Penalised Image Reconstruction Algorithms for Efficient and Consistent Quantification in Emission Tomography

Yu-Jung Tsai

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

in

Medical Physics.

Institute of Nuclear Medicine

Division of Medicine

University College London

January 8, 2019

I, Yu-Jung Tsai, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

With the increased interest in potential clinical applications based on quantitative results, the aim of this study is to improve the quantitative consistency of the reconstructed images in emission tomography (ET).

To achieve practical processing time, a fast convergent quasi-Newton algorithm, preconditioned limited-memory Broyden-Fletcher-Goldfarb-Shanno with boundary constraints (L-BFGS-B-PC), is firstly proposed. Its performance is evaluated using both simulations and three patient datasets. Results show that L-BFGS-B-PC is able to achieve several times faster convergence rate than separable paraboloidal surrogates (SPS). Moreover, the performance is less sensitive to penalty type, penalty strength, noise level and background level, compared to L-BFGS-B.

To be able to improve the image quality and quantitative consistency, an anatomical penalty function is then considered with a spatially-variant penalty strength. Based on results for simulations and data from one patient with inserted pseudo lesions, the spatially-variant penalty reduces the quantitative dependence on the surrounding activity and location. Moreover, it benefits the algorithm convergence rate and its consistency among datasets.

It is important to consider potential misalignment between the functional and anatomical images. For this reason, two approaches that perform alternating penalised image reconstruction and misalignment estimation are therefore proposed. Expanding on the previous work, L-BFGS-B-PC using Parallel Level Sets (PLS) with the spatially-variant penalty strength is used in both approaches. Preliminary results for non-time-of-flight (non-TOF) data simulations demonstrate that both

methods are able to estimate the misalignment and deform the anatomical image accordingly when a proper workflow for the alternating optimisation is applied. By integrating algorithms proposed in this study, both good image quality and consistent quantification can be achieved in a practical processing time.

Impact Statement

A fast convergent penalised image reconstruction algorithm is proposed in this thesis in order to achieve practical processing time. To obtain consistent quantitative results, efforts have been made on eliminating variations of local contrast in individuals, locations and surroundings. The image quality of the reconstructed activity image is also improved by incorporating anatomical information. In consideration of the potential misalignment between the functional and anatomical images in practice, two approaches that perform alternating misalignment estimation and image reconstruction are therefore introduced. All proposed algorithms are thoroughly evaluated with simulated datasets and promising results are obtained. Except for the approaches for solving the misalignment issue, demonstrations with one to three patient datasets are also provided.

The use of the presented methods to improve the quantitative consistency in emission tomography have been demonstrated in this thesis. Their application in practice could benefit the reliability of interpretations based on quantitative results without compromising the patient throughput. With the increased interest in image quantification, these algorithms could be utilised by researchers for exploring potential clinical applications, and could be incorporated into commercial software by manufacturers. The impact of applying methods presented in this thesis could be positive to public health in the medical imaging field on a national and international level.

The presented algorithms are based on proven methods with certain modifications, which allow re-using existing implementations. The feasibility of generalising the proposed strategies to a wider class of reconstruction algorithms and penalty

functions has been discussed. The implementation of the proposed methods should only require small changes to the selected algorithm and prior, which in turn would result in reasonable implications to the manufacturing and economics.

Acknowledgements

This project is supported by UCL Overseas Research studentship 520683, by GE Healthcare, and by the National Institute for Health Research, University College London Hospitals Biomedical Research Centre.

Apart from the sponsors, I would like to take this opportunity to thank my supervisors, Prof. Kris Thielemans, Prof. Simon Arridge and Dr. Alexandre Bousse, for being very supportive along my study. Whenever I got lost and frustrated, they were always on my side to help me with valuable inputs and guidance.

I also want to thank my colleagues, Dr. Debora Salvado and Dr. Richard Brown, for proofreading my upgrade report and this thesis. I cannot be more grateful for your kindness, knowing how much effort and time you have spent with me to improve my English writing.

Thanks all my friends and colleagues at INM (Institute of Nuclear Medicine), especially Alaleh, Ottavia, Debora, Ludovica, Richard and Elise. I will never forget those days we share tears and laughs. Without any of you, I cannot make it so close to the finish line. We had our ups and downs but we were always there for each other. Thank you again for being a part of my life.

A special thank you to my Taiwanese friends in London, Shih-hsin Chen and Tzu-ling Lin. Your support have made London my home away from home. Last but not least, a big thank you goes to my family in Taiwan. Thank you for forgiving me not being around when you need me. Thank you for having believe in me.

Contents

List of Figures

List of Tables

List of Abbreviations

1-D 1-Dimensional

2-D 2-Dimensional

3-D 3-Dimensional

4-D 4-Dimensional

CT Computed Tomography

ET Emission Tomography

FBP Filtered-Back-Projection

FORB Fourier Rebinning

FOV Field-of-View

FWHM Full Width at Half Maximum

HU Hounsfield Units

JTV Joint Total Variation

L-BFGS-B-PC Preconditioned L-BFGS-B

L-BFGS-B Limited-memory Broyden-Fletcher-Goldfarb-Shanno with Boundary Constraints

L-BFGS Limited-memory Broyden-Fletcher-Goldfarb-Shanno

LLPR Linearised Local Perturbation Response

LOR Line of Response

LPR Local Perturbation Response

MC Monte Carlo

ML-EM ML-Expectation Maximization

ML Maximum-Likelihood

MR Magnetic Resonance

OSL One-Step-Late

OS Ordered-Subset

PET Positron Emission Tomography

PLS Parallel Level Sets

PML Penalised Maximum-Likelihood

PMT Photo-Multiplier Tube

RMSE Root-Mean-Square Error

ROI Region-of-Interest

SiPMs Silicon Photo-Multipliers

SNR Signal-to-Noise Ratio

SPECT Single Photon Emission Computed Tomography

SPS Separable Paraboloidal Surrogates

SSRB Single Slice Rebinning

SUV Standardized Uptake Value

TBR True to Background Ratio

TOF Time-of-Flight

TV Total Variation

WCs Wolfe Conditions

Chapter 1

Introduction

In current clinical practice of emission tomography, most diagnostic applications are based on visual interpretation of reconstructed images. With the expansion of its potential clinical use, such as disease follow-up and therapy monitoring [2, 3, 4], there is increased interest in quantification. The reconstructed images are therefore expected to accurately represent the tracer concentration.

The aim of this study is to improve quantitative consistency in emission tomography without compromising the patient throughput in practice. The proposed algorithms are able to achieve consistent convergence rate among different datasets in clinically feasible reconstruction time and good image quality when auxiliary information (e.g. anatomical information) is available.

1.1 Thesis overview

Chapter 2 includes general background on imaging physics and reconstruction methods in ET. Basic principles of one of the anatomical imaging modalities, computed tomography (CT), are also provided.

Chapter 3 investigates the feasibility of using a quasi-Newton optimisation algorithm, L-BFGS-B, for penalised image reconstruction problems in order to achieve practical processing time. For further acceleration, an additional preconditioning technique based on a diagonal approximation of the Hessian is introduced. The convergence rate of L-BFGS-B and the proposed preconditioned algorithm (L-BFGS-B-PC) is evaluated using simulated data with various factors, such as the

noise level, penalty type, penalty strength and background level in the chapter. Data of three ^{18}F-FDG patient acquisitions are also reconstructed.

Chapter 4 explores the use of a spatially-variant penalty strength with an anatomical penalty function to improve both the image quality and quantitative consistency. Since Parallel Level Sets (PLS) is a convex function and has shown promising results in the literature, it is chosen as the representative anatomical penalty and incorporated into the previously proposed preconditioned algorithm (L-BFGS-B-PC). A 2-dimensional (2-D) disc phantom with a hot feature at the centre and a 3-dimensional (3-D) XCAT thorax phantom with lesions inserted in different slices are used respectively to study how surrounding activity and lesion location affect both the visual appearance and quantitative consistency. Anatomical information is provided by the attenuation map or CT images and assumed to be well-aligned with the corresponding activity images. The consistency of the algorithm convergence rate with respect to different data noise and background levels is also investigated using the XCAT phantom. An example reconstruction for a patient dataset with pseudo lesions is used as a demonstration in a clinical context.

Chapter 5 presents two approaches for solving the potential misalignment between the functional and anatomical images. Both methods are based on a recently published joint motion estimation and image reconstruction method. The first approach deforms the anatomical image to align it with the functional one, while the second approach deforms both images to align them with the measured data. The current implementation alternates between image reconstruction and alignment estimation. Expanding on the previous work, PLS with a spatially-variant penalty strength is incorporated into both approaches. To evaluate their performances , simulated non-TOF data generated with a XCAT phantom is used.

Chapter 6 makes general conclusions based on results presented in this thesis, followed by discussion on potential future work. Current and expected publications are listed in the chapter as well.

Chapter 2

Background

In this chapter the basic principles of ET are addressed, summarising its current and possible clinical applications, imaging physics, data acquisition and data organisation. Principles and challenges of image reconstructions in ET using analytic and iterative algorithms are also provided, followed by a discussion on the desired characteristics of reconstruction algorithms in practice. Due to the increasing availability of multi-modality scanners that integrate benefits of functional and anatomical imaging, general principles and current applications for one of the anatomic imaging systems, CT, is also presented.

2.1 Emission tomography

ET is a medical strategy that allows non-invasive observation of metabolic processes *in vivo*. With adequate image processing and analysis methods, it is valuable for the diagnosis of many diseases. The technique has been commonly used in contemporary nuclear medicine [5, 6, 7, 8, 9, 10]. Its application varies with the radioactive tracer and the imaging protocol being used.

Radioactive tracers are analogues of chemicals in the human body with one atom a radioisotope. The analogues are taken up in particular organs or involved in similar physiological processes. Depending on the radioisotope, γ-rays or positrons are emitted during decay. The emissions are then detected by an external scanner specific to the type of emission.

To be able to obtain cross-sectional images of the tracer distribution, two

imaging methods, Single Photon Emission Computed Tomography (SPECT) and Positron Emission Tomography (PET), are applied for γ-ray and positron emitters, respectively. Basic principles, such as data acquisition and image reconstruction for both methods, will be presented in the following sections.

2.1.1 Current and potential clinical applications

ET is useful for imaging objects in a complex background of heterogeneous tissues, such as brain, thorax and abdomen. Currently, its major applications are in the following fields:

- *Oncology* – many cancerous cells show specific accumulation of chemicals or over-expression of receptors, which can be targeted by corresponding tracers. For example, ^{99m}Tc sestamibi has been suggested for early detection of breast cancer in SPECT [11] and ^{18}F-FDG (fluorodeoxy-glucose) is a common tracer of choice in PET as it reflects the high metabolism rate of glucose in malignant cells [12].

- *Cardiology* – myocardial perfusion studies are helpful in the diagnosis of coronary artery disease and heart muscle damage. Cells that are damaged or lack blood supply have no tracer uptake on the images. Examinations are usually carried out using ^{201}Tl in SPECT [13] and ^{13}N-ammonia in PET [14].

- *Neurology* – some neurodegenerative diseases (e.g. Alzheimer's and Parkinson's disease) and psychiatric disorders can be detected using ET. Many tracers that are able to enter the blood brain barrier have been developed in recent decades. For example, ^{99m}Tc-HMPAO (hexamethylpropyleneamine oxime), ^{11}C-PiB (Pittsburgh Compound-B) and related radiotracers labelled with ^{18}F are commonly used tracers for examination of Alzheimer's disease using SPECT [15] and PET [16], respectively.

The application of ET is increasing as novel radioactive tracers are being developed. In recent decades, potential clinical applications based on image quantification are being explored intensively as well. Many studies have proposed using metrics de-

rived from the intensity of the reconstructed images as indicators for disease follow-up, treatment response assessment or receptor occupancy analysis [2, 3, 4].

2.1.2 Scanners and principles

A SPECT scanner detects emissions from radioisotopes decaying by γ-ray emission. It usually consists of two or more detector heads that rotate around the patient to obtain projections from multiple angles. Since γ-rays are emitted in arbitrary directions, physical collimators are required for limiting the incidence angle of the rays, improving the spatial resolution at each projection angle. The mean energy of the emitted γ-rays can vary among different radioisotopes. Therefore, collimators should be changed accordingly.

PET scanners are designed especially for positron-emitting radioisotopes. When interacting with matter, the positron loses its kinetic energy until it annihilates with an electron, creating two photons that are emitted almost back-to-back. The annihilation photons carry the same energy, 511 keV, independent of the labelled positron-emitting radioisotope. To capture these two photons at the same time and register them as an event, coincidence detectors positioned in opposite directions are used. Since the location of the emitter can be traced along the straight line defined by the paired coincidence detectors (line of response, LOR), physical collimators are unnecessary for PET. Instead of rotating them around the patient as in SPECT, current PET scanners adopt stationary detectors arranged in a ring shape to achieve multiple angle acquisition. PET scanners usually consist of parallel detector rings along the transaxial direction. For some of them, thin septa of lead or tungsten are used to block coincidences lying in different detector rings or outside a predefined ring difference.

For converting the energy of γ-rays or annihilation photons to an electric signal, most of SPECT and PET detectors consist of scintillation crystals (scintillator) coupled with a detector for the optical photons, such as a photo-multiplier tube (PMT) and a Silicon photo-multipliers (SiPMs) tube. The optical photons are then converted to electrical signals in the detector.

2.1.3 Data acquisition

When a patient is sent to a nuclear medicine department, a small amount of radiotracer is injected for the required examination. Based on different imaging physics, the data acquisition and organisation processes are different for SPECT and PET.

2.1.3.1 Data acquisition in SPECT

SPECT scanners usually acquire 2-D projections at equally-spaced angular intervals around the patient. One single projection corresponds to the accumulation of the tracer distribution for a specific time frame and along the solid angle limited by the collimator. The cross-sectional images are reconstructed from multiple projections using algorithms that will be discussed in Section 2.3. Generally, the number of projections required for the reconstruction ranges from 64 to 128 in clinical practice.

There are four types of events that may happen in a SPECT scanner, however only valid events provide correct positional information:

- *Valid event* – an event caused by a γ-ray entering directly to the detector without previous interaction and depositing all of its energy at one position.

- *Detector scattered event* – when a γ-ray changes its direction in the detector by Compton scattering. The scattered γ-ray can either escape the detector or interact a second time. In the former case, only part of the γ-ray energy is deposited in the detector, therefore, the event can be rejected by setting a low level energy discriminator. However, in the latter case, the event will be recorded and mispositioned between the two interaction locations. The energy discriminator cannot disregard these events, because all the energy of the ray has been deposited in the detector.

- *Object scattered event* – when a γ-ray changes its direction in any object along its travel to the detector by Compton scattering. For the γ-ray scattered by an object before reaching the detector, the energy discriminator can be used to reject the event, if the scattering angle is large enough for the discriminator to recognise the energy loss caused by the direction change.

- *Septal penetration* – when a γ-ray enters non-parallel to the collimator holes, there is a chance that the ray penetrates one or more septa and interacts with the scintillator, especially for high energy rays. These events cannot be rejected by the discriminator and will lead to mispositional information.

Therefore, the total recorded events $\boldsymbol{g}$ is actually composed of valid events, detector scattered events, object scattered events and events due to septal penetration. This can be expressed using the following equation:

$$\boldsymbol{g} = \boldsymbol{g}^{\text{valid}} + \boldsymbol{g}^{\text{detector scattered}} + \boldsymbol{g}^{\text{object scattered}} + \boldsymbol{g}^{\text{septal penetration}} \tag{2.1}$$

2.1.3.2 Data acquisition in PET

Due to the back-to-back directional characteristic of the annihilation photons, PET scanners record every event belonging to a LOR defined by a pair of coincidence detectors. The ring shape arrangement of the detectors allows simultaneous acquisition of events from various angles. The data acquisition mode of PET can be either 2-D or 3-D. In the 2-D mode, annihilation photons, which would be detected by a pair of detectors belonging to different rings, are blocked by septa. In order to improve the detection sensitivity, septa are retracted in the 3-D acquisition mode. Therefore, LORs lying across different rings are accepted.

In PET, four types of events could be accepted by the paired coincidence detectors in 3D mode. The correct positional information is provided only by valid events:

- *Valid event* – an event occurred from a pair of photons generated by the same annihilation event and detected in a pair of opposite detectors within a specified coincidence timing window without previous interaction.

- *Detector scattered event* – when one or both annihilation photons from an annihilation event change their direction by Compton scattering in the detector. The scattered photon(s) can be rejected by using an energy discriminator if the energy loss caused by the direction change is large enough. Otherwise,

they will be detected in a wrong coincidence detector pair and lead to mispositioning of the event.

- *Object scattered event* – when one or both annihilation photons from an annihilation event change their direction by Compton scattering in any object along their travel to the detector. Similar to the scattering event occurring within the detector, they can lead to mispositioning of the event if the energy loss is too small for an energy discriminator to disregard them.

- *Random* – events caused by accidentally detecting two photons from two unrelated annihilation events in a pair of detectors within the coincidence timing window. Since annihilation events can happen anywhere in the field-of-view (FOV), this kind of mispositioning can lead to apparent LORs far away from the true location.

Similar to the data in SPECT, in PET, the collected events $\boldsymbol{g}$ are composed of valid events, two types of scattered events and randoms:

$$\boldsymbol{g} = \boldsymbol{g}^{\text{valid}} + \boldsymbol{g}^{\text{detectorscattered}} + \boldsymbol{g}^{\text{objectscattered}} + \boldsymbol{g}^{\text{randoms}} \tag{2.2}$$

2.1.4 Data organisation

In 2-D data organisation, a coordinate system (r,s), stationary with respect to the detector, is introduced to reduce the complexity of SPECT and PET data analysis. Given the data collecting angle ϕ, the transformation from the (x,y) coordinate system of the imaged object to (r,s) is illustrated in Figure 2.1 and can be formulated by:

$$r = x\cos\phi + y\sin\phi \tag{2.3}$$

$$s = y\cos\phi - x\sin\phi \tag{2.4}$$

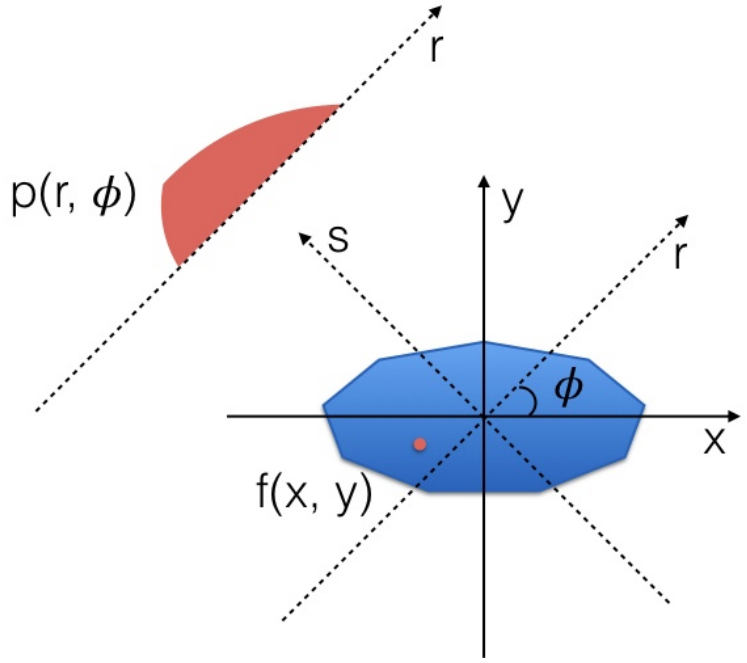

Figure 2.1: Transformation of the coordinate system from (x,y) to (r,s).

2.1.4.1 Data organisation in SPECT

With the help of this coordinate system, a full set of SPECT projection data for a single slice can be described by a 2-D matrix $\boldsymbol{p}$. Each row of the matrix represents the projection acquired at angle ϕ and each element in a row records the number of detected events at location r. The matrix is also referred to as a *sinogram*, since a sinusoidal path down the matrix is observed when imaging a point source at location (x,y) (Figure 2.2). The equation that describes the 2-D relationship between the matrix element $p(r,\phi)$ and the tracer distribution at location (x,y), ignoring physical effects, such as attenuation, is given by [17]:

$$p(r,\phi) = \int_{-\infty}^{\infty} f(r\cos\phi - s\sin\phi, r\sin\phi + s\cos\phi)ds \tag{2.5}$$

For a 3-D volume, the data organisation can be seen as a stack of sinograms that record projections belonging to each slice z in the FOV.

2.1.4.2 Data organisation in PET

As mentioned in Section 2.1.3.2, PET scanners record every single event instead of projections from different angles. To describe PET data in the manner of sinograms,

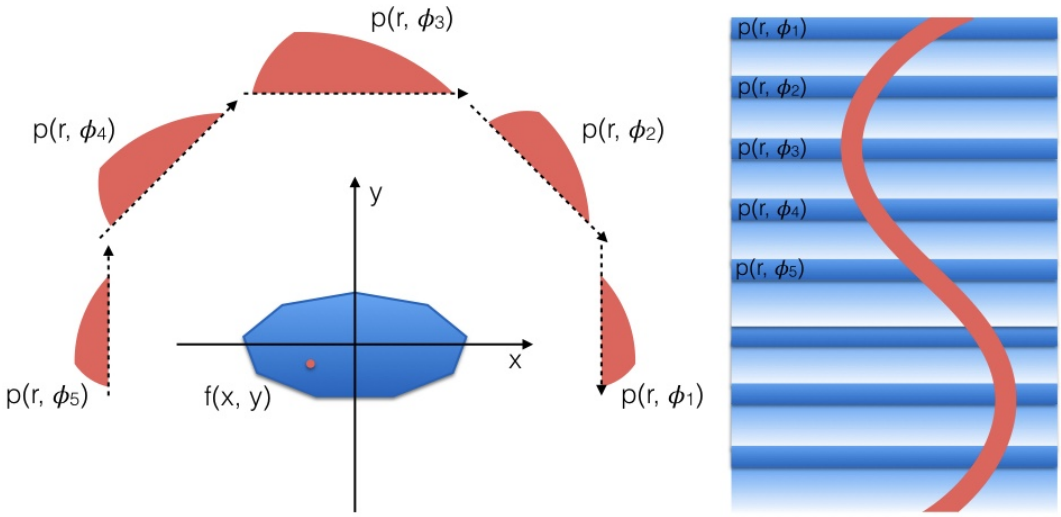

Figure 2.2: Illustration of the sinogram in SPECT.

an extra step is required. For a specific time frame, events are accumulated and combined with adjacent parallel LORs to represent projection data at the particular angle. An example is shown in Figure 2.3.

After the projection construction step, data organisation in 2-D PET is identical to that in SPECT when physical effects are not considered. Data acquired in 3-D mode can also be described by a stack of sinograms with additional sinograms representing projections at different polar angle θ. The coordinate system of scanners in 3-D acquisition is illustrated in Figure 2.4.

2.1.5 Attenuation effect and correction

As γ-rays and annihilation photons can be scattered or absorbed by the patient body before interacting with the detector, events recorded in the projection data will not reflect true activity distribution in the patient. The probability of detection therefore depends on the distance and material that these emissions have to travel through. Given the differences in imaging principles, the measured SPECT and PET data for a single slice considering the attenuation effect can be described as follows:

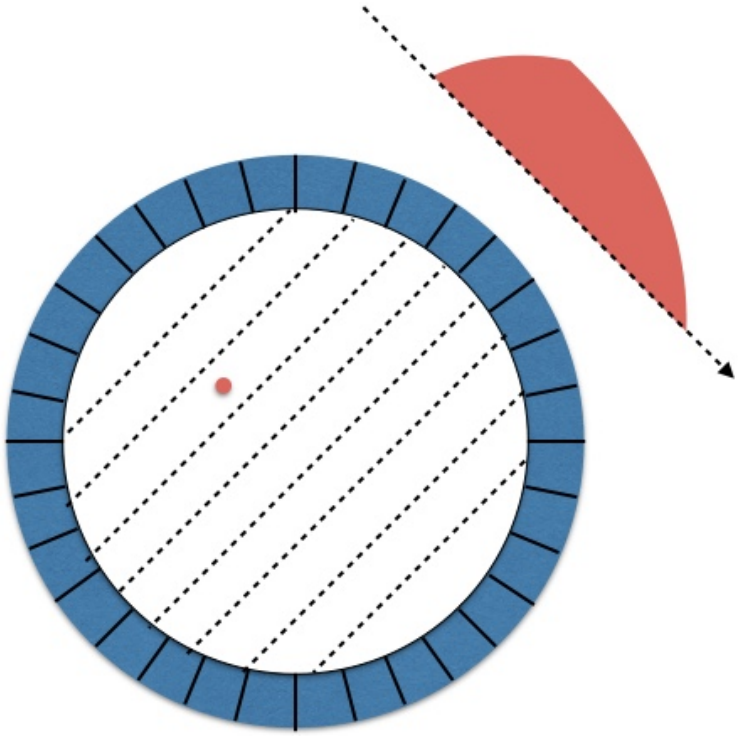

Figure 2.3: An example of constructing one projection from adjacent parallel LORs.

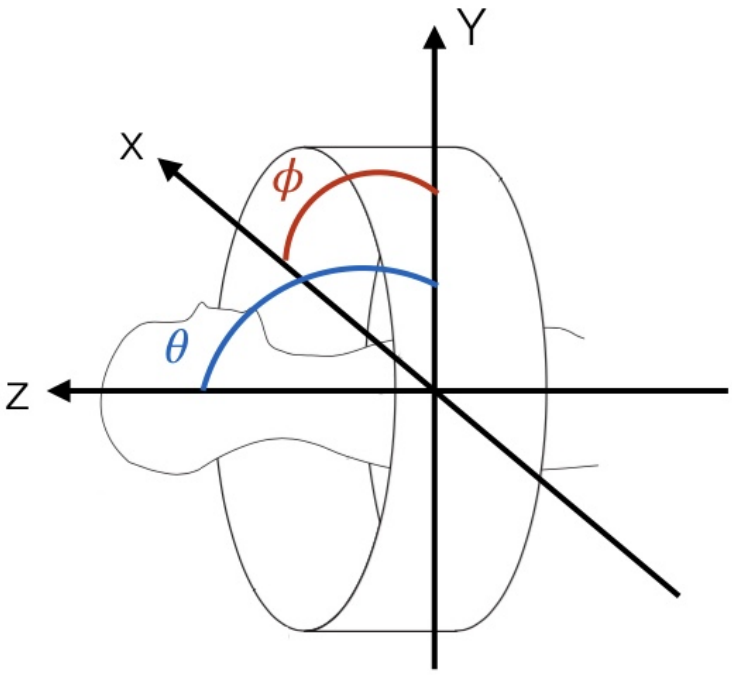

Figure 2.4: Coordinate system for 3-D PET acquisition [1].

SPECT:

$$p(r,\phi) = \int_{-\infty}^{\infty} f(r\cos\phi - s\sin\phi, r\sin\phi + s\cos\phi) e^{-\int_{-\infty}^{s} f(r\cos\phi - s'\sin\phi, r\sin\phi + s'\cos\phi) ds'} ds \tag{2.6}$$

PET:

$$p(r,\phi) = \int_{-\infty}^{\infty} f(r\cos\phi - s\sin\phi, r\sin\phi + s\cos\phi) e^{-\int_{-\infty}^{\infty} f(r\cos\phi - s'\sin\phi, r\sin\phi + s'\cos\phi) ds'} ds \tag{2.7}$$

Without considering the attenuation of the photons, images reconstructed from the data cannot represent the true activity distribution in the object. For emissions surrounded by dense materials, such as bone, the reduction of events due to attenuation is severe and artefacts in the reconstructed images will occur. In addition to interfering with visual interpretation, the image quantification also becomes unreliable. In contemporary nuclear medicine, the reduction can be estimated and compensated by using anatomic information provided by an additional CT or MR scan.

2.2 Anatomical imaging modalities

In contrast to scanners for ET that allow *in vivo* observation of physiological processes, anatomic imaging modalities, such as CT, provide insight into structural detail regarding tissue size, location and morphological change [18, 19, 20]. The anatomical information can be used for performing attenuation correction in ET [21, 22] or assisting diagnosis [23, 24], follow-up [25] and treatment planning in oncology [26, 27]. Although there are other imaging modalities that are able to provide anatomical information, this thesis focuses on the application of CT in ET. The basic principles of CT will be addressed in the following section.

2.2.1 Computed tomography

CT is a technique that generates anatomic cross-sectional images of a patient using external x-ray measurements taken from different angles. Similar to conventional x-ray imaging that acquires only one projection of the patient, CT is based on the

variable absorption of x-rays by different tissues.

2.2.1.1 Basic principles

A CT scanner consists of a motorised table that moves the patient into the imaging position for the required examination and a donut-shaped gantry, on which an x-ray tube is mounted. An x-ray tube is a vacuum tube where electrons emitted from the cathode are accelerated toward the anode using a high potential difference. The electrical input is then converted to x-rays as the electrons interact with the heavy nuclei of the anode and lose their energy. The spectrum of the x-rays depends on the material of the anode and the strength of the accelerating voltage. A typical CT x-ray tube produces photons with energy levels between 20 and 150 keV. For recent generations of CT scanners, the resulting x-rays are distributed into a narrow and fan-shaped beam so that the patient can be completely irradiated at each imaging angle as the tube rotates around the inside of the circular opening.

While passing through the patient, some of the x-rays are absorbed by tissues in their travelling path. The remaining or attenuated x-rays are then measured by the digital x-ray detectors positioned diametrically opposite to the x-ray tube. To be able to convert the registered x-ray to electrical signals, the detector usually consists of a scintillator coupled with a photosensitive device, such as a photodiode. The most common CT scanners in current practice are composed of rotating detectors that allows helical data acquisition. For this kind of acquisition, the table moves through the circular opening of the gantry whilst the x-ray tube and detectors rotate around the patient.

2.2.1.2 Image generation

The voxel value of a CT image represents the ability of x-ray absorption for different tissues in Hounsfield Units (HU), which are calculated based on the linear attenuation coefficient μ of the corresponding tissue:

$$HU = 1000 \times \frac{\mu - \mu_{\text{water}}}{\mu_{\text{water}} - \mu_{air}} \tag{2.8}$$

where μ_{water} and μ_{air} are the linear attenuation coefficient of water and air.

2.2.1.3 Contrast-enhanced CT

The purpose of contrast-enhanced CT is to find pathology by visualising blood flow and washout using non-specific iodinated contrast agents. Depending on the pathology of interest, the optimal contrast injection rate and data acquisition time can vary. For example, helical CT imaging that starts 35 seconds, 70 seconds and 10 minutes after contrast injection at 5 ml/second is routinely used for patient with liver masses [28, 29]. It is especially useful for observing lesions having similar linear attenuation coefficient but different vascular density to surrounding tissues, as the wash-in or clearance rate of the agent is different between them. In practice, the technique is often used to delineate abnormal regions for neoplasm or inflammation in relatively uniform soft tissues, such as liver and lung [30, 31, 32, 33, 34].

2.3 Image reconstruction

The goal of image reconstruction is to obtain a cross-sectional image of the tracer distribution from the corresponding set of projections (sinogram) acquired around the object. To achieve this goal, there are two main categories of algorithms in ET, analytic image reconstruction and iterative image reconstruction. The former considers the reconstruction as a direct inverse problem, without considering any uncertainty, while the latter takes into account the statistical properties of the emission and imaging processes [1]. Principles and challenges for each kind of algorithms will be discussed in this section. However, as this project concentrates on iterative algorithms, analytic ones will be only covered briefly. Although the explanation is given in a single slice manner, the algorithms can be generalised to the case of 3-D PET data.

2.3.1 Analytic algorithms

The most basic approach for image reconstruction is by simple back-projection, which inverts the process by which the projections are created. It distributes events stored in each sinogram element back along the path they are collected from, but in the discrete image plane. Each pixel intersected by the projection path shares the total accumulated events equally, since the exact location information has been

lost during the projection process. By back-projecting all projections acquired from different angles and adding them up together, an approximation of the tracer distribution $f(x,y)$ is obtained. However, images reconstructed with this method are intrinsically blurred because the same activity value is assigned to pixels outside the true activity distribution but intersected by the back-projection path. A ramp filter can therefore be applied to compensate for the blurring effect, and the algorithm becomes filtered-back-projection (FBP). For analytic reconstruction algorithms, data pre-correction for scatter and attenuation effect is required.

Although advantages, such as fast computational speed, have made FBP a popular method for image reconstruction, it has some limitations. First, it introduces severe artefacts if the acquisition is incomplete. Second, serious streak artefacts can be observed for data representing poor counting statistics (e.g. few events). Third, the FBP algorithm cannot be modified to take into account the physical effects of the imaging system, such as the shift-variant intrinsic spatial resolution of the detector.

2.3.2 Iterative algorithms

Due to the ability of modelling the imaging physics and statistics, iterative reconstruction algorithms have become the method of choice in ET. As all available information about the measurement can be incorporated into these methods, a more reliable quantification based on the resulting image is expected. In contrast to analytic methods that obtain images in one step, iterative algorithms adopt successive estimates toward the solution of the objective function (see Section 2.3.2.2 for more information) based on the actual measured data (sinogram(s)) using an *update function*.

2.3.2.1 Imaging model

In this thesis, the imaging model is discretised using voxels, although other discretisations are possible. Given the discretised tracer distribution $\boldsymbol{f} = [f_1, \ldots, f_J]^\top \in \mathbb{R}^J$ and the expected invalid events $\boldsymbol{n} = [n_1, \ldots, n_I]^\top \in \mathbb{R}^I$ (see Section 2.1.3 for more information), the imaging process in ET can be described using the following equa-

tion:

$$\bar{\boldsymbol{g}} = \boldsymbol{A}\boldsymbol{f} + \boldsymbol{n} \tag{2.9}$$

where $\bar{\boldsymbol{g}} = [g_1, \ldots, g_I]^\top \in \mathbb{R}^I$ is the (statistical) mean of the observed data and $\boldsymbol{A}$ is an $I \times J$ transition matrix with each element A_{ij} indicating the probability that an emission from voxel j is detected by detector bin i without scattering. The bin i corresponds to the i^{th} element of the sinogram. The matrix $\boldsymbol{A}$ is also referred to as the system matrix, which characterises the physical system properties, such as resolution and detector sensitivity, in terms of detection probability. The attenuation effect mentioned in Section 2.1.5 is another factor in the calculation of the system matrix $\boldsymbol{A}$.

2.3.2.2 Objective function

In the context of image reconstruction, an objective function quantifies the difference between the measured and estimated data. It reaches the optimal value when the closest estimate to the measured data is found. In ET, the probability of collecting data $\boldsymbol{g} \in \mathbb{N}^I$ given a tracer distribution $\boldsymbol{f}$ can be described using a Poisson model:

$$\boldsymbol{P}(\boldsymbol{f}, \boldsymbol{g}) = \prod_{i=1}^{N} \frac{\bar{g}_i^{g_i} \exp(-\bar{g}_i)}{g_i!} \tag{2.10}$$

where $\bar{g}_i$ is the observed mean value in the i^{th} bin as defined in (2.9). The probability function of $\boldsymbol{g}$ can be interpreted as a likelihood function of $\boldsymbol{f}$. Therefore, maximising the probability (2.10) is equivalent to finding an estimate of the distribution $\hat{\boldsymbol{f}}$ that gives the maximum value of the likelihood. This approach is known as maximum-likelihood (ML), which is a common strategy for iterative reconstructions. Taking the logarithm and omitting terms independent of $\boldsymbol{f}$, the objective function based on the log-likelihood estimation is obtained:

$$L(\boldsymbol{f}, \boldsymbol{g}) = \sum_i g_i \log \bar{g}_i(\boldsymbol{f}) - \bar{g}_i(\boldsymbol{f}) \,. \tag{2.11}$$

maximising L is equivalent to minimising $-L$. The optimisation therefore is referred to as a minimisation problem in the rest of the thesis.

However, image reconstruction using ML estimation is an ill-conditioned problem, resulting in noise amplification as iterations increase [35]. In practice, the noise can be controlled by early termination of the iterative process, at the expense of quantitative accuracy [36], or by incorporation of a penalty term [1, 37]. Instead of optimizing (2.11), penalised ML (PML) image reconstruction minimizes an objective function Φ, which consists of the negative log-likelihood $-L$ and a penalty function R with a parameter β controlling its strength:

$$\Phi(\boldsymbol{f}) = -L(\boldsymbol{f}, \boldsymbol{g}) + \beta R(\boldsymbol{f}) \,. \tag{2.12}$$

The optimisation of the problem can then be addressed as:

$$\hat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f} \geq \boldsymbol{0}} \Phi(\boldsymbol{f}) \,. \tag{2.13}$$

A positivity constraint is normally enforced on $\boldsymbol{f}$, as it represents radioactivity concentration.

Several penalty functions can be used to control noise propagation [38, 39, 40]. It is referred to as the penalty term or prior in this thesis. The most common Gibbs-type penalties, which penalise the difference between voxels in a given neighbourhood $\mathcal{N}$ are firstly applied:

$$R(\boldsymbol{f}) = \frac{1}{2} \sum_j \sum_{k \in \mathcal{N}_j} \omega_{jk} \varphi(f_j - f_k) \tag{2.14}$$

where ω_{jk} indicates the weight between voxel j and its neighbouring voxel k. As a comparison study between different reconstruction conditions and algorithms, two relatively simple potential functions φ are used in Chapter 3:

- *quadratic penalty (QP)* – is a smoothing prior tending to reduce the pixel difference regardless of the presence of true edges.

$$\varphi_{\mathrm{QP}}(x) = x^2 \tag{2.15}$$

- *rescaled log-cosh penalty (LP)* – is a more edge preserving prior [41].

$$\varphi_{\mathrm{LP}}(x) = \frac{1}{\rho^2}\log(\cosh(\rho x)) \tag{2.16}$$

where ρ is a scalar controlling the edge-preservation property of φ_{LP}. The factor $1/\rho^2$ is derived from the second derivative of φ_{LP} for normalisation such that both penalties behave similarly for small $|x|$.

In addition to regularising noise, desired properties, such as sharp edges and smoothness in uniform regions, can also be introduced by using a specially designed penalty function [42, 43]. For example, penalties derived from anatomical CT or Magnetic Resonance (MR) images have been proposed for improving the edge delineation of objects in many studies [38, 44, 45, 46, 47, 48, 49, 50, 51, 52]. In recent years, there have been promising developments on anatomical priors [53, 54]. For instance, Joint Total Variation (JTV) [55] has been used for exploring the edge information based on the magnitude of the image gradients of both the ET and anatomical images. By encouraging strong gradients that are observed in both functional ET and anatomical images, the prior incorporates structural details during the image reconstruction process of ET. The robustness of the estimated image was further improved by considering the orientation of the gradients [56]. With the direction information, the proposed Parallel Level Sets (PLS) shows higher reliability of discovering the structure similarity between these two images. Since it has shown promising results in the literature [56, 57, 58], PLS is chosen as the representative anatomical penalty in this thesis:

$$R(\boldsymbol{f}|\boldsymbol{z}) = \sum_j \sqrt{\varepsilon^2 + \|\,[\nabla \boldsymbol{f}]_j\,\|_2^2 - \langle [\nabla \boldsymbol{f}]_j, [\boldsymbol{\xi}]_j \rangle^2}, \tag{2.17}$$

$$[\boldsymbol{\xi}]_j := \frac{[\nabla \boldsymbol{z}]_j}{\sqrt{\|\,[\nabla \boldsymbol{z}]_j\,\|_2^2 + \eta^2}},\ \varepsilon \text{ and } \eta > 0$$

where ∇ is the gradient operator, $\langle \cdot, \cdot \rangle$ is the Euclidean scaler product, $\boldsymbol{z} = [z_1, \ldots, z_J]^\top \in \mathbb{R}^J$ is the anatomical image and $\|\cdot\|_2$ denotes the ℓ^2-norm. The edge preserving property of the function is modulated by the pair of parameters

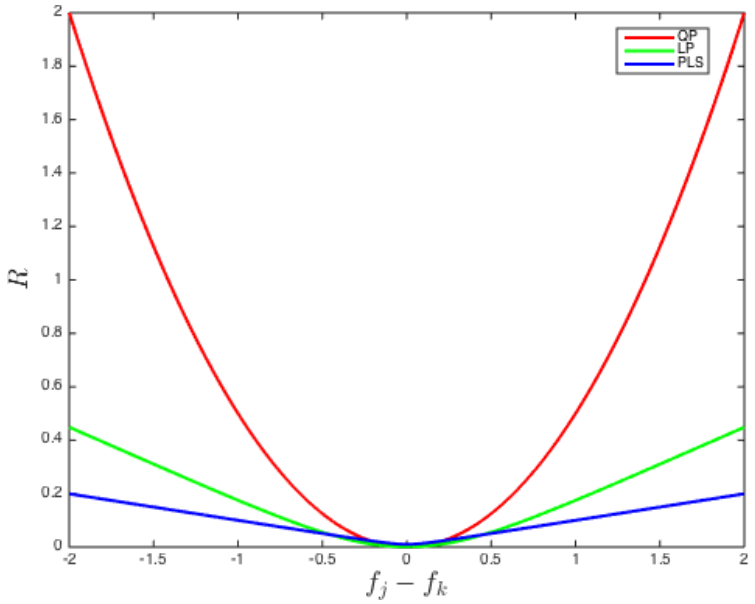

Figure 2.5: Plots of QP, LP and PLS in 1-D. The range of the difference between two adjacent voxels is $[-2,\ 2]$. The scalar ρ in LP and the set of parameters (ε, η) in PLS are 1.8 and (0.01, 0.1), respectively. The gradient of the anatomical image is fixed at 1 for PLS.

(ε, η) [58]. The utilisation of the anatomical information will be studied in Chapter 4 and 5. Note that for the penalties used in this thesis, the objective function Φ is strictly convex [58, 59]. Figure 2.5 illustrates the behaviour of each function in 1-dimensional (1-D), given a fixed range of difference between adjacent voxels.

