Stochastic Optimisation Methods Applied to PET Image Reconstruction

Robert Twyman Skelly

Supervisors:

Prof. Kris Thielemans

Prof. Simon Arridge

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

Medical Physics

Medical Physics and Biomedical Engineering

University College London

September 6, 2022

I, Robert Twyman Skelly, 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

Positron Emission Tomography (PET) is a medical imaging technique that is used to provide functional information regarding physiological processes. Statistical PET reconstruction attempts to estimate the distribution of radiotracer in the body but this methodology is generally computationally demanding because of the use of iterative algorithms. These algorithms are often accelerated by the utilisation of data subsets, which may result in convergence to a limit set rather than the unique solution. Methods exist to relax the update step sizes of subset algorithms but they introduce additional heuristic parameters that may result in extended reconstruction times. This work investigates novel methods to modify subset algorithms to converge to the unique solution while maintaining the acceleration benefits of subset methods.

This work begins with a study of an automatic method for increasing subset sizes, called AutoSubsets. This algorithm measures the divergence between two distinct data subset update directions and, if significant, the subset size is increased for future updates. The algorithm is evaluated using both projection and list mode data. The algorithm's use of small initial subsets benefits early reconstruction but unfortunately, at later updates, the subsets size increases too early, which impedes convergence rates.

The main part of this work investigates the application of stochastic variance reduction optimisation algorithms to PET image reconstruction. These algorithms reduce variance due to the use of subsets by incorporating previously computed subset gradients into the update direction. The algorithms are adapted for the application to PET reconstruction. This study evaluates the reconstruction performance of these algorithms when applied to various 3D non-TOF PET simulated, phantom and patient data sets. The impact of a number of algorithm parameters are explored, which includes: subset selection methodologies, the number of subsets, step size methodologies and preconditioners. The results indicate that these stochastic variance reduction algorithms demonstrate superior performance after only

a few epochs when compared to a standard PET reconstruction algorithm.

Acknowledgements

I would like to begin by acknowledge funding support from GE Healthcare, NIHR UCLH Biomedical Research Centre, and EPSRC-funded UCL Centre for Doctoral Training in Medical Imaging (EP/L016478/1).

There are a number of people I would like to thank for their support, aid and guidance throughout my Ph.D project. First of all, I thank my primary supervisor, Prof. Kris Thielemans, who has been invaluable to me over the past years. Likewise, I thank my secondary supervisor, Prof. Simon Arridge, for all the helpful discussions and valuable input to this research. A special thanks goes to Prof. Brian Hutton for his sharing of knowledge. Thank you to those members at GE, including Sangtae Ahn, Chuck Sterns, Scott Wollenweber and Kuan-Hao Su (Dylan), who provided support for this research. Thank you to Fotis Kotasidis, Irene Burger and Catherine Scott for providing a number of data sets used throughout this research project. Additionally, I would like to thank Alex, Elise, Ludovica, Rebecca, Richard, and the rest of the INM team for your discussions and helpful insights.

Finally, I would like to thank my wife, Evanya, for her constant support throughout my long education and for always believing in me. Without her, I would never have dreamt I could be who I am today.

Again, thank you all!

Impact Statement

Quantification of activity concentrations in PET imaging is essential for monitoring disease progression and treatment effectiveness for a patient's healthcare. Statistical image reconstruction methods have been developed to increase accuracy and precision. However, they suffer from long computation time. Currently commercially available algorithms accelerate computation time by using data subsets, but sacrifice performance over speed.

The novel image reconstruction algorithms investigated in this thesis are proposed as alternatives to standard clinical iterative PET image reconstruction algorithms. Our evaluation demonstrated that the use of stochastic optimisation improves image convergence rates. The proposed methods and their ability to quickly produce stabilised subset image reconstruction sequences may lay the foundation for future research or commercial software at national and international levels. If made available on clinical systems, they may enable more accurate evaluation of disease progression or treatment.

Along with accelerating image reconstruction using these stochastic algorithms, a number of additional software contributions were made to open source software. Primarily contributions were made to the STIR library, which is an open source Multi-Platform Object-Oriented framework for tomographic imaging that is used throughout academic institutions for medical imaging research purposes. Additionally, a pipeline was created to link a Monte Carlo medical imaging simulation software package (GATE) to STIR. The modifications to the STIR library and the development of this STIR-GATE-Connection project will have an impact on international academic studies and enable future research to productively perform accurate PET acquisition simulations for use in quantitative PET research. This may allow for the advancement of novel PET scanner design and physical model development.

Contents

List of Figures

Acronyms

^{18}F-FDG Fluorodeoxyglucose 22

AS AutoSubsets 11

BSREM Block Sequential Regularised EM 13

CaSTOR Customizable and Advanced Software for Tomographic Reconstruction 124

CSI Cosine Similarity Index 11

EM Expectation Maximisation 19

ET Emission Tomography 30

FBP Filtered Back Projection 51

FOV Field-Of-View 26

GATE Geant4 Application for Tomographic Emission 21

GRS Golden Ratio Subsets 11

HDF5 Hierarchical Data Format version 5 28

IAG Incremental Aggregated Gradient 85

keV kilo-electron volt 23

KKT Karush–Kuhn–Tucker 35

L-BFGS Limited Memory Broyden–Fletcher–Goldfarb–Shanno 39

LOR Line Of Response 11

MAP *Maximum a Posteriori* 20

MCIR Motion-Compensated Image Reconstruction 107

ML Maximum-Likelihood 19

MLEM Maximum-Likelihood Expectation Maximisation 36

MRI Magnetic Resonance Imaging 24

NRMSE Normalised Root Mean Squared Error 14

ns nano seconds 26

OS Ordered-Subsets 12

OSEM Ordered-Subsets Expectation Maximisation 44

PET Positron Emission Tomography 19

PML Penalised Maximum-Likelihood 19

PMT Photomultiplier Tubes 24

PS Paraboloidal Surrogate 38

ps pico seconds 26

QP Quadratic Penalty 32

RAMLA Row Action ML Algorithm 44

RB Random Batches 11

RDP Relative Difference Penalty 33

ROI Region of Interest 76

SAG Stochastic Average Gradient 49

SARAH StochAstic Recursive grAdient algoritHm 109

SGA Stochastic Gradient Ascent 48

SiPM Silicon Photomultiplier 24

SNR Signal-to-Noise Ratio 28

SPDHG Stochastic Primal-Dual Hybrid Gradient 74

SPS Separable Paraboloidal Surrogate 37

SSRB Single Slice Re-Binning 131

STIR Software for Tomographic Image Reconstruction 21

SUVpeak Standardised Uptake Value peak 78

SVR Stochastic Variance Reduction 21

SVREM Stochastic Variance Reduction EM 109

SVRG Stochastic Variance Reduction Gradient 49

TOF Time-of-Flight 26

TV Total Variation 32

Chapter 1

Introduction

Positron Emission Tomography (PET) is a medical imaging modality that measures functional and metabolic behaviour in-vivo and is commonly used in the staging of disease, and increasingly for monitoring of therapy and drug development. A small amount of a radioactive compound called a radiotracer is injected into a patient's body and biological processes distribute it throughout the body. The radioactive component decays into a positron that annihilates with an ambient electron, which results in a pair of emitted anti-parallel photons. The emitted photons can provide information regarding specific biochemical processes or blood flows. This information is challenging to obtain from other imaging modalities.

In brief, a PET scanner measures coincidence photon pairs that can be used to estimate the 3D radiotracer distribution for visualisation and quantification. However, this process is complicated by low measurement sensitivity, limitations on injected activity due to dose considerations and scan duration limitations. Statistical methods have been shown to generate desirable performance in the reconstruction of the tracer distribution due to their ability to model physical effects, e.g. attenuation and scanner geometries as well as the stochastic noise in the data. The earliest statistical approach was a Maximum-Likelihood (ML) estimation of the distribution [Rockmore et al. 1976] that was later estimated using a Expectation Maximisation (EM) algorithm [L. A. Shepp et al. 1982]. Many modifications to the ML method have been derived since, including Penalised Maximum-Likelihood (PML) methods. These penalised methods employ some *a priori* information into the statistical model. As an example, this information may encourage local smoothness over individual tissue or organs with strong edges at their boundaries.

1.1 Motivation

Iterative algorithms have become the standard reconstruction method used in clinical practice and remain a focus of academic research. These algorithms involve the cyclical application of a forward projection of the estimated distribution, a comparison to the measured data, and the backprojection of a correction that is used to update the estimate. However, this process for a fully 3D statistical reconstruction is computationally demanding. Although the power of computing systems is persistently increasing, PET image reconstruction will continually face computational challenges due to the increasing number of detectors used, particularly as dynamic PET reconstruction becomes more prevalent in clinical context.

Reconstruction acceleration is often achieved by the use of data subsets, which reduces the computational cost at each iteration of the cycle by using only part of the measured data. This results in an acceleration that is approximately linear in terms of the number of subsets used during early algorithm updates but the convergence rate will begin to decline as the estimate approaches the ML or *Maximum a Posteriori* (MAP) solution. Moreover, the use of subsets often leads to non-convergence, sometimes cycling between different solutions. Consequently, variations between subsequent images produced by these algorithms are observed, which may have adverse effects in quantitative medical evaluation.

1.2 Thesis Objective

The primary aims of this research are to accelerate PML PET image reconstruction and improve algorithm stability by limiting the impact of subset variance on the reconstruction algorithms. To achieve these goals, adaptive subset size and stochastic variance reduction optimisation algorithms are applied to PET reconstruction. Secondary supporting goals are to explore the impact of subset selection methodologies on the performance of various algorithms, to improve a pipeline between a Monte Carlo simulation software and PET reconstruction software, and to contribute towards various improvements of the aforementioned PET reconstruction software.

1.3 Thesis Overview

Chapter 2 includes a general background and overview of PET physics and PML subset image reconstruction algorithms. Additionally, examples of the Stochastic Variance Reduction (SVR) algorithms used in this thesis are described.

The AutoSubsets algorithm is presented in Chapter 3. This algorithm is designed to

be initialised with small subsets that automatically increase in size throughout the iterative process as two distinct subset update directions diverge. The algorithm is fully detailed and its performance is compared to standard fixed subset algorithms using both projection and list mode simulated data. However, the evaluation showed several limitations of the AutoSubsets methodology.

In Chapter 4, three Stochastic Variance Reduction (SVR) methods are implemented into a PET reconstruction framework. These algorithms are frequently applied in other fields of optimisation, including machine learning, and this chapter studies their adaptation to PET reconstruction, and compares their performance to standard subset PET algorithms using simulated data, acquired phantom data and a patient data set. Additionally, the impact of subset selection is investigated for PET image reconstruction.

Chapter 5 details the primary conclusions and achievements of this work along with possible directions of future work related to this thesis. Furthermore, a list of current and expected publications are provided.

Significant software development was involved in this Ph.D. project and is detailed in the first two Appendices. Appendix A details the STIR-GATE-Connection project. The project creates a research pipeline from Geant4 Application for Tomographic Emission (GATE), medical imaging Monte Carlo simulation software, data to the tomographic imaging software Software for Tomographic Image Reconstruction (STIR). This project was developed to compute a number of data sets in this work and to support future research. Appendix B highlights the major contributions to STIR.

Appendix C presents an analysis of introduction of biases in the stochastic variance reduction algorithm framework. Finally, following the AutoSubsets algorithm research presented in Chapter 3, additional studies that modified this algorithm were conducted by another student. These additional studies are described in Appendix D.

Chapter 2

Positron Emission Tomography and Iterative Optimisation

This chapter introduces the basic physics of positron emission followed by a brief overview of the measurement equipment used to acquire PET data. Subsequently, iterative optimisation principles are introduced, examples of iterative PET reconstruction algorithms are given and a discussion of various stochastic algorithms is presented.

