; Fessler, 1994):

$$Z_{WLS}(\lambda) = (n - P\lambda)^{T} \Sigma^{-1} (n - P\lambda), \tag{2.4.13}$$

where $\Sigma = diag(\sigma_1^2, ..., \sigma_N^2)$ is the covariance matrix, which is diagonal since the measurement errors are assumed uncorrelated. A prior (or penalty) $U(\lambda)$ can be utilised as well yielding a *penalized weighted least-squares* (PWLS) objective (Fessler, 1994):

$$Z_{PWLS}(\lambda) = (n - P\lambda)^{T} \Sigma^{-1} (n - P\lambda) + \beta U(\lambda). \tag{2.4.14}$$

The image estimate is then obtained by

$$\hat{\lambda} = \arg\min_{\lambda} Z_{PWLS}(\lambda). \tag{2.4.15}$$

Obviously, various numerical algorithms could be employed for solving (2.4.15), the major consideration being the computational efficiency, resulting in rapid convergence (Pan and Yagle, 1991).

It is beyond doubt that a sophisticated statistical iterative reconstruction has the potential to outperform the transform methods, especially in low count situations, but the distinction from algebraic techniques becomes vague when the WLS approach is involved.

2.4.5. Iterative reconstruction acceleration

The computational cost of an iterative image reconstruction technique is significantly higher than that of a transform method. The efforts to accelerate iterative reconstruction can broadly be classified in two categories:

1) Algorithmic acceleration;

2) Parallel processing techniques.

A number of techniques falling in the first category was mentioned already, e.g. the use of a relaxation parameter or the choice of a faster numerical optimization algorithm. Preconditioning of the gradient-based methods was studied to accelerate convergence rate of iterative reconstruction (Chinn and Huang, 1997). Very fast versions of iterative algorithms can be obtained by rearranging them into a block iterative form (Soares *et al.*, 2000). The interest in block iterative acceleration was sparked by the OS-EM technique proposed in (Hudson and Larkin, 1994). *OS-EM* stands for EM with ordered subsets. It was a modification of the ML-EM algorithm that involved image updates based on data subsets S_m comprised of partial projection data. The procedure is then as follows:

$$\hat{\lambda}_j^{(k+1)}(S_m) = \frac{\hat{\lambda}_j^{(k)}}{\sum_{i \in S_m} p_{ij}} \times \sum_{i \in S_m} \frac{n_i p_{ij}}{\sum_{l=1}^{N} \hat{\lambda}_l^{(k)} p_{il}}, \quad j = \overline{1, M} \tag{2.4.16}$$

and is repeated for all subsets S_m sequentially, yielding a single iteration of OS-EM. Unfortunately, while gaining speed, it lost the convergence properties of ML-EM at the same time. A particular form of OS-EM that uses as many subsets as there are projections, was shown to converge to MLE after making certain assumptions on the relaxation parameter that are met by the training for a particular reconstruction task; this technique was named the row-action maximum likelihood algorithm (RAMLA) (Browne and De Pierro, 1996). The reduction of necessary computations by automatic truncation of the projection set based on data analysis, thus limiting reconstruction to a region-of-interest (ROI) within the FOV was considered as a simple but efficient acceleration technique (Gregor and Huff, 1997).

Another acceleration approach is the algorithm restructuring suitable for parallel computing and subsequent optimization for the dedicated computing hardware (Chen *et al.*, 1991; Johnson *et al.*, 1995; Chen and Lee, 1995; Cruz-Rivera *et al.*, 1995). ART-type algorithm parallelisation was investigated as well and reported in (Rajan *et al.*, 1997). An interesting alternative to dedicated parallel processors is the idea of distributed computing with a network of low-cost workstations, which has been explored in (Olesen *et al.*, 1996).

2.5. 3-D image reconstruction issues

Extending 2-D image reconstruction algorithms to 3-D PET may seem straightforward but it is hardly so in practice (Bendriem and Townsend, 1998). First of

all, a 3-D reconstruction problem calls for special computing approaches due to its size. Parallel or distributed computing as well as the utilisation of dedicated computing hardware have been in the focus of attention primarily fuelled by the needs of fully 3-D reconstruction. Parallel implementation of 3-D reconstruction algorithms has been considered using a number of computing topologies. In addition to the list of relevant references given in the previous section, the use of a network of transputers coupled with non-iterative image reconstruction techniques can be found in (Atkins *et al.*, 1991) and (Comtat *et al.*, 1993).

Several additional limitations render the 3-D case special. These include the significantly increased scatter background (Thompson, 1993), the need to cope with higher count rates (Mazé and Lecomte, 1990), unavailability of the full set of projections of the 3-D object (a similar case in 2-D is referred to as a limited angle problem), and the two orders of magnitude increase in the data set size may strain computing resources beyond capabilities.

The benefit of 3-D data acquisition for more efficient utilisation of the given amount of the radiotracer is obvious. The annihilation photons are not limited to any plane and are emitted isotropically in all directions. Thus, the sensitivity of a PET scanner would be significantly higher with the same radiotracer concentrations if coincidences were registered in oblique TORs also. The signal-to-noise ratio (SNR) could be improved as well (especially in the FOV centre) if the lower energy threshold is high (Karp *et al.*, 1991). Increasing lower energy threshold may not be practical, however, with scintillators having low energy resolution.

Reconstruction of true volumetric data requires special algorithms that are capable of handling truncated projections and varying axial acceptance angle within the FOV (Colsher, 1980). The reprojection method proposed in (Kinahan and Rogers, 1989) is a popular extension of FBP to 3-D imaging. However, a combination of data rebinning reducing the 3-D data to planar sets followed by 2-D reconstruction algorithms has gained most recognition to date as a practical approach to 3-D image reconstruction for whole-body tomographs (Defrise *et al.*, 1997; Comtat *et al.*, 1998). A limitation of this latter approach is the possible discrepancy of photon detection models used during rebinning and reconstruction. Line integrals are exploited during rebinning whereas integrals defined on tubes may be used with 2-D iterative image reconstruction.

The fully 3-D iterative reconstruction for state-of-the-art whole-body scanners has been a daunting task until recently. However, the constant progress in computing hardware gradually helps bridge the gap between the algorithm requirements and the computing platform capabilities. In the case of smaller animal scanners, fully 3-D Bayesian iterative reconstruction has been successfully demonstrated (Qi *et al.*, 1998a). The same method has been adapted for ECAT EXACT HR+ whole-body PET scanner avoiding tremendous memory requirements by exploiting the symmetries present in the system matrix (Qi *et al.*, 1998b).

Chapter 3. Research Questions

The problem of image reconstruction for computerized tomography, and for PET in particular, has had extensive exposure to date, but the research in devising practical and robust algorithms and techniques is far from being over. The research is multifaceted and extends in various directions as we have observed in chapter 2. A broad picture of current approaches to the PET inverse problem solution has been presented but is not meant to be exhaustive. Many interesting details are scattered in the applied literature covering image reconstruction and restoration techniques tuned for certain imaging modalities as well as in the special mathematical literature on inverse problems and numerical analysis, since the state-of-the-art in applications stems from the same mathematical concepts.

We have identified several areas that lacked significant details or possessed the potential that has not been developed to date. These observations yielded related questions building sequentially upon our findings. The points of interest that led to the new developments presented in this dissertation are critically discussed below.

3.1. System matrix

It is a common knowledge that the statistical iterative reconstruction techniques have exhibited the greatest potential to date. Advocating the superiority of statistical foundation of the PET inverse problem, Leahy and Qi say: "While the analytic* approaches result in fast reconstruction algorithms, accuracy of the reconstructed images

* The term "analytic" refers to deterministic transform methods in this context.

is limited by the approximations implicit in the line integral model on which the reconstruction formulae are based. In contrast, the statistical methods ... can adopt arbitrarily accurate models for the mapping between the source volume and the sinograms." (Leahy and Qi, 2000) The statement is appealing but success of the mentioned methods is based on several details that are implicit and are not obvious at first sight.

It is clear that the statistical inverse problem model could be preferable, especially when the counting statistics is the dominant source of data errors. This situation is present at the start of the data acquisition and lasts for a prolonged period of time until the properties of the projection data may be safely approximated by the normal distribution (recall section 1.2.5). However, the superior theoretical basis may or may not be translated into significant advantages when applying statistical methods. The detail that is often overlooked is the "arbitrarily accurate model for the mapping between the source volume and the sinogram". This mapping given by the system matrix is preferably as accurate as possible. "How accurate is enough" and "what are the inherent trade-offs" are the questions that we were not able to find documented answers to.

System modeling is a broad subject that, in principle, should account for every known phenomenon related to data acquisition. The starting point can be found in several papers. The relationship between an X-ray beam and image pixels that yields a matrix of weights is described in (Gordon, 1974) and was used with ART reconstructions. Same ideas were exploited in (Shepp and Vardi, 1982) when deriving a matrix of probabilities for the ML-EM algorithm. Shepp and Vardi mentioned that precomputing the matrix would be an ideal solution but thought that it was "somewhat

inconvenient", which is not surprising taking into account the state-of-the-art in computing for 1982. Instead, they calculated online the width of the intersection of the circle inscribed into a pixel and the strip defined by a given TOR. They also claimed that the choice of weights was not critical.

There are reports that elaborated on the system response modeling for high resolution PET (Lecomte *et al.*, 1984; Schmitt *et al.*, 1988). It was found that the point spread function is asymmetric across the FOV for a system with small discrete detectors and this effect can be approximately described analytically. Huesman and colleagues proposed a data correction procedure involving inversion of the "blurring" matrix to account for crystal penetration before applying the FBP algorithm (Huesman *et al.*, 1989). While accounting for the effect of crystal penetration, this approach resulted in noisier images. The sinogram correction techniques yielding redistribution of counts across the projection are popular and may be considered the standard approach (Hoffman *et al.*, 1989; Liang, 1994). This is the only way to prepare the projection data for the transform algorithms, but the same approach is practised with other techniques that allow, in principle, for more sophisticated system modeling methods, which could improve the quality of reconstruction results (Terstegge *et al.*, 1997). One solution was proposed in (Brix *et al.*, 1997) where a blurring kernel was used to account for the actual measuring process. The detector response was assumed variant radially but invariant tangentially, moreover the model was included as a part of forward projection only, leaving some ambiguity in the issue of optimality of the proposed procedure.

It has been noted in (Llacer and Veklerov, 1989) that if the transition matrix is not accurate enough, the image estimation process appears to bypass the feasibility region.

Despite several hints as to the possible optimal modeling procedure, a practical yet accurate approach to model tubes and the system response at the same time was, nevertheless, not found in the literature. This brought the first issue that had to be clarified:

- **How should the transition matrix be constructed in order to reflect the PET scanner properties adequately and to be used in a statistical iterative image reconstruction efficiently?**

3.2. Stopping rule

The natural and theoretically solid choice among the statistical iterative image reconstruction methods has been the ML-EM algorithm. Despite long computations involved, it is convergent and is guaranteed to yield the ML solution. There are several ways to remedy the inherent noise artefact as we mentioned in section 2.4.4. Terminating iterations with a stopping rule seems an acceptable solution (Gooley and Barrett, 1992) that would not result in the potential loss of resolution, which is hard to escape with Bayesian regularisation and would yield less iterations than would otherwise be performed, i.e. it would potentially require less computations. However, the stopping rules that are based on statistical hypothesis testing (Veklerov and Llacer, 1987; Llacer and Veklerov, 1989; Hebert, 1990) might not find a solution given real data sets. Promising results have been obtained with the procedure described by Coakley who used the concept of cross-validation (CV) (Coakley, 1991). Successful applications of this procedure have been reported in (Llacer *et al.*, 1993; Falcón *et al.*, 1998). The CV procedure of Coakley is rather time consuming, which can potentially limit its practical

usefulness. It is known that hundreds of iterations have to be performed in order to recover an image of fine structure with good accuracy. High accuracy is possible only when many coincidences have been acquired. Otherwise the early onset of the noise artefact prevents from getting a smooth looking image without explicit regularisation. Thus, there is an unknown dependence of the feasibility region on total registered counts. Uncovering the properties of feasible images obtained with a stopping rule may be essential for handling dynamic series of images as well. These considerations triggered the second issue explored in this dissertation:

- **Are there ways to accelerate and simplify the CV stopping rule and make it more appealing and robust for routine application? What are the consequences of applying a stopping rule, especially in the case of dynamic PET series that are assessed visually yet subject to quantitative analysis?**

3.3. SVD potential

The modeling of the relationship between the image representation space and the space of projections is the cornerstone of many image reconstruction algorithms. A technique of computing the system matrix using analytical expressions is a valuable tool. We have observed in section 1.6 that matrix SVD is useful in diagnosing limitations inherent to a given inverse problem. This certainly is an asset especially if an adequately accurate projection model is built into the system matrix and the matrix SVD is available. Moreover, image estimates may be obtained with a regularised pseudo-inverse of the system matrix. The regularisation, which is certainly a major issue with any

reconstruction technique, is somewhat simpler with SVD since it may be limited to singular value spectrum modification in a systematic manner.

There are few reports of practical image reconstruction in CT exploiting SVD as we observed in section 2.4.2. The lack of mainstream interest is mainly due to the "prohibitively" large size of the system matrix. SVD is an extremely useful tool of linear algebra and matrix analysis. Robust algorithms to compute SVD of an arbitrary matrix are available (Press *et al.*, 1992). The regularisation optimality, however, is a subject of research currently pursued in many SVD applications. This brought us to the third set of questions considered in this manuscript:

- **Is practical PET image reconstruction based on the SVD of the system matrix currently feasible? How should the optimal truncation level be derived if TSVD regularisation is applied?**

3.4. Accurate image reconstruction in real time

The fourth and final issue on which this dissertation makes a contribution is derived from our success in solving the previous issue. The severe ill conditioning of the inverse problem arising from the geometry of conventional PET set-up has been isolated by the systematic TSVD regularisation. The TSVD image reconstruction which is an attractive alternative to Fourier transform based methods and which is already fast as it is, lends itself to a potentially faster implementation, that may utilise the availability of the list-mode data. Thus, we propose a procedure that allows

- **PET image reconstruction of the list-mode data potentially in real time! (Canadian patent pending)**

Chapter 4. Results

The research results and the data analysis methods as well as the immediate discussion of the issues explored are given in the following articles included as an integral part of the present dissertation:

[4.1] Selivanov V V, Picard Y, Cadorette J, Rodrigue S, and Lecomte R 2000 Detector response models for statistical iterative image reconstruction in high resolution PET *IEEE Trans. Nucl. Sci.* **47** 1168-75

[4.2] Selivanov V V, Lapointe D, Bentourkia M, and Lecomte R 2001 Cross-validation stopping rule for ML-EM reconstruction of dynamic PET series: Effect on image quality and quantitative accuracy *IEEE Trans. Nucl. Sci.* **48** 883-9

[4.3] Selivanov V V and Lecomte R 2001 Fast PET image reconstruction based on SVD decomposition of the system matrix *IEEE Trans. Nucl. Sci.* **48** 761-7

[4.4] Selivanov V V and Lecomte R 2001 Real-time PET image reconstruction based on regularized pseudo-inverse of the system matrix (to be submitted in revised form to *IEEE Trans. Med. Imag.*)

Later in the text, except within the articles themselves, these papers are referred by their respective numbers in square brackets.

Detector Response Models for Statistical Iterative Image Reconstruction in High Resolution PET

by

Vitali V. Selivanov, Yani Picard, Jules Cadorette,

Serge Rodrigue, and Roger Lecomte

© 2000 IEEE. Reprinted, with permission, from

IEEE TRANSACTIONS ON NUCLEAR SCIENCE

Vol. 47, No. 3, June 2000

Detector Response Models for Statistical Iterative Image Reconstruction in High Resolution PET

V. V. Selivanov, *Student Member, IEEE*, Y. Picard, J. Cadorette, S. Rodrigue, and R. Lecomte, *Member, IEEE*

Metabolic and Functional Imaging Center, Dept. of Nuclear Medicine and Radiobiology, Université de Sherbrooke, Sherbrooke, Québec, Canada

ABSTRACT

One limitation in a practical implementation of statistical iterative image reconstruction is to compute a transition matrix accurately modeling the relationship between projection and image spaces. Detector response function (DRF) in positron emission tomography (PET) is broad and spatially-variant, leading to large transition matrices taking too much space to store. In this work, we investigate the effect of simpler DRF models on image quality in maximum likelihood expectation maximization reconstruction. We studied 6 cases of modeling projection/image relationship: tube/pixel geometric overlap with tubes centered on detector face; same as previous with tubes centered on DRF maximum; two different fixed-width Gaussian functions centered on DRF maximum weighing tube/pixel overlap; same as previous with a Gaussian of the same spectral resolution as DRF; analytic DRF based on linear attenuation of γ-rays in detector arrays weighing tube/pixel overlap. We found that DRF oversimplification may affect visual image quality and image quantification dramatically, including artefact generation. We showed that analytic DRF yielded images of excellent quality for a small animal PET system with long, narrow detectors and generated a transition matrix for 2-D reconstruction that could be easily fitted into the memory of current stand-alone computers.

I. INTRODUCTION

An issue that needs careful consideration in a practical implementation of iterative image reconstruction methods is the construction of a transition matrix that accurately models the relationship between projection and image spaces. In high resolution positron emission tomography (PET) where long and narrow detectors are used, the effect of inter-detector penetration is of great importance and leads to spatially-variant coincident detector response functions (DRF) [1]. Spatially-variant DRFs have been used in filtered backprojection image reconstruction to model [2,3] and to restore [4,5] the point spread function of PET scanners. In statistical iterative image reconstruction, DRFs should be incorporated into the computation of the transition matrix rather than be applied as a pre-correction of projection data in order to preserve the Poisson nature of input data. To date Monte Carlo simulation has been the major means to accurately model DRF spatial dependence [6]. However, by including more physical factors into such a simulation, DRFs are made broader and more complex, resulting in a much larger transition matrix. Thus, simple DRF models are commonly being used in derivation of the transition matrix [7-9], but the consequences of the mismatch between these simple DRF models and the real detector response are poorly understood and have received little attention up to date [10,11]. In this study, we used analytic DRF models instead of the Monte Carlo approach for matrix computation to compare effects of different DRF approximations and, thus, of different transition matrices on reconstructed image quality. A model suitable for a 2-D statistical iterative image reconstruction is proposed. It can be applied with any iterative method using a matrix of weights to state the relationship between measured projections and an image being reconstructed.

II. MATERIALS AND METHODS

A. PET scanner

The Sherbrooke animal PET scanner was used to acquire data. The system consists of 2 ring layers of 3×5×20 mm BGO crystals, 256 per layer, individually coupled to avalanche photodiodes. Data can be acquired in either stationary (without "clam-shell" motion) or double sampling mode (with "clam-shell" motion). Without "clam-shell" motion, the set of parallel tubes-of-response between opposing detectors defines a projection of 32 bins at 256 angles. With the double sampling, 257 angles (with the imaginary detector in the clam opening space) of 32-bin projections are added to the tubes-of-response set. Detailed description of the scanner performance characteristics can be found in [12,13].

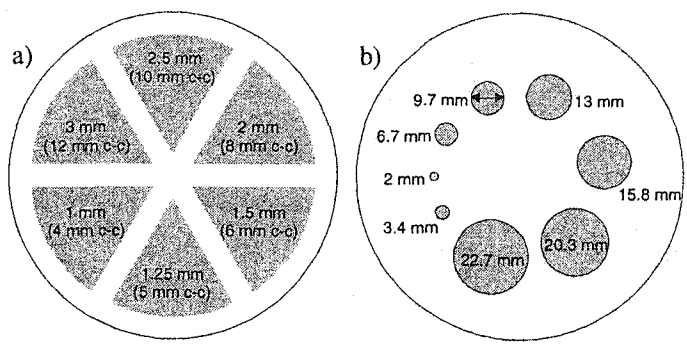

Figure 1: Phantoms used to acquire data with the Sherbrooke animal PET scanner: a) resolution phantom; b) contrast phantom.

0018-9499/00$10.00 © 2000 IEEE

Table 1
Notations used in the paper

Symbol	Definition
d	tube-of-response index
b	image pixel index
$n^*(d)$	number of coincidences detected in the tube-of-response d
$p(b,d)$	transition matrix elements: probability of an annihilation event in pixel b being registered in the tube-of-response d
$p_{corr}(b,d)$	probability $p(b,d)$ corrected for the attenuation and relative detector efficiency in the tube-of-response d
$\lambda_k(b)$	estimate of the mean number of annihilations in pixel b after iteration k
$A(d)$	attenuation correction factor in the tube-of-response d
$E(d)$	detector efficiency factor in the tube-of-response d
$I(d)$	binary function: *1*, if the tube d is reliable; *0*, if the tube d is considered unreliable due to an abnormal detector efficiency (outside of a given value interval)
r_1, r_2	distance of the point for which the DRF is estimated from detectors #1 and #2 of a given detector pair
θ	photon incidence angle (assumed to be the same within a given tube-of-response)
M_0	width of an intrinsic aperture function at a given angle θ
X	point of the tube-of-response cross-section for which the coincident detector response is estimated
$e_0(\theta)$	intrinsic detection efficiency of a detector embedded in an array when irradiated at an angle θ
ε	eccentricity in the position of the source relative to the coincident detectors
$g(x;\theta)$	detector intrinsic aperture function calculated from linear attenuation of γ-rays in the detector array
$f(x,\varepsilon;\theta)$	analytic coincidence detector response function
$\sigma(\theta)$	standard deviation of a Gaussian function at angle θ
$R(\theta)$	spectral resolution of DRF model at angle θ
N_i	total number of pixels in the i-th hot spot of the contrast phantom
M_i	mean value within the i-th hot spot of the contrast phantom
M_{max}	mean value in the hot spot of the contrast phantom having the largest count density

B. Phantoms

PET data were acquired with two phantoms. The first one was a *resolution phantom* made of a 110 mm diameter Lucite cylinder having six sections with holes of diameter 1, 1.25, 1.5, 2, 2.5, 3 mm at a distance of 4 times the hole diameter center-to-center (*Figure 1a*). A total of 5.37 million coincidence events were recorded in double sampling mode with this phantom for the two planes used in reconstruction (3.03 million for the upper plane and 2.34 million for the lower one). This phantom was used for visual inspection of reconstructed images.

The second one was a *contrast phantom* of 110 mm diameter also made of Lucite and having holes of diameter 2, 3.4, 6.7, 9.7, 13, 15.8, 20.3, 22.7 mm located on a circumference at a distance of 28 mm from the center (*Figure 1b*). A total of 1.78 million coincidence events were recorded in stationary mode with this phantom for the two planes (0.95 million for the upper plane and 0.83 million for the lower one).

C. Reconstruction Method

Images were reconstructed on a grid of 128×128 pixels (0.95-mm pixel side) with maximum likelihood expectation maximization (ML-EM) method [14]. Iterative equation with corrections for detector efficiency and attenuation was the following:

$$\lambda_{k+1}(b)=\frac{\lambda_k(b)}{\sum_d p_{corr}(b,d)}\sum_d \frac{n^*(d)\,p_{corr}(b,d)}{\sum_{b'}\lambda_k(b')\,p_{corr}(b',d)} \qquad (1)$$

where

$$p_{corr}(b,d)=I(d)\frac{p(b,d)}{A(d)E(d)} \qquad (2)$$

See *Table 1* for the explanation of notation used in the equations.

D. Detector Response Function

Analytic DRF approximation introduced in [1,2] was used. The following equation defines detector response in any point of the field-of-view (FOV), as illustrated in *Fig. 2*:

$$f(X,\varepsilon,\theta)=\frac{(1+\varepsilon)^2}{\varepsilon M_0^2 e_0^2}\int_x g(x;\theta)\times g\big((1+\varepsilon)X-\varepsilon x;\theta\big)dx \qquad (3)$$

where

$$e_0(\theta)=\frac{1}{M_0}\int_x g(x;\theta)dx \qquad (4)$$

$$\varepsilon=\begin{cases} r_1/r_2, & r_1\le r_2 \\ r_2/r_1, & r_1>r_2 \end{cases} \qquad (5)$$

In order to check accuracy of the analytic DRF, experimental measurement of the in-plane DRFs was performed. A line source of ^{22}Na having an active diameter of 0.84 mm was placed axially in the FOV and moved radially from the center to the edge of the FOV in steps of 0.2185 mm to yield a total of 350 samples covering slightly more than one FOV radius. Complete projection data were acquired at each source position. Data for all detector pairs orthogonal to the radial line along which the source moved were plotted as a function of distance from FOV center. These are referred to as the experimental DRFs.

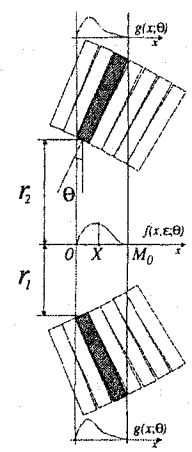

Figure 2: Geometric model used for the calculation of the analytic DRF (eq. 3).

E. Transition Matrix

Several transition matrix modifications were derived as follows. Contribution of a given pixel into a tube-of-response was computed as a sum of linear integrals of DRF

approximation along the tube over the pixel, at a distance of 0.1 mm. It was, therefore, proportional to the pixel and the tube intersection area weighed by a function approximating DRF. The contributions of a given pixel to all tubes-of-response were normalized to yield probabilities $p(b,d)$. Six different DRF models of increasing complexity were investigated:

IA-c: intersection area of a pixel and a tube-of-response without any weighing, tubes were set to a fixed width of 3 mm (equal to a detector width) and aligned with the faces of coincident detectors;

IA-p: same as *Case IA-c*, but tubes were aligned with the corresponding analytic DRF peak;

G-3: same as *Case IA-p*, but a depth-invariant Gaussian function with 3 mm full-width at half-maximum (FWHM) was used to weigh pixels within a tube; tube width was set equal to the Gaussian's full-width at tenth-maximum (FWTM);

G-2: same as *Case G-3*, but the FWHM of a Gaussian function was set to 2 mm (equal to the scanner intrinsic resolution in the FOV center);

G-var: same as *Case IA-p*, but a Gaussian function of the same spectral resolution [1] as the corresponding analytic DRF with ε=1 was used to weigh pixels within a tube, i.e. Gaussian resolution varied radially but was constant tangentially

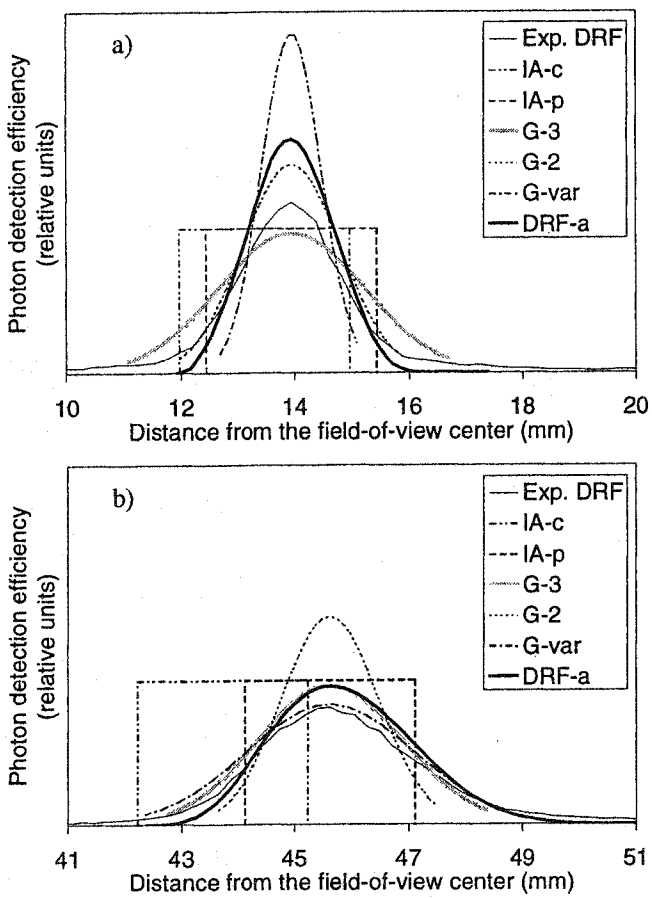

Figure 3: Two examples of six different DRF models in comparison to the experimental DRF when ε=1: a) θ=4.9°; b) θ=16.2°.

$$R(\theta)=\frac{\left[\int_x f(x,1;\theta)dx\right]^2}{\int_x f(x,1;\theta)^2 dx}=2\sigma(\theta)\sqrt{\pi} \qquad (6)$$

tube width was set equal to the Gaussian's FWTM;

DRF-a: analytic DRF was used to weigh intersection area of a pixel and a tube-of-response; tubes of variable width extend over the entire projected width of detectors (see *Figure 2*).

Two examples of functions used to approximate detector response when r_1=r_2 and photon incidence angle θ=4.9° and θ=16.2° are presented in comparison to the experimental DRF in *Figure 3*.

The transition matrices were pre-computed and stored in files. Each matrix element was encoded using 6 bytes (2 bytes for the column index and 4 bytes for the value). Matrix size for each case and the two sampling modes of the Sherbrooke animal PET scanner is given in *Table 2*. During reconstruction matrix elements were corrected for detector efficiency, including preventing unreliable tubes of response from taking part in reconstruction. Finally, the matrices were corrected for attenuation in a cylindrical water phantom using analytical attenuation coefficients (0.098 cm^{-1}). Scatter and random coincidences were ignored.

Table 2
Transition matrix size (non-zero elements only), MB

Case	Stationary mode	Double sampling mode
IA-c	24.45	48.98
IA-p	19.74	39.52
G-3	32.10	64.29
G-2	23.35	46.76
G-var	23.91	47.81
DRF-a	33.44	66.92

F. Stopping Rule

Since the goal of this study was to show the dependence of reconstruction results on DRF, it was chosen to use a reconstruction procedure with as little assumptions as possible. In particular, regularization (through penalty or priors) which may have acted as a low-pass filter smoothing the image and masking artefacts was avoided. Rather, two series of image reconstruction were performed that used different strategies to stop iterations. The cross-validation procedure [8,15] was used in the first one (Section III-A). Two adjacent planes acquired simultaneously were reconstructed as two independent data sets drawn from the same radioactivity distribution. This could be done given the axial structure of the phantom. Poisson nature of the measured samples (the two planes) was preserved, since data from the PET scanner were not pre-corrected.

Reconstructing one of the two samples, log-likelihood of the image estimate was checked, as if another sample was the source of the image. Iterations were terminated for this plane, if the log-likelihood decreased after the current iteration. This ensured that both samples of projections could be drawn from the same map of emission intensities given the same probability distribution of photon counting statistics. The two image estimates were added to obtain the final image.

Only the contrast phantom was reconstructed in the second series (Section III-B). A region-of-interest (ROI) was placed over the largest hot spot of the phantom and standard deviation was evaluated within the ROI after each iteration. Iterations were terminated when the same level of noise was reached. The two planes mentioned above were reconstructed separately and the two resulting image estimates were also summed to yield the final image. The two planes were used in order to assess the results of reconstruction on exactly the same data as in the first series. Sinograms for the two planes were not summed into a single data set in order not to compromise the detector efficiency correction step, since it is different for the two planes. In the third series (not reported here), a fixed number of iterations was also used with every DRF model for a given phantom [16].

G. Image Evaluation

Quantitative estimates of relative recovery factor RF_i and relative standard deviation STD_i were calculated from the reconstructed images of the contrast phantom according to the formulae:

$$RF_i = 100 \times \frac{M_i}{M_{\max}} \tag{7}$$

$$STD_i = 100 \times \frac{1}{M_i} \times \sqrt{\frac{\sum_{b \in ROI_i} (\lambda_k(b) - M_i)^2}{N_i - 1}} \tag{8}$$

See *Table 1* for the explanation of notation used in the equations. Values of N_i (total number of pixels in the i-th hot spot of the contrast phantom) were the following, for different hot spot size: 4, 9, 37, 69, 137, 225, 349 and 421.

III. Results

A. First Reconstruction Series

In this series two coplanar data sets were reconstructed independently and the stopping point for ML-EM iterations was defined by the cross-validation procedure [8,15]. Versions of ML-EM exploiting different transition matrices performed different number of iterations (see *Table 3*). Resolution phantom needed almost twice the number of iterations with the transition matrix of *Case G-3* than the other cases to satisfy the stopping rule. Therefore, reconstruction time was longer due to larger number of iteration and because matrix size was one of the largest in this case.

Use of different matrices, i.e. DRF models, yielded visible differences in reconstructed images. Resolution phantom images are presented in *Figure 4*. Note, that the one reconstructed with the matrix of case *IA-c* showed the poorest rendering of radioactivity distribution due to activity shift in the image towards the FOV center introduced by the tube-of-response definition (see *Figure 3*). Round spots of the phantom appeared of irregular shape in the image (*Figure 4, IA-c*). The region containing spots of 1-mm diameter was distorted and did not reproduce phantom structure (*Figure 5, IA-c*). Tube alignment with the DRF maximum helped to produce a more realistic image (*Figure 4, IA-p*). But a rectangular DRF approximation resulted in non-round hot spots. Using the 3 mm FWHM Gaussian DRF model, round spots are obtained in the image, but of less diameter than that expected from the phantom dimensions (*Figure 4, G-3*). The stopping rule allowed significantly more iterations for the resolution phantom image when the transition matrix of case *G-3* was used (see *Table 3*). The recovery of the size of large spots and low noise at the same time are remarkable. But the region with the finest details appeared very close to the background relative to the peak values found in the image (*Figure 5, G-3*) and is, therefore, almost invisible on a standard gray-scale image (*Figure 4, G-3*). Image processing allowed visualization of the 1-mm spots, but the structure of this phantom region was not reproduced, some spots were washed out during iterations. Use of the 2 mm FWHM Gaussian yielded an artefact visible near the FOV periphery (*Figure 4, G-2*). Spots of 2 mm in diameter and more appeared elongated towards the center. The shape of these spots resembled that observed in case *IA-p* (*Figure 4, IA-p*). Similar for these two cases was the artificial reduction of the radial extent of the peripheral tubes-of-response introduced by the DRF approximations (see *Figure 3*). No artefact (found in resolution phantom images of cases *IA-p* and *G-2*) was present in the reconstructed image when the Gaussian model took into account radial resolution change (*Figure 4, G-var*). But noise on the background appeared very pronounced in this case. Apparently, it was the result of narrowing artificially the central tubes-of-response, hence the

Table 3
Number of iterations suggested by the cross-validation stopping rule

Case	Contrast phantom		Resolution phantom	
	Plane #1	Plane #2	Plane #1	Plane #2
IA-c	163	187	559	578
IA-p	103	87	491	446
G-3	155	148	841	803
G-2	76	66	382	334
G-var	141	120	365	407
DRF-a	144	129	387	393

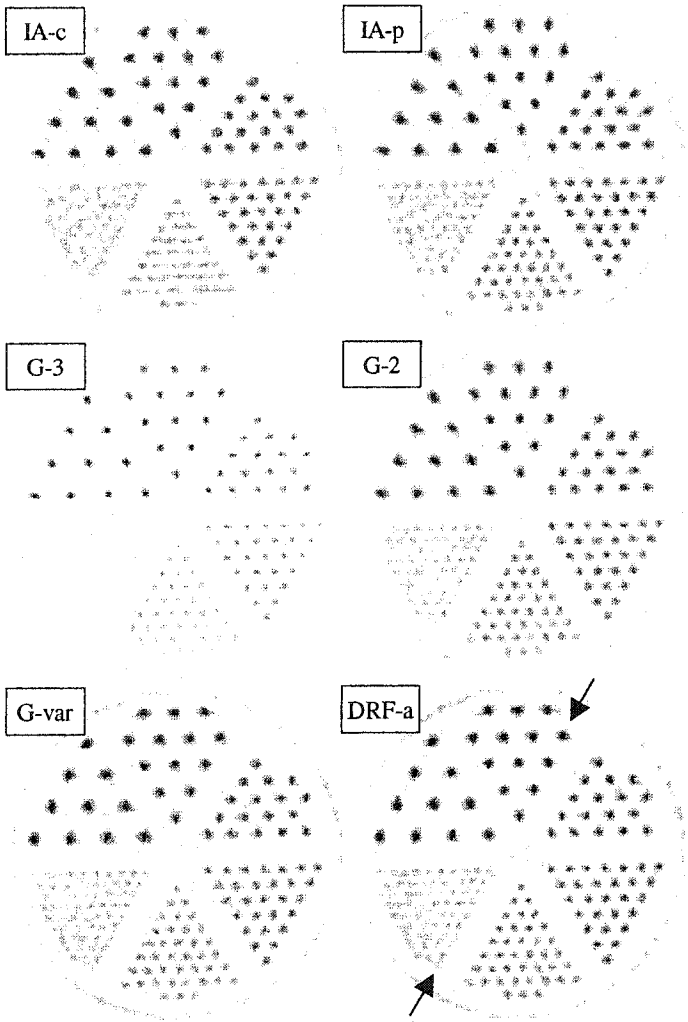

Figure 4: Resolution phantom reconstructed using six different DRF approximations and the cross-validation procedure to stop ML-EM iterations (first reconstruction series); arrows show the line along which the profile presented in Figure 5 was taken.