2.3.2.3 Update function

The update function, which is also referred to as the optimisation algorithm, finds a better estimate $\hat{\boldsymbol{f}}$ from a given estimate based on the comparison result of the objective function. Many different approaches have been proposed for the optimisation. One of the most widely used methods that converges to ML solution is ML-expectation maximisation (ML-EM) [60]. It makes the next step to the solution with current estimate $\hat{\boldsymbol{f}}^{(t)}$ by:

$$\hat{f}_j^{(t+1)} = \hat{f}_j^{(t)} \left(\frac{\sum_i A_{ij} \frac{g_i}{\sum_{j'} A_{ij'} \hat{f}_{j'}^{(t)} + n_i}}{\sum_i A_{ij}} \right) \tag{2.18}$$

The equation can also be written in an additive form as follows:

$$\hat{f}_j^{(t+1)} = \hat{f}_j^{(t)} + \frac{\hat{f}_j^{(t)}}{\sum_i A_{ij}} \underbrace{\left(\sum_i \frac{A_{ij}}{\sum_{j'} A_{ij'} \hat{f}_{j'}^{(t)} + n_i} (g_i - (\sum_{j'} A_{ij'} \hat{f}_{j'}^{(t)} + n_i)) \right)}_{\nabla L} \tag{2.19}$$

Both of these expressions can be interpreted as a process that updates the current estimate with a correction term (in the parenthesis). Note that the correction term in the additive form is actually the gradient of the log-likelihood function (2.11). The computation of ML-EM is readily parallelisable with respect to projections. However, its application is limited in practice, since it converges very slow. Although it is not necessary to converge to the ML-solution due to the ill-conditioned property of the problem, the algorithm requires a large number of iterations to obtain a visually appealing image.

To improve the convergence rate, the OS (ordered-subset)-EM algorithm was proposed [61]. By using only a subset of the projection data per update ("sub-iteration"), the algorithm can be about an order of magnitude faster than ML-EM. Although it is fast to compute and also parallelisable, OS-EM normally does not converge [61]. The estimation oscillates among the solutions for the sets of incomplete data (limit-cycle problem). In particular, using a large number of subsets can introduce artefacts and hence interfere with lesion detection [62].

Instead of using OS, accelerated EM methods use a larger step at each iteration [63, 64, 65, 66, 67]. The improvement of the convergence rate is linear in the step size. However, the convergence of the accelerated methods is hard to prove. To preserve the convergence, a manually selected relaxation parameter was introduced in [68]. Although the algorithm can achieve a fast convergence rate while the parameter is optimised, there are no rules for choosing the parameter. Moreover, it can be sensitive to many factors, such as data scale and object geometry [68].

With only a minor modification in the correction term of (2.18) using the one-step-late (OSL) approach [69], both ML-EM and OS-EM can be applied to PML optimisation problems incorporating any differentiable penalty function. However,

the inconsistency between the log-likelihood and penalty terms makes the algorithm unstable and divergent for a larger penalty strength [70]. Although other modified ML-EM algorithms [71] or separable paraboloidal surrogates (SPS) [72] can directly incorporate the penalty term into a closed-form update of the image without suffering from convergence issues, the application of both strategies is limited by the need to find a convex surrogate function that lies above the original function and easier to solve. The image update scheme of SPS can be summarised as follows (see Appendix A for more information):

$$\begin{aligned} \boldsymbol{f}_{t+1} &= \boldsymbol{f}_t - \hat{\boldsymbol{H}}_t^2 \nabla\Phi(\boldsymbol{f}_t) \\ \hat{\boldsymbol{H}}_t &= \operatorname{diag}\left\{ \boldsymbol{A}^\top \boldsymbol{X}_t \boldsymbol{A}\mathbf{1} + \beta R_\varphi(\boldsymbol{f}_t)\mathbf{1} \right\}^{-\frac{1}{2}} \end{aligned} \tag{2.20}$$

where $R_\varphi(x) = \nabla R(x)/x$ provides the second order information of the penalty function R of the form in (2.14) with the potential function φ and $\boldsymbol{X}_t$ is a vector with the same length as the measured data $\boldsymbol{g}$. Note that subscripts are used to indicate the iteration number instead of the voxel location and the estimated $\hat{\boldsymbol{f}}$ has been replaced by $\boldsymbol{f}$ in this and the following sections for compactness. Given the current estimate data $\hat{\boldsymbol{g}}_t = \boldsymbol{A}\boldsymbol{f}_t + \boldsymbol{n}$, $\boldsymbol{X}_t$ can be computed by:

$$[\boldsymbol{X}_t]_i = \begin{cases} \left[2\frac{\Phi_i(0) - \Phi_i([\hat{\boldsymbol{g}}_t]_i) + \nabla\Phi_i([\hat{\boldsymbol{g}}_t]_i)[\hat{\boldsymbol{g}}_t]_i}{[\hat{\boldsymbol{g}}_t]_i^2} \right] & \text{if } [\hat{\boldsymbol{g}}_t]_i > 0 \\ \nabla^2\Phi_i(0) & \text{if } [\hat{\boldsymbol{g}}_t]_i = 0 \end{cases} \tag{2.21}$$

Another alternative is to employ the generic steepest-descent optimisation algorithm to find the local solution along the gradient of the penalised log-likelihood function by using a line search:

$$\begin{aligned} \boldsymbol{f}_{t+1} &= \boldsymbol{f}_t - \alpha^\star \nabla\Phi(\boldsymbol{f}_t), \\ \alpha^\star &= \arg\min_{\alpha \geq 0} (\Phi(\boldsymbol{f}_t) - \alpha\nabla\Phi(\boldsymbol{f}_t)),\ \boldsymbol{f} \geq 0 \end{aligned} \tag{2.22}$$

With a good line search algorithm, steepest descent can show fast initial convergence rate but often slows down while approaching the final solution as the direction

defined by the gradient can lead to a zigzag path to the solution for ill-conditioned problems (Figure 2.6).

Instead of using merely the gradient, Newton's method [73] defines a better search direction with the help of the Hessian matrix, which is calculated from the second partial derivatives of Φ. Given the current estimate $\boldsymbol{f}_t$ at iteration t, a quadratic polynomial approximation of the objective function Φ in the neighbourhood of $\boldsymbol{f}_t$ is defined as:

$$q_t(\boldsymbol{f}) = \Phi(\boldsymbol{f}_t) + \boldsymbol{d}_t^\top \nabla\Phi(\boldsymbol{f}_t) + \frac{1}{2}\boldsymbol{d}_t^\top \boldsymbol{H}_t \boldsymbol{d}_t \,. \tag{2.23}$$

where $\boldsymbol{d}_t = \boldsymbol{f} - \boldsymbol{f}_t$ is the search direction and $\boldsymbol{H}_t$ is the Hessian matrix at $\boldsymbol{f}_t$. The optimisation is therefore performed by minimising this quadratic model instead of the objective function. By taking a gradient of the model, the local minimiser is found at $\boldsymbol{d}_t = -\boldsymbol{H}_t^{-1}\nabla\Phi(\boldsymbol{f}_t)$ and the next update of Newton's method is achieved with:

$$\boldsymbol{f}_{t+1} = \boldsymbol{f}_t + \alpha^\star \boldsymbol{d}_t \,. \tag{2.24}$$

With information inherent in $\boldsymbol{H}$, the algorithm shows not only fast convergence but also consistent performance for different datasets. However, $\boldsymbol{H}$ in large scale problems is usually too large to calculate, store in memory or invert. To overcome this, quasi-Newton algorithms that use approximations for $\boldsymbol{H}^{-1}$ were developed [74]. The use of a quasi-Newton algorithm for PML reconstructions is further discussed in Chapter 3.

2.3.3 Desired properties of reconstruction algorithms

A reconstruction algorithm should provide good image quality with accurate and precise intensity values for research and clinical application. Since both visual interpretation and quantitative analysis can be affected by the spatial resolution of the reconstructed image, the desired algorithm should be capable of encouraging high resolution at edges and smoothness in uniform regions as well. To have reliable reconstructed images without compromising patient throughput, algorithms that show fast convergence rate are preferable. In addition, the performance of the algorithm

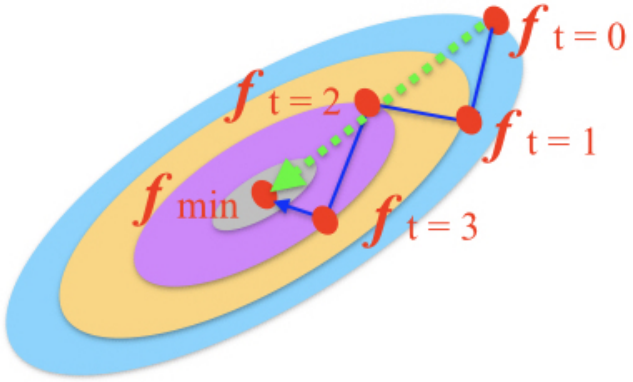

Figure 2.6: Illustration of the optimisation path for an ill-conditioned problem using the steepest descent algorithm. The contour-plot represents the objective function and the blue lines are the search directions defined by the gradient at each estimate. The shortest direction is denoted by the dashed green arrow.

should be independent of as many factors as possible to ensure predictable run-time in a clinical environment.

With a good choice for the penalty, penalised image reconstruction algorithms are able to provide both appealing images and reliable quantification results [75]. However, the image visual appearance and quantitative accuracy and precision are functions of iterations when the applied algorithm has not converged yet. As shown in Figure 2.7, sharper edges are observed in the image at higher iteration than that at lower iteration. Moreover, the quantification error can be reduced significantly when a higher iteration number is used (Figure 2.8), especially for regions with low activity level (e.g. region 3 and 4). Since each optimisation algorithm approaches the PML solution in a different way, the required number of iterations for achieving convergence varies between algorithms and applications. This makes ensuring similar visual quality and quantitative accuracy and precision for different data and reconstruction conditions even more challenging in practice.

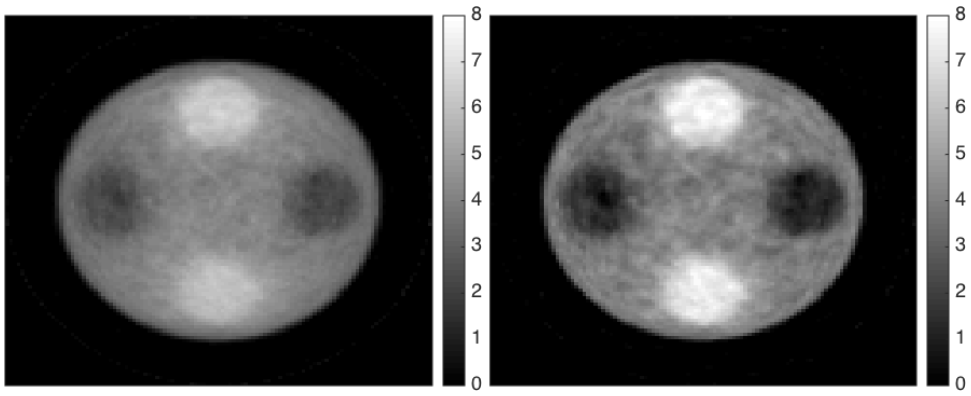

Figure 2.7: An example of how iteration number affects image appearance. The left image is terminated at the 20$^{\text{th}}$ iteration and the right one is at the 50$^{\text{th}}$. The simulation condition was the same as in Section 3.5.1 with total counts of 594 K. The applied reconstruction algorithm was SPS with the quadratic penalty and $\beta = 0.1$. A summary of the algorithm can be found in Appendix A.

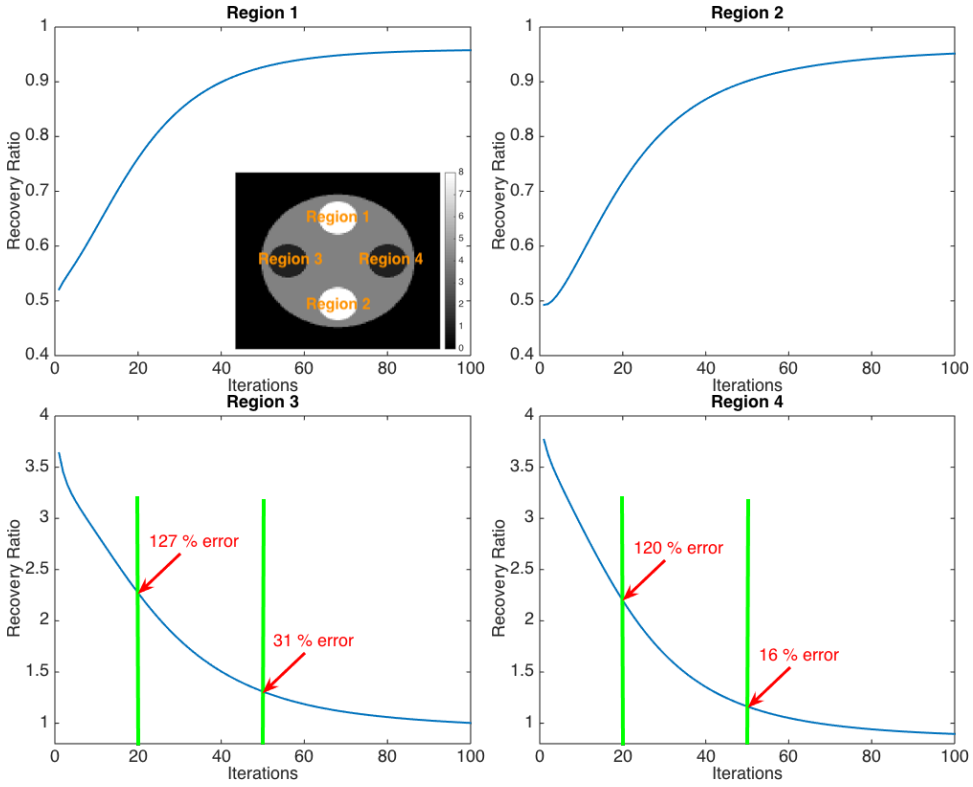

Figure 2.8: An example of how iteration number affects quantitative accuracy. Value 1 means 100% recovery. The recovery ratio was calculated by drawing a square region-of-interest (ROI) in the centre of each circular spot with a side length of 26.4 mm. See the caption in Figure 2.7 for details of the simulation and reconstruction settings.

Chapter 3

Fast quasi-Newton algorithms for PML reconstruction in ET

In this chapter a brief description of a quasi-Newton optimisation algorithm, L-BFGS-B, is given. To circumvent its potential slow convergence rate for ill-conditioned or large-scale reconstruction problems in ET, a preconditioner based on a diagonal approximation of the Hessian is introduced. Basic principles of preconditioning and the derivation of the preconditioned algorithm, L-BFGS-B-PC, are therefore addressed. Performance evaluations for both algorithms with digital and real patient data are provided, followed by discussion and conclusions.

3.1 Introduction

With the additional information contained in the Hessian matrix, Newton's method is able to define a better search direction at each iteration, leading to a faster convergence rate. However, as mentioned in Section 2.3.2.3, the Hessian matrix is usually unavailable for image reconstruction problems. Instead of calculating the true Hessian, quasi-Newton optimisation methods use approximations of the Hessian to define the search direction. A popular example of a quasi-Newton method is the L-BFGS algorithm [76, 77], which approximates the inverse of the Hessian based on the gradient information obtained in the previous few iterations. L-BFGS has been extended to allow box constraints on the variables that are to be estimated (L-BFGS-B) [74, 78]. Since the amount of memory the algorithm requires can be

controlled by the user and scales linearly with the dimensionality of the problem, it has become the most popular quasi-Newton method for optimising non-linear problems [79]. It is widely used in machine learning but not yet in PML image reconstruction. As L-BFGS-B constructs approximations of the inverse Hessian by using only the gradient information, the algorithm should be able to handle any differentiable penalty term. This enables the incorporation of many non-convex penalty functions, such as the joint entropy priors [80].

However, the application of L-BFGS-B for PML image reconstruction is limited by the dependence of its convergence rate on image and data scale [78, 81]. This chapter concentrates on improving the performance of L-BFGS-B by introducing better initialisation and additional diagonal preconditioning. Previously, Kaplan et al. used L-BFGS-B with a preconditioner for accelerating simultaneous estimation of activity and attenuation distributions in SPECT [82]. A constant value was chosen as the preconditioner to rescale the activity estimate. The algorithm showed a faster convergence rate in most cases when both the transformed activity and attenuation were in a similar scale. However, since the scale of the activity varies with application and individual dataset, the preconditioner had to be tuned accordingly by trial and error. In this work, a more general diagonal preconditioner, based on the second partial derivative of the objective function, is used. With the help of the extra information, the penalised reconstruction problem is transformed to a better-conditioned form, which is then incorporated into the L-BFGS-B optimisation process. The resulting algorithm is denoted as L-BFGS-B-PC.

3.2 L-BFGS-B

In this section, the main ideas behind L-BFGS-B are described. More detail can be found in [74] and [83].

3.2.1 Unconstrained optimisation

Substituting $\boldsymbol{H}_t$ in (2.23) with an approximation of the inverse of the Hessian matrix $\boldsymbol{B}_t$, the polynomial approximation of Φ in the neighbourhood of $\boldsymbol{f}_t$ becomes:

$$q_t(\boldsymbol{f}) = \Phi(\boldsymbol{f}_t) + \boldsymbol{d}_t^\top \nabla\Phi(\boldsymbol{f}_t) + \frac{1}{2}\boldsymbol{d}_t^\top \boldsymbol{B}_t^{-1}\boldsymbol{d}_t\,. \tag{3.1}$$

The matrix $\boldsymbol{B}_t$ is calculated by L-BFGS using limited memory. In order to compute the matrix/vector products with $\boldsymbol{B}_t$ efficiently [84], the algorithm does not store $\boldsymbol{B}_t$ directly, but represents it by a pair of lower-dimensional correction matrices, which record the change of the update and the gradient of Φ in the previous few iterations. A description of the construction of $\boldsymbol{B}_t$ is given in Appendix C.

When $\boldsymbol{B}_t$ is positive definite, q_t has a unique minimiser $\boldsymbol{f}^\star$:

$$\boldsymbol{f}^\star = \boldsymbol{f}_t - \boldsymbol{B}_t\nabla\Phi(\boldsymbol{f}_t)\,. \tag{3.2}$$

Since the polynomial approximation (3.1) is local, $\boldsymbol{f}^\star$ cannot be used as an update for the minimisation of Φ. Instead, an update $\boldsymbol{f}_{t+1}$ along the line segment $\{\boldsymbol{f}_t + \alpha\boldsymbol{d}_t^\star, \alpha \in [0,1]\}$ with $\boldsymbol{d}_t^\star = \boldsymbol{f}^\star - \boldsymbol{f}_t = -\boldsymbol{B}_t\nabla\Phi(\boldsymbol{f}_t)$ is sought:

$$\boldsymbol{f}_{t+1} = \boldsymbol{f}_t + \alpha^\star\boldsymbol{d}_t^\star\,. \tag{3.3}$$

To ensure convergence and sufficient progress, the step length $\alpha^\star$ is generally obtained using a "backtracking" algorithm, which gradually decreases α from an initial value $\alpha^{\text{init}} \leq 1$ until the Wolfe conditions (WCs) are met [85]:

$$\Phi(\boldsymbol{f}_t + \alpha\boldsymbol{d}_t^\star) \leq \Phi(\boldsymbol{f}_t) + \lambda_1\alpha\nabla\Phi(\boldsymbol{f}_t)^\top\boldsymbol{d}_t^\star \tag{3.4}$$

$$\|\nabla\Phi(\boldsymbol{f}_t + \alpha\boldsymbol{d}_t^\star)^\top\boldsymbol{d}_t^\star\|_2 \leq \lambda_2\|\nabla\Phi(\boldsymbol{f}_t)^\top\boldsymbol{d}_t^\star\|_2 \tag{3.5}$$

where $0 < \lambda_1 < \lambda_2 < 1$ and $\|\cdot\|_2$ is the ℓ^2-norm. In this study, λ_1 and λ_2 were set to 10^{-4} and 0.9 as in [73]. Since both the objective function and its gradient have to be computed for each new α (as shown in (3.4)–(3.5)), in ET extra forward

and back-projection operations are required when applying a line search. Note that when $\alpha^\star$ satisfies the WCs and the current estimated $\boldsymbol{B}_t$ is positive-definite, the new estimated L-BFGS matrix $\boldsymbol{B}_{t+1}$ is necessarily positive-definite [73].

3.2.2 Boundary constraints

L-BFGS was extended to L-BFGS-B [74, 78] to be able to handle minimisation with box constraints. The search direction is computed by solving the constrained problem corresponding to (3.1):

$$\boldsymbol{f}^\dagger = \arg\min q_t(\boldsymbol{f}) \quad \text{subject to} \quad \boldsymbol{l} \leq \boldsymbol{f} \leq \boldsymbol{u} \tag{3.6}$$

where $\boldsymbol{l}$ and $\boldsymbol{u}$ denote the lower and upper bounds of the problem, respectively. In this chapter, solving (3.6) was achieved following the method proposed in [74], which utilises the active constraints[1] defined by the generalised Cauchy point[2].

A lower bound $\boldsymbol{l} = \boldsymbol{0}$ was used to impose the non-negativity constraint of the image reconstruction problem in this study, while the upper bound was set to infinity. The line-search is performed in the direction $\boldsymbol{d}_t^\dagger = \boldsymbol{f}^\dagger - \boldsymbol{f}_t$. Similarly to the unbounded case, a backtracking algorithm is used to find a solution $\alpha^\dagger$ that satisfies the WCs. By convexity, the update is guaranteed to satisfy the boundary constraints.

For well-conditioned and small-scale problems, L-BFGS-B is expected to produce a minimiser with fast convergence rate, since $\boldsymbol{B}$ is a non-diagonal matrix that takes into account the inter-variable correlation. However, the limited memory approximations that are introduced can lead to low accuracy of $\boldsymbol{B}$ and slow convergence for ill-conditioned or large-scale problems [84].

3.3 Preconditioned L-BFGS-B

This work proposes to circumvent the potential slow convergence of L-BFGS-B via preconditioning. Preconditioning is a general strategy that transforms the problem into a new coordinate system where it is easier to solve (3.6) [86]. Given $\boldsymbol{f}$ the

[1] Variables whose value at the generalised Cauchy point is at lower or upper bound define the active constraints.

[2] The generalised Cauchy point is defined as the first local minimiser along the gradient direction within the trust region for a quadratic model.

original estimate, the transformation is described as:

$$\tilde{\boldsymbol{f}} = \boldsymbol{D}\boldsymbol{f} \tag{3.7}$$

where the preconditioner, $\boldsymbol{D}$, is used as the transformation matrix. To deal with the new estimate $\tilde{\boldsymbol{f}}$, the objective function and its derivatives should be transformed accordingly:

$$\begin{aligned} \tilde{\Phi}(\tilde{\boldsymbol{f}}) &= \Phi(\boldsymbol{f}) = \Phi(\boldsymbol{D}^{-1}\tilde{\boldsymbol{f}}) \\ \nabla\tilde{\Phi}(\tilde{\boldsymbol{f}}) &= \boldsymbol{D}^{-1}\nabla\Phi(\boldsymbol{D}^{-1}\tilde{\boldsymbol{f}}) \\ \tilde{\boldsymbol{H}}(\tilde{\boldsymbol{f}}) &= \boldsymbol{D}^{-1}\boldsymbol{H}(\boldsymbol{D}^{-1}\tilde{\boldsymbol{f}})\boldsymbol{D}^{-1} \end{aligned} \tag{3.8}$$

where $\boldsymbol{H}(\boldsymbol{D}^{-1}\tilde{\boldsymbol{f}})$ and $\tilde{\boldsymbol{H}}(\tilde{\boldsymbol{f}})$ respectively denote the Hessians of Φ and $\tilde{\Phi}$ evaluated at $\boldsymbol{D}^{-1}\tilde{\boldsymbol{f}}$ and $\tilde{\boldsymbol{f}}$. In consideration of the computational efficiency, we suggested to use a diagonal preconditioner. As the transformation has become a voxel-wise rescaling process, modifications on the box-constraints are not required. Since L-BFGS-B will have to restart the approximation process for constructing $\boldsymbol{B}_t$ every time the preconditioner is updated, it is essential to keep D fixed during a sufficient number of iterations to have enough history information. In this study, the use of a precomputed (and fixed) preconditioner D was chosen.

3.3.1 Preconditioner

As an initial investigation, the "precomputed denominator" of relaxed OS-SPS introduced in [59] was considered. This preconditioner is an approximation of the Hessian of a quadratic surrogate of the objective function Φ (see [59] for more information):

$$\boldsymbol{D}_1 = \operatorname{diag}\left\{\boldsymbol{A}^\top \operatorname{diag}\left\{\frac{1}{(\boldsymbol{g}+1)}\right\}\boldsymbol{A}\mathbf{1} + \beta\nabla^2 R(\boldsymbol{f})\mathbf{1}\right\}^{\frac{1}{2}}, \tag{3.9}$$

where $\operatorname{diag}\{\cdot\}$ is an operator that constructs a diagonal matrix from a vector, $\mathbf{1}$ is a vector of ones and $\nabla^2 R$ is the second derivative (*i.e.*, the Hessian) of the penalty

function. The latter is essentially a constant for quadratic penalties. Note that a constant value 1 has been added to the denominator to make it positive-definite. A summary of OS-SPS can be found in Appendix B.

Since the preconditioner (3.9) is calculated with the measured data in the denominator, its performance is sensitive to low counts. For adapting to noisy data, the following preconditioner, derived based on the Hessian of the objective function at the initial estimate, is therefore proposed in this thesis:

$$\boldsymbol{D}_2 = \operatorname{diag}\left\{\boldsymbol{A}^\top \operatorname{diag}\left\{\frac{\boldsymbol{g}}{(\boldsymbol{A}\boldsymbol{f}^{\text{init}}+\boldsymbol{n})^2}\right\}\boldsymbol{A}\mathbf{1} + \beta\nabla^2 R(\boldsymbol{f}^{\text{init}})\mathbf{1}\right\}^{\frac{1}{2}} \tag{3.10}$$

where $\boldsymbol{f}^{\text{init}}$ is the initial guess. Note that the performance of $\boldsymbol{D}_2$ will be affected by the initial guess $\boldsymbol{f}^{\text{init}}$. Choosing a better initial guess can therefore improve the convergence rate of L-BFGS-B-PC by starting closer to the solution and also by improving the preconditioner $\boldsymbol{D}_2$.

Both preconditioners $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ can be interpreted as diagonal approximations of the square root of the Hessian of the objective function.

As $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ are not updated between iterations, the overall computational demand of L-BFGS-B-PC is similar to that of L-BFGS-B. The performance of both preconditioners are compared in terms of convergence rate in Section 3.6.2.

3.4 Algorithm implementation

The implementation of L-BFGS-B employed in this study was originally proposed in [74]. Pseudo-code that summarises the implementation can be found in Algorithm 1. In Algorithm 1, WC refers to the backtracking algorithm to find a step length $\alpha^\star$ that satisfies the WCs (3.4) and (3.5), whereas ApproxInvHess refers to the Hessian inverse compact approximation method described in the Appendix C. Since it has been observed that a satisfactory approximation of $\boldsymbol{H}^{-1}$ can be obtained based on a few previous iterations [84], a history length $m = 5$ was maintained for constructing $\boldsymbol{B}_t$. In an attempt to take the scale of the variables into account, Byrd

et al. initialised the line search step at the first iteration as [74]:

$$\alpha_0^{\text{init}} = \min\left(\frac{1}{\|\nabla\Phi(\boldsymbol{f}^{\text{init}})\|_2}, 1\right). \tag{3.11}$$

Note that the maximum α_0^{init} is set to 1 to remain within the boundary. All other line searches in the algorithm are initialised from $\alpha_t^{\text{init}} = 1$ as they start from the boundary. Although the initialisation (3.11) is fine for certain scales of $\boldsymbol{f}$, for some problems it can lead to suboptimal step length at the first iteration if α_0^{init} is too small. The log-likelihood term and its gradient were implemented in MATLAB and imported to the implementation of L-BFGS-B of Byrd et al [74] through a MATLAB interface.

Algorithm 1: Pseudo-code for L-BFGS-B

Input: Data $\boldsymbol{g}$, Φ, $\nabla\Phi$, initial $\boldsymbol{f}^{\text{init}}$, α_0^{init}, β, λ_1, λ_2, m
Output: Estimated tracer distribution $\boldsymbol{f}$
$\boldsymbol{f}_0 \leftarrow \boldsymbol{f}^{\text{init}}$;
$\boldsymbol{z}_0 \leftarrow \nabla\Phi(\boldsymbol{f}_0)$;
$\boldsymbol{B} \leftarrow \mathbf{Id}$;
for $t = 0, \ldots, \text{MaxIter} - 1$ **do**
  Define $q\colon \boldsymbol{x} \mapsto (\boldsymbol{x} - \boldsymbol{f}_t)^\top \boldsymbol{z}_t + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{f}_t)^\top \boldsymbol{B}^{-1}(\boldsymbol{x} - \boldsymbol{f}_t)$;
  $\boldsymbol{f}^\star \leftarrow \arg\min_{\boldsymbol{x} \geq \boldsymbol{0}} q(\boldsymbol{x})$;
  $\boldsymbol{d}^\star \leftarrow \boldsymbol{f}^\star - \boldsymbol{f}_t$;
  if $t = 0$ **then**
    $\alpha^{\text{init}} \leftarrow \alpha_0^{\text{init}}$;
  else
    $\alpha^{\text{init}} \leftarrow 1$;
  end
  $\alpha^\star \leftarrow \text{WC}(\Phi, \nabla\Phi, \boldsymbol{f}_t, \boldsymbol{d}^\star, \alpha^{\text{init}}, \lambda_1, \lambda_2)$;
  $\boldsymbol{f}_{t+1} \leftarrow \boldsymbol{f}_t + \alpha^\star \boldsymbol{d}^\star$;
  $\boldsymbol{z}_{t+1} \leftarrow \nabla\Phi(\boldsymbol{f}_{t+1})$;
  $m' \leftarrow \min(t+1, m)$;
  $\boldsymbol{B} \leftarrow \text{ApproxInvHess}\,(\boldsymbol{f}_s, \boldsymbol{z}_s, s \in \{t+1-m', \ldots, t+1\})$;
end
$\boldsymbol{f} \leftarrow \boldsymbol{f}_{\text{MaxIter}}$;

For the proposed L-BFGS-B-PC, the same L-BFGS-B implementation was used, with the transformed objective and gradient functions programmed in MATLAB. Since the lack of the scale information of the variables is supplemented by the preconditioner $\boldsymbol{D}$, it is unnecessary for L-BFGS-B-PC to use the suboptimal first step length (3.11) as L-BFGS-B does. In other words, $\alpha_0^{\text{init}} = 1$ is a good choice for the preconditioned problem. After the modification, L-BFGS-B-PC is able to

update the current estimate with a reasonably optimal step length at every iteration. The use of this modification to speed up the initial line search, and therefore the initial convergence rate have been verified. More information can be found in Section 3.6.1. Algorithm 2 shows a pseudo-code summary of the L-BFGS-B-PC implementation.

Algorithm 2: Pseudo-code for L-BFGS-B-PC

Input: Data $\boldsymbol{g}$, Φ, $\nabla\Phi$, initial $\boldsymbol{f}^{\text{init}}$, β, λ_1, λ_2, m
Output: Estimated tracer distribution $\boldsymbol{f}$
$\boldsymbol{f} \leftarrow \boldsymbol{f}^{\text{init}}$;
$\boldsymbol{D} \leftarrow \boldsymbol{D}_1$ **or** $\boldsymbol{D}_2$;
$\tilde{\boldsymbol{f}} \leftarrow \boldsymbol{D}\boldsymbol{f}$;
$\alpha_0^{\text{init}} \leftarrow 1$;
Define $\tilde{\Phi}\colon \boldsymbol{x} \mapsto \Phi(\boldsymbol{D}^{-1}\boldsymbol{x})$;
Define $\nabla\tilde{\Phi}\colon \boldsymbol{x} \mapsto \boldsymbol{D}^{-1}\nabla\Phi(\boldsymbol{D}^{-1}\boldsymbol{x})$;
$\tilde{\boldsymbol{f}} \leftarrow$ L-BFGS-B$(\boldsymbol{g}, \tilde{\Phi}, \nabla\tilde{\Phi}, \tilde{\boldsymbol{f}}, \alpha_0^{\text{init}}, \beta, \lambda_1, \lambda_2, m)$;
$\boldsymbol{f} \leftarrow \boldsymbol{D}^{-1}\tilde{\boldsymbol{f}}$;

3.5 Data

A 2-D disc phantom was used for verifying the modification of the first line search in L-BFGS-B-PC and studying the performance of the algorithm using $\boldsymbol{D}_1$ or $\boldsymbol{D}_2$ as the preconditioner. The investigation of the best initial image for $\boldsymbol{D}_2$ and the performance comparison between L-BFGS-B and L-BFGS-B-PC was conducted with a more realistic 3-D XCAT phantom [87]. To demonstrate the feasibility with practical applications, sample reconstructions with three patient datasets are also presented.

3.5.1 Digital disc phantom

The phantom was a 2-D disc (radius = 98.28 mm) containing 4 circular inserts, 2 hot spots and 2 cold spots (Figure 3.1 (left)). The radius of each sphere was 26.37 mm. The activity ratios of the cold and hot regions to the background were 0.25 and 2, respectively. For each group of activity levels, different attenuation materials simulating the effects of bone and soft tissue were applied to each spot (Figure 3.1 (right)). The attenuation of the background region was set to that of

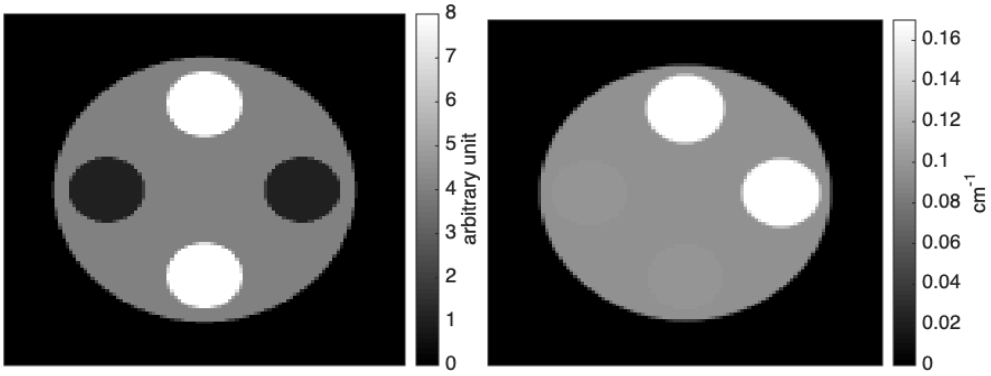

Figure 3.1: The 2-D disc phantom (left) and the corresponding attenuation map (right).

water. The projection data were generated using STIR [88] projectors to simulate the GE Discovery STE in 2D. A uniform projection with a constant intensity was added to the generated data to simulate the background events. To assess possible noise effects, two datasets with different levels of Poisson noise were generated. The number of background events was scaled accordingly such that both datasets had the same true-to-background event ratio (TBR) = 1.23. The total counts of the datasets were 29.3 K and 594 K.

3.5.2 Digital XCAT phantom

A 3-D volume from the XCAT torso phantom [87] was cropped to a $192 \times 192 \times 47$ matrix with voxel size of 3.125 mm in all directions. A slice of the phantom and the corresponding attenuation map are shown in Figure 3.2. This image was forward projected, taking attenuation into account, into 3-D sinograms corresponding to data from the GE Discovery STE in 3-D acquisition mode [89] using C functions provided by the scanner manufacturer. The background events were simulated by adding a constant value to the generated sinograms. For assessing the noise effects, three datasets (G0) with total counts S_{tot} of 52 M, 261 M and 1305 M were generated. The composition of each dataset is summarised in Table 3.1. Note that each of them had the same TBR = 0.74.

The possible effects from the background were investigated by introducing 4 more datasets, which can be divided into two groups. The first group (G1) had the same total counts as the data with S_{tot} = 261 M counts, but had 5 times lower or higher TBR, achieved by adjusting both background S_{bg} and true events S_{true}. As

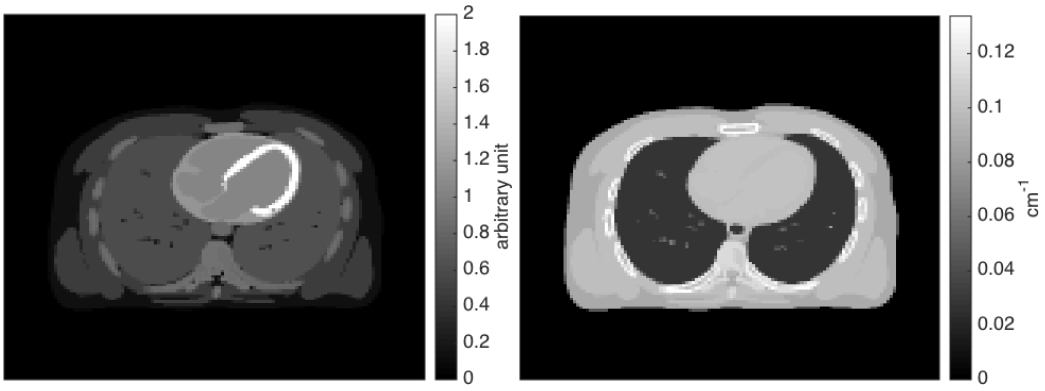

Figure 3.2: A slice of the XCAT phantom (left) and the corresponding attenuation map (right).

Table 3.1: A summary of the simulated data for evaluating the influence of the background.

		S_{true}	S_{bg}	S_{tot}
G0	TBR = 0.74	22 M	30 M	52 M
	TBR = 0.74	111.2 M	149.8 M	261 M
	TBR = 0.74	555.8 M	749.2 M	1305 M
G1	TBR = 0.15	33.7 M	227.3 M	261 M
	TBR = 3.71	205.6 M	55.4 M	261 M
G2	TBR = 0.15	111.2 M	749.2 M	860.4 M
	TBR = 3.71	111.2 M	30 M	141.2 M

S_{tot} was unchanged, there were less S_{true} in the data with higher S_{bg}. For the other group (G2), we kept S_{true} the same as that in the data with S_{tot} = 261 M counts, but changed S_{bg} by 5 times lower or higher. The total count of the data in G2 after adding the background were S_{tot} = 141.2 M and S_{tot} = 860.4 M, respectively. Note that these two groups had identical TBR for the same background level: TBR = 0.15 for the high background data and TBR = 3.71 for the low background data. Table 3.1 shows a summary of the simulated data for evaluating the influence of the background as well.

3.5.3 Patient data

Data used for this retrospective study included three patient datasets of the thorax acquired on the GE Discovery STE PET/CT scanner. For each study, a cine-CT scan (140 kVp, 60 mA, 4 s duration, 0.5 s rotation period, 0.45 s time between reconstructed images, 9 bed positions, 8 axial slices per bed position) was performed, followed by a PET scan in fully 3-D mode. The CT scan was used for the attenua-

tion correction. The acquisition was started 1 hour after the injection of 315 MBq of ^{18}F-FDG and patient consent was collected beforehand. The total counts of the PET data were S_{tot} = 181 M, 255 M and 355 M, respectively. The vendor-provided software was then used to bin the PET data into sinograms and to model the corresponding detection efficiency, attenuation, scatter and randoms.

3.6 Initial investigation

In this section, the modification of the first line search in L-BFGS-B-PC is verified, followed by dependency evaluation of the proposed algorithm on data noise level when different preconditioners are applied. Since the preconditioner $\boldsymbol{D}_2$ is affected by the initial image $\boldsymbol{f}^{\text{init}}$, a sub-study that seeks the reasonable initialisation for $\boldsymbol{D}_2$ is also presented.

Simulated data were used in this section. For the disc phantom, QP with $\beta = 0.1$ was chosen as the penalty function and the neighbourhood structure was defined as the closest 4 pixels. The applied penalty function for the XCAT phantom was QP with $\beta = 4$ and the neighbourhood was defined as the closest 6 voxels. The reconstructed images were 111×111 matrices with voxel size of 3.125×3.125 mm^2 for the disc phantom and $192 \times 192 \times 47$ matrices with voxel size of $3.646 \times 3.646 \times 3.27$ mm^3 for the XCAT phantom, respectively. Note that the voxel size used for image reconstruction was different from that for digital phantom generation to avoid artefacts induced by discretisation.

Plots of the objective function values against the total number of projection operations were used for evaluating the performance of different reconstruction or data simulation conditions. Each forward and back-projection of the full set of data was counted separately. We used the number of projection operations instead of the iteration numbers as it represents the computational demand, especially for algorithms involving a line search. Note that we did not plot results from the initial point to improve clarity in the plots. The first point of each line in this section therefore represents the objective function value after the first iteration and the corresponding number of projection operations is sum of the required projection operations for

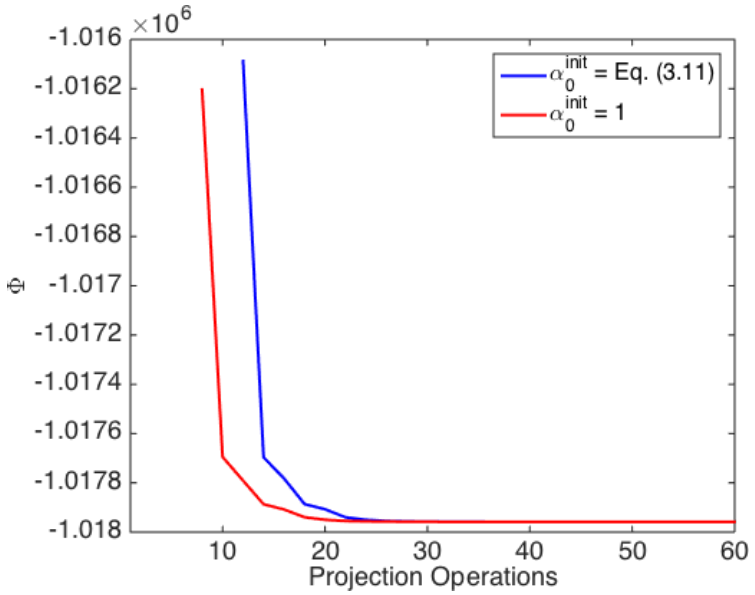

Figure 3.3: The objective function values plotted against the number of projection operations for L-BFGS-B-PC using (3.11) and 1 as the initial step length of the first line search.

constructing the initial image and for completing the first iteration of L-BFGS-B-PC.