2.1 Positron Emission Tomography

PET is an in-vivo non-invasive imaging modality used to observe functional and metabolic processes [Kumar et al. 2010]. PET reconstructions provide spatial information regarding the distribution density of a biological radiotracer. This information is used in medical disease staging and monitoring [Anand et al. 2009]. A radiotracer is a chemical compound where one or more of its atoms is a radionuclide. The radionuclides used in PET are positron emitting particles. The most commonly used radiotracer is Fluorodeoxyglucose (^{18}F-FDG) ($t_{1/2} = 109.8$min). It behaves similarly to glucose and therefore may be used to study glucose uptake in tissues. A commonly application of ^{18}F-FDG PET scans is the detection and quantification of lesions [Sah et al. 2017; Bollapragada et al. 2018]. Other radiotracer molecules exist and are employed for various in-vivo studies of biological mechanisms. For example, Rubidium-82 ($t_{1/2} = 76.4$s) is another radionuclide that is used as a reliable tool for the diagnosis of coronary arterial disease and microvascular disease [Chilra et al. 2017].

Positron emission and its subsequent annihilation is a physical process that results in the emission of two photons in (approximately) opposite directions. These photons may be detected with sophisticated external equipment. After multiple positron annihilations, a large data set of photon pair measurements is acquired. Using a variety of computational

methods, this measured data is reconstructed into an estimation of the tracer distribution that produced it. The physics of a positron emission and the resulting photon detection and coincidence measurements are the focus of the following sections.

2.1.1 Positron Emission Physics

Positron (e^+) emission typically occurs in 'proton-rich' radioisotopes, e.g., Carbon-11 [^{11}C], Oxygen-15 [^{15}O], and Fluorine-18 [^{18}F]. A proton in a nucleus decays into a neutron, positron e^+, and electron neutrino (ν_e) [Bailey et al. 2014; Conti et al. 2016]. The proton's decay leads to a change of proton number in the nucleus, e.g. Fluorine-18 ($t_{1/2} \approx 110$min) [Allisy-Roberts et al. 2008, p. 136] decays into Oxygen-18:

$$^{18}_{9}\mathrm{F} \rightarrow ^{18}_{8}\mathrm{O} + e^+ + \nu_e. \tag{2.1}$$

The positron is emitted from the nucleus with some kinetic energy and travels a short distance. The majority of its kinetic energy is deposited to nearby matter via Coulomb interactions as the positron travels [Bailey et al. 2014]. The average distance between emission and annihilation is known as positron range. As the positron transfers the majority of its kinetic energy, the likelihood of annihilation with an electron e^- increases. Once the positron and an electron annihilate, two 511 kilo-electron volt (keV) photons are emitted, which propagate in approximately opposite directions. This process is detailed in Figure 2.1. Some residual kinetic energy, associated with either the positron or electron prior to the annihilation, is responsible for a small directional perturbation to the anti-parallel photon emission.

In general, photons traversing a material may be attenuated, i.e., absorbed or scattered. The survival probability P of a photon travelling through a uniform medium without absorption or scattering is given by

$$P = e^{-\mu L}, \tag{2.2}$$

where μ is the linear attenuation coefficient of the medium and L is the length traversed. Photons are attenuated by either photoelectric absorption, Rayleigh scattering or Compton scattering [Bailey et al. 2014]. The former results in the photons ceasing to exist and the energy is absorbed by electrons, however, this rarely occurs for high energy photons. Rayleigh scattering is an elastic interaction as the photon loses essentially none of its energy but is scattered over a small angle. Compton scattering involves the photons transferring some

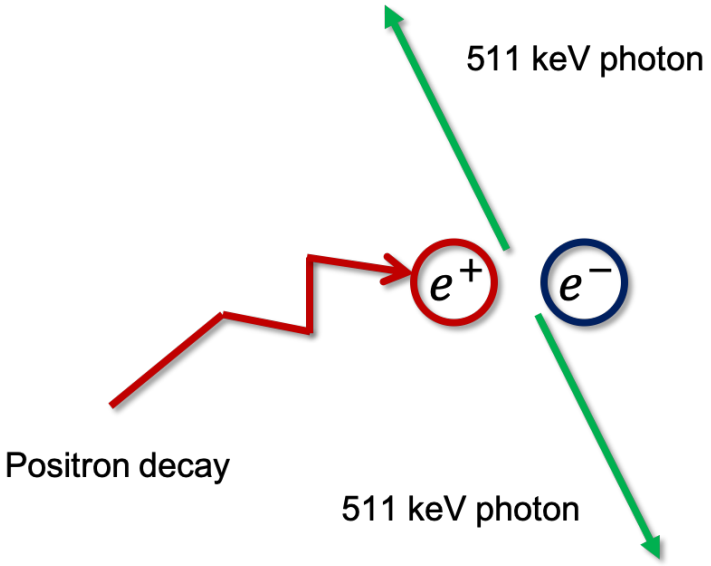


Figure 2.1: An illustration of positron-electron annihilation and the emission of a photon-pair in opposite directions.

energy to the attenuating material and a change occurs to the photon's direction of travel.

2.1.2 Photon Detection in PET

Clinical PET scanners are constructed from rings of photon detectors that are usually stacked to form a cylinder. A positron emitting radioactive object of interest (e.g. a patient or phantom) is placed within the centre of the cylinder. Photons emitted from this object are measured by the photon detectors of the scanner.

A scanner's photon detectors are comprised of scintillating crystals and a photodetector, such as Photomultiplier Tubes (PMT) or Silicon Photomultiplier (SiPM) [Pan et al. 2019; Meikle et al. 2021]. Scintillating crystals fully or partially absorb high energy photons (i.e. approximately 511 keV) and exhibit luminescence, which is the production of a cascade of many low energy photons. A photomultiplier tube is a photosensitive device that converts low energy photons, typically around the visible energy range such as those produced by the scintillating crystals, into a measurable electrical signal. Therefore, photon detectors convert photons into an electrical signal where spatial, temporal, and energy information may be measured and recorded. SiPMs can improve on the timing resolution of scanners and can also operate in a magnetic field that allows for usage in a Magnetic Resonance Imaging (MRI) [Vos et al. 2017]. Thus, SiPMs have replaced PMTs in the majority of current top-of-the-range commercial scanners PET systems.

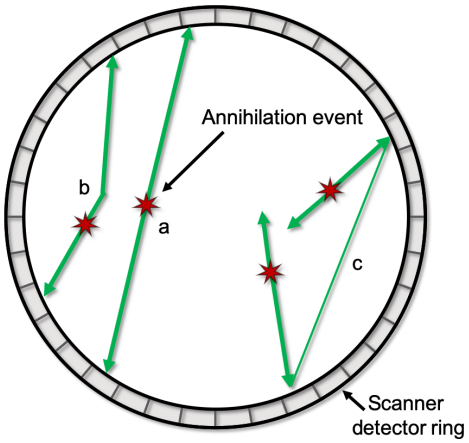


Figure 2.2: An illustration of different types of photon coincidence events. a) The detection of a pair of photons from a common annihilation event. b) The detection of a pair of photons but one photon has been scattered. c) A coincidence event corresponding to detection of a photon pair from two different annihilations

A pair of gamma photons interacting with two different photon detectors within a short period of time is recorded as a coincidence event by a PET scanner. For any given coincidence event, one might assume that the two detected photons were emitted from a common positron annihilation and were emitted in opposite directions. Thus, the annihilation is assumed to have occurred along the straight line connecting the two detectors, known as a LOR. In Figure 2.2, a) illustrates the measurement of two photons from a common annihilation event and are therefore a true coincidence event. While this assumption provides a good approximation, several physical factors may falsify this assumption, i.e., scattered and random coincidence events.

As aforementioned in Section 2.1.1, photons scattered by an attenuating material will travel along an altered path [Allisy-Roberts et al. 2008, p. 134-135]. Figure 2.2.b depicts two photons detected within the coincidence window but one photon was scattered prior to detection. These coincidence measurements are known as scattered events and the assigned LOR will not accurately represent the path the two photons traversed. Likely, this LOR will not overlay the photon emission location. However, photons may lose energy when scattered. Some PET scanners may measure the incidence photon energy deposited into the crystal and reject events that are outside a predetermined energy range.

The third type of coincidence measurement is a random coincidence. This type occurs when two photons, which are emitted from different annihilation events, are designated as a coincidence. Random (coincidence) events are illustrated by Figure 2.2.c [Allisy-Roberts

et al. 2008, p. 134-135]. The resulting LOR will not overlay either of the emissions.

It is commonly assumed that a true coincidence LOR intersects with the site of the positron emission. However, positron range makes it unlikely that the LOR will correctly intercept the positron's emission location. Furthermore, the non-colinearity, due to the small angular deviation from the anti-parallel photon emission, may result in further misalignment between the LOR and the true positron emission location [Iriarte et al. 2016].

The coincidence timing window is approximately 5 nano seconds (ns) in modern scanners [Bettinardi et al. 2011]. However, in modern systems the timing resolution Δt of a scanner may be in the order of a few hundred pico seconds (ps). This allow for further localisation along the LOR by measuring the difference between incident times of the two photons at their respective detectors. The spatial uncertainty ΔL along the LOR is given by

$$\Delta L = \frac{c\Delta t}{2}, \tag{2.3}$$

where $c \approx 3 \times 10^8$ ms^{-1} is the speed of light in a vacuum and Δt is the discrepancy between the two measured photon detection times. This allows for Time-of-Flight (TOF) measurements. This data is commonly used in commercial clinical PET.

2.1.3 Data Acquisition and Storage

Millions of coincidence events may be recorded during the course of a typical PET scan with spatial, temporal and energy information measured for each photon-pair.

2.1.3.1 Projection Data

Traditionally, this measured data may be stored in sets of sinograms. Sinograms are 2D matrices (N_s by N_θ) used to represent the number of coincidence events measured between various detector pairs [Bailey et al. 2014]. Each row of a sinogram corresponds to a set of N_θ parallel LORs connecting detector pairs around a scanner's ring [Allisy-Roberts et al. 2008, p. 135-136]. These parallel LORs/projection angles θ are often known as scanner views and are evenly spaced over 180° for most cylindrical PET systems. The N_s columns of a sinogram correspond to the off-centre distance s of the LORs [Allisy-Roberts et al. 2008, p. 135-136]. These are non-equidistant displacements that decrease towards the edges of the Field-Of-View (FOV) due to the ring's arc, which is taken into account during a reconstruction [Bailey et al. 2014].

Modern PET scanners stack rings next to one another to form a cylinder of N_R detector

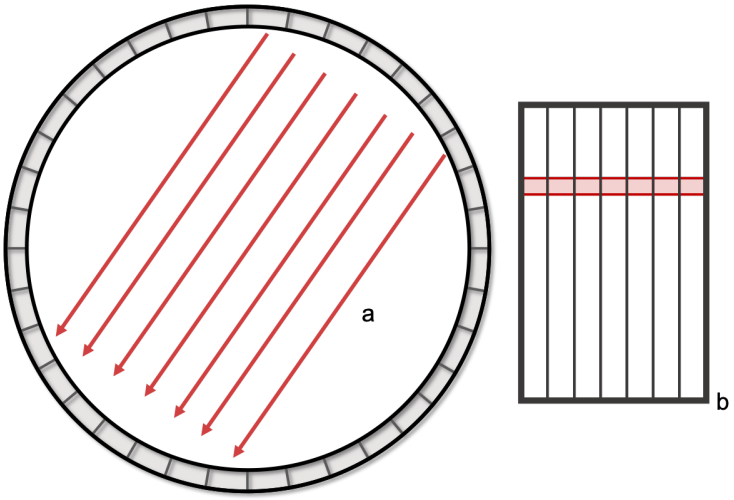


Figure 2.3: An illustration of a scanner view and the corresponding sinogram row. a) A set of parallel LORs drawn inside a scanner. b) A 2D sinogram matrix and the highlighted row corresponds to the set of parallel LORs shown in a).

rings, allowing for a variation of acquisition modalities. Historically, stacked ring systems measured data in 2D slices with photon coincidence pairs measured by detectors in the same ring. Septa were introduced between rings to shield out-of-plane photons from entering the detectors and reduce the number of random coincidence events measured [Bailey et al. 2014].

Modern PET scanners collect data in 3D and do not incorporate the septa into their design. This allows for inter-ring coincidence event measurements with a maximum ring difference of $N_R - 1$ that results in a maximum of N_R^2 sinograms. Moreover, the number of coincidence events recorded increases significantly [Bailey et al. 2014]. Although acquiring coincidence measurements from oblique angles increases a scanner's sensitivity and reduces statistical noise in reconstructed images, it increases the coincidence scatter fraction [Allisy-Roberts et al. 2008, p. 136]. However, highly oblique sinograms generally record fewer counts because of solid angle effects and longer distances for photons to travel, increasing the chance of attenuation.