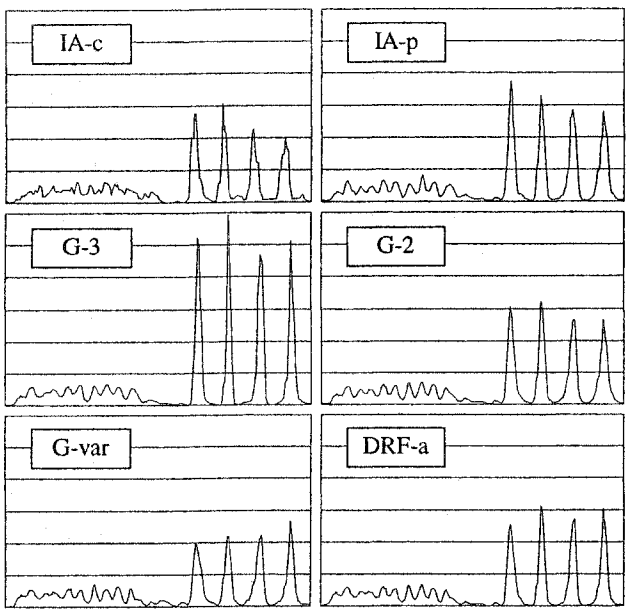

Figure 5: Profiles of the resolution phantom (first reconstruction series); direction marked with arrows in Figure 4.

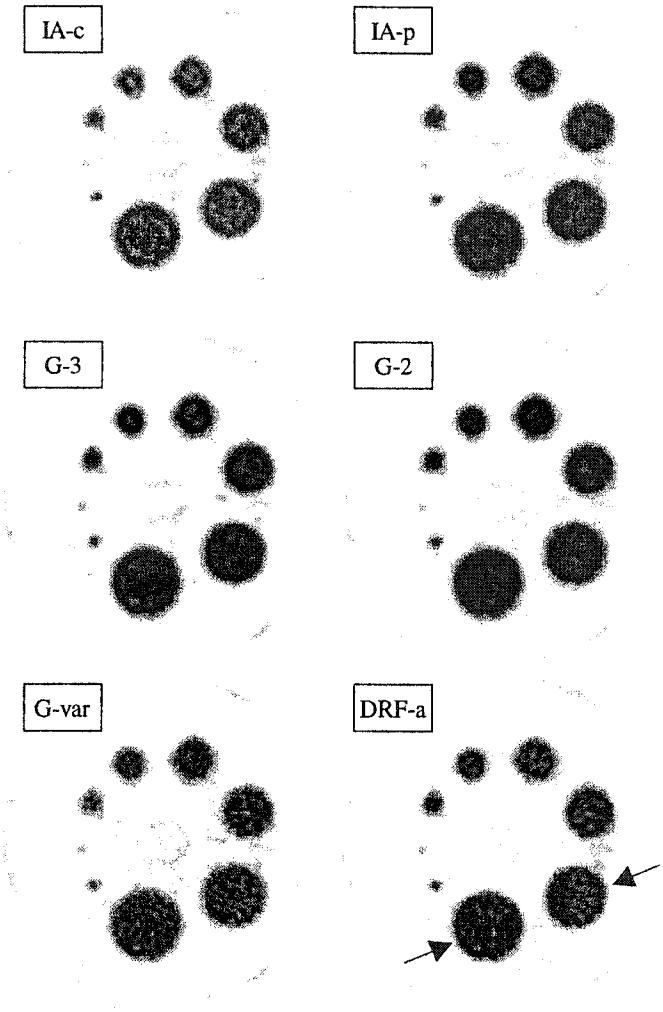

Figure 6: Contrast phantom reconstructed using six different DRF approximations and the cross-validation procedure to stop ML-EM iterations (first reconstruction series); arrows show the line along which the profiles presented in Figures 7 and 11 were taken.

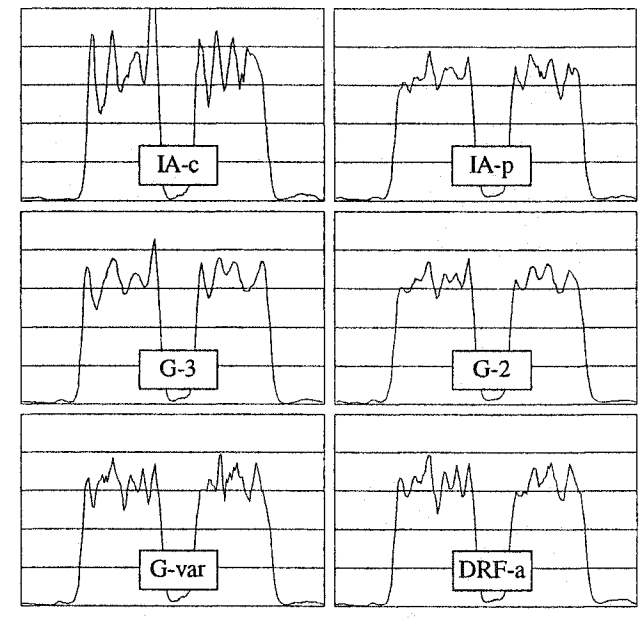

Figure 7: Profiles of the contrast phantom illustrating the edge artefact in Cases IA-c and G-3 (first reconstruction series); direction marked with arrows in Figure 6.

Gaussian approximation from the measured DRF was large (see *Figure 3a*), whereas the Gaussian in the peripheral tubes-of response was a good match of the measured DRF (*Figure 3b*). Finally, the analytic DRF ensured the best match of reconstructed resolution phantom image to the known phantom geometry with reasonably low noise in the image (*Figure 4, DRF-a*).

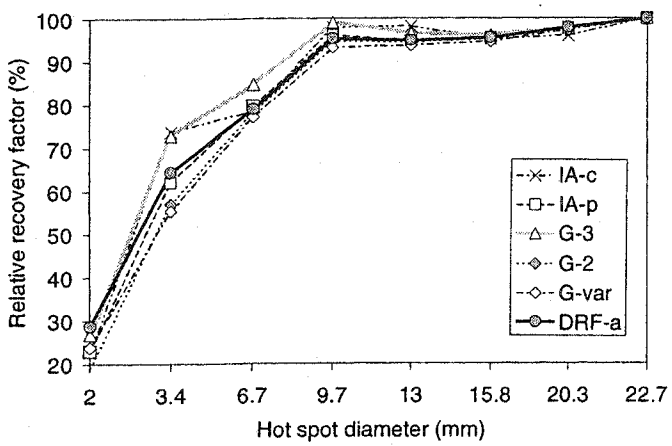

Figure 8: Relative recovery factor as a function of hot spot size, calculated from the contrast phantom images reconstructed using six different DRF models (first reconstruction series).

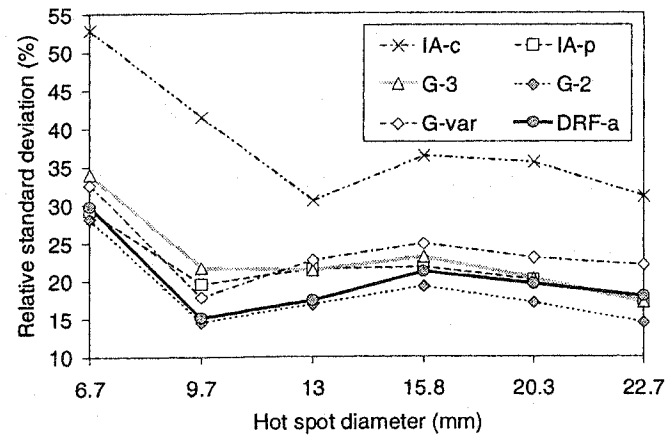

Figure 9: Relative standard deviation as a function of hot spot size calculated from the contrast phantom images reconstructed using six different DRF models (first reconstruction series).

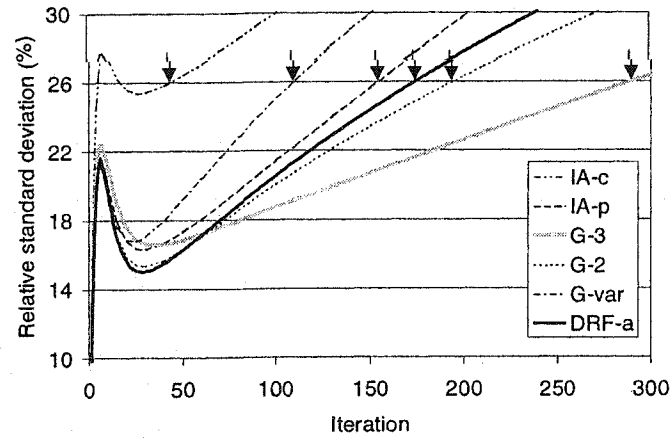

Figure 10: Relative standard deviation within the largest hot region of the contrast phantom as a function of iteration number (second reconstruction series).

Comparing quantitative data drawn from the reconstructed images of the contrast phantom (*Figure 6*), we found that in two cases unstable activity concentration recovery was present due to an artefact which amplified intensity at the edge of a region with uniform radioactivity concentration (*Figures 6,7; IA-c, G-3*). Recovery factor appeared to be a non-monotonic function of the hot spot diameter in these cases (*Figure 8, IA-c* and *G-3*). Reconstructed image of the contrast phantom also exhibited the poorest noise properties when the simplest DRF model was exploited (*Figure 9, IA-c*). The best recovery of small spots was obtained when using analytic DRF (*Figure 8, DRF-a*). Only case *G-2* was superior to case *DRF-a* in producing a smooth image estimate in this series (*Figure 9*).

B. Second Reconstruction Series

Only the contrast phantom was reconstructed in this series. Iterations were stopped at 26% relative standard deviation within the largest hot spot of the phantom (see *Figure 10*). Significantly different number of iterations was performed, ranging from 44 using the DRF approximation of case *IA-c* to 291 with the DRF of case *G-3*. Note that at the local minimum at around 30 iterations, case *IA-c* had much higher noise level than the rest of the cases.

Contrast phantom images in this series also showed the edge artefact in two cases (*Figure 11, IA-c* and *G-3*), though it became less severe in case *IA-c* due to early termination of iterations in the present series. Due to the small number of iterations in this case recovery of small spots was very poor (*Figure 12, IA-c*) as opposed to improved recovery factors

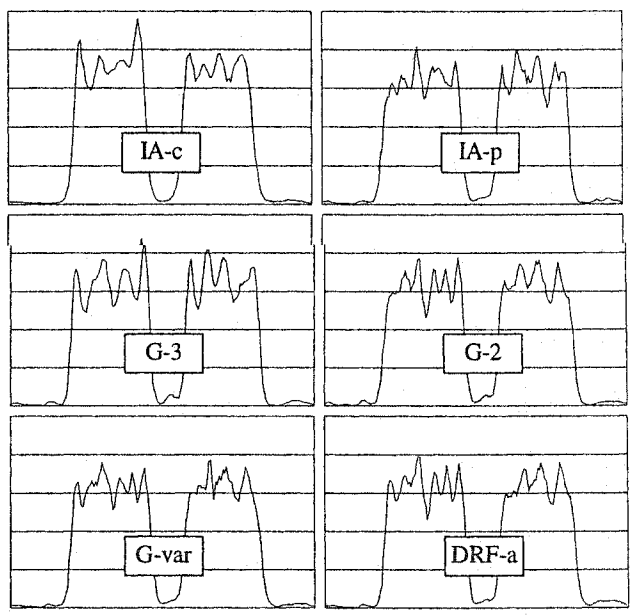

Figure 11: Profiles of the contrast phantom illustrating the edge artefact in Cases IA-c and G-3 (second reconstruction series); direction marked with arrows in Figure 6.

for the small hot regions in case of the 3-mm Gaussian DRF, which was allowed iterating longer (*Figure 12, G-3*). A non-monotonic recovery factor variation with the spot size as in the previous series was observed (*Figure 12, G-3*).

Noisy estimation of small structures using the simple DRF of case *IA-c* was observed when comparing the noise level in the hot spots of different size. The lowest level of noise appeared in cases using the 2-mm Gaussian function and the analytic detector response models (*Figure 13; G-2, DRF-a*).

IV. Discussion

The transition matrix used in statistical iterative image reconstruction should properly include the instrument features, such as resolution variations throughout the field-of-view. Use of measured DRF is very unlikely (at least for our scanner) for two reasons. First, this function extends far beyond the projected width of detectors, hence the resulting matrix size would be impractical. Second, if one tries to use functions measured on a particular tube-of-response set (it's a non-trivial engineering task to measure detector response everywhere in the field-of-view) the small deviations from ideal position in the ring will most probably result in misalignment of the rest of the tubes (unmeasured). *A priori* this will degrade the results of reconstruction if compared to the ones obtained with the perfect ring geometry.

Analytic DRF based on linear attenuation of γ-rays in a detector array is a good approximation of experimental detector response asymmetry while it has a finite base equal to the radial detector projection. Due to this fact the DRF model is very attractive for high precision simulation of the transition matrix that describes a high resolution PET system with long and narrow detectors. We have shown that using this analytic model, artefact free images exhibiting reasonably low noise can be obtained.

It has been shown that the use of an inappropriate detector response model for the transition matrix computation can generate artefacts. Hence close reproduction of radioactivity distribution in the reconstructed images is not possible. This happened when the DRF models of cases *IA-c, IA-p* and *G-2* were used. The edge artefact observed in cases *IA-c* and *G-3* and reported in the literature [17] was previously attributed to fundamental properties of unconstrained maximum likelihood reconstruction. It appears from our study that a more realistic DRF model should be used to cure the problem. The complexity of the DRF model and the computation load arising from that is not an issue at stake if a pre-computed matrix of probabilities is used in reconstruction. Therefore, an adequate attention should be given to a proper description of a PET scanner through the transition matrix.

Presented results were based on two different ways to terminate iterations. The third one was used in an earlier presentation of this work [16], when the same number of

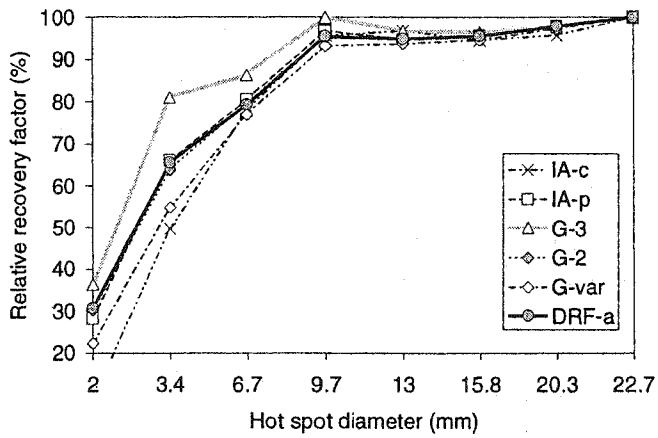

Figure 12: Relative recovery factor as a function of hot spot size calculated from the contrast phantom images reconstructed using six different DRF models (second reconstruction series).

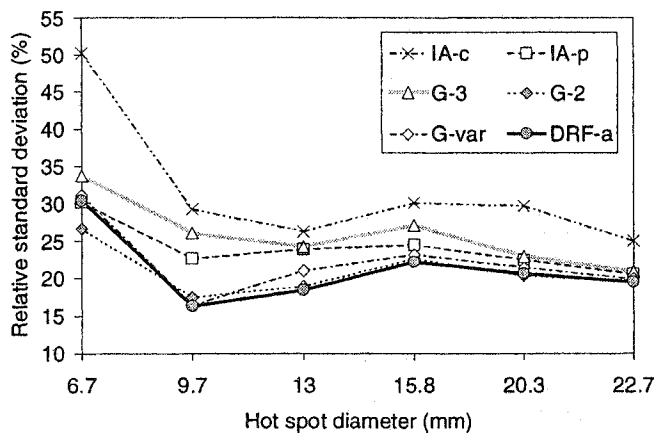

Figure 13: Relative standard deviation as a function of hot spot size calculated from the contrast phantom images reconstructed using six different DRF models (second reconstruction series).

iterations was performed for a given phantom with every DRF model. In practice, explicit regularization can be used to control noise effectively in the form of a prior or penalty function instead of stopping iterations at a specific point. In that case the reconstructed image would be forced to exhibit properties introduced by the prior, among them convergence to a special class of solutions, bias and reduced resolution. Therefore, regularization has been avoided in this study. Otherwise, the effect of a DRF model would have been coupled with that of a prior and would be hard to analyze. Only after ensuring the best resolution recovery through physical modeling should regularization be employed. Then its sole purpose would be to control noise and not to mask artefacts created by an inappropriate DRF model. Remarkably, it was observed that the simplest possible DRF, case IA-c here, which is not uncommon in practical implementations, would inevitably yield images with resolution far poorer than the intrinsic scanner resolution.

It is known that ML-EM is a nonlinear algorithm, i.e. small details are recovered at later iterations as compared to large image structures [18]. DRF model incorporated into

the transition matrix has an impact on the convergence of ML-EM as well. It was observed in [16] that, for the same iteration number, the described analytic DRF model ensured superior contrast and recovery of small structures at the limit of scanner resolution (2 mm), at the same time ensuring the lowest noise level as compared to the other DRF models studied here. In fact, very similar artefacts and relative differences in reconstructed images were observed using the different ways to stop ML-EM iterations, which gives us confidence that the observed results followed from the used DRF model and were fairly insensitive to the stopping points. Needless to say that a proper DRF, such as the analytic model proposed in this work, could be implemented together with a regularization scheme for practical reconstruction of medical images.

V. Conclusion

It has been shown that the use of an oversimplified detector response model in an iterative reconstruction can generate artefacts in reconstructed images. The analytic function based on linear attenuation of γ-rays in a detector array used as a detector response approximation helps to obtain reconstructed images of significantly better quality as compared to conventionally used models. This analytic DRF takes into account resolution variation in the field-of-view and is easy to compute from the geometry and the attenuation properties of the detectors.

References

[1] R. Lecomte, D. Schmitt, G. Lamoureux, "Geometry study of a high resolution PET detection system using small detectors", *IEEE Trans. Nucl. Sci.*, vol.31, no.1, pp.556-561, 1984

[2] D. Schmitt, B. Karuta, C. Carrier, R. Lecomte, "Fast point spread function computation from aperture functions in high-resolution PET", *IEEE Trans. Med. Imag.*, vol.7, no.1, pp.2-12, 1988

[3] B. Karuta, R. Lecomte, "Effect of detector weighting functions on the point spread function of high resolution PET tomographs: A simulation study", *IEEE Trans. Med. Imag.*, vol.11, no.3, pp.379-385, 1992

[4] R. Huesman, E. Salmeron, J. Baker, "Compensation for crystal penetration in high resolution positron emission tomography", *IEEE Trans. Nucl. Sci.*, vol.36, no.1, pp.1100-1107, 1989

[5] Z. Liang, "Detector response restoration in image reconstruction of high resolution positron emission tomography", *IEEE Trans. Med. Imag.*, vol.13, no.2, pp.314-321, 1994

[6] E. Veklerov, J. Llacer, E. J. Hoffman, "MLE reconstruction of a brain phantom using a Monte Carlo transition matrix and a statistical stopping rule", *IEEE Trans. Nucl. Sci.*, vol.35, pp.603-607, 1988

[7] R. E. Carson, Y. Yan, B. Chodkowski, T. K. Yap, M. E. Daube-Witherspoon, "Precision and accuracy of regional radioactivity quantitation using the maximum likelihood EM reconstruction algorithm", *IEEE Trans. Med. Imag.*, vol.13, no.3, pp.526-537, 1994

[8] J. Llacer, E. Veklerov, K. J. Coakley, E. J. Hofman, J. Nunez, "Statistical analysis of maximum likelihood estimator images of human brain FDG PET studies", *IEEE Trans. Med. Imag.*, vol.12, no.2, pp.215-231, 1993

[9] J. Ollinger, A. Goggin, "Maximum likelihood reconstruction in fully 3D PET via the SAGE algorithm", *Proc. IEEE Nucl. Sci. Symp. & Med. Imag. Conf.*, vol.3, pp.1594-1598, 1996

[10] A. Terstegge, S. Weber, H. Herzog, H. W. Müller-Gärtner, H. Halling, "High resolution and better quantification by tube of response modelling in 3D PET reconstruction", *Proc. IEEE Nucl. Sci. Symp. & Med. Imag. Conf.*, vol.3, pp.1603-1607, 1996

[11] J. Qi, R. Leahy, S. Cherry, A. Chatziioannou, T. Farquhar, "High-resolution 3D Bayesian image reconstruction using the microPET small-animal scanner", *Phys. Med. Biol.*, vol.43, pp.1001-1013, 1998

[12] R. Lecomte, J. Cadorette, P. Richard, S. Rodrigue, D. Rouleau., "Design and engineering aspects of a high resolution positron tomograph for small animal imaging", *IEEE Trans. Nucl. Sci.*, vol.41, no.4, pp.1446-1452, 1994

[13] R. Lecomte, J. Cadorette, S. Rodrigue, D. Lapointe, D. Rouleau, M. Bentourkia, R. Yao, P. Msaki, "Initial results from the Sherbrooke avalanche photodiode positron tomograph", *IEEE Trans. Nucl. Sci.*, vol.43, no.3, pp.1952-1957, 1996

[14] L. A. Shepp, Y. Vardi, "Maximum likelihood reconstruction for emission tomography", *IEEE Trans. Med. Imag.*, vol.1, no.2, pp.113-122, 1982

[15] K. J. Coakley, "A cross-validation procedure for stopping the EM algorithm and deconvolution of neutron depth profiling spectra", *IEEE Trans. Nucl. Sci.*, vol.38, no.1, pp.9-15, 1991

[16] V. Selivanov, Y. Picard, J. Cadorette, S. Rodrigue, R. Lecomte, "Detector response models for statistical iterative image reconstruction in high resolution PET", *Proc. IEEE Nucl. Sci. Symp. & Med. Imag. Conf.*, pp.1377-1381, 1998

[17] D. Snyder, M. Miller, L. Thomas Jr., D. Politte, "Noise and edge artifacts in maximum-likelihood reconstructions for emission tomography", *IEEE Trans. Med. Imag.*, vol.6, no.3, pp.228-238, 1987

[18] J. S. Liow, S. C. Strother, "The convergence of object-dependent resolution in maximum likelihood based tomographic image reconstruction", *Phys. Med. Biol.*, vol.38, no.1, pp.55-70, 1993

Cross-Validation Stopping Rule for ML-EM Reconstruction of Dynamic PET Series: Effect on Image Quality and Quantitative Accuracy

by

Vitali V. Selivanov, David Lapointe, M'hamed Bentourkia,

and Roger Lecomte

© 2001 IEEE. Reprinted, with permission, from

IEEE TRANSACTIONS ON NUCLEAR SCIENCE

Vol. 48, No. 3, June 2001

Cross-Validation Stopping Rule for ML-EM Reconstruction of Dynamic PET Series: Effect on Image Quality and Quantitative Accuracy

Vitali V. Selivanov, *Student Member, IEEE*, David Lapointe, M'hamed Bentourkia, *Member, IEEE*, and Roger Lecomte, *Member, IEEE*

***Abstract*—A major shortcoming of the maximum likelihood expectation maximization (ML-EM) method for reconstruction of dynamic positron emission tomography (PET) images is to decide when to stop the iterative process for image frames with largely different statistics and activity distributions. A widespread practice to overcome this problem involves overiteration of an image estimate followed by smoothing. In this paper, we investigate the qualitative and quantitative accuracy of the cross-validation procedure (CV) as a stopping rule, in comparison to overiteration and postfiltering, for the reconstruction of phantom and small animal dynamic ^{18}F-fluorodeoxyglucose PET data acquired in two-dimensional mode. The CV stopping rule ensured visually acceptable image estimates with balanced resolution and noise characteristics. However, quantitative accuracy required some minimum number of counts per image. The effect of the number of ML-EM iterations on time–activity curves and metabolic rates of glucose extracted from image series is discussed. A dependence of the CV defined number of iterations on projection counts was found that simplifies reconstruction and reduces computation time.**

***Index Terms*—Cross validation, image reconstruction, image sequence analysis, iterative methods, maximum likelihood estimation, positron emission tomography (PET), stopping rule.**

I. Introduction

THE USE of maximum likelihood via expectation maximization (ML-EM) image reconstruction [1] is not very popular nowadays due to the common belief that it is intrinsically unstable and inevitably results in images exhibiting high variance. Bayesian framework [2], [3] and other regularizing approaches [4], [5] were studied to avoid excessive noise amplification during the iterative statistical estimation. Criteria were proposed to stop iterations before noise artifact develops [6], [7]. Another concern is the slow convergence of the ML-EM algorithm when compared to the proposed acceleration techniques [8], [9], which results in relatively long reconstruction time. The ordered subset acceleration, abbreviated OS-EM, has become very popular even with the lack of proven convergence for the real-world noisy data [9]. Unfortunately, the performance and quantitative accuracy of these new methods depend on newly introduced parameters (prior, penalty function, number of subsets, or relaxation factor) that require optimization, which still does not guarantee unbiased estimation. Therefore, it is relevant to address the issues of unconstrained likelihood maximization since it has the attractive property of being asymptotically unbiased.

This paper addresses the issue of choosing the number of iterations in ML-EM image reconstruction automatically in order to ensure diagnostically useful image appearance and quantitatively accurate measurements of radiotracer concentration. A widespread practice to overcome this problem involves overiteration of an image estimate and its subsequent smoothing, which eliminates to some extent the need for a stopping rule and reduces the effect of the noise artifact. However, this procedure reduces the contrast and the spatial resolution. In the case of dynamic positron emission tomography (PET) studies, series of frames with significantly different statistics and radiotracer distributions must be reconstructed. It is obvious that an arbitrary number of ML-EM iterations is not optimal; thus, visually acceptable and quantitatively reliable image estimates are hard to obtain at the same time. Yet, it is important not to iterate longer than really necessary for providing suitable image estimates in order to minimize the reconstruction time. In this paper, we investigate the robustness of the cross-validation (CV) procedure [7] as a stopping rule for ML-EM reconstruction of dynamic PET image series.

II. Materials and Methods

A. PET Scanner

PET scans were performed with the Sherbrooke high-resolution animal tomograph. The system consists of two ring layers of 3 × 5 × 20 mm BGO crystals, 256 per layer, individually coupled to avalanche photodiodes. Two direct planes and one cross-plane can be acquired simultaneously. The transaxial field-of-view (FOV) diameter is 118 mm, and reconstructed image resolution in the FOV center is 2.1 mm. A sinogram consists of 32 bin projections at 256 angles in the acquisition mode without sampling motion used in this study. A detailed description of the scanner can be found in [10] and [11].

Manuscript received November 9, 1999; revised February 24, 2001. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada and by the Medical Research Council of Canada under Grant MT-15348. The work of V. V. Selivanov was also supported in part by the Québec Ministry of Education under a scholarship.

V. V. Selivanov, M. Bentourkia, and R. Lecomte are with the Department of Nuclear Medicine and Radiobiology, Université de Sherbrooke, Sherbrooke QC J1H 5N4, Canada (e-mail: vitali@tep.crc.usherb.ca; mohamed@tep.crc.usherb.ca; rlecomte@courrier.usherb.ca).

D. Lapointe was with the Department of Nuclear Medicine and Radiobiology, Université de Sherbrooke, Sherbrooke QC J1H 5N4, Canada. He is now with CADx Medical Systems, Laval QC H7V 4A7, Canada.

Publisher Item Identifier S 0018-9499(01)05071-7.

0018–9499/01$10.00 © 2001 IEEE

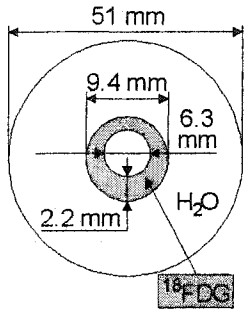

Fig. 1. Phantom modeling the left ventricle of a rat heart used to acquire data for assessing quantitative accuracy of different reconstruction protocols.

B. PET Data

Data were acquired with a phantom modeling the left ventricle of a rat heart (Fig. 1). The asymmetric annular compartment shown in gray in Fig. 1 was filled with a water solution of ^{18}F-fluorodeoxyglucose (FDG). A series of 27 frames with duration from 3 s to 40 min were acquired.

Dynamic PET scan of the heart region was also performed with a normal 300-g rat injected with a bolus of 8.14×10^7 Bq (2.2 mCi) of FDG over 19 s. The following series was acquired for 51 min, starting at the beginning of injection (seconds $\times$ frames): 5×24, 180×8, 300×5.

C. Reconstruction Method

Images were reconstructed with the ML-EM method on a grid of 128×128 pixels with pixel size of 0.95 mm. The following iterative equation was used to compute the image estimate:

$$\hat{\lambda}_{k+1}(b) = \frac{\hat{\lambda}_k(b)}{\sum_{d=1}^{D} p_{corr}(b,\, d)} \sum_{d=1}^{D} \frac{n^*(d) p_{corr}(b,\, d)}{\sum_{b'=1}^{B} \hat{\lambda}_k(b') p_{corr}(b',\, d)}; \quad b = 1, \ldots, B \tag{1}$$

where

$$p_{corr}(b,\, d) = \frac{I(d)}{E(d)}\, p(b,\, d). \tag{2}$$

See Table I for the explanation of notation used in the equations.

Relative detector efficiencies were applied to weigh the transition matrix elements. A model of the spatially variant detector response specific to the scanner was also included into the transition matrix. Analytic detector response functions based on linear attenuation of γ-rays in a detector array were used to weigh intersection area of a pixel and a tube-of-response (TOR). Tubes had a variable width extending over the entire detector projection onto a radial line perpendicular to the respective TOR [12]. Attenuation, scatter, and random events were not included in the model in this study.

The cross-validation procedure was applied using two new data sets $n_1^*(d)$ and $n_2^*(d)$ created by the process of thinning from the measured sample of projections $n^*(d)$ for the two direct planes [13]. The process involved taking each count in each projection bin and assigning it to one or another of the new data halves $n_i^*(d)$, according to the outcome of a random process of equal probability for each of the two outcomes. It resulted in two Poisson distributed data sets, with means that are one-half of the mean of the original set. For the cross-plane, data splitting was not necessary, since the scanner already provides two independent data sets. Reconstructing one of the two samples, e.g., $n_1^*(d)$, log-likelihood $l(\hat{\lambda}_k)$ of the image estimate $\hat{\lambda}_k$ was checked as if another sample $n_2^*(d)$ was the source of the image, i.e.,

$$l(\hat{\lambda}_k) = \sum_{d=1}^{D} \left(-\hat{n}(d) + n_2^*(d) \ln \hat{n}(d) - \ln\left[n_2^*(d)!\right]\right) \tag{3}$$

where forward projection $\hat{n}(d)$ of the image estimate $\hat{\lambda}_k$ for tube-of-response d was

$$\hat{n}(d) = \sum_{b=1}^{B} \hat{\lambda}_k(b) p_{\mathrm{corr}}(b,\, d). \tag{4}$$

TABLE I
NOTATION USED

Symbol	Definition
d	tube-of-response index
b	image pixel index
D	total number of tubes-of-response
B	total number of image pixels
$n^*(d)$	number of coincidences detected in the tube-of-response d
$p(b,d)$	elements of the transition matrix: probability of an annihilation event in pixel b being registered in the tube d
$p_{corr}(b,d)$	probability $p(b,d)$ corrected for the relative detector efficiency [see (2)]
$\hat{\lambda}_k(b)$	estimate of the mean number of annihilations in pixel b after iteration k
$E(d)$	detector efficiency factor in the tube-of-response d
$I(d)$	binary function: *1*, if the detector pair d is reliable; *0*, if it has an abnormal detector efficiency
$n_i^*(d)$	sample of projections after the process of thinning (i=1,2)
$l(\hat{\lambda}_k)$	natural logarithm of the likelihood function for image estimate $\hat{\lambda}_k$ [see (3)]
$\hat{n}(d)$	forward projection of an image estimate for the tube-of-response d [see (4)]
a_i	parameters of a power function found by fitting a curve (i=1,2) [see (5)]
$\hat{\lambda}(b)$	estimate of the value of image pixel b after reconstruction completion
N	total number of pixels in a given ROI of phantom image
c_t	decay corrected count rate in the ROI for phantom image [see (6)]
f_t	normalization factor accounting for F-18 decay and time frame length [see (7)]
$T_{1/2}$	F-18 half-life (in seconds)

The iterative process was terminated for the sample if the log-likelihood started to decrease. This ensured that both samples of projections, $n_1^*(d)$ and $n_2^*(d)$, could have been drawn from the same map of emission intensities given the same probability distribution of photon counting statistics. Log-likelihood analogous to (3) was traced, while reconstructing $n_2^*(d)$, but as if $n_1^*(d)$ was the source of the image. The two image estimates obtained after reconstructing $n_1^*(d)$ and $n_2^*(d)$ were added to yield the final image.

The number of iterations was also computed using an analytic dependence (later referred to as CVF) represented as a power function

$$\text{Iterations} = a_1 \times (\text{Total counts in projections})^{a_2}. \quad (5)$$

Parameters a_1 and a_2 in the above equation were fitted to the CV defined number of iterations with the Levenberg–Marquardt method [14].