3.6.1 Verification of the modified initial step length

Data with $S_{\text{tot}} = 594$ K for the disc phantom was used for verifying the modification of the initial step length in L-BFGS-B-PC. All reconstructions were initialised from 1 full iteration of OS-EM with 35 subsets and $\boldsymbol{D}_2$ was chosen as the preconditioner. As shown in Figure 3.3, a faster convergence rate of the objective function value is observed when the modification is applied.

3.6.2 Comparison between $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$

Simulated data with total counts of 594 K and 29.3 K for the disc phantom were reconstructed by L-BFGS-B-PC using $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ as the preconditioner. Again, the algorithm was initialised from 1 full iteration of OS-EM with 35 subsets. The initial images for both count levels can be found in Figure 3.4. As shown in Figure 3.5, $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ are similar to each other for the high count dataset. However, when the noisy dataset was used, the overall scale of the preconditioners was quite different (see the bottom row of Figure 3.5) as the ”+1” term, instead of the data itself, had

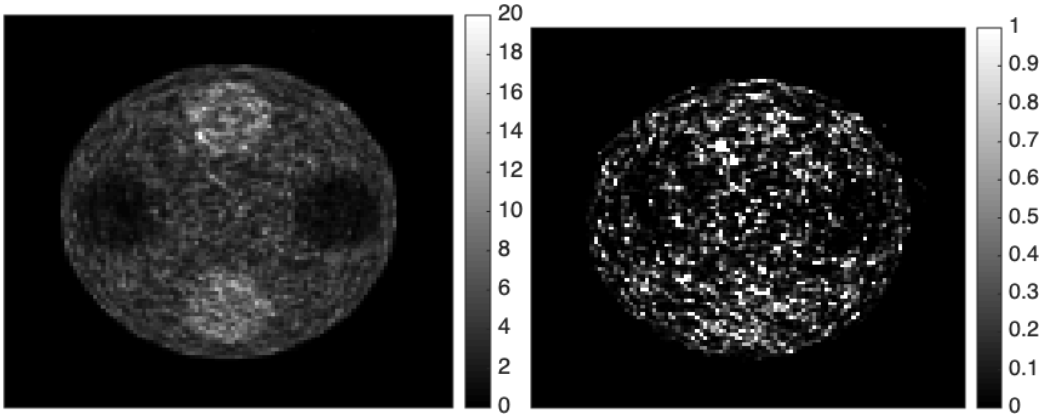

Figure 3.4: Initial images reconstructed from data with total counts of 594 K (left) and 29.3 K (right). The scale for each image was set according to the different scaling factors leading to different noise levels.

made a dominant contribution to $\boldsymbol{D}_1$.

The performance comparison between $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ in terms of convergence rate is shown in Figure 3.6. Similar convergence rates are observed for the high counts data. However, when the total counts was reduced to 29.3 K, the preconditioner $\boldsymbol{D}_2$ outperforms $\boldsymbol{D}_1$. This is because of the wrong information carried by $\boldsymbol{D}_1$ had led to a suboptimal step length in early iterations. To eliminate the influence of data noise level, $\boldsymbol{D}_1$ is excluded in future reconstructions by L-BFGS-B-PC. In the rest of this chapter, $\boldsymbol{D}$ is used to represent $\boldsymbol{D}_2$ for simplicity.

3.6.3 Best initial image

As mentioned in Section 3.3.1, initialising reconstruction algorithms with an image closer to the final solution could increase their speed, especially for the proposed L-BFGS-B-PC with the preconditioner in (3.10). To avoid increasing the overall computational cost significantly, an initial image reconstructed by OS-type algorithms is suggested. In this study, the use of OS-EM [61] is investigated as this algorithm is widely used in practice.

To simplify the problem of finding the best initial image, a two-part study was conducted. In the first part of the study, one full iteration of OS-EM with 8 different numbers of subsets (1, 2, 5, 7, 10, 14, 35 and 70) were employed to speed up the convergence rate. The number of subsets was then fixed to the optimal found in the first part and the number of full iterations was increased from one to two to

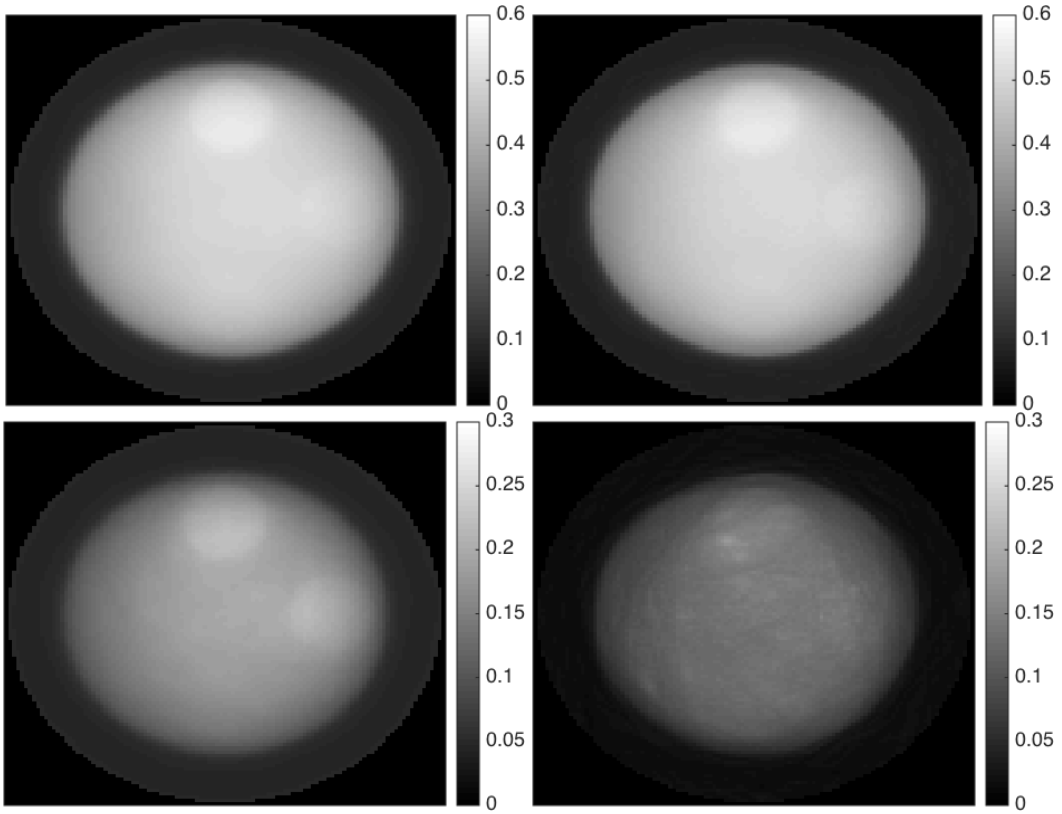

Figure 3.5: The preconditioner $\boldsymbol{D}_1$ (left column) and $\boldsymbol{D}_2$ (right column) for data with total counts of 594 K (upper row) and 29.3 K (bottom row). As the image scale was partially determined by the strength of the penalty and the experimental "+1" term for $\boldsymbol{D}_1$, it was set for achieving good visual comparison instead of reflecting the difference in data noise level. Note that images for data representing the same noise level are displayed in the same scale.

assess if the performance can be improved even further. The reconstruction was then performed with L-BFGS-B-PC initialised with the images described above. Note that the initial images were reconstructed without using a penalty function. All initial conditions were evaluated using the digital XCAT phantom dataset with S_{tot} = 261 M counts (TBR = 0.74).

As shown in Figure 3.7, the convergence rate of the objective function value was improved as the number of subsets was increased to 35. The convergence trend for 70 subsets (there were only 4 projections in one subset) was quite different from the others. Therefore, we chose 35 as the highest number of subsets and increased the full iteration number. Since the performance was not improved any further after one full iteration (Figure 3.7, bottom), all reconstructions afterwards were initialised

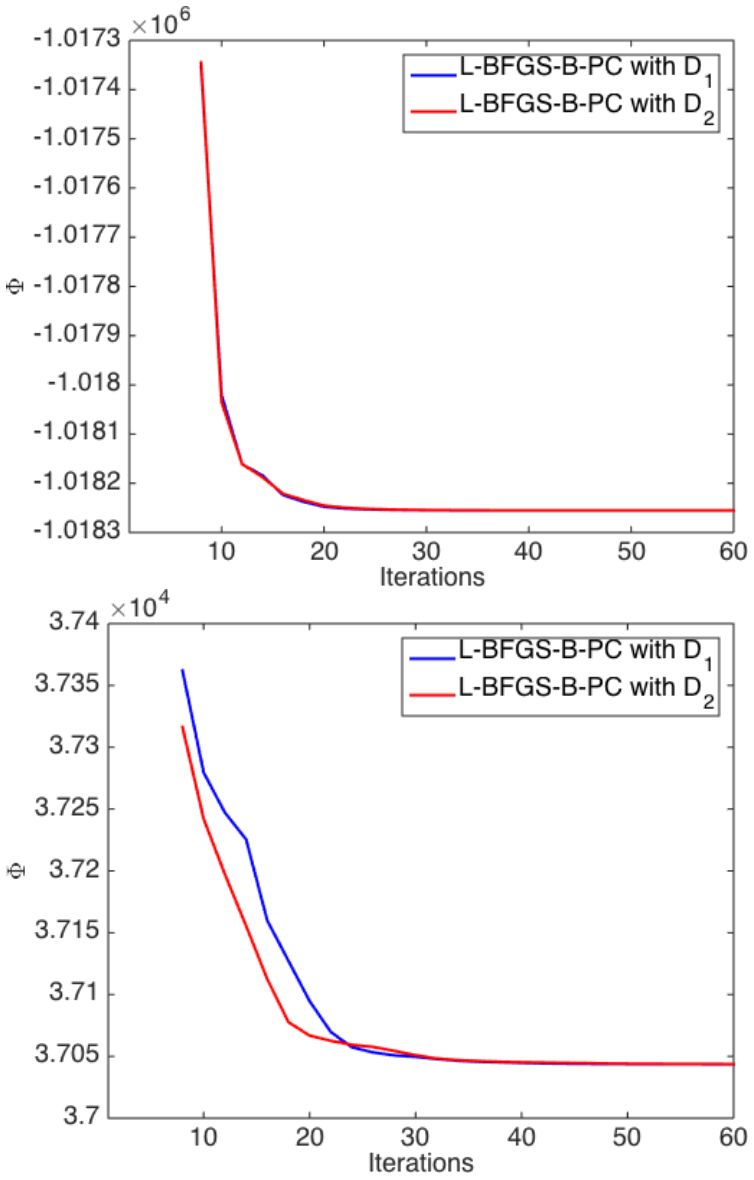

Figure 3.6: The objective function values plotted against the total projection operations for L-BFGSB-PC using $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$ as the preconditioner. Results for data with 594 K and 29.3 K counts are presented in the top and bottom plots, respectively.

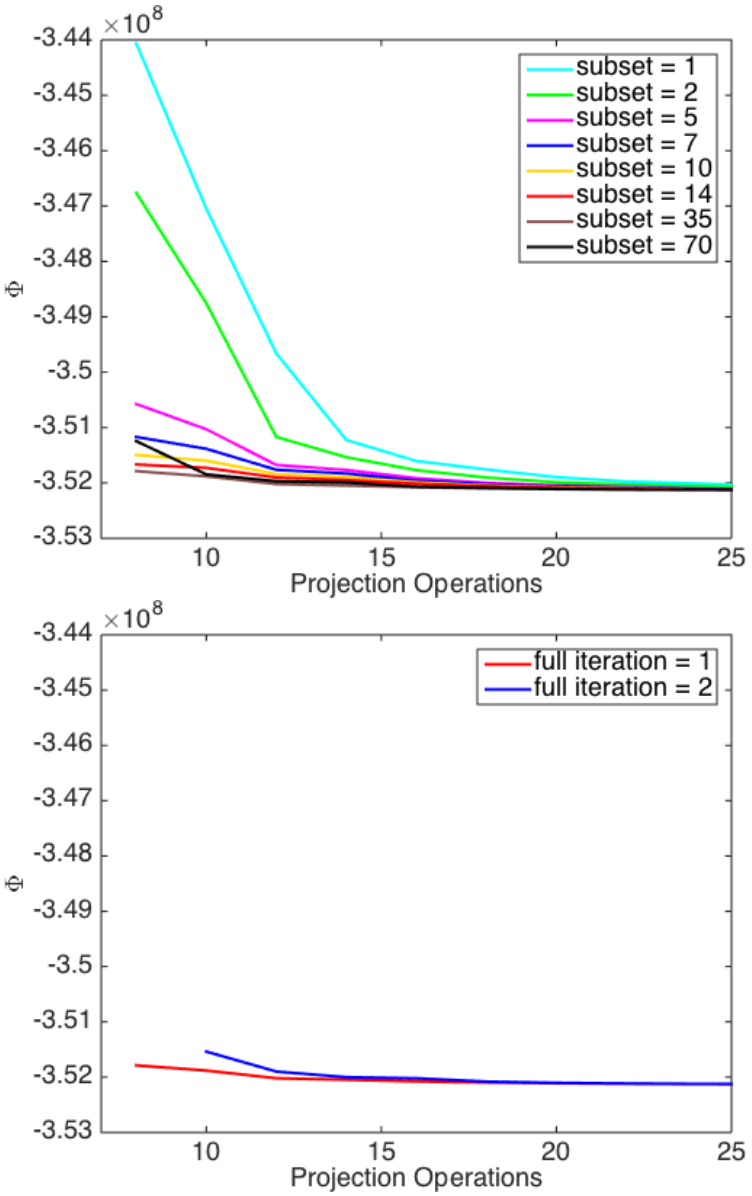

Figure 3.7: The objective function values plotted against projection operations for L-BFGS-B-PC initialised from one full iteration of OS-EM with various subsets (top) and 2 different full iterations of OS-EM with 35 subsets (bottom).

by 1 iteration of OS-EM with 35 subsets.

3.7 Performance comparison between L-BFGS-B & L-BFGS-B-PC

The performance of L-BFGS-B and L-BFGS-B-PC is compared in this section. Either QP or LP was incorporated into the reconstruction algorithm and the penalty

neighbourhood structure was defined as the closest 6 voxels. The scalar ρ in LP was fixed at 1.8, based on a visual comparison with images from QP, so as to have an apparent edge preserving effect. All reconstructions were initialised with the best initial image found in Section 3.6.3 and the reconstructed images had $192 \times 192 \times 47$ voxels with voxel size of $3.646 \times 3.646 \times 3.27$ mm^3.

The evaluation was conducted in terms of visual comparison and a convergence estimate M that measures the difference between the current reconstruction and the converged solution $\boldsymbol{f}_{\mathrm{c}}$. The metric was defined as:

$$M_t = \sqrt{\frac{1}{\mathrm{N}} \frac{\|\boldsymbol{f}_t - \boldsymbol{f}_{\mathrm{c}}\|_2^2}{(\bar{\boldsymbol{f}}_{\mathrm{c}})^2}} \tag{3.12}$$

where N is the number of voxels in the volume and $\bar{\boldsymbol{f}}_{\mathrm{c}}$ is the mean value of all voxels in $\boldsymbol{f}_{\mathrm{c}}$. Fast decrease of M indicates fast convergence rate to the solution. For the converged image $\boldsymbol{f}_{\mathrm{c}}$ in (3.12), the output of SPS [72] at high iteration number is used, since the convergence of this algorithm has been well-established. To reduce the total computational cost, the output of L-BFGS-B-PC with 40 iterations was applied as the initial image for SPS. We then ran SPS for 15000 iterations and investigated the change of visual appearance and objective function values. Since no significant change was observed after 14000 iterations for all datasets and image reconstruction conditions used in this study, the image obtained with (L-BFGS-B-PC initialised) SPS at the 15000^{th} iteration was chosen as the converged image $\boldsymbol{f}_{\mathrm{c}}$. The image-based measurement defined in (3.12) is more reliable than the objective function value as the latter might have reached an apparently saturated value while the image is still changing, especially for low value of the penalty strength. Therefore, this study concentrates on the convergence of M values for performance comparison in following sections.

To be able to compare the convergence rate among different datasets, the required number of projection operations and the corresponding iterations for achieving "practical" convergence were computed. The corresponding iteration number

was determined by:

$$t_M^\star = \min\{t : M_t \leq 0.01\} . \tag{3.13}$$

3.7.1 Visual comparison

An initial evaluation with a visual comparison of a slice of the reconstructed images from both L-BFGS-B and L-BFGS-B-PC at different iterations was used to see if the changes in the convergence rate are relevant.

Figure 3.8 shows reconstructed images of the XCAT data with S_{tot} = 261 M counts and TBR = 0.74 at the 5$^{\text{th}}$, 10$^{\text{th}}$ and 15$^{\text{th}}$ iteration for L-BFGS-B and L-BFGS-B-PC. The reconstructions were performed with QP and $\beta = 20$. Comparing images at the same iteration, we found those from L-BFGS-B-PC represent better contrast and object delineation than images reconstructed by the other algorithm.

Images for L-BFGS-B and L-BFGS-B-PC at iterations that achieve convergence of M values (3.13) are shown in Figure 3.9, with the converged image from SPS for comparison. Profiles along the central row of each image are also provided. As indicated in the figure, both algorithms are able to converge visually to the same image and profile as SPS does.

3.7.2 Convergence rate

Quantitative evaluation used plots of M values against the total number of projection operations, *i.e.*, the number of projection operations in both the initial OS-EM iterations and L-BFGS-B or L-BFGS-B-PC. Results from the same dataset and reconstruction configuration as in the previous visual comparison section were used. Comparison of the convergence rate of M values for L-BFGS-B and L-BFGS-B-PC is given in Figure 3.10. Results for SPS are also provided. As shown in the plot, both L-BFGS-B and L-BFGS-B-PC achieved convergence rates that were several times faster than SPS. By introducing a preconditioner, the proposed L-BFGS-B-PC shows the ability to converge rapidly compared to L-BFGS-B. Although only plots from one simulation condition are provided, similar behaviour was observed for all data and reconstruction configurations in this chapter.

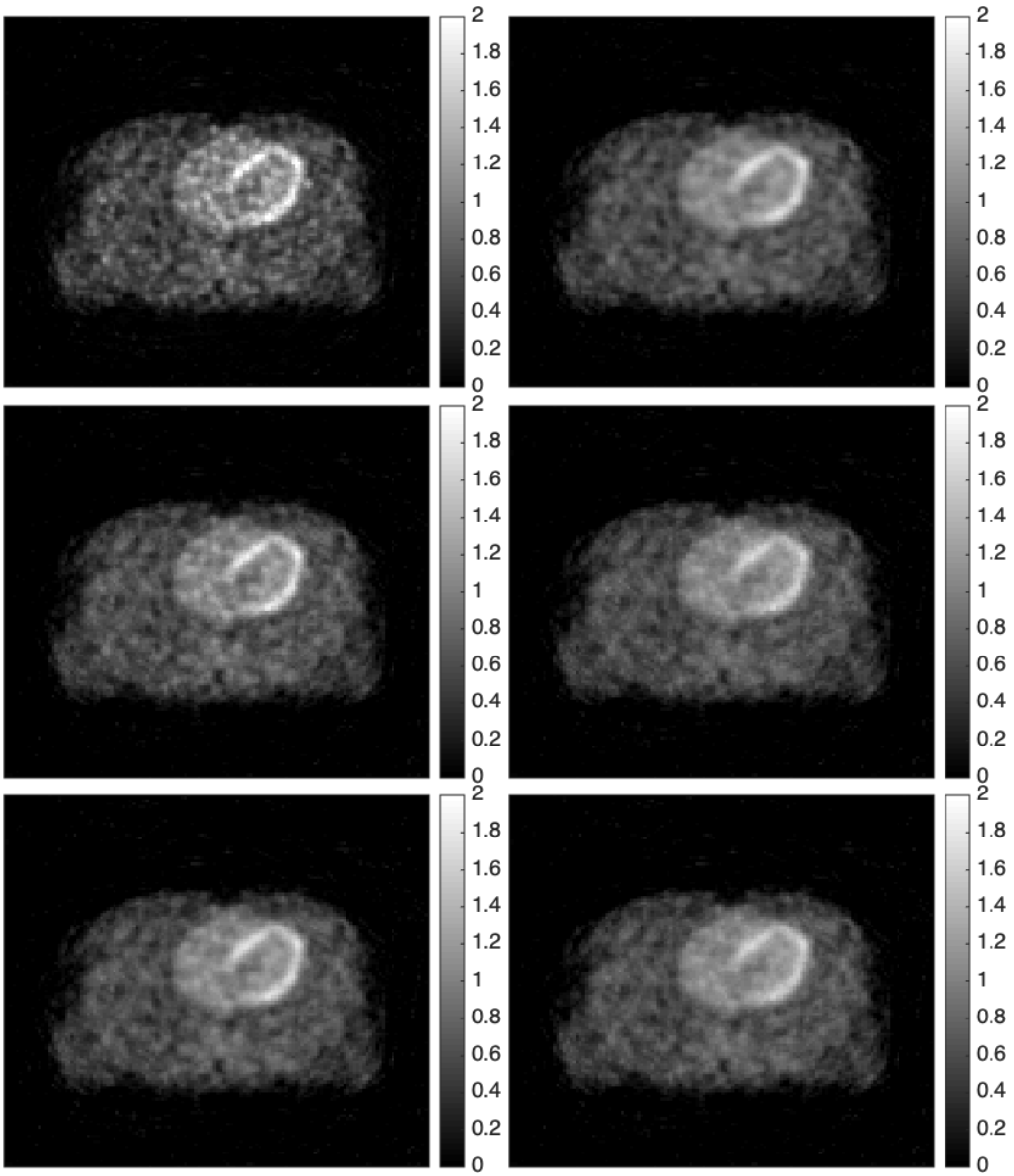

Figure 3.8: A slice of the data with 261 M counts (TBR = 0.74) reconstructed by L-BFGS-B (left column) and L-BFGS-B-PC (right column) at the 5^{th} (first row), 10^{th} (second row) and 15^{th} (third row) iteration.

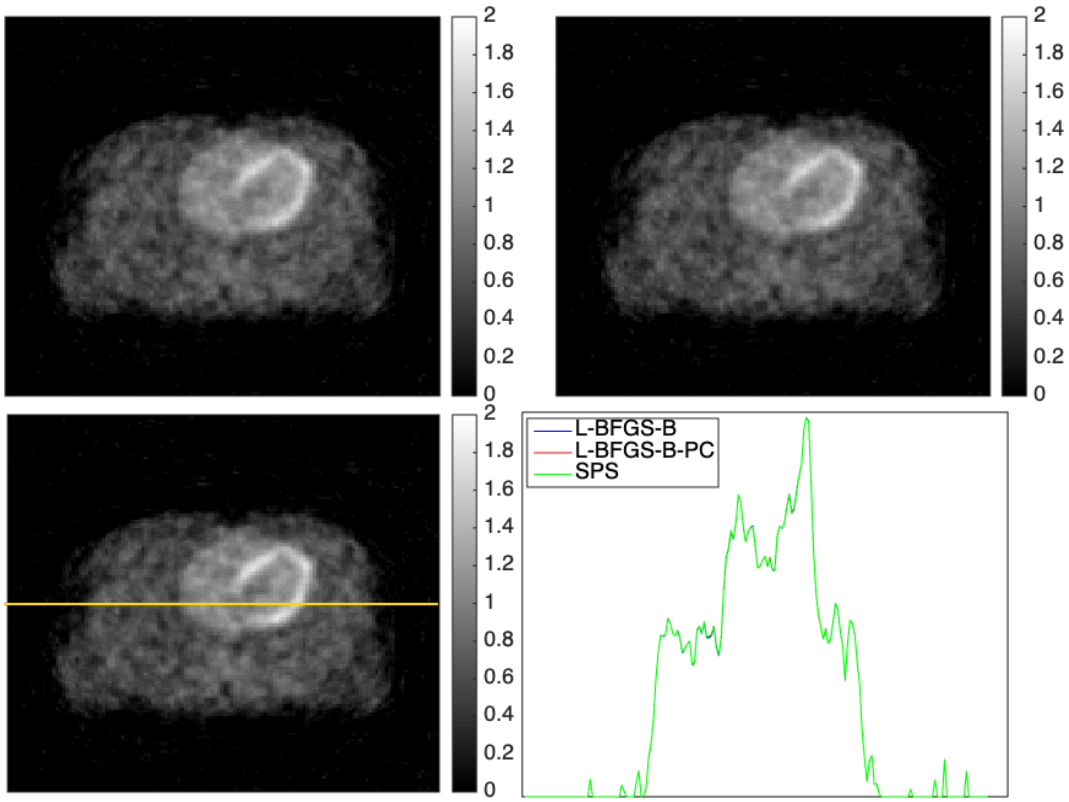

Figure 3.9: A slice of images that achieves convergence of M values for L-BFGS-B (at the 44^{th} iteration) (top-left) and L-BFGS-B-PC (at the 24^{th} iteration) (top-right). The converged image from SPS is also shown for comparison (bottom-left). Profiles along the central row of all images are also provided (bottom-right).

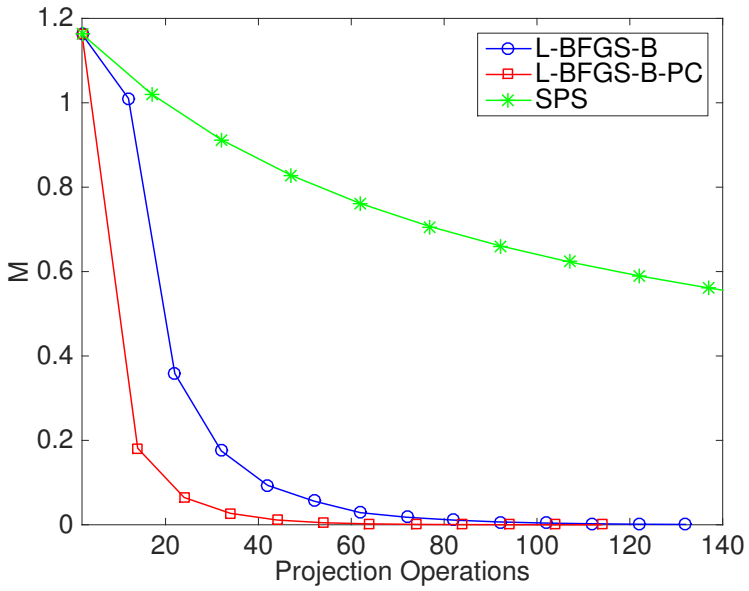

Figure 3.10: A comparison of the convergence rate of M values for SPS, L-BFGS-B and L-BFGS-B-PC with respect to the total projection operations.

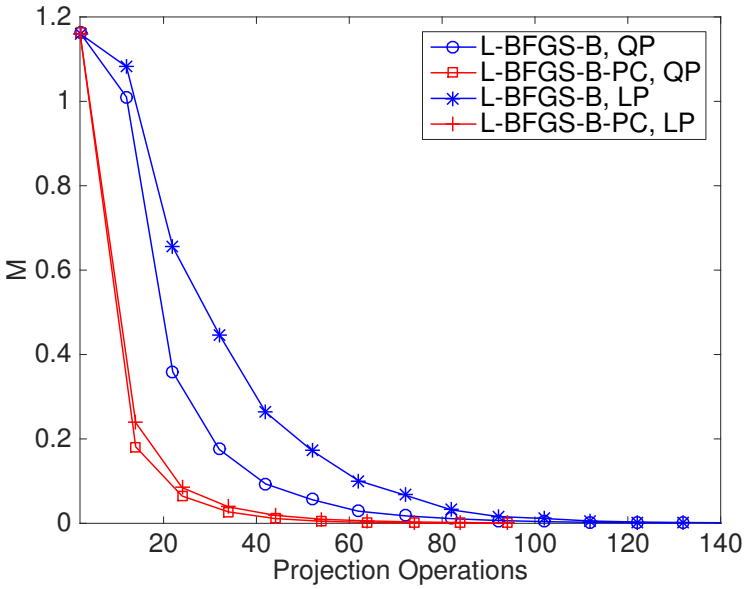

Figure 3.11: *M* values plotted against the total projection operations.

3.7.3 Convergence dependence on different factors

Assessments with respect to penalty type, penalty strength, noise level and true-to-background ratio were then performed to investigate the performance consistency of the algorithms.

3.7.3.1 Penalty type, penalty strength & noise level

Simulated data with S_{tot} = 52 M, 261 M and 1305 M for the XCAT phantom (representing high, medium and low noise level) were reconstructed with L-BFGS-B and L-BFGS-B-PC with both $\beta = 4$ and $\beta = 20$ to investigate the effect of noise levels and penalty strength. The smoothing QP or edge preserving LP penalty functions were used for evaluating the performance dependence on the penalty type.

An example comparison of the convergence rate of *M* values for L-BFGS-B and L-BFGS-B-PC is given in Figure 3.11 using the dataset of 261 M total counts and the penalty strength $\beta = 20$. In addition to achieving superior convergence rate of *M* value to L-BFGS-B, the proposed L-BFGS-B-PC also shows extremely small difference in performance when QP or LP was used. Consistent results were obtained for other simulated conditions.

For each image reconstruction and data simulation condition, Table 3.2 lists the required number of projection operations for achieving convergence according

Table 3.2: The required number of projection operations and iterations for achieving convergence of M values for different penalty types, penalty strengths and noise levels.

		L-BFGS-B	L-BFGS-B-PC
QP	$\beta = 4$	162 / 132 / 82[1] (79 / 64 / 40)[2]	94 / 64 / 44 (45 / 29 / 20)
	$\beta = 20$	122 / 92 / 72 (59 / 44 / 35)	64 / 54 / 44 (30 / 24 / 20)
LP	$\beta = 4$	552 / 182 / 132 (274 / 89 / 65)	94 / 84 / 64 (45 / 40 / 30)
	$\beta = 20$	182 / 112 / 92 (89 / 54 / 44)	74 / 54 / 84 (35 / 25 / 39)

[1] Values listed from left to right and separated by a slash are the required numbers of projection operations for problems with noise level from high to low.
[2] Values listed in parentheses are the corresponding number of iterations.

to (3.13). Values for simulation conditions with noise level from high to low are shown from left to right and separated by a slash. The corresponding number of iterations was also listed in parentheses. Except for the low noise data reconstructed using L-BFGS-B-PC with LP and $\beta = 20$ (the possible cause is discussed in Section 3.8), both algorithms generally required more operations (or iterations) to satisfy the convergence criterion as the noise level was increased or the penalty strength was decreased. Note that reconstructing with LP led to a slower convergence rate than when using QP (with the same β).

3.7.3.2 Background levels

The data for the XCAT phantom simulating different background levels in both groups of fixed S_{tot} and fixed number of S_{true} were used to study the influence of the background on the convergence rate. The results were compared with those from the data with S_{tot} = 261 M counts and TBR = 0.74 ($S_{\text{true}} = 111.2$ M and $S_{\text{bg}} = 149.8$ M). Since the dependence on penalty type and strength were included above, the data were reconstructed with only QP and $\beta = 4$ for both algorithms. The convergence rate was evaluated by plotting M values against the total projection operations (Figure. 3.12) and by listing the required number of projection operations to reach

Table 3.3: The required numbers of projection operations and iterations for achieving convergence of M values for data with different S_{tot}, S_{true} and background levels.

	L-BFGS-B	L-BFGS-B-PC
S_{tot} = 261 M, TBR = 0.74	132 (64)[1]	64 (29)
S_{tot} = 261 M, TBR = 3.71	132 (65)	64 (31)
S_{tot} = 261 M, TBR = 0.15	142 (70)	64 (31)
S_{tot} = 141.2 M, TBR = 3.71	272 (135)	84 (39)
S_{tot} = 860.4 M, TBR = 0.15	92 (46)	44 (21)

[1] The required numbers of projection operations and the corresponding number of iterations for each reconstruction are listed together with the latter in parentheses.

the convergence of M values (Table 3.3). The former shows the convergence rate in early iterations while the latter quantifies this at late iterations. For data with the same S_{tot}, convergence rate is observed to be higher in early iterations for higher TBR values (*i.e.*, the more true events) (Figure 3.12, top). However, an opposite trend is obtained when S_{tot} is increased with the background level. The presence of the background helps the convergence rate in early iterations when the same number of S_{true} are collected (Figure 3.12, bottom).

Considering the convergence rate at later iterations, we found that data with the same S_{tot} can reach the criterion (3.13) at almost the same iteration, regardless of the change in the background level. For data with a fixed number of S_{true} but increasing TBR, more iterations are needed to achieve the convergence of M values (Table 3.3). Despite the observed dependence on various factors, L-BFGS-B-PC shows a relatively consistent performance and outperforms L-BFGS-B.

3.7.4 Demonstration with patient data

Finally, the feasibility of applying both algorithms in a clinical context was assessed on three patient datasets. Again, QP was used as the penalty function with a fixed $\beta = 20$. A coronal view of one patient dataset from each algorithm at the iteration that achieves criterion (3.13) are shown in Figure 3.13 as an example. Profiles along the central slice of both images are also provided. As in the simulation study, the algorithms are able to converge to visually identical images. This was also the case for the other two patient datasets (not shown). Figure 3.14 shows the M

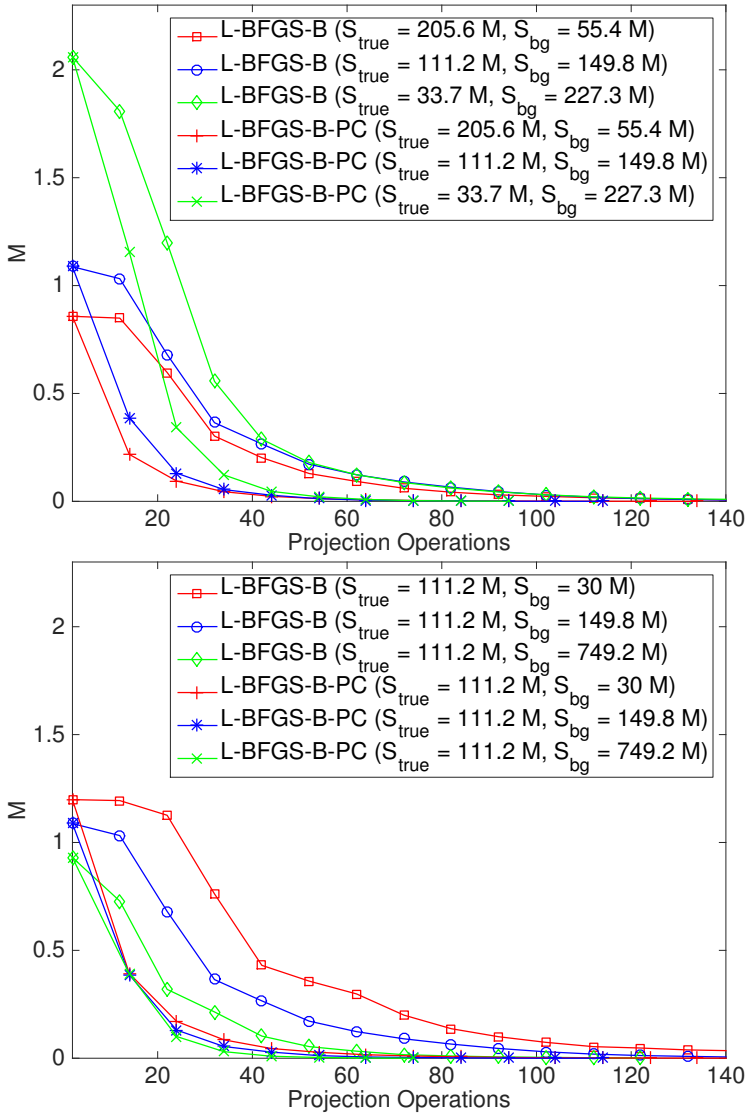

Figure 3.12: M values plotted against the total projection operations for data with a fixed number of S_{tot} (top) and data with a fixed number of S_{true} (bottom) but different background levels.

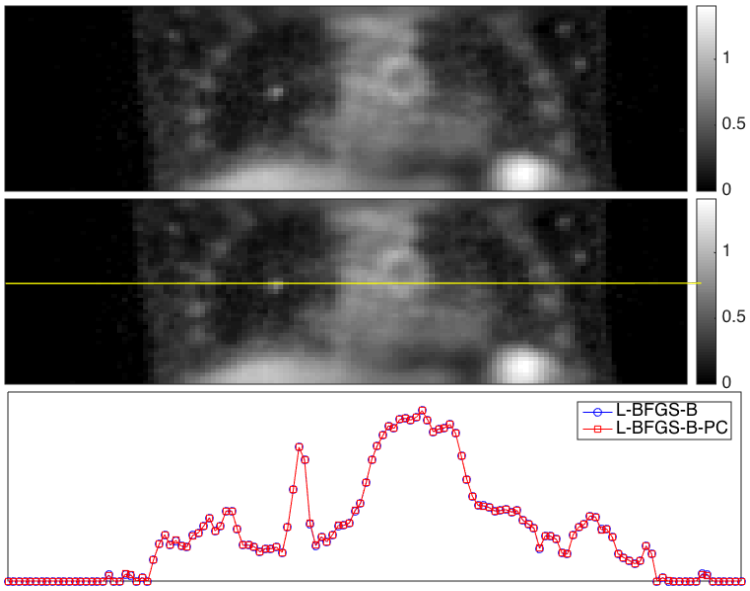

Figure 3.13: A coronal view of images for L-BFGS-B at the 260th iteration (top) and L-BFGS-B-PC at the 35th iteration (middle) from one patient data. Profiles along the central slice of both images are also provided (bottom).

Table 3.4: The required projection and iteration numbers for achieving convergence of M values for three patient datasets.

	L-BFGS-B	L-BFGS-B-PC
S_{tot} = 181 M	662 (330)	74 (36)
S_{tot} = 257 M	752 (375)	94 (45)
S_{tot} = 355 M	522 (260)	74 (35)

values plotted against the total projection operations for each algorithm. Faster initial convergence is achieved for data with higher S_{tot}, which is similar to what was observed in Table 3.2. The required projection operations for achieving the convergence of M values are listed in Table 3.4. Based on the results in Figure 3.14 and Table 3.4, L-BFGS-B-PC shows faster convergence rate than L-BFGS-B in all cases and its performance is much less sensitive to noise level.

3.8 Discussion

The feasibility of using L-BFGS-B and L-BFGS-B-PC in PML reconstruction problems in ET is demonstrated in this study. Both L-BFGS-B and L-BFGS-B-PC are able to converge to visually and numerically identical solutions as SPS does (Figure 3.9), but with different speeds. For the evaluation of the computational demand,

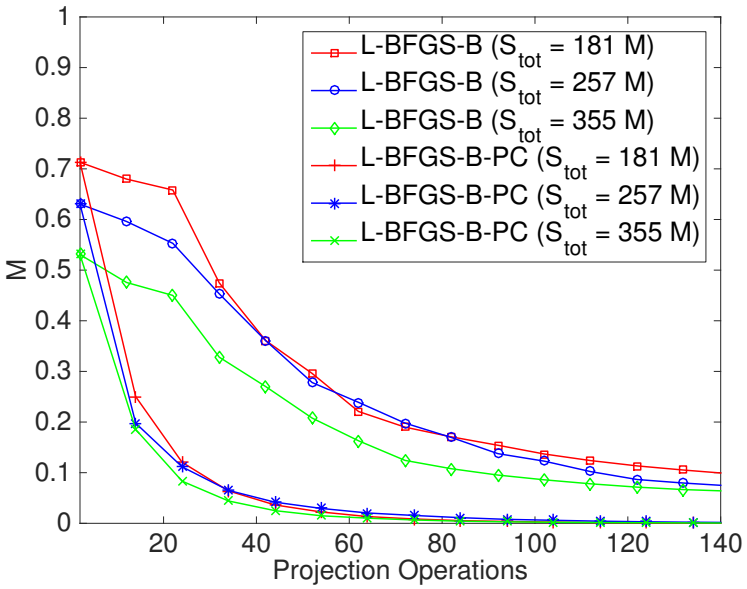

Figure 3.14: *M* values plotted against the total projection operations for all patient data.

we used the total projection operations instead of the computation time because the algorithms were not implemented using the same programming language and the projection operation is the time-limiting factor. For example, L-BFGS-B was implemented in a combination of C, Fortran and MATLAB while SPS was implemented in MATLAB but using the vendor provided projectors programmed in C. Therefore, except for the visual comparison that requires results at certain iterations (Figure 3.8, 3.9 and 3.14), plots and tables based on projection operations were utilised to compare the computational demand between different algorithms. In terms of memory demand, however, both L-BFGS-B and L-BFGS-B-PC require more memory for storing the correction matrices used to represent $\boldsymbol{B}_t$ compared to SPS. The required extra memory is approximately twice the product of the total number of voxels J and the maintained history length m (see Appendix C for more information). As a precomputed preconditioner has to be stored as well for the proposed L-BFGS-B-PC, it uses slightly more memory than L-BFGS-B. Since the bias of penalised image reconstruction algorithms is determined by the applied penalty and its weight over the whole objective function, the performance comparison in accuracy can vary with the selected datasets and reconstruction conditions. Therefore, we did not measure the absolute accuracy for simulation studies where the true

solutions were known in this work.

In Section 3.7.2, the convergence rate of L-BFGS-B and L-BFGS-B-PC was also compared with SPS [72]. We found that both L-BFGS based algorithms showed the ability to converge over 100 times faster than SPS in terms of M values. Although the convergence rate of SPS can be further improved by using ordered subsets, its performance depends on the relaxation parameter. This makes the comparison of the convergence rate difficult, especially at late iterations. Therefore, SPS was not included for further comparison in this study.