In this work, this 3D data is referred to as projection data, these are collections of sinograms at various oblique angles. Furthermore, TOF information may be encoded in an extra dimension of the projection data set.

2.1.3.2 List Mode Data

List mode is an alternative data storage method whereby each coincidence event measured, and its properties, are stored as a unique entry. As millions of events are measured in a

typical PET scan, list mode files can become large and data compression is often required. The exact data structure of a list mode data set can vary throughout the PET field. Example file formats are ROOT data [Brun et al. 1997] and Hierarchical Data Format version 5 (HDF5) [Wadhwa et al. 2018; Wadhwa et al. 2021].

List mode data may be reconstructed directly or processed post-scan into a projection data format for reconstruction [Efthimiou et al. 2019]. In this thesis, a list mode data set is considered to be a set of coincidence events $\{e_l\}_{l=1}^{N_e}$, where N_e is the number of events and l is an index. Each event e_l encodes the LOR position along with other information, e.g., detection time, energy and TOF information. While both ROOT and HDF5 list mode PET data are used in this thesis, list mode reconstruction is only conducted using ROOT data sets, which are read by the STIR library [Thielemans et al. 2012]. See Section 2.3.4.4 for more information regarding list mode reconstruction. HDF5 list mode data sets, acquired by GE scanners, are used in this work but the data is converted into STIR compatible interfile projection data for image reconstruction, see Sections 4.2.3 and 4.2.4.

2.2 Basis of Iterative Reconstruction

Iterative PET reconstruction attempts to obtain an estimate of the unknown tracer distribution $\boldsymbol{x}$ that best represents the measured data $\boldsymbol{y}$ [L. A. Shepp et al. 1982]. However, the presence of noise complicates this estimation. Analytical methods, based on mathematical inversion, exist to approximate the tracer distribution but they do not incorporate any model of the noise [Bailey et al. 2014].

Iterative PET reconstruction methods rely on physical modelling factors to model noise in the measured data [Bailey et al. 2014]. With the use of these models, iterative methods typically supersede analytical methods with improved tracer distribution estimate spatial resolutions and Signal-to-Noise Ratios (SNRs) while maintaining quantitative results [Iriarte et al. 2016]. However, these improvements come at the cost of significantly increased computation requirements. Yet, with the advent of more powerful computers and the general statistical improvements gained, iterative methods have become the primary reconstruction methodology used in clinical practice. Reconstruction may still be slow for large data sets and this thesis aims to address this issue.

In general, iterative algorithms cyclically compare the current estimate of the distribution to the measured data and compute a quantitative update to the estimate [Erdoğdu 2019]. Ideally, this updated distribution better represents the measured data. Many varia-

tions of this general concept exist, including some that: incorporate *a priori* information, accelerate the reconstruction and/or use a better model of various physical processes. The remainder of this section provides additional details of iterative methods. Specific focus is given to current iterative PET reconstruction methods, regularisation, subsets and stochastic optimisation algorithms.

2.2.1 The Forward Model

Iterative PET reconstruction is model-based and requires the application of a system model $\boldsymbol{A} \in \mathbb{R}_{\geq 0}^{N_v \times N_b}$ to map between distribution estimate space $\mathbb{X} \in \mathbb{R}_{\geq 0}^{N_v}$ and measured projection data space $\mathbb{Y} \in \mathbb{R}_{\geq 0}^{N_b}$, with N_v and N_b representing the number of image voxels and number of projection data elements (bins) respectively [Qi et al. 2006]. Note, $\mathbb{R}_{\geq 0}$ denotes non-negative real numbers. The system matrix $\boldsymbol{A}$ models the probability of an un-scattered emission from the i^{th} voxel of the unknown distribution being detected by the j^{th} projection data element. Therefore, the reconstruction problem may be approximately reduced to solving $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ for $\boldsymbol{x}$, where $\boldsymbol{x} \in \mathbb{X}$ and $\boldsymbol{y} \in \mathbb{Y}$. PET measures the number of photon coincidence events on a set of detectors, hence $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{A}$ are positive semi-definite.

The system matrix $\boldsymbol{A}$ may be designed to model physical properties of the PET system in a number of stages, e.g. using the product of several matrices, given by

$$\boldsymbol{A} = \boldsymbol{A}_{det.sens.}\boldsymbol{A}_{attenuation}\boldsymbol{A}_{for.proj.}\boldsymbol{A}_{positron}, \tag{2.4}$$

where: $\boldsymbol{A}_{det.sens.}$ models the intrinsic detector sensitivity, $\boldsymbol{A}_{attenuation}$ is an attenuation model, $\boldsymbol{A}_{for.proj.}$ models the geometric forward projection factors associated with the PET scanner and $\boldsymbol{A}_{positron}$ models the positron range [Iriarte et al. 2016].

The presence of noise complicates this measured data model. Therefore, additional data models are incorporated into the system model, i.e., expected scatter events $\boldsymbol{s} \in \mathbb{Y}$ and expected random events $\boldsymbol{r} \in \mathbb{Y}$, see Section 2.1.2. The affine forward model of the statistical mean $\bar{\boldsymbol{y}}(\boldsymbol{x})$ is given by

$$\bar{\boldsymbol{y}}(\boldsymbol{x}) = \mathbb{E}[\boldsymbol{y}] = \boldsymbol{A}\boldsymbol{x} + \bar{\boldsymbol{b}}, \tag{2.5}$$

where $\bar{\boldsymbol{b}} = \bar{\boldsymbol{s}} + \bar{\boldsymbol{r}}$ are the expected background events [Fessler et al. 1999; Qi et al. 2006]. Each individual projection data element is computed as $\bar{y}_j(\boldsymbol{x}) = \boldsymbol{A}_j\boldsymbol{x} + \bar{b}_j$.

2.2.2 Likelihood Function

Iterative methods require a measure of performance to optimise, known as an objective function $\Phi(\boldsymbol{x})$. The objective function provides a quantitative comparison between the measured data $\boldsymbol{y}$ and the expected data $\bar{\boldsymbol{y}}(\boldsymbol{x})$. With no regularisation, the function is considered optimised when $\boldsymbol{y}$ is best represented by $\bar{\boldsymbol{y}}(\boldsymbol{x})$.

In Emission Tomography (ET), a Poisson model is employed, which allows for a modelling of noise [Qi et al. 2006]. The Poisson model is given by

$$P(\boldsymbol{y}|\bar{\boldsymbol{y}}(\boldsymbol{x})) = \prod_{j\in\mathcal{J}} e^{-\bar{y}_j(\boldsymbol{x})} \frac{\bar{y}_j(\boldsymbol{x})^{y_j}}{y_j!}, \tag{2.6}$$

where $\mathcal{J}$ represents all the measured data indices. This probability function is concave with respect to $\boldsymbol{x}$ and the maximiser is known as the ML estimate. The natural logarithm of Equation (2.6), dropping terms independent of $\boldsymbol{x}$, leads to an expression of the log-likelihood function of $\boldsymbol{x}$ given by

$$L(\boldsymbol{y}|\boldsymbol{x}) = \sum_{j\in\mathcal{J}} y_j \log(\bar{y}_j(\boldsymbol{x})) - \bar{y}_j(\boldsymbol{x}). \tag{2.7}$$

As log is a monotonic function, the ML solution is also realised by maximising $L(\boldsymbol{y}|\boldsymbol{x})$.

The first and second partial derivatives of the log-likelihood function are given by

$$\nabla L(\boldsymbol{y}|\boldsymbol{x})_i = \frac{\partial L(\boldsymbol{y}|\boldsymbol{x})}{\partial x_i} = \sum_{j\in\mathcal{J}} a_{ji}\left(\frac{y_j}{\sum_k a_{jk}x_k + \bar{b}_j} - 1\right) \tag{2.8a}$$

$$\nabla^2 L(\boldsymbol{y}|\boldsymbol{x})_{i,l} = \frac{\partial^2 L(\boldsymbol{y}|\boldsymbol{x})}{\partial x_i \partial x_l} = -\sum_{j\in\mathcal{J}} \frac{a_{ji}a_{jl}y_j}{(\sum_k a_{jk}x_k + \bar{b}_j)^2}. \tag{2.8b}$$

Both equation (2.6) and equation (2.7) are concave functions for all $\boldsymbol{x} \in \mathbb{X}$ as the second derivative (Hessian) is negative semi-definite [Ahn et al. 2003; Qi et al. 2006]. Several PET reconstruction algorithms attempt to optimise the objective function $\Phi(\boldsymbol{x}) = L(\boldsymbol{y}|\boldsymbol{x})$ to identify the ML solution, see Section 2.3. Modified log-likelihood functions may be required for global concavity and convergence of some of the later discussed iterative algorithms [Ahn et al. 2003].

2.2.3 Penalty Function

The previously discussed PET problem is ill-posed. This results in image noise amplification and an estimated distribution that is not representative of the truth [Leahy et al. 2000]. An optimisation algorithm may be regularised with the addition of a weighted penalty term

into the objective function to introduce *a priori* information and/or early stopping [Arridge et al. 2019]. The former of these regularisation methods is the focus in this work and may be used to reduce noise properties and/or incorporate some anatomical structure into the distribution [Webster Stayman et al. 2000; Ehrhardt et al. 2016]. Additionally, other regularisation methods exist, such as the kernel reconstruction method and post-filtering but these are not considered in this work [L. A. Shepp et al. 1982; Wang et al. 2015].

The objective function $\Phi(\boldsymbol{x})$ may be formulated as a weighted sum of the log-likelihood function $L(\boldsymbol{y}|\boldsymbol{x})$ and a penalty function $R(\boldsymbol{x})$, given by

$$\Phi(\boldsymbol{x}) = L(\boldsymbol{y}|\boldsymbol{x}) - \beta R(\boldsymbol{x}), \tag{2.9}$$

where β is a scaling factor between the likelihood term and penalty. Assuming convexity of $R(\boldsymbol{x})$, the optimiser of the objective is known as the MAP solution, which has some bias from the ML solution.

Priors typically incorporate some *a priori* information. An intuitive example of such information for PET image reconstruction is that the tracer uptake is approximately uniform over a given tissue volume but abrupt changes may be found at the boundaries [Qi et al. 2006]. Therefore, penalty functions are often designed with two primary, yet conflicting, objectives: noise suppression via local smoothing and retention of edge structures in the image. These are local regularisation properties that are commonly enforced by comparing a voxel to other voxels in its local neighbourhood $\mathcal{N}$ with some energy function $\rho(x_i, x_l)$. Hence, the regularisation term $R(\boldsymbol{x})$ in Equation (2.9) is often given as

$$R(\boldsymbol{x}) = \sum_{i=1}^{N_v} \sum_{l \in \mathcal{N}_i} \kappa_i \kappa_l w_{i,l} \rho(x_i, x_l), \tag{2.10}$$

where $\mathcal{N}_i$ represents a neighbourhood of voxels about the i^{th} voxel, κ_i and κ_l are spatially variant scalar values for the i^{th} and l^{th} voxels and $w_{i,l}$ is the neighbourhood weighting between voxels i and l [Geman et al. 1984].

2.2.3.1 Spatially Variant Penalty Strengths

Selection of the aforementioned weighting variables $w_{i,j}$ can be constant or vary between voxel indices. A spatially variable penalty strength, controlled by κ_i and κ_l, can reduces the dependence of local perturbation on surrounding activity and location. An example of this

is $\kappa(\boldsymbol{x})$, which is given by

$$\kappa(\boldsymbol{x}) \triangleq \sqrt{\boldsymbol{A}^\top \text{diag}\left[\frac{\boldsymbol{y}}{\bar{\boldsymbol{y}}^2(\boldsymbol{x})}\right]\boldsymbol{A}\mathbf{1}}, \tag{2.11}$$

where $\kappa \in \mathbb{X}$ [Tsai et al. 2020]. [i] Intuitively, these weights encourage greater usage of the prior when little measured data is available for the given voxel while discouraging its usage when data is present. Furthermore, the usage of these spatially variant penalty strength weights can improve convergence rates of regions of interest within data sets.

2.2.3.2 Example Potential Functions

The **Quadratic Penalty (QP)** is popular potential function, given by

$$\rho_{\text{QP}}(x_i, x_l) = \frac{(x_i - x_l)^2}{2}. \tag{2.12}$$

This potential reduces image noise by smoothing local regions but also reduces spatial resolution [Somayajula et al. 2005].