Images were also reconstructed with a fixed number of ML-EM iterations. Postreconstruction filtering/smoothing was performed by

1) convolving image estimate with a Gaussian mask of 7 × 7 pixels, 2.2 mm full-width at half-maximum, which is equal to the scanner resolution close to the field-of-view center (GAUSS);
2) filtering with a Butterworth low-pass filter of the first order with cutoff frequency locus at 33% of the frequency range (BW1);
3) filtering with a Butterworth low-pass filter of the first order with cutoff frequency locus at 67% of the frequency range (BW2).

For comparison, images were also reconstructed with filtered backprojection (FBP) using a ramp filter with cutoff frequency satisfying the Nyquist criteria. Detector response model was not included into FBP reconstruction and data were precorrected for detector efficiency.

D. Image Analysis

Two regions-of-interest (ROI) were drawn on images reconstructed with 75 iterations on the basis of phantom dimensions: one covering the hot ring that represented the myocardium (see Fig. 1) and another one centered on the inner region representing the blood pool and covering 50% of it. The decay corrected count rate c_t in the ROI was evaluated as

$$c_t = \frac{1}{f_t} \times \frac{1}{N} \times \sum_{b \in \text{ROI}} \hat{\lambda}(b) \quad (6)$$

where

$$f_t = \int e^{-(\ln 2/T_{1/2}) \times \tau} \, d\tau \quad (7)$$

is a normalization factor accounting for ^{18}F decay and time frame length (in seconds). See Table I for the explanation of notation used in these equations.

Contrast in the phantom images was computed as

$$\frac{\text{Contrast} = c_t(\text{myocardium}) - c_t(\text{blood pool})}{c_t(\text{myocardium}).} \quad (8)$$

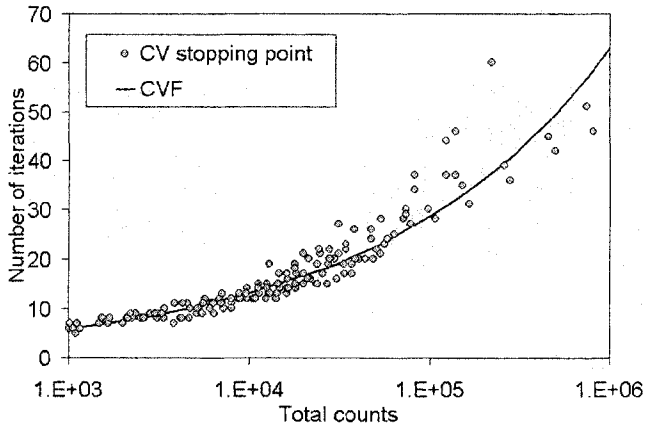

Fig. 2. Number of ML-EM iterations, defined by the automatic CV procedure, as a function of the total counts in phantom projections (three planes together). The fitted power function (CVF) was a reasonably accurate approximation to the CV results.

Rat images were corrected for the radionuclide decay and analyzed using ROIs drawn manually in the last image frame of the series. ROIs were placed on the left ventricular blood pool, as well as on anterior, septal, inferior, and lateral regions of the myocardium. The same ROIs were used after reconstructing images with ML-EM, but ROIs were adjusted for use with FBP images since visible differences in image geometry were observed due to the absence of a detector response model in FBP reconstruction. The input curve was extracted from the blood pool ROI of the respective image series.

The three-compartment FDG model [15] was used to fit every ROI of the heart muscle, and estimates of the rate constants K_1, k_2, k_3, and k_4 were obtained with the Levenberg–Marquardt algorithm [14]. Regional metabolic rate of glucose consumption (rMRGlc) was estimated as follows:

$$r\text{MRGlc} \sim \frac{K_1 k_3}{k_2 + k_3}. \quad (9)$$

III. Results

A. Phantom Study

Visual assessment of images confirmed that the stopping rule invariably ensured low-noise images. But iterations were terminated prematurely for images containing less than 10^5 events, based on the expected level of details in the reconstructed image, i.e., proper thickness of the "myocardium wall" and low estimated radioactivity in the "blood pool." The number of iterations defined by the CV for phantom reconstruction and the fitted power function (CVF case) are presented in Fig. 2.

The decay-corrected count rate c_t within the "myocardial" ROI as a function of the number of events in projection data is given in Fig. 3. Different reconstruction approaches produced different absolute values. A constant value was expected for a given reconstruction method irrespective of the input data statistics. It was found that a fixed number of ML-EM iterations with or without postreconstruction smoothing, as well as FBP, produced more uniform values. Decrease in recovery accuracy was obvious with the stopping rule with low-count data. This

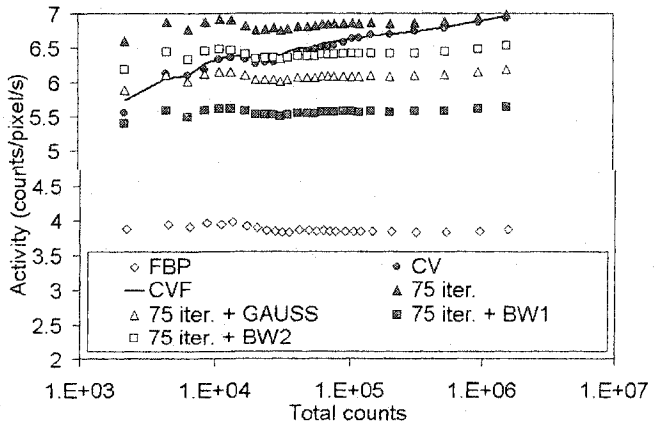

Fig. 3. Decay-corrected count rate c_t in the phantom ROI as a function of the total counts in the projection data (cross-plane only). A constant value was expected for a given reconstruction method irrespective of the input data statistics. Fixed number of ML-EM iterations as well as FBP produced more uniform values.

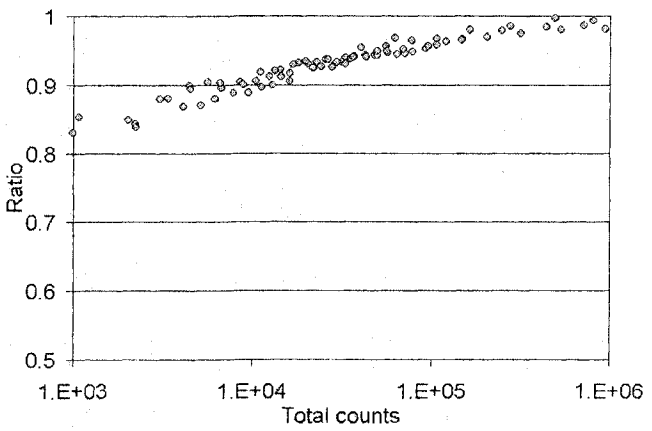

Fig. 4. Ratio of c_t for ML-EM with the cross-validation to c_t for 200 ML-EM iterations (phantom, three planes together). CV stopped iterations to preserve image smoothness, but signal recovery suffered in low-count cases.

trend is illustrated in Fig. 4, where the ratio of c_t after ML-EM with CV to c_t after 200 ML-EM iterations is reported. The ratio presented in Fig. 4 gives a clue about how far from the asymptotic value the cross-validation procedure stopped, since the increase in regional activity was marginal even after 75 iterations. The difference was not statistically significant when more than 10^5 counts were reconstructed with the CV. Curve fitting procedure yielded real numbers that had been rounded to integers to represent iteration numbers. Therefore, rounding to an equal or greater number was performed, yielding marginally better quantitative results in the CVF case. Low estimates of regional activity with FBP may be explained by the lower resolution and by the streak artifact, which captured some counts outside of the ROIs.

The contrast was best with high iteration number. But postreconstruction filtering reduced it significantly (see Fig. 5). Contrast recovery was better with the CV than with overiteration followed by filtering in high-count images and worse with low total counts. FBP images had the poorest contrast.

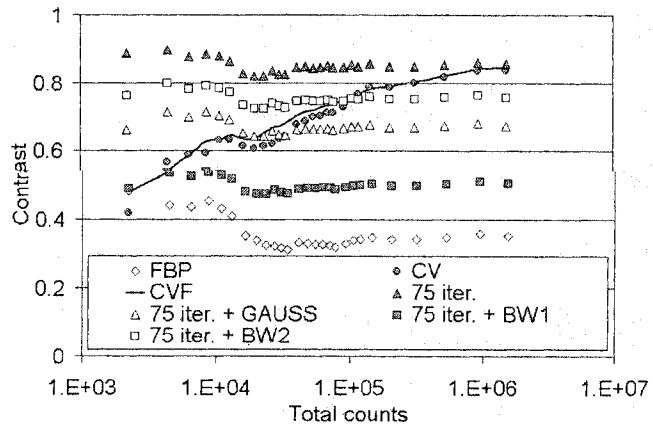

Fig. 5. Phantom image contrast as a function of the total counts in the projection data (cross-plane only). The contrast after 75 iterations was superior. Contrast delivered with CV depended on image statistics.

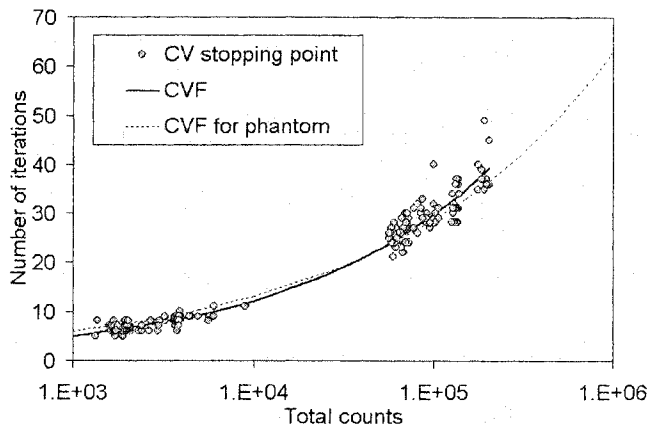

Fig. 6. Dependence of the number of ML-EM iterations, defined by the automatic CV procedure, on the total counts for the rat series (CV stopping point). The power function fitted to the CV stopping points (CVF) is compared to the function derived for the phantom (CVF for phantom), given as well in Fig. 2.

B. Animal Study

Similar dependence of the iteration number on the number of registered events exhibited by the CV stopping rule in phantom reconstruction was observed in the case of dynamic rat series. It is presented in Fig. 6 together with the fitted power function (CVF) and the function fitted previously for the phantom data. With the animal data, FDG was gradually redistributed throughout the regions. The phantom did not model this dynamic aspect. Therefore, a close match of the two functions extracted for the given range of counts from two different reconstructed sequences is remarkable.

Visual image quality was judged upon the contrast and detail resolution versus noise. For a dynamic series, the preferred reconstruction strategy was the one that satisfied those criteria for high- and low-count images. Selected images of the rat series are presented in Fig. 7. The first column gives a low-count image 15–20 s after FDG injection. It corresponds to the peak of radiotracer concentration in plasma. Hence, the two ventricles of the heart are expected to be visible. But since the activity is distributed in the entire vascular system early after injection, lungs

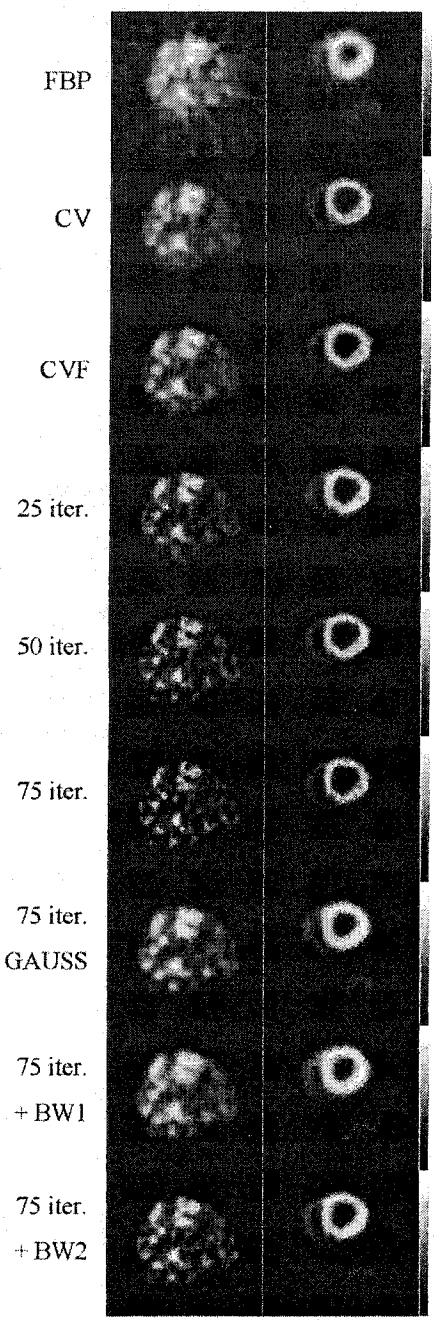

Fig. 7. Selected frames from the dynamic rat scan, reconstructed with different techniques. (Left) 15–20 s after injection (≈ 10 000 events), peak of FDG concentration in the blood plasma, two ventricles of the heart visible. (Right) 46–51 min after injection (≈ 200 000 events), FDG accumulated in the myocardium.

are also visible as a large area of activity below the ventricles in this slice. The second column shows an image at the latest stage of scanning, 50 min after FDG injection, when the radiotracer accumulated in the myocardium.

Very different image quality was observed after different number of iterations and with different postreconstruction filters. Two reconstruction strategies resulted in the most acceptable low- and high-count images as judged visually: the CV procedure and the analytically defined number of iterations (CVF). The rest of the cases yielded either a well-defined

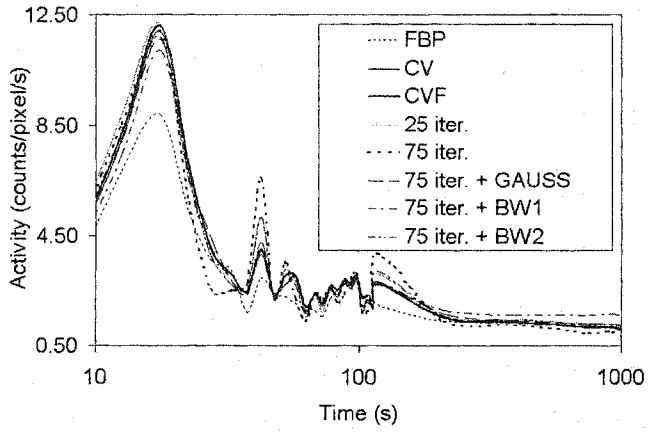

Fig. 8. Early part of the blood pool time–activity curves. Nonfiltered images obtained with 75 iterations exhibit the highest curve variations due to the increased noise in the low-count image frames. Image filtering increased TAC smoothness. CV and CVF yielded reasonably accurate TACs.

blood pool at the beginning of the scan or a good definition of the myocardium with little spillover from the heart muscle to the blood pool, but not both at the same time. Nonfiltered or mildly filtered images showed noisy activity distribution in early frames, whereas heavily filtered images yielded reduced resolution and blurring of structural details in late images. It should be noted that the choice of a filter and its parameters played a more important role than the number of iterations with later high-count frames when a postreconstruction filter was applied, whereas these were equally important with low-count images.

The early part of the left ventricular blood pool time–activity curve (TAC) obtained from the reconstructed rat series is shown in Fig. 8 with the logarithmic time scale for better peak visualization. The late part of the TAC is presented in Fig. 9. Fifty iterations + filter and 75 iterations + filter yielded similar results. Therefore, only the latter was reported. Bounces in the curves are attributed to statistical frame-to-frame variations due to noisy reconstructed images and possible rat movement that may have changed heart position slightly. Nonfiltered images obtained with high iteration number exhibited the lowest spillover from the myocardium to the blood pool later in the series but the highest variations in early frames. Filtering amplified the spillover effect, making the TAC extracted from the blood pool ROI less reliable. An external blood sampling would be necessary to obtain an input function for these images of suboptimal resolution. The cross-validation yielded better TAC than postfiltering, but the activity spillover later in the series was still significant as compared to 75 iteration images. It is interesting that the early TAC part around the peak was not very different for all studied reconstruction cases except FBP.

The various reconstruction schemes not only produced visually different images but yielded significantly different rMRGlc values as well (Fig. 10). The pattern of values across the four myocardial ROIs was somewhat similar in most cases. An interesting finding was that rMRGlc absolute value constantly increased as a function of the number of ML-EM iterations in the tested iteration range. This is illustrated with 25, 50, and 75 iterations in Fig. 10. This fact was a result of improved image reso-

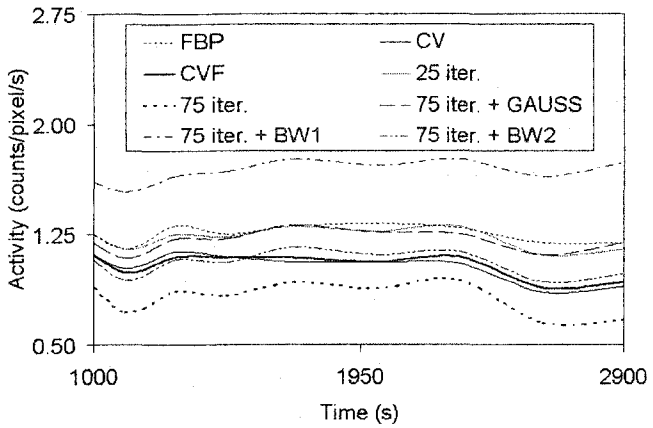

Fig. 9. Late part of the blood pool time–activity curves. Nonfiltered images obtained with 75 iterations exhibited the lowest activity spillover from the myocardium to the blood pool ROI. Filtering amplified the spillover effect making the TAC extracted from the blood pool ROI less reliable. CV and CVF yielded more accurate TACs than 75 iterations + filtering, but the estimated blood pool activity concentration was still higher than after 75 iterations.

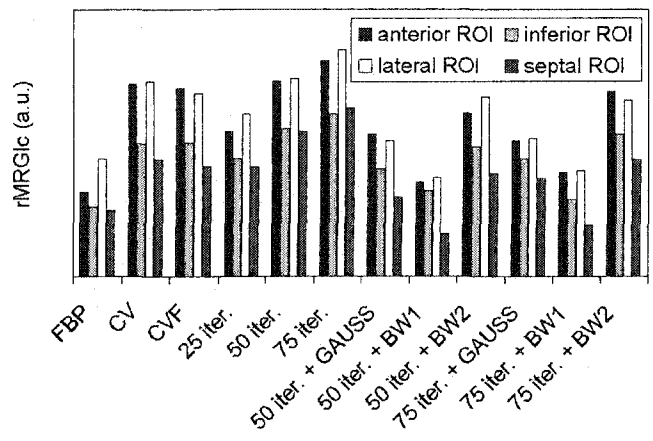

Fig. 10. rMRGlc estimates computed for myocardial ROIs using TACs of the blood pool as input functions. rMRGlc absolute value constantly increased as a function of the number of ML-EM iterations (without image filtering). Changes in the rMRGlc value and, more important, in the inter-ROI pattern were significant if heavy image filtering was applied to the image series.

lution and contrast with higher iteration number. Changes in this pattern were obvious after image filtering, e.g., estimated glucose consumption for the septum was significantly lower after image filtering than in nonfiltered cases. The CV and CVF cases yielded values in between those produced after 25 and 50 iterations. It was not surprising since the maximum number of iterations with CV and CVF did not exceed 50 (see Fig. 6). But the reduced contrast with early low-count frames resulted in noticeably lower estimate of glucose consumption in the septal ROI relative to the other myocardial segments. Estimates of rMRGlc values in the FBP case were low due to poor image contrast and lower level of estimated regional radioactivity concentration, as illustrated in Fig. 3.

IV. Discussion

Quantitative reliability of images was limited by the total registered counts. Statistical variations in low-count data samples resulted in some uncertainty of regional image quantitation (Figs. 3 and 5). The CV stopping rule was found to optimize diagnostically useful signal-to-noise ratio (SNR), but at the same time to compromise regional recovery accuracy with low-count data.

Overiteration followed by filtering did not handle low-count and high-count data equally well. Either the noise artifact was still significant in low-count images or the resolution in the images that had more counts was seriously compromised. Mild filtering could not mask the noise artifact completely when a low-count image was significantly overiterated, whereas heavy filtering yielded poor resolution in high-count images. An adaptive filtering similar to the approach reported in [16], where the filter is to be chosen for each particular image instead of applying a single filter to the sequence, could probably have been more helpful. Sophisticated regularization within iterative reconstruction could have been advantageous also. But it would obviously affect quantitation as well.

The number of iterations as a parameter of unconstrained ML-EM is important to optimize in making the method application successful, since it has an effect on the reconstruction time and on qualitative and quantitative image analysis. The CV stopping rule could be helpful with high-count dynamic series, but in a low-count situation it tends to sacrifice quantitative accuracy in favor of low-noise images. Whether the sacrifice of accuracy would be significant with compartment modeling if just a small fraction of frames had low total counts is not clear. It is possible that optimization of the frame length (after data acquisition if list-mode data are available) could reduce inaccuracy of the glucose consumption estimates and make the CV an acceptable reconstruction strategy for quantitative image analysis.

Postreconstruction image filtering also profoundly affects quantitative accuracy, and the effect of a fixed iteration number is still not negligible. Smoothing PET image sequences for compartment modeling after reconstructing with a fixed number of iterations changes the outcome of model fitting due to image contrast reduction. Moreover, this effect does not seem to be consistent for all ROIs (Fig. 10). Though the simple phantom study showed consistency of the applied filters with respect to activity and contrast recovery, the subsequent model analysis exhibits nonlinear dependence. Therefore, it is questionable whether image filtering should be performed at all before compartment modeling, apart from helping at the ROI delineation step.

The reconstruction time was reduced when using the stopping rule as compared to the overiteration accompanied by filtering even with the data splitting step, since the number of iterations adapted to the image statistics. Applying the analytic dependence exhibited by CV further reduced the reconstruction time, since the data splitting and reconstruction of the two data sets corresponding to the same image were eliminated.

Given the current popularity of iterative image reconstruction methods and the large variations in values describing metabolism when varying parameters of the reconstruction method (ML-EM in this study), there is a need for standardization in this field. The reconstruction technique should be compared to an acknowledged "standard" that would reveal the dependence of clinically useful quantitation on the technique parameters that can be varied. It is worth noting that even the widely accepted

FBP has several parameters that affect image quantitation, e.g., filter type, cutoff (or rolloff) frequency, applied data corrections, and data rebinning.

V. Conclusion

Results of this study showed the importance of optimizing the number of iterations when ML-EM reconstruction is used. The CV stopping rule was superior to overiteration and filtering ensuring balanced resolution, noise and quantitative accuracy of image estimates with projection sets having more than some minimal total counts. No image filtering was necessary after reconstruction to improve visual image perception in this case. But quantitative signal recovery was compromised in favor of diagnostically useful SNR in images with low total counts.

The described analytic dependence on total counts (CVF) simplified utilization of the stopping rule and accelerated reconstruction. It can be used instead of the original CV procedure for high-count data, when the CV yields reasonably accurate images. In a dynamic series with many low-count frames, CV is not recommended for the whole series, but reconstruction of several high-count frames at the end of the series can give a clue as to the minimum fixed number of iterations required to provide suitable images for ROI delineation without filtering/smoothing.

CV application is not limited to dynamic series only. In the case of a single-frame static emission scan, e.g., for standard uptake value calculation, CV can be used to avoid subjective decision-making at the image reconstruction step (if ML-EM is the algorithm of choice), since it would ensure optimal SNR *and* quantitative accuracy, provided the image has sufficient total counts.

Acknowledgment

The authors thank an anonymous referee for valuable comments and suggestions that significantly improved this paper.

References

[1] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imag.*, vol. MI-1, no. 2, pp. 113–122, 1982.

[2] E. Levitan and G. T. Herman, "A maximum *a posteriori* probability expectation maximization algorithm for image reconstruction in emission tomography," *IEEE Trans. Med. Imag.*, vol. MI-6, no. 3, pp. 185–192, 1987.

[3] E. Ü. Mumcuoğlu, R. M. Leahy, and S. R. Cherry, "Bayesian reconstruction of PET images: Methodology and performance analysis," *Phys. Med. Biol.*, vol. 41, no. 9, pp. 1777–1807, 1996.

[4] D. L. Snyder, M. I. Miller, L. J. Thomas Jr., and D. G. Politte, "Noise and edge artifacts in maximum-likelihood reconstructions for emission tomography," *IEEE Trans. Med. Imag.*, vol. MI-6, no. 3, pp. 228–238, 1987.

[5] J. S. Liow and S. C. Strother, "Noise and signal decoupling in maximum likelihood reconstructions and Metz filter for PET brain images," *Phys. Med. Biol.*, vol. 39, no. 4, pp. 735–750, 1994.

[6] J. Llacer and E. Veklerov, "Feasible images and practical stopping rules for iterative algorithms in emission tomography," *IEEE Trans. Med. Imag.*, vol. 8, no. 2, pp. 186–193, 1989.

[7] K. J. Coakley, "A cross-validation procedure for stopping the EM algorithm and deconvolution of neutron depth profiling spectra," *IEEE Trans. Nucl. Sci.*, vol. 38, no. 1, pp. 9–15, 1991.

[8] R. M. Lewitt and G. Muehllehner, "Accelerated iterative reconstruction for positron emission tomography based on the EM algorithm for maximum likelihood estimation," *IEEE Trans. Med. Imag.*, vol. MI-5, no. 1, pp. 16–22, 1986.

[9] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 601–609, 1994.

[10] R. Lecomte, J. Cadorette, P. Richard, S. Rodrigue, and D. Rouleau, "Design and engineering aspects of a high resolution positron tomograph for small animal imaging," *IEEE Trans. Nucl. Sci.*, vol. 41, no. 4, pp. 1446–1452, 1994.

[11] R. Lecomte, J. Cadorette, S. Rodrigue, D. Lapointe, D. Rouleau, M. Bentourkia, R. Yao, and P. Msaki, "Initial results from the Sherbrooke avalanche photodiode positron tomograph," *IEEE Trans. Nucl. Sci.*, vol. 43, no. 3, pp. 1952–1957, 1996.

[12] V. V. Selivanov, Y. Picard, J. Cadorette, S. Rodrigue, and R. Lecomte, "Detector response models for statistical iterative image reconstruction in high resolution PET," *IEEE Trans. Nucl. Sci.*, vol. 47, no. 3, pp. 1168–1175, 2000.

[13] J. Llacer, E. Veklerov, K. J. Coakley, E. J. Hofman, and J. Nunez, "Statistical analysis of maximum likelihood estimator images of human brain FDG PET studies," *IEEE Trans. Med. Imag.*, vol. 12, no. 2, pp. 215–231, 1993.

[14] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge, U.K.: Cambridge Univ. Press, 1988, pp. 542–547.

[15] M. E. Phelps, S. C. Huang, E. J. Hoffman, C. Selin, L. Sokoloff, and D. E. Kuhl, "Tomographic measurement of local cerebral glucose metabolic rate in humans with (F-18)2-fluoro-2-deoxy-D-glucose: Validation of method," *Ann. Neurol.*, vol. 6, no. 5, pp. 371–388, 1979.

[16] M. A. King, R. B. Schwinger, P. W. Doherty, and B. C. Penney, "Two-dimensional filtering of SPECT images using the Metz and Wiener filters," *J. Nucl. Med.*, vol. 25, no. 11, pp. 1234–1240, 1984.

Fast PET Image Reconstruction Based on SVD Decomposition of the System Matrix

by

Vitali V. Selivanov and Roger Lecomte

© 2001 IEEE. Reprinted, with permission, from

IEEE TRANSACTIONS ON NUCLEAR SCIENCE

Vol. 48, No. 3, June 2001

Fast PET Image Reconstruction Based on SVD Decomposition of the System Matrix

Vitali V. Selivanov, *Student Member, IEEE,* and Roger Lecomte, *Member, IEEE*

***Abstract*—Data filtering based on matrix pseudo-inverse is a well-known but not yet appreciated means of tomographic image reconstruction. In the present work, the feasibility of image reconstruction based on singular value decomposition (SVD) of the system matrix for animal two-dimensional positron emission tomography is demonstrated. Analytic detector response function accounting for the noninvariant spatial system response is explicitly included into the system matrix. Regularization of the SVD-based solution with the singular spectrum truncation (TSVD solution) derived from spatial resolution analysis is proposed. TSVD reconstruction is fast except for the matrix decomposition step, which is performed once for a given scanner geometry. Reconstructed image quality and quantitation are compared to those obtained with filtered backprojection (FBP) and iterative maximum likelihood technique. With the constant progress of computing power, TSVD image reconstruction may become a viable alternative to FBP for routine clinical applications.**

***Index Terms*—Image reconstruction, positron emission tomography, singular value decomposition.**

I. Introduction

DATA filtering based on matrix pseudo-inverse is a well-known means of image reconstruction. An early treatment applied to the problem of image reconstruction in emission imaging can be found in [1]. Image reconstruction based on pseudo-inversion of matrices with singular value decomposition (SVD) has been applied to x-ray computed tomography by Shim and Cho, though their principal interest was in imaging with a limited number of projections [2]. Other attempts met with the insufficient computer power of the day to solve the problems of practical size and the shift of interest toward iterative solutions [3], [4]. The matrix pseudo-inversion step is very demanding in terms of the numerical precision and power of computing hardware. A large condition number of the system matrix in tomography reflects the ill-posed nature of the tomographic image reconstruction problem. Thus, the solution using system matrix pseudo-inverse would be very sensitive to noise in the experimental data, and explicit regularization is essential [5], [6]. In recent years, the usefulness of the matrix pseudo-inverse approach with singular value spectrum modification has been evaluated in SPECT [7], [8]. In positron emission tomography (PET), image reconstruction has also been attempted using special formulation of matrix SVD with the matrix decomposition step performed online with a supercomputer [9]. In the present work, the feasibility of fast image reconstruction based on the SVD of the system matrix is demonstrated with animal two–dimensional (2-D) PET. The system matrix is derived using the analytical model of spatially varying detector response, and matrix decomposition is performed once for a given scanner geometry and image grid. A systematic analysis of image resolution as a function of the number of singular values included in the solution based on SVD has been performed. The regularization with the singular value spectrum truncation based on spatial resolution analysis is proposed.

Manuscript received November 5, 2000; revised March 26, 2001. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada and by the Canadian Institutes of Health Research. The work of V. V. Selivanov was also supported in part by a scholarship from the Québec Ministry of Education.

The authors are with the Department of Nuclear Medicine and Radiobiology, Université de Sherbrooke, Sherbrooke, QC J1H 5N4, Canada (e-mail: vitali@tep.crc.usherb.ca) (e-mail: rlecomte@courrier.usherb.ca).

Publisher Item Identifier S 0018-9499(01)05101-2.

II. Theory

PET imaging may be modeled by a set of the following Fredholm integral equations of the first kind:

$$b_i = \int_\Omega r_i(x) f(x) dx, \quad i = 1, \ldots, N \tag{1}$$

where

- $x \in \Omega \subset R^n$ — point of Euclidean n-dimensional space ($n = 2$ or $n = 3$), support region Ω is associated with the PET scanner field-of-view (FOV);
- $f(x)$ — unknown tracer distribution;
- b_i — projection of the tracer distribution, an approximation of which is actually measured by the detector pair i;
- $r_i(x)$ — spatially variant point response function, which may also be considered (after appropriate normalization) as the probability for an event originating at point x being registered in detector pair i;
- N — total number of tubes-of-response (active detector pairs).

It is customary to solve the inverse problem numerically utilizing discrete representations of Ω, thus, it may be written in the matrix form

$$Pf = \hat{b} \tag{2}$$

where

- $P = \{p_{ij}: i = 1, \ldots, N;\ j = 1, \ldots, M\}$ — system matrix that includes the model of the point response function for a given scanner;

0018–9499/01$10.00 ©2001 IEEE

$\hat{b} = \{\hat{b}_i : i = 1, \ldots, N\}$ vector-column of projections actually measured by the scanner and commonly modeled using Poisson distributed random variables;

$f = \{f_i : i = 1, \ldots, M\}$ vector-column (image defined on pixels) that has to be found;

M total number of image pixels.

The system of linear equations (2) has no solution in the strict sense due to the imperfect system modeling, the discrete nature of acquired PET projections, hence, inconsistency of (2), and due to the noise in experimental data. The situation may be, in principle, complicated by the improper choice of the image grid if it results in $\text{rank}P < M$. Thus, only an approximate solution is feasible.

One may argue that the least squares solution is appropriate when measured projections are precorrected for accidental coincidences and the statistical errors are no longer Poisson [10]. Alternatively, we may take into account the results of Rockmore and Macovski [11] who have shown that assuming the Poisson nature of the underlying physical process the maximum likelihood estimate of an image may be found using the matrix pseudo-inverse. The only conflicting assumption used in [11] is that the system matrix has the full rank and its pseudo-inverse may be found directly which is hardly practical. Fortunately, there is another approach to matrix pseudo-inversion.

Any matrix P can be decomposed [12] into

$$P = UDV^{\mathrm{T}} \tag{3}$$

where $U = \{u_{ij} : i = 1, \ldots, N;\ j = 1, \ldots, M\}$ and $V = \{v_{ij} : i, j = 1, \ldots, M\}$ are orthogonal matrices and $D = \{\mu_{ij} : i, j = 1, \ldots, M;\ \mu_{ij} = 0 \text{ if } i \neq j\}$ is a diagonal matrix containing singular values $\mu_i \equiv \mu_{ii}$, $i = 1, \ldots, M$. This factored matrix representation is known as the SVD.

One may find the minimum norm least squares solution of (2) using

$$\hat{f} = P^{+}\hat{b} \tag{4}$$

where

$$P^{+} = VD^{+}U^{\mathrm{T}} \tag{5}$$

is the pseudo-inverse of P [12]. The range of singular values for a real-world system matrix can be very wide. Some singular values (they are presented as a nonincreasing sequence called singular value spectrum) can be very small (or even zeros). Thus, the condition number

$$c = \frac{\max_i \mu_i}{\min_i \mu_i} \tag{6}$$

of the system matrix can be very high (even infinity if the matrix is singular). This ill-conditioning of the system matrix is the main reason why the pseudo-inverse can hardly be found directly even if P has the full rank. Moreover, the solution directly exploiting the pseudo-inverse would be very sensitive to noise. One simple regularization approach is the truncation of the singular value spectrum at some index T and removal of very small values μ_i, $i = T + 1, \ldots, M$ from the solution, leaving

$$\hat{f} = \left\{ \hat{f}_k = \sum_{i=1}^{T} v_{ki}\mu_i^{-1} \sum_{j=1}^{N} u_{ji}\hat{b}_j; \quad k = 1, \ldots, M; T \leq M \right\} \tag{7}$$

which is known as the TSVD solution. The TSVD solution may be viewed as a special case of a general regularization approach involving singular value spectrum modification given by

$$\hat{f} = \left\{ \hat{f}_k = \sum_{i=1}^{M} v_{ki} g(\mu_i) \sum_{j=1}^{N} u_{ji}\hat{b}_j; \quad k = 1, \ldots, M \right\} \tag{8}$$

where, in the TSVD case, $g(\mu_i)$ is set to

$$g(\mu_i) = \begin{cases} \mu_i^{-1}, & i \leq T, \\ 0, & i > T \end{cases}. \tag{9}$$

III. Materials and Methods

PET data were acquired with the Sherbrooke high-resolution animal tomograph [13]. There were 8192 active tubes-of-response per slice in the acquisition mode used in this paper, i.e., $N = 8192$. A set of measurements was performed with a line source of ^{22}Na having an active diameter of 0.84 mm. It was placed axially within the field-of-view (FOV) and moved radially across the FOV in steps of 5 mm covering almost one FOV diameter. More than 1.5×10^6 coincident events were registered at each source position. A phantom of 110-mm diameter made of Lucite and having holes of diameter 2, 3.4, 6.7, 9.7, 13, 15.8, 20.3, and 22.7 mm located on a circumference at a distance of 28 mm from the center was also scanned. A total of 3.6×10^6 coincident events were recorded. Data were corrected for random coincidences, detector efficiency, and, in the phantom case, for attenuation.