In this study, the penalty strength β was selected based on an initial investigation, where the reconstructed image was visually compared with respect to a series of increasing strengths from 1 to 20 (not shown). According to the apparentness of the smoothing or edge preserving properties of the reconstructed image, we chose $\beta = 4$ and $\beta = 20$ as the weak and strong penalty strengths, respectively. The dependence of the convergence rate of L-BFGS-B and L-BFGS-B-PC on penalty strength as well as penalty type and noise level was assessed with simulated data in Section 3.7.3.1. We observed that faster convergence rate was achieved generally with a strong penalty strength, smoothing prior and low noise level data for both algorithms (Table 3.2). In terms of convergence rate of M values, the proposed L-BFGS-B-PC outperformed L-BFGS-B for all datasets that have been evaluated. Its performance was also less sensitive to the change of penalty strength, penalty type and data noise level. In particular, L-BFGS-B-PC achieved convergence within 100 projection operations for all simulations, even for the noisy dataset (*i.e.,* the simulated data with S_{tot} = 52 M). The results suggest that the proposed algorithm can even be used in cases where the noise level is high, such as in gated or dynamic studies.

As observed in the top plot of Figure 3.12, the change in background level for data with the same S_{tot} can affect the convergence rate in early iterations. From the plot, this can be explained at least partially by the fact that the initial OS-EM image was further away from the final solution for a higher background. Other algorithms, less sensitive to TBR for initialisation, might decrease this effect. In

the case where we fixed S_{true} but changed the number of background events, both L-BFGS-B and L-BFGS-B-PC were able to achieve a faster convergence rate in early (Figure 3.12, bottom) and late iterations (Table 3.3) as TBR decreased. As the behaviour is somewhat counter-intuitive, investigations for the possible causes are required. Despite the performance dependence on those factors, at later iterations, the proposed L-BFGS-B-PC is more consistent compared to L-BFGS-B (Table 3.2 and Table 3.3).

For the patient data study, the dataset with the highest S_{tot} achieved the fastest convergence rate for both L-BFGS-B and L-BFGS-B-PC at late iterations, consistent with what has been observed from the simulation study. However, the slowest convergence rate of M values was observed for the data with medium S_{tot}. Since the performance could be affected by many factors, such as patient size and scatter fraction, a more comprehensive evaluation with more patients would be necessary.

Recall that the number of projection operations includes both the forward and back-projections in the combined OS-EM and L-BFGS, and the line search. As shown in the tables, the number of projection operations for achieving the convergence criterion is only slightly larger than twice the number of iterations for both algorithms. This means that the line search subroutine did not involve many projections and so required minimum computational burden. In other words, the initial step length satisfied the WCs (3.4) and (3.5) for almost every iteration. As mentioned in Section 3.4, both algorithms initialise the line search with a step length of 1 after the first iteration. With this step length, the algorithms make a direct approach from the current estimate to the local solution of (3.6) as described in Section 3.2.2 with $\boldsymbol{d}_t^\dagger = \boldsymbol{f}^\dagger - \boldsymbol{f}_t$. The backtracking of the embedded line search takes place only when the algorithm is about to converge. To find a smaller step length, a certain decreasing pattern predefined by the backtracking algorithm is considered. However, depending on the adopted decrease scheme, the backtracking might not be able to find the step length that minimises the objective function for the reconstruction algorithm at the current estimate. This might be the cause of the unexpected slow convergence rate observed in Table 3.2 for the last entry (*i.e.*, the required projec-

tion operations for achieving the convergence of M values for the low noise data reconstructed by L-BFGS-B-PC with LP and $\beta = 20$). Further optimisation of the line search is beyond the scope of this study.

In the implementation of L-BFGS-B and L-BFGS-B-PC, the iterative process is terminated automatically when no step length $\alpha^\star > 0$ that satisfies WCs can be found in 20 trials of backtracking. Although not shown, the resulting images from both algorithms at the termination point were compared with that from SPS at 15000^{th} iteration for all reconstructions performed in this study. The image reconstructed by L-BFGS-B or L-BFGS-B-PC was visually identical to that from SPS and the corresponding M value was even smaller than what was defined for the practical convergence (*i.e.*, 0.01). This suggests that the resulting image at the termination point of both algorithms could be a reasonable substitution for $\boldsymbol{f}_{\text{c}}$ in (3.12) when the converged image from SPS at high iteration is not available. Therefore, we have used this a stopping criterion in the next chapter.

The primary motivation of incorporating a preconditioner into L-BFGS-B is to have an initial estimate of the second derivative associated with the problem. By utilising the extra information from the start, L-BFGS-B-PC is able to solve the reconstruction problem rapidly and shows consistent performance for different data conditions and reconstruction configurations. Although the current study concentrated on L-BFGS-B, the strategy could be applied to other algorithms as well.

The preconditioner (3.9) or (3.10) was used in this study. Additional information will be provided by $\boldsymbol{B}_t$ after a few iterations and the influence of the preconditioner will become less significant. This implies that the preconditioner does not need to be a precise approximation of the square root of the Hessian. Therefore, the algorithm should be able to benefit from other fixed diagonal approximations of the square root of the Hessian. For example, by expressing ML-EM in a gradient descent form, a diagonal matrix with elements equal to a normalized version of the current estimate was obtained in [90]. This was used as motivation for using this diagonal matrix as a preconditioner to improve the convergence rate of a conjugate gradient algorithm [91]. In that paper, the preconditioner was updated at each itera-

tion. However, in order for L-BFGS-B to benefit from the previous iterations when constructing $\boldsymbol{B}_t$, the current estimate can be replaced by the initial image as for the proposed preconditioner so that the preconditioner becomes pre-computable.

Depending on the application, 2 to 10 full iterations of OS-EM with 24 to 32 subsets are often used for PET image reconstruction in contemporary clinical practice. The reconstruction process is terminated early to control the noise amplification as iterations increase at the cost of quantitative accuracy and reliability. The computational demand is equivalent to 4 to 20 projection operations and the required memory is the same as the length of the image vector J. By comparison, L-BFGS-B and the proposed L-BFGS-B-PC with current settings require at least 10 times larger memory than that for OS-EM for storing the correction matrices and the preconditioner. Based on the results shown in this chapter, the computational demand for L-BFGS-B-PC can be around 5 to 25 times higher than that for OS-EM when images at the practical convergence are desirable. The computational cost for L-BFGS-B can be as high as 200 times of that for OS-EM if achieving the practical convergence is necessary. Although the increase of memory and computational burden has implications to the workflow and economics of the procedure in clinic, both better image quality in terms of noise regularisation and more reliable quantitative results can be expected when either L-BFGS-B or L-BFGS-B-PC is applied and sufficient projection operations are allowed.

In this chapter, QP and LP were used as the penalty functions since both are convex and twice differentiable. This supports the use of L-BFGS-B, which approximates the local estimate of the second derivative of a function by differences of first derivatives. In the case where the function being minimised is differentiable but not twice differentiable at some point (e.g. the Huber functional [92]), it is likely that the L-BFGS-B algorithm will have difficulty. However, this issue is beyond the scope of this study. The use of the proposed L-BFGS-B-PC with a convex and twice differentiable anatomical prior will be explored in next chapter.

3.9 Conclusions

The performance of L-BFGS-B for penalised reconstruction problems in ET is investigated with simulated and real patient data in this study. Its convergence rate can be considerably improved by introducing a diagonally-scaled preconditioner (L-BFGS-B-PC), combined with good initialisation and the modification of the first line search. Since the proposed preconditioner can be precomputed, the overall computational demand of L-BFGS-B-PC is similar to that of L-BFGS-B. In addition to showing faster convergence rate than L-BFGS-B, the performance of L-BFGS-B-PC, in terms of the image-based metric M, is less sensitive to penalty type, penalty strength, data noise level and background level. These encouraging results indicate the potential usefulness of L-BFGS-B-PC for achieving high quantitative accuracy with acceptable reconstruction time. As L-BFGS-B-PC converges faster than other algorithms investigated here, in the case of SPS substantially faster, it will be used as the reconstruction algorithm in the rest of the thesis.

Chapter 4

A spatially-variant penalty strength with anatomical priors for PML reconstruction in ET

In this chapter, the use of the proposed L-BFGS-B-PC with an anatomical prior to achieve both fast convergence rate and good image quality is explored. To further improve its quantitative consistency, a spatially-variant penalty strength is incorporated. The potential benefits of using the method with regards to algorithm performance are studied as well. A brief description of the method is addressed in Section 4.2, followed by evaluations using digital phantoms and a real patient dataset with inserted lesions. Discussion and conclusions are provided based on the results.

4.1 Introduction

As addressed in Section 2.3.2.2, the use of an image reconstruction algorithm based on PML is one of the strategies to control noise amplification as iterations increase [1, 37, 69]. Desired properties such as sharp edges and smoothness in uniform regions can also be encouraged by appropriate choice of the penalty function [42, 43]. However, the trade-off between noise and resolution in reconstructed images can vary with the individual dataset [93], making the tuning of the weight between the likelihood term and the penalty term of the objective function difficult. In addition, the effective penalty strength at each voxel is determined by its activity level,

as well as that of the surrounding voxels [94]. This makes visual comparison and quantitative analysis for regions with different activity levels or at different locations difficult even for lesions in the same image.

These issues have been studied extensively for quadratic priors, and analytical predictions of image resolution and variance are available [94, 95, 96, 97, 98]. The dependence of lesion interpretation on location, surrounding object and therefore the individual can be largely eliminated by modifying the weights of the penalty function using analytical models [94, 96, 98, 99, 100, 101, 102]. Moreover, since the reconstruction properties have become almost object independent, one can tabulate the relationship between the overall penalty weight and the corresponding resolution by performing reconstructions using a set of Monte Carlo (MC) simulated or experimental data [94].

In contrast to the large literature on quadratic penalisation, limited studies exist for edge-preserving priors. These priors aim to encourage high resolution at edges and smoothness in uniform regions. To be able to predict local resolution (and variance), the analytical models proposed initially for quadratic penalties were further generalised for non-quadratic ones in [103]. The local resolution was characterized using the local perturbation response (LPR). Instead of trying to achieve uniform resolution, the authors of [103] proposed a similar modification of the penalty weights to obtain uniform LPR across the FOV for lesions having similar local features. Since the local contrast of the "reconstructed" lesion, as determined by LPR, is one of the factors that affects visual comparison and quantitative accuracy of the lesion [104, 105], reducing the dependence of LPR on location or surroundings could help to improve consistency and reliability in many applications. For example, it could benefit PET ^{18}F-FDG scans for treatment response evaluation [106, 107, 108] or gross tumour volume delineation [109, 110]. Since the modification is essentially a spatial normalisation method for the penalty function, it is referred to as a spatially-variant penalty strength in this study.

4.2 Spatially-variant penalisation

In [103], the use of the modified penalty function was validated with an edge-preserving Huber prior using 2-D MC simulations. However, there is increased interest in priors using anatomical information to tune local properties of the penalty. This chapter extends the theory developed in [103] to a wider class of penalty functions such that it can be applied to any anatomical priors defined in a given neighbourhood structure. To adapt the modification to 3-D PET reconstruction, a different type of spatially-variant penalisation scheme is proposed in Section 4.2.2. Further approximation for reducing the sensitivity of the scheme to data noise level is made in this study as well.

4.2.1 Generic penalty function

To extend the application to a wider range of penalty functions, this study focuses on penalty function R of a generic form:

$$R(\boldsymbol{f}) = \sum_j \phi_j(\boldsymbol{f}), \tag{4.1}$$

where $\boldsymbol{f} = [f_1, \ldots, f_J]^\top \in \mathbb{R}^J$ is the activity image and each $\phi_j(\boldsymbol{f})$ only depends on f_j and voxels in a given neighbourhood $\mathcal{N}_j$ of voxel j. Such penalty functions include standard priors of the form $\phi_j(\boldsymbol{f}) = \sum_{k\in\mathcal{N}_j} \varphi(f_j - f_k)$ as well as local anatomical priors such as the Bowsher prior [111] or the Parallel Level Sets (PLS) prior [58]. Since PLS has shown promising results in literature [58] and its convexity that facilitates numerical optimisation is well-established, it is chosen to be the representative anatomical prior in this study. Using the generic expression in (4.1), PLS can be rewritten as:

$$\begin{aligned} \phi_j(\boldsymbol{f}|\boldsymbol{z}) &= \sqrt{\varepsilon^2 + \| [\nabla \boldsymbol{f}]_j \|_2^2 - \langle [\nabla \boldsymbol{f}]_j, [\boldsymbol{\xi}]_j \rangle^2}, \\ [\boldsymbol{\xi}]_j &\coloneqq \frac{[\nabla \boldsymbol{z}]_j}{\sqrt{\| [\nabla \boldsymbol{z}]_j \|_2^2 + \eta^2}}, \ \varepsilon \text{ and } \eta > 0, \end{aligned} \tag{4.2}$$

where ∇ is the gradient operator, $\langle \cdot, \cdot \rangle$ is the Euclidean scaler product, $\boldsymbol{z} = [z_1, \ldots, z_J]^\top \in \mathbb{R}^J$ is the anatomical image and $\| \cdot \|_2$ denotes the ℓ^2-norm. Its edge preserving property is modulated by the pair of parameters (ε, η) [58].

4.2.2 Linearised local perturbation response

The theoretical description of the image resolution proposed in [103] for non-quadratically penalised image reconstruction is summarised in this section. By further exploring the property of the Fisher information matrix, a different type of the spatially-variant penalty strength suitable for 3-D PET reconstruction is derived. To reduce its sensitivity to data noise level, an approximation that avoids calculating the inverse of the measured data is also proposed.

Let $\boldsymbol{f}_{\text{true}}$ be a "true" activity image consisting of a background $\boldsymbol{f}_{\text{b}}$ and a signal of interest $\boldsymbol{f}_{\text{s}}$:

$$\boldsymbol{f}_{\text{true}} = \boldsymbol{f}_{\text{b}} + \boldsymbol{f}_{\text{s}} \,. \tag{4.3}$$

One way to quantify resolution properties around the signal $\boldsymbol{f}_{\text{s}}$ is the *local perturbation response* (LPR) [103]:

$$\bar{\hat{\boldsymbol{f}}}_{\text{s}} \triangleq E[\hat{\boldsymbol{f}}_{\text{true}}] - E[\hat{\boldsymbol{f}}_{\text{b}}], \tag{4.4}$$

where $E[\hat{\boldsymbol{f}}_{\text{true}}]$ and $E[\hat{\boldsymbol{f}}_{\text{b}}]$ are the mean reconstructed images for the noisy measurements with and without the presence of the signal. For reasonably high signal-to-noise ratio (SNR), the mean perturbation $\bar{\hat{\boldsymbol{f}}}_{\text{s}}$ defined in (4.4) can be approximated by the linearised LPR (LLPR) [103]:

$$\breve{\boldsymbol{f}}_{\text{s}} \triangleq \breve{\boldsymbol{f}}_{\text{true}} - \breve{\boldsymbol{f}}_{\text{b}}, \tag{4.5}$$

where

$$\breve{\boldsymbol{f}}_{\text{true}} \triangleq \hat{\boldsymbol{f}}(\bar{\boldsymbol{g}}(\boldsymbol{f}_{\text{true}})) \text{ and } \breve{\boldsymbol{f}}_{\text{b}} \triangleq \hat{\boldsymbol{f}}(\bar{\boldsymbol{g}}(\boldsymbol{f}_{\text{b}})) \,. \tag{4.6}$$

Given $\bar{\boldsymbol{g}} = [g_1, \ldots, g_I]^\top \in \mathbb{R}^I$ the noiseless measurement, $\breve{\boldsymbol{f}}_{\text{true}}$ and $\breve{\boldsymbol{f}}_{\text{b}}$ are the noiseless estimates when the signal is present and absent, respectively. This approximation is based on the assumption that the applied image reconstruction algorithm is

approximately linear for high SNR. With suitable Taylor expansions, the relationship between the LLPR $\breve{\boldsymbol{f}}_{\mathrm{s}}$ and the true perturbation $\boldsymbol{f}_{\mathrm{s}}$ can be described as:

$$\left[\boldsymbol{F}(\breve{\boldsymbol{f}}_{\mathrm{b}})+\beta\boldsymbol{\Omega}(\breve{\boldsymbol{f}}_{\mathrm{b}};\breve{\boldsymbol{f}}_{\mathrm{s}})\right]\breve{\boldsymbol{f}}_{\mathrm{s}}\approx\boldsymbol{F}(\breve{\boldsymbol{f}}_{\mathrm{b}})\boldsymbol{f}_{\mathrm{s}}\,, \tag{4.7}$$

where $\boldsymbol{F}(\boldsymbol{f})\in\mathbb{R}^{J\times J}$ is the Fisher information matrix for estimating $\boldsymbol{f}$

$$\boldsymbol{F}(\boldsymbol{f})\triangleq\boldsymbol{A}^{\top}\mathrm{diag}\left\{\frac{1}{\bar{\boldsymbol{g}}(\boldsymbol{f})}\right\}\boldsymbol{A} \tag{4.8}$$

with $\mathrm{diag}\{\cdot\}$ an operator that constructs a diagonal matrix from a vector and $\boldsymbol{A}\in\mathbb{R}^{I\times J}$ the system matrix. In [103], $\boldsymbol{\Omega}(\breve{\boldsymbol{f}}_{\mathrm{b}};\breve{\boldsymbol{f}}_{\mathrm{s}})\in\mathbb{R}^{J\times J}$ is defined as:

$$\boldsymbol{\Omega}(\breve{\boldsymbol{f}}_{\mathrm{b}};\breve{\boldsymbol{f}}_{\mathrm{s}})\triangleq\int_0^1\nabla^2 R(\breve{\boldsymbol{f}}_{\mathrm{b}}+\tau\breve{\boldsymbol{f}}_{\mathrm{s}})\,\mathrm{d}\tau \tag{4.9}$$

such that

$$\boldsymbol{\Omega}(\breve{\boldsymbol{f}}_{\mathrm{b}};\breve{\boldsymbol{f}}_{\mathrm{s}})\breve{\boldsymbol{f}}_{\mathrm{s}}=\nabla R(\breve{\boldsymbol{f}}_{\mathrm{b}}+\breve{\boldsymbol{f}}_{\mathrm{s}})-\nabla R(\breve{\boldsymbol{f}}_{\mathrm{b}})\,. \tag{4.10}$$

The notation ∇^2 represents the Hessian operator.

Although $\boldsymbol{F}(\boldsymbol{f})$ is a non-diagonal matrix, it is concentrated about its diagonal. The image of each row (or column) vector of $\boldsymbol{F}(\boldsymbol{f})$ is assumed to have a similar shape (*i.e.* $1/r$ and $1/r^2$ blurring kernel for 2D and 3D PET acquisition, respectively) and its peak centre value is approximatively proportional to the sum of the image. The value depends on the activity distribution and the spatial variations in sensitivity for a shift-variant system. To investigate the spatial variance, it is useful to define an alternative matrix that separates $\boldsymbol{F}(\boldsymbol{f})$ into a data dependent part $\boldsymbol{\kappa}(\boldsymbol{f})$ and a data independent part $\boldsymbol{F}_0(\boldsymbol{f})$:

$$\boldsymbol{F}_0(\boldsymbol{f})\triangleq\mathrm{diag}\{\boldsymbol{\kappa}(\boldsymbol{f})\}^{-1}\boldsymbol{F}(\boldsymbol{f})\mathrm{diag}\{\boldsymbol{\kappa}(\boldsymbol{f})\}^{-1}\,, \tag{4.11}$$

where $\boldsymbol{\kappa}(\boldsymbol{f})$ is chosen such that $\boldsymbol{F}_0(\boldsymbol{f})$ is an approximately shift-invariant matrix, desirable for 3D PET reconstruction. For example, the following $\boldsymbol{\kappa}(\boldsymbol{f})$[1] was used

[1] In the notation of Qi & Leahy [98], this is κ/ν.

in [98] in the case where detector blurring is not modelled:

$$\kappa_j(\boldsymbol{f}) \triangleq \sqrt{\sum_i A_{i,j}^2 / \bar{g}_i(\boldsymbol{f})}, \quad \forall j = 1, \ldots, J. \tag{4.12}$$

This choice leads to an approximately shift-invariant $\boldsymbol{F}_0(\boldsymbol{x})$ with diagonal elements being 1.

As $\kappa_j(\boldsymbol{f})$ varies slowly with j, the following simplification of (4.11) is valid around voxel j:

$$\boldsymbol{F}_0(\boldsymbol{f})\boldsymbol{e}_j \approx \kappa_j^{-2}(\boldsymbol{f})\boldsymbol{F}(\boldsymbol{f})\boldsymbol{e}_j \tag{4.13}$$

where $\boldsymbol{e}_j$ is the j-th unit vector of length J. With (4.13) and assuming that $\breve{\boldsymbol{f}}_{\mathrm{s}}$ is concentrated on voxel l, (4.7) can be further approximated by:

$$\left[\boldsymbol{F}_0(\breve{\boldsymbol{f}}_{\mathrm{b}}) + \frac{\beta}{\kappa_l^2(\breve{\boldsymbol{f}}_{\mathrm{b}})}\boldsymbol{\Omega}(\breve{\boldsymbol{f}}_{\mathrm{b}}; \breve{\boldsymbol{f}}_{\mathrm{s}})\right]\breve{\boldsymbol{f}}_{\mathrm{s}} \approx \boldsymbol{F}_0(\breve{\boldsymbol{f}}_{\mathrm{b}})\boldsymbol{f}_{\mathrm{s}}. \tag{4.14}$$

Inspired by the "precomputed denominator" in [59], the property of $\boldsymbol{F}(\boldsymbol{f})$ is further explored and a different type of $\boldsymbol{\kappa}(\boldsymbol{f})$ using the square root of the row-sums of $\boldsymbol{F}(\boldsymbol{f})$ is chosen. In matrix notation:

$$\boldsymbol{\kappa}(\boldsymbol{f}) \triangleq \sqrt{\boldsymbol{F}(\boldsymbol{f})\,\mathbf{1}} = \sqrt{\boldsymbol{A}^\top \mathrm{diag}\left\{\frac{1}{\bar{\boldsymbol{g}}(\boldsymbol{f})}\right\}\boldsymbol{A}, \mathbf{1}}, \tag{4.15}$$

where $\mathbf{1}$ is a vector of ones. Note that this type of $\boldsymbol{\kappa}(\boldsymbol{f})$ is computationally efficient as it requires only forward and back-projection operations. Substituting (4.15) into (4.13), the row-sums of $\boldsymbol{F}_0(\boldsymbol{f})$ will be approximately equal to 1:

$$\boldsymbol{F}_0(\boldsymbol{f})\,\mathbf{1} \approx \mathrm{diag}\{\boldsymbol{\kappa}(\boldsymbol{f})\}^{-2}\boldsymbol{F}(\boldsymbol{f})\,\mathbf{1} \approx \mathbf{1}. \tag{4.16}$$

In other words, this type of $\boldsymbol{\kappa}(\boldsymbol{f})$ also leads to a shift-invariant $\boldsymbol{F}_0(\boldsymbol{f})$ for 3-D PET reconstruction as the image of each row (or column) vector of $\boldsymbol{F}_0(\boldsymbol{f})$ would have a similar shape and peak centre value.

Since $\boldsymbol{\kappa}(\boldsymbol{f})$ in (4.15) is calculated with the mean or noiseless measurement

$\bar{\boldsymbol{g}}(\boldsymbol{f})$ which is unavailable in practice, a plug-in method that substitutes $\bar{\boldsymbol{g}}(\boldsymbol{f})$ by the noisy measurement $\boldsymbol{g}$ was introduced in [103]:

$$\tilde{\boldsymbol{\kappa}} \triangleq \sqrt{\boldsymbol{A}^\top \mathrm{diag}\left\{\frac{1}{\boldsymbol{g}}\right\}\boldsymbol{A}\mathbf{1}}\,, \tag{4.17}$$

However, as it requires a backprojection of the inverse of the measured data, its performance is sensitive to the noise level. The following approximation is therefore proposed in this study:

$$\hat{\boldsymbol{\kappa}} \triangleq \sqrt{\boldsymbol{A}^\top \mathrm{diag}\left\{\frac{\boldsymbol{g}}{\hat{\boldsymbol{g}}^2(\boldsymbol{f}^{\text{init}})}\right\}\boldsymbol{A}\mathbf{1}}\,, \tag{4.18}$$
$$\hat{\boldsymbol{g}}(\boldsymbol{f}^{\text{init}}) = \boldsymbol{A}\boldsymbol{f}^{\text{init}} + \boldsymbol{n}$$

where $\boldsymbol{f}^{\text{init}}$ is the initial image and $\boldsymbol{n} = [n_1, \ldots, n_I]^\top \in \mathbb{R}^I$ is the expected background events, including randoms and scatter. As $\boldsymbol{A}$ and $\boldsymbol{A}^\top$ are smoothing operators that reduce the noise in $\boldsymbol{f}^{\text{init}}$ and $\boldsymbol{g}$, both $\tilde{\boldsymbol{\kappa}}$ and $\hat{\boldsymbol{\kappa}}$ are reasonable approximations of $\boldsymbol{\kappa}(\boldsymbol{f})$ defined in (4.15) when the initial image is close to the solution and the noise level of data $\boldsymbol{g}$ is reasonably low.

4.2.3 Modified prior for data independent LLPR

The dependence of LLPR on data in PML reconstructions is due to the presence of $\kappa_l^2(\breve{\boldsymbol{f}}_{\text{b}})$ in (4.14). Several approaches have been proposed to modify R to eliminate the data-dependency for priors of the form $\phi_j(\boldsymbol{f}) = \sum_{k\in\mathcal{N}_j} \varphi(f_j - f_k)$ [94, 103]. This study proposes the following modification for priors of the more general form (4.1):

$$\tilde{R}(\boldsymbol{f}) = \sum_j \hat{\kappa}_j^2 \phi_j(\boldsymbol{f})\,. \tag{4.19}$$

Given $\mathscr{O}_l$ a neighbourhood of voxel l containing the non-zero indices of the corresponding LLPR and $\tilde{f}_{\text{s}}$ the LLPR using $\tilde{R}$, (4.10) can be rewritten as:

$$\tilde{\boldsymbol{\Omega}}(\breve{\boldsymbol{f}}_{\text{b}}; \tilde{\boldsymbol{f}}_{\text{s}})\tilde{\boldsymbol{f}}_{\text{s}} = \sum_{j\in\mathscr{O}_l} \hat{\kappa}_j^2 \boldsymbol{c}_j(\breve{\boldsymbol{f}}_{\text{b}}; \tilde{\boldsymbol{f}}_{\text{s}}) \tag{4.20}$$

where $\tilde{\boldsymbol{\Omega}}$ is defined as in (4.9) by replacing R with $\tilde{R}$ and

$$\boldsymbol{c}_j(\breve{\boldsymbol{f}}_{\mathrm{b}};\tilde{\boldsymbol{f}}_{\mathrm{s}}) = \nabla\phi_j(\breve{\boldsymbol{f}}_{\mathrm{b}}+\tilde{\boldsymbol{f}}_{\mathrm{s}}) - \nabla\phi_j(\breve{\boldsymbol{f}}_{\mathrm{b}})\,.$$

Since $\boldsymbol{c}_j(\breve{\boldsymbol{f}}_{\mathrm{b}};\tilde{\boldsymbol{f}}_{\mathrm{s}})$ has non-zero entries $[\boldsymbol{c}_j(\breve{\boldsymbol{f}}_{\mathrm{b}};\tilde{\boldsymbol{f}}_{\mathrm{s}})]_m$ only for m in the neighbourhood of l and $\hat{\kappa}_l$ varies slowly with location, (4.20) is further approximated by:

$$\begin{aligned}\tilde{\boldsymbol{\Omega}}(\breve{\boldsymbol{f}}_{\mathrm{b}};\tilde{\boldsymbol{f}}_{\mathrm{s}})\tilde{\boldsymbol{f}}_{\mathrm{s}} &\approx \hat{\kappa}_l^2 \sum_{j\in\mathscr{O}_l} \boldsymbol{c}_j(\breve{\boldsymbol{f}}_{\mathrm{b}};\tilde{\boldsymbol{f}}_{\mathrm{s}}) \\ &\approx \hat{\kappa}_l^2 \boldsymbol{\Omega}(\breve{\boldsymbol{f}}_{\mathrm{b}};\tilde{\boldsymbol{f}}_{\mathrm{s}})\tilde{\boldsymbol{f}}_{\mathrm{s}}\,. \end{aligned} \tag{4.21}$$

With $\kappa_l(\breve{\boldsymbol{f}}_{\mathrm{b}}) \approx \hat{\kappa}_l$ around l, substituting (4.21) into (4.14) gives:

$$\left[\boldsymbol{F}_0(\breve{\boldsymbol{f}}_{\mathrm{b}}) + \beta\tilde{\boldsymbol{\Omega}}(\breve{\boldsymbol{f}}_{\mathrm{b}};\tilde{\boldsymbol{f}}_{\mathrm{s}})\right]\tilde{\boldsymbol{f}}_{\mathrm{s}} \approx \boldsymbol{F}_0(\breve{\boldsymbol{f}}_{\mathrm{b}})\boldsymbol{f}_{\mathrm{s}}\,, \tag{4.22}$$

This result shows that the modified penalty $\tilde{R}$ is able to eliminate the LLPR dependence on data. Note that it still depends on the local properties introduced by the penalty.

4.3 Algorithm implementation

The modified penalty function (4.19) with the spatially-variant penalty strength (4.18) was incorporated into a fast convergent reconstruction algorithm, L-BFGS-B-PC, which was proposed in Section 3.3 [44, 112, 113]. The algorithm performed L-BFGS-B [74, 78] in a transformed coordinate system to circumvent its potential slow convergence and sensitivity to global scale factors. As addressed in Section 3.3.1, the transformation matrix (or preconditioner) $\boldsymbol{D}$ can be chosen to be the square root of a diagonal approximation of the Hessian of the objective function Φ, which consists of the Hessian of the likelihood function (2.11) and the Hessian of the penalty function (4.1). Since (4.18) is also related to a diagonal approximation of the Hessian of (2.11), it has been used to construct $\boldsymbol{D}$ to reduce the total computational burden. As it is not necessary that $\boldsymbol{D}$ is a precise approximation of the Hessian of Φ [113] (see Section 3.8 for more information), the Hessian of the

penalty function was further replaced by a small constant ζ for simplicity and to avoid division by zero problem. The transformation can therefore be described as:

$$\boldsymbol{f}' = \boldsymbol{D}\boldsymbol{f}, \quad \boldsymbol{D} = \mathrm{diag}\left\{ \hat{\boldsymbol{\kappa}}^2 + \zeta \right\}^{\frac{1}{2}}, \tag{4.23}$$

where $\zeta = 10^{-4}$ in this study.

An update at each iteration is then found along a search direction $\boldsymbol{d}$ with the transformed image $\boldsymbol{f}'$:

$$\boldsymbol{f}'_{t+1} = \boldsymbol{f}'_t + \alpha^\star \boldsymbol{d}_t, \tag{4.24}$$

where $\alpha^\star$ is the step length and $\boldsymbol{d}_t = -\boldsymbol{B}_t \nabla\Phi'(\boldsymbol{f}'_t)$ with $\boldsymbol{B}_t$, an approximation of the inverse of the Hessian of Φ at $\boldsymbol{f}'_t$, constructed by L-BFGS-B. Here $\nabla\Phi'(\boldsymbol{f}'_t)$ is the transformed gradient for Φ at $\boldsymbol{f}'_t$. In this section, the subscript t indicates the iteration number for compactness. To ensure convergence and sufficient progress, $\alpha^\star$ is generally obtained by performing a backtracking algorithm in which a series of gradually decreasing α from an initial value $\alpha^{\mathrm{init}} \leq 1$ are tested until the Wolfe conditions are met:

$$\Phi'(\boldsymbol{f}'_t + \alpha\boldsymbol{d}_t) \leq \Phi'(\boldsymbol{f}'_t) + \lambda_1 \alpha \nabla\Phi'(\boldsymbol{f}'_t)^\top \boldsymbol{d}_t \tag{4.25}$$

$$\| \nabla\Phi'(\boldsymbol{f}'_t + \alpha\boldsymbol{d}_t)^\top \boldsymbol{d}_t \|_2 \leq \lambda_2 \| \nabla\Phi'(\boldsymbol{f}'_t)^\top \boldsymbol{d}_t \|_2, \tag{4.26}$$

where $\Phi'(\boldsymbol{f}'_t) = \Phi(\boldsymbol{f}_t)$ and $0 < \lambda_1 < \lambda_2 < 1$. As in [73], λ_1 and λ_2 were set to 10^{-4} and 0.9, respectively. The algorithm was stopped if no step length $\alpha^\star > 0$ that satisfies the Wolfe conditions can be found in 20 trials of backtracking. This stopping criterion has been evaluated in Chapter 3 [113]. Detailed information about the L-BFGS-B-PC implementation were presented in Chapter 3.

Since both the objective function and its gradient have to be computed for each candidate α, extra forward and back-projection operations are needed for the backtracking.

4.4 Data

In this study, the performance of the anatomical prior, PLS, with and without the new approximation of $\boldsymbol{\kappa}(\boldsymbol{x})$ in (4.19) is investigated. The potential of applying the modified PLS to improve the quantitative consistency and algorithm convergence rate are investigated with digital phantoms. A patient dataset with inserted pseudo lesions is also used as an example reconstruction closer to the clinical context.

4.4.1 Digital disc phantom

To demonstrate that the uniformity of the local perturbation for lesions in different surroundings can be improved by applying the modified PLS, a 2-D disc phantom with a sphere (value = 3) inserted right at the centre of a large hot (value = 5) or cold (value = 1) uniform region was used. The phantom was a 111×111 matrix with voxel size of 2.397×2.397 mm^2 and the diameter of the sphere was 21.573 mm. An attenuation map ($\boldsymbol{\mu}$) was also used to provide anatomical information, consisting of 0.096 and 0.172 cm^{-1} for the feature and the surrounding matter, respectively. Figure 4.1 shows the phantom with a high activity surrounding and the corresponding attenuation map as an example. Note that, relative to the surrounding region, the sphere had the same absolute activity difference and, hence, the effect of PLS would be the same for the spheres in both cold and hot surroundings.

The projection data were generated by using STIR [88] projectors to simulate data from a GE Discovery STE in 2D acquisition mode, without considering the Poisson noise. Other physical effects, such as attenuation and system blurring (modelled as smoothing in image space with full width at half maximum (FWHM) = 5.2 mm in tangential and radial directions), were simulated. A uniform projection with a constant intensity was added to the generated data to simulate the background events. It is equivalent to 90% and 64% of the total prompts for the phantom with cold and hot surroundings. For analysis purposes (see Section 4.5.2 for more information), data generated using the same phantom but with no sphere at the centre of both activity and anatomical images were also simulated.

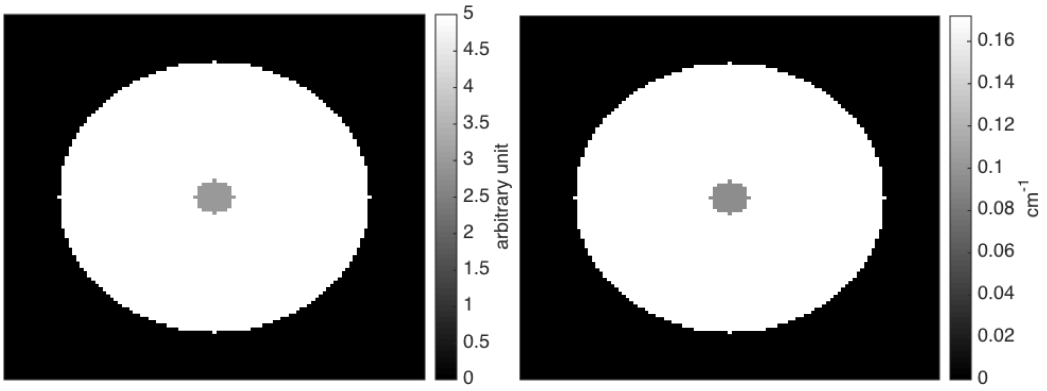

Figure 4.1: The activity distribution of the 2-D disc phantom with a high activity surrounding (left) and the corresponding attenuation map (right).

4.4.2 Digital XCAT phantom

The more realistic XCAT phantom [87] was used to generate data representing typical ^{18}F-FDG scans in the thorax. The resulting volume was of size $152 \times 152 \times 47$ with voxel size 3.125 mm in all directions. The phantom was rescaled to have voxel intensities between 0 to 2.

To evaluate the dependence of local perturbation on location, 6 hot lesions with diameter of $9.375 \times 9.375 \times 9.375$ mm^3 were inserted in different slices (see Figure 4.2). None of the lesions were in the central slice and 2 of them were in the liver. The uptake of the liver was either high (value $= 1.6$) or low (value $= 0.4$) to simulate change of surroundings for lesions in the liver. Each lesion had the same absolute difference to its surrounding in both activity (difference $= 0.8$) and anatomical images (difference $= 20$ Hounsfield units (HU)) hence the same influence of the anatomical prior. As lesions in liver would have a similar linear attenuation coefficient to liver, our simulations roughly correspond to using CT with injected contrast to provide anatomical information. Note that all lesions in the activity image were exactly aligned with those in the anatomical image. Figure 4.2 shows the XCAT phantom with high liver uptake and its corresponding contrast enhanced CT image as an example. The phantom was forward projected using vendor supplied software into sinograms simulating data from the GE Discovery STE in 3-D acquisition. The background events were simulated by adding a constant value to the generated sinograms. The amount of background events was equivalent to 56% and 55% of total

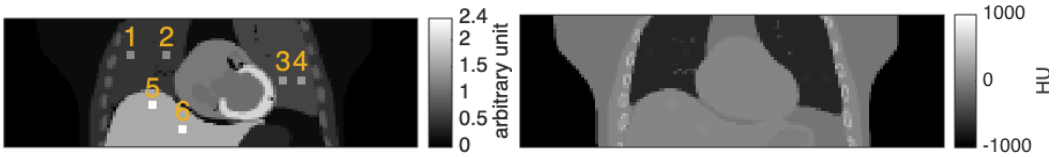

Figure 4.2: The central coronal view of the XCAT phantom with high liver uptake (top) and the corresponding contrast enhanced CT image used to provide anatomical information (bottom).

prompts for phantom with low and high liver uptake, respectively. Similar to the disc phantom, Poisson noise was not considered and the phantom with and without lesions was used to generate datasets for analysis purposes.

The phantom with high liver uptake was also used for investigating the influence of the modified PLS on the performance of L-BFGS-B-PC with respect to different noise and background levels. These factors had been proven to influence the convergence rate of L-BFGS-B-PC when using non-anatomical priors with spatially invariant weights in the previous chapter. For assessing effects of the noise level, three datasets (G0) with total counts S_{tot} of 50 M, 250 M and 1252 M were generated. The Poisson noise model was used. The true-to-background ratio (TBR) for each of them was fixed at 0.6. The possible influence of background events was studied using four more datasets, which were divided into two groups. For the first group (G1), each dataset had the same total counts as the data representing medium noise level, *i.e.*, $S_{\text{tot}} = 250$ M, but with 5 times lower or higher TBR. For data in the second group (G2), we kept S_{true} the same as that in the data with $S_{\text{tot}} = 250$ M counts, but changed S_{bg} such that it was 5 times lower or higher. Note that these two groups had identical TBR for the same background level: TBR = 0.12 for the high background data and TBR = 3 for the low background data. A summary of the simulated data representing different noise and background levels can be found in Table 4.1. All simulations took into account the attenuation effect and system blurring using FWHM = 5.2 mm in tangential and radial directions and 5.7 mm in the trans-axial direction.

Table 4.1: A summary of the simulated data for evaluating the influence of the background.

		S_{true}	S_{bg}	S_{tot}
G0	TBR = 0.6	19 M	31 M	50 M
	TBR = 0.6	94 M	156 M	250 M
	TBR = 0.6	469 M	783 M	1252 M
G1	TBR = 3	187 M	63 M	250 M
	TBR = 0.12	27 M	223 M	250 M
G2	TBR = 3	94 M	31 M	125 M
	TBR = 0.12	94 M	783 M	877 M

4.4.3 Patient data

A patient dataset of the thorax region acquired on the GE Discovery STE PET/CT scanner was used for this retrospective study. The patient was injected with 315 MBq of ^{18}F-FDG approximately 1 hour before the scan started. The acquisition included a cine-CT scan (140 kVp, 60 mA, 4 s duration, 0.5 s rotation period, 0.45 s time between reconstructed images, 9 bed positions, 8 axial slices per bed position), followed by a PET scan in fully 3-D mode. The total number of counts in the PET data was $S_{\text{tot}} = 219$M. Since the true lesion value and location are unknown, three pseudo lesions (two in the liver and one in the right lung of the patient, none of these in the central slice) with diameter of 9.375 mm were forward projected, and attenuated using vendor-provided software and added to the measured data for quantitative evaluation. Poisson noise was considered during data simulation. The difference between each pseudo lesion and its surroundings was 0.8 MBq/cc. The corresponding contrast enhanced CT with lesions inserted at the same locations was simulated from the average cine-CT to provide anatomical information. Based on the results in [114, 115, 116], an absolute difference of 150 HU to the surroundings was assigned to all pseudo lesions in the simulated enhanced CT image. The attenuation change induced by the inserted lesions was ignored. The vendor-provided software was also used for binning the PET data into sinograms and modelling the detection efficiency, attenuation, scatter and randoms.

4.5 Evaluation

The reconstruction conditions for all datasets are described in this section. In addition, two metrics of reconstruction algorithm performance: one to measure the improvement of consistency in lesion quantification and one to assess algorithm convergence rate are introduced. The former is mainly evaluated with respect to different surroundings and locations, while the latter is studied using simulations representing different noise and background levels.