Total Variation (TV) is a regularisation function used in optimisation that acts on the gradient of $\boldsymbol{x}$ [Rudin et al. 1992; Ehrhardt et al. 2019]. The TV penalty given by

$$\rho_{\text{TV}}(x_i, x_l) = |x_i - x_l|, \tag{2.13}$$

and is commonly used in denoising images and the neighbourhood is only comprised of voxels adjacent to the i^{th}. However, the TV penalty is non-differentiable and non-smooth. Optimisation using this penalty requires proximal operators [Parikh et al. 2014] or modification to smooth the function, e.g. the Huber penalty [Nikolova et al. 2001], given by

$$\rho_{\text{H}}(x_i, x_l) = \begin{cases} \frac{(x_i - x_l)^2}{2} & \text{if } |x_i - x_l| \leq \upsilon \\ \upsilon|x_i - x_l| - \upsilon^2/2, & \text{otherwise,} \end{cases} \tag{2.14}$$

for some $\upsilon > 0$.

Log-cosh is another penalty that has been previously used in PET PML reconstruction

[i] Utilities to compute both this κ and an alternative lower cost approximation were added to the STIR library as part of this work, see Section B.

[Green 1990]. The penalty, given by

$$\rho_{\mathrm{LC}}(x_i, x_l) = \upsilon^2 \log\cosh\left(\frac{x_i - x_l}{\upsilon}\right), \tag{2.15}$$

acts similarly to the Huber function with the trade-off between linear (edge-preservation) and quadratic (smoothing) controlled by their respective $\upsilon > 0$ values.

Relative Difference Penalty (RDP)[ii] is given by

$$\rho_{\mathrm{RDP}}(x_i, x_l) = \frac{(x_i - x_l)^2}{(x_i + x_l) + \upsilon|x_i - x_l| + \varepsilon}, \tag{2.16}$$

where $\upsilon \geq 0$ is used to control edge preservation and $\varepsilon \geq 0$ [Ahn et al. 2015; Nuyts et al. 2002]. The preservation of strong edges improves quantification while also providing visually appealing images [Asma et al. 2012a; Asma et al. 2012b]. The RDP is implemented into the Q.Clear reconstruction algorithm deployed by General Electric on its recent PET scanners [Ross 2014].

The $\varepsilon > 0$ term is a modification to the RDP and is included in this work because of the function's non-smooth behaviour when either $x_i = 0$ and $x_l \to 0$, or $x_l = 0$ and $x_i \to 0$. The selection of ε is heuristic but, if small enough, its impact on the estimated distribution is negligible.

Many of the presented energy functions can be written as $\rho(|x_i - x_j|)$, except for the RDP. This makes the RDP incompatible with many surrogate algorithms that are discussed later in this text.

2.2.4 Objective Function Properties

The objective functions in this thesis are a composite of the concave log-likelihood data fit and convex regularisation, given by Equation (2.9). As a result a unique solution $\hat{\boldsymbol{x}} \in \mathbb{X}$ exists, given some assumptions regarding $\bar{\boldsymbol{y}} > 0$ if $\boldsymbol{y} > 0$. The function is also lower bounded by $x_i \geq 0$ because $\boldsymbol{x} \in \mathbb{X}$. A theoretical upper bound $U = U(\boldsymbol{y}) \in (0, \infty)$ is described by Ahn et al. 2003 that allows for the solution to exist in a bounded set $\mathbb{B}$. Additional modifications were proposed for the log-likelihood function, given by Equation (2.7), that allows for the handling of instances when $\bar{y}_j = 0$ and $y_j > 0$ for some j [iii]. However, this only occurs when $\bar{b}_j = 0$ and $\boldsymbol{A}_j \boldsymbol{x} = 0$, which is not realistic for PET reconstruction [Qi et al. 2006].

[ii] During this work, both the Logcosh prior and RDP were added to STIR, including function-gradient consistency tests, see Appendix B

[iii] The adopted convention is that $0\log(0) = 0$ and $\log(0) = -\infty$

With these modification to the log-likelihood, $\Phi(\boldsymbol{x})$ is Lipschitz continuous on $\mathbb{B}$ if there exists some $\mathcal{L} > 0$, which satisfies

$$\frac{\|\Phi(\boldsymbol{x}) - \Phi(\boldsymbol{x}^\dagger)\|}{\|\boldsymbol{x} - \boldsymbol{x}^\dagger\|} \leq \mathcal{L}, \quad \forall \boldsymbol{x}, \boldsymbol{x}^\dagger \in \mathbb{B}. \tag{2.17}$$

$\Phi(\boldsymbol{x})$ is differentiable over $\mathbb{B}$, the derivatives are bounded and $\nabla\Phi(\boldsymbol{x})$ is Lipschitz continuous [Ahn et al. 2003].

2.3 Optimisation

This section describes methods of objective function optimisation and is divided into subsections of update direction, step size, and subsets methods via the splitting of the objective function.

2.3.1 General Objective Function Optimisation

Maximisation of an objective function $\Phi(\boldsymbol{x})$ may be written as

$$\hat{\boldsymbol{x}} = \underset{\boldsymbol{x} \in \mathbb{X}}{\operatorname{argmax}}\, \Phi(\boldsymbol{x}). \tag{2.18}$$

A global maximiser $\hat{\boldsymbol{x}}$ satisfies

$$\Phi(\hat{\boldsymbol{x}}) \geq \Phi(\boldsymbol{x}), \quad \forall \boldsymbol{x} \in \mathbb{X}. \tag{2.19}$$

If the objective function is strictly concave, then there is a unique global maximiser [Nocedal et al. 2006]. If $\Phi(\boldsymbol{x})$ is twice continuously differentiable, the functions gradient $\nabla\Phi(\boldsymbol{x})$ and Hessian $\nabla^2\Phi(\boldsymbol{x})$ may be evaluated [Nocedal et al. 2006]. The second-order Taylor expansion of the objective function is given by

$$\Phi(\boldsymbol{x} + \boldsymbol{p}) \approx \Phi(\boldsymbol{x}) + \boldsymbol{p}^T \nabla\Phi(\boldsymbol{x}) + \frac{1}{2}\boldsymbol{p}^T \nabla^2\Phi(\boldsymbol{x})\boldsymbol{p}, \tag{2.20}$$

where $\boldsymbol{p} \in \mathbb{R}^n$ is some small perturbation to $\boldsymbol{x}$. Ignoring any constraints, a local maximiser is a stationary point, i.e. $\nabla\Phi(\hat{\boldsymbol{x}}) = \boldsymbol{0}$ [Nocedal et al. 2006]. In a system constrained to non-negative space, like in the PET problem, the maximiser must satisfy the

Karush–Kuhn–Tucker (KKT) conditions [Erdoğdu 2019], given by

$$\nabla\Phi(\hat{\boldsymbol{x}})_i = 0, \text{if } \hat{\boldsymbol{x}}_i > 0 \tag{2.21a}$$

$$\nabla\Phi(\hat{\boldsymbol{x}})_i \leq 0, \text{if } \hat{\boldsymbol{x}}_i = 0. \tag{2.21b}$$

These conditions imply that it is possible for $\|\nabla\Phi(\hat{\boldsymbol{x}})\|_2 > 0$ but voxels are at the boundary.

Iterative optimisation algorithms generate a sequence of iterates $\{\boldsymbol{x}_k\}_{k=0}^{K}$ from some initial guess $\boldsymbol{x}_0$ [Erdoğdu 2019]. A sequence of estimates that increase the value of $\Phi(\boldsymbol{x})$ at each update is sought but not always guaranteed. An algorithm may be terminated after a sufficient increase of Φ has been achieved or $\hat{\boldsymbol{x}}$ has been closely approximated, e.g., after K iterates.

2.3.2 Ascent Direction of Differentiable Concave Functions

Generally, sequential iterates are generated by

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k, \tag{2.22}$$

where α_k is a step size and $\boldsymbol{p}$ is a search direction or vector. This vector is said to be an ascent direction of $\Phi(\boldsymbol{x})$ if there exists some $\delta > 0$ that satisfies

$$\Phi(\boldsymbol{x} + \alpha\boldsymbol{p}) \geq \Phi(\boldsymbol{x}), \quad \forall\alpha \in (0, \delta). \tag{2.23}$$

For any differentiable concave function, an ascent direction satisfies

$$\frac{\nabla\Phi(\boldsymbol{x})^T \boldsymbol{p}}{|\nabla\Phi(\boldsymbol{x})\|_2 \|\boldsymbol{p}\|_2} = \cos(\theta) > 0, \quad \text{if} \|\nabla\Phi(\boldsymbol{x})\|_2 > 0, \tag{2.24}$$

where $\theta \in [0, \pi]$, and $\|\cdot\|_2$ is the l^2-norm [Nocedal et al. 2006]. Thus, the simplest method to determine a suitable ascent direction is steepest gradient ascent, where $\boldsymbol{p}_k = \nabla\Phi(\boldsymbol{x}_k)$, or its normalised form $\boldsymbol{p}_k = \nabla\Phi(\boldsymbol{x}_k)/|\nabla\Phi(\boldsymbol{x}_k)|$ [Bailey et al. 2014].

One complication of Poisson model gradient ascent PET algorithms is the requirement for $\boldsymbol{x} \in \mathbb{X}$, defined in Section 2.2.1. Therefore, a non-negativity constraint my be required for iterative algorithms, which may be practically enforced after each update as a projection operator $P_{\boldsymbol{x}\geq 0}[\cdot]$ that projects from $\mathbb{R}^{N_v} \to \mathbb{X}$. Thus the constrained gradient ascent equation

is given by

$$\boldsymbol{x}_{k+1} = P_{\boldsymbol{x}\geq 0}[\boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k]. \tag{2.25}$$

2.3.2.1 Preconditioning

The performance of steepest gradient ascent is limited by the shape and scaling of $\Phi(\boldsymbol{x})$. A function is poorly scaled if a perturbation to $\boldsymbol{x}$ in a direction results in a larger change in the value of Φ than a perturbation in an alternative direction [Nocedal et al. 2006]. Steepest gradient ascent optimisation of the log-likelihood function is highly sensitive to poorly scaled objectives and small α_k values are required for convergence.

A diagonal positive definite preconditioner may be used to re-scale the optimisation problem by weighting the parameters so perturbations in all directions lead to similar changes in $\Phi(\boldsymbol{x})$. This results in the option to use larger α_k values. A preconditioned steepest ascent update equation may be written as

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{D}(\boldsymbol{x}_k)\nabla\Phi(\boldsymbol{x}_k), \tag{2.26}$$

where $\boldsymbol{D}(\boldsymbol{x}_k)$ is a preconditioning matrix, which may be parameterised by $\boldsymbol{x}_k$. Generally, this matrix is positive definite because of the requirement to maintain an ascent direction, i.e., $\boldsymbol{p}_k = \boldsymbol{D}(\boldsymbol{x}_k)\nabla\Phi(\boldsymbol{x}_k)$ satisfying Equation (2.24).

An example of a preconditioner used in iterative PET reconstruction is the diagonal EM preconditioner. This preconditioner is attained from the Maximum-Likelihood Expectation Maximisation (MLEM) algorithm and is given by

$$\boldsymbol{D}_{\text{EMoriginal}}(\boldsymbol{x}) = \text{diag}\left\{\frac{\boldsymbol{x}}{\boldsymbol{A}^\top \boldsymbol{1}}\right\}, \tag{2.27}$$

where the division is element-wise, $\boldsymbol{A}^\top \boldsymbol{1}$ is the back projection of a uniform vector of ones and $\text{diag}\{\cdot\}$ is an operator that converts a vector into a diagonal array [L. A. Shepp et al. 1982; Ahn et al. 2003]. The MLEM update formula is given by

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \frac{\boldsymbol{x}_k}{\boldsymbol{A}^\top \boldsymbol{1}}\nabla L(\boldsymbol{y}|\boldsymbol{x}_k), \tag{2.28}$$

where $\alpha_k = 1$, the EM preconditioner $\boldsymbol{D}_{\text{EMoriginal}}(\boldsymbol{x}_k)$ is used and $\Phi(\boldsymbol{x}_k) = L(\boldsymbol{y}|\boldsymbol{x}_k)$ (i.e., $\beta = 0$) [Lewitt et al. 1986]. A property of the MLEM algorithm is that it naturally constrains $\boldsymbol{x} \in \mathbb{X}$. While the MLEM algorithm guarantees convergence to the ML solution, it is practically

slow and requires significant computation to do so [Qi et al. 2006]. Furthermore, the ML solution of a PET reconstruction generally exhibits high noise properties and the resulting images are often not clinically desirable.