The matrix P was derived based on geometrically modeled tubes-of-response (intersection of which with the FOV is sometimes referred to as natural pixels [14]) and took into account the spatially variant analytical detector response model derived from the linear attenuation of γ-rays in a detector array [15]. The contributions of a given square pixel of the image grid to all tubes-of-response were normalized to yield probabilities

$$\sum_{k=1}^{N} p_{kj} = 1, \quad j = 1, \ldots, M \tag{10}$$

that ensured the global count preserving property of TSVD reconstruction.

Grids of 64×64, 80×80, and 96×96 pixels were utilized for comparison, the limit being defined by the maximum available amount of fast random access memory (RAM) on the workstation (512 MB) in order to ensure a reasonably fast SVD of the system matrix. SVD decomposition was performed with 64-bit floating-point arithmetics using the approach given in

[16]. Only pixels within a circle inscribed into the square FOV were taken into account. Matrices of 8192×3332, 8192×5180, and 8192×7400 were decomposed for the above-mentioned image grids, respectively.

The truncation index T was defined after the spatial resolution analysis performed with the "point" sources reconstructed with TSVD according to the following methodology. T took values from 1 to M sequentially and a profile of the point source image reconstructed with a given T was fitted with a Gaussian function by the Levenberg–Marquardt method [17] in two in-plane directions: radial and tangential. These estimates will be called *local resolution estimates*. Negative values were replaced with zeros in the profile before the fitting procedure. *Global resolution estimates* were obtained by fitting the local resolution estimates sampled at regular intervals along the FOV diameter (a total of 21 points with 5-mm steps in this study) to a straight line assuming that the dependence on the distance from the FOV center is linear. Matching radial resolution (global estimate), given by the full-width at half-maximum (FWHM) at the FOV center, to the intrinsic scanner resolution (1.9 mm) yielded fixed T. Only radial estimates were used to derive the truncation index here since the radial and tangential global reconstructed resolution estimates are highly correlated at the FOV center.

For comparison phantom data were also reconstructed with the filtered backprojection (FBP) method [18] using a ramp filter with cut-off frequency satisfying the Nyquist criteria. The iterative maximum likelihood with expectation maximization (ML-EM) technique [19] was used as well to reconstruct the phantom data. A 128×128-pixel grid was used with FBP and ML-EM. ML-EM utilized the system matrix P described above. FBP obviously had to use the conventional Radon transform PET model.

Relative recovery factors were calculated from the hot spot mean density in the phantom images

$$RF_i = \frac{\bar{H}_i}{\bar{H}_{\max}} \times 100\% \tag{11}$$

where $\bar{H}_i$ is the mean density value within the ith hot spot of the phantom, and $\bar{H}_{\max} = \max_i \bar{H}_i$.

IV. Results

The singular value spectra of the system matrix P for the Sherbrooke animal PET scanner and image grids of 64×64, 80×80, and 96×96 pixels are reported in Fig. 1. The condition number c of the system matrix was (approximately) 4442, 51 784, and 15 585 310 with the above-mentioned image grids, respectively. Even more instructive is the log–log plot displayed in Fig. 2, which shows the existence of a "plateau" of singular values and an abrupt turning point revealing the spectrum tail responsible for the severe ill conditioning of the reconstruction inverse problem. The inclusion of this tail in the matrix pseudo-inverse solution results in enormous noise amplification.

An example of the local resolution analysis obtained by varying truncation index T is reported in Fig. 3. The point source was located in the FOV center and the image was reconstructed with TSVD on a 96×96 pixel grid in this case.

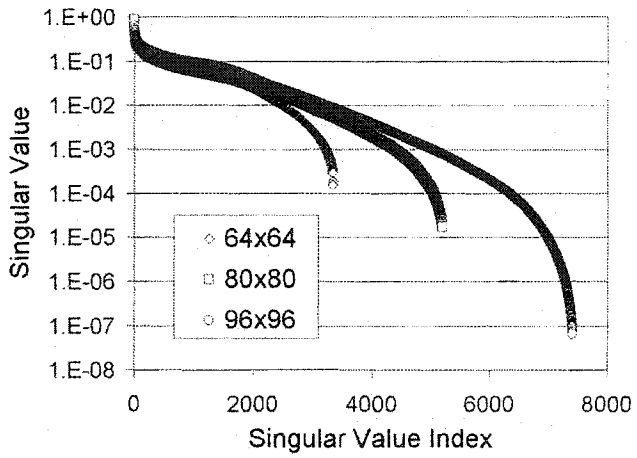

Fig. 1. Singular value spectra of the system matrix for the Sherbrooke animal PET scanner and three different image grids.

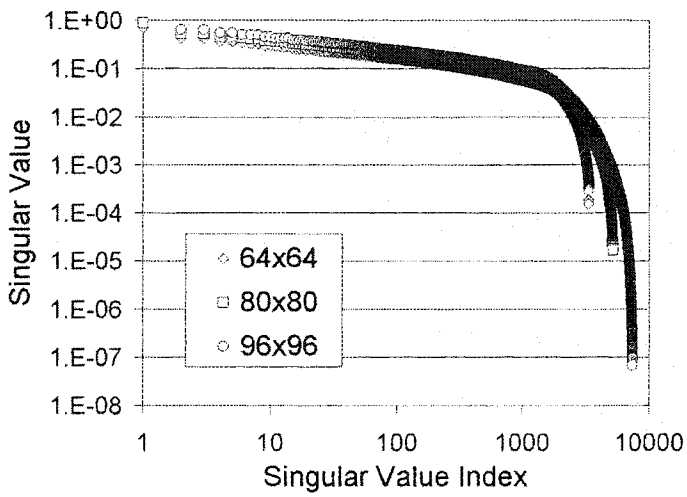

Fig. 2. Same as Fig. 1 on a log–log scale showing the existence of a "plateau" of singular values and an abrupt drop-off of the spectrum tail responsible for the severe ill-conditioning of the reconstruction inverse problem.

A very small percentage of the points was excluded from the chart since the fit was not found for a few indexes T or the uncertainty in the computed local resolution estimate was very high (actually, estimates having standard deviation higher than 1% of the estimated value were dropped). Matching radial (local or global) resolution in reconstructed images to the intrinsic scanner resolution in the FOV center suggested using only a fraction of the singular value spectrum. The values of truncation index T derived using the spatial resolution analysis are summarized in Table I. The numbers shown in Table I should not be considered the solution for any measuring instrument but rather serve as guidance. There are associated errors, first due to the nature of PET data acquisition, then due to the image reconstruction method, and finally due to the fitting procedure. In addition, the imaging geometry, the model of detector response, and the structure of the image grid are also important factors, since they are explicitly accounted for in the system matrix. Only spatial resolution analysis performed with a particular imaging geometry is able to provide the valid answer in each case.

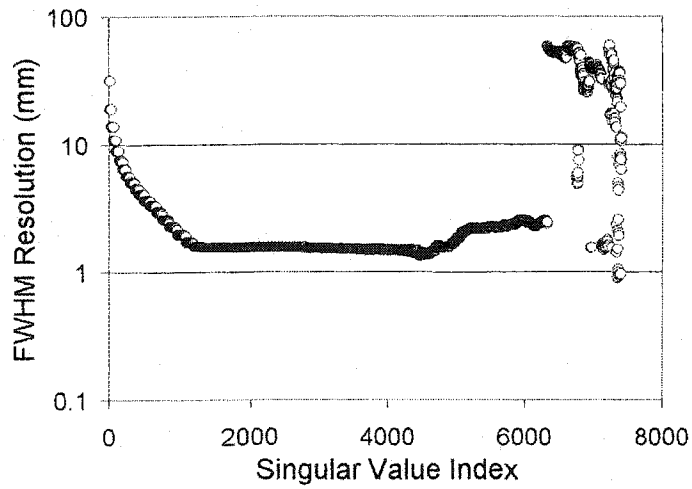

Fig. 3. Radial FWHM local resolution estimates of a point source in the FOV center, reconstructed with TSVD with varying truncation index T, 96×96-pixel image.

TABLE I
TRUNCATION INDEX T BASED ON SPATIAL RESOLUTION ANALYSIS (USING GLOBAL RESOLUTION ESTIMATES)

Image grid	Active pixels	Optimal T
64×64	3332	822
80×80	5180	1053
96×96	7400	1217

TSVD reconstruction using small T yielded low-resolution images, which exhibited a severe ringing artifact [Fig. 4(a)]. In general, point source image quality and resolution did not change noticeably when picking T from the middle of the singular value spectrum [Fig. 4(b)] and making small deviations ΔT. Therefore, it was feasible to set the reconstructed image resolution somewhat higher than the intrinsic scanner resolution (1.9 mm) by increasing index T obtained with the spatial resolution analysis if some moderate noise amplification was tolerated. Inclusion of most of the singular values into the solution resulted in very noisy images resembling the point source very little, and the noise artifact was dominant [Fig. 4(c)].

Global and local resolution estimates across the FOV derived using point source images reconstructed with TSVD (T = 1217, 96×96-pixel image) are reported in Fig. 5.

Phantom images reconstructed with FBP, ML-EM, and two cases of TSVD are presented in Fig. 6. TSVD images (96×96 pixels) were enlarged to show them at the same size as FBP and ML-EM images (128×128 pixels). TSVD images have a distinct appearance exhibiting a streak-like artifact resembling the one observed with the FBP image. But at the comparable resolution level [Fig. 6(c)] the nonactive background is somewhat cleaner with TSVD as compared to that of the FBP image.

Image profiles through the largest holes of the phantom are shown in Fig. 7. The FBP image had the lowest variance within the uniformly hot regions and the blurriest appearance at the same time. This may be attributed in part to the data rebinning and interpolation that reduced resolution and correlated data noise. Note that the projection data interpolation has not been performed except with FBP. ML-EM after 100 iterations exhibited the developing noise artifact, which was possibly aggravated by the system matrix rank deficiency in this case. Image resolution after 100 ML-EM iterations was visibly higher than with the other methods. TSVD yielded rather noisy hot spots as well, which was in part the result of the sampling limitations.

Recovery factor analysis performed using reconstructed phantom images (shown in Fig. 6) is reported in Fig. 8. TSVD

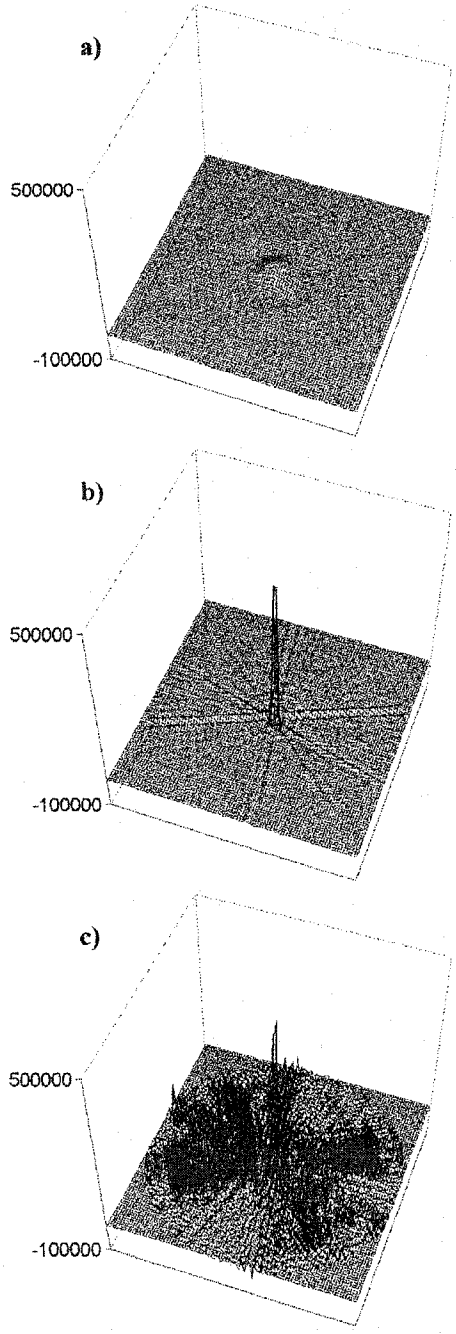

Fig. 4. Point source in the FOV center reconstructed with TSVD, 96×96 pixel image: a) T = 100; b) T = 1217; and c) T = 4000. Estimated global radial FWHM resolution in the FOV center: a) 8.7 mm; b) 1.9 mm; and c) 1.6 mm.

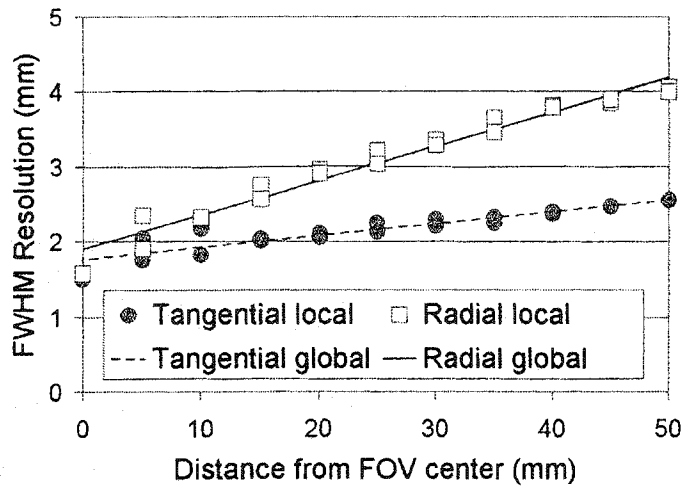

Fig. 5. FWHM resolution estimates (local and global), derived using point source images reconstructed with TSVD, $T = 1217$, 96×96 pixel image.

yielded consistent recovery coefficients that were lower than with ML-EM after 100 iterations, but higher than with FBP reconstruction.

V. Discussion

The SVD of the system matrix in PET, apart from precise numerical diagnostics of the tomographic reconstruction ill-conditioning with a given detection system and image grid, provides a linear and very fast reconstruction means. All the factors related to PET system configuration are accounted for once during the system matrix derivation. One should distinguish between the factors depending on the day-to-day system performance as well as the object being imaged and the factors specific to a given scanner that do not change (detector bank geometry, crystal properties as well as image grid used for reconstruction). The latter may be taken into account in the system matrix, and there is no point in correcting data for these latter factors online, i.e., every time the reconstruction is launched. Therefore, the time-consuming step of system matrix decomposition needs to be performed only once for a given system and image grid. As a result, necessary data corrections for detector efficiency, randoms, object scatter, and attenuation should be performed using approaches widely adopted with FBP reconstruction.

Despite its high computational requirements, the SVD of a huge matrix for a state-of-the-art PET system operating in 2-D mode is within the reach of current state-of-the-art computing platforms. A workstation based on a single Intel Pentium III (500 MHz) processor running under a Windows NT Workstation operating system was used to successfully perform matrix SVD decomposition. The workstation had 512 MB of RAM and took days to complete calculations in the case of 96×96-pixel image (8192×7400 system matrix), but was hardly the optimal platform for such a task. Note that one needs to perform this offline matrix SVD calculation only once as long as the same detection system configuration and FOV pixelization are utilized. Thus, the manufacturer of commercial PET systems may perform matrix decomposition using special computing hardware

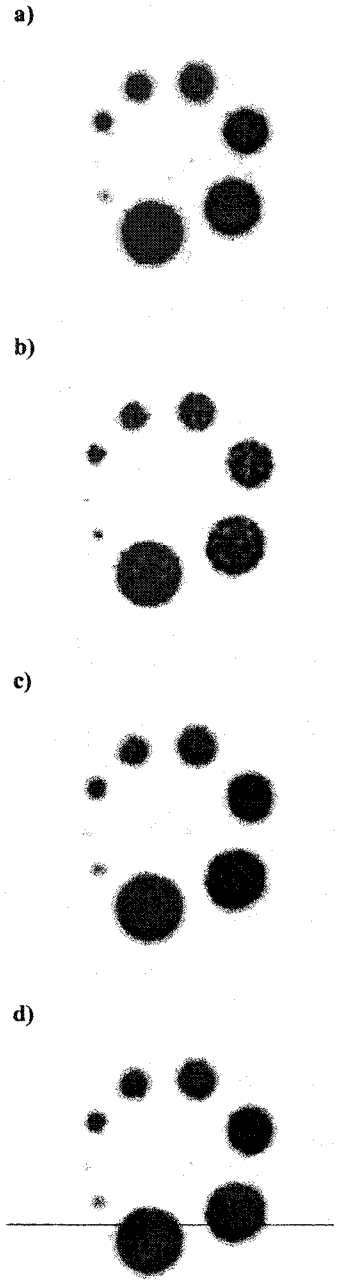

Fig. 6. Phantom images reconstructed with: a) FBP; b) ML-EM (100 iterations); c) TSVD, matching FBP (global) resolution ($T = 950$); and d) TSVD, matching global reconstructed image resolution to the intrinsic scanner resolution in the FOV center ($T = 1217$). TSVD images (96×96 pixels) were enlarged to show them at the same size as FBP and ML-EM images (128×128 pixels). The line in d) represents the direction of the profile shown in Fig. 7.

covering at once all the similar systems. Necessity to cope with the large total size of the three resulting SVD pieces (or one combined pseudo-inverse-like matrix) is the price for the simplicity and speed of the online reconstruction.

Proper truncation of the singular value spectrum sets the meaningful balance between the signal and noise components in the image estimate. It should be noted here that the ill-conditioning of the tomographic inverse problem aggravates

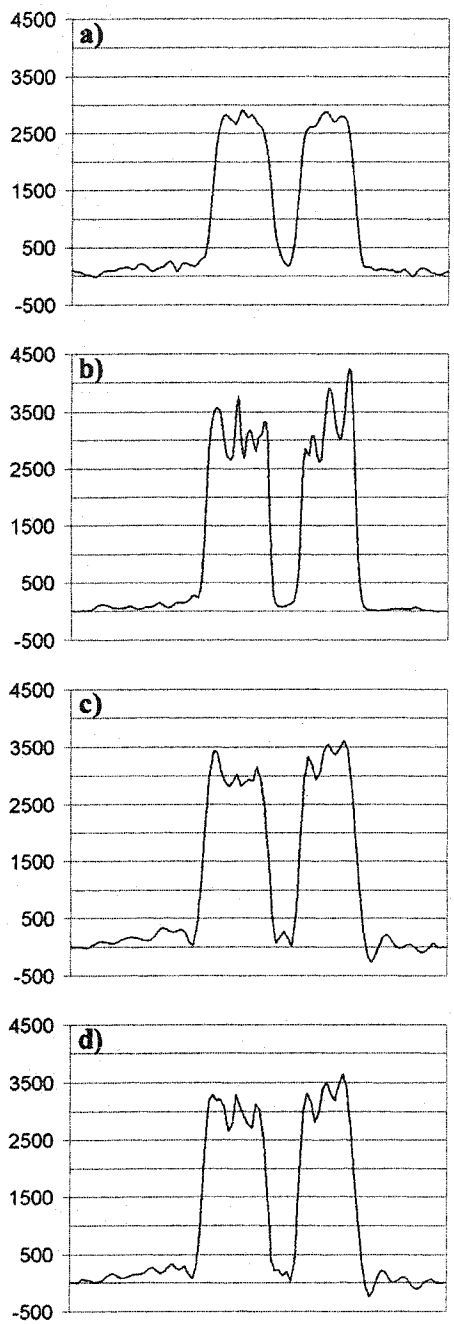

Fig. 7. Image profiles through the two largest holes of the phantom along the line shown in Fig. 6(d): a) FBP; b) ML-EM, 100 iterations; c) TSVD, $T = 950$; and d) TSVD, $T = 1217$.

difficulties in handling the statistical noise of the measured data. Therefore, perfect statistical algorithms working in theory (ML-EM is an example) do not work as expected with real data and explicit regularization becomes mandatory. One is able to get rid of ill-conditioning in an explicit manner with TSVD by not using the smallest singular values.

Index T, as we have shown, also sets the tradeoff between noise and resolution. Results on the preferred singular value spectrum truncation obtained with the spatial resolution analysis correlate strongly with the shape of the spectrum revealed by the graph in Fig. 2. The spectrum tail responsible for the severe ill-conditioning, which is clearly seen in Fig. 2, has to be truncated (or modified if another regularizing approach is preferred) in order to produce meaningful images. The presence of the abrupt turning point in the spectrum which was found to correlate with the optimal reconstructed image resolution suggests that the truncation index may also be found approximately with spectrum analysis only. Here by optimal resolution we mean the equality of reconstructed image resolution to the actual intrinsic resolution delivered by a given scanner with all the inherent resolution nonuniformity within the FOV.

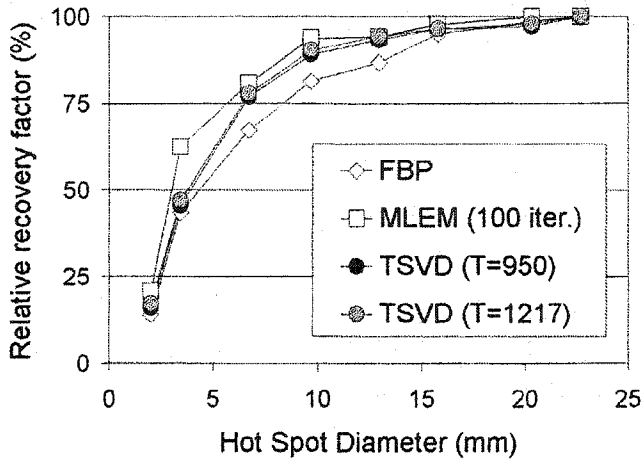

Fig. 8. Recovery factors calculated using reconstructed phantom images shown in Fig. 6.

The TSVD reconstruction shares some drawbacks with FBP: negative values in the image estimate and streak-like artefact with low-count images. But it also has very attractive benefits.

- The commonly used mathematical model of PET data acquisition using the Radon transform may be replaced with a model utilizing tubes-of-response or natural pixels, which has stronger physical grounds.
- The detector response can be easily included in the reconstruction, which eliminates the assumption of a spatially invariant system response exploited by FBP.
- Data rebinning within the imaging plane is not necessary since the actual geometry of a given system is utilized.
- One may set the resolution of reconstructed images based on the prior resolution analysis by changing the truncation index T or based on another regularization technique involving singular value spectrum modification.
- TSVD reconstruction is potentially faster than FBP and is definitely faster than any iterative image reconstruction technique, as it comes to just one matrix to vector multiplication if the regularization approach by singular value spectrum truncation (or modification) is set once for the routine utilization.

VI. Conclusion

Linear and fast image reconstruction based on the system matrix SVD is an attractive reconstruction approach. The computational burden and numerical stability of SVD decomposition of huge matrices may be overcome with current

state-of-the-art computing platforms. In this work the feasibility of TSVD image reconstruction was demonstrated for small animal high-resolution 2-D PET.

A regularization methodology was proposed based on the spatial resolution analysis. The peculiarity of the singular value spectrum shape was found which revealed the relationship between the optimal spectrum truncation to be used with TSVD and the optimal reconstructed image resolution. This relationship suggests that the spectrum truncation index may also be derived based on the analysis of the singular value spectrum.

TSVD reconstruction has a number of benefits that are utilized in iterative image reconstruction techniques but are hard to account for using FBP. TSVD utilizing the proposed regularization approach delivered images of comparable visual quality, but somewhat higher quantitative accuracy than FBP. Based on the results of this work, TSVD may be a viable alternative to FBP image reconstruction for routine clinical applications.

ACKNOWLEDGMENT

The authors would like to thank S. Rodrigue and J. Cadorette, staff members of the Metabolic and Functional Imaging Center at the Université de Sherbrooke, for the invaluable help in PET data acquisition. The authors also acknowledge illuminating conversations with M. Defrise during the IEEE Medical Imaging Conference in Lyon, France.

REFERENCES

[1] T. F. Budinger and G. T. Gullberg, "Three-dimensional reconstruction in nuclear medicine emission imaging," *IEEE Trans. Nucl. Sci.*, vol. 21, pp. 2–20, June 1974.

[2] Y. S. Shim and Z. H. Cho, "SVD pseudoinversion image reconstruction," *IEEE Trans. ASSP*, vol. 29, pp. 904–909, 1981.

[3] J. Llacer, "Tomographic image reconstruction by eigenvector decomposition: Its limitations and areas of applicability," *IEEE Trans. Med. Imag.*, vol. 1, pp. 34–42, 1982.

[4] J. Llacer and J. D. Meng, "Matrix-based image reconstruction methods for tomography," *IEEE Trans. Nucl. Sci.*, vol. 32, pp. 855–864, 1985.

[5] M. Bertero, C. D. Mol, and E. R. Pike, "Linear inverse problems with discrete data. I: General formulation and singular system analysis," *Inverse Problems*, vol. 1, pp. 301–330, 1985.

[6] ——, "Linear inverse problems with discrete data. II: Stability and regularization," *Inverse Problems*, vol. 4, no. 3, pp. 573–594, 1988.

[7] M. F. Smith, C. E. Floyd Jr, R. J. Jaszczak, and R. E. Coleman, "Reconstruction of SPECT images using generalized matrix inverses," *IEEE Trans. Med. Imag.*, vol. 11, pp. 165–175, 1992.

[8] M. F. Smith, "Generalized matrix inverse reconstruction for SPECT using a weighted singular value spectrum," *IEEE Trans. Nucl. Sci.*, vol. 43, pp. 2008–2017, June 1996.

[9] J. R. Baker, T. F. Budinger, and R. H. Huesman, "Generalized approach to inverse problems in tomography: Image reconstruction for spatially variant systems using natural pixels," *Critical Rev. Biomed. Eng.*, vol. 20, no. 1/2, pp. 47–71, 1992.

[10] J. A. Fessler, "Penalized weighted least-squares image reconstruction for positron emission tomography," *IEEE Trans. Med. Imag.*, vol. 13, pp. 290–300, 1994.

[11] 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.

[12] G. Strang, *Linear Algebra and Its Applications*, 2nd ed: Harcourt Brace Jovanovich, 1980, pp. 137–145.

[13] R. Lecomte, J. Cadorette, S. Rodrigue, D. Lapointe, D. Rouleau, M. Bentourkia, R. Yao, and P. Msaki, "Initial results from the Sherbrooke avalanche photodiode positron tomograph," *IEEE Trans. Nucl. Sci.*, vol. 43, pp. 1952–1957, June 1996.

[14] M. H. Buonocore, W. R. Brody, and A. Macovski, "A natural pixel decomposition for two-dimensional image reconstruction," *IEEE Trans. Biomed. Eng.*, vol. 28, pp. 69–78, 1981.

[15] V. V. Selivanov, Y. Picard, J. Cadorette, S. Rodrigue, and R. Lecomte, "Detector response models for statistical iterative image reconstruction in high resolution PET," *IEEE Trans. Nucl. Sci.*, vol. 47, pp. 1168–1175, June 2000.

[16] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing*: Cambridge Univ. Press, 1988, pp. 60–72.

[17] ——, *Numerical Recipes in C: The Art of Scientific Computing*: Cambridge Univ. Press, 1988, pp. 542–547.

[18] S. W. Rowland, "'Computer implementation of image reconstruction formulas' in Image Reconstruction from Projections: Implementation and Applications," *Topics in Applied Physics*, vol. 32, pp. 9–79, 1979.

[19] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Med. Imag.*, vol. 1, pp. 113–122, 1982.

To be submitted

List-mode PET Image Reconstruction Based on Regularized Pseudo-Inverse of the System Matrix

Vitali V. Selivanov and Roger Lecomte, *Member, IEEE*

Metabolic and Functional Imaging Center, Clinical Research Centre,

Université de Sherbrooke, Sherbrooke, Québec, Canada, J1H 5N4

This work was supported in part by the Natural Sciences and Engineering Research Council of Canada, and by the Canadian Institutes of Health Research. V.V.S. was also supported in part by a scholarship from the Québec Ministry of Education.

For all correspondence, please contact:

Roger Lecomte, Ph.D.
Departement of Nuclear Medicine and Radiobiology
Faculty of Medicine, Université de Sherbrooke
Metabolic & Functional Imaging Center, CRC/CHUS Room 1864
3001, 12th Avenue N.
Sherbrooke (Québec)
Canada J1H 5N4
Phone: (810) 346-1110 x.14608
Fax: (819) 820-6490
Email: rlecomte@courrier.usherb.ca

Abstract

The feasibility of tomographic image reconstruction by projection data filtering based on the pseudo-inverse of the system matrix has recently been demonstrated in high-resolution animal positron emission tomography (PET). A regularisation approach using truncation of the singular-value spectrum based on the systematic spatial-resolution analysis has been proposed and successfully applied. In the present paper, we show how list-mode image reconstruction can be achieved using the regularized pseudo-inverse of the system matrix. An update of the current image estimate can be obtained using one column of the regularized pseudo-inverse matrix to account for the next registered event, thus allowing, in principle, for instant visualization of the radioactivity distribution while the object is still being scanned. The reconstruction process involves just one vector-to-vector addition in the simplest case, or one scalar-to-vector multiplication followed by the vector addition if data corrections are performed online. Computed estimates converge to the minimum-norm least-squares solution of the regularized inverse problem when sufficient total counts are acquired to fulfill the assumption of the normal (Gaussian) data error distribution. Only the total number of pixels in the discrete image determines the computational expenses of image updating for a single registered event. Algorithm data storage requirements are discussed. Limited-angle tomography and non-traditional detection geometry may be handled using the described reconstruction approach as well. The proposed method was tested with the list-mode PET data.

I. INTRODUCTION

Image restoration using matrix inversion was applied early on for radionuclide imaging [1], but was considered impractical for tomographic image reconstruction at the time [2]. Later, a maximum likelihood approach was proposed for emission imaging, taking into account stochastic data nature and assuming measurements are described by a Poisson process [3]. The latter approach involved matrix pseudo-inverse but did not take into account the severe ill-conditioning of the image reconstruction problem. Matrix pseudo-inversion via Singular Value Decomposition (SVD) has also been investigated in x-ray computed tomography [4]. The use of SVD enables explicit regularisation by spectral filtering, which potentially improves conditioning of the inverse problem. There are a number of regularisation approaches involving SVD, and optimization of the regularisation parameter is always required [5]-[9].

Image reconstruction based on the matrix SVD has been applied in single-photon emission computed tomography (SPECT) [10]-[12] and in positron emission tomography (PET) [13], but a practical reconstruction technique has not been shown. Recently, the feasibility of fast tomographic image reconstruction by projection data filtering based on the regularized pseudo-inverse of the system matrix has been demonstrated in high-resolution animal PET. A regularisation approach based on systematic spatial resolution analysis and truncation of the singular value spectrum was proposed and successfully applied [14].

Previous attempts at reconstructing tomographic data in real time were based on the principle of superposition of filter functions for individual events derived from the filtered-backprojection (FBP) algorithm. Several variants of such "real-time"

reconstruction algorithms were considered for conventional two-dimensional (2-D) image reconstruction [15], time-of-flight positron emission tomography (TOFPET) [16], and 3-D PET [17]-[19]. The implementations usually involved simplifying assumptions, such as reordering of filtering and backprojection, and required special dedicated hardware to cope with the computational load of backprojection onto a grid of voxels online [20]-[22]. System modelling was not considered and was not feasible in real time with the proposed techniques.

In this paper, we show how potentially real-time PET image reconstruction can be achieved using a regularized pseudo-inverse of the system matrix that includes a proper model of the system response. The proposed method was successfully tested with list-mode 2-D PET data.

II. Background Theory

A. Inverse Problem

The process of measuring tomographic "projections" in emission tomography is stochastic in nature. It is widely accepted that Poisson distribution adequately describes the counting statistics for large quantities of radioactive nuclei. Thus, the following model may describe an idealized PET imaging process [23]:

$$b_i^* = \int_{\Omega} p_i(x) f(x) dx, \; i = 1,...,N, \qquad (1)$$

where

$x \in \Omega \subset R^n$ point of a Euclidean n-dimensional space (n=2 or n=3), Ω is the support region associated with the PET scanner field of view (FOV);

$f(x)$ unknown tracer distribution that is described by the means of Poisson point processes;

b_i^* expected value of the counter (in the probabilistic sense) corresponding to the detector pair i (yielding a single bin of a discrete projection of the tracer map at a given angle); a sample of the N-dimensional random variable b having expectation vector $b^* = (b_1^*, b_2^*, ..., b_N^*)^T$ is measured; under ideal conditions, components of the random variable are considered independent and Poisson distributed (this assumption will be discussed in more detail below);

$p_i(x)$ spatially varying point response function, which, after appropriate normalization, becomes the probability of an event originating at point x being registered by detector pair i;

N total number of tubes of response (active detector pairs).

System (1) is a set of Fredholm integral equations of the first kind [7] incorporating stochastic assumptions. In order to facilitate numerical solution, $f(x)$ is usually represented as

$$f(x) \cong \sum_{i=1}^{M} f_i \phi_i(x), \qquad (2)$$

where

$\{\phi_i(x), i = 1, ..., M\}$ expansion function set spanning the desired functional space; in the absence of additional constraints on the smoothness of $f(x)$, the space $L_2(\Omega)$ of square-integrable functions defined on Ω would be

relevant, but one has to resort to a narrower space of finite dimensionality, a subspace of $L_2(\Omega)$, to ensure computational tractability;

$\{f_i, i = 1,..., M\}$ scalar expansion coefficients;

M total number of expansion functions.

Taking into account (2), the system (1) is reduced to a simple matrix form:

$$b^* = Pf \qquad (3)$$

where

$f = (f_1, f_2,..., f_M)^{\mathrm{T}}$ vector of the coefficients in expansion (2), which is conventionally referred to as the image;

P $N \times M$ real-valued system matrix incorporating the model of spatially varying system response derived by mapping $p_i(x)$ onto the image grid and translating to probabilities.