4.5.1 Reconstruction conditions

Based on our previous study (see Section 3.6.3 for more information), L-BFGS-B-PC initialised by 1 full iteration of OS-EM with 35 subsets was adopted for image reconstruction. The initial image was also used to calculate the spatially-variant penalty strength $\hat{\boldsymbol{\kappa}}$ and the preconditioner $\boldsymbol{D}$. Each dataset with and without using the proposed $\hat{\boldsymbol{\kappa}}$ to modify PLS was reconstructed. For simulated data, the reconstructed image had the same matrix size as the corresponding phantom. The reconstructed voxel size was 3.125×3.125 mm^2 for the disc phantom and $3.947 \times 3.947 \times 3.27$ mm^3 for the XCAT phantom. The reconstructed image for the patient datasets had $192 \times 192 \times 47$ voxels with voxel size of $3.125 \times 3.125 \times 3.27$ mm^3. The system blurring was also modelled during reconstruction. To exclude the dependence of the penalty function on the selection of parameters, the parameter set (ε, η) in PLS was chosen according to the scale of the anatomical and activity images, respectively. The strengths of ε and η were kept to 12.5% and 25% of the intensity difference between the lesion and its surroundings in anatomical and activity images for all reconstructions. Therefore, the set of parameters was (0.25, 0.019 cm^{-1}) for the disc phantom, (0.1, 5 HU) for the XCAT phantom and (0.1, 37.5 HU) for the real patient data. A different global penalty strength $\hat{\beta} = \beta \hat{\kappa}_0^2$ was defined for reconstructions without applying $\hat{\boldsymbol{\kappa}}$, where $\hat{\kappa}_0$ represents the value at the centre of $\hat{\boldsymbol{\kappa}}$ calculated with a reference dataset. Therefore, the effective penalty strength at the image centre is identical in both reconstructions with and without using $\hat{\boldsymbol{\kappa}}$ when the reference dataset is considered. The influence of the global penalty strength on quantification is discussed further in Section 4.7. For different evalua-

tions, a different reference dataset was selected (see Section 4.6 for more information). For reconstructions using $\hat{\boldsymbol{\kappa}}$, the global penalty strength β was fixed at 1 for the disc phantom and 10^{-3} for the noiseless XCAT datasets. A stronger $\beta = 10^{-2}$ was chosen for data with noise, including the real patient data. The strength β for each dataset was determined based on an initial investigation, where the reconstructed image was visually compared with respect to a series of candidate strengths (not shown). The evaluated set of strengths started from 10^{-4} and increased by a factor of 10 to 10^{1}. Note that the selected strength for the disc phantom datasets was much stronger than that for other datasets in this study.

4.5.2 Analysis methods

The potential benefits of using the modified PLS were investigated in terms of visual interpretation, local perturbation and algorithm convergence rate. To be able to quantify the last two features, two metrics were adopted to measure the local contrast recovery ratio (*CR*) and the distance between the converged image and current estimate (*M*). The latter was initially introduced in Section 3.7. Given the reconstructed image with lesion(s) $\hat{\boldsymbol{f}}^{\text{true}}$ and that without lesion(s) $\hat{\boldsymbol{f}}^{\text{bg}}$, the metrics are defined as:

$$CR = \frac{|\text{ROImean}(\hat{\boldsymbol{f}}_{\text{c}}^{\text{true}} - \hat{\boldsymbol{f}}_{\text{c}}^{\text{bg}})|}{\text{True Contrast}} \times 100\%, \tag{4.27}$$

$$M_t = \sqrt{\frac{1}{\text{N}} \frac{\|\hat{\boldsymbol{f}}_t^{\text{true}} - \hat{\boldsymbol{f}}_{\text{c}}^{\text{true}}\|_2^2}{\text{mean}(\hat{\boldsymbol{f}}_{\text{c}}^{\text{true}})^2}}, \tag{4.28}$$

where $\text{ROImean}(\cdot)$ is an operator that calculates the mean value of the ROI, "True Contrast" is the assigned activity difference between the lesion and its background and N is the number of voxels. The subscript t indicates the iteration number and c denotes that the converged image of L-BFGS-B-PC is considered (see Section 4.3 for the definition of convergence of L-BFGS-B-PC). The ROIs were drawn in the centre of the sphere or lesion with size of 9×9 voxels for the disc phantom and $3 \times 3 \times 3$ voxels for both the XCAT phantom and real patient data. Note that the local perturbation was evaluated using noiseless data and the convergence rate was

investigated using data representing different noise and background levels.

To further explore the performance of the algorithm with and without applying the modified PLS, the required number of projection operations and the corresponding iterations for achieving "practical" convergence were computed as well. As was proposed in Section 3.7, the corresponding iteration number was determined by:

$$t_M^\star = \min\{t : M_t \leq 0.01\} . \tag{4.29}$$

4.6 Results

The differences between $\boldsymbol{\kappa}(\boldsymbol{f})$, $\tilde{\boldsymbol{\kappa}}$ and $\hat{\boldsymbol{\kappa}}$ was evaluated using simulated datasets in the presence and absence of noise. Quantitative evaluations for the local perturbation was achieved by comparing the *CR* value of each sphere or lesion at convergence. For quantifying the algorithm convergence rate, plots of the convergence measure M against the total number of projection operations were used. One projection operation means a forward or back-projection of the full dataset. We chose the number of projection operations instead of iterations as it takes into account the extra computational demand induced by the backtracking algorithm. Fast decrease of M values indicates fast convergence rate to the solution of the reconstruction problem. The use of L-BFGS-B-PC with the modified PLS was then demonstrated using a patient dataset.

4.6.1 Comparison between $\kappa_j(\boldsymbol{f})$, $\tilde{\kappa}_j$ and $\hat{\kappa}_j$

To demonstrate that the proposed $\hat{\boldsymbol{\kappa}}$ calculated with the suggested initialisation is a reasonable approximation of $\boldsymbol{\kappa}(\boldsymbol{f})$, we first investigated the difference between them using simulated noiseless datasets. Figure 4.3 shows both values for the central coronal plane of the XCAT phantom with high liver uptake. Profiles along the central point of each image are also provided. As illustrated in the figure, $\hat{\boldsymbol{\kappa}}$ is nearly identical to $\boldsymbol{\kappa}(\boldsymbol{f})$. Although not shown, similar results were obtained for both the disc phantom and XCAT phantom with low liver uptake.

The comparison between $\tilde{\boldsymbol{\kappa}}$ and $\hat{\boldsymbol{\kappa}}$ was then conducted with a simulated dataset for the XCAT phantom with high liver uptake as well but in the presence of noise.

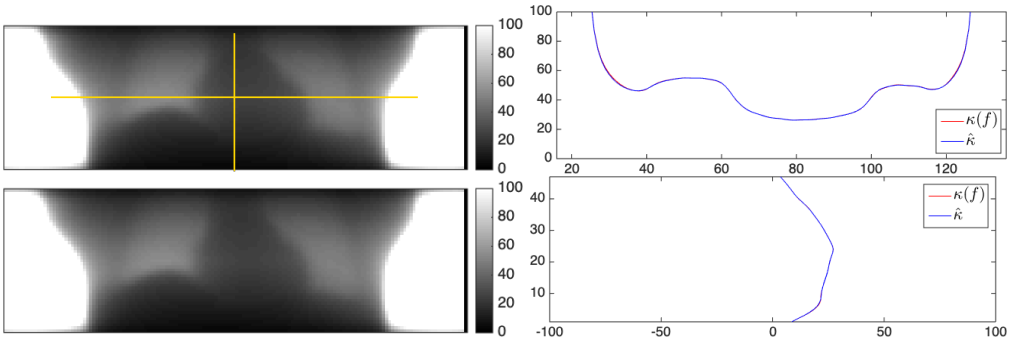

Figure 4.3: $\kappa_j(\boldsymbol{f})$ and $\hat{\kappa}_j$ values (left column, top and bottom images, respectively) for the central coronal plane of the XCAT phantom with high liver uptake. Horizontal and vertical profiles through the central point of each map are also provided (right column, top and bottom profiles, respectively).

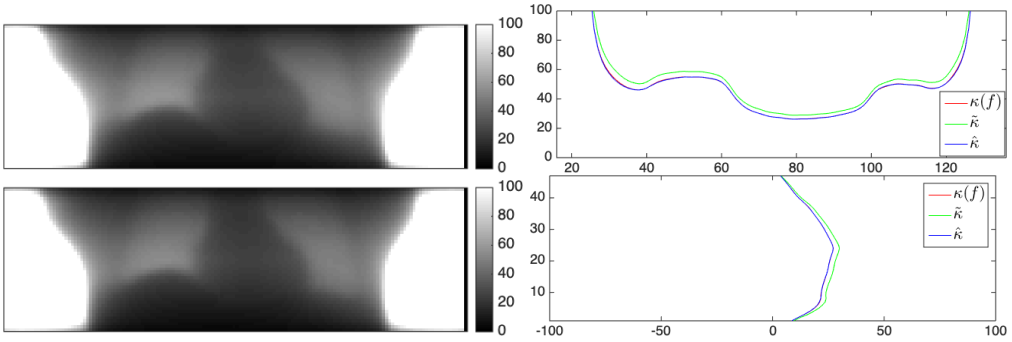

Figure 4.4: $\tilde{\kappa}_j$ and $\hat{\kappa}_j$ values (left column, top and bottom images, respectively) for the central coronal plane of the XCAT phantom with high liver uptake. Poisson noise was considered and the total counts of the data was 250 M. Horizontal and vertical profiles through the central point of each map are also provided (right column, top and bottom profiles, respectively), compared with those of $\kappa_j(\boldsymbol{f})$ calculated with the corresponding noiseless dataset.

The total counts of the data was 250 M. More detail regarding the data can be found in Table 4.1. For each approximation of $\boldsymbol{\kappa}(\boldsymbol{f})$, the central coronal plane and profiles along its central point are provided in Figure 4.4. Results were compared with those from $\boldsymbol{\kappa}(\boldsymbol{f})$ calculated with the corresponding noiseless dataset. Although the central coronal planes for both approximations are visually identical, profiles for $\tilde{\boldsymbol{\kappa}}$ and $\hat{\boldsymbol{\kappa}}$ are not the same. The former show higher values than those for $\boldsymbol{\kappa}(\boldsymbol{f})$, while the latter are nearly indistinguishable from those for $\boldsymbol{\kappa}(\boldsymbol{f})$. The results support the feasibility of using $\hat{\boldsymbol{\kappa}}$, instead of $\tilde{\boldsymbol{\kappa}}$, as an alternative to $\boldsymbol{\kappa}(\boldsymbol{f})$.

Table 4.2: *CR* values for sphere in different surroundings.

	hot surrounding	cold surrounding
with $\hat{\boldsymbol{\kappa}}$	53	51
without $\hat{\boldsymbol{\kappa}}$	41	54

4.6.2 Dependence of contrast ratio on surroundings & location

The same noiseless datasets for both digital phantoms were used to study the sensitivity of *CR* to different surroundings and locations. The data generated using the phantom with hot surroundings or high liver update was selected as the reference dataset such that the global penalty strength was the same for both cases with different surrounding activities, regardless of the use of $\hat{\boldsymbol{\kappa}}$ or not. This is helpful in evaluating the influence of applying a constant or spatially-variant penalty strength to datasets for similar objects but with different activity distribution as currently performed in clinical practice [117, 118]. Converged images for the disc phantom are shown in Figure 4.5. The sphere has a relatively low visual contrast to hot surroundings when $\hat{\boldsymbol{\kappa}}$ is not applied (Figure 4.5 (b)). However, the benefit of using $\hat{\boldsymbol{\kappa}}$ to preserve visual contrast is less significant for the sphere in low activity surroundings (Figure 4.5 (d)). The corresponding *CR* values are listed in Table 4.2 where a lower sensitivity to the change of the surrounding activity is observed when $\hat{\boldsymbol{\kappa}}$ is applied.

Similar evaluations were performed using the XCAT phantom to investigate the dependence of *CR* on different lesion locations. Figure 4.6 shows the central coronal view of the converged images reconstructed with and without applying $\hat{\boldsymbol{\kappa}}$ for the phantom with high liver uptake. The corresponding difference images are also provided to assist the visual comparison. As shown in the figure, a relatively uniform visual contrast for lesions at different locations is observed in reconstructions using $\hat{\boldsymbol{\kappa}}$ (Figure 4.6 (a) and (c)). When $\hat{\boldsymbol{\kappa}}$ is not applied (Figure 4.6 (b) and (d)), the lesions near the end slices has a lower visual contrast (indicated with purple arrows). Although not shown, similar results were obtained for data simulated with low liver uptake. *CR* values for each lesion under different data simulation and reconstruction conditions are calculated and listed in Table 4.3. Consistent with the visual comparison, the variation of the contrast recovery ratio in locations is reduced

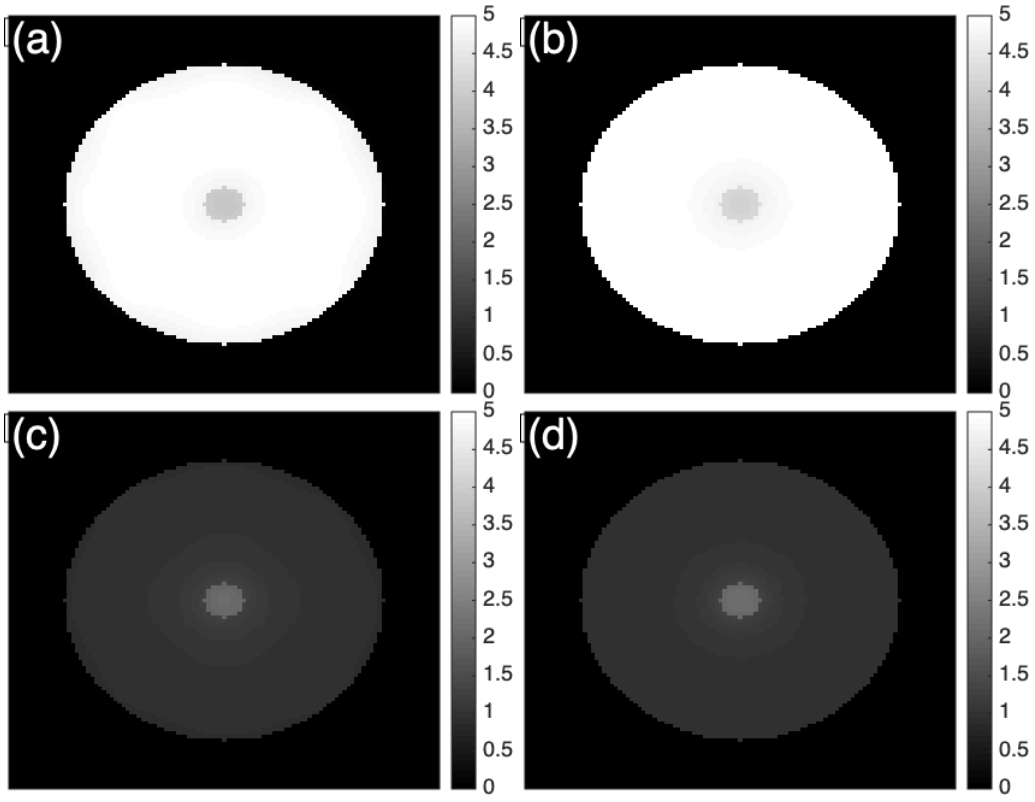

Figure 4.5: Converged images for the disc phantom with high (top row) or low (bottom row) activity uptake in surroundings. Images reconstructed with and without using $\hat{\boldsymbol{\kappa}}$ are presented in left and right columns, respectively. Note that the same image scale [0, 5] is used for assisting the visual comparison.

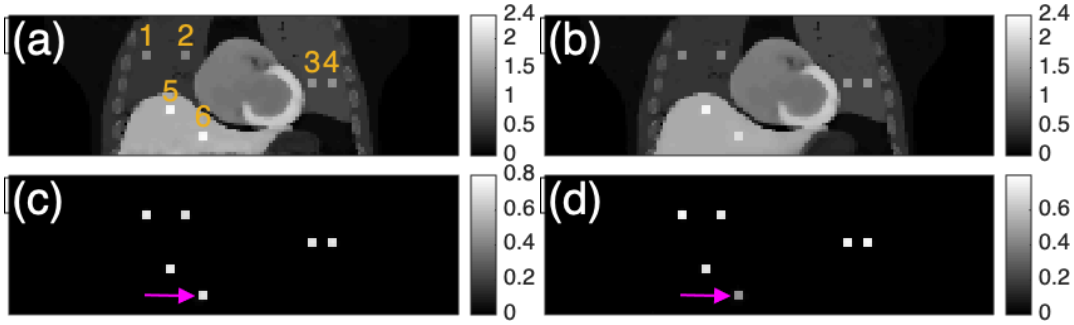

Figure 4.6: Converged images reconstructed with (a) and without (b) using $\hat{\boldsymbol{\kappa}}$ for the XCAT phantom having high liver uptake. The corresponding difference images are given in (c) and (d), respectively.

when $\hat{\boldsymbol{\kappa}}$ is used. However, the influence of surrounding activity is not obvious for the XCAT phantom as lesion 5 and 6 show similar *CR* values when the activity level of their surroundings (liver) is changed, regardless of $\hat{\boldsymbol{\kappa}}$ being applied or not.

Table 4.3: *CR* values for lesions at different locations.

		1	2	3	4	5	6	mean ± std
with $\hat{\boldsymbol{\kappa}}$	HL[1]	91	87	89	90	92	88	89 ± 1.78
	CL	91	88	89	90	88	86	
without $\hat{\boldsymbol{\kappa}}$	HL	95	93	95	96	89	59	88 ± 13.35
	CL	96	93	95	96	88	61	

[1] HL and CL stand for hot and cold liver uptake, respectively.

Table 4.4: Required projection and iteration numbers for achieving convergence of M values for datasets representing different noise levels.

	with $\hat{\boldsymbol{\kappa}}$	without $\hat{\boldsymbol{\kappa}}$
$S_{\text{tot}} =$ 50 M	82 (39)	412 (200)
$S_{\text{tot}} =$ 250 M	92 (45)	686 (334)
$S_{\text{tot}} =$ 1252 M	132 (65)	812 (398)

4.6.3 Dependence of convergence rate on noise & background level

Simulated data with $S_{\text{tot}} = 50$ M, 250 M and 1252 M, representing high, medium and low noise levels, were reconstructed with and without using $\hat{\boldsymbol{\kappa}}$ to study the influence of the noise level. The $\hat{\boldsymbol{\kappa}}$ for the dataset with $S_{\text{tot}} = 250$ M was chosen as the reference. Figure 4.7 shows the central coronal view of the converged image of each reconstruction. Noisier end slices of the reconstructed images are observed when the modified PLS is applied compared to those without using $\hat{\boldsymbol{\kappa}}$. The effect becomes less apparent as the noise level of the data decreases. In addition to visual comparison, the performance of the algorithm was also evaluated by plotting convergence estimates M against the total projection operations (Figure 4.8) and by listing the required number of projection operations and iterations to achieve the practical convergence defined in (4.29) (Table 4.4). The former illustrates the convergence rate in early iterations while the latter gives an insight to late iterations. The convergence rate is improved significantly when applying $\hat{\boldsymbol{\kappa}}$. Moreover, it also reduces the difference in convergence rate for data with different noise levels.

The effect of the background level on the performance of L-BFGS-B-PC with and without using $\hat{\boldsymbol{\kappa}}$ was investigated with data in both groups of fixed S_{tot} (G1) and fixed S_{true} (G2). The results were compared to those from the data having medium

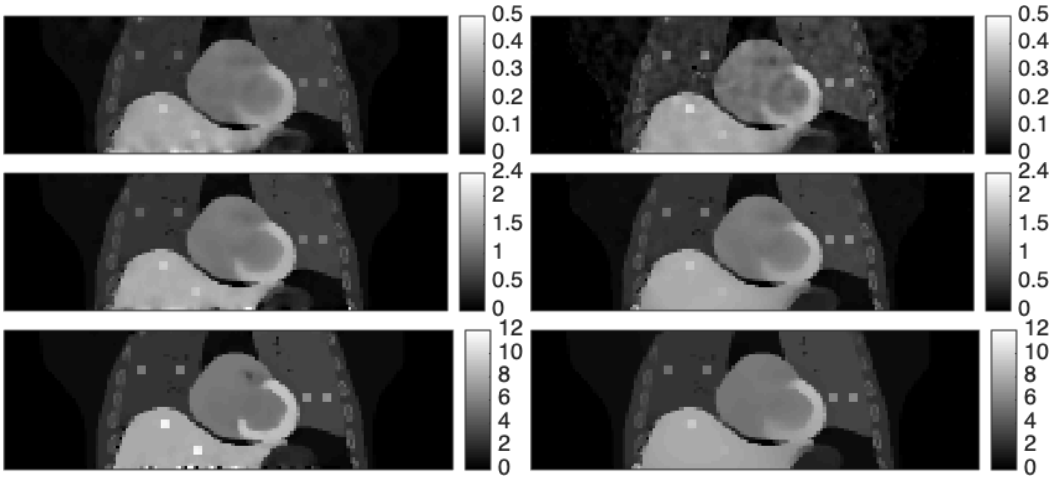

Figure 4.7: Converged images reconstructed with (left column) and without (right column) using $\hat{\boldsymbol{\kappa}}$ for the XCAT phantom with high liver uptake. From top to bottom row are images for high, medium and low noise level. The scale for each image was set according to the different scaling factors leading to different noise levels.

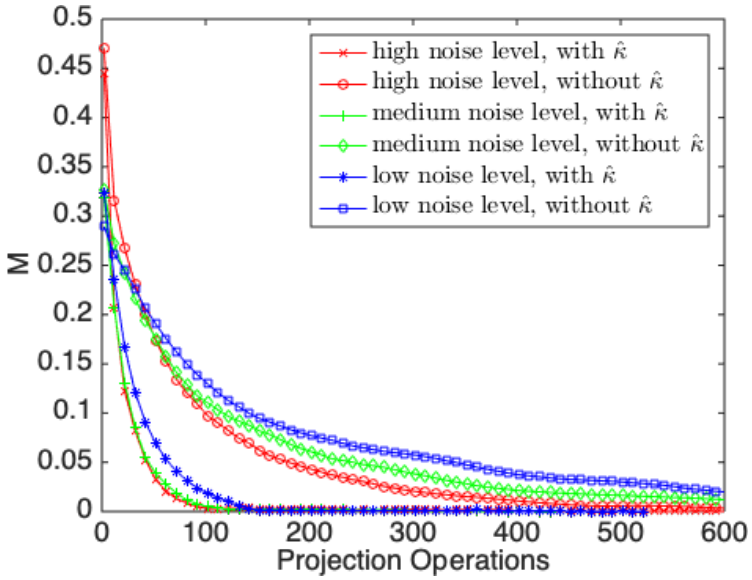

Figure 4.8: M values plotted against the total number of projection operations for datasets representing different noise levels.

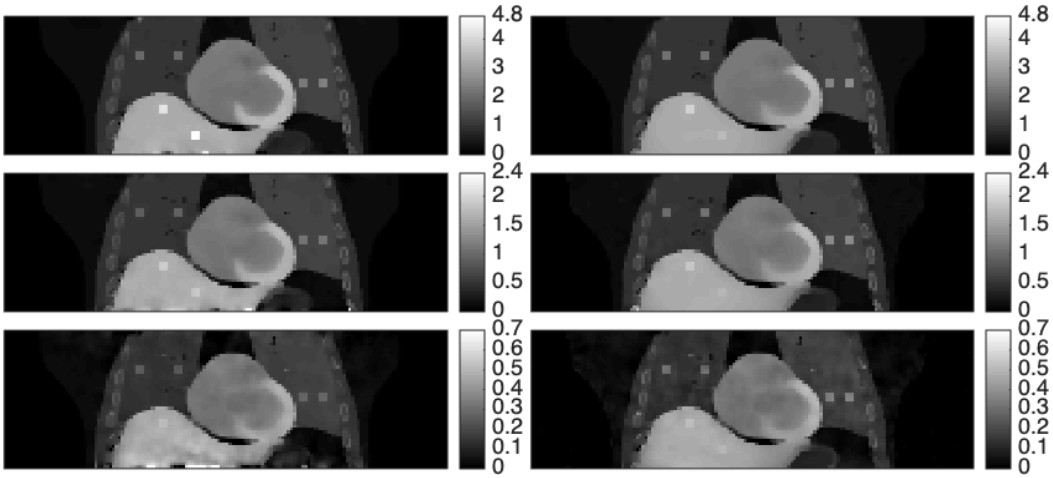

Figure 4.9: Converged images reconstructed with (left column) and without (right column) using $\hat{\boldsymbol{\kappa}}$ for data in G1. From top to bottom row are images for high, medium and low TBR. The scale for each image was set according to the different scaling factors leading to different number of true events.

Table 4.5: Required projection and iteration numbers for achieving practical convergence of M values for datasets representing different background levels.

		with $\hat{\boldsymbol{\kappa}}$	without $\hat{\boldsymbol{\kappa}}$
G1	TBR = 3	82 (40)	644 (313)
	TBR = 0.12	92 (43)	522 (251)
G2	TBR = 3	82 (40)	512 (253)
	TBR = 0.12	142 (56)	624 (300)

noise and background level ($S_{\text{tot}} = 250$ M, $S_{\text{true}} = 94$ M, $S_{\text{bg}} = 156$ M and TBR = 0.6). The reference strength $\hat{\kappa}_0$ was also calculated with this dataset. For both groups of data, the converged images are slightly more affected by noise when the modified PLS is used (Figure 4.9 and 4.10). In terms of convergence rate, plots of M values against the total number of projection operations for each dataset in both groups are provided in Figure 4.11. The required number of projection operations for achieving the practical convergence of M values are listed in Table 4.5. Based on the results, the use of the spatially-variant penalty strength shows the ability to improve not only the convergence speed in both early and late iterations but also the consistency of the convergence rate among data with different background levels.

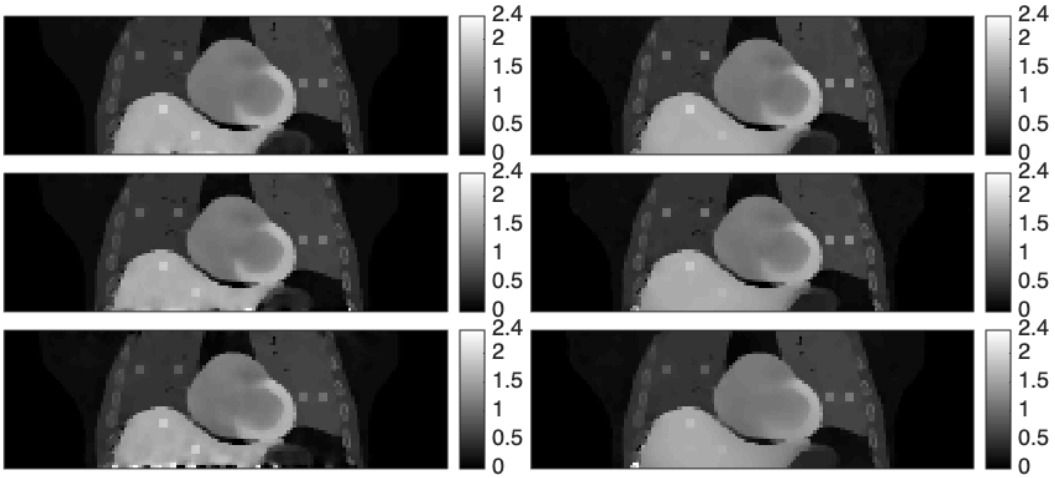

Figure 4.10: Converged images reconstructed with (left column) and without (right column) using $\hat{\boldsymbol{\kappa}}$ for data in G2. From top to bottom row are images for high, medium and low TBR.

Table 4.6: *CR* values for each pseudo lesion.

	1	2	3	mean $\pm$ std
with $\hat{\boldsymbol{\kappa}}$	41	53	44	46 ± 6.25
without $\hat{\boldsymbol{\kappa}}$	55	49	17	40 ± 20.43

Table 4.7: Required projection and iteration numbers for achieving practical convergence of M values for the patient dataset with lesions.

	with $\hat{\boldsymbol{\kappa}}$	without $\hat{\boldsymbol{\kappa}}$
$S_{\text{tot}} = 219$ M	42 (20)	262 (129)

4.6.4 Demonstration with patient data

Figure 4.12 shows the coronal plane through the centre of the inserted lesions for both reconstructions with and without using $\hat{\boldsymbol{\kappa}}$ for the patient data. A relatively uniform visual contrast for lesions at different locations is observed when $\hat{\boldsymbol{\kappa}}$ is available. *CR* values for all lesions in both reconstruction conditions are listed in Table 4.6. Consistent with the visual comparison, a smaller variation in the contrast recovery ratio is obtained for data reconstructed using $\hat{\boldsymbol{\kappa}}$. The required number of projection operations for achieving the practical convergence (4.29) is also provided in Table 4.7. Again, the convergence rate of the reconstruction algorithm is substantially improved when using $\hat{\boldsymbol{\kappa}}$.

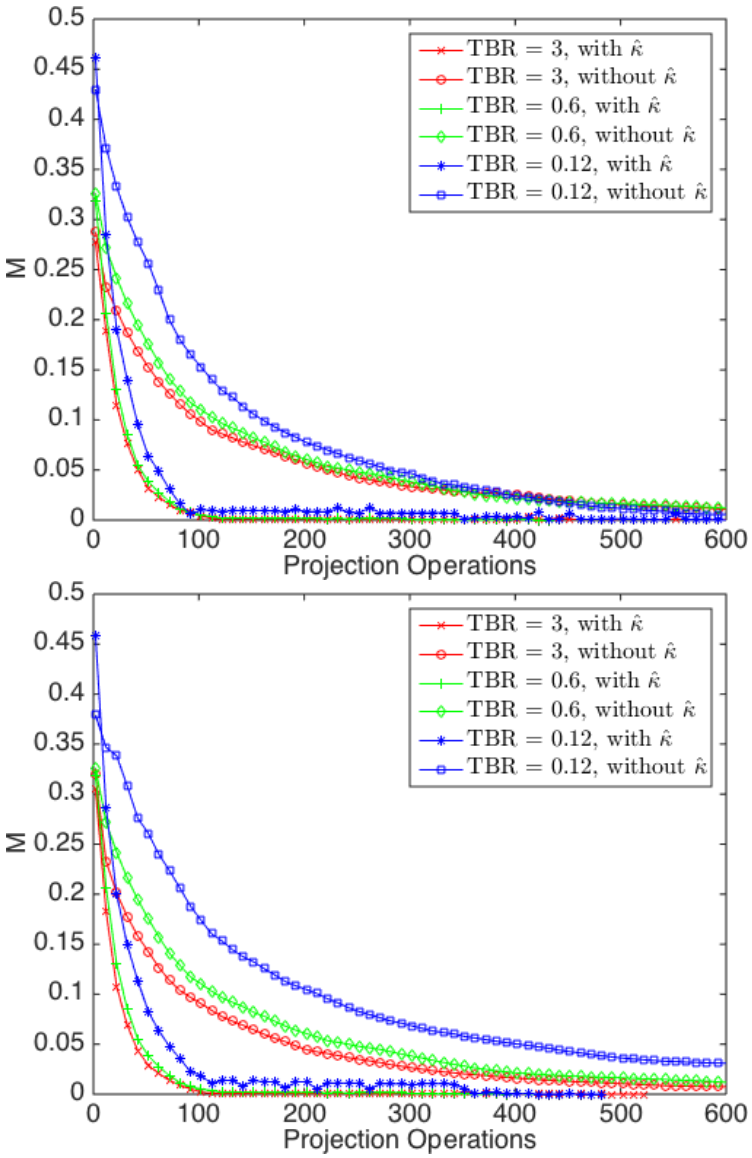

Figure 4.11: *M* values plotted against the total number of projection operations for data sets in G1 (top) and G2 (bottom).

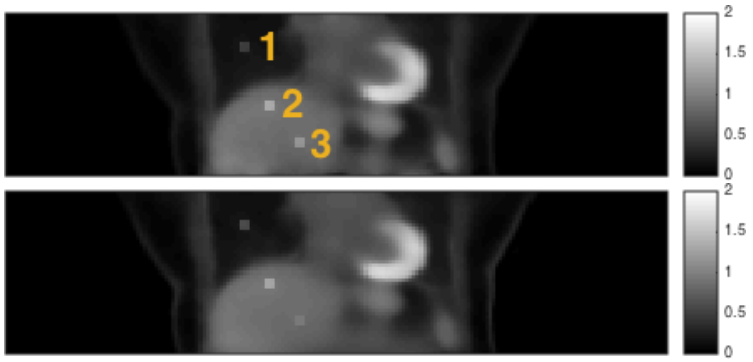

Figure 4.12: Converged images reconstructed with (top) and without (bottom) using $\hat{\boldsymbol{\kappa}}$ for the patient dataset with inserted pseudo lesions. The lesions are marked from 1 to 3 as illustrated in the top image.

4.7 Discussion

This study demonstrates the feasibility of using the spatially-variant penalty strength $\hat{\boldsymbol{\kappa}}$ with an anatomical penalty to improve the uniformity of contrast (and LPR) for lesions in different surroundings or at different locations. Its potential benefits to the convergence rate and convergence consistency are also explored with a previously proposed reconstruction algorithm, L-BFGS-B-PC.

Although the performance of the proposed $\hat{\boldsymbol{\kappa}}$ can be influenced by the initial image, we have demonstrated that the difference between $\boldsymbol{\kappa}(\boldsymbol{f})$ and $\hat{\boldsymbol{\kappa}}$ is small with the suggested initial condition (1 full iteration of OS-EM with 35 subsets) (Figure 4.3) using noiseless datasets. As other "plug-in" methods have been suggested to substitute $\bar{\boldsymbol{g}}(\boldsymbol{f})$ in (4.15) by the noisy measurement [103], the difference was also compared with one noisy dataset with total counts of 250 M and the results were essentially the same (Figure 4.4). Given the proposed $\hat{\boldsymbol{\kappa}}$ does not require the collected data to be inverted, it has no potential division-by-zero problem compared to the "plug-in" methods. Moreover, since $\hat{\boldsymbol{\kappa}}$ was reused to construct the transformation matrix in L-BFGS-B-PC, no additional computation was required. The computational efficiency combined with reduced sensitivity to noise make the proposed $\hat{\boldsymbol{\kappa}}$ a practical substitution for $\boldsymbol{\kappa}(\boldsymbol{f})$. However, as the SNR of the noisy dataset used for the comparison was reasonably high, further investigations on the highest limit of data noise level for applying $\hat{\boldsymbol{\kappa}}$ to improve the uniformity of local contrast and algorithm performance are necessary.

When the preconditioner $\boldsymbol{D}$ was constructed, the spatially-variant penalty strength $\hat{\boldsymbol{\kappa}}$ was reused and the Hessian of the penalty function was replaced with a small constant $\zeta = 10^{-4}$. This approximation of the Hessian of the objective function is suboptimal for large β and the convergence rate of the algorithm might become slow. Since a relatively small β was chosen for the reconstructions used for performance evaluation (*i.e.*, reconstructions for the XCAT phantom and patient dataset), this approximation is reasonable in this study. For future investigation on convergence rate with large β, it would be helpful to take into account the strength of β while constructing $\boldsymbol{D}$.

As discussed in [103], the dependence of the local perturbation on surrounding activity can be explained using the analytical model defined in (4.14). For any voxel j in the image, the local perturbation around j depends on the strength of $\kappa_j(\breve{\boldsymbol{f}}_{\mathrm{b}})$. When the activity level of $\breve{\boldsymbol{f}}_{\mathrm{b}}$ is high around j, the number of rays crossing through the voxel would be large, leading to a small $\kappa_j(\breve{\boldsymbol{f}}_{\mathrm{b}})$ hence a strong smoothness over features around j. For reasonable penalty strengths, the use of PLS should help to achieve high *CR* values as it is designed to prevent smoothing across the anatomical boundaries that correspond with edges in activity images. However, to demonstrate the dependence of quantification consistency on surroundings, an unrealistically high $\beta = 1$ was used for the disc phantom. The strong smoothness across the boundaries leads to a redistribution of the activity. Therefore, the *CR* values for the disc phantom were only around 50% in this study (Table 4.2). The dependence of the contrast recovery on surroundings was not obvious for the XCAT phantom (lesions 5 and 6 in Table 4.3). This is because the selected penalty strength was reasonable, judging from the small differences in *CR* values for lesions 5 and 6 in high and low liver uptake.

In fact, for a uniform activity distribution with homogeneous attenuation effect, $\kappa_j(\breve{\boldsymbol{f}}_{\mathrm{b}})$ can be interpreted as an index of the spatial variations in sensitivity. Using the spatially-variant penalty strength to modify PLS is therefore roughly equivalent to applying sensitivity compensation across the FOV. This explains the low *CR* value for the lesion at end slices when $\hat{\boldsymbol{\kappa}}$ was not used (lesion 6 in Table 4.3). In the case of

applying $\hat{\boldsymbol{\kappa}}$ in the presence of Poisson noise and edge information, there would be a trade-off between the image noise level and the uniformity of LPR across the FOV, especially for lesions at end slices. Fortunately, as multiple bed positions are often used in clinical practice, the increased noise level of end slices can be compensated by averaging across neighbouring bed positions.

Regardless of the use of $\hat{\boldsymbol{\kappa}}$, the LPR is always affected by the local property of the object introduced by the penalty function. It varies with the global strength β that controls the influence of the penalty over whole objective function and parameters that determine the weight of the edge information in the anatomical or functional image. For example, since β used for the patient dataset ($\beta = 10^{-2}$) was stronger than that for noiseless data generated using the XCAT phantom ($\beta = 10^{-3}$), a lower *CR* value was observed for lesions in the patient dataset (Table 4.3 and 4.6). The XCAT phantom was simulated such that it had a similar activity distribution to the patient data in the thorax region and the set of parameters, (ε, η), in PLS were scaled accordingly to enforce similar local effect in both reconstructions. When the same β is chosen, lesions in the patient data are able to converge to similar *CR* values as those in the XCAT phantom (not shown). However, the reconstructed image becomes noisy due to the reduction of the penalty strength. The main motivation of rescaling ε and η in PLS according to the scale of the functional and anatomical images is to have a similar local influence from the penalty function for different datasets. In this way, the comparison of *CR* values depends mainly on the global parameter β and the use of the spatially-variant penalty strength $\hat{\boldsymbol{\kappa}}$. As the smoothness enforced by PLS in a uniform region (or also across the boundary when β is large) is modulated by ε and η, the benefit of using $\hat{\boldsymbol{\kappa}}$ might be affected by these two parameters. This means that the redistribution of the activity across the boundaries for large β might be less severe with small ε such that using $\hat{\boldsymbol{\kappa}}$ would be less beneficial. However, discussions on the dependence of the quantitative accuracy and image quality on parameters that control the global strength of the penalty or the strength of the local information are beyond the scope of this study.

The convergence rate of the algorithm can be improved significantly (Fig. 4.8

and 4.11) when $\hat{\boldsymbol{\kappa}}$ is applied. Moreover, the variability of the algorithm performance for different datasets is reduced. As shown in Table 4.4 and 4.5, all data representing different noise and background levels show the ability to achieve practical convergence in 150 projection operations. This implies that a more reliable quantitative comparison among different datasets can be obtained if a fixed number of projection operations is chosen.

For the XCAT phantom and patient dataset, simulated contrast enhanced CT images were used to provide anatomical information. This could have wide-ranging implications to the work-flow and economics of the procedure in clinical use. The overall radiation dose received by the patient can also be increased due to the additional contrast enhanced CT scan. Although this might not be an issue for lesions involving density change as the attenuation map can be used to provide the corresponding anatomical information, the potential benefits and limitations of applying the proposed strategy in clinical practice should be considered.

4.8 Conclusions

In this study, the use of a spatially-variant penalty strength with a convex anatomical penalty function to reduce the dependence of local perturbation on surrounding activity and location has been demonstrated. The proposed weighting scheme for the penalisation can be precomputed. Moreover, when using L-BFGS-B-PC, the weights can be reused for constructing the preconditioner such that the overall computational demand remains unchanged. Based on the results for simulated data, the effective convergence rate of M values is considerably improved when the spatially-variant penalisation is applied. Moreover, the variation of the convergence rate for different data noise and background levels is largely reduced when the weighting is taken into account. The idea is further demonstrated using a real patient dataset with inserted pseudo lesions in different slices. Consistent with the simulation results, significant improvement in quantitative consistency and algorithm performance is observed.

Chapter 5

Approaches for joint misalignment estimation and PML reconstruction using anatomical priors in ET

In the previous chapter, the anatomical information is assumed to be exactly aligned with the activity distribution. To take into account the potential misalignment between the functional and anatomical images, two approaches that perform alternating misalignment estimation and image reconstruction are proposed in this chapter. Both methods are evaluated using simulated data with respect to different factors, such as algorithm workflow and initialisation.

5.1 Introduction

PML image reconstruction using penalties derived from anatomical images, such as CT or MR images, has been shown to be effective in improving object delineation and reducing quantitative error in many studies [38, 44, 45, 46, 47, 48, 49, 50, 51, 68]. However, to utilise the structural information without incurring artefacts, a good alignment between the anatomical and functional images is essential [48, 51, 119, 56]. This is challenging in practice because these images are most likely obtained separately or sequentially. Even with a multimodality scanner that performs simultaneous acquisition (e.g. Siemens mMR system and GE Signa PET/MR), good alignment between the acquisitions is still difficult to achieve due

to the different time scales of the scans. In particular, for the misalignment involving object deformation, for instance due to respiration, the assumption is even harder to satisfy. To take into account the mismatch in regions interfered by non-rigid motion, accurate image segmentation and co-registration (which are both difficult and time consuming) are often required [120, 121].

In many cases, anatomical information can be derived from the attenuation image. Since the literature on the misalignment issue between a priori anatomical information and the functional information is quite limited, we instead seek ideas from a similar but previously studied problem in thoracic PET imaging, in which a potentially misaligned CT or MR-derived attenuation map is used for the attenuation correction. Although these methods are intended to be used for having a better attenuation corrected image from emission data, they offer insights into resolving the misalignment between the anatomical attenuation map and the functional emission image. As for PML image reconstruction using anatomical priors, the misalignment induced by patient respiratory degrades resolution of the reconstructed image and introduces artefacts where large movement or deformation of organs is observed [122, 123].