As a consequence of the non-negativity constraint, voxels may have value 0 and thus the original EM preconditioner does not allow for any future change, regardless of the value of $\nabla\Phi(\boldsymbol{x})_i$. A simple modification that ensures the aforementioned diagonal EM preconditioner is positive definite is given by

$$\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}) = \text{diag}\left\{\frac{\boldsymbol{x}+\delta}{\boldsymbol{A}^\top\boldsymbol{1}}\right\}, \tag{2.29}$$

where δ is a small positive vector. For convenience, this modified EM preconditioner shall be referred to as $\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_k)$ throughout its usage in this work. This modification does not have the same convergence proof for MLEM but the impact of the additional δ is expected to be negligible if δ is small enough[iv].

Another preconditioner that has been used in PET reconstruction relates to the Separable Paraboloidal Surrogate (SPS) algorithm, somewhat described in Section 2.3.2.2. This preconditioner takes the form

$$\boldsymbol{D}_{\text{SPSmod}}(\boldsymbol{x}_{\text{init}}) = \text{diag}\left\{\boldsymbol{A}^\top\text{diag}\left\{\frac{\boldsymbol{y}}{(\boldsymbol{A}\boldsymbol{x}_{\text{init}}+\bar{\boldsymbol{b}})^2}\right\}\boldsymbol{A}\boldsymbol{1}+\beta\nabla^2 R(\boldsymbol{x}_{\text{init}})\boldsymbol{1}\right\}^{-1}, \tag{2.30}$$

which is the row-sum of the Hessian of the objective at an early image estimate $\boldsymbol{x}^{\text{init}}$ and $\nabla^2 R(\boldsymbol{x})$ is the second derivative of the penalty function [Ahn et al. 2003; Tsai et al. 2016; Tsai et al. 2018] This preconditioner can be precomputed but requires three (two forward and one backward) projection operations. Additionally, algorithm performance will vary with the selection of $\boldsymbol{x}^{\text{init}}$. A lower computational cost version also exists, given by

$$\boldsymbol{D}_{\text{SPS}} = \text{diag}\left\{\boldsymbol{A}^\top\text{diag}\left\{\frac{\boldsymbol{1}}{\boldsymbol{y}+\boldsymbol{1}}\right\}\boldsymbol{A}\boldsymbol{1}\right\}^{-1}, \tag{2.31}$$

but the division by $\boldsymbol{y}$ makes this preconditioner sensitive to low count measured data. Both these preconditioners are related to the $\kappa(\boldsymbol{x})$ computation in Equation (2.11) but with the

[iv] Alternative methodologies forgo this EM preconditioner modification and instead project of voxels $x_i \leq 0$ to a small positive value δ value [Ahn et al. 2003]. However, this requires special handling of step sizes and is not considered in this work.

additional regularisation term.

2.3.2.2 Surrogate Methods

Thus far, the presented methodologies have highlighted ascent direction computations using preconditioned gradient ascent class algorithms. "Functional substitution", "optimisation transfer" or "surrogate methods" are an alternative to gradient ascent. These algorithms monotonically increase the objective function, accommodate the non-negativity constraint and ensure convergence to $\hat{\boldsymbol{x}}$ [Fessler et al. 1999; Qi et al. 2006]. Generally, these algorithms construct a paraboloidal surrogate function of the objective function at each iteration from the current $\boldsymbol{x}_k$. This surrogate function lies below the objective function for $\boldsymbol{x} \in \mathbb{X}$. This function, which may be quadratic (e.g. Paraboloidal Surrogate (PS) [Fessler et al. 1999]), is easily maximised to determine the next iteration $\boldsymbol{x}_{k+1}$. Surrogate algorithms can optimise a regularised objective function but the penalty function requires an analytical surrogate [De Pierro 1995]. Of the penalty functions presented in Section 2.2.3, the QP and log-cosh penalty functions have known surrogate functions[v].

The aforementioned MLEM algorithm was originally derived as a surrogate method [Lange et al. 1984; L. A. Shepp et al. 1982; Qi et al. 2006]. The PS method is a quadratic surrogate algorithm that provides a local functional approximation of the objective function about $\boldsymbol{x}_k$, which allows convergence of penalised objective functions [Fessler et al. 1999]. An analysis of global convergence of such methods using convex penalty functions for PET reconstruction has been presented [Ahn et al. 2003]. A primary limitation of surrogate methods is their slow convergence to the solution if the surrogate does not represent the function well.

2.3.2.3 Second Order Optimisation

A method for determining an improved ascent direction is Newton's method, which is derived from the second order Taylor expansion approximation, given by Equation (2.20). Assuming the strictly concavity of Φ, the Newton ascent direction is obtained by identifying the stationary point of the Taylor expansion [Nocedal et al. 2006]. The ascent direction

[v] A parabolic surrogate function for the log-cosh penalty function was added to STIR as part of this work, see Section B.

is thus given by

$$0 = \nabla\Phi(\boldsymbol{x}) + \nabla^2\Phi\boldsymbol{p} \tag{2.32a}$$

$$\boldsymbol{p} = -(\nabla^2\Phi)^{-1}\nabla\Phi(\boldsymbol{x}) \tag{2.32b}$$

and the Newton update by

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}^k + \alpha_k(\nabla^2\Phi_k)^{-1}\nabla\Phi(\boldsymbol{x}_k). \tag{2.33}$$

Newton's method generally accelerates convergence compared to standard first order gradient ascent [Nocedal et al. 2006]. This Newton ascent equation may be considered a method of preconditioned gradient ascent, due to the application of an operator upon $\nabla\Phi(\boldsymbol{x})$ that results in better conditioned update steps. This preconditioner is non-diagonal and is positive definite. Experimental evidence demonstrates that Newton's method performs best with $\alpha_k \approx 1$ for a quadratic problem [Erdoğdu 2019].

In the instance of high-dimensional problems, the inversion, computation, and even storage of $\nabla^2\Phi$ may be impractical. Quasi-Newton methods address this problem by not directly utilising the Hessian (or its inverse) but making an approximation of $(\nabla^2\Phi)^{-1}$ [Erdoğdu 2019].

Limited Memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) [Byrd et al. 1995] is a Quasi-Newton optimisation algorithm that has been applied to the PET reconstruction problem [Tsai et al. 2018]. Curvature information from a small number of previous updates is stored and $\nabla\Phi(\boldsymbol{x})$ is modified (preconditioned) by a series of linear operations. In the aforementioned PET application, the authors used a modification of this algorithm, named L-BFGS-B, to enforce non-negative boundary constraints on $\boldsymbol{x}$ [Zhu et al. 1997]. In an attempt to better condition the initial updates of the optimisation problem, an additional modification to the algorithm (L-BFGS-B-PC) was made by preconditioning the gradient steps used to update the Hessian with either Equation 2.30 or 2.31, see Section 2.3.2.1. The practical performance of the L-BFGS-B-PC algorithm, especially during early updates, was significantly improved compared to L-BFGS-B. However, in both of these algorithms, to ensure the $\boldsymbol{x}_{k+1}$ remains in a feasible region and $\Phi(\boldsymbol{x}_{k+1}) \geq \Phi(\boldsymbol{x}_k)$, an in-exact line search was employed, see Section 2.3.3.3. Line searches increase the number of objective function evaluations required. This can significantly slow reconstruction rates.

2.3.3 Step Size

Once an optimisation method has identified a search direction, the magnitude of the step to be applied at each update needs to be determined. If a step size is too small, many updates may be required to make sufficient progress. Alternatively, step sizes that are too large may result in a non-convergent algorithm. This section presents a variety of methods that may be used to determine step sizes at each update of Equation (2.22).

2.3.3.1 Fixed Step Size

The simplest step size selection method is the use of a constant scalar α and set $\alpha_k = \alpha$ at each update. Some methods, such as Newton and MLEM, may reliably converge with a fixed step size $\alpha = 1$ [Nocedal et al. 2006; Qi et al. 2006]. However, these are the exception to the majority of cases.

In gradient ascent optimisation of a concave problem, convergence using a fixed step size is only guaranteed[vi] if $\alpha_k = \alpha \leq 1/2\mathcal{L}$, where $\mathcal{L}$ is the global Lipschitz value of a Lipschitz continuous function [Nesterov 2004; Nocedal et al. 2006]. Thus, gradient ascent optimisation, given by Equation (2.22), of the log-likelihood function is either non-convergent because $\alpha > \alpha_{\mathcal{L}}$, or slow because $\alpha < \alpha_{\mathcal{L}}$ is small. The preconditioning discussed in Section 2.3.2.1 can significantly assist in reducing the impact of this by better scaling the objective function.

2.3.3.2 Relaxation Sequence

A second step size method that may allow for gradient ascent algorithms to converge is to use a sequence of relaxing step sizes, where $\alpha_k > \alpha_{k+1} \geq 0 \; \forall k \in (0, \infty)$ [Nesterov 2004; Qi et al. 2006]. Such a sequence does not require the computation of $\mathcal{L}$ and for this reason it is commonly used in many optimisation applications [Qi et al. 2006; Defazio et al. 2014]. While conditional on the exact optimisation algorithm, the step size sequences are generally required to satisfy two conditions, given by

$$\sum_{k=0}^{\infty} \alpha_k^2 < \infty, \tag{2.34a}$$

$$\sum_{k=0}^{\infty} \alpha_k = \infty. \tag{2.34b}$$

[vi] Although convergence to an optimal solution of a convex problem is only guaranteed if $\alpha_k = \alpha \leq 1/2\mathcal{L}$, convergence may occur using larger step sizes, e.g. $\Phi(x) = -x^2$, gradient ascent with $\alpha < 2/\mathcal{L}$ will converge.

An example of such a relaxation sequence is given by

$$\alpha_k = \frac{\alpha_0}{\eta k + 1}, \tag{2.35}$$

where α_0 is an initial step size and η the relaxation coefficient [Browne et al. 1996]. However, practical convergence rates of relaxed gradient methods are highly sensitive to these two heuristic relaxation parameters and suboptimal optimisation may be observed if poorly selected [Qi et al. 2006].

2.3.3.3 Inexact Line Search Methods

The two previous step size selection methodologies either require the computation of $\mathcal{L}$ for convergence or are applied without respect to the data with parameters that require tuning via trial and error. Inexact line search methods are another step size selection method and were briefly mentioned in Section 2.3.2.3. These methods select a step size $\alpha_k > 0$ that guarantees $\Phi(\boldsymbol{x} + \alpha_k \boldsymbol{p}_k) > \Phi(\boldsymbol{x})$, assuming $\boldsymbol{p}_k$ is an ascent direction that satisfies Equation (2.24) [Erdoğdu 2019]. Line searches attempt to find α_k that maximises a 1D sub-optimisation problem, given by

$$\hat{\alpha}_k = \underset{\alpha_k \geq 0}{\operatorname{argmax}} \left(\Phi \left(\boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k \right) \right), \tag{2.36}$$

where, due to the concavity of Φ, the optimum $\hat{\alpha}_k$ is the unique global maximiser[vii] [Nocedal et al. 2006].