In practice, one has to estimate f based on a single realization of b. This yields the following problem:

$$b_\varepsilon = Pf + \varepsilon . \qquad (4)$$

Here b_ε is the measured data (sinogram), and $\varepsilon = (\varepsilon_1,..., \varepsilon_N)^{\mathrm{T}}$ is a vector of measurement errors. It is known that a sample drawn from the Poisson distribution having a large mean would be hard to distinguish from one extracted from the Gaussian (normal) distribution. The "large" mean is not so large in practice, e.g. [24] advises that "for values of the mean greater than about 10, the Gaussian distribution closely approximates the shape of the Poisson distribution." Therefore, it may be appropriate to assume that

data are drawn from the Gaussian distribution if the mean is large enough. It has been shown also that statistical errors are no longer Poisson if measured projections are pre-processed to correct for accidental coincidences, scanner sensitivity, or dead time before reconstruction. In this case, a weighted least-squares solution [25] or transformation into noise equivalent counts followed by maximum-likelihood reconstruction [26] was proposed. We assume that $\varepsilon_i \sim N(0,\sigma_i)$, i.e. the errors are drawn from normal distribution. At this point we assume that errors ε_i, $i = 1,...,N$ are identically distributed by assigning a single variance $\sigma^2 = \max_i \sigma_i^2$. This allows for easy solution but potentially worsens the resulting image variance estimates. Let us give extreme examples, e.g. a few point sources scattered across the FOV would result in emission data that would be reconstructed with suboptimal precision due to this latter assumption, and a flood source occupying the whole FOV would be the best candidate to produce emission data satisfying the latter assumption. Given the assumptions introduced above, the model (4) would be a valid approximation of (3) for high total counts.

B. Solution Using the Pseudo-inverse Matrix

A well-known solution of (4) that is unique is the minimum-norm least-squares estimate:

$$f_{MNLS} = P^{+}b, \qquad (5)$$

where P^{+} is the Moore-Penrose pseudo-inverse. If P has full rank, pseudo-inverse may be computed directly as:

$$P^{+} = \left(P^{T}P\right)^{-1}P^{\mathrm{T}}, \qquad (6)$$

where T stands for transpose. However, it is more general to take an approach that does

not assume the invertibility of $P^{\mathrm{T}}P$.

It is well known that any matrix P can be decomposed into a product of three special matrices [27]:

$$P = UDV^{\mathrm{T}}, \qquad (7)$$

where

$D = diag(\mu_1, \mu_2, ..., \mu_M)$ diagonal matrix of singular values, which are usually ordered so that $\mu_1 \geq \mu_2 \geq ... \geq \mu_M \geq 0$; the ratio μ_1/μ_M is referred to as the condition number of matrix P and measures the inverse problem conditioning;

$U = (u_{ij})_{i=1,...,N;\ j=1,...,M}$ $N \times M$ matrix with orthonormal columns, the columns corresponding to the non-zero singular values form an orthonormal basis spanning the range of matrix P;

$V = (v_{ij})_{i=1,...,M;\ j=1,...,M}$ $M \times M$ matrix with orthonormal columns, the columns corresponding to the zero singular values (if any) form an orthonormal basis for the nullspace of matrix P.

This matrix representation is known as the SVD, and the set of singular values is called the singular-value spectrum. The minimum-norm least-squares solution (5) may be found with the pseudo-inverse based on the SVD of the matrix P:

$$P^{+} = VD^{+}U^{\mathrm{T}}, \qquad (8)$$

where D^{+} is a diagonal matrix containing the reciprocals of the respective positive singular values, i.e.

$$D^{+} = diag\left(d_i^{+}\right),\ \ d_i^{+} = \begin{cases} \mu_i^{-1}, & if\ \mu_i > 0 \\ 0, & if\ \mu_i = 0 \end{cases}. \quad (9)$$

Let k be the total number of non-zero singular values. The minimum variance of the resulting image estimate obtained with the pseudo-inverse based on the SVD can be given by [28]:

$$\sigma^2\left(\hat{f}_i\right) = \sum_{j=1}^{k}\left(\frac{V_{ij}}{\mu_j}\right)^2,\ i = 1,...,M\ . \quad (10)$$

The variance estimate given in (10) describes the uncertainty coming from the inverse problem conditioning, which may dominate the total variance estimate (with the data variance included) for moderately or severely ill-conditioned cases, i.e. when the condition number of the system matrix is high.

Truncation of the singular value spectrum at some non-zero value having index T (for regularisation) results in a "truncated" SVD (TSVD) solution:

$$\hat{f} = \left\{\hat{f}_k : \hat{f}_k = \sum_{i=1}^{T} v_{ki}\mu_i^{-1}\sum_{j=1}^{N} u_{ji}b_j;\ k = 1,...,M;\ T \le M\right\}. \quad (11)$$

A more general regularisation approach of which (11) is a special case, is modification of the singular spectrum in order to diminish the effect of inverting very small singular values (and zeros if any). The solution then can be written as:

$$\hat{f} = \left\{\hat{f}_k : \hat{f}_k = \sum_{i=1}^{M} v_{ki}g(\mu_i)\sum_{j=1}^{N} u_{ji}b_j;\ k = 1,...,M\right\}, \quad (12)$$

where $g(\mu_i)$ is the function used to modify the singular value spectrum. Such a solution is referred to as the modified SVD (MSVD) solution. An example of MSVD can be found in [12]. The choice

$$g(\mu_i) = \begin{cases} 1, & \text{if } i \le T \\ 0, & \text{otherwise} \end{cases} \qquad (13)$$

yields equation (11). Another well-known example is the Tikhonov-Phillips filter function [29]:

$$g(\mu_i) = \frac{\mu_i}{\mu_i^2 + \eta}, \qquad (14)$$

where η is a free parameter

An SVD-based solution can be used to reconstruct data acquired with an arbitrary 2-D as well as 3-D system geometry. A major limitation however is the numerical decomposition of huge matrices, which becomes feasible with the current general-purpose computing hardware, as reported recently [14]. Note that the matrix SVD calculation has to be performed only once as long as the same detection system configuration and the chosen image expansion with finite series or voxels are utilized. Object independent factors such as the spatially varying system response utilizing models of positron emission and photon detection should be explicitly included during the computation of the system matrix. Other factors such as normalization for detector efficiency and corrections for attenuation in tissue can be accounted for by multiplying the columns of the regularized pseudo-inverse matrix with the appropriate correction factors [30].

III. List-mode Image Reconstruction

Rearranging the summation order in (11), one gets

$$\hat{f}_k = \sum_{j=1}^{N}\left[\sum_{i=1}^{T} v_{ki}\mu_i^{-1}u_{ji}\right]b_j;\ k=1,...,M\ . \qquad (15)$$

Let $\tilde{P}^+$ be the "truncated" pseudo-inverse:

$$\tilde{P}^+ = \left\{\tilde{p}_{kj}:\ \tilde{p}_{kj} \equiv \sum_{i=1}^{T} v_{ki}\mu_i^{-1}u_{ji};\ k=1,...,M;\ \ j=1,...,N\right\}, \quad (16)$$

then (11) becomes

$$\hat{f} = \tilde{P}^+ b\ . \qquad (17)$$

Let $b(t)$ be a sinogram containing a total of t counts, i.e.

$$b(t) = \left\{b_i(t) \equiv b_i : \sum_{j=1}^{N} b_j = t;\ i=1,...,N\right\}. \qquad (18)$$

Let $\Delta b^{(s)} = \{\Delta b_i : i=1,...,N\}$ be the data update vector such that

$$\Delta b_i \equiv \begin{cases} 0,\ i=1,...,s-1 \\ 1,\ i=s \\ 0,\ i=s+1,...,N \end{cases} \qquad (19)$$

and

$$b^{(s)}(t) \equiv b(t-1) + \Delta b^{(s)}. \qquad (20)$$

In other words, $b^{(s)}(t) = \{b_i^{(s)}(t): i=1,...,N\}$ is the sinogram that differs from $b(t-1)$ by only one count in a bin having index s, i.e.

$$b_i^{(s)}(t) = \begin{cases} b_i(t-1),\ i=1,...,s-1 \\ b_i(t-1)+1,\ i=s \\ b_i(t-1),\ i=s+1,...,N \end{cases} \qquad (21)$$

The respective solution given by (17) for $b^{(s)}(t)$

$$\hat{f}^{(s)}(t) = \tilde{P}^+ b^{(s)}(t) \qquad (22)$$

may be found as follows:

$$f_k^{(s)}(t) = \sum_{j=1}^{N} \tilde{p}_{kj} b_j^{(s)}(t) = \sum_{j=1}^{s-1} \tilde{p}_{kj} b_i(t-1) + \tilde{p}_{ks}\left[b_s(t-1)+1\right] + \sum_{j=s+1}^{N} \tilde{p}_{kj} b_i(t-1) = \\ = \sum_{j=1}^{N} \tilde{p}_{kj} b_i(t-1) + \tilde{p}_{ks};\ k = 1,...,M \qquad (23)$$

Thus, the new image estimate is simply a sum of the previous image estimate and one column of the matrix $\tilde{P}^+$, i.e.

$$f_k^{(s)}(t) = f_k(t-1) + \tilde{p}_{ks},\ k = 1,...,M\ . \qquad (24)$$

The event-by-event reconstruction process using the regularized pseudo-inverse matrix is illustrated schematically in Fig. 1. A blank image can be used to initialize reconstruction $f = \{f_k = 0;\ k = 1,...,M\}$ and (24) is applied afterwards to obtain the current image estimate in real-time by adding one column of $\tilde{P}^+$ to the previous image estimate. The column index corresponds to the index of the tube of response where the last event was registered. The resulting image estimate is the same as the TSVD solution given by (11).

Solution involving the system matrix SVD and a different regularisation approach (independent of the data) can be obtained via updating as well using (24) and taking into account that the entries of the regularized pseudo-inverse $\tilde{P}^+$ in this case will be given by

$$\tilde{p}_{kj} \equiv \sum_{i=1}^{M} v_{ki}\, g(\mu_i)\, u_{ji};\ k = 1,...,M;\ \ j = 1,...,N\ . \qquad (25)$$

The proposed approach is general and can be applied to any estimation problem in which data are obtained by counting or by incremental measurements and that is amenable to the solution exploiting a (pseudo-)inverse matrix. As mentioned, object independent factors such as a spatially varying system response are already explicitly included in the elements of $\tilde{P}^+$, while other factors such as the normalization for detector

efficiency and the object-dependent corrections can be accounted for by multiplying the columns of $\tilde{P}^+$ with the correction factors. With the proposed real-time image reconstruction approach, such data corrections can be applied on the fly. To do this, the following update equation would be used instead of (24):

$$f_k^{(s)}(t) = f_k(t-1) + C_s \times \tilde{p}_{ks},\ k = 1,...,M\ . \quad (26)$$

Here C_s is the correction factor assigned to the tube-of-response s.

In a similar manner, random coincidences can be corrected for by the subtraction of one column of the matrix $\tilde{P}^+$ (scaled with C_s if corrections on the fly are performed) from the previous image estimate, or simply by skipping the next image update given by (24) or (26) if a random coincidence has been registered (but not corrected for) in the same tube of response.

IV. Materials and Methods

PET data were acquired with the Sherbrooke high-resolution animal tomograph [31]. A prototype real-time acquisition and reconstruction system is being developed in Sherbrooke [32], [33] but is not yet operational. Therefore, conventional data acquisition was used in this study and processing was started when a scan was complete. A phantom of 110 mm diameter made of Lucite and having holes of diameters 2, 3.4, 6.7, 9.7, 13, 15.8, 20.3, and 22.7 mm located on a circumference at a distance of 28 mm from the center of the phantom was filled with a solution of ^{18}F and scanned. A total of 9.5×10^5 coincident events were recorded with the plane used in the reconstruction experiments.

The acquired sinogram was rearranged with a filter using a random number

generator to mimic a sequence of events, as it would have been produced in the list-mode acquisition. A "list-mode" sequence of single events yielding the acquired sinogram was produced as a result. Another "list-mode" sequence has been produced to mimic the acquisition with a bank of detectors registering counts at a single angle only, yielding, after full rotation, a sequence of all angles attainable with the full detector ring.

The matrix P that takes into account the spatially varying system response was computed for the Sherbrooke animal PET scanner and image grids of 64×64 and 96×96 pixels. This step involving system modeling is crucial for adequate representation of the continuous distribution $f(x)$ with the chosen set of expansion functions. In this work, analytic functions based on linear attenuation of γ-rays in a detector array mapped onto the grid of square pixels were used [34]. The details of numerical SVD computation and the systematic singular value spectrum truncation approach were described elsewhere [14]. Corrections for detector efficiency, random counts, and attenuation were applied according to (26).

The expected pixel-wise standard deviation in the image estimate was calculated for TSVD as:

$$\sigma\left(\hat{f}_i;T\right)=\sqrt{\sum_{j=1}^{T}\left(\frac{V_{ij}}{\mu_j}\right)^2},\ i=1,...,M\ . \qquad (27)$$

V. NUMERICAL EXPERIMENTS

The progress of the phantom image reconstructed with the proposed incremental TSVD approach and the complete ring of detectors is shown in Fig. 2. Low image statistics results in poor images as expected, but as the number of total counts grows the

image estimate improves gradually. Finally, in the last few frames, acceptable estimate of the radioactivity map are observed. If such an image sequence is made available during acquisition, it could be used to decide when a scan might be terminated.

Images obtained with the proposed method and the (simulated) rotating banks of detectors acquiring data at one projection angle at a time are shown in Fig. 3. This sequence of images has a very distinct appearance. These results with the limited number of angles confirm that the proposed method might be useful with rotating gamma cameras as well.

The expected standard deviation map with the TSVD reconstruction was calculated using (27). It was computed for several truncation levels and for the grids of 64×64 and 96×96 pixels. For these estimates to be meaningful, the measured data have to satisfy the i.i.d. normal error assumptions. The radial cross-sections of the two grids are shown in Fig. 4 and Fig. 5, respectively. The calculated maps exhibit rotational symmetry and are, therefore, fully described by the presented profiles. These charts give an idea of the baseline of uncertainties that are expected in the image estimate, since these "intrinsic" standard deviations will be increased by the statistical uncertainty calculated from the measured data. Note the progressive growth of the standard deviation as the truncation index T grows. This illustrates the effects of increased inverse problem ill conditioning and confirms the necessity of regularisation [14]. The values of the standard deviation appear large, e.g. they fall in the range of 4 to 7 counts (for the 64×64 image and the optimal truncation level of T=822) as shown in Fig. 4. However, when applied to the example presented in Fig. 2 and compared to the estimated number of counts per pixel within the six largest hot spots of the phantom (which is in the range of 6000 to 8000 in

the last frame of Fig. 2), it is clear that the relative standard deviation is not high. Estimated value of the standard deviation is lower for TSVD reconstructed images representing the same FOV but having a larger number of pixels, as we see by comparing Fig. 4 and Fig. 5. This is not surprising since for the same input data the estimated absolute pixel values are usually higher in images of lower dimensions, which compensates for higher estimated variance.

VI. Discussion

Image reconstruction performed using the proposed algorithm can allow for the instant visualization of an image estimate while a patient is being scanned. This would make early identification of problems related to data acquisition easier. Subject positioning would be facilitated, thus ensuring that the desired region-of-interest (ROI) is completely in the FOV. This is particularly useful for scanners with limited axial fields of view, such as the original microPET [35] and the Sherbrooke animal PET scanner [31]. If one has an image online, one can stop scanning as soon as the data statistics are sufficient to provide answers to given questions, or one can extend data acquisition as needed. One result might be increased patient throughput.

All image reconstruction methods make a number of assumptions on the nature of the data and, hence, yield emission density estimates having predictable properties only with the data satisfying (at least approximately) those assumptions. FBP is a deterministic method and is derived ignoring the statistical data nature, which becomes a clear disadvantage in the case of emission image reconstruction. FBP employs low-pass filtering, trading resolution for improved image variance. We note that the FBP algorithm can be converted into matrix form and be adapted to process individual events.

Doing this would yield essentially the same algorithm as the one proposed in this paper, with the exception of the content of the matrix $\tilde{P}^{+}$. The columns of the matrix $\tilde{P}^{+}$ for the FBP case would approximate the inverse Radon transform and would be obtained by calculating the backprojection of a single filter function (in spatial domain representation) properly centered on a given line-of-response. The precursor of this procedure was suggested in [15], but pre-calculation of the backprojected filter functions was not considered, since sufficient data storage was not available 20 years ago. Emphasis was put on developing hardware for fast online backprojection instead [20]-[22]. The equivalence of FBP and backprojection filtering was utilized in the proposed implementations [16,19,22]. However, the usefulness of the latter approach is limited, as the image has to be filtered in the Fourier domain every time prior to display. Pre-computing the intersections of the lines-of-response with voxels used by Di Sciascio *et al.* [18,19,22] reduces the amount of online computation. We pre-compute the system matrix as well but our approach based on modelling of the detector response function [34] is more general than the Radon transform model exploited with the Fourier based techniques. Application of the system model in real time would be challenging with the previously described "real-time" image reconstruction techniques since it has to be limited to rebinning the data. This would involve additional computation if FBP adaptation was used for event-by-event reconstruction. The image reconstruction method based on the system matrix SVD, on the other hand, has the ability to include complex system models, if it utilizes an appropriately pre-computed system matrix. The resulting image is free of spatial distortions if the spatially varying system response has been accounted for [34].

Regularisation with truncation of the singular value spectrum implies that the calculated solution is limited primarily by the scanner and its ability to acquire data having high information content, the latter reflected in the respective singular value spectrum. This may result in sub-optimal image reconstructions with arbitrary (low count) data sets as noted by Demoment [36], but it becomes an additional safeguard during decision making when image is monitored in real-time. One should keep in mind that the goal of emission scanning is to obtain a *definitive* answer to a given question, be it diagnostic or research. It would be naive to expect a meaningful estimate of the unknown emission density based on a few counts with any reconstruction technique. Whether there exist a minimum number of counts that would be sufficient for a given situation remains unknown. The authors are not aware of a universal criterion that would be robust in practice with arbitrary emission densities, except "the more counts the better". The proposed method utilizing TSVD is capable of real-time image reconstruction and provides a way of monitoring data statistics sufficiency.

The goals of fast (but safe) scanning and image reconstruction integrity are always in conflict. One could introduce some a priori assumptions or constraints on the estimated emission map that would yield a solution with favorable properties, but the quantitative accuracy would be questionable unless the data outweigh the constraints. With the proposed real time TSVD reconstruction, poor images with low total counts would improve gradually yielding feasible estimates as the data statistics tends toward the normal distribution. More robust image estimation could be achieved by taking into account the variance of the error distribution derived from the data. This leads to the weighted least-squares method.

It is worth noting that the computational expenses of image updating for *a single plane* will be the same for a small animal scanner as for a whole-body scanner as long as the same image grid size is used and the total counting rate is the same. Only the size of the matrix to be stored in the fast random access memory (RAM) would be different, due to the differences in the total number of tubes-of-response involved. Examples of the system matrix sizes for three different image sizes and a single plane of the Sherbrooke animal PET scanner and of the ECAT EXACT HR+ scanner are given in Table I. It is assumed that matrix entries are represented with the standard four-byte floating-point numbers. The square grid of pixels is assumed truncated so as to leave only relevant pixels within the circular FOV in order to eliminate the corners of the square image from solution computation [34]. Thus, only about 80% of the pixels contained within the useful FOV are used during calculations. The size of RAM (in bytes) required for the data segment during SVD computation would be higher than the number reported in Table I and could be easily calculated as

$$D = M \times (N + M + 1) \times X , \qquad (28)$$

where X is the number of bytes used for the floating point number representation. Somewhat higher amounts of RAM than given by these estimates will be used in real implementations since additional RAM would be necessary for bookkeeping purposes.

A simple way to reduce the data storage requirements and the number of calculations would be to reconstruct a ROI instead of the whole FOV. An efficient way of reducing the size of the pseudo-inverse matrix would be to use natural pixels for image representation [37], which would yield the complete rotational symmetry of the system matrix in 2-D and most likely of the pseudo-inverse matrix as well. If that is the case,

then only the data set corresponding to a single projection angle could be stored and the rest may be restored online easily. This complete rotational symmetry is lacking with conventional square pixels, but even then there are a few symmetries that could be exploited [38]. Having the image estimate expanded in another set of pixels (voxels) would require remapping onto the conventional grid for image display.

Needless to say real-time reconstruction could be followed by a more conventional offline reconstruction technique to confirm or verify the results of the online reconstruction if need arises and the complete data set is available. However, the proposed reconstruction technique based on the regularized pseudo-inverse of the system matrix has the potential to outperform the conventional FBP in speed, image quality, and quantitative accuracy [14] as long as the computational resources used for image reconstruction are sufficient to accommodate the former.

TSVD image reconstruction is linear, i.e. yields images that are the linear combinations of the data. This is an advantage if image quantification is required. For example, a challenging issue with ML-EM is the non-linear convergence rates when recovering high and low frequency image components. As a result, images of questionable quantitative accuracy are obtained, unless a sufficiently high number of iterations is performed [39] and the impact of the regularisation approach is known exactly. To date, the high computational cost has limited practical application of iterative techniques, in spite of their ability to produce images with higher contrast and improved variance for data sets with low total counts.

This paper is concerned primarily with image reconstruction for PET, but the proposed reconstruction algorithm, as given in section III, is valid for SPECT as well. In

the case of SPECT, a specific challenge would be the calculation of the system matrix based on the appropriate system model. However, the usefulness of the proposed real time method for SPECT is not as clear as in PET, since the set of projections remains incomplete until the end of data acquisition. This can be overcome by adjusting the acquisition protocol, e.g. by speeding-up the rotation to make several full turns instead of a single one, or by increasing the angular steps at which measurements are performed and completing the data set during additional rotations. Nevertheless, the matrix pseudo-inverse approach may prove worthwhile since its qualitative performance is not worse than that of the conventional FBP, which is currently the workhorse of computerized tomography.

Finally, we note that sequential projection data acquisition in the case of rotating PET tomographs or SPECT gamma cameras lends itself to another image reconstruction approach referred to as concurrent reconstruction [40], [41]. The image estimation may be started before the data collection is complete in these special cases. Therefore, reconstructed images are obtained earlier than with conventional reconstruction protocols. However, the described TSVD technique is fast whether it is performed in the event-by-event scheme as proposed in this paper or by simple matrix multiplication [14].

VII. Conclusion

A novel event-by-event linear image reconstruction based on the regularized pseudo-inverse of the system matrix has been proposed. The algorithm provides a fast way to update an image potentially in real time. It also allows for the possibility of instant visualization of a radioactivity distribution during PET data acquisition. Current image

estimates are updated with one column of the regularized pseudo-inverse matrix to account for the next registered event. The process involves just one vector-to-vector addition in the simplest case, or one-scalar-to vector multiplication followed by the vector addition if data corrections are performed on the fly. The computed image estimate is the minimum-norm least-squares solution of the regularized inverse problem. The quality of the image improves gradually as more total counts are acquired. The proposed reconstruction method provides a way to quickly check subject positioning and to monitor data statistics sufficiency. Limited-angle tomography and tomography using non-traditional detection geometries might be handled using the described real-time image reconstruction approach as well.

VIII. REFERENCES

[1] P. Cahill and L. M. Blau, "The preliminary application of a matrix method for radionuclide imaging", *J. Nucl. Med.*, vol. 11, no. 10, pp. 613-615, 1970.

[2] T. F. Budinger and G. T. Gullberg, "Three-dimensional reconstruction in nuclear medicine emission imaging", *IEEE Trans. Nucl. Sci.*, vol. 21, no. 3, pp. 2-20, June 1974.

[3] 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.

[4] Y. S. Shim and Z. H. Cho, "SVD pseudoinversion image reconstruction", *IEEE Trans. ASSP*, vol. 29, no. 4, pp. 904-909, 1981.

[5] M. Defrise and C. De Mol, "A note on stopping rules for iterative regularization methods and filtered SVD," in *Inverse problems, an interdisciplinary study* (P.C. Sabatier ed.), Advances in Electronics and Electron Physics, supplement 19, Academic Press, 1987, pp. 261-268.

[6] M. Bertero, C. De Mol, and E. R. Pike, "Linear inverse problems with discrete data. II: Stability and regularisation", *Inverse Problems*, vol. 4, no. 3, pp. 573-594, 1988.

[7] P. C. Hansen, "Numerical tools for analysis and solution of Fredholm integral equations of the first kind", *Inverse Problems*, vol. 8, pp. 849-72, 1992.

[8] P. C. Hansen and D. P. O'Leary, "The use of the L-curve in the regularization of discrete ill-posed problems", *SIAM J. Sci. Comput.*, vol. 14, no. 6, pp. 1487-1503, November 1993.

[9] P. Xu, "Truncated SVD methods for discrete linear ill-posed problems", *Geophys. J. Int.*, vol. 135, no. 2, pp. 505-514, 1998.

[10] M. F. Smith, C. E. Floyd, Jr., R. J. Jaszczak, and R. E. Coleman, "Reconstruction of SPECT images using generalized matrix inverses", *IEEE Trans. Med. Imag.*, vol. 11, no. 2,

pp. 165-175, 1992.

[11] G. T. Gullberg, Y.-L. Hsieh, and G. L. Zeng, "An SVD reconstruction algorithm using a natural pixel representation of the attenuated Radon transform", *IEEE Trans. Nucl. Sci.*, vol. 43, no. 1, pp. 295-303, February 1996.

[12] M. F. Smith, "Generalized matrix inverse reconstruction for SPECT using a weighted singular value spectrum", *IEEE Trans. Nucl. Sci.*, vol. 43, no. 3, pp. 2008-2017, 1996.

[13] J. R. Baker, T. F. Budinger, and R. H. Huesman, "Generalized approach to inverse problems in tomography: Image reconstruction for spatially variant systems using natural pixels", *Critical Rev. Biomed. Eng.*, vol. 20, no. 1/2, pp. 47-71, 1992.

[14] V. V. Selivanov and R. Lecomte, "Fast PET image reconstruction based on SVD decomposition of the system matrix", *IEEE Trans. Nucl. Sci.*, vol. 48, no. 3, pp. 761-767, June 2001.

[15] J. A. McIntyre, "Computer assisted tomography without a computer", *IEEE Trans. Nucl. Sci.*, vol. NS-28, no. 1, pp. 171-173, 1981.

[16] E. A. Philippe, N. Mullani, W. Wong, and R. Hartz, "Real-time image reconstruction for time-of-flight positron emission tomography (TOFPET)", *IEEE Trans. Nucl. Sci.*, vol. NS-29, no. 1, pp. 524-528, 1982.

[17] R. Clack, D. Townsend, and M. Defrise, "An algorithm for three-dimensional reconstruction incorporating cross-plane rays", *IEEE Trans. Med. Imag.*, vol. 8, no. 1, pp. 32-42, 1989.

[18] E. Di Sciascio, R. Guzzardi, and D. Marino, "Proposal of a real-time reconstruction processor for 3-D PET", *Conference Record of 1992 IEEE Medical Imaging Conference*, vol. 2, pp. 921-924, 1992.

[19] E. Di Sciascio, A. R. Manni, R. Guzzardi, L. Leopizzi, and F. Resta, "Event by event 3D-PET reconstruction algorithm for a dedicated hardware architecture: Preliminary results",

Conference Record of 1994 IEEE Medical Imaging Conference, vol. 3, pp. 1192-1196, 1994.

[20] R. Hartz, E. Philippe, W.-H. Wong, K. Yerian, and N. Mullani, "Implementation of real-time image reconstruction", 1982 Workshop on Time-of-Flight Tomography, *IEEE Proc.* CH1791-3/82, pp. 153-160.

[21] R. Hartz, D. Bristow, and N. Mullani, "A real-time TOFPET slice-backproject engine employing dual AM29116 microprocessors", *IEEE Trans. Nucl. Sci.*, vol. 32, no. 1, pp. 839-842, February 1985.

[22] E. Di Sciascio, A. R. Manni, and R. Guzzardi, "Design of an architecture for real-time 3D PET imaging", *Real-Time Imaging*, vol. 4, pp. 255-262, 1998.

[23] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography", *IEEE Trans. Med. Imag.*, vol. 1, no. 2, pp. 113-122, October 1982.

[24] P. R. Bevington and D. K. Robinson, *Data reduction and error analysis for the physical sciences*, 2nd ed., WCB/McGraw-Hill, 1992, p. 63.

[25] J. A. Fessler, "Penalized weighted least-squares image reconstruction for positron emission tomography", *IEEE Trans. Med. Imag.*, vol. 13, no. 2, pp. 290-300, 1994.

[26] J. Nuyts, C. Michel, and P. Dupont, "Maximum-likelihood expectation-maximization reconstruction of sinograms with arbitrary noise distribution using NEC-transformations", *IEEE Trans. Med. Imag.*, vol. 20, no. 5, pp. 365-375, May 2001.

[27] G. Strang, *Linear algebra and its applications*, 2nd ed., Harcourt Brace Jovanovich, 1980, pp. 137-145.

[28] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes in C: The art of scientific computing*, 2nd ed., Cambridge University Press, 1992, pp. 676-679.

[29] C. R. Vogel and J. G. Wade, "Iterative SVD-based methods for ill-posed problems", *SIAM*

J. Sci. Comput., vol.15, no.3, pp.736-754, May 1994.

[30] V. Selivanov, G. Léger, and R. Lecomte, "Real-time image reconstruction for emission tomography", *J. Nucl. Med.*, vol. 42, no. 5, p. 138P, May 2001 Supplement.

[31] R. Lecomte, J. Cadorette, S. Rodrigue, D. Lapointe, D. Rouleau, M. Bentourkia, R. Yao, and P. Msaki, "Initial results from the Sherbrooke avalanche photodiode positron tomograph", *IEEE Trans. Nucl. Sci.*, vol. 43, no. 3, pp. 1952-1957, 1996.

[32] M. D. Lepage, G. Léger, J. Cadorette, S. Rodrigue, and R. Lecomte, "FPGA/DSP-based coincidence unit and data acquisition system for the Sherbrooke animal PET scanner", *2000 IEEE Nuclear Science Symposium Conference Record CD*, IEEE Catalog 00CH37149C, Topics: Trigger and Data Acquisition, Session: Data Acquisition, pp. 12.173 – 12.175, 2001.

[33] M. D. Lepage, J.-D. Leroux, V. Selivanov, J. Cadorette, and R. Lecomte, "Design of a prototype real-time image reconstruction system for PET imaging", *2001 IEEE Nuclear Science Symposium and Medical Imaging Conference*, San Diego, CA, 4-10 November 2001.

[34] V. V. Selivanov, Y. Picard, J. Cadorette, S. Rodrigue, and R. Lecomte, "Detector response models for statistical iterative image reconstruction in high resolution PET", *IEEE Trans. Nucl. Sci.*, vol. 47, no. 3, pp. 1168-1175, June 2000.

[35] S. R. Cherry, Y. Shao, R. W. Silverman, K. Meadors, S. Siegel, A. Chatziioannou, J. W. Young, W. F. Jones, J. C. Moyers, D. Newport, A. Boutefnouchet, T. H. Farquhar, M. Andreaco, M. J. Paulus, D. M. Binkley, R. Nutt, and M. E. Phelps, "MicroPET: A high resolution PET scanner for imaging small animals", *IEEE Trans. Nucl. Sci.*, vol. 44, no. 3, pp. 1161-1166, June 1997.

[36] G. Demoment, "Image reconstruction and restoration: Overview of common estimation structures and problems", *IEEE Trans. ASSP*, vol. 37, no. 12, pp. 2024-2036, December

1989.

[37] M. H. Buonocore, W. R. Brody, and A. Macovski, "A natural pixel decomposition for two-dimensional image reconstruction", *IEEE Trans. Biomed. Eng.*, vol. 28, no. 2, pp. 69-78, 1981.

[38] B. Desjardins and R. Lecomte, "Parallel approach to iterative tomographic reconstruction for high resolution PET imaging", *1997 IEEE Nuclear Science Symposium Conference Record*, ed. O. Nalcioglu, IEEE Catalog 97CH36135, Vol. II, 1998, pp. 1551-1555.

[39] V. V. Selivanov, D. Lapointe, M. Bentourkia, and R. Lecomte, "Cross-validation stopping rule for ML-EM reconstruction of dynamic PET series: Effect on image quality and quantitative accuracy", *IEEE Trans. Nucl. Sci.*, vol. 48, no. 3, pp. 883-889, June 2001.

[40] J. K. Brown, B. H. Hasegawa, and T. F. Lang, "Iterative concurrent reconstruction algorithms for emission computed tomography", *Phys. Med. Biol.*, vol. 39, no. 7, pp. 1113-1132, July 1994.

[41] C. Comtat, R. W. Dobinson, M. Egger, A. H. Scheurer, F. F. de Jong, C. Joseph, C. Morel, A. J. E. van Nieuwenborg, J. L. Pagès, D. W. Townsend, and M. I. Williams, "On-line parallel processing for a rotating positron tomograph operated in 3D mode", *Nucl. Instr. Meth. Phys. Res. A*, vol. 351, pp. 527-533, 1994.

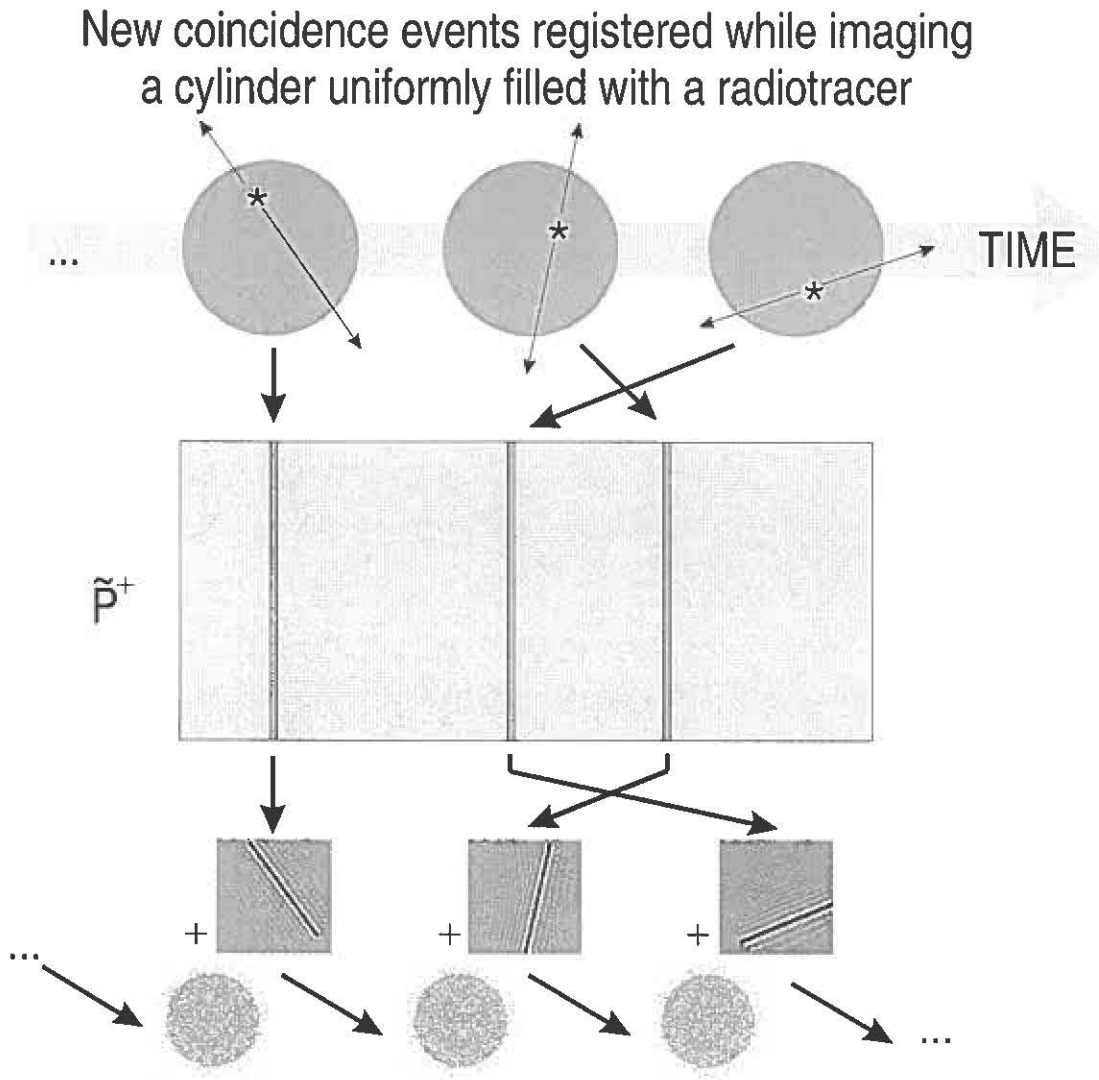

Fig. 1. Schematic of the event-by-event reconstruction process using the regularized pseudo-inverse matrix. The incoming coincidence events trigger image update. The update is performed by adding a single column of $\tilde{P}^+$ to the current image estimate. Tubes of response have unique indexes, and the index of the tube of response where the event was registered is also the index of a column of $\tilde{P}^+$ that has to be used for the update.