This chapter will concentrate on imaging of the thorax, where respiratory motion is a known problem [124]. One strategy to tackle the respiratory motion is to sort the acquired data from both modalities into several gates where no motion is assumed in each of them. The gated data are then paired up according to their breathing phases estimated from the data themselves or an external tracking system [125]. In addition to reconstructing these gated data pairs individually and then registering them to a reference respiratory phase [126], one can also incorporate the corresponding attenuation information into the 4-dimensional (4-D) reconstruction algorithms [127, 128]. However, as these methods rely on regular and consistent breathing pattern during both scans [129], they are not able to cope with the residual misalignment caused by irregular breathing or other general motion of the patient. Moreover, they imply the need of special scans to obtain the gated anatomical information. This can increase patient total dose or prolong the overall examining time,

depending on the applied anatomical imaging modality. To adapt to irregular breathing patterns, another strategy applies individual motion model to deform the input attenuation map [130]. The model can be derived from other imaging data, such as dynamic CT and MR [131], as well as the non-attenuation corrected PET [132]. However, the former approach has the potential problem of propagating the error in the model estimation to the final reconstructed image, while the latter shows the dependence of the performance on the tracer distribution and data statistics. The other way around these issues is to use population-based deformation models [133, 134]. However, in addition to the need of being combined with other approaches, they have not been convincingly shown to work in practice [130]. The application of applying an individual model to PML image reconstruction using anatomical priors therefore faces similar challenges.

Algorithms that allow simultaneous estimation of the activity distribution and the corresponding attenuation map from the respiratory gated PET data have been proposed [135, 136] in recent years. These methods do not rely on assumptions about the breathing cycle or a pre-estimated motion model. Therefore, they have the potential to be applied to different misalignment problems without suffering from the error propagation issue. However, since the problem is very ill-conditioned, some *a priori* knowledge about the intensity distribution of the attenuation map is required. This can compromise the benefit of using anatomical information during the image reconstruction as the intensity is restricted to several values and most of the anatomical details are lost. Besides, severe cross-talk between the estimated activity and attenuation map is observed in non-TOF PET. Although the artefact can be eliminated when TOF data is available, the *a priori* knowledge about the intensity distribution is still necessary [136].

In contrast to seeking the attenuation map aligned with the emission image, a different joint estimation algorithm incorporates a warp matrix that deforms both the activity distribution and the attenuation map into the objective function [137, 138]. By optimising the objective function using an alternating process between motion estimation and image reconstruction, the motion compensation and attenuation cor-

rection are achieved simultaneously. The optimisation can be applied to both non-TOF and TOF data [139] without incurring artefacts but with a significantly improved convergence rate when TOF data is available. The algorithm uses a smoothing penalty function on the motion field to help the convergence rate of the motion estimation. The *a priori* knowledge of the attenuation distribution is not necessary any more. This study motivates our idea of applying the warp matrix to an anatomical prior, such as the previously mentioned PLS (see Section 2.3.2.2). Extending on the existing method, two approaches that account for the misalignment between the functional and anatomical images by incorporating the warp matrix into the penalised objective function are proposed in Section 5.3. As a special case of the application, we will only study the alignment of one PET position with a single CT derived attenuation map, which is also used to provide anatomical information.

5.2 Objective function without considering the potential misalignment

In this section, the objective function for penalised image reconstruction using anatomical priors without considering the potential misalignment between the anatomical and emission information (see Section 2.3.2.2) is briefly summarised. Given the emission image $\boldsymbol{f} = [f_1, \ldots, f_J]^\top \in \mathbb{R}^J$, the anatomical image $\boldsymbol{z} = [z_1, \ldots, z_J]^\top \in \mathbb{R}^J$ and the measured data $\boldsymbol{g} = [g_1, \ldots, g_I]^\top \in \mathbb{R}^I$, the objective function can be written as:

$$\Phi(\boldsymbol{f}) = -L(\boldsymbol{f}, \boldsymbol{g}) + \beta R(\boldsymbol{f}|\boldsymbol{z}), \tag{5.1}$$

where L is the log-likelihood and R is the penalty function with a parameter β controlling its strength. As in the previous chapter, the modified PLS in (4.19) is chosen as a representative anatomical prior for achieving both good image quality

and uniform local contrast in this study:

$$
\begin{aligned}
R(\boldsymbol{f}|\boldsymbol{z}) &= \sum_j \kappa_j^2 \sqrt{\varepsilon^2 + \| [\nabla \boldsymbol{f}]_j \|_2^2 - \langle [\nabla \boldsymbol{f}]_j, [\boldsymbol{\xi}]_j \rangle^2}, \\
\boldsymbol{\kappa} &= \left[\boldsymbol{A}^\top \mathrm{diag}\left\{ \frac{\boldsymbol{g}}{(\boldsymbol{A}\boldsymbol{f}^{\mathrm{init}} + \boldsymbol{n})^2} \right\} \boldsymbol{A}\mathbf{1} \right]^{\frac{1}{2}}, \\
[\boldsymbol{\xi}]_j &:= \frac{[\nabla \boldsymbol{z}]_j}{\sqrt{\| [\nabla \boldsymbol{z}]_j \|_2^2 + \eta^2}}, \ \varepsilon \text{ and } \eta > 0
\end{aligned} \tag{5.2}
$$

where $\boldsymbol{A} \in \mathbb{R}^{I \times J}$ represents the system matrix, $\boldsymbol{n} \in \mathbb{R}^I$ is the expected background events vector, $\mathbf{1}$ is a vector of ones and $\boldsymbol{f}^{\mathrm{init}}$ is the initial image. The edge preserving property of PLS is modulated by the pair of parameters (ε, η). The notation ∇ is the gradient operator, $\langle \cdot, \cdot \rangle$ is the Euclidean scalar product and $\| \cdot \|_2$ denotes the ℓ^2-norm.

5.3 Objective function considering the potential misalignment

Two approaches that account for the misalignment between the functional and anatomical images in penalised image reconstruction using anatomical priors are proposed in this section. Both approaches are based on a joint motion estimation and image reconstruction method proposed recently for dealing with the mismatch between the attenuation map and the PET image in respiratory gated PET/CT [137]. Instead of applying a quadratic penalty function to enforce smoothness on the reconstructed activity images as in [137], an anatomical penalty calculated with the attenuation map is employed to improve the image quality in this study. The main difference between these two approaches is that the first approach Φ_1 (Approach I) deforms the anatomical image (*i.e.* the attenuation map) to align it with the functional image, while the second approach Φ_2 (Approach II) deforms both images to align them with the measured data. As in [137], we use uniform cubic B-splines for image interpolation and deformation, which will be briefly reviewed first. A description of the uniform cubic B-splines is given in Appendix D. Despite the use

of an anatomical prior, Approach II is quite similar to the method introduced in [137] and Approach I is essentially a simplified algorithm to Approach II. We also modified Approach II to impose positivity on image values as opposed to B-spline coefficients.

5.3.1 Image interpolation and deformation based on cubic B-splines

Assume a continuous image function s can be represented as a linear combination of basis functions centred on a voxel grid $\mathsf{C} = \{\boldsymbol{r}_k,\ k = 1,\ldots,\ N\}$ that coincides with the voxel centres (see Figure 5.1 for an example in 2-D):

$$s(\boldsymbol{r}) = \sum_{k=1}^{N} s_k B(\frac{\boldsymbol{r}-\boldsymbol{r}_k}{\sigma_1}), \tag{5.3}$$

where s_k is the B-spline coefficient of the basis function centred on voxel k, $\boldsymbol{r} = (x,y,z)$ is the index vector in the 3-D Cartesian coordinate system, $B(\boldsymbol{r}) = b(x)b(y)b(z)$ is an interpolating function based on the cubic B-splines b and σ_1 represents the voxel-spacing. The discretised image can therefore be represented as a collection of the B-spline coefficients $\dot{\boldsymbol{s}} = (s_j)_{j=1}^{N}$. Note that the cubic B-splines are non-zero at several nodes. Therefore, the B-spline coefficients are not identical to the image values at the grid nodes. The deformation of $\dot{\boldsymbol{s}}$ is then achieved by deforming the continuous image function s followed by a re-sampling on C for every voxel j:

$$[\boldsymbol{W}\boldsymbol{s}]_j = \sum_{k=1}^{N} s_k B(\frac{v(\boldsymbol{r}_j)-\boldsymbol{r}_k}{\sigma_1}), \tag{5.4}$$

where $\boldsymbol{W}$ is a square matrix with each element $[\boldsymbol{W}]_{j,k} \triangleq B(\frac{v(\boldsymbol{r}_j)-\boldsymbol{r}_k}{\sigma_1})$ and v is the warping function. Given $\tilde{\mathsf{C}} = \{\tilde{\boldsymbol{r}}_l,\ l = 1,\ldots,\ Q\}$ a uniform sub-grid of C with Q grid nodes, the function v_θ can be parametrised by a collection of the deformation

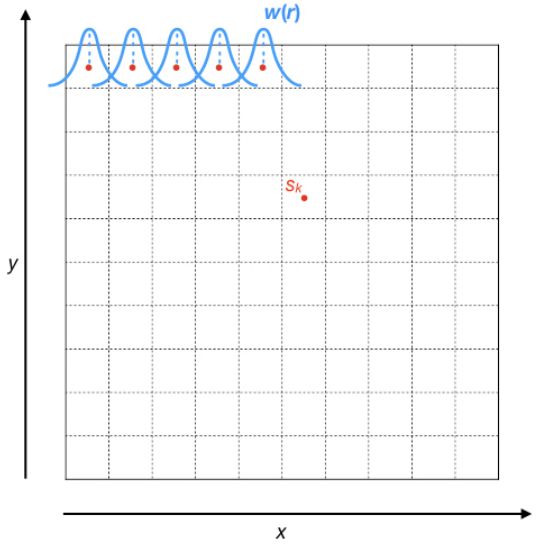

Figure 5.1: Example of representing a 2-D image function s using basis functions.

coefficients $\boldsymbol{\theta} = (\boldsymbol{\theta}^x,\ \boldsymbol{\theta}^y,\ \boldsymbol{\theta}^z)$:

$$v_\theta(\boldsymbol{r}) \triangleq \boldsymbol{r} + \begin{bmatrix} \sum_{l=1}^{Q} \theta_l^x B(\frac{\boldsymbol{r}-\tilde{\boldsymbol{r}}_l}{\sigma_2}) \\ \sum_{l=1}^{Q} \theta_l^y B(\frac{\boldsymbol{r}-\tilde{\boldsymbol{r}}_l}{\sigma_2}) \\ \sum_{l=1}^{Q} \theta_l^z B(\frac{\boldsymbol{r}-\tilde{\boldsymbol{r}}_l}{\sigma_2}) \end{bmatrix} = \begin{bmatrix} v^x(\boldsymbol{r}) \\ v^y(\boldsymbol{r}) \\ v^z(\boldsymbol{r}) \end{bmatrix} \tag{5.5}$$

where σ_2 is the distance between two grid nodes.

5.3.2 Approach I

The first approach optimises an objective function Φ_1 that considers the deformed anatomical image. Assume the attenuation map $\boldsymbol{\mu}$, represented as a collection of the B-spline coefficients $\dot{\boldsymbol{\mu}} = (\mu_j)_{j=1}^N$, is used to provide anatomical information as well, Φ_1 is given by:

$$\Phi_1(\boldsymbol{f}, \dot{\boldsymbol{\mu}}, \boldsymbol{\theta}) = -L(\boldsymbol{f}, \boldsymbol{g}, \boldsymbol{W}\dot{\boldsymbol{\mu}}) + \beta R(\boldsymbol{f}|\boldsymbol{W}\dot{\boldsymbol{\mu}}) + \gamma QP(\boldsymbol{\theta}), \tag{5.6}$$

where $QP(\boldsymbol{\theta})$ is a quadratic penalty on the deformation coefficients for reducing the influence of noise and γ is a constant that controls its strength. Note that the

misalignment between $\boldsymbol{\mu}$ and $\boldsymbol{f}$ affects the optimisation through the attenuation correction in the log-likelihood L and the incorporated anatomical penalty R as both functions use the warped attenuation map $\boldsymbol{W}\dot{\boldsymbol{\mu}}$ as inputs. Since this approach does not require deforming the activity image, $\boldsymbol{f}$ in (5.6) represents a vector of image values. The positivity constraint on $\boldsymbol{f}$ can therefore be achieved by performing constrained image reconstruction. In contrast, the attenuation map is represented as a collection of coefficients $\dot{\boldsymbol{\mu}}$ for image warping using B-splines. This could lead to negative values in $\boldsymbol{W}\dot{\boldsymbol{\mu}}$ as we are optimising the B-spline coefficients. However, since small negative values in $\boldsymbol{W}\dot{\boldsymbol{\mu}}$ would become attenuation factors very close to one, they have been left unchanged in this study. Although the deformation of the attenuation map implies the change of the scatter distribution, the estimated background events are fixed during the optimisation process for simplicity. The warped attenuation map should be in the same space as the activity image when the objective function in (5.6) is optimised.

5.3.3 Approach II

Instead of seeking to align the warped attenuation map with the reconstructed activity image, the second approach deforms both the anatomical and functional images in order to obtain an estimate that optimises the objective function Φ_2:

$$\Phi_2(\dot{\boldsymbol{f}}, \dot{\boldsymbol{\mu}}, \boldsymbol{\theta}) = -L(\boldsymbol{W}\dot{\boldsymbol{f}}, \boldsymbol{g}, \boldsymbol{W}\dot{\boldsymbol{\mu}}) + \beta R(\boldsymbol{f}|\boldsymbol{\mu}) + \gamma QP(\boldsymbol{\theta}) + \delta E(\boldsymbol{W}\dot{\boldsymbol{f}}), \tag{5.7}$$

where

$$E(\boldsymbol{s}) = \sum_{n=1}^{N'} \min(0,\ s(v_\theta(\hat{\boldsymbol{r}}_n)))^2 \tag{5.8}$$

is a barrier function that penalises negative values in $\boldsymbol{W}\dot{\boldsymbol{f}}$ [140]. Given $\hat{\mathsf{C}} = \{\hat{\boldsymbol{r}}_n,\ n = 1, \ldots,\ N'\}$ a finer grid that contains a finite number of uniformly spaced locations in each interval of two grid nodes in the voxel grid C, the function computes the spline values (*i.e.*, the image values) centred on the finer grid and penalises the square of any negatives. The strength of E is determined by the parameter δ. In this study, we defined the distance between two adjacent locations in $\hat{\mathsf{C}}$ equal to one quarter of the grid spacing used for the image. Although the log-likelihood requires the

warped emission and anatomical images as inputs, the penalty function is calculated with the image value of the non-warped ones. Since the motivation of using an anatomical prior is to encourage edges in the emission image corresponding to those in the anatomical one, finding common edges in the warped or non-warped images is essentially a similar optimisation problem. In other words, calculating the anatomical penalty function with $\boldsymbol{W}\dot{\boldsymbol{f}}$ and $\boldsymbol{W}\dot{\boldsymbol{\mu}}$ should lead to the same solution. When the objective function in (5.7) is optimised, both the emission and anatomical images will be aligned to a virtual space where the corresponding estimated data is very close to the measured one.

5.4 Algorithm implementation

The optimisation of both approaches is implemented as an alternating process that includes a misalignment estimation subroutine and a penalised image reconstruction subroutine. Pseudo-code that summarises the implementation can be found in Algorithm 3. The workflow is defined by the number of inner iterations (InnerIter1 and InnerIter2 in Algorithm 3) for these two subroutines and the number of outer iterations (OuterIter in Algorithm 3) that controls the repetition of the alternating process. We applied L-BFGS for unconstrained optimisation (misalignment estimation in both approaches and image reconstruction in the second approach) and L-BFGS-B [74] for the positivity constrained image reconstruction in Approach I (see Section 3.2 for more information regarding the optimisation algorithms). To improve the convergence rate of the penalised image reconstruction, a preconditioner proposed in (4.23) was also incorporated into both approaches [113]. OS-EM is used to reconstruct the initial activity image $\boldsymbol{f}_0$ of the whole process. The misalignment estimation is then initialised by $\boldsymbol{f}_0$ and the attenuation map $\boldsymbol{\mu}$. The implementation of the misalignment estimation employed in this study was originally proposed in [137]. Every time the misalignment estimation is done, a new initial image $\boldsymbol{f}_{\text{InitInner}}$ for the penalised image reconstruction is recomputed using OS-EM, taking into account the current estimated misalignment. The spatially variant penalty strength $\boldsymbol{\kappa}$ is recomputed at every outer iteration as well.

Algorithm 3: Pseudo-code for the alternating process

Input: Data $\boldsymbol{g}$, attenuation map $\boldsymbol{\mu}$ and strength of each penalty function (*i.e.*, the set of parameters (β,γ) for Approach I and (β,γ,δ) for Approach II

Output: Estimated tracer distribution $\boldsymbol{f}$ and B-spline deformation coefficient $\boldsymbol{\theta}$

$\Phi \leftarrow \Phi_1(\beta,\gamma)$ or $\Phi_2(\beta,\gamma,\delta)$;
$\nabla\Phi \leftarrow \nabla\Phi$;
$\boldsymbol{\theta}_0 \leftarrow \mathbf{0}$;
$\boldsymbol{f}_0 \leftarrow$ OS-EM$(\boldsymbol{g},\boldsymbol{\mu},\boldsymbol{\theta}_0)$;
for $t = 0,\ldots,$ OuterIter -1 **do**
  misalignment estimation;
  $\boldsymbol{\theta}_{t+1} \leftarrow$ L-BFGS$(\boldsymbol{g},\boldsymbol{f}_t,\boldsymbol{\mu},\Phi,\nabla\Phi,\boldsymbol{\theta}_t,$ InnerIter1);
  image reconstruction;
  $\boldsymbol{f}_{\text{InitInner}} \leftarrow$ OS-EM$(\boldsymbol{g},\boldsymbol{\mu},\boldsymbol{\theta}_{t+1})$;
  $\boldsymbol{\kappa} \leftarrow [\boldsymbol{A}^\top \text{diag}\{\frac{\boldsymbol{g}}{(\boldsymbol{A}\boldsymbol{f}_{\text{InitInner}}+\boldsymbol{n})^2}\}\boldsymbol{A}\mathbf{1}]^{\frac{1}{2}}$;
  $\boldsymbol{D} \leftarrow \text{diag}\{\boldsymbol{\kappa}^2 + 10^{-4}\}^{\frac{1}{2}}$;
  $\boldsymbol{f} \leftarrow \boldsymbol{D}\boldsymbol{f}_{\text{InitInner}}$;
  Define $\tilde{\Phi}$: $\boldsymbol{u} \mapsto \Phi(\boldsymbol{D}^{-1}\boldsymbol{u})$;
  Define $\nabla\tilde{\Phi}$: $\boldsymbol{u} \mapsto \boldsymbol{D}^{-1}\nabla\Phi(\boldsymbol{D}^{-1}\boldsymbol{u})$;
  if Approach I **then**
    $\boldsymbol{f}_{\text{InnerIter2}} \leftarrow$ L-BFGS-B$(\boldsymbol{g},\boldsymbol{f},\tilde{\Phi},\nabla\tilde{\Phi},\boldsymbol{\theta}_{t+1},$ InnerIter2) ;
  else
    $\boldsymbol{f}_{\text{InnerIter2}} \leftarrow$ L-BFGS$(\boldsymbol{g},\boldsymbol{f},\tilde{\Phi},\nabla\tilde{\Phi},\boldsymbol{\theta}_{t+1},$ InnerIter2) ;
  end
  $\boldsymbol{f}_{t+1} \leftarrow \boldsymbol{D}^{-1}\boldsymbol{f}_{\text{InnerIter2}}$;
end

5.5 Data

Two XCAT phantoms representing different respiratory phases and the corresponding $\boldsymbol{\mu}$ maps (Figure 5.2) were produced to simulate different PET positions. Both phantoms were a $128 \times 128 \times 47$ matrix with voxel size of 3.906 mm in all directions. The set of images at end inspiration was used to generate data similar to that from a GE Discovery STE in 3-D non-TOF acquisition mode. The number of projection angles has been down sampled from 280 to 140 to accelerate the computation. Both approaches were initially evaluated with a noiseless dataset. To further assess their performance in the presence of noise, a dataset with total counts of 161 M was also simulated using the Poisson noise model. Note that all simulations took into account the attenuation effect and system blurring using FWHM = 5.2 mm in tangential and radial directions and 5.7 mm in trans-axial direction. To simulate the misalignment between functional and anatomical images, the attenuation map

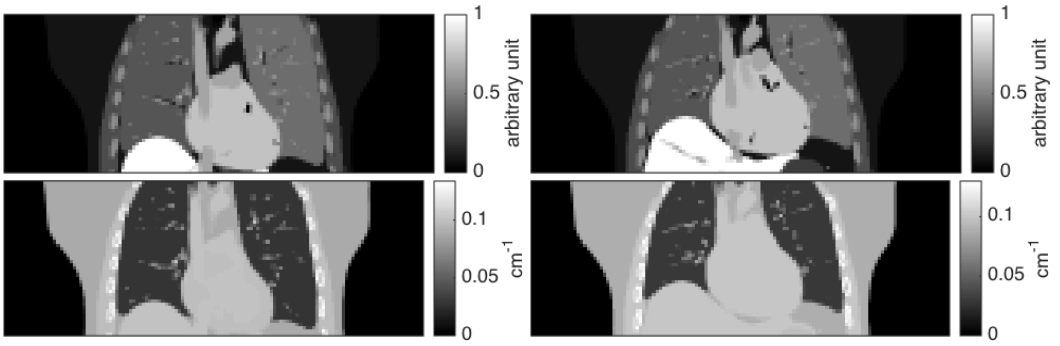

Figure 5.2: The central coronal view of the XCAT phantoms representing end inspiration (top left) and expiration (top right). The corresponding $\boldsymbol{\mu}$ maps are also provided (bottom).

at end expiration was used as the initial input for both attenuation correction and misalignment estimation.

5.6 Evaluation

The general reconstruction conditions are given in this section. Both approaches should be able to find a warped attenuation map similar to that used for data generation (*i.e.*, $\boldsymbol{\mu}$ at end inspiration). Therefore, the evaluation is focused on the performance of the misalignment subroutine of each approach.

5.6.1 Reconstruction conditions

The selection of each penalty strength was based on an initial investigation, where the difference between the warped and target $\boldsymbol{\mu}$ maps at OuterIter = 30 was visually compared with respect to a given set of candidate strengths (not shown). We studied the strength of one penalty at a time using the noiseless dataset and the value from the set giving the best visual alignment was recorded. Table 5.1 lists the candidate strengths for each penalty and approach in this study. As a result, for the noiseless dataset, the set of parameters that determines the strength of each penalty function was $(\beta,\gamma) = (7\times10^{-3}, 10^{-4})$ for Φ_1 and $(\beta,\gamma,\delta) = (7\times10^{-3}, 10^{-4}, 10^{-1})$ for Φ_2. A stronger $\beta = 2\times10^{-1}$, roughly equivalent to β for the noiseless dataset times the change of the image scale (30 times larger), was chosen when the data with noise was considered for both approaches. The strength of other penalty function(s) remained the same as for the noiseless dataset. The parameter pair (ε, η) in PLS was

Table 5.1: Candidate strengths for each penalty function and approach.

	β	γ	δ
Approach I	$[7\times10^{-4}, 7\times10^{-3}, 7\times10^{-2}]$	$[10^{-2}, 10^{-4}, 10^{-6}]$	N/A
Approach II	$[7\times10^{-4}, 7\times10^{-3}, 7\times10^{-2}]$	$[10^{-2}, 10^{-4}, 10^{-6}]$	$[10^{-1}, 10^{-2}, 10^{-3}]$

Figure 5.3: The central coronal view of the difference image between the input and target $\boldsymbol{\mu}$ maps.

fixed at (10^{-1}, 10^{-2}). The distance between two grid nodes for the deformation model was 6 voxels. The alternating process as well as the image reconstruction subroutine at every outer iteration were initialised by one full iteration of OS-EM with 14 subsets if not explicitly stated otherwise. Up to 100 outer iterations were used for both approaches. Each reconstructed activity image and the warped attenuation map had $128 \times 128 \times 47$ voxels with voxel size of $4.687 \times 4.687 \times 3.27$ mm^3.

5.6.2 Analysis methods

The difference image between the warped and target $\boldsymbol{\mu}$ maps at a given outer iteration was used to evaluate the performance of the misalignment subroutine. To be able to quantify the performance, the root-mean-square errors (RMSE) was calculated with the difference images and plotted against the outer iteration numbers. As the lungs are the main target of the respiratory motion alignment, a mask was applied to the difference images and the RMSE was computed only in the lungs as well to reflect the misalignment estimation of small structures of the lungs. Figure 5.3 shows the central coronal view of the difference image between the target and input $\boldsymbol{\mu}$ maps to give an idea regarding the simulated misalignment at OuterIter = 0.

5.7 Initial investigation

The alternating optimisation process has many parameters which must be specified. To find reasonable settings for the parameters in both approaches, their performance of misalignment estimation was initially evaluated using the noiseless dataset. We started with finding the workflow that provides satisfactory results for both approaches such that the performance evaluation was conducted at a given outer iteration with the same number of iterations for each subroutine. The influence of incorporating the anatomical prior, PLS, on misalignment estimation was studied by substituting the $\boldsymbol{\mu}$ map at end expiration with a uniform image when calculating PLS, therefore it is equivalent to using a (smooth) TV prior. To investigate if taking into account the anatomical prior in the initialisation of the whole process is beneficial, the optimisation process was also initialised from one full iteration of OS-EM with 14 subsets, followed by different iterations of L-BFGS-B-PC. The anatomical prior in L-BFGS-B-PC was again calculated with the $\boldsymbol{\mu}$ map at end expiration.

5.7.1 Workflow optimisation

A two-part study was conducted to find the workflow in common for both approaches. In the first part of the study, we used 1 inner iteration for the misalignment estimation and explored the minimum iterations required for the image reconstruction subroutine to obtain satisfactory results. The studied inner iteration numbers for the image reconstruction subroutine were 1, 5, 10 and 20 (InnerIter2 = 1, 5, 10 and 20). We then fixed the iteration number for the image reconstruction to the limit found in the first part and increased the number of iterations for the misalignment estimation from 1 to 5, 15 or 30 (InnerIter1 = 1, 5, 15 and 30) to assess if the performance of the misalignment estimation can be improved by using a higher inner iteration number. The number of outer iterations that controls the repetition of the alternating process was fixed at 100 (OuterIter = 100) such that the workflow was determined by the selected inner iteration number for each subroutine to reduce the number of parameters to investigate.

For the first sub-study where the inner iteration number for the misalignment

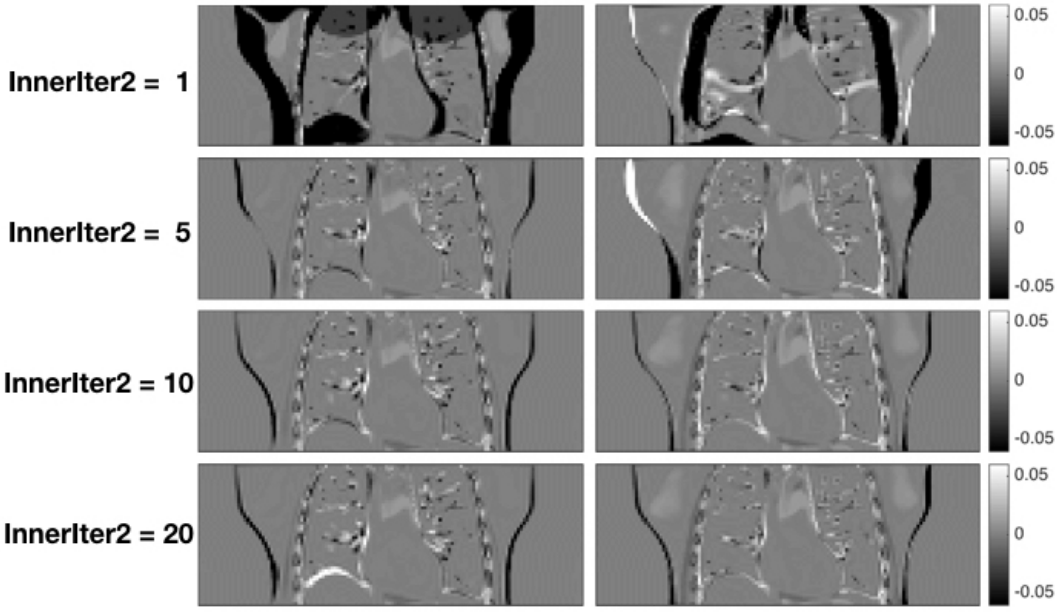

Figure 5.4: The central coronal view of the difference images between the target and $\boldsymbol{W\mu}$ maps for Approach I (left column) and (right column) Approach II at 100 outer iteration. The applied workflows were 1 inner iteration for the misalignment estimation and 1 (top row), 5 (second row), 10 (third row) and 20 (bottom row) inner iterations for the image reconstruction.

estimation was fixed at 1, the central coronal view of the difference images between the warped and target $\boldsymbol{\mu}$ maps at OuterIter = 100 are shown in Figure 5.4. For both approaches, using InnerIter2 = 1 was problematic, resulting in severe distortion of structures in the warped $\boldsymbol{\mu}$ map (Figure 5.4, top row). Satisfactory results were obtained with Approach I when InnerIter2 = 5 or 10 was chosen (Figure 5.4, left column, second and third rows). However, when a higher InnerIter2 = 20 was applied, the misalignment around the diaphragm (Figure 5.4, left column, bottom row) was still observed after 100 outer iterations. In contrast to Approach I, the performance of the misalignment estimation of Approach II was improved as InnerIter2 increased (Figure 5.4, right column). When InnerIter2 $\geq$ 10 was used, the algorithm was able to realign the input $\boldsymbol{\mu}$ map to the target one at OuterIter = 100.

Interestingly, some observations based on the overall RMSE contradicted the visual comparison (Figure 5.5, top row). For example, the plots of the overall RMSE for Approach I with InnerIter2 = 10 and 20 seemed to have the same value at OuterIter = 100 (Figure 5.5, top left). In addition, the lowest overall RMSE at

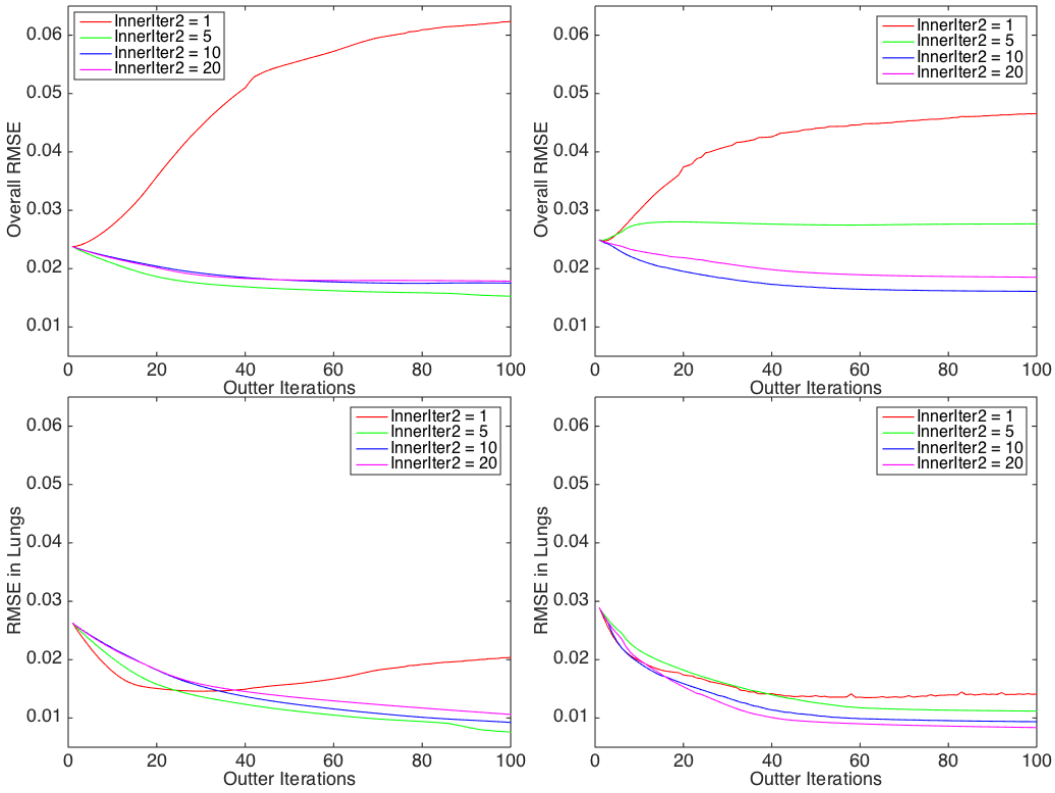

Figure 5.5: The overall RMSE (top row) and the RMSE in the lungs (bottom row) plotted against the outer iteration numbers for Approach I (first column) and II (second column). The applied workflows were 1 inner iteration for the misalignment estimation and 1, 5, 10 or 20 inner iterations for the image reconstruction.

OuterIter = 100 was achieved by the workflow with InnerIter2 = 5 and 10 for Approach I and II, respectively (Figure 5.5, top row). The RMSE in the lungs was more consistent with the visual observation from the difference images. As shown in the bottom row of Figure 5.5, for both approaches, the RMSE in the lungs tended to reach a higher error when InnerIter2 = 1 was used. The value at OuterIter = 100 decreased as InnerIter2 increased, except for the workflow with InnerIter2 = 20 for Approach I. In terms of the convergence rate, except for the workflow with InnerIter2 = 1, Approach II was able to reach a relatively stable overall RMSE and RMSE in the lungs after 60 outer iterations. Although Approach I showed the ability to achieve a stable overall RMSE after 60 outer iterations for workflow with InnerIter2 $>$ 1 as well, it required more than 100 outer iterations to reach a stable RMSE in the lungs for all evaluated workflows with different numbers of InnerIter2.

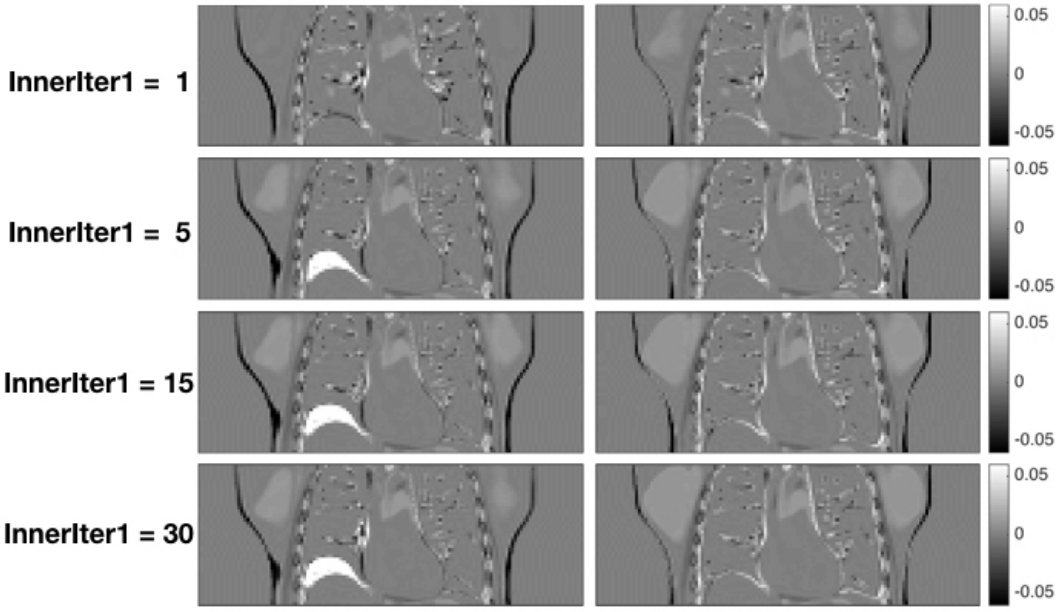

Figure 5.6: The central coronal view of the difference images between the target and $\boldsymbol{W\dot{\mu}}$ maps for Approach I (left column) and (right column) Approach II after 100 outer iterations. The applied workflows were 10 inner iteration for the image reconstruction and 1 (top row), 5 (second row), 15 (third row) and 30 (bottom row) inner iterations for the misalignment estimation.

Since both approaches provided visually and numerically good results when InnerIter2 = 10 was used, we defined InnerIter2 = 10 as the best number of iterations for the image reconstruction subroutine and increased the iteration number used in the misalignment estimation subroutine. Figure 5.6 shows the central coronal view of the difference image between the warped and target $\boldsymbol{\mu}$ maps for both approaches with various InnerIter1 but a fixed InnerIter2 = 10. The misalignment around the diaphragm region became apparent for Approach I at OuterIter = 100 as InnerIter1 > 1 was chosen (Figure 5.6, left column). For Approach II, however, the difference images at OuterIter = 100 were visually identical, regardless of the number of the applied InnerIter1 (Figure 5.6, right column).

The RMSE in the lungs for Approach I and II with different InnerIter1 plotted against the outer iteration numbers are given in Figure 5.7. We omitted evaluations based on the overall RMSE in the rest of the chapter as we were more interested in the alignment of small structures of the lungs for respiratory motion estimation. Consistent with the visual comparison, for the first approach, the smallest RMSE

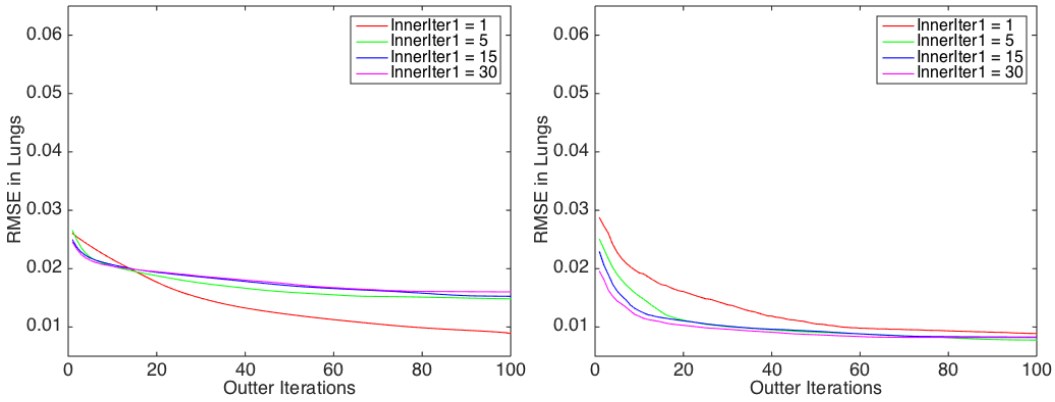

Figure 5.7: The RMSE in the lungs plotted against the outer iteration numbers for Approach I (left) and II (right). The applied workflows were 10 inner iteration for the image reconstruction and 1, 5, 15 or 30 inner iterations for the misalignment estimation.

in the lungs at OuterIter = 100 was achieved by the workflow with InnerIter1 = 1. When Approach II was adopted, all workflows with different InnerIter1 settings were able to achieve a similar RMSE in the lungs at OuterIter = 100. The convergence rate of the misalignment estimation of Approach II was improved as InnerIter1 increased. However, the performance of Approach I in terms of the convergence rate of the RMSE in the lungs seemed insensitive to the change of the inner iteration number for the misalignment estimation when the applied InnerIter1 was larger than 1.

Based on the results shown in this section, 1 iteration of misalignment estimation (InnerIter1 = 1), followed by 10 iterations of image reconstruction (InnerIter2 = 10) was defined as the workflow that provides satisfactory results for these two approaches when the noiseless dataset is considered. The corresponding central coronal view of the reconstructed functional images and the warped attenuation maps at OuterIter = 100 are provided in the top rows of Figure 5.8 and 5.9, respectively. The optimisation with the same settings but using the target $\boldsymbol{\mu}$ map as the input was also performed for both approaches. The bottom rows of Figure 5.8 and 5.9 shows the reconstructed images and the deformed attenuation maps for the aligned functional and anatomical information. Due to the implied interpolation of image warping and re-sampling using B-splines, all images in Figure 5.8 were visually similar to the target activity image but somewhat smoother. Regardless of the

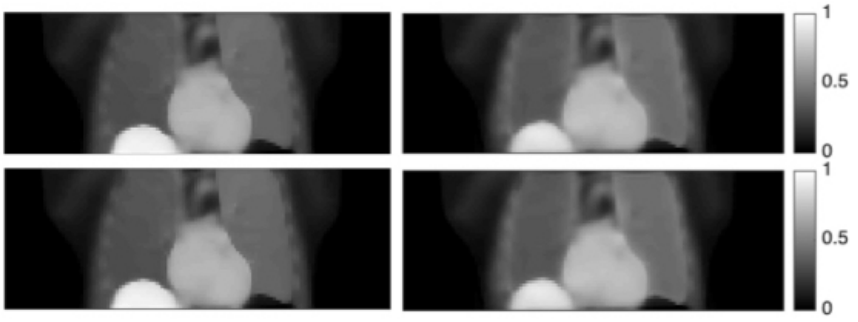

Figure 5.8: The central coronal view of the activity images at 100 outer iteration for Approach I (left column) and II (right column) with (top row) and without (bottom row) using the misaligned anatomical information. The applied workflow was 1 inner iteration for the misalignment estimation and 10 inner iterations for the image reconstruction.

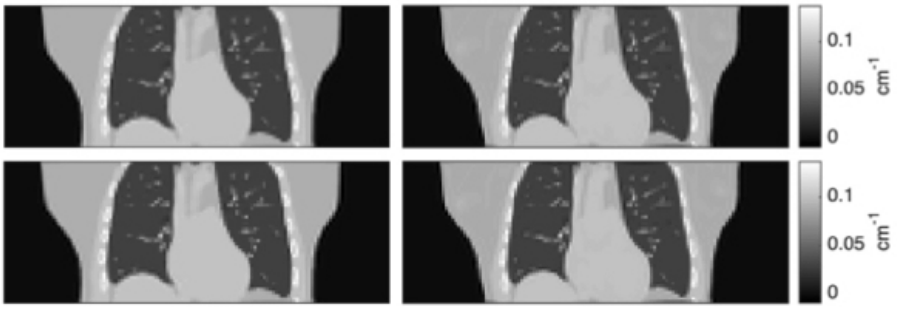

Figure 5.9: The central coronal view of the warped attenuation maps at 100 outer iteration for Approach I (left column) and II (right column) with (top row) and without (bottom row) using the misaligned anatomical information. The applied workflow was 1 inner iteration for the misalignment estimation and 10 inner iterations for the image reconstruction.

applied approach and the alignment of the functional and anatomical images, the warped attenuation maps shown in Figure 5.9 were visually identical and very close to the target one. As an initial investigation, the applied outer iteration number was fixed at 100 throughout this chapter, although it was not sufficient for Approach I to achieve a stable RMSE in the lungs.