The optimisation of this problem may be performed iteratively by inexact line search algorithms [Nocedal et al. 2006]. One example of an inexact line search algorithm is to incrementally decrease α_k from an initial value until the Wolfe Conditions are met. The Wolfe Conditions are given by

$$\Phi(\boldsymbol{x}_k + \alpha \boldsymbol{p}_k) \leq \Phi(\boldsymbol{x}) + \lambda_1 \alpha \nabla \Phi(\boldsymbol{x}_k)^\top \boldsymbol{p}_k \tag{2.37a}$$

$$\left\| \nabla \Phi(\boldsymbol{x}_k + \alpha \boldsymbol{p}_k)^\top \boldsymbol{p}_k \right\|_2 \leq \lambda_2 \left\| \nabla \Phi(\boldsymbol{x}_k)^\top \boldsymbol{p}_k \right\|_2, \tag{2.37b}$$

where $\lambda_1 \in (0, 1)$ and $\lambda_2 \in (\lambda_1, 1)$ [Nocedal et al. 2006]. Example values of these parameters are $\lambda_1 = 10^{-4}$ and $\lambda_2 = 0.9$. Equation (2.37a) is known as the Armijo condition and it enforces a sufficient decrease in the objective function value. The curvature condition

[vii] Due to floating/double point precision issues when computing $\Phi(\boldsymbol{x})$, the function may appear to be not concave. STIR's computation of $L(\boldsymbol{y}|\boldsymbol{x})$ was modified from float to double point precision to minimise the impact of numerical precision issues when computing line searches, see Appendix B

is given by Equation (2.37b) and it ensures the next gradient, at $\boldsymbol{x}_k + \alpha \boldsymbol{p}_k$, is sufficiently orthogonal to the current $\nabla\Phi(\boldsymbol{x}_k)$.

The primary limitation of inexact line searches is the significant computation required to ensure that Equation (2.37a) and Equation (2.37b) are satisfied. Many line search iterations may be required before an acceptable $\hat{\alpha}_k$ is identified. This can involve numerous objective function evaluations that are computationally expensive and can significantly slow optimisation algorithms.

2.3.4 Subsets

The computation of an objective function $\Phi(\boldsymbol{x}) = \Phi_{\mathcal{J}}(\boldsymbol{x})$ and its first and second order derivatives may be computationally expensive for a large data-set $\mathcal{J}$. If the objective function is a sum of separable sub-objective functions $\Phi_j(\boldsymbol{x})$, e.g., Equation (2.7), $\Phi_{\mathcal{S}}(\boldsymbol{x})$ may be evaluated, where $\mathcal{S} \subset \mathcal{J}$. Assuming the computational cost of evaluating $\Phi_j(\boldsymbol{x})\ \forall j \in \mathcal{J}$ is constant, the computational cost of evaluating $\Phi_{\mathcal{S}}(\boldsymbol{x})$ is $|\mathcal{J}|/|\mathcal{S}|$ times less than the evaluation of $\Phi_{\mathcal{J}}(\boldsymbol{x})$. This may also be true for first derivative/gradient evaluations. Therefore, use of subsets in an optimisation algorithm can substantially accelerate reconstruction algorithms.

Consider the log-likelihood function presented in Equation (2.7). Standard PET subset methods divide $\mathcal{J}$ into a set of M subsets, i.e., $\{\mathcal{S}_m\}_{m=1}^{M}$, where m is a subset index. The log-likelihood function may be evaluated for a subset $\mathcal{S}_m$ of the data as

$$L_{\mathcal{S}_m}(\boldsymbol{y}|\boldsymbol{x}) = \sum_{j \in \mathcal{S}_m} L_j(\boldsymbol{y}|\boldsymbol{x}) = \sum_{j \in \mathcal{S}_m} y_j \log(\bar{y}_j(\boldsymbol{x})) - \bar{y}_j(\boldsymbol{x}). \tag{2.38}$$

All data is contained within this set of subsets, i.e.,

$$\bigcup_{m=1}^{M} \mathcal{S}_m = \mathcal{J}, \tag{2.39}$$

and the subsets are constrained to be disjoint, i.e.,

$$\mathcal{S}_m \cap \mathcal{S}_n = \emptyset, \forall m \neq n. \tag{2.40}$$

This allows for the log-likelihood to be written as the sum of subset objective functions, given by

$$L(\boldsymbol{y}|\boldsymbol{x}) = \sum_{m=1}^{M} L_{\mathcal{S}_m}(\boldsymbol{y}|\boldsymbol{x}). \tag{2.41}$$

The first derivative of $L(\boldsymbol{y}|\boldsymbol{x})$ may be written as $\nabla L(\boldsymbol{y}|\boldsymbol{x}) = \sum_{j\in\mathcal{J}} \nabla L_j(\boldsymbol{y}|\boldsymbol{x})$ and therefore subset gradients may be written in same manner as the function, given by

$$\nabla L_m(\boldsymbol{y}|\boldsymbol{x}) = \nabla L_{\mathcal{S}_m}(\boldsymbol{y}|\boldsymbol{x}) = \sum_{j\in\mathcal{S}_m} \nabla L_j(\boldsymbol{y}|\boldsymbol{x}) = \sum_{j\in\mathcal{S}_m} \boldsymbol{A}_j^\top \left(\frac{y_j}{\boldsymbol{A}_j\boldsymbol{x} + \bar{b}_j} - 1 \right), \tag{2.42}$$

where $\nabla L_m(\boldsymbol{y}|\boldsymbol{x}) = \nabla L_{\mathcal{S}_m}(\boldsymbol{y}|\boldsymbol{x})$ is used to simplify the notation. Consequently, the number of projection operations ($\boldsymbol{A}_j$ and $\boldsymbol{A}_j^\top$) is directly related to the size of $\mathcal{S}_m$.

Sub-objective functions of an objective function with regularisation can be formulated in numerous ways. The formulation used in this thesis includes a fraction of the penalty term into each sub-objective function, i.e.,

$$\Phi_{\mathcal{J}}(\boldsymbol{x}) = \sum_{j\in\mathcal{S}_m} \left(L_{\mathcal{S}_m}(\boldsymbol{y}|\boldsymbol{x}) + \frac{\beta}{M} R(\boldsymbol{x}) \right). \tag{2.43}$$

It is assumed that the computational cost of evaluating $R(\boldsymbol{x})$ and $\nabla R(\boldsymbol{x})$ is negligible with respect to computing $L_{\mathcal{S}_m}(\boldsymbol{y}|\boldsymbol{x})$ and $\nabla L_{\mathcal{S}_m}(\boldsymbol{y}|\boldsymbol{x})$ and so repeated evaluations are possible. However, alternatives exist, i.e., the penalty term may be treated as an additional subset that can be selected at an algorithm iteration [Ehrhardt et al. 2019].

Consider the optimisation problem where each of the subset objective functions $\Phi_m(\boldsymbol{x})$ are constructed from concave functions, e.g. $\Phi_m(\boldsymbol{x}) = L_m(\boldsymbol{y}|\boldsymbol{x})$. Each $\Phi_m(\boldsymbol{x})$ will have its own unique solution $\hat{\boldsymbol{x}}_m$. In certain instances, it is possible that $\hat{\boldsymbol{x}}_m$ will be equal for all m but this consistency is rarely realised due to variance between subset data. Therefore, a degree of variance is generally found between the subset solutions. Additionally, the more varied the subset data, the larger the expected variance between the $\hat{\boldsymbol{x}}_m$ solutions. Increasing M results in smaller $|\mathcal{S}_m|$ for these PET methods and therefore the variance between $\hat{\boldsymbol{x}}_m$ solutions is expected to be greater.

Subset optimisation algorithms generally select a single subset at each update m_k, based on a selection function. This function may cycle through an ordered sequence with respect to k (as discussed in this section) or, if using a stochastic sampling method, return a random variable index (Section 2.3.5) [Bertsekas 2011]. The general constrained preconditioned subset gradient update is given by

$$\boldsymbol{x}_{k+1} = P_{\boldsymbol{x}\geq 0}\left[\boldsymbol{x}_k + \alpha_k \boldsymbol{D}_{m_k}(\boldsymbol{x}) \nabla \Phi_{m_k}(\boldsymbol{x}_k)\right], \tag{2.44}$$

where the preconditioner $\boldsymbol{D}_{m_k}(\boldsymbol{x})$ may be parameterised by the subset selection m_k.

Many works in PET literature refer to an algorithm update as a "sub-iteration" and M updates as an "iteration". This naming convention is not used in this thesis. A single application of Equation (2.44) (or a comparable equation) is referred to as an (image) update or iteration. Similarly, an epoch is defined as M subset updates or, more generally, equivalent to the projection computational cost of evaluating $\nabla L(\boldsymbol{y}|\boldsymbol{x})$ for a given data set. Note, the penalty computation is not considered in the definition of an epoch because it is generally observed that the computation of the penalty function and its gradient is significantly cheaper than the projection operations $\boldsymbol{A}$ and $\boldsymbol{A}^\top$ due to the size of $\boldsymbol{y}$ being much greater than $\boldsymbol{x}$.

The linear acceleration of a subset algorithm is only realised during early updates when far from the solution [Hudson et al. 1994; Qi et al. 2006; Ruder 2016]. During these early updates, $\nabla\Phi_m(\boldsymbol{x})$ is approximately parallel to $\nabla\Phi(\boldsymbol{x})$. When $\boldsymbol{x}$ approaches the space of the solution set $\{\hat{\boldsymbol{x}}_m\}_m^M$, the directions of $\nabla\Phi_m(\boldsymbol{x})$ for different m's begin to diverge from one another. This may lead to deceleration of the convergence rate. Furthermore, these updates are not guaranteed to satisfy Equation (2.24). As a result, subset algorithms are generally non-convergent when implemented with a constant step size $\alpha > 0$. Cyclical subset selection sequence algorithms are often observed to enter a limit cycle, with cycle "size" proportional to α [Byrne 1998; Hudson et al. 1994; Qi et al. 2006]. A stochastic subset selection sequences, implemented with α, have no such guarantees [Ruder 2016].

Different step size generation sequences may be used to achieve convergence, such as step size relaxation (Section 2.3.3.2) or inexact line searches (Section 2.3.3.3). However, there are practical issues with line search methods as they generally require multiple evaluations of $\Phi(\boldsymbol{x})$ and/or $\Phi_m(\boldsymbol{x})$. These evaluations are expensive, which consequently negates the purpose of the reduced computational cost of subset algorithms. Step size relaxation sequences, applied to subset algorithms, can ensure convergence to a single solution. Such a sequence reduces the size of the limit set to zero and the sequence of estimates converges to a single estimate as $k \to \infty$. Stochastic subset sampling methods will also converge with step size relaxation [Ruder 2016]. However, algorithm convergence rates will depend on the heuristic selection of α_0 and η in Equation (2.35).

Examples of iterative subset PET reconstruction algorithms are: Ordered-Subsets Expectation Maximisation (OSEM) (Section 2.3.4.1), Row Action ML Algorithm (RAMLA)

(Section 2.3.4.2) and BSREM (Section 2.3.4.3).

2.3.4.1 OSEM

Hudson et al. 1994 proposed a subset modification to the MLEM algorithm, Equation (2.28), known as OSEM. The OSEM update equation is given by

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k D_{m_k}(\boldsymbol{x}_k) \nabla L_{m_k}(\boldsymbol{y}|\boldsymbol{x}_k), \tag{2.45}$$

where $\alpha_k = 1$ and $D_{m_k}(\boldsymbol{x}) = \frac{\boldsymbol{x}_k}{\boldsymbol{A}_{m_k}^{\top}\mathbf{1}}$. OSEM's preconditioner is a subset dependant modification of the EM preconditioner, given by Equation (2.27), that may be precomputed at algorithm initialisation.

The OSEM algorithm constructs subsets of regularly spaced scanner projection angles, which correspond to a set of equally spaced rows across all sinograms in the projection data, see Section 2.1.2 [Hudson et al. 1994]. Generally, this implies that $|\mathcal{S}_m| = |\mathcal{J}|/M$. This subset construction criteria was suggested to balance the subsets to minimise variance in the probability of each voxel being detected by different subsets [Hudson et al. 1994; Qi et al. 2006]. In practice, balanced subsets are rarely realised because of non-symmetric attenuation and non-uniformity in detector sensitivities [Qi et al. 2006].

In an attempt to compensate for unbalanced subsets, it is recommended that sequentially selected subsets are as dissimilar as possible from those recently applied [Herman et al. 1993; Hudson et al. 1994]. A cyclical subset selection sequence was suggested and sequential subsets should be selected to correspond to a direction of maximum variability from previously applied subsets in the sequence [Hudson et al. 1994]. An example of an $M = 8$ subset case would lead to the cyclical application of the sequence[viii] $\{1,5,3,7,2,6,4,8\}$ [Herman et al. 1993]. The aforementioned construction of balanced subsets and the subset selection method are two practical steps taken in an attempt to reduce the variance between $\nabla\Phi_m(\boldsymbol{x})$ when $\boldsymbol{x}_k$ approaches the limit set. This methodology is referred to as Ordered-Subsets (OS) for the remainder of this work and is generally implemented in all projection data iterative PET subset algorithms, including the following RAMLA and BSREM examples.