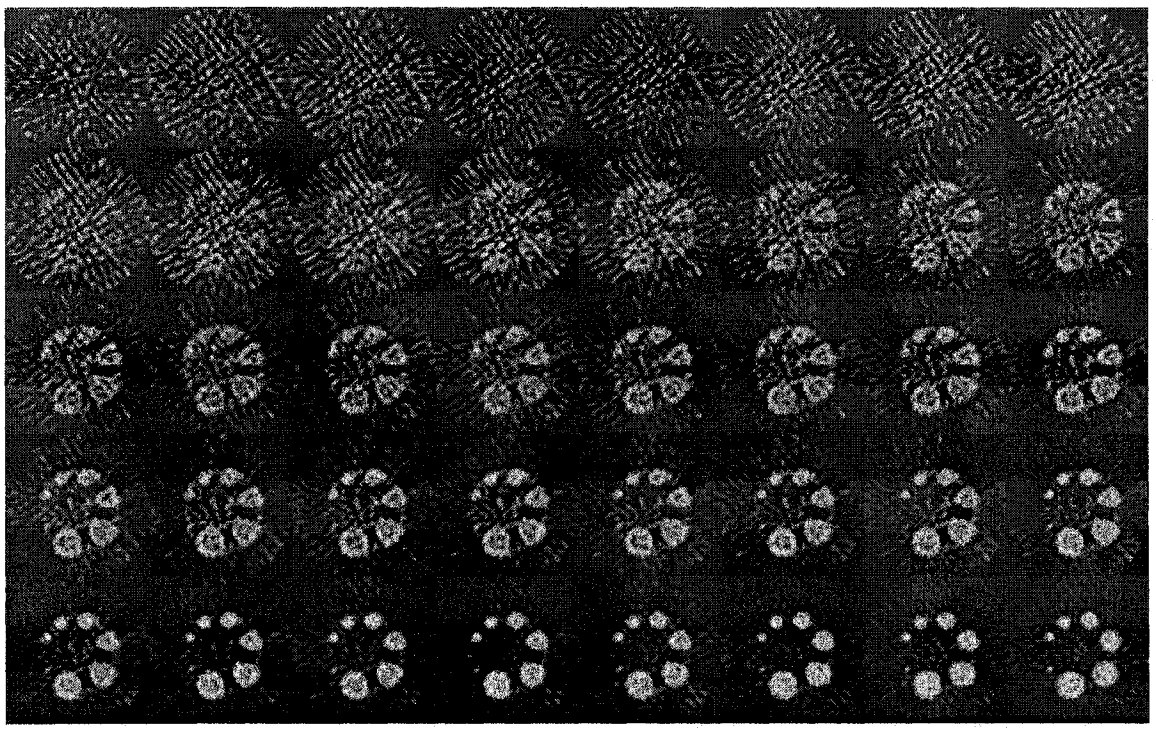

Fig. 2. Sequence of images demonstrating the use of the incremental TSVD reconstruction for a system having a complete ring of detectors. From left to right and from top row to the bottom, consecutive estimates of the radioactivity distribution are shown as approximately 23800 additional coincidences were registered (64×64 pixel images). Data acquired with the Sherbrooke animal PET scanner were used in this experiment. See section IV for the description of data acquisition and processing.

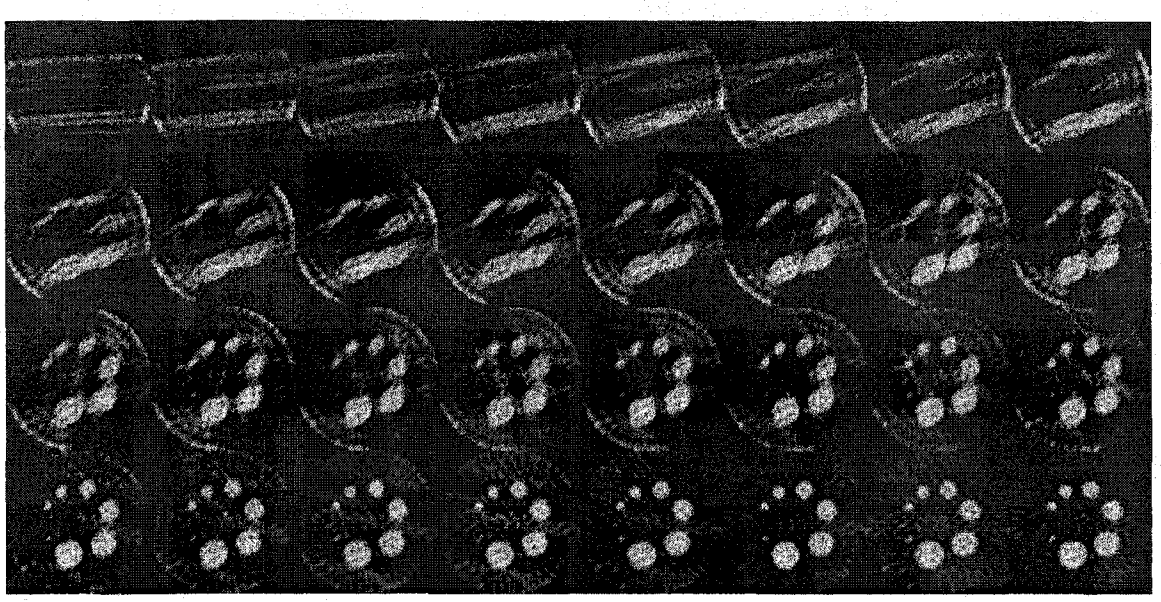

Fig. 3. Sequence of images demonstrating the use of the incremental TSVD reconstruction for rotating banks of detectors, an incomplete ring, acquiring data at one projection angle at a time. From left to right, consecutive estimates of radioactivity distribution are shown when 8 additional adjacent projection angles (out of 256 angles total) having approximately 29200 additional counts are included in the reconstructed image (64×64 pixel images). See section IV for the description of data acquisition and processing that yielded the list mode data characteristic of a scanner with rotating banks of detectors.

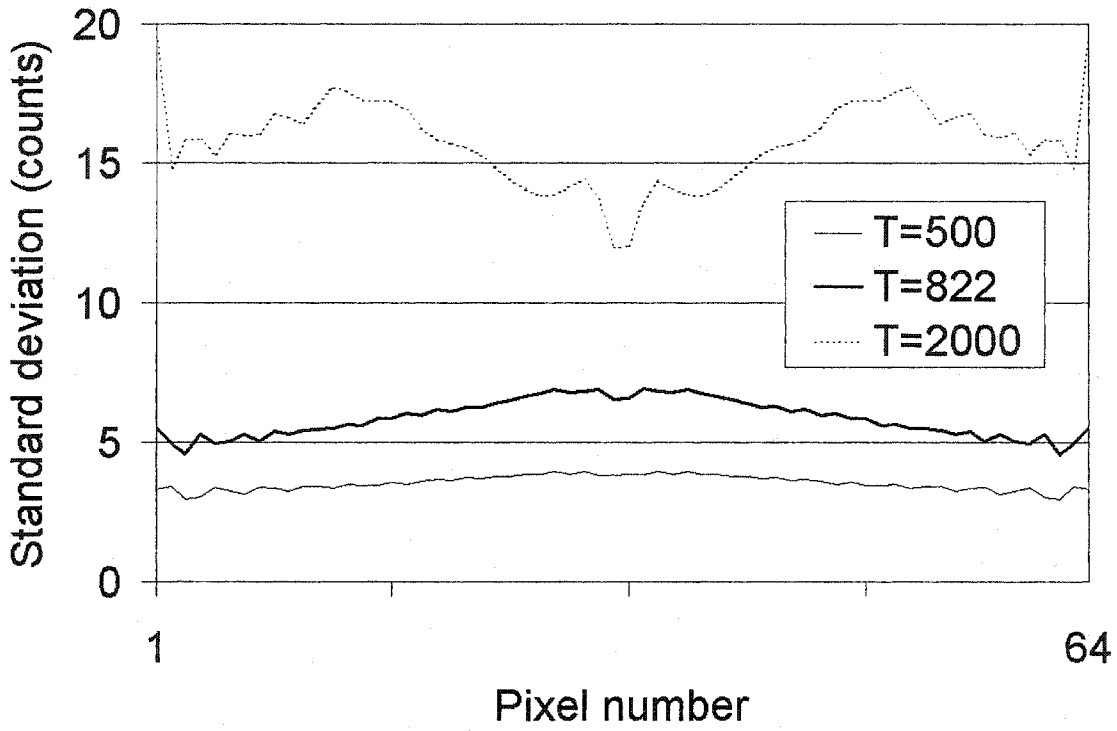

Fig. 4. Radial profiles through the expected standard deviation maps with the TSVD reconstruction for three different truncation levels and the image of 64×64 pixels. Data have to satisfy i.i.d. normal error assumptions for these estimates to be valid. The calculated maps exhibit rotational symmetry and are, therefore, fully described by the presented profiles. Note the progressive growth of the uncertainty in pixel value estimation as the truncation index T grows. T=822 was found optimal based on the spatial resolution analysis reported earlier [14].

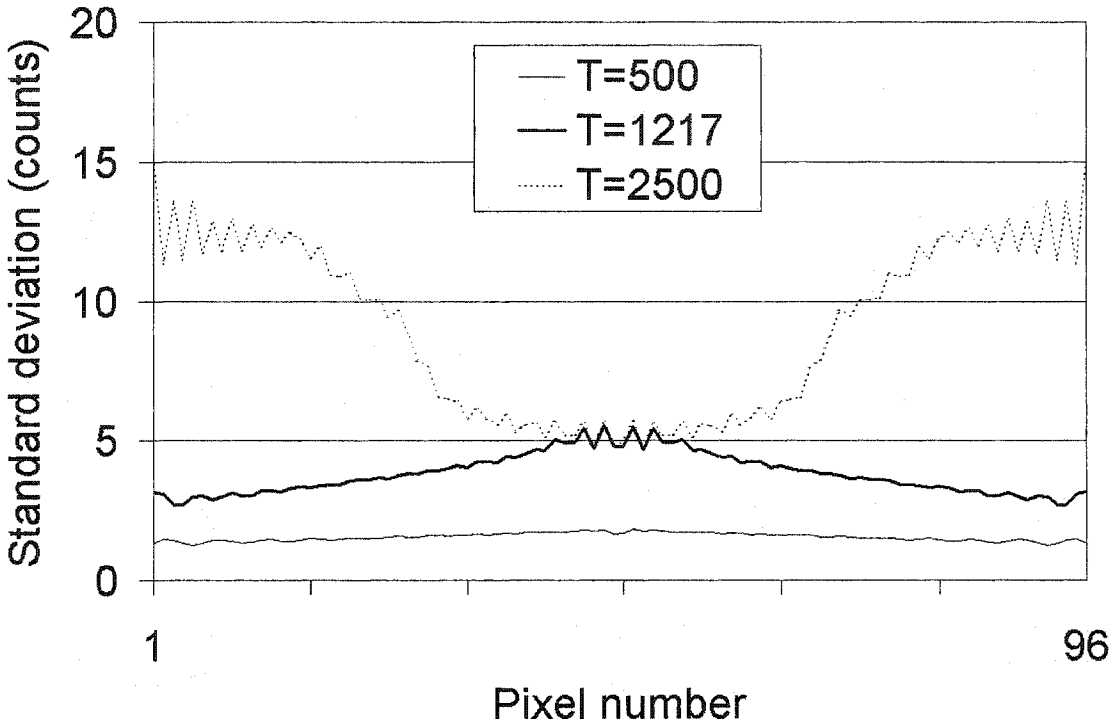

Fig. 5. Radial profiles through the expected standard deviation maps with the TSVD reconstruction for three different truncation levels and the image of 96×96 pixels. T=1217 was found optimal based on the spatial resolution analysis reported previously [14]. Comparable level of ill conditioning is attained with higher T for images representing the same FOV but having larger number of pixels. This yields lower variance for a given pixel when comparing two optimal truncation levels for different image grids (see estimates for T=822 shown in Fig. 3 and T=1217 in this figure).

TABLE I

EXPECTED SIZE OF THE (REGULARIZED) PSEUDO-INVERSE MATRIX FOR VARIOUS SINOGRAM SIZE (FOUR-BYTE FLOATING POINT NUMBER REPRESENTATION IS ASSUMED, CORNERS OF THE SQUARE IMAGE GRID ARE OMITTED)

Sinogram size / Image size	Sherbrooke animal PET scanner		ECAT EXACT HR+	
	256 × 32	513 × 32	288 × 144	576 × 144
64×64	104.1 MB	208.7 MB	527.1 MB	1.0 GB
96×96	231.3 MB	463.4 MB	1.1 GB	2.3 GB
128×128	409.5 MB	820.6 MB	2 GB	4.1 GB

Chapter 5. Discussion

Ill-posed problems are a topic of an interdisciplinary interest. The non-exclusive list of applications includes mathematical physics, geodesy, geomagnetism, geophysics, seismology, non-destructive evaluation and medical imaging. The theory developed for certain applications could be relevant in other fields. However, there are issues specific to the application and crucial for success. PET is similar to other tomographic imaging techniques in many ways, but to reach its full potential and to extract maximum information from the data, PET has to use precise, yet practical, algorithms and techniques.

In the present work we were concerned primarily with the image reconstruction problems arising in high resolution PET, which is currently represented with state-of-the-art animal scanners. Animal scanners are the instruments dedicated to imaging small animals and are at the forefront of radiopharmaceutical and clinical research (Hichwa, 1994; Tornai *et al.*, 1999). Several issues have been covered that range from the direct problem analysis to the robust solution of the inverse problem allowing near real-time image reconstruction.

5.1. Detector response as a system model

The system response is a combination of factors. The geometric system design together with the physical properties of the scintillation and detector packing materials define the intrinsic point spread function of the system. These features fundamentally limit the tomograph spatial resolution. It is essential to model the system, hence solve the

direct problem, to set an adequate basis for the inverse problem solution. Reconstruction methods using the system matrix have an advantage over the transform methods as the model can be incorporated into the precomputed matrix instead of rearranging or filtering the sinogram on-the-fly. The time to compute a matrix is not a primary limiting factor, since such computation is performed once and the matrix can be reused until the system configuration changes, which is not a frequent occurrence if a stationary detection system is used. Thus, an accurate model of the system response, which would require hours or, maybe, days of computation can be used instead of data rebinning techniques. A program in C used to compute the system matrix and perform ML-EM reconstruction is given in Appendix B.

The absence of the system response model in the basic transform methods necessitates additional online computing in order to perform a series of corrections for geometric image distortions that result from the mispositioning of the TORs relative to their actual physical presence. A consequence of several overlaid approximations is the compromised geometric accuracy that contributes e.g. to the difficulties of multi-modality image registration (Kops *et al.*, 1999). The suboptimal representation of fine detail off the FOV center, i.e. the decreased image resolution, is another possible effect of image distortion.

The possible refinement of the system response model based on the analytical DRF, used for the system matrix calculation in the present work, may be the inclusion of the models of detector scatter, positron range as well as the non-collinearity of the γ-rays. The same technique of weighting the intersection of the TOR with a pixel can be utilised to superimpose additional models. This would certainly broaden the TOR's spatial

coverage that would potentially entail additional computation overhead diminishing the system matrix sparseness. Potential benefits would be based on the accuracy and the limitations of those models as we have shown in [4.1].

5.2. To stop or not to stop...

In [4.2] we have found an interesting dependence that approximates the cross-validation stopping rule and makes it easier to obtain an image estimate belonging to the feasibility region. This technique can be very useful in static PET applications or when quantitation is not required since it tends to optimise the signal-to-noise ratio and yields visually appealing images. However, conventional tracer kinetic modeling should not be used with the CV stopping rule, because the basic assumption of constant measurement volume is violated with dynamic images having variable spatial resolution over time. The stopping rule attempts to mask a limitation that cannot be ignored. It is related to the effect of temporal data "binning". Changes in the intensity of radioactive decay in time are averaged within a frame if the data are assembled in a sequence of time frames. For an experimentalist, it is tempting to decrease frame duration in order to increase temporal resolution. However, it results in low count data as fewer events are counted during a shorter time interval given that all other experimental parameters are the same, and image estimation is affected when lowering total counts as statistical noise becomes dominant.

Dynamic PET studies are based on sequences of images that are currently reconstructed frame-by-frame, but treating the data as a 4-D spatio-temporal problem has, probably, a brighter future. Several approaches have been investigated lately. A MAP algorithm was derived applying the 4-D prior (Lalush *et al.*, 1997). A PWLS spatio-

temporal criterion was optimised in (Wernick *et al.*, 1999), with the Karhunen-Loève (KL) transformation first applied along the time axis, reconstruction performed in the KL domain, and then performing an inverse KL transformation. The KL transform is tightly related to the eigenvector problem applied to the covariance matrix of the time behaviour of the image sequence. The resulting eigenvalue analysis yields a regularisation problem analogous to singular value spectrum filtering. These approaches have been compared to 4-D Fourier filtering in (Lalush *et al.*, 2000). Another promising approach is to employ the list-mode data, which are collected using an age-old acquisition mode (Snyder and Politte, 1983) that was reborn recently as it is more efficient for fully 3-D data acquisition. There are a few reports on continuous time image estimation with statistical iterative methods using list-mode data (Reader *et al.*, 1998; Nichols *et al.*, 1999). Conventionally, image reconstruction and kinetic modeling are performed sequentially. An alternative approach involving the direct calculation of parametric images was investigated recently, see e.g. (Matthews *et al.*, 1997; Meikle *et al.*, 1998).

An important reminder can be drawn from the experiments with the stopping rule. An image reconstruction algorithm does not create any information; it is simply a tool to recover information present in the data. The best an algorithm can do is to extract nearly all information, in the form of an image, but if the data do not contain much information, there is little an algorithm can do without additional guidance. If reliable external information was indeed available, then its inclusion as a prior would help, but in the absence of additional information two logical assumptions have to be utilised: the image is expected to be a (more or less) smooth function and the statistical errors are (approximately) Poisson. The cross-validation stopping rule does a perfect job in the

latter case. However, it is the responsibility of an experimentalist to obtain as much projection data, i.e. coincidence counts, as necessary to produce a reliable image estimate. Otherwise, subjective assumptions rather than objective data would dominate the image estimation process and the resulting image might be of little value.

5.3. Regularisation is the key

The problem of image reconstruction in PET is generally ill posed. A few assumptions could transform it to a well posed one, and expanding an image with finite series could be one of those. However, special attention has to be paid to find out if the problem remains severely ill-conditioned or not. The noise artefact, developing with the unconstrained ML-EM, is a well-known example of the undesirable effect of severe ill-conditioning. The ill-posedness of the PET inverse problem in general or the severe ill-conditioning in some discrete-to-discrete cases are the main reason for explicit regularisation. In the absence of the *a priori* information about the solution and the variance of error, the most attractive approach for choosing a regularisation parameter often employs cross validation in one form or another (Vogel, 1986; Coakley, 1991; Golub and von Matt, 1997).

The non-linearity of the iterative reconstruction techniques and resulting quantification issues are sometimes a drawback and a linear reconstruction approach may be preferable. Though it is believed that there is no closed form solution of the ML problem for the Poisson likelihood, the solution that stands closest has been shown in (Rockmore and Macovski, 1976). Its obvious limitation is the inability to handle ill-posed problems, which makes the proposed method impractical. An ideal test of problem

conditioning is via examination of the singular value spectrum of the system matrix. A direct approach to image reconstruction becomes possible when having the matrix SVD at hand. It is interesting to see the link between the MLE solution for Poisson data and the pseudo-inverse matrix, which was derived by Rockmore and Macovski. See equations (A9-A10) in (Rockmore and Macovski, 1976). Remember that the pseudo-inverse is easily obtained given matrix SVD.

Most regularisation methods may be represented as special cases of the approximate inverse studied in (Louis, 1996). TSVD is an approximation to the Moore-Penrose pseudo-inverse P^{+}. Regularisation with TSVD limits the dimensionality of the solution space. This is equivalent to projecting the solution onto a "significant" subspace spanned by the remaining right singular vectors. TSVD has a number of undesirable properties. It produces image estimates containing negative values, which do not have physical interpretation. The extension of TSVD solution incorporating the non-negativity of the true sought-for parameters has been suggested in (de Villiers *et al.*, 1999). Another limitation is the ringing artefact that becomes more obvious as fewer basis vectors are used in the solution; see figures 4a and 4b in [4.3]. Despite these inconveniences optimised TSVD appears to be a viable alternative at least to the conventional FBP and coupled with the proposed list-mode reconstruction described in [4.4] can be useful for a number of clinical tasks.

5.4. How prohibitive is the SVD computation?

The direct SVD computation is feasible with the relatively small 2-D problems. An example of the combined size of the SVD pieces is listed in Table I in [4.4]. The time to

perform matrix decomposition would increase in a non-linear fashion as well. It is possible to trade the size of the image grid for reducing the system matrix size to acceptable limits. Nevertheless, proposed list-mode reconstruction might be useful for real-time monitoring, even if the resulting image estimates would be less than optimal in terms of resolution and detailed distribution visualisation. Another important consideration is the matrix ill conditioning, which, combined with the finite precision of digital computing, leads to the accumulation of the numerical errors. Therefore, it is essential to increase the mantissa size, if the condition number is close to the precision limit of the digital number representation.

The real challenge is to obtain the matrix SVD for huge 3-D problems that arise with scanners dedicated to whole body imaging of humans. It is beyond doubt that further progress in computing hardware as well as parallel and distributed computing techniques would render such problems much easier. In the meantime, one has to employ some ingenuity to reduce the problem size and, hence, the amount of computation. One solution may be in the computation of partial SVD, i.e. finding singular vectors with the corresponding large singular values only, as suggested e.g. in (Vogel and Wade, 1994). Equipped with the knowledge of the singular value spectrum shape shown in figure 2 in [4.3], which is characteristic for optimal regularisation as we have observed, one can terminate the matrix decomposition as soon as the plateau of singular values has been recovered. Another approach discussed e.g. in (Fierro and Hansen, 1995) involves rank-revealing decompositions that allow obtaining approximations to the TSVD solution. On the other hand, if the matrix product approach to system modeling is utilised

(Mumcuoğlu *et al.*, 1997), then recent results of Golub and colleagues may be helpful for computing the SVD of a product (Golub *et al.*, 2000).

5.5. Real time PET imaging prospects

There have been several attempts to reach the ambitious goal of real-time image reconstruction. Fourier based techniques were the basis of these developments. Backprojection on an event-by-event basis has been utilised in (McIntyre, 1981; Hartz *et al.*, 1985). The high speed implementation of DFM has been addressed by developing capable hardware and a claim has been made that this DFM implementation was two orders of magnitude faster than convolution based algorithms (Brantner *et al.*, 1997). However, the use of a Fourier transform based algorithm yields all the drawbacks inherent to this reconstruction method. The associated difficulties of system modeling are hard, if possible, to overcome especially if the real-time implementation is the goal.

The accurate system model explicitly built into the system matrix as shown in [4.1] was the basis of our approach. TSVD is certainly a capable image reconstruction technique as we have shown in [4.3]. The observation that the same image estimate can be obtained incrementally with the list-mode data yields a very fast image reconstruction approach described in [4.4]. The benefit of performing the incremental image update proposed in [4.4] as opposed to just matrix to vector multiplication with TSVD [4.3] would be appreciated if the counting rate is lower than the threshold which can be derived for a particular system matrix and the speed of computing hardware. For example the number of updates x corresponding to individual events using equation (25)

in [4.4] that would produce the same computation load as the direct matrix to vector multiplication would be as follows:

$$x = \frac{M \times (D \times \chi_{\times} + [D-1] \times \chi_{+})}{M \times \chi_{+}} = D \times \left(\frac{\chi_{\times}}{\chi_{+}} + 1 \right) - 1, \tag{5.5.1}$$

where $\chi_{\times}$ and χ_{+} are the time necessary to perform a single operation of multiplication and addition, respectively. A similar estimate for the update performed according to (27) in [4.4] yields a more complicated relationship:

$$x = \frac{D \times \chi_{\times} + M \times (D \times \chi_{\times} + [D-1] \times \chi_{+})}{M \times (\chi_{\times} + \chi_{+})} = D \times \left(\frac{\chi_{\times}}{M \times (\chi_{\times} + \chi_{+})} + 1 \right) - \frac{\chi_{+}}{\chi_{\times} + \chi_{+}}. \tag{5.5.2}$$

The equations (5.5.1) and (5.5.2) give a rough estimate assuming that all operations are performed sequentially and the matrix multiplication is not optimised, e.g. operations are performed for all the TORs even if they had no counts, which could, in principle, be avoided. On the other hand an optimised implementation of the algorithm using a finely tuned data acquisition system serving data to the image reconstruction engine in real time would be a prerequisite for real-time imaging (Lepage *et al.*, 2001).

Statistical iterative image reconstruction techniques exploiting the availability of list-mode data have been proposed as well (Snyder and Politte, 1983; Reader *et al.*, 1998; Parra and Barrett, 1998; Nichols *et al.*, 1999). The resulting algorithms are not targeted for real-time implementations since there is a fundamental issue with the number of iterations, which we have explored with the help of a stopping rule in [4.2]. It should be remembered that there exists no single recipe for the fastest yet most accurate image reconstruction that would be equally useful in any situation. First and foremost, a

practical algorithm must provide a reliable basis for definitive answers to clinical questions. TSVD applied for real-time image monitoring can be useful in a number of situations, e.g. ability to address the issue of data sufficiency objectively cannot be overemphasized. It is needless to say that a post-acquisition image reconstruction method might be invoked to validate the findings of online estimation.

Conclusion

Several topics in PET image reconstruction have been explored. The following contributions have been made in the course of the work presented in this dissertation:

- A viable system matrix model has been developed that incorporated the analytic detector response function based on linear attenuation of γ-rays in a detector array. It has been demonstrated that the use of an oversimplified system model for the computation of the system matrix results in image artefacts.
- The dependence on total counts was modelled analytically which simplified the utilisation of the cross-validation stopping rule and accelerated reconstruction. It can be utilised instead of the original CV procedure for high-count projection data, when the CV yields reasonably accurate images. In a dynamic series with many low-count frames, the CV is not recommended for the whole series. However, the reconstruction of several high-count frames at the end of the series can help determine the minimum fixed number of iterations required to provide suitable images for ROI delineation without smoothing or filtering.
- A new regularisation methodology based on the SVD of the system matrix was proposed based on the spatial resolution analysis. A characteristic property of the singular value spectrum shape was found that revealed the relationship between the optimal truncation level to be used with the TSVD reconstruction and the optimal reconstructed image resolution.

- A novel event-by-event linear image reconstruction based on the regularised pseudo-inverse of the system matrix has been developed. The algorithm provides a fast way to update an image potentially in real time and allows, in principle, for the instant visualisation of the radioactivity distribution while the object is still being scanned. The computed image estimate is the minimum-norm least-squares solution of the regularised inverse problem.

Acknowledgements

I thank my father, Victor Selivanov, and my mother, Natalie Selivanova, for their inspiration and giving me a good start in life.

I am very grateful to my research director and dissertation supervisor, Professor Roger Lecomte, who introduced me into the exciting world of medical imaging and provided me with an opportunity to become a part of the PET imaging laboratory at the Université de Sherbrooke. I am indebted for his inspiring guidance and openness throughout the course of my studies and research in Sherbrooke.

My sincere thanks go to the co-authors of the articles that became a part of this dissertation for contributing their expertise to this project. These are Jules Cadorette, Serge Rodrigue, Yani Picard, David Lapointe, M'hamed Bentourkia, and, of course, Roger Lecomte.

I wish to express my gratitude to Germain Léger for his curiosity that sparked our discussion of the matrix-based image reconstruction and led to the development of the list-mode data reconstruction with TSVD.

Many thanks go to all members of the PET imaging laboratory in Sherbrooke, who have contributed in various ways, but especially to David Lapointe, Jules Cadorette, Catherine Pépin, Martin Lepage, and M'hamed Bentourkia, for their warm and friendly personalities that made our joint work enjoyable.

I gratefully acknowledge support from the Québec Ministry of Education through a scholarship that made my stay in Québec easier and much more productive.

Finally, I am very grateful to my wife, Svetlana Selivanova, and our two kids, Anastasia and Alexandre, for their support and numerous positive emotions that make my life worth living.

Appendix A

Are the data Poisson or Gaussian distributed after all?

It has been commonly assumed that a Poisson model is appropriate for PET data (Rockmore and Macovski, 1976; Snyder *et al.*, 1981; Lange and Carson, 1984; Vardi *et al.*, 1985). No doubt, this is true with ideal data. As far as the practice is concerned, the Poisson law may well be replaced by the normal distribution for some "real" data sets, as has been discussed in section 1.2.5. Fortunately, there are stronger grounds for this empirical assertion.

The theory of probability contains a remarkable result called the *central limit theorem* (CLT). CLT is a set of results unified by common principles, the in depth probabilistic analysis may be found e.g. in (Araujo and Giné, 1980). In a simple form exploited in mathematical statistics, the theorem states that the sum of a large number of independent observations from the same distribution has, under certain conditions, an approximate normal distribution. The normal approximation steadily improves as the number of observations increases. This has important implications for statistical PET data analysis.

For example, let η be a Poisson distributed r.v. having $E[\eta]=20$. It may be thought of as the sum Y of twenty random samples drawn from a Poisson distribution with the mean equal to 1. The following random variable,

$$W=\frac{(Y-20)}{\sqrt{20}} \tag{A.1}$$

can be characterized approximately with the standard normal distribution, i.e. $W \sim \mathrm{N}(0,1)$. In general, if Y has a Poisson distribution with the mean Λ, then the distribution of

$$W = \frac{(Y - \Lambda)}{\sqrt{\Lambda}} \tag{A.2}$$

is approximately $N(0,1)$ if Λ is sufficiently large (Hogg and Tanis, 1988). Thus, Poisson distributed random variables can be approximated with specially chosen Gaussian distributed random variables for image reconstruction with high total count data.