5.7.2 Influence of incorporating anatomical priors on misalignment estimation

To investigate whether the use of an anatomical prior is beneficial for either approaches, reconstructions without considering any structural information were performed as well. This was achieved by substituting the anatomical image (*i.e.*, atten-

uation map) with a uniform image when calculating the penalty function. Note that the misaligned $\boldsymbol{\mu}$ map was still used for the attenuation correction. The performance of both approaches with and without considering anatomical information was evaluated using the difference image between the warped and target $\boldsymbol{\mu}$ maps and the RMSE in the lungs. The workflow found in the previous section was applied in all reconstructions.

As observed in the central coronal view of the difference images at OuterIter = 100 (Figure 5.10 and 5.11, bottom row), both Approach I and II were able to estimate the misalignment and warp the input attenuation map accordingly, regardless of the presence of the structural information. However, in terms of the convergence rate, these two approaches had different responses to the use of the anatomical information. For the first approach, incorporating the structural information degraded the convergence rate of the misalignment estimation. The central coronal view of the difference images at OuterIter = 20 and 60 for the reconstructions without using anatomical information showed less apparent misalignment around the diaphragm region compared to those for the reconstructions considering the anatomical information (Figure 5.10, top and second rows). By contrast, Approach II was able to achieve a faster convergence rate when the additional structural information was available (Figure 5.11, top and second rows). These observations were further demonstrated by the RMSE in the lungs plotted against the outer iteration numbers (Figure 5.12). When the anatomical information was considered, Approach II reached a stable RMSE in the lungs after around 60 outer iterations, while Approach I required more than 80 outer iterations to achieve that.

5.7.3 Influence of incorporating anatomical priors in initialisation on misalignment estimation

For both Approach I and II, the alternating process was initialised from one full iteration of OS-EM with 14 subsets, which did not take into account the anatomical information. As we have demonstrated that the presence of the anatomical details can affect the performance, we then substituted the initialisation of the whole process with one iteration of OS-EM (with 14 subsets), followed by 5 or 10 iterations

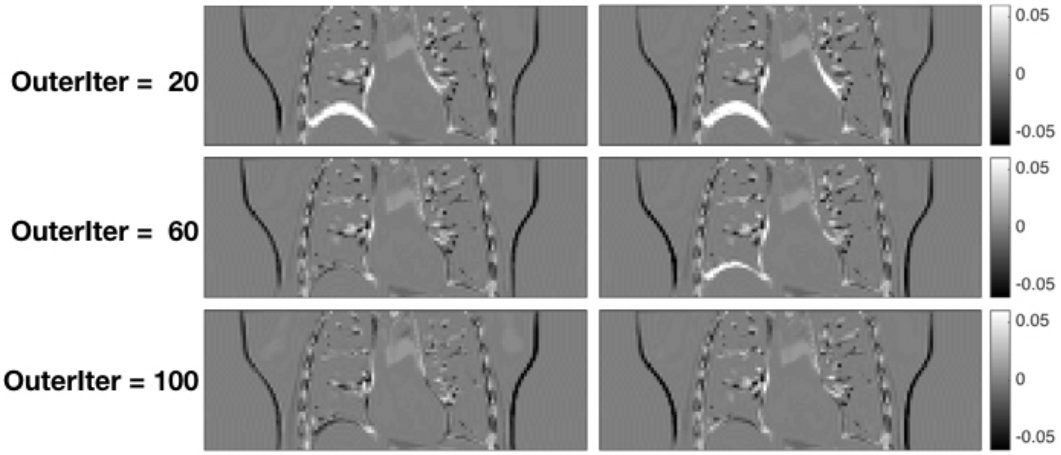

Figure 5.10: The central coronal view of the difference images between the target and $\boldsymbol{W\dot{\mu}}$ maps for Approach I at 20 (top row), 60 (second row) and 100 (bottom row) outer iteration. The results for the reconstructions without and with considering the anatomical information are shown in the left and right column, respectively.

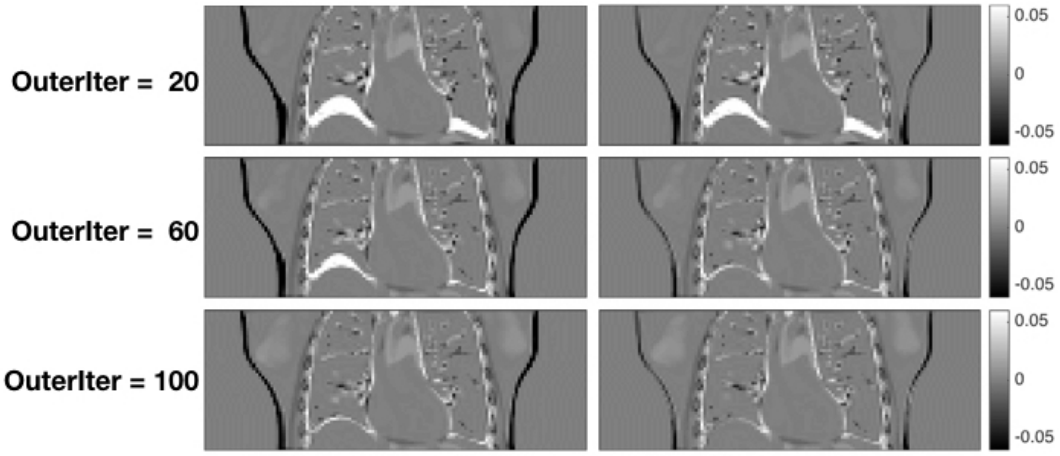

Figure 5.11: The central coronal view of the difference images between the target and $\boldsymbol{W\dot{\mu}}$ maps for Approach II at 20 (top row), 60 (second row) and 100 (bottom row) outer iteration. The results for the reconstructions without and with considering the anatomical information are shown in the left and right column, respectively.

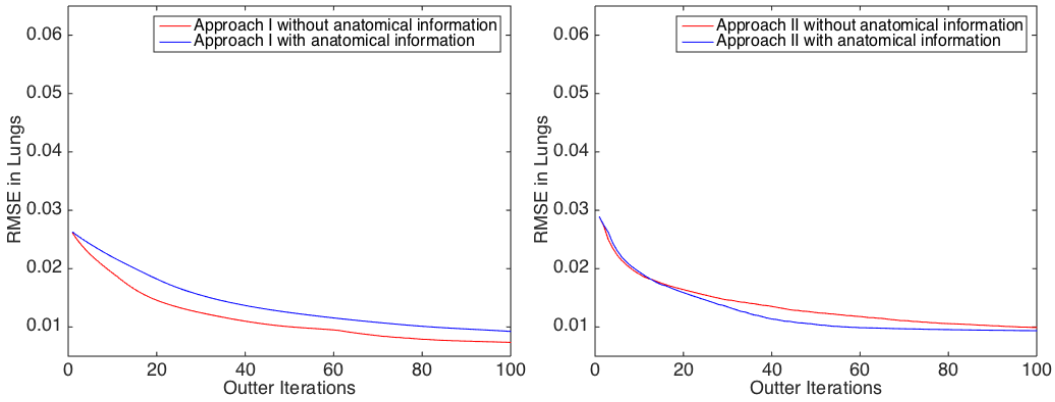

Figure 5.12: The RMSE in the lungs plotted against the outer iteration numbers for Approach I (left) and II (right) without and with considering the anatomical information.

of L-BFGS-B-PC to introduce the structural information into the initial activity image. After the initialisation, the workflow from Section 5.7.1 (1 iteration for the misalignment estimation subroutine and 10 iterations for the penalised image reconstruction subroutine) was applied to the alternating process. Note that the image reconstruction subroutine was still initialised from one iteration of OS-EM with 14 subsets at every outer iteration.

The sensitivity of both approaches to different initialisations was evaluated using the difference image between the warped and target $\boldsymbol{\mu}$ maps (Figure 5.13) and the RMSE in the lungs plotted against the outer iteration numbers (Figure 5.14). For each approach, the difference images at OuterIter = 100 were visually identical and the plots of RMSE in the lungs were almost on top of each other for different initial conditions. Based on these results, we conclude that both Approach I and II were not sensitive to the initial activity image for the alternating process when the above workflow was applied.

5.8 Preliminary investigation on noisy data

The workflow giving satisfactory results for both approaches found with the noiseless dataset was further investigated on the noisy dataset. Recall that a stronger strength $\beta = 2 \times 10^{-1}$ was used for the anatomical penalty function in the presence of noise. As the convergence rate of L-BFGS-B-PC varies with the strength of

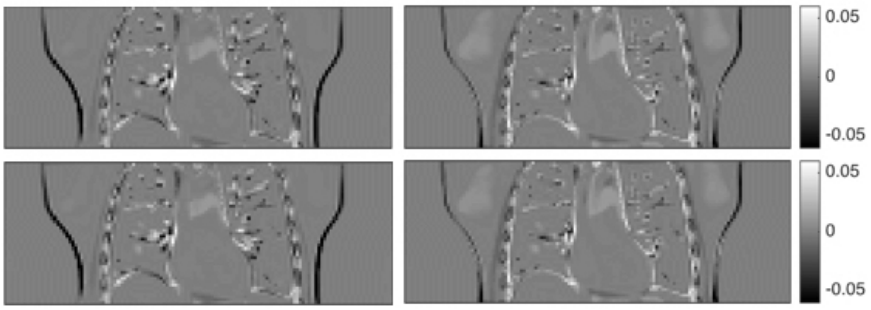

Figure 5.13: The central coronal view of the difference images between the target and $\boldsymbol{W\dot{\mu}}$ maps for Approach I (left) and II (right) at 100 outer iteration. The results for the alternating process initialising from 1 full iteration of OS-EM, followed by 5 or 10 iterations of L-BFGS-B-PC are shown in the top and bottom rows, respectively.

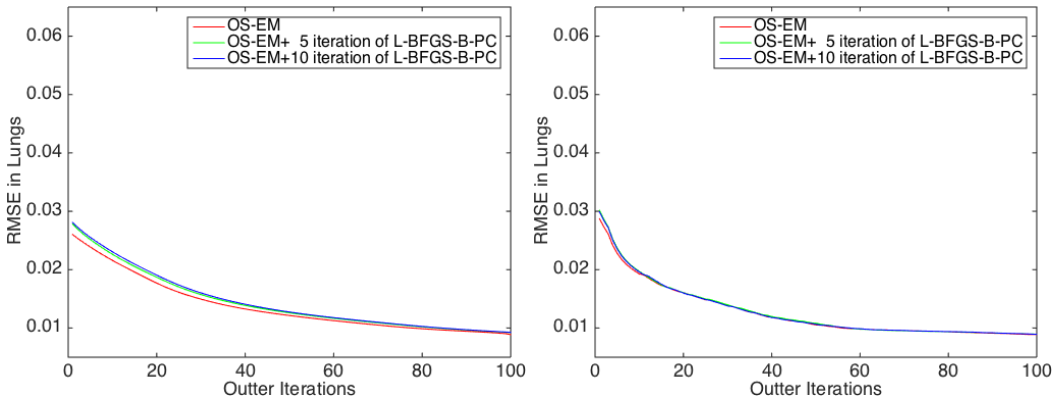

Figure 5.14: The RMSE in the lungs plotted against the outer iteration numbers for Approach I (left) and II (right) with different initialisations.

the penalty function and data noise level (see Section 3.7.3.1), the alternating process was also performed with a higher InnerIter2 = 20 or 40 for both approaches. The number of inner iterations for the misalignment subroutine was kept the same (InnerIter1 = 1) as well as the number of outer iterations (OuterIter = 100).

Figure 5.15 shows the central coronal view of the difference image between the warped and target $\boldsymbol{\mu}$ maps for each reconstruction condition at OuterIter = 100. As observed in the figure, the first approach still suffered from the misalignment issue at 100 outer iteration when InnerIter2 = 10 or 20 were applied (Figure 5.15, left column, top and second images). Consistent with the results for the noiseless dataset (Figure 5.4), the performance of the misalignment estimation of the second approach was less sensitive to the applied number of InnerIter2. Note that the

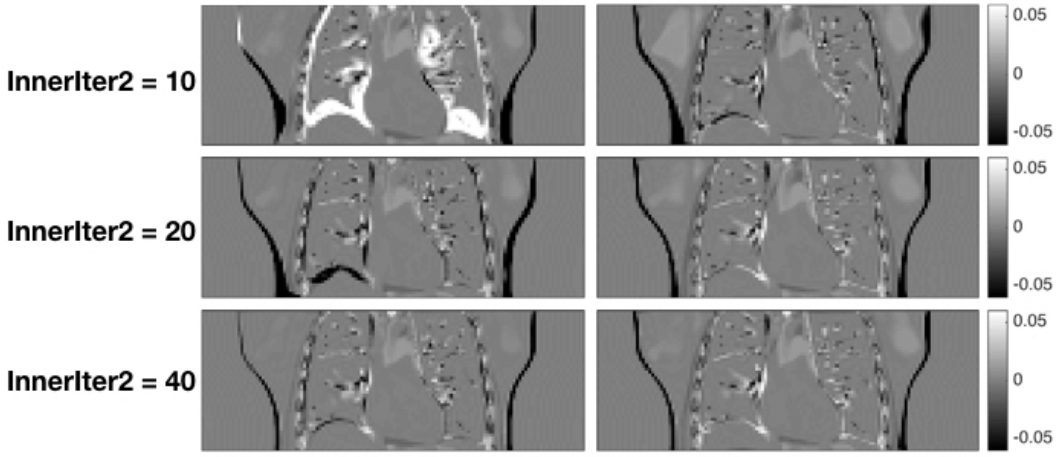

Figure 5.15: The central coronal view of the difference images between the target and $\boldsymbol{W}\dot{\boldsymbol{\mu}}$ maps for Approach I (left column) and (right column) Approach II at 100 outer iteration. The applied workflows were 1 inner iteration for the misalignment estimation and 10 (top row), 20 (middle row), 40 (bottom row) inner iterations for the image reconstruction.

alignment of the contour of the object was improved as InnerIter2 increased for both approaches. The RMSE in the lungs plotted against the outer iteration numbers support our observations from the difference images (Figure 5.16). Moreover, Approach II converged faster than Approach I for all evaluated workflows. The corresponding reconstructed functional images and the warped attenuation maps for InnerIter2 = 40 were also provided in Figure 5.17. As shown in the top row, the reconstructed activity images are smoother than the target one due to the implied interpolation and the use of a stronger penalty strength. These two approaches were able to achieve visually identical warped attenuation maps with the applied workflow. However, residual misalignment in the lungs was still observed for both approaches at OuterIter = 100 (Figure 5.17, bottom row).

5.9 Discussion

The potential misalignment between functional and anatomical images is the main concern for incorporating an anatomical prior into image reconstruction. Expanding on the algorithm proposed in [137], two approaches that perform alternating misalignment estimation and penalised image reconstruction using anatomical priors are proposed. In this study, we used the $\boldsymbol{\mu}$ map for attenuation correction and

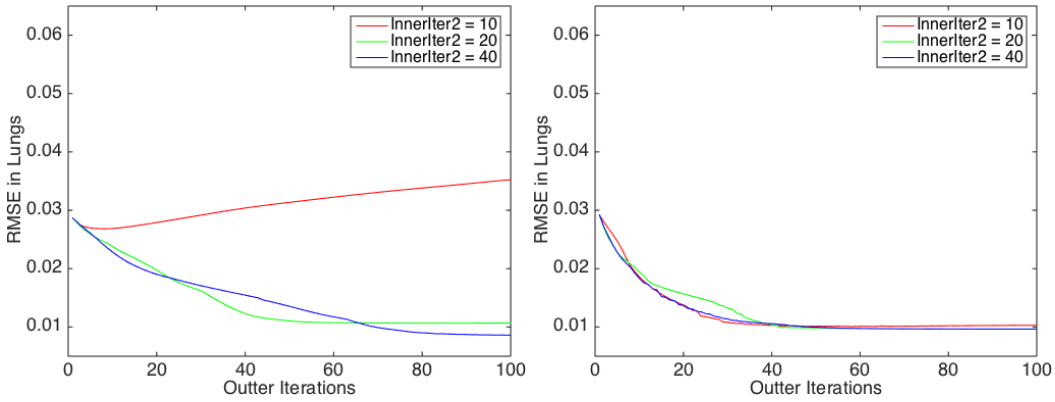

Figure 5.16: The RMSE in the lungs plotted against the outer iteration numbers for Approach I (left) and II (right). The applied workflows were 1 inner iteration for the misalignment estimation and 10, 20 or 40 inner iterations for the image reconstruction.

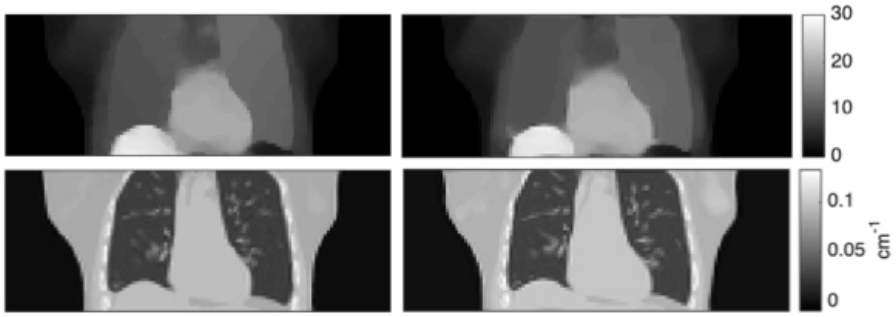

Figure 5.17: The central coronal view of the activity images for Approach I (left) and (right) Approach II at 100 outer iteration (top row). The corresponding warped attenuation maps are also provided (bottom row). The applied workflows were 1 inner iteration for the misalignment estimation and 40 inner iterations for the image reconstruction.

anatomical prior calculation. Therefore, the optimisation was influenced by the misalignment through the attenuation correction and the incorporated penalty function. We have demonstrated that both approaches are able to estimate the misalignment and warp the anatomical image accordingly, but with a different convergence rate, depending on the applied workflow and if the anatomical information is included as a prior. Since the deformation of the attenuation map will lead to a different scatter distribution, the estimated background events should be updated accordingly during the optimisation. However, the effect was assumed to be small and ignored for simplicity in current study. To achieve accurate quantification, performing active scatter correction based on the update of the estimated background events might be

necessary. In practice, this could be done by re-estimating the scatter after a number of iterations of the current algorithm.

The same as in [137], we have chosen to use the uniform cubic B-splines to model the image deformation. As no priori knowledge regarding the misalignment is required, it should be able to adapt to various misalignment scenarios caused by different types of motion effect. However, due to the implied interpolation, it also leads to a smooth deformation field which might not be able to accurately model sliding motion, such as the pleural sliding on the inside of the ribcage. To improve the accuracy of the misalignment estimation and make maximum use of the anatomical information, investigating other differentiable deformation models specific to different applications should be included in future work. However, as long as the deformation field would be parametrisable using a linear sum of basis-functions, the methods of this work would be applicable.

The evaluation in this study was focused on the performance of the misalignment estimation of each approach. The central coronal view of the difference image between the warped and target $\boldsymbol{\mu}$ maps at different outer iterations were used for visual comparison. For quantitative assessment, the overall RMSE and RMSE in the lungs were calculated with the difference images and plotted against the number of outer iterations. We found that some observations based on the overall RMSE did not agree with the visual comparison. This might be because the overall RMSE takes into account the residual misalignment around the object contour, which is affected by the applied number of inner iterations for the image reconstruction subroutine (Figure 5.4 and 5.15). As the alignment of small structures in the lungs is the main target for respiratory motion estimation, most results were compared in terms of RMSE in the lungs only. Since the objective functions of these two approaches are not comparable, seeking metrics that measure the "practical" convergence would be helpful in demonstrating the convergence of the algorithms.

As an initial investigation, the maximum outer iteration number that controls the repetition of the alternating process was set to 100 to reduce the number of parameters to optimise. Therefore, the workflow of both approaches was determined

by the number of iterations for the misalignment estimation (InnerIter1) and image reconstruction (InnerIter2) subroutines. To study how the change of the workflow influences the performance of the proposed approaches, initial evaluations on noiseless datasets were conducted in Section 5.7.1. As shown in the left column of Figure 5.4 and 5.6, satisfactory visual alignment was obtained at 100 outer iterations when InnerIter1 = 1 and InnerIter2 = 5 or 10 were chosen for the first approach, although the corresponding RMSE in the lungs indicated that a higher number of outer iterations was required for achieving a stable value (Figure 5.5 and 5.7, left). The use of a larger InnerIter1 or InnerIter2 slowed down convergence rate of the misalignment estimation. When a sufficient number of iterations for the image reconstruction subroutine was applied (InnerIter2 $\geq$ 10), Approach II showed the ability to achieve a good alignment in 60 outer iterations (Figure 5.5 and 5.7, right). Its performance in terms of convergence rate was improved as InnerIter1 or InnerIter2 increased. In general, Approach II was less sensitive to the change of workflow compared to Approach I. For a given workflow, Approach II outperformed Approach I as it required less outer iterations to achieve a reasonable result. In this study, the performance was evaluated at given outer iterations. Since one iteration of motion estimation is computationally more expensive than one iteration of image reconstruction, a workflow (1 iteration of misalignment estimation, followed by 10 iterations of image reconstruction) that gave good results for both approaches was sought to be able to compare their performance. Further improvement in convergence rate can be expected by optimising the workflow specifically for each approach. A measure that reflects the computational demand of each approach is therefore required.

The use of the anatomical penalty function improved the convergence rate of the misalignment estimation for Approach II but slowed it down for Approach I (Figure 5.10 and 5.11). To study the response of the proposed approaches to initialisations considering the anatomical information, the initial activity image for the whole alternating process was also reconstructed with one full iteration of OS-EM, followed by a few iterations of L-BFGS-B-PC with an anatomical prior, PLS. With

the workflow found in Section 5.7.1, both approaches showed similar performance to that using one iteration of OS-EM alone for initialisation (Figure 5.13 and 5.14). Given OS-EM is computationally efficient and easy to implement, one iteration of OS-EM with 14 subsets was therefore chosen as the initial condition for the alternating process in all other reconstructions in this study.

The workflow found in Section 5.7.1 with noiseless data was further investigated on one noisy dataset. Based on the result shown in Figure 5.15, both approaches required a higher number of iterations for the image reconstruction subroutine (InnerIter2 ≥ 40) to achieve good results in 100 outer iterations. Since the strength of the anatomical penalty function was increased in order to regularise noise as well, it is hard to attribute the cause to the presence of noise or the change of penalty strength. The observation also implies that the required number of InnerIter2 should be optimised according to these factors. Although further investigation on data representing different noise levels and reconstructed with different conditions is necessary, we found that using the workflow with InnerIter1 = 1 and InnerIter2 = 40 should be sufficient for all simulations applying either approach in this study. Consistent with results for the noiseless dataset, Approach II showed lower sensitivity to the change of workflow and outperformed Approach I in terms of the convergence rate of misalignment estimation (Figure 5.16). For all evaluated workflows with different numbers of InnerIter2, it was able to achieve a stable RMSE in the lungs in 60 outer iterations.

For both approaches, the strength of each penalty function was chosen based on the performance of the misalignment estimation. However, together with the implied interpolation of image warping and re-sampling using B-splines, the selected penalty strengths have led to over-smoothed functional images. In the current investigation, we focused on the visual and numerical differences between the warped and target $\boldsymbol{\mu}$ maps, evaluations on the reconstructed activity images were not included. The over-smoothing in the reconstructed activity image can be overcome by running a final image reconstruction with settings optimised for the estimation of the functional image. Note that in the case of a single gate, this can be done by

running the final reconstruction using the warped attenuation map at the last outer iteration with a proper penalty strength for both approaches. However, for multiple gated datasets, the warp matrix $\boldsymbol{W}$ should be incorporated into the final image reconstruction for Approach II.

In this study, the feasibility of applying the proposed approaches to estimate the misalignment in PML image reconstruction using anatomical priors was demonstrated with one gate of non-TOF PET data in which no motion was assumed. The motivation of incorporating an anatomical penalty is to improve the visual appearance and quantification of the reconstructed activity image by utilising the additional anatomical information. It should be able to introduce smoothness in uniform regions and sharpen edges observed in both functional and anatomical images. The more and stronger the common edges are, the sharper reconstructed images can be achieved. Therefore, applying anatomical priors to non-gated PET data is not beneficial as there will be substantial motion over the acquisition, leading to an overall smoothness across boundaries of the object. In the current study, we have adapted the strategy often used in CT or MR derived attenuation correction that down-samples the anatomical image to match the resolution of PET and surrenders the structural details carried by the high-resolution anatomical image. However, based on the evaluation shown in [141], a sharper reconstructed activity image can be expected by reconstructing the PET image at the same voxel size as the anatomical image instead and exploring the common edges in the resolution of the anatomical one. The benefit of using TOF data to misalignment estimation was studied in [139] where a similar algorithm was applied to obtain reconstructed activity images with aligned attenuation correction. As Approach I and II are extensions of that algorithm, practical convergence of misalignment estimation in less outer iterations can be expected when TOF data is available. Moreover, by exploring the similarity between gates and utilising the whole data statistics, the misalignment estimation of both approaches should also be able to benefit from multiple gates of data.

5.10 Conclusions

Two approaches for solving the potential misalignment between the functional and anatomical images in penalised image reconstruction using anatomical priors have been proposed in this study. The main difference between them is that the first approach deforms the anatomical image to align it with the functional one, while the second approach deforms both images to align them with the measured data. Both approaches were implemented using alternation between misalignment estimation and image reconstruction. The results demonstrated that both methods are able to estimate the misalignment and deform the anatomical image accordingly when a proper workflow for the alternating optimisation is applied. Moreover, the second approach shows the ability to converge to the correct alignment faster than the first approach and is less sensitive to variations in the workflow. These encouraging results indicate that it is possible to align functional and anatomical information, enabling the use of anatomical priors in practice.

Chapter 6

General Conclusions

With the increased interest in quantitative analysis in emission tomography (ET), reconstruction algorithms that provide both good image quality and accurate intensity values are desirable. Due to the ability of taking into account the statistical nature of the data, iterative algorithms have become the method of choice for the state-of-the-art imaging systems. However, as image properties such as resolution and noise are functions of iterations, the visual appearance and quantitative accuracy vary with iterations when the applied algorithm has not converged yet. In fact, the image properties also change between individuals and lesions with different local features. To ensure predictable run-time in a clinical environment, image reconstruction algorithms, independent of as many factors as possible, are therefore preferable. The applied algorithm should also show fast convergence rate to avoid compromising patient throughput. Since using anatomical priors to improve image quality and quantitative accuracy has shown promise in literature, the algorithm is expected to be capable of incorporating anatomical information as well.

The aim of this thesis was to improve the quantitative consistency of ET by investigating the use of advanced optimisation algorithms for penalised image reconstruction. The proposed algorithms have shown the potential to achieve practical feasibility in time as well as consistent performance in convergence rate and local response. The algorithms are also capable of incorporating anatomical information. Taking into account the potential misalignment between the activity and anatomical images, two approaches that perform alternating misalignment estimation and pe-

nalised image reconstruction were proposed. The evaluation of the proposed strategies and algorithms had been concentrated on PET data. The main conclusions and contributions of the study are highlighted in this chapter, followed by a discussion of potential future work. The publications that have arisen from this work and from collaborations with other researchers are also listed.

6.1 Summary of main conclusions

To achieve practical processing time, a fast convergent quasi-Newton algorithm, L-BFGS-B-PC, was proposed in Chapter 3. This algorithm was based on the standard L-BFGS-B algorithm but we incorporated a preconditioner specific to ET. Since the proposed preconditioner could be precomputed, the computational cost of L-BFGS-B-PC per iteration was similar to that of L-BFGS-B. The performance of L-BFGS-B-PC was evaluated using both simulations and three patient datasets. In addition to achieving over 100 times faster convergence rate than SPS, a previously published algorithm for penalised image reconstruction, its performance in terms of run-time was also less sensitive to various factors, such as the noise level, penalty type, penalty strength and background level, compared to the non-preconditioned L-BFGS-B. Although the current work concentrated on L-BFGS-B, the proposed strategy of introducing a preconditioner based on knowledge of the ET objective function could be applied to other optimisation algorithms as well.

To be able to improve the image quality and quantitative consistency, an anatomical penalty function was then incorporated with a spatially-variant penalty strength in Chapter 4. The weighting scheme was previously proposed for quadratic priors and then generalised the non-quadratic ones in the literature. In Chapter 4, the theory was further extended to a wider class of penalty functions such that it could be applied to any anatomical prior that utilises local properties within a pre-defined neighbourhood structure. In addition, we introduced a different type of weighting scheme which is suitable for 3-D PET reconstruction and can be easily precomputed. As the weighting was also reused to construct the preconditioner in L-BFGS-B-PC, the overall computational demand remained virtually unchanged.

The proposed weighting scheme was evaluated with an anatomical prior based on Parallel Level Sets (PLS). Since PLS had shown promising results in literature and its convexity that facilitates numerical optimisation is well-established, it was chosen as the representative anatomical penalty in this thesis. Based on results for various simulations and data from one patient with inserted pseudo lesions, the spatially-variant penalty reduced the quantitative dependence on the surrounding activity and location. Moreover, substantial improvement in convergence rate and its consistency was also observed.

As the use of anatomical penalties in practice could be limited by the potential misalignment between the functional and anatomical images, two approaches that perform alternating misalignment estimation and penalised image reconstruction using anatomical priors were proposed in Chapter 5. Both approaches were based on a joint motion estimation and image reconstruction method proposed recently for dealing with the mismatch between the attenuation map and the PET image in respiratory gated PET/CT. Expanding on our previous work, L-BFGS-B-PC using PLS with the spatially-variant penalty strength was used for the image reconstruction. In terms of the misalignment compensation, the first approach deformed the anatomical image to align it with the functional one while the second approach deforms both images to align them with the measured data. Preliminary results for non-TOF data simulations demonstrated that both methods were able to estimate the misalignment and deform the anatomical image accordingly when a proper workflow for the alternating optimisation was applied. The second approach also showed the ability to converge to the correct alignment faster than the first one, and with less dependence on the algorithm workflow. Both methods were further evaluated with a simulated noisy dataset. In the presence of noise, a higher iteration number for the image reconstruction subroutine was required to obtain satisfactory results.

6.2 Original contributions

The main contributions provided by the author among the research presented in this thesis are summarised below.

6.2.1 Algorithm development

- Derivation and evaluation of the precomputed preconditioner proposed for L-BFGS-B-PC.
- Evaluation and modification of the initial step length of the line search subroutine of L-BFGS-B.
- Evaluation of the performance of SPS, L-BFGS-B and L-BFGS-B-PC.
- Generalisation of the use of the spatially-variant penalty strength to a wider class of edge preserving penalty functions.
- Exploration of the improvement in algorithm convergence rate and convergence consistency when using the spatially-variant penalty strength.
- Development of the two approaches for alternating misalignment estimation and penalised image reconstruction including anatomical priors based on a previously proposed joint reconstruction algorithm.
- Evaluation of the performance dependence of these approaches on workflows and initialisations.

6.2.2 Algorithm implementation

- Integration of the L-BFGS-B implementation provided by other experts with projectors and penalty functions implemented in STIR and the GE proprietary PET Toolbox via MATLAB.
- Implementation of the proposed preconditioner and the transformed objective function and its gradient for L-BFGS-B-PC along with the modification to the line search subroutine of L-BFGS-B.
- Implementation of SPS.

- Implementation of PLS and its gradient for both 2-D and 3-D penalised image reconstruction.
- Implementation of the proposed spatially-variant penalty strength.
- Implementation of the two approaches for alternating misalignment estimation and penalised image reconstruction including anatomical priors based on a previously work from Dr. Alexandre Bousse.

6.3 Suggested future work

We have proposed a preconditioned image reconstruction algorithm, L-BFGS-B-PC, that shows much faster convergence rate than SPS and also faster than the non-preconditioned L-BFGS-B. To improve the image quality and quantitative consistency, an anatomical penalty function with a spatially-variant penalty strength was further incorporated into the proposed algorithm. The potential misalignment between functional and anatomical images were taken into account and two approaches that perform alternating misalignment estimation and penalised image reconstruction were proposed. To be able to prototype these ideas, implementations used in this study are not optimised yet. For example, as mentioned in Section 3.8, the L-BFGS-B was implemented in a combination of C, Fortran and MATLAB. Moreover, we have been using functions with MATLAB interfaces provided by the scanner manufacturer or other experts. By optimising the implementation of the proposed algorithms using the same programming language on a single platform, their performance in terms of the required memory or computation time can be improved. To enable evaluation on a wider dataset, it would also be useful to incorporate the implementation into open source distributions, such as Software for Tomographic Image Reconstruction (STIR) [88].

The algorithms or strategies proposed in this thesis have been demonstrated with simulated and a few patient datasets. Clearly, a thorough validation with more real data is required to demonstrate usefulness in future applications in the clinic. In addition, as the algorithm performance and the quantitative accuracy can be affected by the penalty strength and other parameters that determine the edge-preserving

property of the penalty (*e.g.*, ε and η in PLS), future work should also include parameter optimisation with respect to different applications. As the optimisation problem becomes less ill-conditioned when TOF data is available, we also expect the performance of the proposed methods can be further improved with TOF data.

Recall that L-BFGS-B initialises the line search with a step length of 1 after the first iteration. With this step length, the algorithm makes a direct approach to the local solution as described in Section 3.2.2. The backtracking of the embedded line search takes place only when the algorithm is about to converge and a smaller step length is obtained by following a predefined decreasing pattern. However, depending on the adopted decreasing scheme, the backtracking might not be able to find the step length that minimises the objective function at the current estimate, leading to an unexpected slow convergence rate. By substituting this generic line search with other more sophisticated ones in the literature [142, 143], the algorithm performance might be speeded up even more. This can also benefit other algorithms based on L-BFGS-B in this thesis, for example, the fast convergent algorithm L-BFGS-B-PC presented in Chapter 3 and the image reconstruction subroutine of the joint optimisation approaches presented in Chapter 5.

As discussed in Section 3.8, other image reconstruction algorithms should be able to benefit from preconditioning as well. In addition to applying the proposed preconditioner to different reconstruction algorithms, one can also explore the use of other fixed diagonal preconditioners. For example, by expressing ML-EM in a gradient descent form, a diagonal matrix with elements equal to a normalised version of the current estimate can be obtained. The matrix was used as a preconditioner to improve the convergence rate of a conjugate gradient algorithm in [91]. We can then replace the current estimate by the initial image so that it can be precomputed.

One of the motivations of applying penalised image reconstruction is to regularise noise propagation when iterations increase. In order to obtain uniform local perturbation response (*i.e.*, uniform local contrast), independent of lesion surroundings and location, a spatially-variant penalty strength was used to modify the applied penalty function in Chapter 4. The benefit of using the modified prior in quantitative

consistency has been demonstrated with noiseless datasets. However, as changes in the local resolution will influence the noise, bias and variance analysis using noise realisations is required for understanding the effect of the spatially-variant penalisation on overall resolution and noise trade-off. In Chapter 4, the proposed algorithm, L-BFGS-B-PC, was used as the representative reconstruction method. To be able to generalise the application of the spatially-variant penalty strength for improving convergence rate and its consistency, further demonstration of the strategy with other penalised reconstruction algorithms is necessary.

In Chapter 5, two approaches that take into account the potential misalignment between functional and anatomical images were proposed. The algorithms have been evaluated using simulated data with respect to different workflows, initialisations and the use of the PLS anatomical prior. Both approaches showed the ability to estimate the mismatch and warp the anatomical image accordingly. Although the current study simulated the misalignment between a single functional and anatomical imaging position, the proposed approaches should be able to be applied to multiple gated datasets with small modifications. To investigate the effectiveness of applying the proposed approaches in improving quantitative accuracy, evaluations using phantoms with inserted features or patient datasets with pseudo-lesions should also be included. In our current work, we assumed that the attenuation correction and anatomical image used for PLS were the same. However, the algorithms can be easily generalised to other user cases where the anatomical image is independent of the attenuation image. One particular case of this might be reconstructions of PET/CT data using an anatomical prior calculated with MR images.

Throughout Chapter 4 and 5, PLS was used as the representative anatomical prior. To demonstrate that the algorithms and strategies presented in this thesis can be generalised to a wider range of priors, evaluations based on reconstructions applying different edge preserving penalty functions can be useful. While most of the work here is done using PET examples and PET data, the results, with appropriate validation and possible reparameterisation, should also apply to SPECT imaging, where the fundamental theory for image reconstruction is very similar.

6.4 List of publications

In this section, publications that have arisen from work presented in this thesis and from collaborations with other researchers are listed.

6.4.1 Publications

Y.-J Tsai, C. W. Stearns, S. Ahn, A. Bousse, M. J. Ehrhardt, B. F. Hutton, S. Arridge, and K. Thielemans. "Fast quasi-Newton algorithms for penalized reconstruction in emission tomography and further improvements via preconditioning." *IEEE Transactions on Medical Imaging*, $37(4) : 1000 - 10$, 2018.

6.4.2 Submitted publication

Y.-J Tsai, G. Schramm, S. Ahn, A. Bousse, S. Arridge, J. Nuyts, B. F. Hutton, C. W. Stearns and K. Thielemans. "Benefits of using a spatially-variant penalty strength with anatomical priors in PET reconstruction" *IEEE Transactions on Medical Imaging*, submitted August 2018.

D. Deidda, N. Karakatsanis, P. Robson, **Y.-J. Tsai**, N. Efthimiou, K. Thielemans, Z. Fayad, R. Aykroyd and C. Tsoumpas. "Hybrid PET-MR List-Mode Kernelized Expectation Maximization Reconstruction" *Inverse Problems*, submitted May 2018.

6.4.3 Published conference proceedings

Y.-J Tsai, G. Schramm, J. Nuyts, S. Ahn, C. W. Stearns, A. Bousse, S. Arridge and Kris Thielemans. "Spatially-variant strength for anatomical priors in PET reconstruction." In *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, IEEE, 2017.

Y.-J Tsai, A. Bousse, C. W. Stearns, S. Ahn, B. F. Hutton, S. Arridge, and K. Thielemans. "Performance improvement and validation of a new MAP reconstruction algorithm." In *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, IEEE, 2016.

Y.-J Tsai, A. Bousse, M. J. Ehrhardt, B. F. Hutton, S. Arridge, and K. Thielemans. "Performance evaluation of MAP algorithms with different penalties, object geometries and noise levels." In *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, IEEE, 2015.

6.4.4 Published conference abstracts

R. Brown, B. A. Thomas, A. Rashidnasab, **Y.-J Tsai**, D. Deidda, E. Ovtchinnikov, E. Pasca, C. da Costa-Luis, B. F. Hutton, C. Tsoumpas, K. Thielemans. "Motion-corrected PET reconstruction with SIRF." In *7th Conference on PET-MRI and SPECT-MRI (PSMR)*, 2018.

L. Brusaferri, A. Bousse, **Y.-J Tsai**, D. Atkinson, S. Ourselin, B. F. Hutton, S. Arridge and K. Thielemans. "Attenuation estimation using non-TOF PET scattered photon energy information and an anatomical MRI prior." In *7th Conference on PET-MRI and SPECT-MRI (PSMR)*, 2018.

D. Deidda, N. Karakatsanis, P. M. Robson, **Y.-J Tsai**, N. Efthimiou, K. Thielemans, Z. A. Fayad, R. G. Aykroyd, C. Tsoumpas. "A comparison on two ways of including MR information into the PET reconstruction algorithm: The Hybrid Kernel and the Parallel Level Sets." In *7th Conference on PET-MRI and SPECT-MRI (PSMR)*, 2018.

6.4.5 Accepted conference abstract

Y.-J Tsai, A. Bousse, S. Ahn, C. W. Stearns, S. Arridge, B. F. Hutton and K. Thielemans. "Algorithms for solving misalignment issues in penalized PET/CT reconstruction using anatomical priors." In *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, IEEE, 2018.

Appendix A

The Separable Paraboloidal Surrogate (SPS) algorithm

In [72], SPS was proposed for penalised image reconstruction. Ignoring the non-negativity constraint on $\boldsymbol{f}$, the image update scheme of the algorithm can be summarised as follows with $\nabla\Phi$ the gradient of the objective function given in (2.12):

$$\begin{aligned} \boldsymbol{f}_{t+1} &= \boldsymbol{f}_t - \hat{\boldsymbol{D}}_t^2 \nabla\Phi(\boldsymbol{f}_t) \\ \hat{\boldsymbol{D}}_t &= \operatorname{diag}\left\{ \boldsymbol{A}^\top \boldsymbol{X}_t \boldsymbol{A}\mathbf{1} + \beta R_\varphi(\boldsymbol{f}_t)\mathbf{1} \right\}^{-\frac{1}{2}} \end{aligned} \tag{A.1}$$

where $\hat{\boldsymbol{D}}_t$ can be interpreted as a square root of a diagonal approximation of $\boldsymbol{H}^{-1}$, $R_\varphi(x) = \nabla R(x)/x$ provides the second order information of the penalty function R of the form in (2.14) with the potential function φ and $\boldsymbol{X}_t$ is a vector with the same length as the measured data $\boldsymbol{g}$. Given the current estimate data $\hat{\boldsymbol{g}}_t = \boldsymbol{A}\boldsymbol{f}_t + \boldsymbol{n}$, $\boldsymbol{X}_t$ can be computed by:

$$[\boldsymbol{X}_t]_i = \begin{cases} \left[2\frac{\Phi_i(0) - \Phi_i([\hat{\boldsymbol{g}}_t]_i) + \nabla\Phi_i([\hat{\boldsymbol{g}}_t]_i)[\hat{\boldsymbol{g}}_t]_i}{[\hat{\boldsymbol{g}}_t]_i^2} \right] & \text{if } [\hat{\boldsymbol{g}}_t]_i > 0 \\ \nabla^2\Phi_i(0) & \text{if } [\hat{\boldsymbol{g}}_t]_i = 0 \end{cases} \tag{A.2}$$

Note that as $\hat{\boldsymbol{D}}_t$ has to be updated at every iteration, leading to expensive computation, practical applications of SPS are limited. However, SPS has been shown to converge to the PML solution.