As aforementioned, subset algorithms may converge to a single estimate when step sizes are relaxed. However, relaxed-OSEM does not converge to the ML solution because

[viii] This sequence is not perfect because the subset selected following $m = 8$ is $m = 1$. These subsets may be considered similar to one another.

of $\boldsymbol{D}_{m_k}(\boldsymbol{x}_k)$'s dependency on m_k [Browne et al. 1996; Qi et al. 2006]. Instead, OSEM is commonly implemented with $\alpha_k = 1$ with early stopping and post image reconstruction smoothing [Qi et al. 2006]. This implementation of OSEM, like its MLEM parent given by Equation (2.28), constrains $\boldsymbol{x}$ to non-negative space.

2.3.4.2 RAMLA

A convergent iterative PET reconstruction algorithm for ML estimation (using a relaxed step size) is RAMLA [Browne et al. 1996], given by

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{x}_k \nabla L_{m_k}(\boldsymbol{y}|\boldsymbol{x}_k). \tag{2.46}$$

The primary discrepancy between RAMLA and OSEM is RAMLA's preconditioner is subset invariant and only scales the gradient by the previous estimate, i.e., $\boldsymbol{D}(\boldsymbol{x}_k) = \boldsymbol{x}_k$. RAMLA's convergence to the ML solution is only guaranteed when step sizes are relaxed after each cyclical epoch.

2.3.4.3 BSREM

RAMLA was extended into a PML framework, known as BSREM, which allows for the inclusion of non-zero regularisation in $\Phi(\boldsymbol{x})$ [De Pierro et al. 2001]. The BSREM update formula is given by

$$\boldsymbol{x}_{k+1} = P_{\boldsymbol{x}\geq 0}\left[\boldsymbol{x}_k + \alpha_k D(\boldsymbol{x}_k)\nabla\Phi_{m_k}(\boldsymbol{x}_k)\right], \tag{2.47}$$

where $P_{\boldsymbol{x}\geq 0}[\cdot]$ projects the negative elements of $\boldsymbol{x}$ to zero, $D(\boldsymbol{x}_k)$ is the orginal EM preconditioner given by Equation (2.27) and M scales such that updates direction magnitudes of various numbers of subset configurations are approximately equal. Similar to RAMLA, BSREM includes a subset invariant preconditioner but reintroduces the division by a tomographic "sensitivity" term.

Global convergence of this algorithm requires some assumptions regarding the positivity and boundedness of $\Phi(\boldsymbol{x})$ [Ahn et al. 2003]. Hence, the projection operation and step size relaxation is required because of the inclusion of the prior. Additionally, modifications to the log-likelihood function are necessary. These include the lower bound changes discussed in Section 2.3.3.1 and an upper bound U on $\boldsymbol{x}$. However, this U value is only required for the theoretical convergence and, in practice, it can be very large.

2.3.4.4 List Mode Reconstruction

Only projection based reconstruction algorithms have been considered thus far. However, as aforementioned in Section 2.1.3, PET data can be stored in list mode formats. Reconstruction of this list mode data is possible using the aforementioned algorithms with the modified list mode likelihood gradient computation equation, given by

$$\nabla L^{\mathrm{LM}}_{\mathcal{S}_m}(\Lambda|\boldsymbol{x}) = \sum_{e\in\mathcal{S}_m} \boldsymbol{A}^{\top}_{j(e)} \left(\frac{1}{\bar{y}_{j(e)}(\boldsymbol{x})} \right) - \boldsymbol{A}^{\top}\mathbf{1}, \tag{2.48}$$

where Λ is the measured list mode event data, $e \in \mathcal{S}_m$ is the coincidence event e in the m^{m} list mode data subset and $j(e)$ is a function that ascertains the LOR index of event e [Barrett et al. 1997; Parra et al. 1998]. Note, if $M = 1$, $\mathcal{S}_1 = \{e_l\}_{l=1}^{N_e}$, otherwise, consider the following subset methodology for list mode data.

Subset construction of list mode data can be performed in a number of ways. One method might be to bin events into sets of various projection angles, this is equivalent to the OS methodology for projection data. However, an alternative methodology is considered in this thesis. The N_e events are divided into M blocks of events based on their acquisition time. Each of these blocks is referred to as a subset and a constraint might be enforced that requires each subset to be of (approximately) equal size. This list mode subset construction methodology is comparable to the one-pass list mode reconstruction algorithm's method [Reader et al. 2002]. That list mode reconstruction algorithm indicated that, with a reasonable number of subsets, only a single pass through the data is required for a suitable reconstruction of the PET data.

2.3.5 Stochastic Methods

Stochastic methods also rely on the separability of the objective function, similar to Section 2.3.4. Stochastic sampling involves assigning finite probability to the selection of each data sample or subset.

These algorithms have seen great success in machine learning training - optimisation problems that may involve millions of individual training sets with significant redundancies between various examples [Roux et al. 2012; Ruder 2016]. In machine learning, a single data pair (sample) may be used as a sample at each iterative update. This results in frequent, but highly varied, updates that can cause significant fluctuations in objective function value and parameter estimations [Roux et al. 2012; Ruder 2016]. To reduce update variance,

subsets of data may be constructed to contain more than a single sample, i.e., batches or subsets. This allows for the usage of parallelism (present in most modern computer architectures) when computing update directions from a set of samples, which may accelerate computation [Schmidt et al. 2017].

Two types of stochastic algorithms are presented in this subsection: stochastic gradient ascent and stochastic variance reduction algorithms. However, many additional stochastic algorithms exist, some of which are discussed in Section 4.6.8.

2.3.5.1 Stochastic Gradient Ascent

Stochastic Gradient Ascent (SGA) is the simplest stochastic method and is a stochastic/subset modification to Equation (2.22) [Roux et al. 2012]. At each update k, an update direction $\boldsymbol{p}_{m_k}(\boldsymbol{x}) = \nabla\Phi_{m_k}(\boldsymbol{x})$ is computed. The SGA update formula is therefore given by

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \nabla\Phi_{m_k}(\boldsymbol{x}), \tag{2.49}$$

where m_k is selected by a stochastic subset selection function. Sampling m_k from a uniform distribution yields an unbiased estimate of the full gradient, i.e.,

$$\mathbb{E}[M\nabla\Phi_m(\boldsymbol{x})|\boldsymbol{x}] = \sum_{m=1}^{M} \nabla\Phi_m(\boldsymbol{x}) = \nabla\Phi(\boldsymbol{x}), \tag{2.50}$$

where the expectation is taken with respect to m, where the expected value of a random variable with a finite number of outcomes is a weighted average of all possible outcomes [Roux et al. 2012].

The convergence rate of this algorithm is at least as fast as full gradient methods when not close to the solution, due to the low cost of updates [Bottou et al. 2004]. However, this slows as the $\boldsymbol{x}_k$ approaches $\hat{\boldsymbol{x}}$. Using a relaxing step size sequence, which satisfies Equation (2.34), Stochastic Gradient Ascent (SGA) has been shown to converge to a solution [Bottou et al. 2018].

One may consider the application of a preconditioned-SGA algorithm in PET reconstruction to be analogous to OSEM, RAMLA, or BSREM, but with a stochastic subset selection at each update. This concept is explored in Chapter 4 where performance comparisons between stochastic subset and OS sampling schemes are made.

2.3.5.2 Stochastic Variance Reduction Algorithms

The variance between subset data induces variance between update directions for different m selections. This impedes convergence rates of subset/stochastic algorithms as $\boldsymbol{x}_k$ approaches $\hat{\boldsymbol{x}}$ [Roux et al. 2012]. A novel class of stochastic variance reduction techniques has been proposed to combat this by approximating $\nabla\Phi(\boldsymbol{x}_k)$ with a low computational cost gradient estimate $\tilde{\nabla}_{k,m}(\boldsymbol{x})$. This estimate is expected to have reduced variance, with respect to the selection of m, compared to $\nabla\Phi_m(\boldsymbol{x}_k)$ [Gower et al. 2020]. Generally, $\tilde{\nabla}_{k,m_k}(\boldsymbol{x}_k)$ is computed from a set $\mathcal{G}$ of M previously evaluated subset gradients (usually stored in computer memory), i.e., $\mathcal{G} = \{\boldsymbol{g}_m\}_{m=1}^{M}$, where $\boldsymbol{g}_m \in \mathbb{R}^{N_v}$.

Three SVR algorithms are investigated in this work: Stochastic Average Gradient (SAG), SAGA and Stochastic Variance Reduction Gradient (SVRG). Many other algorithms exist in this class, some of which are discussed in Section 5.3.3. Under certain convexity assumptions, SVR methods converge linearly in expectation. Compare this to traditional subset algorithms, e.g., Equation (2.44), which demonstrate sub-linear convergence [Gower et al. 2020]. The SVR algorithms are more complex than the SGA method. The generalised form of the update direction is given by

$$p_k = \tilde{\nabla}_{k,m_k}(\boldsymbol{x}_k) = \xi\left(\nabla\Phi_{m_k}(\boldsymbol{x}_k) - \boldsymbol{g}_{m_k}\right) + \sum_{\mu=1}^{M} \boldsymbol{g}_\mu, \tag{2.51}$$

where ξ is a scalar value and different algorithms may implement various values[ix]. The SAGA and SVRG algorithms set $\xi = M$, which results in an unbiased estimator of the gradient, i.e., $\nabla\Phi(\boldsymbol{x}_k) = \mathbb{E}[\tilde{\nabla}_{k,m_k}(\boldsymbol{x}_k)]$ [Gower et al. 2020; Driggs et al. 2020b]. The SAG algorithm is a biased estimator because $\xi = 1$ [Roux et al. 2012]. This is further detailed in Appendix C.

Additionally, SAG and SAGA differ from SVRG in the methodology for updating the $\boldsymbol{g}_m$ terms. For SAG and SAGA, after $\nabla\Phi_{m_k}(\boldsymbol{x}_k)$ is computed (see Equation (2.51)) and $\boldsymbol{x}_{k+1}$ updated, the m^{th} entry of $\mathcal{G}$ is updated to $\nabla\Phi_{m_k}(\boldsymbol{x}_k)$:

$$\boldsymbol{g}_\mu = \begin{cases} \nabla\Phi_{m_k}(\boldsymbol{x}_k) & \text{if } \mu = m_k, \\ \boldsymbol{g}_{k,\mu} & \text{otherwise.} \end{cases}, \quad \forall \mu. \tag{2.52}$$

[ix] Note, the formulation of (2.51) is subtly different to the original works of Johnson et al. 2013 and Defazio et al. 2014. This is because the objective function in the said works are formulated as $\Phi(\boldsymbol{x}) = \frac{1}{M}\sum_{m=1}^{M}\Phi_m(\boldsymbol{x})$. Instead, this work uses $\Phi(\boldsymbol{x}) = \sum_{m=1}^{M}\Phi_m(\boldsymbol{x})$, see Equation (2.41).

Thus, SAG and SAGA maintain a record of the most recently computed subset gradient for each subset. Pseudo-code for the SAG and SAGA algorithms is given by Algorithm 1, with $\xi = 1$ and $\xi = M$ respectively.

SVRG implements a different method for updating the entries of $\mathcal{G}$. After a period of γM updates, an anchor image $\tilde{\boldsymbol{x}}$ is re-set and all $\boldsymbol{g}_m = \nabla\Phi_m(\tilde{\boldsymbol{x}}),\ \forall\ m \in \mathcal{M}$. This periodic computation is equivalent to the computation of the full gradient (i.e., an epoch). However, only a single image update is applied. This allows for the evaluation of the objective function at $\tilde{\boldsymbol{x}}$ without additional projection operations. This is likely a unique property of this class of subset algorithm. For convex optimisation problems $\gamma = 2$ is heuristically suggested [Johnson et al. 2013]. The SVRG pseudo-code is given in Algorithm 2.