Appendix B

```
/*
 *   mlem-v3.c
 *
 *   9 July 1998    Selivanov Vitali
 *
 *   ML-EM image reconstruction (sinogram 32*257, image 128*128). A tube is
 *   the space between two detectors within their geometric boundaries. Any
 *   number of frames (n) in the sinogram.
 *
 *   Syntax: mlem-v3 NumberOfIterations < sinogram(n*32*257) > outfile
 *
 *   Compile: cc -O -o mlem-v3 mlem-v3.c -lm /usr/local/hips/lib/libhips.a
 */

#include <stdio.h>
#include <stdlib.h>
#include <math.h>
#include <string.h>
#include <sys/mode.h>
#include "/hosts/russia/batch/hipl_format.h"

/* Parameters of a PET ring and reconstructed image */

#define sqrtB           128          /* square root of total number of pixels */
#define BTotal          16384        /* total number of pixels in image
                                        (sqrtB*sqrtB) */
#define DTotal          256          /* total number of detector units */
#define TTotal          8192         /* number of tubes */
#define RRadius         157.0        /* radius of a detector ring in mm */
#define WDet            3.0          /* detector width in mm */
#define LDet            20.0         /* detector length in mm */
#define FOVRadius       60.0         /* field of view radius */
#define FOVDiameter     120.0        /* field of view diameter (2*FOVRadius) */

/* Parameters for tubes simulation */

#define NumOfBins       32           /* number of bins in a sinogram */
#define Start1          48           /* number of the first detector that
                                        forms the first tube */
#define Start2          208          /* number of the second detector that
                                        forms the first tube */

/* Parameters for CAF simulation */

#define IAFlen          33           /* max number of different IAFs */
#define MRowLen         600          /* max matrix row length */
#define noIAFinmem     -3            /* "no IAF in memory" flag */
#define CrossStepInt    0.1          /* step in computing CAF */

typedef unsigned short IFormat;      /* format for representing tube index */

/* structures for data storage */

typedef struct { float Xc; float Yc; float XRin; float XRout;
                 float XLin; float XLout; float YRin; float YRout;
                 float YLin; float YLout; float A; } DetectorXY;
typedef DetectorXY PETRing[DTotal];  /* set of detectors */

typedef struct { float Ac; float Bc; float Cc;
                 float Ain; float Bin; float Cin; float Aout; float Bout; float Cout;
                 int D1; int D2; int IAFpnt; float IncAng; } TubeLine;
typedef TubeLine PETTubes[TTotal]; /* set of tubes */

typedef struct { int N; IFormat *index; float *value; } SparseVector;

typedef struct { float *X; float *F; char *IncAng; float E0;
                 int N; } simIAF;    /* a simulated IAF */
typedef simIAF IAFset[IAFlen];       /* Intrinsic Apperture Function set */

typedef enum {Inside,Outside} Check;
typedef Check Image[BTotal];

/* advance functions' definition */
```

```
void  process_param(int argc, char **argv);
void  my_exit(int err);
void  CreateMatrix(char *fn);
void  WriteMatrix(SparseVector *m, char* fn, int npix);
void  CreateRingClosed(PETRing *Ring);
int   CreateTubes(PETRing *R, PETTubes *PT, int *NIAF);
float ResponseInt();
int   Section(float C,float *x1,float *y1,float *x2,float *y2);
int   CreateIAF(float angle, int n);
float CAF(float x, float E, simIAF *iaf);
int   CreateGrid(Check* Grid, Check* mask);


/* public variables */

char     *fname = "/hosts/data/ml-em/prob128x128-v3";  /* matrix filename */
int       fltSize = sizeof(float);
int       IFSize  = sizeof(IFormat);/* ideally should be checked in matrix */
int       intSize = sizeof(int);
int       NumIter;                  /* number of iterations to perform */
int       NumOfFrames;              /* number of frames in the sinogram */
int       boxinmat;                 /* sqrt of number of boxes in matrix */
float     XL,XR,YU,YD;              /* coordinates of a box being processed */
PETRing  *D;                        /* reference to a set of detectors in use */
TubeLine *T;                        /* reference to a tube being processed */
char     *Progname;
struct    header im;                /* input HIPS file header */
IAFset    RespFuncC;                /* set of IAF for clam shell closed */

void main(int argc, char **argv)
{
   FILE         *f;              /* transition matrix file reference */
   int           i,j,k,l;        /* counters */
   int           it;             /* current iteration number */
   int           linewidth;      /* number of affecting tubes for a given box */
   int           xor,yor,x3h;    /* coordinates of pixel when turning image */
   IFormat       RLen;           /* length of matrix row */
   float        *Pic;            /* image estimate */
   float        *Pic1;           /* final image */
   float        *n_t;            /* result of 1st matrix multiplication */
   float        *del;            /* image correction vector */
   float        *n;              /* measured data (sinogram) */
   float        *ntale;          /* tale of the frame which has no data */
   float         CTotal;         /* total number of counts in the sinogram */
   SparseVector *Mat;            /* matrix of probabilities */
   SparseVector *V;              /* current line of matrix */
   int           AffPix;         /* number of pixels in circular image */
   Image         ImageGrid;      /* image grid defining participating pixels */
   Image         Mask;           /* mask for the edge artifact elimination */
   float         total;          /* total number of counts in image estimate */
   float         artifact;       /* counts captured by edge artifact */
   float         correct;        /* correction factor for artifact elimination */
   float         Picj;           /* temporary variable */

   process_param(argc,argv);

   /* open and read transition matrix */

   f = fopen(fname,"rb");
   if (f==NULL) {
      fprintf(stderr,"Matrix %s not found\n",fname);
      CreateMatrix(fname);
      f = fopen(fname,"rb");
   }
   AffPix = CreateGrid(&ImageGrid[0],&Mask[0]);
   fprintf(stderr,"Reading matrix %s ... ",fname);
   fread(&boxinmat,intSize,1,f);
   if (boxinmat!=sqrtB) my_exit(5);
   Mat = (SparseVector *) calloc(AffPix,sizeof(SparseVector));
   if (Mat==NULL) my_exit(1);
   for (i=0;i<AffPix;i++)
   {
      if (fread(&RLen,IFSize,1,f)!=1) my_exit(2);
      V = &Mat[i];
      (*V).N = (int) RLen;
      (*V).index = (IFormat *) calloc(RLen,IFSize);
      if ((*V).index==NULL) my_exit(1);
      (*V).value = (float *) calloc(RLen,fltSize);
      if ((*V).value==NULL) my_exit(1);
      if (fread((*V).index,IFSize,RLen,f)!=RLen) my_exit(2);
      if (fread((*V).value,fltSize,RLen,f)!=RLen) my_exit(2);
```

```
}
fclose(f);
fprintf(stderr,"Done\n");

/* allocate memory space */

n_t = (float *) calloc(TTotal,fltSize);
del = (float *) calloc(AffPix,fltSize);
n = (float *) calloc(TTotal,fltSize);
ntale = (float *) calloc(NumOfBins,fltSize);
Pic = (float *) calloc(AffPix,fltSize);
Pic1 = (float *) calloc(BTotal,fltSize);
if ((n_t==NULL)||(del==NULL)||(n==NULL)||
    (Pic==NULL)||(Pic1==NULL)||(ntale==NULL)) my_exit(1);

/* reading a sinogram */

for (l=0;l<NumOfFrames;l++)
{
   if (pread(0,n,TTotal*fltSize)!=TTotal*fltSize) my_exit(3);
   if (pread(0,ntale,NumOfBins*fltSize)!=NumOfBins*fltSize) my_exit(3);

   /* filling the image estimate */

   CTotal = 0;
   for (i=0;i<TTotal;i++) CTotal += n[i];
   if (CTotal==0)  fprintf(stderr,"Null number of counts in frame #%d!\n",l);

   /* could skip reconstruction part if CTotal==0 ! but not today! */

   Picj = CTotal / AffPix;
   for (i=0;i<AffPix;i++) Pic[i] = Picj;

   /* iterations */

   for (it=0;it<NumIter;it++)
   {
      /* multiply image estimate to the transition matrix */

      for (i=0;i<TTotal;i++) n_t[i] = 0.0;
      for (j=0;j<AffPix;j++)
      {
         V = &Mat[j];
         linewidth = (*V).N;
         Picj = Pic[j];
         for (k=0;k<linewidth;k++) n_t[(*V).index[k]] += (*V).value[k]*Picj;
      }

      /* divide measured data by the above obtained vector */

      for (i=0;i<TTotal;i++)
         n_t[i] = n[i] / n_t[i];

      /* multiply transition matrix to the correction vector */

      for (j=0;j<AffPix;j++)
      {
         V = &Mat[j];
         linewidth = (*V).N;
         del[j] = 0.0;
         for (k=0;k<linewidth;k++)
            { del[j] += (*V).value[k] * n_t[(*V).index[k]]; }
      }

      /* multiply image estimate to the correction vector */

      for (i=0;i<AffPix;i++) Pic[i] *= del[i];
      fprintf(stderr,"Frame %d: Iteration %d of %d done    \r",
                     l,it+1,NumIter);

   } /* next iteration */

   /* turn image clockwize 3 h and write to output */

   j = 0;
   for (i=0;i<BTotal;i++) {
       xor = i % sqrtB;
       yor = i / sqrtB;
       x3h = sqrtB - yor - 1;
       Pic1[sqrtB*xor+x3h] = (ImageGrid[i]==Inside) ? Pic[j++] : 0.0;
   }
```

```
      /* elimination of the edge artifact */

      total = artifact = 0.0;
      for (i=0;i<BTotal;i++) {
         total += Pic1[i];
         if (Mask[i]==Inside) {
            artifact += Pic1[i];
            Pic1[i] = 0;
         }
      }
      correct = total / (total-artifact);
      total = 0.0;
      for (i=0;i<BTotal;i++) {
         Pic1[i] *= correct;
         total += Pic1[i];
      }
      if (write(1,Pic1,BTotal*fltSize)!=BTotal*fltSize) my_exit(4);
   }
   fprintf(stderr,"\n");

   /* free allocated memory */

   free(n_t);
   free(del);
   free(n);
   free(ntale);
   free(Pic);
   free(Pic1);
   for (i=0;i<AffPix;i++)
   {
      V = &Mat[i];
      free((*V).index);
      free((*V).value);
   }
   free(Mat);
   exit(0);
}


/* processing input parameters */

void process_param(int argc, char **argv)
{
   char *szBuff;

   if (argc!=2) {
      fprintf(stderr,
         "\nSyntax: %s NumberOfIterations < sinogram(n*32*257) > outfile\n\n",
         argv[0]);
      exit(0);
   }
   Progname = strsave(*argv);
   NumIter = atoi(argv[1]);
   if (NumIter<=0) NumIter = 1;
   read_header(&im);
   if (im.pixel_format!=PFFLOAT) my_exit(6);
   if (im.cols!=32||im.rows!=257) my_exit(7);
   NumOfFrames = im.num_frame;
   szBuff=malloc(130);
   sprintf(szBuff,"%s  grid: %d x %d  frames: %d",argv[0],sqrtB,sqrtB,
           im.num_frame);
   im.cols = sqrtB;
   im.rows = sqrtB;
   update_header(&im,1,&szBuff);
   write_header(&im);
   free(szBuff);
}


/* emergency exit */

void my_exit(int err)
{
   switch (err) {
      case 1:
         fprintf(stderr,"\nNot enough memory!\n");
         break;
      case 2:
         fprintf(stderr,"\nError reading matrix!\n");
         break;
```

```
      case 3:
         fprintf(stderr,"\nError reading sinogram!\n");
         break;
      case 4:
         fprintf(stderr,"\nError writing to output!\n");
         break;
      case 5:
         fprintf(stderr,"Matrix isn't for %d*%d image!\n",sqrtB,sqrtB);
         break;
      case 6:
         fprintf(stderr,"Data must be float!\n");
         break;
      case 7:
         fprintf(stderr,"Not a 32*257 sinogram!\n");
   }
   exit(err);
}


/* CreateGrid() marks pixels of image grid that should not be used in
   calculations; returns number of pixels in circular image to reconstruct */

int CreateGrid(Check* Grid, Check* mask)
{
   float step,xl,xr,yu,yd,sFOV,sFOVm,sxl,sxr,syu,syd,tmp;
   int   j,ix,iy,TPix;

   TPix = 0;
   step = (float) FOVDiameter / sqrtB;       /* side of a box in the image */
   sFOV = (float) FOVRadius * FOVRadius;
   tmp = FOVRadius - step*sqrt(2.0);
   sFOVm = tmp * tmp;
   for (j=0;j<BTotal;j++)
   {
      ix = j % sqrtB;
      iy = j / sqrtB;
      xl = - FOVRadius + ix * step;
      xr = xl + step;
      yu = FOVRadius - iy * step;
      yd = yu - step;
      sxl = xl * xl;
      sxr = xr * xr;
      syu = yu * yu;
      syd = yd * yd;
      if ( ((sxl+syu-sFOV)>0) && ((sxr+syu-sFOV)>0) &&
           ((sxl+syd-sFOV)>0) && ((sxr+syd-sFOV)>0) )
      {
         Grid[j] = Outside;
         mask[j] = Outside;
      } else {
         Grid[j] = Inside;
         TPix++;
         mask[j] = ( ((sxl+syu-sFOVm)>0) && ((sxr+syu-sFOVm)>0) &&
                     ((sxl+syd-sFOVm)>0) && ((sxr+syd-sFOVm)>0) ) ? Inside : Outside;
      }
   }
   return TPix;
}


/* create matrix for exact detectors' position */

void CreateMatrix(char *fn)
{
   PETTubes       TubesC;
   PETRing        RingClosed;
   float          stepXY,weight,xx,gx,e0;
   int            t,r,i,j,k,cnt,tmp,eq;
   int            NumOfTubesC,NumOfIAFC;
   char           *name,*s;
   SparseVector  *m2;
   FILE          *p;
   int            APix;
   Image          IGrid;
   Image          AMask;

   fprintf(stderr,"Simulating ring and tubes ... ");
   APix = CreateGrid(IGrid,AMask);
   CreateRingClosed(&RingClosed);
   for (i=0;i<IAFlen;i++) { RespFuncC[i].IncAng = (char *) malloc(9); }
   NumOfTubesC = CreateTubes(&RingClosed,&TubesC,&NumOfIAFC);
```

```
/* reading RFs */

s = (char *) malloc(2);
sprintf(s,"%d",fltSize);
for (i=0;i<NumOfIAFC;i++)
{
   name = (char *) malloc(20);
   strcpy(name,RespFuncC[i].IncAng);
   strcat(name,".iaf.");
   strcat(name,s);
   p = fopen(name,"rb");
   if (p!=NULL)
   {
      fread(&tmp,intSize,1,p);
      RespFuncC[i].N = tmp;
      RespFuncC[i].X = (float *) calloc(tmp,fltSize);
      if (RespFuncC[i].X==NULL) fprintf(stderr,"Memory allocation problem\n");
      RespFuncC[i].F = (float *) calloc(tmp,fltSize);
      if (RespFuncC[i].F==NULL) fprintf(stderr,"Memory allocation problem\n");
      fread(RespFuncC[i].X,fltSize,tmp,p);
      fread(RespFuncC[i].F,fltSize,tmp,p);
      fclose(p);
      e0 = gx = xx = 0;
      for (j=0;j<tmp;j++)
      {
         e0 += (gx+RespFuncC[i].F[j])*(RespFuncC[i].X[j]-xx)/2;
         xx = RespFuncC[i].X[j];
         gx = RespFuncC[i].F[j];
      }
      RespFuncC[i].E0 = e0 / RespFuncC[i].X[tmp-1];
   }
   else
   {
      fprintf(stderr,"[CreateMatrix] Can't find file %s",name);
      exit(0);
   }
   free(name);
}

/* end of reading RFs */

stepXY = (float)FOVDiameter / (float)sqrtB;
fprintf(stderr,"Done (%d different IAFs used)\nCreating matrix for ring closed :\n",NumOfIAFC);
m2 = (SparseVector *) calloc(APix,sizeof(SparseVector));
if (m2==NULL) my_exit(1);
for(j=0;j<APix;j++)
{
   m2[j].index = (unsigned short*) calloc(MRowLen,sizeof(unsigned short));
   m2[j].value = (float*) calloc(MRowLen,sizeof(float));
   if ((m2[j].index==NULL)||(m2[j].value==NULL)) my_exit(1);
   m2[j].N = 0;
}
cnt = k = 0;
D = &RingClosed;
for (j=0;j<sqrtB;j++)
{
   YU = FOVRadius  - j * stepXY;
   YD = YU - stepXY;
   for (i=0;i<sqrtB;i++)
   {
      if (IGrid[k++]==Inside)
      {
         XL = -FOVRadius + i * stepXY;
         XR = XL + stepXY;
         for (t=0;t<NumOfTubesC;t++)
         {
            T = &TubesC[t];
            weight = ResponseInt();
            if (weight>0.0)
            {
               m2[cnt].value[m2[cnt].N] = weight;
               m2[cnt].index[m2[cnt].N++] = (unsigned short) t;
            }
         }
         fprintf(stderr,"pixel %d    \r",++cnt);
      }
   }
}
fprintf(stderr,"\nDone       \n");
WriteMatrix(m2,fn,APix);
```

```
   for (i=0;i<APix;i++)
   {
      free(m2[i].index);
      free(m2[i].value);
   }
   free(m2);
   sprintf(s,"%d",fltSize);
   name = (char *) malloc(20);
   for (i=0;i<NumOfIAFC;i++)
   {
      strcpy(name,RespFuncC[i].IncAng);
      strcat(name,".iaf.");
      strcat(name,s);
      remove(name);
   }
   for (i=0;i<IAFlen;i++) free(RespFuncC[i].IncAng);
   free(name);
   free(s);
}


/* write matrix to disk */

void WriteMatrix(SparseVector *m, char* fn, int npix)
{
   IFormat N;
   float   sum;
   int     i,j,Grid;
   FILE   *p;

   fprintf(stderr,"Writing to file %s ...\n",fn);
   Grid = sqrtB;
   p = fopen(fn,"wb");
   fwrite(&Grid,intSize,1,p);
   for(i=0;i<npix;i++)
   {
      N = (IFormat) m[i].N;
      sum = 0;
      for(j=0;j<N;j++) { sum += m[i].value[j]; }
      for(j=0;j<N;j++) { m[i].value[j] /= sum; }
      fwrite(&N,IFSize,1,p);
      fwrite(m[i].index,IFSize,N,p);
      fwrite(m[i].value,fltSize,N,p);
   }
   fclose(p);
}


/* simulate detector ring */

void CreateRingClosed(PETRing *Ring)
{
   float CurAng,AngStep,x1,x2,y1,y2,cosA,sinA,a,b;
   int   i;

   AngStep = 2.0 * M_PI / (float) DTotal;
   CurAng = AngStep / 2.0;
   for (i=0;i<DTotal;i++)
   {
      cosA = cos(CurAng);
      sinA = sin(CurAng);
      x1 = (float) RRadius * cosA;
      y1 = (float) RRadius * sinA;
      x2 = (float) (RRadius + LDet) * cosA;
      y2 = (float) (RRadius + LDet) * sinA;
      (*Ring)[i].Xc = x1;
      (*Ring)[i].Yc = y1;
      a = -sinA * (float) WDet / 2.0;
      b = cosA * (float) WDet / 2.0;
      (*Ring)[i].XRin = (*Ring)[i].Xc - a;
      (*Ring)[i].YRin = (*Ring)[i].Yc - b;
      (*Ring)[i].XLin = (*Ring)[i].Xc + a;
      (*Ring)[i].YLin = (*Ring)[i].Yc + b;
      (*Ring)[i].XRout = x2 - a;
      (*Ring)[i].YRout = y2 - b;
      (*Ring)[i].XLout = x2 + a;
      (*Ring)[i].YLout = y2 + b;
      (*Ring)[i].A = CurAng;
      CurAng += AngStep;
   }
}
```

```
/* simulate tubes of response */

int CreateTubes(PETRing *R, PETTubes *PT, int *NIAF)
{
   DetectorXY *Ri,*Rj;
   TubeLine   *PTk;
   float       d1,d2,a,b,B_A;
   int         i,j,k,l,m,n,flag,shift1,shift2,res;

   k = n = flag = shift1 = shift2 = 0;
   for (l=0;l<DTotal;l++)
   {
      for (m=0;m<NumOfBins;m++)
      {
         if (k>=TTotal)
            { printf("[CreateTubes] TTotal exceeded !\n"); exit(1); }
         i = Start1 + m + shift1;
         if (i>=DTotal) i = i % DTotal;
         j = Start2 - m + shift2;
         if (j>=DTotal) j = j % DTotal;
         Ri = &(*R)[i];
         Rj = &(*R)[j];
         a = (*Ri).XRin - (*Rj).XLin;
         b = (*Ri).YRin - (*Rj).YLin;
         d1 = a * a + b * b;
         a = (*Ri).XLin - (*Rj).XRin;
         b = (*Ri).YLin - (*Rj).YRin;
         d2 =  a * a + b * b;
         PTk = &(*PT)[k];
         (*PTk).Ac = (*Rj).Yc - (*Ri).Yc;
         (*PTk).Bc = (*Ri).Xc - (*Rj).Xc;
         (*PTk).Cc = (*Rj).Xc * (*Ri).Yc - (*Ri).Xc * (*Rj).Yc;
         (*PTk).D1 = i;
         (*PTk).D2 = j;
         if (d1==d2)          /* parallel detectors */
         {
            (*PTk).Ain = (*Rj).YLin - (*Ri).YRin;
            (*PTk).Bin = (*Ri).XRin - (*Rj).XLin;
            (*PTk).Cin = (*Rj).XLin * (*Ri).YRin - (*Ri).XRin * (*Rj).YLin;
            (*PTk).Aout = (*Rj).YRin - (*Ri).YLin;
            (*PTk).Bout = (*Ri).XLin - (*Rj).XRin;
            (*PTk).Cout = (*Rj).XRin * (*Ri).YLin - (*Ri).XLin * (*Rj).YRin;
            (*PTk).IncAng = 0.;
         }
         else
         {
            if (d1>d2)    /* detectors close with left and right (D1-D2) sides */
            {
               (*PTk).Ain = (*Rj).YLin - (*Ri).YRin;
               (*PTk).Bin = (*Ri).XRin - (*Rj).XLin;
               (*PTk).Cin = (*Rj).XLin * (*Ri).YRin - (*Ri).XRin * (*Rj).YLin;
               (*PTk).Aout = (*Rj).YRout - (*Ri).YLout;
               (*PTk).Bout = (*Ri).XLout - (*Rj).XRout;
               (*PTk).Cout = (*Rj).XRout * (*Ri).YLout -
                             (*Ri).XLout * (*Rj).YRout;
            }
            else          /* detectors close with right and left (D1-D2) sides */
            {
               (*PTk).Ain = (*Rj).YRin - (*Ri).YLin;
               (*PTk).Bin = (*Ri).XLin - (*Rj).XRin;
               (*PTk).Cin = (*Rj).XRin * (*Ri).YLin - (*Ri).XLin * (*Rj).YRin;
               (*PTk).Aout = (*Rj).YLout - (*Ri).YRout;
               (*PTk).Bout = (*Ri).XRout - (*Rj).XLout;
               (*PTk).Cout = (*Rj).XLout * (*Ri).YRout -
                             (*Ri).XRout * (*Rj).YLout;
            }
            B_A = fabs((*Rj).A - (*Ri).A);
            (*PTk).IncAng = (B_A<M_PI) ? (M_PI-B_A)/2 : (B_A-M_PI)/2;
         }
         res = CreateIAF((*PTk).IncAng,n);
         (*PTk).IAFpnt = (res==noIAFinmem) ? n++ : res;
         k++;
      }
      if (flag==0) { flag = 1; shift1++; }
           else { flag = 0; shift2++; }
   }
   *NIAF = n;
   return k;
}
```

```
/* ResponseInt() computes the impact of a given box on tube counts
   (sum of unweighed lengths of intersection of sample lines imposed
   through the tube at a distance of 'CrossStepInt' from each other);
   uses public variables XL,XR,YU,YD,T,D

   input       :   reference to a detector set
   return code :   weighted value of intesection of tube with box          */


float ResponseInt()
{
   float   I,x,y,x1,y1,x2,y2,r1,r2,r,x3,y3,tmp,tmp1,tmp2,CstepX,CstepY;
   float   E1,E2,Er1,Er2,D1Xc,D2Xc,D1Yc,D2Yc,c,c3,C0,d1,pnt1,pnt2;
   int     secN,i;
   simIAF *iaf;

   I = 0;

   /* compute distance between the centers of detector faces */

   D1Xc = (*D)[(*T).D1].Xc;
   D2Xc = (*D)[(*T).D2].Xc;
   D1Yc = (*D)[(*T).D1].Yc;
   D2Yc = (*D)[(*T).D2].Yc;
   r1 = D1Xc - D2Xc;
   r2 = D1Yc - D2Yc;
   r = sqrt(r1 * r1 + r2 * r2);
   iaf = &RespFuncC[(*T).IAFpnt];

   if (fabs((*T).Bc)<0.01)   /* line through detectors "parallel" to Y */
   {
      x1 = -(*T).Cin / (*T).Ac;
      x2 = -(*T).Cout / (*T).Ac;
      tmp = (x1>x2) ? x1 - x2 : x2 - x1;
      CstepX = (x1>x2) ? -CrossStepInt : CrossStepInt;
      secN = (int) (tmp / CrossStepInt);
      Er1 = fabs(D1Yc - YU);
      Er2 = r - Er1;
      E1 =  (Er1<Er2) ? Er1/Er2 : Er2/Er1;
      Er1 = fabs(D1Yc - YD);
      Er2 = r - Er1;
      E2 =  (Er1<Er2) ? Er1/Er2 : Er2/Er1;
      for (i=1;i<secN;i++)
      {
         x3 = x1 + i * CstepX;
         if ((x3>=XL)&&(x3<=XR))
         {
            d1 = (x1>x3) ? x1 - x3 : x3 - x1;
            pnt1 = CAF(d1,E1,iaf);
            pnt2 = CAF(d1,E2,iaf);
            I += (pnt1 + pnt2) / 2;
         }
      }
      I *= YU - YD;
   }
   else
   {
      if (fabs((*T).Ac)<0.01)  /* line through detect. "parallel" to X */
      {
         y1 = -(*T).Cin / (*T).Bc;
         y2 = -(*T).Cout / (*T).Bc;
         tmp = (y1>y2) ? y1 - y2 : y2 - y1;
         CstepY = (y1>y2) ? -CrossStepInt : CrossStepInt;
         secN = (int) (tmp / CrossStepInt);
         Er1 = fabs(D1Xc - XR);
         Er2 = r - Er1;
         E1 =  (Er1<Er2) ? Er1/Er2 : Er2/Er1;
         Er1 = fabs(D1Xc - XL);
         Er2 = r - Er1;
         E2 =  (Er1<Er2) ? Er1/Er2 : Er2/Er1;
         for (i=1;i<secN;i++)
         {
            y3 = y1 + i * CstepY;
            if ((y3>=YD)&&(y3<=YU))
            {
               d1 = (y1>y3) ? y1 - y3 : y3 - y1;
               pnt1 = CAF(d1,E1,iaf);
               pnt2 = CAF(d1,E2,iaf);
               I += (pnt1 + pnt2) / 2;
```

```
      }
   }
   I *= XR - XL;
}
else   /* line through detectors not "parallel" to X or Y axes */
{
   if ((*T).Bc>5)
   {
      x = (float)RRadius/2.0;
      y = -((*T).Ac*x+(*T).Cin) / (*T).Bc;
   }
   else
   {
      y = (float)RRadius/2.0;
      x = -((*T).Bc*y+(*T).Cin) / (*T).Ac;
   }
   c = (*T).Ac * y - (*T).Bc * x;
   y2 = ((*T).Ac * c - (*T).Bc * (*T).Cout) /
        ((*T).Ac * (*T).Ac + (*T).Bc * (*T).Bc);
   x2 = ((*T).Ac * y2 - c) / (*T).Bc;
   tmp1 = y - y2;
   tmp2 = x - x2;
   tmp = sqrt(tmp2 * tmp2 + tmp1 * tmp1);
   secN = (int) (tmp / CrossStepInt);
   CstepX = fabs(CrossStepInt * r2/ r);
   CstepX = (x>x2) ? -CstepX : CstepX;
   CstepY = fabs(CrossStepInt * r1 / r);
   CstepY = (y>y2) ? -CstepY : CstepY;
   for (i=1;i<secN;i++)
   {
      x3 = x + i * CstepX;
      y3 = y + i * CstepY;
      C0 = -((*T).Ac * x3 + (*T).Bc * y3);
      if (Section(C0,&x1,&y1,&x2,&y2)!=0)
      {
         /* project (x1,y1) onto the "central" line of the tube */

         c3 = (*T).Ac * y1 - (*T).Bc * x1;
         y3 = ((*T).Ac * c3 - (*T).Bc * (*T).Cc) /
              ((*T).Ac * (*T).Ac + (*T).Bc * (*T).Bc);
         x3 = ((*T).Ac * y3 - c3) / (*T).Bc;
         tmp = D1Xc - x3;
         tmp1 = D1Yc - y3;
         Er1 = sqrt(tmp * tmp + tmp1 * tmp1);
         Er2 = r - Er1;
         E1 =  (Er1<Er2) ? Er1/Er2 : Er2/Er1;

         /* project (x1,y1) onto the inner line of the tube */

         c3 = (*T).Ain * y1 - (*T).Bin * x1;
         y3 = ((*T).Ain * c3 - (*T).Bin * (*T).Cin) /
              ((*T).Ain * (*T).Ain + (*T).Bin * (*T).Bin);
         x3 = ((*T).Ain * y3 - c3) / (*T).Bin;
         tmp = x1 - x3;
         tmp1 = y1 - y3;
         d1 = sqrt(tmp * tmp + tmp1 * tmp1);
         pnt1 = CAF(d1,E1,iaf);

         /* project (x2,y2) onto the "central" line of the tube */

         c3 = (*T).Ac * y2 - (*T).Bc * x2;
         y3 = ((*T).Ac * c3 - (*T).Bc * (*T).Cc) /
              ((*T).Ac * (*T).Ac + (*T).Bc * (*T).Bc);
         x3 = ((*T).Ac * y3 - c3) / (*T).Bc;
         tmp = D1Xc - x3;
         tmp1 = D1Yc - y3;
         Er1 = sqrt(tmp * tmp + tmp1 * tmp1);
         Er2 = r - Er1;
         E2 =  (Er1<Er2) ? Er1/Er2 : Er2/Er1;

         /* project (x2,y2) onto the inner line of the tube */

         c3 = (*T).Ain * y2 - (*T).Bin * x2;
         y3 = ((*T).Ain * c3 - (*T).Bin * (*T).Cin) /
              ((*T).Ain * (*T).Ain + (*T).Bin * (*T).Bin);
         x3 = ((*T).Ain * y3 - c3) / (*T).Bin;
         tmp = x2 - x3;
         tmp1 = y2 - y3;
         d1 = sqrt(tmp * tmp + tmp1 * tmp1);
         pnt2 = CAF(d1,E2,iaf);
```

```
                /* calculate the distance between (x1,y1) and (x2,y2) */

                tmp = x1 - x2;
                tmp1 = y1 - y2;
                d1 = sqrt(tmp * tmp + tmp1 * tmp1);
                I += d1 * (pnt1 + pnt2) / 2;
             }
          }
       }
    }
    return I;
}


/* Section() finds points of intersection of a given line with a given box;
   uses public variables XL,XR,YU,YD,T

   return code: 0 if there is no such points (x1,y1,x2,y2 undefined)
                1 if there are such points (x1,y1,x2,y2 returned)          */


int Section(float C,float *x1,float *y1,float *x2,float *y2)
{
   float x,y;
   int   flag;

   flag = 0;
   x = -((*T).Bc * YU + C) / (*T).Ac;
   if ((x>=XL)&&(x<=XR)) { *x1 = x; *y1 = YU; flag++; }
   x = -((*T).Bc * YD + C) / (*T).Ac;
   if ((x>=XL)&&(x<=XR))
   {
      if (flag==0) { *x1 = x; *y1 = YD; }
             else { *x2 = x; *y2 = YD; }
      flag++;
   }
   if (flag==2) { return 1; }
   y = -((*T).Ac * XL + C) / (*T).Bc;
   if ((y>=YD)&&(y<=YU))
   {
      if (flag==0) { *x1 = XL; *y1 = y; }
             else { *x2 = XL; *y2 = y; }
      flag++;
   }
   if (flag==2) { return 1; }
   y = -((*T).Ac * XR + C) / (*T).Bc;
   if ((y>=YD)&&(y<=YU))
   {
      if (flag==0) { *x1 = XR; *y1 = y; }
             else { *x2 = XR; *y2 = y; }
      flag++;
   }
   if (flag==2) { return 1; } else { return 0; }
}

/*
   simulation of intrinsic aperture function for a given detector and incidence
   angle (adaptation of FORTRAN program SIMLIN), contains settings for phisical
   parameters inside; different IAF saved in different files;
   sets and uses public variables RespFuncC

   input       :   incidence angle;
                   number of defferent IAF encountered up to the moment
   return code :   error code noIAFinmem or index of the element in array
                   RespFuncC or RespFuncO that contains the IAF already
*/

int CreateIAF(float angle, int n)
{
   float ddmu[2],dx1[512],bcent[512][2],prof[2],dmu[2];
   float proft,detect,septa,ecasep,xlarge,epeint,x,tgp,cp,smu;
   float dbx,xbo,deca,reg,ssmu,dyep,dysep,ynt1,h,y,xmu,dy,dify;
   int   nreg,i,numbc,npl,j,js,go;
   char *nom,*name,*s;
   FILE *p;

   name = (char *) malloc(9);
   nom = (char *) malloc(15);
   s = (char *) malloc(2);
   sprintf(name,"%.3f",angle);
   strcpy(nom,name);
```

```
   strcat(nom,".iaf.");
   sprintf(s,"%d",fltSize);
   strcat(nom,s);
   p = fopen(nom,"r");
   if (p!=NULL)
   {
      fclose(p);
      free(nom);
      free(s);
      i = 0;
      while ( (strcmp(RespFuncC[i].IncAng,name)!=0) && (i<n) ) { i++; }
      if (i>=n)
      {
         strcpy(RespFuncC[n].IncAng,name);
         free(name);
         return noIAFinmem;
      }
      else
      {
         free(name);
         return i;
      }
   }
   p = fopen(nom,"wb");
   dmu[0]=0.697;    /* Lineare attenuation coefficient (cm-1): inner layer */
   dmu[1]=.9581;    /* Lineare attenuation coefficient (cm-1): outer layer */
   smu=0.697;       /* Lineare attenuation coefficient (cm-1) of septa */
   numbc = 2;       /* IAF for inner layer = 1, for outer layer = 2 */
   proft = 21;      /* Total depth of the layer (mm) */
   prof[0] = 1;     /* Depth of inner layer (mm) */
   prof[1] = 20;    /* Depth of outer layer (mm) */
   xlarge = 3.8;    /* Inter-detector distance (mm) */
   epeint = 0.05;   /* Width of air between septa and crystal (mm) */
   septa = 0.7;     /* SEPTA thickness (mm) */
   ecasep = 0;      /* SEPTA gap (mm) */
   dmu[0] = -dmu[0] / 10.;
   dmu[1] = -dmu[1] / 10.;
   smu = -smu / 10.;
   ddmu[0] = dmu[0] * 10.;
   ddmu[1] = dmu[1] * 10.;
   ssmu = smu * 10.;
   cp = cos(angle);
   tgp = sin(angle) / cp;
   if (tgp==0) { tgp = .0001; };
   dbx = (xlarge / 32.) / cp;
   xbo = proft  * tgp + xlarge;
   detect = xlarge - septa - 2 * epeint;
   deca = detect - .99 * dbx;
   npl = xbo / dbx;
   for (i=0;i<npl;i++)
   {
      bcent[i][0] = 0.;
      bcent[i][1] = 0.;
      reg = deca / xlarge +75.;
      nreg = reg;
      x = (reg - nreg) * xlarge;

/* ************ it was subroutine discon ************************************ */

      dyep = epeint / tgp;
      dysep = septa / tgp;

      ynt1 = 1.;
      h = 0.;
      j = 1;
      js = 1;

      if (septa==0.)
      {
         do
         {
            go = 1;
            y = h;
            j = js;
            if (y<proft)
            {
               if (x>=detect)
               {
                  nreg += 1;
                  h = y + (xlarge - x) / tgp;
                  x = 0;
```

```
            if (h>=prof[0]) { js = 2; go = 0; }
         }
         if (go!=0)
         {
            xmu = dmu[j-1];
            if (j==2)
            {
               dy = (detect - x) / tgp;
               h = y + dy;
               x = detect;
               if (h>=proft) { dy = proft - y; }
               js = 1;
            }
            else
            {
               dy = (detect - x) / tgp;
               h = y + dy;
               if (h>=prof[0])
               {
                  dy = prof[0] - y;
                  x += dy * tgp;
                  h = y + dy;
                  js = 2;
               }
               else
               {  x = detect; }
            }

            dify = ynt1 * (1 - exp(xmu*dy/cp));
            ynt1 -= dify;

            if (nreg!=75) { go = 0; } else {  if (x==0) { go = 0; } }
            if (go!=0)
            {
               bcent[i-1][j-1] = dify;
               if (j!=2) { go = 0; }
            }
         }
      }
   } while (go==0);
}
else
{
   do
   {
      go = 1;
      y = h;
      j = js;
      if (y<proft)
      {
         if (x>=detect)
         {
            nreg += 1;
            h = y + (xlarge - x) / tgp;
            x = 0;
            if (h>=prof[0]) { js = 2; }
            if ((h<=(ecasep+dyep))||(h>=proft))
            { go = 0; }
            else
            {
               xmu = smu;
               dy = h - ecasep - dyep;
               if (dy>dysep) { dy = dysep; }
            }
         }
         else
         {
            xmu = dmu[j-1];
            if (j==2)
            {
               dy = (detect - x) / tgp;
               h = y + dy;
               x = detect;
               if (h>=proft) { dy = proft - y; }
               js = 1;
            }
            else
            {
               dy = (detect - x) / tgp;
               h = y + dy;
               if (h>=prof[0])
```

```
                    {
                        dy = prof[0] - y;
                        x += dy * tgp;
                        h = y + dy;
                        js = 2;
                    }
                    else
                    {  x = detect; }
                }
            }
            if (go!=0)
            {
                dify = ynt1 * (1 - exp(xmu*dy/cp));
                ynt1 -= dify;

                if (nreg!=75) { go = 0; } else { if (x==0.) { go = 0; } }
                if (go!=0)
                {
                    bcent[i-1][j-1] = dify;
                    if (j!=2) { go = 0; }
                }
            }
        }
    } while (go==0);
}

/* ************************************************************************** */

    deca -= dbx;
  }
  for (i=0;i<npl;i++)
  {
    dx1[i] = cp * dbx * (i + 1);
  }
  fwrite(&npl,intSize,1,p);
  fwrite(&dx1[0],fltSize,npl,p);
  for (i=0;i<npl;i++) { fwrite(&bcent[i][numbc-1],fltSize,1,p); }
  strcpy(RespFuncC[n].IncAng,name);
  fclose(p);
  free(name);
  free(nom);
  free(s);
  return noIAFinmem;
}


/*
   computes the coincidence response function for a given tube and position
   in the tube using simulated IAF

   input       :   distance from the point being evaluated to the inner line
                   of tube;
                   ratio of distances from the point being evaluated to lines
                   crossing center of detector faces and perpendiculare to the
                   tube E (E<=1, E=1 when the point is half-way between
                   detectors);
                   reference to IAF
   return code :   value of coincedence response function for the point

*/


float CAF(float x, float E, simIAF *iaf)
{
  float min,max,iv,g,g1,g2,t;
  int   dim,j,first,k;

  dim = (*iaf).N;
  min = (*iaf).X[0];
  max = (*iaf).X[dim-1];
  iv = 0;
  first = 0;
  for (j=0;j<dim;j++)
  {
    t = x + E * (x - (*iaf).X[j]);
    if ((t>=min) && (t<=max))
    {
      k = 0;
      while (((*iaf).X[k]<=t) && (k<dim)) { k++; }
      k--;
      g = (t==(*iaf).X[k]) ? (*iaf).F[k] :
```

```
                                (*iaf).F[k] + ((*iaf).F[k+1] - (*iaf).F[k]) *
                                (t - (*iaf).X[k])/((*iaf).X[k+1] - (*iaf).X[k]);
            if (first==0)
            {
               first++;
               g1 = (*iaf).F[j] * g;
            }
            else
            {
               g2 = (*iaf).F[j] * g;
               iv += (g1 + g2) / 2 * ((*iaf).X[j] - (*iaf).X[j-1]);
               g1 = g2;
            }
         }
      }
      iv *= (1+E)*(1+E)/(E*max*max*(*iaf).E0*(*iaf).E0);
      return iv;
}
```

References

Anastasio M A, Pan X, and Clarkson E 2001 Comments on the filtered backprojection algorithm, range conditions, and the pseudoinverse solution *IEEE Trans. Med. Imag.* **20** 539-42

Araujo A and Giné E 1980 *The central limit theorem for real and Banach valued random variables* (New York: Wiley)

Atkins M S, Murray D, and Harrop R 1991 Use of transputers in a 3-D positron emission tomograph *IEEE Trans. Med. Imag.* **10** 276-83

Baker J R, Budinger T F, and Huesman R H 1992 Generalized approach to inverse problems in tomography: Image reconstruction for spatially variant systems using natural pixels *Critical Rev. Biomed. Eng.* **20** 47-71

Barrett H H 1999 Topics in the mathematics of computed tomography *Notes of a short course presented at the 1999 IEEE Nucl. Sci. Symp. and Med. Imag. Conf.*, Seattle, USA (Three draft chapters from Barrett H H and Myers K J *Foundations of image science* to be published by John Wiley, Inc.)