Appendix B

Relaxed Ordered Subsets SPS (OS-SPS)

To circumvent the high computational demand, the authors in [72] also proved the feasibility of applying the OS scheme and a precomputed denominator (OS-SPS). However, as for OS-EM method, OS-SPS suffers from the limit-cycle problem and requires a relaxation parameter to preserve convergence [59]. Ignoring the non-negativity constraint on $\boldsymbol{f}$, the update scheme of the relaxed OS-SPS can be expressed by:

$$\boldsymbol{f}_{t+1} = \boldsymbol{f}_t - \alpha_t \tilde{\boldsymbol{D}}^2 \nabla\Phi(\boldsymbol{f}_t),\ \alpha_t = \frac{\sigma}{1+t\gamma} \quad \text{and} \tag{B.1}$$

$$\tilde{\boldsymbol{D}} = \operatorname{diag}\left\{ \boldsymbol{A}^\top \operatorname{diag}\left\{ \frac{1}{\boldsymbol{g}} \right\} \boldsymbol{A}\mathbf{1} + \beta \nabla^2 R(\boldsymbol{f}_t)\mathbf{1} \right\}^{-\frac{1}{2}}$$

where $\tilde{\boldsymbol{D}}$ can be interpreted as a square root of a diagonal approximation of $\boldsymbol{H}^{-1}$ and α_t is the relaxation parameter. Although the algorithm can cope with ordered subsets, the update equation (B.1) has been written without subsets for simplicity. The decrease in α_t has to be chosen to satisfy certain conditions for convergence. The given scheme satisfies those conditions if the constants σ and γ are > 0. However, choosing optimal σ and γ for fast convergence is an unsolved problem.

Appendix C

Construction of the approximation of the inverse of the Hessian using a pair of correction matrices

This section describes the ApproxInvHess step in Algorithm 1. At every iteration t, the corresponding correction matrices consisting of gradient information in the last m iterations are expressed as follows:

$$\boldsymbol{S}_t = [\boldsymbol{s}_{t-m}, \ldots, \boldsymbol{s}_{t-1}],\ \boldsymbol{Y}_t = [\boldsymbol{y}_{t-m}, \ldots, \boldsymbol{y}_{t-1}] \tag{C.1}$$

where $\boldsymbol{s}_t = \boldsymbol{f}_{t+1} - \boldsymbol{f}_t$ and $\boldsymbol{y}_t = \nabla\Phi(\boldsymbol{f}_{t+1}) - \nabla\Phi(\boldsymbol{f}_t)$. These matrices can be used to find the second order behaviour of the objective function and therefore to calculate approximations of the Hessian. Based on the compact representations described in [84], the approximation of $\boldsymbol{H}^{-1}$ at iteration t can be written as follows:

$$\boldsymbol{B}_t \equiv \frac{1}{Q}\boldsymbol{I} + \bar{\boldsymbol{W}}_t \bar{\boldsymbol{M}}_t \bar{\boldsymbol{W}}_t^{\top} \tag{C.2}$$

where

$$
\begin{aligned}
\bar{\boldsymbol{W}}_t &\equiv \begin{bmatrix} \frac{1}{Q}\boldsymbol{Y}_t & \boldsymbol{S}_t \end{bmatrix}, \\
\bar{\boldsymbol{M}}_t &\equiv \begin{bmatrix} \boldsymbol{0} & -\boldsymbol{R}_t^{-1} \\ -\boldsymbol{R}_t^{-\top} & \boldsymbol{R}_t^{-\top}(\boldsymbol{V}_t + \frac{1}{Q}\boldsymbol{Y}_t^{\top}\boldsymbol{Y}_t\boldsymbol{R}_t^{-1}) \end{bmatrix}, \\
[\boldsymbol{R}_t]_{kl} &= \begin{cases} \boldsymbol{s}_{t-m-1+k}^{\top}\boldsymbol{y}_{t-m-1+l} & \text{if } k \leq l \\ \boldsymbol{0} & \text{otherwise} \end{cases}
\end{aligned}
$$

with $\boldsymbol{V}_t = \text{Diag}[\boldsymbol{s}_{t-m}^{\top}\boldsymbol{y}_{t-m}, \ldots, \boldsymbol{s}_{t-1}^{\top}\boldsymbol{y}_{t-1}]$, $k,l = 1,\ldots,m$ and Q is a constant [74]. The representation of $\boldsymbol{B}_t$ is efficient in terms of memory and computation time as $\bar{\boldsymbol{W}}_t$ is a $J \times 2m$ matrix and $\bar{\boldsymbol{M}}_t$ is $2m \times 2m$, where J is the number of voxels and $m = 5$ in this study. In practice, the algorithm does not compute and store $\boldsymbol{B}_t$ directly. Instead, it uses the correction matrices so that the product $\boldsymbol{B}_t \nabla\Phi(\boldsymbol{f}_t)$ can be calculated efficiently by applying the unrolling technique described in [84].

To initialise the construction of $\boldsymbol{B}_1$, the current implementation performs gradient descent at the first iteration to find the first pair of correction vectors, $\boldsymbol{S}_1 = [\boldsymbol{s}_0]$ and $\boldsymbol{Y}_1 = [\boldsymbol{y}_0]$. For iteration $t < m$, the corresponding $\boldsymbol{B}_t$ is calculated with only t pairs of gradient information.

Appendix D

Uniform cubic basis splines (B-splines)

The main structure of a cubic B-spline is summarised below. More information can be found in [144]. A cubic B-spline is a piecewise polynomial function consisting of four segments (Figure D.1):

$$
\begin{aligned}
b_0(u) &= \frac{1}{6}(1-3u+3u^2-u^3) \\
b_1(u) &= \frac{1}{6}(4-6u^2+3u^3) \\
b_2(u) &= \frac{1}{6}(1+3u+3u^2 3u^3) \\
b_3(u) &= \frac{1}{6}u^3 \,.
\end{aligned}
$$

Combining segments of these cubic functions in the interval $u = [0,\ 1]$, a positive, symmetric and continuously differentiable basis function is obtained (Figure D.2). In Chapter 5, the image function is represented as a linear combination of basis functions centred on a equally-spaced grid that coincides with the voxel centres. In other words, it is represented as a series of uniformly distributed cubic B-splines.

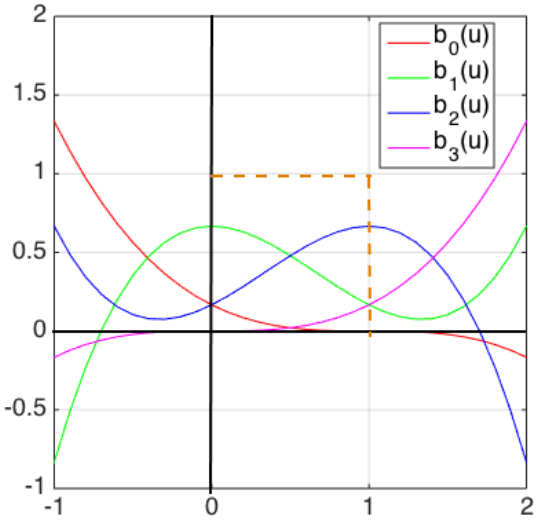

Figure D.1: Plots of the segments.

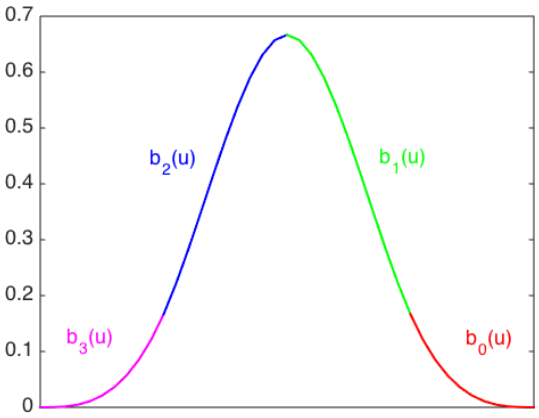

Figure D.2: Example of a basis function.

Bibliography

[1] A. Alessio and P. Kinahan. *PET image reconstruction*, volume 1. Amsterdam Elsevier, 2nd edition, 2006.

[2] D. L. Bailey and K. P. Willowson. An evidence-based review of quantitative SPECT imaging and potential clinical applications. *J Nucl. Med.*, 54(1):83–9, 2013.

[3] J. Y. Y. Ngeow, R. H. H. Quek, D. C. E. Ng, S. W. Hee, M. Tao, L. C. Lim, Y. H. Tan, and S. T. Lim. High SUV uptake on FDG-PET/CT predicts for an aggressive B-cell lymphoma in a prospective study of primary FDG-PET/CT staging in lymphoma. *Annals of Oncology*, 20(9):1543–7, 2009.

[4] B. Bai, J. Bading, and P. S Conti. Tumor quantification in clinical positron emission tomography. *Theranostics*, 3(10):787–801, 2013.

[5] R. Kumar, D. Halanaik, and A. Malhotra. Clinical applications of positron emission tomography-computed tomography in oncology. *Indian J. Cancer*, 47(2):100–19, 2010.

[6] E. Quak, A. CM van de Luijtgaarden, L.-F. de Geus-Oei, W. TA van der Graaf, and W. JG Oyen. Clinical applications of positron emission tomography in sarcoma management. *Expert Rev. Anticancer Ther.*, 11(2):195–204, 2011.

[7] Y. F. Tai and P. Piccini. Applications of positron emission tomography (PET) in neurology. *J Neurol. Neurosurg Psychiatry*, 75(5):669–76, 2004.

[8] K. A. Driver, A. E. Atchley, P. Kaul, and S. Borges-Neto. Single photon emission computed tomography myocardial imaging: clinical applications and future directions. *Minerva Cardioangiol.*, 57(3):333–347, 2009.

[9] M. D. Devous Sr. Single-photon emission computed tomography in neurotherapeutics. *NeuroRx.*, 2(3):237–49, 2005.

[10] D. J. Smith and J. T. Cavanagh. The use of single photon emission computed tomography in depressive disorders. *Nucl. Med. Commun.*, 26(3):197–203, 2005.

[11] I. Khalkhali, J. K. Baum, J. Villanueva-Meyer, S. L. Edell, L. G. Hanelin, C. E. Lugo, R. Taillefer, L. M. Freeman, C. E. Neal, A. M. Scheff, J. L. Connolly, S. J. Schnitt, M. J. Houlihan, J. S. Sampalis, and S. B. Haber. ^{99m}Tc Sestamibi breast imaging for the examination of patients with dense and fatty breasts: multicenter study. *Radiology*, 222(1):149–55, 2002.

[12] B. G. M. Durie, A. D. Waxman, A. D'Agnolo, and C. M. Williams. Whole-body ^{18}F-FDG PET identifies high-risk Myeloma. *J Nucl. Med.*, 43(11):1457–63, 2002.

[13] G. A. Beller. Myocardial perfusion imaging with Thallium-201. *J Nucl. Med.*, 35(40):674–80, 1994.

[14] M. Fiechter, J. R. Ghadri, C. Gebhard, T. A. Fuchs, A. P. Pazhenkottil, R. N. Nkoulou, B. A. Herzog, C. A. Wyss, O. Gaemperli, and P. A. Kaufmann. Diagnostic value of 13N-Ammonia myocardial perfusion PET: added value of myocardial flow reserve. *J Nucl. Med.*, 53(8):1230–4, 2012.

[15] M. B. Imran, R. Kawashima, S. Awata, K. Sato, S. Kinomura, S. Ono, M. Sato, and H. Fukuda. Tc-99m HMPAO SPECT in the evaluation of Alzheimer's disease: correlation between neuropsychiatric evaluation and CBF images. *J Neurol. Neurosurg. Psychiatry*, 66(2):228–32, 1999.

[16] D. A. Wolk, Z. Zhang, S. Boudhar, C. M. Clark, M. J. Pontecorvo, and S. E. Arnold. Amyloid imaging in Alzheimer's disease: comparison of Florbetapir and Pittsburgh Compound-B positron emission tomography. *J Neurol. Neurosurg. Psychiatry*, 83(9):923–6, 2012.

[17] P. P. Bruyant. Analytic and iterative reconstruction algorithms in SPECT. *J Nucl. Med.*, 43(10):1343–58, 2002.

[18] S. N. Histed, M. L. Lindenberg, E. Mena, B. Turkbey, P. L. Choyke, and K. A. Kurdziel. Review of functional/anatomic imaging in oncology. *Nucl. Med. Commun.*, 33(4):349–61, 2012.

[19] A. H. Maurer. Combined imaging modalities: PET/CT and SPECT/CT. *Health Phys.*, 95(5):571–6, 2008.

[20] D. W. Townsend. Multimodality imaging of structure and function. *Phys. Med. Biol.*, 53(4):R1–R39, 2008.

[21] G. Wagenknecht, H.-J. Kaiser, F. M. Mottaghy, and H. Herzog. MRI for attenuation correction in PET: methods and challenges. *Semin. Nucl. Med.*, 26(1):99–113, 2013.

[22] J. P. J. Carneya and D. W. Townsend. Method for transforming CT images for attenuation correction in PET/CT imaging. *Med. Phys.*, 33(4):976–83, 2006.

[23] S. E. Jung, J. M. Lee, S. E. Rha, J. Y. Byun, J. I. Jung, and S. T. Hahn. CT and MR imaging of ovarian tumors with emphasis on differential diagnosis. *RadioGraphics*, 22(6):1305–25, 2002.

[24] W. R. Webb, C. Gatsonis, E. A. Zerhouni, R. T. Heelan, G. M. Glazer, I. R. Francis, and B. J. McNeil. CT and MR imaging in staging non-small cell bronchogenic carcinoma: report of the Radiologic Diagnostic Oncology Group. *Radiology*, 178(3):705–13, 1991.

[25] C. Dromain, T. de Baere, D. Elias, V. Kuoch, M. Ducreux, V. Boige, P. Petrow, A. Roche, and R. Sigal. Hepatic tumors treated with percutaneous radio-frequency ablation: CT and MR imaging follow-up. *Radiology*, 223(1):255–62, 2002.

[26] R. K. T. Haken, A. F. Thornton Jr., H. M. Sandler, M. L. LaVigne, D. J. Quint, B. A. Fraass, M. L. Kessler, and D. L. McShan. A quantitative assessment of the addition of MRI to CT-based, 3-D treatment planning of brain tumors. *Radiology*, 25(2):121–33, 1992.

[27] D. Delbeke, H. Schöder, W. H. Martin, and R. L. Wahl. Hybrid imaging (SPECT/CT and PET/CT): improving therapeutic decisions. *Semin. Nucl. Med.*, 39(5):308–40, 2009.

[28] J. H. Oliver and R. L. Baron. Helical biphasic contrast-enhanced CT of the liver: technique, indications, interpretation, and pitfalls. *Radiology*, 201(1):1–14, 1996.

[29] G. Brancatelli, R. L. Baron, M. S. Peterson, and W. Marsh. Helical CT screening for HCC in patients with cirrhosis: frequency and causes of false-positive interpretation. *AJR Am J Roentgenol.*, 180(4):1007–14, 2003.

[30] D. J. Godfrey, B. N. Patel, J. D. Adamson, E. Subashi, J. K. Salama, and M. Palta. Triphasic contrast enhanced CT simulation with bolus tracking for pancreas SBRT target delineation. *Pract. Radiat. Oncol.*, 7(6):489–97, 2017.

[31] A. C. Pfannenberg, P. Aschoff, K. Brechtel, M. MÏler, R. Bares, F. Paulsen, J. Scheiderbauer, G. Friedel, C. D. Claussen, and S M. Eschmann. Low dose non-enhanced CT versus standard dose contrast-enhanced CT in combined PET/CT protocols for staging and therapy planning in non-small cell lung cancer. *Eur. J Nucl. Med. Mol.*, 34(1):36–44, 2007.

[32] G. Antoch, L. S. Freudenberg, T. Beyer, A. Bockisch, and J. Debatin. To enhance or not to enhance? 18F-FDG and CT contrast agents in dual-modality 18F-FDG PET/CT. *J Nucl. Med.*, 45(1):56S–65S, 2004.

[33] C. R. Li, Y. Z. Li, Y. M. Li, and Y. S. Zheng. Dynamic and contrast enhanced CT imaging of lung carcinoma, pulmonary tuberculoma, and inflammatory pseudotumor. *Eur. Rev. Med. Pharmacol. Sci.*, 21(1):1588–92, 2017.

[34] K. M. Horton, F. M. Corl, and E. K. Fishman. CT evaluation of the colon: inflammatory disease. *RadioGraphics*, 20(2):419–30, 2000.

[35] V. P. Hart II. *The application of tomographic reconstruction techniques to ill-conditioned inverse problems in atmospheric science and biomedical imaging*. PhD thesis, Utah State University, 2012.

[36] J. Dutta, S. Ahn, and Q. Li. Quantitative statistical methods for image quality assessment. *Theranostics.*, 3(10):741–56, 2013.

[37] J. M. Ollinger and J. A. Fessler. Positron-emission tomography. *IEEE Signal Processing Mag.*, 14(1):43–55, 1997.

[38] K. Vunckx, A. Atre, K. Baete, A. Reilhac, C. M. Deroose, K. V. Laere, and J. Nuyts. Evaluation of three MRI-based anatomical priors for quantitative PET brain imaging. *IEEE Trans. Med. Imag.*, 31(3):599–612, 2012.

[39] K. M. Hanson. Introduction to Bayesian image analysis. In *Proc. SPIE*, volume 20, pages 716–31, 1993.

[40] B. Bai, Q. Li, and R. M. Leahy. MR guided PET image reconstruction. *Semin. Nucl. Med.*, 43(1):30–44, 2013.

[41] P. J. Green. Bayesian Reconstructions From Emission Tomography Data Using a Modified EM Algorithm. *IEEE Trans. Med. Imag.*, 9(1):84–93, 1990.

[42] L. Yang, J. Zhou, A. Ferrero, R. D. Badawi, and J. Qi. Regularization design in penalized maximum-likelihood image reconstruction for lesion detection in 3D PET. *Phys. Med. Biol.*, 59(2):403–19, 2013.

[43] G. Wang and J. Qi. Penalized likelihood PET image reconstruction using patch-based edge-preserving regularization. *IEEE Trans. Med. Imag.*, 31(12):2194–204, 2012.

[44] Y.-J. Tsai, H.-M. Huang, C.-Y. Chou, W. Wang, and I.-T. Hsiao. Effective anatomical priors for emission tomographic reconstruction. *J Med. Biol. Eng.*, 35(1):52–61, 2015.

[45] K. Sakaguchi, H. Shinohara, T. Hashimoto, and T. Yokoi. An iterative reconstruction using median root prior and anatomical prior from the segmented μ-Map for count-limited transmission data in PET imaging. *Ann. Nucl. Med.*, 22(4):269–79, 2008.

[46] G. Gindi, M. Lee, A. Rangarajan, and I. G. Zubal. Bayesian reconstruction of functional images using anatomical information as priors. *IEEE Trans. Med. Imag.*, 12(4):670–80, 1993.

[47] B. Lipinski, H. Herzog, E. Rota Kops, W. Oberschelp, and H. W. Müller-Gärtner. Expectation maximization reconstruction of positron emission tomography images using anatomical magnetic resonance information. *IEEE Trans. Med. Imag.*, 16(2):129–36, 1997.

[48] C. Comtat, P. E. Kinahan, J. A. Fessler, T. Beyer, D. W. Townsend, M. Defrise, and C. Michel. Clinically feasible reconstruction of 3D whole-body PET/CT data using blurred anatomical labels. *Phys. Med. Biol.*, 47(1):1–20, 2002.

[49] Y. Mameuda and H. Kudo. New anatomical-prior-based image reconstruction method for PET/SPECT. *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf. Rec.*, 6:4142–8, 2007.

[50] J. Nuyts, K. Baete, D. Bequé, and P. Dupont. Comparison between MAP and postprocessed ML for image reconstruction in emission tomography when anatomical knowledge is available. *IEEE Trans. Med. Imag.*, 24(5):667–75, 2005.

[51] P. P. Bruyant, H. C. Gifford, G. Gindi, and M. A. King. Numerical observer study of MAP-OSEM regularization methods with anatomical priors for lesion detection in 67-Ga images. *IEEE Trans. Nucl. Sci.*, 51(1):193–7, 2004.

[52] J. E. Bowsher, V. E. Johnson, T. G. Turkington, R. J. Jaszczak, C. E. Floyd Jr., and R. E. Coleman. Bayesian reconstruction and use of anatomical a priori information for emission tomography. *IEEE Trans. Med. Imag.*, 15(5):673–86, 1996.

[53] A. Mehranian, M. A. Belzunce, F. Niccolini, M. Politis, C. Prieto, F. Turkheimer, A. Hammers, and A. J. Reader. PET Image Reconstruction Using Multi-Parametric Anato-Functional Priors. *Phys. Med. Biol.*, 62(15):5975–6007, 2017.

[54] G. Schramm, M. Holler, A. Rezaei, K. Vunckx, F. Knoll, K. Bredies, F. Boada, and J. Nuyts. Evaluation of Parallel Level Sets and Bowshers Method as Segmentation-Free Anatomical Priors for Time-of-Flight PET Reconstruction. *IEEE Trans. Med. Imag.*, 37(2):590–603, 2018.

[55] E. Haber and M. H. Gazit. Model fusion and joint inversion. *Surv. Geophys.*, 34(5):675–95, 2013.

[56] M. J. Ehrhardt and S. R. Arridge. Vector-valued image processing by parallel level sets. *IEEE Trans. Image Processing*, 23(1):9–18, 2014.

[57] M. J. Ehrhardt, K. Thielemans, L. Pizarro, D. Atkinson, S. Ourselin, B. F. Hutton, and S. R. Arridge. Joint reconstruction of PET-MRI by exploiting structural similarity. *Inverse Problems*, 31(1):1–23, 2015.

[58] M. J. Ehrhardt, P. Markiewicz, M. Liljeroth, A. Barnes, V. Kolehmainen, J. S. Duncan, L. Pizarro, D. Atkinson, B. F. Hutton, S. Ourselin, K. Thielemans, and S. R. Arridge. PET reconstruction with an anatomical MRI prior using parallel level sets. *IEEE Trans. Med. Imag.*, 35(9):2189–99, 2016.

[59] S. Ahn and J. A. Fessler. Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms. *IEEE Trans. Med. Imag.*, 22(5):613–26, 2003.

[60] L. A. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Trans. Med. Imag.*, 1(2):113–22, 1982.

[61] H. M. Hudson and R. S. Larkin. Accelerated image reconstruction using ordered subsets of projection data. *IEEE Trans Med. Imag.*, 13(4):601–9, 1994.

[62] A. M. Morey and D. J. Kadrmas. Effect of varying number of OSEM subsets on PET lesion detectability. *J Nucl. Med. Technol.*, 41(4):268–73, 2013.

[63] E. Tanaka, N. Nohara, T. Tomitani, and M. Yamamoto. *Utilization of non-negativity constraints in reconstruction of emission tomograms. in: information processing in medical imaging*. Dordrecht Springer, 1996.

[64] D. Hwang and G. L. Zeng. Convergence study of an accelerated ML-EM algorithm using bigger step size. *Phys. Med. Biol.*, 51(2):237–52, 2006.

[65] L. Kaufman. Implementing and accelerating the EM algorithm for positron emission tomography. *IEEE Trans Med. Imag.*, 6(1):37–51, 1987.

[66] 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.*, 5(1):16–22, 1986.

[67] E. Tanaka. A fast reconstruction algorithm for stationary positron emission tomography based on a modified EM algorithm. *IEEE Trans. Med. Imag.*, 6(2):98–105, 1987.

[68] J. Browne and A. B. de Pierro. A row-action alternative to the EM algorithm for maximizing likelihood in emission tomography. *IEEE Trans. Med. Imag.*, 15(5):687–99, 1996.

[69] P. J. Green. On use of the EM for penalized likelihood estimation. *J. Roy. Stat. Soc. B*, 52(3):443–52, 1990.

[70] A. R. De Pierro and M. E. B. Yamagishi. Fast EM-like methods for maximum "a posteriori" estimates in emission tomography. *IEEE Trans. Med. Imag.*, 20(4):280–8, 2001.

[71] A. R. De Pierro. A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography. *IEEE Trans. Med. Imag.*, 14(1):132–7, 1995.

[72] H Erdoğan and J. A. Fessler. Ordered subsets algorithms for transmission tomography. *Phys. Med. Biol.*, 44(1):2835–51, 1999.

[73] J. Nocedal and S. J. Wright. *Numerical optimization*. Springer, 2nd edition, 2006.

[74] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited memory algorithm for bound constrained optimization. *SIAM J Sci. Comput.*, 16(5):1190–208, 1995.

[75] S. Ahn, S. G. Ross, E. Asma, J. Miao, X. Jin, L. Cheng, S. D. Wollenweber, and R. M. Manjeshwar. Quantitative comparison of OSEM and penalized likelihood image reconstruction using relative difference penalties for clinical PET. *Phys. Med. Biol.*, 60(15):5733–51, 2015.

[76] H. Matthies and G. Strang. The solution of nonlinear finite element equations. *SIAM J Sci. Comput.*, 14(11):1613–26, 1979.

[77] J. Nocedal. Updating quasi-Newton matrices with limited storage. *Math. Comp.*, 35(151):773–82, 1980.

[78] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal. Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software*, 23(4):550–60, 1997.

[79] G. Andrew and J. Gao. Scalable training of L1-regularized log-linear models. In *International Conference on Machine Learning*, pages 33–40, 2007.

[80] S. Somayajula, C. Panagiotou, A. Rangarajan, Q. Li, S. R. Arridge, and R. M. Leahy. PET image reconstruction using information theoretic anatomical priors. *IEEE Trans. Med. Imag.*, 30(3):537–49, 2011.

[81] Y.-J. Tsai, A. Bousse, M. J. Ehrhardt, B. F. Hutton, S. Arridge, and K. Thielemans. Performance evaluation of MAP algorithms with different penalties, object geometries and noise levels. *IEEE Nucl. Sci. Symp. and Med. Imag. Conf. Rec.*, 2015.

[82] M. S. Kaplan, D. R. Haynor, and H. Vija. A differential attenuation method for simultaneous estimation of SPECT activity and attenuation distributions. *IEEE Trans. Nucl. Sci.*, 46(3):535–41, 1999.

[83] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal. L-BFGS-B: Fortran subroutines for large-scale bound constrained optimization. Technical report, EECS Tech. Rep. NAM12. Northwestern University, Evanston, IL., 1995.

[84] R. H. Byrd, J. Nocedal, and R. B. Schnabel. Representations of quasi-Newton matrices and their use in limited memory methods. *Mathematical Programming*, 63(1–3):129–56, 1994.

[85] J. J. Moré and D. J. Thuente. On line search algorithms with guarenteed sufficient decrease. *ACM Trans. Math. Software*, 20(3):286–307, 1994.

[86] T. Allahviranloo, R. G. Moghaddam, and M. Afshar. Comparison theorem with modified Gauss-Seidel and modified Jacobi methods by M-matrix. *Journal of Interpolation and Approximation in Scientific Computing*, 2012:1–8, 2012.

[87] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. W. Tsui. 4D XCAT phantom for multimodality imaging research. *Med. Phys.*, 37(9):4902–15, 2010.

[88] K. Thielemans, C. Tsoumpas, S. Mustafovic, T. Beisel, P. Aguiar, N. Dikaios, and M. W. Jacobson. STIR: software for tomographic image reconstruction release 2. *Phys. Med. Biol.*, 57:867–83, 2012.

[89] M. Teräs, T. Tolvanen, J. J. Johansson, J. J. Williams, and J. Knuuti. Performance of the new generation of whole-body PET/CT scanners: Discovery STE and Discovery VCT. *Eur. J. Nucl. Med. Mol. Imaging*, 34(10):1683–92, 2007.

[90] K. Lange, M. Bahn, and R. Little. A theoretical study of some maximum likelihood algorithms for emission and transmission tomography. *IEEE Trans. Med. Imag.*, 6(2):106–14, 1987.

[91] E. Ü. Mumcuoğlu, R. M. Leahy, S. R. Cherry, and Z. Zhou. Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET images. *IEEE Trans. Med Imag.*, 13(4):687–701, 1994.

[92] J. A. Fessler, E. P. Ficaro, N. H. Clinthorne, and K. Lange. Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction. *IEEE Trans. Med. Imag.*, 16(2):166–175, 1997.

[93] D. M. Higdon, J. E. Bowsher, V. E. Johnson, T. G. Turkington, D. R. Gilland, and R. J. Jaszczak. Fully Bayesian estimation of Gibbs hyperparameters for emission computed tomography data. *IEEE Trans. Med. Imag.*, 16(5):516–26, 1997.

[94] J. A. Fessler and W. L. Rogers. Spatial resolution properties of penalized-likelihood image reconstruction: space-invariant tomographs. *IEEE Trans. Image Processing*, 5(9):1346–58, 1996.

[95] J. A. Fessler. Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): applications to tomography. *IEEE Trans. Med. Imag.*, 5(3):493–506, 1996.

[96] Q. Li, E. Asma, J. Qi, J. R. Bading, and R.M. Leahy. Accurate estimation of the Fisher information matrix for the PET image reconstruction problem. *IEEE Trans. Medical Imaging*, 23(9):1057–64, 2004.

[97] J. Qi and R. M. Leahy. A theoretical study of the contrast recovery and variance of MAP reconstructions from PET data. *IEEE Trans. Med. Imag.*, 18(4):293–305, 1999.

[98] J. Qi and R. M. Leahy. Resolution and noise properties of MAP reconstruction for fully 3-D PET. *IEEE Trans. Med. Imag.*, 19(5):493–506, 2000.

[99] J. W. Stayman and J. A. Fessler. Regularization for uniform spatial resolution properties in penalized-likelihood image reconstruction. *IEEE Trans. Med. Imag.*, 19(6):601–15, 2000.

[100] J. W. Stayman and J. A. Fessler. Compensation for nonuniform resolution using penalized-likelihood reconstruction in space-variant imaging systems. *IEEE Trans. Med. Imag.*, 23(3):269–84, 2004.

[101] J. Qi and R. H. Huesman. Penalized maximum-likelihood image reconstruction for lesion detection. *Phys. Med. Biol.*, 51(16):4017–29, 2006.

[102] J. Nuyts and J. A. Fessler. A penalized-likelihood image reconstruction method for emission tomography, compared to postsmoothed maximum-likelihood with matched spatial resolution. *IEEE Trans. Med. Imag.*, 22(9):1042–52, 2003.

[103] S. Ahn and R. M. Leahy. Analysis of resolution and noise properties of nonquadratically regularized image reconstruction methods for PET. *IEEE Trans. Med. Imag.*, 27(3):413–24, 2008.

[104] P. Cheebsumon, F. H. P. van Velden, M. Yaqub, V. Frings, A. J. de Langen, O. S. Hoekstra, A. A. Lammertsma, and R. Boellaard. Effects of image characteristics on performance of tumor delineation methods: a test-retest assessment. *J. Nucl. Med.*, 52(10):1550–8, 2011.

[105] A. Rahmim, J. Qi, and V. Sossi. Resolution modeling in PET imaging: theory, practice, benefits, and pitfalls. *Med Phys.*, 40(6):064301–15, 2013.

[106] V. Evilevitch, F. Eilber, W. Tap, M. Allen-Auerbach, M. Phelps, J. Czernin, and W. Weber. Influence of ROI definition on the results of FDG-PET studies in patients with sarcomas. *J. Nucl. Med.*, 48(Suppl 2):145P, 2007.

[107] M. Vanderhoek, S. B. Perlman, and R. Jeraj. Impact of the definition of peak standardized uptake value on quantification of treatment response. *J. Nucl. Med.*, 53(1):4–11, 2012.

[108] J. F. Eary, D. S. Hawkins, E. T. Rodler, and E. U. Conrad III. ^{18}F-FDG PET in sarcoma treatment response imaging. *Am. J. Nucl. Med. Mol. Imaging*, 1(1):47–53, 2011.

[109] M. Brændengen, K. Hansson, C. Radu, A. Siegbahn, H. Jacobsson, and B. Glimelius. Delineation of gross tumor volume (GTV) for radiation treatment planning of locally advanced rectal cancer using information from MRI or FDG-PET/CT: a prospective study. *Int. J. Radiat. Oncol. Biol. Phys.*, 81(4):439–45, 2011.

[110] D. E. Spratt, R. Diaz, J. McElmurray, I. Csiki, D. Duggan, B. Lu, and D. Delbeke. Impact of FDG PET/CT on delineation of the gross tumor volume for radiation planning in non-small-cell lung cancer. *Clin. Nucl. Med.*, 35(4):237–43, 2010.

[111] J. E. Bowsher, H. Yuan, et al. Utilizing MRI information to estimate F18-FDG distributions in rat flank tumors. *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf. Rec.*, 4:2488–2492, 2004.

[112] Y.-J. Tsai, A. Bousse, C. W. Stearns, S. Ahn, B. F. Hutton, S. Arridge, and K. Thielemans. Performance improvement and validation of a new MAP reconstruction algorithm. *IEEE Nucl. Sci. Symp. and Med. Imag. Conf. Rec.*, 2016.

[113] Y.-J. Tsai, A. Bousse, M. J. Ehrhardt, C. W. Stearns, S. Ahn, B. F. Hutton, S. Arridge, and Thielemans K. Fast quasi-Newton algorithms for penalized reconstruction in emission tomography and further improvements via preconditioning. *IEEE Trans. Med. Imag.*, 37(4):1000–10, 2018.

[114] Y. E. Chung, K. W. Kim, J. H. Kim, J. S. Lim, Y. T. Oh, J.-J. Chung, and M.-J. Kim. Optimal delay time for the hepatic parenchymal enhancement at the multidetector CT examination. *J. Comput. Assist Tomogr.*, 30(2):182–8, 2006.

[115] F. Gao, M. Li, Y. Sun, L. Xiao, and Y. Hua. Diagnostic value of contrast-enhanced CT scans in identifying lung adenocarcinomas manifesting as GGNs (ground glass nodules). *Medicine (Baltimore)*, 96(43):1–6, 2017.

[116] H. Ohkubo, Y. Kanemitsu, T. Uemura, O. Takakuwa, M. Takemura, K. Maeno, Y. Ito, T. Oguri, N. Kazawa, R. Mikami, and A. Niimi. Normal lung quantification in usual interstitial pneumonia pattern: the impact of threshold-based volumetric CT analysis for the staging of idiopathic pulmonary fibrosis. *PLoS One*, 11(3):1–13, 2016.

[117] E. J. Teoh, D. R. McGowan, R. E. Macpherson, K. M. Bradley, and F. V. Gleeson. Phantom and clinical evaluation of the Bayesian penalized likelihood reconstruction algorithm Q.Clear on an LYSO PET/CT system. *J. Nucl. Med.*, 56(9):1447–52, 2015.

[118] J. Lantos, E. S. Mittra, C. S. Levin, and A. Iagaru. Standard OSEM vs. regularized PET image reconstruction: qualitative and quantitative comparison using phantom data and various clinical radiopharmaceuticals. *Am. J. Nucl. Med. Mol. Imaging*, 8(2):110–8, 2018.

[119] S. Kulkarni, P. Khurd, I. Hsiao, L. Zhou, and G. Gindi. A channelized Hotelling observer study of lesion detection in SPECT MAP reconstruction using anatomical priors. *Phys. Med. Biol.*, 52(12):3601–17, 2007.

[120] J. Liao and J. Qi. PET image reconstruction with anatomical prior using multiphase level set method. *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf. Rec.*, 2007.

[121] O. Camara, G. Delso, O. Colliot, A. Moreno-Ingelmo, and I. Bloch. Explicit incorporation of prior anatomical information into a nonrigid registration of thoracic and abdominal CT and 18-FDG whole-body emission PET images. *IEEE Trans. Med. Imag.*, 26(2):164–78, 2007.

[122] Q. Xu, K. Yuan, and D. Ye. Respiratory motion blur identification and reduction in ungated thoracic PET imaging. *Phys. Med. Biol.*, 56(14):4481–98, 2011.

[123] M. J. Nyflot, T. C. Lee, A. M. Alessio, S. D. Wollenweber, C. W. Stearns, S. R. Bowen, and P. E. Kinahan. Impact of CT attenuation correction method on quantitative respiratory-correlated (4D) PET/CT imaging. *Med. Phys.*, 42(1):110–20, 2015.

[124] A. Pépin, J. Daouk, P. Bailly, S. Hapdey, and M. Meyer. Management of respiratory motion in PET/computed tomography: the state of the art. *Nucl. Med. Commun.*, 35(2):113–22, 2014.

[125] S. A. Nehmeh and Y. E. Erdi. Respiratory motion in positron emission tomography/computed tomography: a review. *Semin. Nucl. Med.*, 38(3):167–76, 2008.

[126] S. A. Nehmeh, Y. E. Erdi, T. Pan, A. Pevsner, K. E. Rosenzweig, E. Yorke, G. S. Mageras, H. Schoder, P. Vernon, O. Squire, H. Mostafavi, S. M. Larson, and J. L. Humm. Four-dimensional (4D) PET/CT imaging of the thorax. *Med. Phys.*, 31(12):3179–86, 2004.

[127] R. Manjeshwar, X. Tao, E. Asma, and K. Thielemans. Motion compensated image reconstruction of respiratory gated PET/CT. *IEEE Int. Symp. Biomed. Imag.*, pages 674–7, 2006.

[128] F. Qiao, T. Pan, J. W Clark Jr, and O. R. Mawlawi. A motion-incorporated reconstruction method for gated PET studies. *Phys. Med. Biol.*, 51(15):3769–83, 2006.

[129] B. K. Teo, B. Saboury, R. Munbodh, J. Scheuermann, D. A. Torigian, H. Zaidi, and A. Alavi. The effect of breathing irregularities on quantitative accuracy of respiratory gated PETCT. *Med. Phys.*, 39(12):7390–7, 2012.

[130] J. R. McClelland, D. J. Hawkes, T. Schaeffter, and A. P. King. Respiratory motion models: a review. *Med. Image Anal.*, 17(1):19–42, 2013.

[131] R. Manber, K. Thielemans, B. F. Hutton, S. Wan, J. McClelland, A. Barnes, S. Arridge, S. Ourselin, and D. Atkinson. Joint PET-MR respiratory motion models for clinical PET motion correction. *Phys. Med. Biol.*, 61(17):6515–30, 2016.

[132] F. Kalantari and Jing Wang. Attenuation correction in 4D-PET using a single-phase attenuation map and rigidity-adaptive deformable registration. *Med. Phys.*, 44(2):522–32, 2017.

[133] S. J. McQuaid, T. Lambrou, and B. F. Hutton. A novel method for incorporating respiratory-matched attenuation correction in the motion correction of cardiac PET-CT studies. *Phys. Med. Biol.*, 56(10):2903–15, 2011.

[134] H. Fayad, J. F. Clement, T. Pan, C. Roux, C. C. L. Rest, O. Pradier, and D. Visvikis. Towards a generic respiratory motion model for 4D CT imaging of the thorax. *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf. Rec.*, pages 3975–9, 2009.

[135] J. Nuyts, P. Dupont, S. Stroobants, R. Benninck, L. Mortelmans, and P. Suetens. Simultaneous maximum a posteriori reconstruction of attenuation and activity distributions from emission sinograms. *IEEE Trans. Med. Imag.*, 18(5):393–403, 1999.

[136] A. Rezaei, M. Defrise, G. Bal, C. Michel, M. Conti, C. Watson, and J. Nuyts. Simultaneous reconstruction of activity and attenuation in time-of-flight PET. *IEEE Trans. Med. Imag.*, 31(12):2224–33, 2012.

[137] A. Bousse, O. Bertolli, D. Atkinson, S. Arridge, S. Ourselin, B. F. Hutton, and K. Thielemans. Maximum-likelihood joint image reconstruction/motion estimation in attenuation-corrected respiratory gated PET/CT using a single attenuation map. *IEEE Trans. Med. Imag.*, 35(1):217–28, 2016.

[138] A. Rezaei, C. Michel, M. E. Casey, and J. Nuyts. Simultaneous reconstruction of the activity image and registration of the CT image in TOF-PET. *Phys. Med. Biol.*, 61(4):1852–74, 2016.

[139] A. Bousse, O. Bertolli, D. Atkinson, S. Arridge, S. Ourselin, B. F. Hutton, and K. Thielemans. Maximum-likelihood joint image reconstruction and motion estimation with misaligned attenuation in TOF-PET/CT. *Phys. Med. Biol.*, 61(3):L11–19, 2016.

[140] T. E. Nichols, J. Qi, E. Asma, and R. M. Leahy. Spatiotemporal reconstruction of list-mode PET data. *IEEE Trans. Med. Imag.*, 21(4):396–404, 2002.

[141] M. A. Belzunce, A. Mehranian, J. Bland, and A. J. Reader. ML and MAP PET reconstruction with MR-voxel sizes for simultaneous PET-MR. *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf. Rec.*, 2017.

[142] M. Ahookhosh and S. Ghaderi. On efficiency of nonmonotone Armijo-type line searches. *Appl. Math. Model.*, 43:170–90, 2017.

[143] G. Yuan, Z. Wei, and X. Lu. Global convergence of BFGS and PRP methods under a modified weak Wolfe-Powell line search. *Appl. Math. Model.*, 47:811–25, 2017.

[144] M. Unser, A. Aldroubi, and M. Eden. B-Spline Signal Processing: Part I-Theory. *IEEE Trans. Signal Proc.*, 41(2):821–33, 1993.