Bendriem B and Townsend D W (eds.) 1998 *The Theory and Practice of 3D PET* (Dordrecht: Kluwer Academic Publishers)

Bertero M, De Mol C, and Pike E R 1988 Linear inverse problems with discrete data. II: Stability and regularisation *Inverse Problems* **4** 573-94

Bevington P R and Robinson D K 1992 *Data reduction and error analysis for the physical sciences* 2nd edn 63 (WCB/McGraw-Hill)

Bortfeld T and Oelfke U 1999 Fast and exact 2D image reconstruction by means of Chebyshev decomposition and backprojection *Phys. Med. Biol.* **44** 1105-20

Brantner S, Young R C D, Budgett D, and Chatwin C R 1997 High-speed tomographic reconstruction employing Fourier methods *Real-Time Imaging* **3** 255-74

Brix G, Doll J, Bellemann M E, Trojan H, Haberkorn U, Schmidlin P, and Ostertag H 1997 Use of scanner characteristics in iterative image reconstruction for high resolution positron emission tomography studies of small animals *Eur. J. Nucl. Med.* **24** 779-86

Browne J and De Pierro A R 1996 A row-action alternative to the EM algorithm for maximizing likelihoods in emission tomography *IEEE Trans. Med. Imag.* **15** 687-99

Budinger T F and Gullberg G T 1974 Three-dimensional reconstruction in nuclear medicine emission imaging *IEEE Trans. Nucl. Sci.* **21** 2-20

Budinger T F, Gullberg G T, and Huesman R H 1979 Emission computed tomography *Image Reconstruction from Projections: Implementation and Applications* (Herman G T ed.) *Topics in Applied Physics* **32** 147-246 (Berlin: Springer-Verlag)

Budinger T F 1998 PET instrumentation: What are the limits? *Seminars in Nuclear Medicine* **28** 247-67

Buonocore M H, Brody W R, and Macovski A 1981 A natural pixel decomposition for two-dimensional image reconstruction *IEEE Trans. Biomed. Eng.* **28** 69-78

Cahill P and Blau L M 1970 The preliminary application of a matrix method for radionuclide imaging *J. Nucl. Med.* **11** 613-5

Caponnetto A and M Bertero 1997 Tomography with a finite set of projections: singular value decomposition and resolution *Inverse Problems* **13** 1191-1205

Casey M E and Hoffman E J 1986 Quantitation in positron emission computed tomography: 7. A technique to reduce noise in accidental coincidence measurement and coincidence efficiency calibration *J. Comput. Assist. Tomogr.* **10** 147-246

Censor Y 1983 Finite series-expansion reconstruction methods *Proceedings of the IEEE* **71** 409-19

Chen C M, Lee S Y, and Cho Z H 1991 Parallelization of the EM algorithm for 3-D PET image reconstruction *IEEE Trans. Med. Imag.* **10** 513-22

Chen C M and Lee S Y 1995 Optimal data replication: A new approach to optimizing parallel EM algorithms on a mesh-connected multiprocessor for 3D PET image reconstruction *IEEE Trans. Nucl. Sci.* **42** 1235-45

Chinn G and Huang S C 1997 A general class of preconditioners for statistical iterative reconstruction of emission computed tomography *IEEE Trans. Med. Imag.* **16** 1-10

Cho Z H, Chan J K, Ericksson L, Singh M, Graham S, MacDonald N S, and Yano Y 1975 Positron ranges obtained from biomedically important positron-emitting radionuclides *J. Nucl. Med.* **16** 1174-6

Cho Z H, Chen C M, and Lee S Y 1990 Incremental algorithm – a new fast backprojection scheme for parallel beam geometries *IEEE Trans. Med. Imag.* **9** 207-17

Coakley K J 1991 A cross-validation procedure for stopping the EM algorithm and deconvolution of neutron depth profiling spectra *IEEE Trans. Nucl. Sci.* **38** 9-15

Coakley K J and Llacer J 1992 An optimized maximum likelihood approach to positron emission tomography image reconstruction *Proc. 1991 IEEE Nucl. Sci. Symp.* **3** 2084-8

Colsher J G 1980 Fully three-dimensional positron emission tomography *Phys. Med. Biol.* **25** 103-15

Comtat C, Morel C, Defrise M, and Townsend D W 1993 The FAVOR algorithm for 3D PET data and its implementation using a network of transputers *Phys. Med. Biol.* **38** 929-44

Comtat C, Kinahan P E, Defrise M, Michel C, and Townsend D W 1998 Fast reconstruction of 3D PET data with accurate statistical modeling *IEEE Trans. Nucl. Sci.* **45** 1083-9

Cormack A M 1963 Representation of a function by its line integrals, with some radiological applications *J. Appl. Phys.* **34** 2722-7

Cormack A M 1964 Representation of a function by its line integrals, with some radiological applications. II *J. Appl. Phys.* **35** 2908-13

Cormack A M 1980 Early two-dimensional reconstruction and recent topics stemming from it *Science* **209** 1482-6

Cormack A M 1992 75 years of Radon transform *J. Comput. Assist. Tomogr.* **16** 673

Cruz-Rivera J L, Di Bella E V R, Wills D S, Gaylord T K, and Glytsis E N 1995 Parallelized formulation of the maximum likelihood-expectation maximization algorithm for fine-grain message-passing architectures *IEEE Trans. Med. Imag.* **14** 758-62

Daube-Witherspoon M E and Carson R E 1991 Unified deadtime correction model for PET *IEEE Trans. Med. Imag.* **10** 267-75

Davison M E 1983 The ill-conditioned nature of the limited angle tomography problem *SIAM J. Appl. Math.* **43** 428-48

Defrise M and De Mol C 1987 A note on stopping rules for iterative regularization methods and filtered SVD *Inverse problems, an interdisciplinary study* (Sabatier P C ed.) *Advances in Electronics and Electron Physics* (supplement 19) 261-8 (Academic Press)

Defrise M, Kinahan P E, Townsend D W, Michel C, Sibomana M, and Newport D F 1997 Exact and approximate rebinning algorithms for 3-D PET data *IEEE Trans. Med. Imag.* **16** 145-58

Demoment G 1989 Image reconstruction and restoration: Overview of common estimation structures and problems *IEEE Trans. ASSP* **37** 2024-36

Dempster A P, Laird N M, and Rubin D B 1977 Maximum likelihood from incomplete data via the EM algorithm (with discussion) *J. R. Statist. Soc.* **39** 1-38

Derenzo S E, Moses W W, Huesman R H, and Budinger T F 1993 Critical instrumentation issues for resolution smaller than 2 mm, high sensitivity brain PET (with discussion) *Quantification of Brain Function. Tracer Kinetics and Image Analysis in Brain PET* (Uemura *et al.* eds.) 25-41 (Elsevier)

de Villiers G D, McNally B, and Pike E R 1999 Positive solutions to linear inverse problems *Inverse Problems* **15** 615-35

Egger M L, Joseph C, and Morel C 1998 Incremental beamwise backprojection using geometrical symmetries for 3D PET reconstruction in a cylindrical scanner geometry *Phys. Med. Biol.* **43** 3009-24

Falcón C, Juvells I, Pavía J, and Ros D 1998 Evaluation of a cross-validation stopping rule in MLE SPECT reconstruction *Phys. Med. Biol.* **43** 1271-83

Fessler J A 1994 Penalized weighted least-squares image reconstruction for positron emission tomography *IEEE Trans. Med. Imag.* **13** 290-300

Fierro R D and Hansen P C 1995 Accuracy of TSVD solutions computed from rank-revealing decompositions *Numer. Math.* **70** 453-71

Franklin J N 1970 Well-posed stochastic extensions of ill-posed linear problems *J. Math. Anal. Appl.* **31** 682-716

Geman S and McClure D E 1987 Statistical methods for tomographic image reconstruction *Bulletin of the International Statistical Institute* **52** (book 4) 5-21

Germano G and Hoffman E J 1991 An investigation of methods of pileup rejection for 2-D array detectors employed in high resolution PET *IEEE Trans. Med. Imag.* **10** 223-7

Gilliam D S, Lund J R, and Vogel C R 1990 Quantifying information content for ill-posed problems *Inverse Problems* **6** 725-36

Golub G H and von Matt U 1997 Generalized cross-validation for large-scale problems *J. Comput. Graph. Statist.* **6** 1-34

Golub G, Solna K, and Van Dooren P 2000 Computing the SVD of a general matrix product/quotient *SIAM J. Matrix Anal. Appl.* **22** 1-19

Gooley T A and Barrett H H 1992 Evaluation of statistical methods of image reconstruction through ROC analysis *IEEE Trans. Med. Imag.* **11** 276-83

Gordon R 1974 A tutorial on ART *IEEE Trans. Nucl. Sci.* **21** 78-93

Gordon R and Herman G T 1974 Three-dimensional reconstruction from projections: A review of algorithms *Int. Rev. Cytol.* **38** 111-51

Gottlieb D, Gustafsson B, and Forssén P 2000 On the direct Fourier method for computer tomography *IEEE Trans. Med. Imag.* **19** 223-32

Gregor J and Huff D A 1997 A focus-of-attention preprocessing scheme for EM-ML PET reconstruction *IEEE Trans. Med. Imag.* **16** 218-23

Gullberg G T, Hsieh Y-L, and Zeng G L 1996 An SVD reconstruction algorithm using a natural pixel representation of the attenuated Radon transform *IEEE Trans. Nucl. Sci.* **43** 295-303

Hansen P C 1992 Numerical tools for analysis and solution of Fredholm integral equations of the first kind *Inverse Problems* **8** 849-72

Hansen P C and O'Leary D P 1993 The use of the L-curve in the regularization of discrete ill-posed problems *SIAM J. Sci. Comput.* **14** 1487-1503

Hartz R, Bristow D, and Mullani N 1985 A real-time TOFPET slice-backproject engine employing dual AM29116 microprocessors *IEEE Trans. Nucl. Sci.* **32** 839-42

Hebert T J 1990 Statistical stopping criteria for iterative maximum likelihood reconstruction of emission images *Phys. Med. Biol.* **35** 1221-32

Herman G T 1980 *Image reconstruction from projections: The fundamentals of computerized tomography* (New York: Academic Press)

Herman G T and Meyer L B 1993 Algebraic reconstruction techniques can be made computationally efficient *IEEE. Trans. Med. Imag.* **12** 600-9

Hichwa R 1994 Are animal scanners really necessary for PET? *J. Nucl. Med.* **35** 1396-7

Hoffman E J, Huang S C, Phelps M E, and Kuhl D E 1981 Quantitation in positron emission computed tomography: 4. Effect of accidental coincidences *J. Comput. Assist. Tomogr.* **5** 391-400

Hoffman E J, Huang S C, Plummer D, and Phelps M E 1982 Quantitation in positron emission computed tomography: 6. Effect of nonuniform resolution *J. Comput. Assist. Tomogr.* **6** 987-99

Hoffman E J, Guerrero T M, Germano G, Digby W M, and Dahlbom M 1989 PET system calibrations and corrections for quantitative and spatially accurate images *IEEE Trans. Nucl. Sci.* **36** 1108-12

Hogg R V and Tanis E A 1988 *Probability and statistical inference* 3rd edn 304-21 (New York: Macmillan)

Hoh C K, Schiepers C, Seltzer M A, Gambhir S S, Silverman D H S, Czernin J, Maddahi J, and Phelps M E 1997 PET in oncology: Will it replace the other modalities? *Seminars in Nuclear Medicine* **27** 94-106

Hounsfield G N 1973 Computerized transverse axial scanning (tomography): Part I. Description of system *Brit. J. Radiol.* **46** 1016-22

Hounsfield G N 1976 Historical notes on computerized axial tomography *J. Can. Ass. Radiol.* **27** 135-42

Hounsfield G N 1980 Computed medical imaging *Science* **210** 22-8

Hsieh Y-L, Gullberg G T, Zeng G L, and Huesman R H 1996 Image reconstruction using a generalized natural pixel basis *IEEE Trans. Nucl. Sci.* **43** 2306-19

Hudson H M and Larkin R S 1994 Accelerated image reconstruction using ordered subsets of projection data *IEEE Trans. Med. Imag.* **13** 601-9

Huesman R H, Salmeron E M, and Baker J R 1989 Compensation for crystal penetration in high resolution positron tomography *IEEE Trans. Nucl. Sci.* **36** 1100-7

Iinuma T A and Nagai T 1967 Image restoration in radioisotope imaging systems *Phys. Med. Biol.* **12** 501-9

Johnson C A, Yan Y, Carson R E, Martino R L, and Daube-Witherspoon M E 1995 A system for the 3D reconstruction of retracted-septa PET data using EM algorithm *IEEE Trans. Nucl. Sci.* **42** 1223-27

Jones W F, Byars L G, and Casey M E 1990 Design of a super fast three-dimensional projection system for positron emission tomography *IEEE Trans. Nucl. Sci.* **37** 800-4

Kak A C and Slaney M 1988 *Principles of Computerized Tomographic Imaging* 49-112 (New York: IEEE Press)

Karp J S, Kinahan P E, and Mankoff D A 1991 Positron emission tomography with a large axial acceptance angle: signal-to-noise considerations *IEEE Trans. Med. Imag.* **10** 249-55

Karuta B and Lecomte R 1992 Effect of detector weighting functions on the point spread function of high-resolution PET tomographs: A simulation study *IEEE Trans. Med. Imag.* **11** 379-85

Kaufman L 1987 Implementing and accelerating the EM algorithm for positron emission tomography *IEEE Trans. Med. Imag* **6** 37-51

Kaufman L 1993 Maximum likelihood, least squares, and penalized least squares for PET *IEEE Trans. Med. Imag.* **12** 200-14

Kawata S and Nalcioglu O 1985 Constrained iterative reconstruction by the conjugate gradient method *IEEE Trans. Med. Imag.* **4** 65-71

Kazantsev I G, Van de Walle R, and Lemahieu I 2000 Ridge functions, natural pixels and minimal norm reconstruction *IEEE Trans. Nucl. Sci.* **47** 1118-22

Kearfott K J 1985 Practical considerations (Comment) *J. Amer. Stat. Assoc.* **80** 26-8

Kinahan P E and Rogers J G 1989 Analytic 3D image reconstruction using all detected events *IEEE Trans. Nucl. Sci.* **36** 964-8

Kops E R, Schmitz T, and Herzog H 1999 Geometric accuracy of reconstructed PET images *Proc. 1998 IEEE Nucl. Sci. Symp. & Med. Imag. Conf* **3** 1904-6

Lalush D S, Cui L, and Tsui B M W 1997 A priori motion models for four-dimensional reconstruction in gated cardiac SPECT *Proc. 1996 IEEE Nucl. Sci. Symp. & Med. Imag. Conf.* **3** 1923-7

Lalush D S, Segars W P, and Tsui B M W 2000 A unified approach to 4D gated myocardial SPECT reconstruction *Proc. 1999 IEEE Nucl. Sci. Symp. & Med. Imag. Conf.* M12-5 (CD-ROM, IEEE Catalog Number: 99CH37019C, ISSN: 1082-3654)

Lange K and Carson R 1984 EM reconstruction algorithms for emission and transmission tomography *J. Comput. Assist. Tomogr.* **8** 306-16

Lange K, Bahn M, and Little R 1987 A theoretical study of some maximum likelihood algorithms for emission and transmission tomography *IEEE Trans. Med. Imag.* **6** 106-14

Leahy R M and Qi J 2000 Statistical approaches in quantitative positron emission tomography *Statistics and Computing* **10** 147-65

Lecomte R, Schmitt D, and Lamoureux G 1984 Geometry study of a high resolution PET detection system using small detectors *IEEE Trans. Nucl. Sci.* **31** 556-561

Lepage M D, Leroux J-D, Selivanov V, Cadorette J, and Lecomte R 2001 Design of a prototype real-time image reconstruction system for PET imaging *Proc. 2001 IEEE Nucl. Sci. Symp. & Med. Imag. Conf.* IEEE Catalog # 02CH37310C (CD-ROM) M9C-15

Levin C S and Hoffman E J 1999 Calculation of positron range and its effect on the fundamental limit of positron emission tomography system spatial resolution *Phys. Med. Biol.* **44** 781-99

Levitan E and Herman G T 1987 A maximum a posteriori probability expectation maximization algorithm for image reconstruction in emission tomography *IEEE Trans. Med. Imag.* **6** 185-92

Lewitt R M 1983 Reconstruction algorithms: Transform methods *Proceedings of the IEEE* **71** 390-408

Lewitt R M 1992 Alternatives to voxels for image representation in iterative reconstruction algorithms *Phys. Med. Biol.* **37** 705-16

Lewitt R M and Muehllehner G 1986 Accelerated iterative reconstruction for positron emission tomography based on the EM algorithm for maximum likelihood estimation *IEEE Trans. Med. Imag.* **5** 16-22

Liang Z, Jaszczak R, and Greer K 1989 On Bayesian image reconstruction from projections: uniform and nonuniform a priori source information *IEEE Trans. Med. Imag.* **8** 227-35

Liang Z 1994 Detector response restoration in image reconstruction of high resolution positron emission tomography *IEEE Trans. Med. Imag.* **13** 314-21

Links J M 1998 Advances in nuclear medicine instrumentation: considerations in the design and selection of an imaging system *Eur. J. Nucl. Med.* **25** 1453-66

Liow J S and Strother S C 1993 The convergence of object-dependent resolution in maximum likelihood based tomographic image reconstruction *Phys. Med. Biol.* **38** 55-70

Llacer J 1979 Theory of imaging with a very limited number of projections *IEEE Trans. Nucl. Sci.* **26** 596-602

Llacer J and Veklerov E 1989 Feasible images and practical stopping rules for iterative algorithms in emission tomography *IEEE Trans. Med. Imag.* **8** 186-93

Llacer J 1990 On the validity of hypothesis testing for feasibility of image reconstruction *IEEE Trans. Med. Imag.* **9** 226-30

Llacer J, Veklerov E, Coakley K J, Hofman E J, and Nunez J 1993 Statistical analysis of maximum likelihood estimator images of human brain FDG PET studies *IEEE Trans. Med. Imag.* **12** 215-31

Louis A K 1986 Incomplete data problems in X-ray computerized tomography I. Singular value decomposition of the limited angle transform *Numer. Math.* **48** 251-62

Louis A K 1996 Approximate inverse for linear and some nonlinear problems *Inverse Problems* **12** 175-90

Louis A K and Natterer F 1983 Mathematical problems of computerized tomography *Proceedings of the IEEE* **71** 379-89

Mandelkern M A 1995 Nuclear techniques for medical imaging: Positron Emission Tomography *Annu. Rev. Nucl. Part. Sci.* **45** 205-54

Marr R B 1974 On the reconstruction of a function on a circular domain from a sampling of its line integrals *J. Math. Anal. Appl.* **45** 357-74

Matej S and Lewitt R M 1996 Practical considerations for 3-D image reconstruction using spherically symmetric volume elements *IEEE Trans. Med. Imag.* **15** 68-78

Matthews J, Bailey D, Price P, and Cunningham V 1997 The direct calculation of parametric images from dynamic PET data using maximum-likelihood iterative reconstruction *Phys. Med. Biol.* **42** 1155-73

Mazé A and Lecomte R 1990 Analytical study of the effect of collimation on the performance of PET cameras in 3-D imaging *IEEE Trans. Nucl. Sci.* **37** 823-31

McIntyre J A 1981 Computer assisted tomography without a computer *IEEE Trans. Nucl. Sci.* **28** 171-3

Meikle S R, Matthews J C, Cunningham V J, Bailey D L, Livieratos L, Jones T, Price P 1998 Parametric image reconstruction using spectral analysis of PET projection data *Phys. Med. Biol.* **43** 651-66

Melcher C L 2000 Scintillation crystals for PET *J. Nucl. Med.* **41** 1051-5

Msaki P, Bentourkia M, and Lecomte R 1996 Scatter degradation and correction models for high-resolution PET *J. Nucl. Med.* **37** 2047-9

Muehllehner G 1976 Resolution limit of positron cameras (with replies by Phelps M E and Hoffman E J; by Cho Z H) *J. Nucl. Med.* **17** 757-8

Mumcuoğlu E Ü, Leahy R M, and Cherry S R 1996 Bayesian reconstruction of PET images: methodology and performance analysis *Phys. Med. Biol.* **41** 1777-807

Mumcuoğlu E Ü, Leahy R M, Cherry S R, and Hoffman E 1997 Accurate geometric and physical response modeling for statistical image reconstruction in high resolution PET *Proc. 1996 IEEE Nucl. Sci. Symp. and Med. Imag. Conf.* **3** 1569-73

Natterer F 1980 Efficient implementation of optimal algorithms in computerized tomography *Math. Meth. Appl. Sci.* **2** 545-55

Natterer F 1999 Numerical methods in tomography *Acta Numerica* **8** 107-41 (Cambridge: Cambridge University Press)

Nichols T E, Qi J, and Leahy R M 1999 Continuous time dynamic PET imaging using list mode data *Proc. 16th Conf. Inf. Process. Med. Imag.* (Kuba *et al.* eds.) 98-111 (http://neuroimage.usc.edu//papers/pdf_files/IMPI99_PET.pdf)

Nunez J and Llacer J 1990 A fast Bayesian reconstruction algorithm for emission tomography with entropy prior converging to feasible images *IEEE Trans. Med. Imag.* **9** 159-71

Olesen S P, Gregor J, Thomason M G, and Smith G T 1996 EM-ML PET reconstruction on multiple processors with reduced communications *Int. J. Imag. Syst. Tech.* **7** 215-23

Pan T-S and Yagle A E 1991 Numerical study of multigrid implementations of some iterative image reconstruction algorithms *IEEE Trans. Med. Imag.* **10** 572-88

Parra L and Barrett H H 1998 List–mode likelihood: EM algorithm and image quality estimation demonstrated on 2-D PET *IEEE Trans. Med. Imag.* **17** 228-35

Peng H and Stark H 1989 One-step image reconstruction from incomplete data in computer tomography *IEEE Trans. Med. Imag.* **8** 16-31

Peters T M 1981 Algorithms for fast back- and re-projection in computed tomography *IEEE Trans. Nucl. Sci.* **28** 3641-7

Phillips P R 1989 Bayesian statistics, factor analysis, and PET images – Part I: Mathematical background *IEEE Trans. Med. Imag.* **8** 125-32

Press W H, Teukolsky S A, Vetterling W T, and Flannery B P 1992 *Numerical recipes in C: The art of scientific computing* 2nd ed. (Cambridge University Press)

Qi J, Leahy R M, Cherry S R, Chatziioannou A, and Farquhar T H 1998a High-resolution 3D Bayesian image reconstruction using microPET small-animal scanner *Phys. Med. Biol.* **43** 1001-13

Qi J, Leahy R M, Hsu C, Farquhar T H, and Cherry S R 1998b Fully 3D Bayesian image reconstruction for the ECAT EXACT HR+ *IEEE Trans. Nucl. Sci.* **45** 1096-1103

Radon J 1986 On the determination of functions from their integrals values along certain manifolds *IEEE Trans. Med. Imag.* **5** 170-6 (Translated from the original text that appeared in 1917: *Berichte Saechsische Akademie des Wissenschaften* **29** 262-77)

Raichle M E 1998 Imaging the mind *Seminars in Nuclear Medicine* **28** 278-89

Rajan K, Patnaik L M, and Ramakrishna J 1997 High-speed parallel implementation of a modified PBR algorithm on a DSP-based EH topology *IEEE Trans. Nucl. Sci.* **44** 1658-72

Reader A J, Erlandsson K, Flower M A, and Ott R J 1998 Fast accurate iterative reconstruction for low-statistics positron volume imaging *Phys. Med. Biol.* **43** 835-46

Rockmore A J and Macovski A 1976 A maximum likelihood approach to emission image reconstruction from projections *IEEE Trans. Nucl. Sci.* **23** 1428-32

Rowland S W 1979 Computer implementation of image reconstruction formulas *Image Reconstruction from Projections: Implementation and Applications* (Herman G T ed.) *Topics in Applied Physics* **32** 9-79 (Berlin: Springer-Verlag)

Saha G B, MacIntyre W J, and Go R T 1992 Cyclotrons and Positron Emission Tomography radiopharmaceuticals for clinical imaging *Seminars in Nuclear Medicine* **22** 150-61

Sastry S and Carson R E 1997 Multimodality Bayesian algorithm for image reconstruction in positron emission tomography: A tissue composition model *IEEE Trans. Med. Imag.* **16** 750-61

Schmitt D, Karuta B, Carrier C, and Lecomte R 1988 Fast point spread function computation from aperture functions in high-resolution positron emission tomography *IEEE Trans. Med. Imag.* **7** 2-12

Shepp L A and Vardi Y 1982 Maximum likelihood reconstruction for emission tomography *IEEE Trans. Med. Imag.* **1** 113-22

Smith M F, Floyd C E Jr, Jaszczak R J, and Coleman R E 1992 Reconstruction of SPECT images using generalized matrix inverses *IEEE Trans. Med. Imag.* **11** 165-75

Smith M F 1996 Generalized matrix inverse reconstruction for SPECT using a weighted singular value spectrum *IEEE Trans. Nucl. Sci.* **43** 2008-17

Snyder D L, Thomas L J Jr, and Ter-Pogossian M M 1981 A mathematical model for positron-emission tomography systems having time-of-flight measurements *IEEE Trans. Nucl. Sci.* **28** 3575-83

Snyder D L and Politte D G 1983 Image reconstruction from list-mode data in an emission tomography system having time-of-flight measurements *IEEE Trans. Nucl. Sci.* **30** 1843-49

Snyder D L and Miller M I 1985 The use of sieves to stabilize images produced with the EM algorithm for emission tomography *IEEE Trans. Nucl. Sci.* **32** 3864-72

Soares E J, Byrne C L, and Glick S J 2000 Noise characterization of block-iterative reconstruction algorithms: I. Theory *IEEE Trans. Med. Imag.* **19** 261-70

Sorenson J A and Phelps M E 1987 *Physics in nuclear medicine* 2nd edn (Orlando: Grune & Stratton)

Strang G 1980 *Linear algebra and its applications* 2nd edn. 137-45 (Harcourt Brace Jovanovich)

Tanaka E 1987 A fast reconstruction algorithm for stationary positron emission tomography based on a modified EM algorithm *IEEE Trans. Med. Imag.* **6** 98-105

Terstegge A, Weber S, Herzog H, Müller-Gärtner H W, and Halling H 1997 High resolution and better quantification by tube of response modelling in 3D PET reconstruction *Proc. 1996 IEEE Nucl. Sci. Symp. & Med. Imag. Conf* **3** 1603-7

Thompson C J and Peters T M 1981 A fractional address accumulator for fast back-projection *IEEE Trans. Nucl. Sci.* **28** 3648-50

Thompson C J 1993 The problem of scatter correction in positron volume imaging *IEEE Trans. Med. Imag.* **12** 124-32

Tornai M P, Jaszczak R J, Turkington T G, and Coleman R E 1999 Small animal PET: Advent of a new era of PET research *J. Nucl. Med.* **40** 1176-9

Vardi Y, Shepp L A, and Kaufman L 1985 A statistical model for positron emission tomography (with comments by Carson R E and Lange K; by Herman G T, Censor Y, Gordon D, and Lewitt R M; by Kearfott K J; by Laird N M and Maher E F; by Rubin D B) *J. Amer. Stat. Assoc.* **80** 8-37

Veklerov E and Llacer J 1987 Stopping rule for the MLE algorithm based on statistical hypothesis testing *IEEE Trans. Med. Imag.* **6** 313-9

Veklerov E and Llacer J 1990 The feasibility of images reconstructed with the method of sieves *IEEE Trans. Nucl. Sci.* **37** 835-41

Vogel C R 1986 Optimal choice of a truncation level for the truncated SVD solution of linear first kind integral equations when the data are noisy *SIAM J. Numer. Anal.* **23** 109–17

Vogel C R and Wade J G 1994 Iterative SVD-based methods for ill-posed problems *SIAM J. Sci. Comput.* **15** 736-54

Volkow N D, Rosen B, and Farde L 1997 Imaging the living human brain: Magnetic resonance imaging and positron emission tomography *Proc. Natl. Acad. Sci. USA* **94** 2787-8

Wahba G 1980 Ill posed problems: Numerical and statistical methods for mildly, moderately, and severely ill posed problems with noisy data *Tech. Report No.595* Department of Statistics - University of Wisconsin - Madison

Wahba G 1987 Three topics in ill posed inverse problems in *Inverse and Ill-Posed Problems* (Engl M and Groetsch G eds.) 37-51 (Academic Press)

Waldén J 2000 Analysis of the direct Fourier method for computer tomography *IEEE Trans. Med. Imag.* **19** 211-22

Wernick M N, Infusino E J, and Milošević M 1999 Fast spatio-temporal image reconstruction for dynamic PET *IEEE Trans. Med. Imag.* **18** 185-95

Wu C F J 1983 On the convergence properties of the EM algorithm *Annals of Statistics* **11** 95-103

Xu P 1998 Truncated SVD methods for discrete linear ill-posed problems *Geophys. J. Int.* **135** 505-14

Xu X L, Liow J S, and Strother S C 1993 Iterative algebraic reconstruction algorithms for emission computed tomography: A unified framework and its application to positron emission tomography *Med. Phys.* **20** 1675-84

Index

A

algebraic reconstruction technique 35
algorithms
  deterministic 26
  iterative 26
  one-step 26
  series-expansion 26
  statistical 26
  transform 26
ART *See* algebraic reconstruction technique

B

backprojection 20
  filtered 30
Bayesian approach 40

C

coincidence event 4
convolution method 30
count 7
cross-validation 34

E

emission computed tomography 6
expansion functions 14

F

feasibility region 39
first-kind Fredholm integral equation 9

I

image 6
  estimate 7, 31
  feasible 39
  reconstruction 5, 7
  reconstruction in PET 13
  representation space 14
  space 9

L

line-of-response 4
LOR *See* line-of-response

M

matrix
  condition number 23
  ill conditioned 23
  pseudo-inverse 34
  singular 23
  system 20
  transition 20
maximum likelihood estimate 37
method
  direct Fourier 30
  summation 28
ML-EM 37
model
  deterministic 12
  stochastic 12

N

noise artefact 38
Nyquist sampling rate 31

O

operator
  injective 21
  invertible 22
  singular 22
  surjective 22
  the domain of 21
  the nullspace of 21
  the range of 21
orthonormal basis 14
OS-EM 42

P

penalised weighted least-squares 41
pixel 15
  natural 16
  polar 15
problem
  direct 20
  ill posed 10
  inverse 7
  mildly ill posed 23
  severely ill posed 23
  well posed 10
projection
  bins 18
  data 4
  of an object 18

R

radiopharmaceutical 3
radiotracer 3
ramp filter 31
regularisation 24

S

singular
- *matrix* 23
- *value* 22
- *value decomposition* 22
- *value spectrum* 23

singular vectors
- *left* 22
- *right* 22

sinogram 18
SPECT 5
statistical model 10
stopping rule 36, 38
SVD *See* singular value decomposition

T

theorem
- *Bayes'* 40
- *central limit* 125
- *central slice* 29
- *Fourier slice* 29
- *projection slice* 29

TOR *See* tube-of-response
tracer 3
transform
- *Fourier* 29
- *inverse Fourier* 29
- *Radon* 8
- *X-ray* 8

truncation level 34
TSVD 34
tube-of-response 16

V

voxel 15

W

weighted least-squares 41