Algorithm 1: SAG/SAGA Algorithm

Input : $\boldsymbol{x}_0 \in \mathbb{R}^{N_v}$, $\Phi(\boldsymbol{x})$, M, K, ξ and $\{\alpha_k > 0\}_{k=0}^{K}$
Output: $\boldsymbol{x}_K$

Store $\boldsymbol{g}_m = \mathbf{0}, \quad \forall m \in \{1,2,...,M\}$
for $k = 0, 1, \ldots, K$ **do**
  Choose m_k uniformly in $\{1,2,\ldots,M\}$
  Compute $\nabla\Phi_{m_k}(\boldsymbol{x}_k)$
  $\tilde{\nabla}_{k,m_k}(\boldsymbol{x}_k) \leftarrow \xi\left(\nabla\Phi_{m_k}(\boldsymbol{x}_k) - \boldsymbol{g}_{m_k}\right) + \sum_{\mu=1}^{M}\boldsymbol{g}_\mu$
  Store $\boldsymbol{g}_{m_k} \leftarrow \nabla\Phi_{m_k}(\boldsymbol{x}_k)$
  $\boldsymbol{x}_{k+1} \leftarrow \boldsymbol{x}_k + \alpha_k\tilde{\nabla}_{k,m_k}(\boldsymbol{x}_k)$ *# Update step*
end

2.4 Desirable Properties of PET Reconstruction Algorithms

This chapter has described PET physics and technology basics, statistical modelling for PET and general properties of gradient ascent-type optimisation algorithms. This section provides a summary of practical requirements and suggestions for iterative PET reconstruction algorithms in clinical practice.

The primary application of PET image reconstruction is the diagnosis and staging of disease in various medical fields, e.g. Oncology, Cardiology, and Neurology [Durie et al. 2002; Fiechter et al. 2012; Wolk et al. 2012]. A result of this medical application is the desire for reduced patient scan durations and fast but reliable image reconstruction algorithms. This will allow for greater patient throughput and a more widespread and potentially improved quality of care. Clinical practice also often aims to reduce the delivered (injected) radioactive dose to minimise the risk to patients [Caribé et al. 2019].

Algorithm 2: SVRG Algorithm

Input : $\boldsymbol{x}_0 \in \mathbb{R}^{N_v}$, $\Phi(\boldsymbol{x})$, M, K, $\{\alpha_k > 0\}_{k=0}^{K}$ and $\gamma > 0$
Output: $\boldsymbol{x}_K$

for $k = 0, 1, \ldots, K$ **do**
  if $k \mod \gamma M \equiv 0$ **then**
    $\tilde{\boldsymbol{x}} \leftarrow \boldsymbol{x}_k$
    for $m \in \{1, \ldots, M\}$ **do**
      Compute $\nabla\Phi_m(\tilde{\boldsymbol{x}})$
      Store $\boldsymbol{g}_m \leftarrow \nabla\Phi_m(\tilde{\boldsymbol{x}})$
    end
    $\tilde{\nabla}^{\mathrm{SVRG}}_{k,m_k}(\boldsymbol{x}_k) \leftarrow \sum_{m=1}^{M} \boldsymbol{g}_m$
  else
    Choose m_k uniformly in $\{1, 2, \ldots, M\}$
    Compute $\nabla\Phi_{m_k}(\boldsymbol{x}_k)$
    $\tilde{\nabla}^{\mathrm{SVRG}}_{k,m_k}(\boldsymbol{x}_k) \leftarrow M\left(\nabla\Phi_{m_k}(\boldsymbol{x}_k) - \boldsymbol{g}_{m_k}\right) + \sum_{\mu=1}^{M} \boldsymbol{g}_\mu$
  end
  $\boldsymbol{x}_{k+1} \leftarrow \boldsymbol{x}_k + \alpha_k \tilde{\nabla}^{\mathrm{SVRG}}_{k,m_k}(\boldsymbol{x}_k)$ *# Update step*
end

While shorter duration PET scans and reduced injected dose are beneficial for clinical throughput and patient care, the acquired PET data contains fewer counts and a lower SNR [Caribé et al. 2019]. In addition, compound PET image reconstruction applications, e.g., motion correction and kinetics modelling, may further reduce SNR by dividing acquired PET data sets into several bins [Nehmeh et al. 2002; Gallezot et al. 2019]. In recent years, with the implementation of the Q.Clear reconstruction algorithm, clinical practice has been embracing PML image reconstruction [Ahn et al. 2015; Ross 2014; Howard et al. 2017]. Q.Clear utilises the RDP, along with post-smoothing, to increase contrast recovery and organ uniformity over a range of PET tracers [Lantos et al. 2015]. With suitable hyper-parameter selection, PML optimisation algorithms can achieve improved image quality compared to conventional OSEM with post-smoothing or Filtered Back Projection (FBP) reconstructions [Teoh et al. 2015].

As discussed in Section 2.3.4, faster reconstructions may be achieved with the use of subsets but this is complicated by limit cycle behaviour. Reconstruction algorithms may be terminated after only a few epochs because of reconstruction duration limits or as a form of regularisation. This is common for OSEM, but is not desirable for PML algorithms as early termination may result in greater quantification errors. However, clinical throughput limits

available computation time per reconstruction. Therefore, even with a convergent subset algorithm, e.g., BSREM, convergence may not be achieved in practice.

The aforementioned factors motivate the use of convergent, fast and penalised iterative reconstruction algorithms. Reconstructed images realised after a minimal number of epochs should be both visually appealing and quantitatively accurate for clinical reporting.

Chapter 3

AutoSubsets

Iterative PET reconstruction is often accelerated with the use of subset algorithms, as aforementioned in Section 2.3.4. RAMLA and BSREM are examples of convergent subset algorithms due to the implementation of a relaxing step size sequence and their subset independent preconditioners. However, the practical performance of these algorithms is sensitive to the heuristic relaxation parameters.

In this chapter, another form of relaxation is investigated. This method involves the automatic increase of the size of subsets used throughout a reconstruction after initialisation with small subsets. The aim of this method is to realise significant initial algorithm acceleration and reduce limit cycle behaviour at later updates. The method is designed to automatically detect when subset size should be increased at each algorithm update. The algorithm is named AutoSubsets and is applied to three PET data sets: two projection data and one list mode. This method is an extension of an algorithm presented in Thielemans et al. 2015 and a fraction of this work has been previously presented in Twyman et al. 2019.

3.1 Introduction

The optimisation of smooth, convex and data separable objective functions, given by Equation (2.41), is considered in this chapter. As previously discussed in section 2.3.4, fixed step size subset algorithms realise fast initial convergence rates but do not converge to a stationary point. Instead, the solutions become bound within some closed limit set with non-zero size. Relaxing step size sequences ensure convergence to a single solution by reducing the size of the limit set to zero as $k \to \infty$.

An alternative method to ensure a sequence converges would be to run a number of updates using a non-convergent subset algorithm before switching to a non-subset convergent algorithm [Ahn et al. 2003]. This method takes advantage of the fast initial convergence

rates of subset algorithms. This transition may be gradual, beginning with many subsets and reducing the number throughout a reconstruction at predetermined update numbers. An example of this methodology would be to perform a number of OSEM epochs[i] before switching to MLEM. The determination regarding when to make these transitions is usually heuristic and therefore likely sub-optimal.

Subset size at each algorithm update k is denoted by n_k. Given the union and disjoint restrictions on subset construction in general PET algorithms, given by Equation (2.39) and Equation (2.40) respectively, reducing the number of subsets increases n_k. Using fewer but larger subsets generally results in improved subset balance but more computational effort is required to evaluate $\nabla\Phi_m(\boldsymbol{x})$.

An algorithm that increases n_k based on the divergence between two subset update directions was previously proposed by Thielemans et al. 2015. This divergence measurement was motivated by intuition. As a subset algorithm approaches the limit set, the two subset search directions $\boldsymbol{p}_{k,m}$ begin to diverge from one another. After this divergence measurement, the two $\boldsymbol{p}_{k,m}$ vectors are combined into a single search direction and $\boldsymbol{x}_{k+1}$ is computed. If a divergence threshold is met, the algorithm increases n_{k+1}, otherwise it remains the same. As the algorithm progresses, n_k is incrementally increased until the two subsets contain all the data, i.e., $n_k = |\mathcal{J}|/2$, as the subsets are constrained to be disjoint and satisfying Equation (2.40). This adaptive subset size methodology was designed so that it could be applied to any subset algorithm and example reconstructions using OSEM and OSSPS were provided in Thielemans et al. 2015.

While a useful proof of concept, this subset methodology remained largely unexplored with only investigation into non-PML objective function reconstructions using OSEM and OSSPS [Thielemans et al. 2015]. As the ML solution is generally noisy, convergence is undesirable. Furthermore, the numerical experiments only illustrated algorithm performance using a simulated 2D Shepp-Logan phantom [L A Shepp et al. 1974] using 120 scanner views. Additionally, traditional OS subset construction was considered, as described in Section 2.3.4.1. The algorithm was initialised with a subset size of $n_0 = 2$ projection angles and was limited to increase n_{k+1} to factors of half the number of projection angles.

This chapter presents the AutoSubsets (AS) algorithm. AS is a modified version of this aforementioned adaptive algorithm and is evaluated using data from simulated 3D PET

[i] An epoch is defined as equivalent projection computation as computing $\nabla L(\boldsymbol{x})$, see Section 2.3.4

scans using PML objective functions. Additionally, AS modifies how subsets are handled and investigates two subset selection methods based on projection angles and one based on temporal/event-by-event list mode data.

3.2 Algorithm Methodology

The AS algorithm is a modification of standard subset iterative reconstruction algorithms. Standard methods compute a single gradient direction at each update, using a single subset of the data, and subsets are predefined at algorithm initialisation. AS differs from these methods. At each update k, AS samples the data set twice. Two subsets, $\mathcal{S}_{k,1}$ and $\mathcal{S}_{k,2}$, are constructed with size $n_k < |\mathcal{J}|/2$ by a subset selection function, denoted Γ. Subset construction methods are further discussed in section 3.2.1. Two update directions, $\boldsymbol{p}_{\mathcal{S}_{k,1}}$ and $\boldsymbol{p}_{\mathcal{S}_{k,2}}$, are computed from $\boldsymbol{x}_k$ using the respective subsets. The divergence between these two vectors is computed using the Cosine Similarity Index (CSI) $S_c(\boldsymbol{p}_{\mathcal{S}_{k,1}}, \boldsymbol{p}_{\mathcal{S}_{k,2}})$, given by

$$S_c = S_c(\boldsymbol{p}_{\mathcal{S}_{k,1}}, \boldsymbol{p}_{\mathcal{S}_{k,2}}) = \frac{\boldsymbol{p}_{\mathcal{S}_{k,1}} \cdot \boldsymbol{p}_{\mathcal{S}_{k,2}}}{\|\boldsymbol{p}_{\mathcal{S}_{k,1}}\|_2 \|\boldsymbol{p}_{\mathcal{S}_{k,2}}\|_2}, \tag{3.1}$$

where $(\cdot)$ is the dot product of two vectors. For any update, if $S_c(\boldsymbol{p}_{\mathcal{S}_{k,1}}, \boldsymbol{p}_{\mathcal{S}_{k,2}}) < T$, where T is a heuristic lower threshold value, the subset size is increased[ii] $n_{k+1} = \tau n_k$, where $\tau > 1$ is also heuristic. The maximum size of the subsets $n_{\max}$ is limited to $n_{\max} := |\mathcal{J}|/2$, which would result in half the data in each subset.

The determination of $\boldsymbol{p}_{\mathcal{S}_{k,1}}$ and $\boldsymbol{p}_{\mathcal{S}_{k,2}}$ will depend on the reconstruction algorithm template being implemented. At this time, various diagonally-preconditioned gradient ascent algorithms, given by Equation (2.44), are considered for the AS algorithm. Hence, $\boldsymbol{p}_{\mathcal{S}_{k,m}} = \boldsymbol{D}_{\mathcal{S}_{k,m}}(\boldsymbol{x}_k) \nabla \Phi_{\mathcal{S}_{k,m}}(\boldsymbol{x}_k) \; \forall m \in \{1,2\}$.

To compute the updated image, AS combines the update directions for each subset sample. Note that this implies that the update is computed with a subset of size $2n_k$. To account for any preconditioner subset dependence, the iteration's update direction $\boldsymbol{p}_k$ is computed as

$$\boldsymbol{p}_k = \left(\boldsymbol{D}_{\mathcal{S}_{k,1}}(\boldsymbol{x}_k) + \boldsymbol{D}_{\mathcal{S}_{k,2}}(\boldsymbol{x}_k)\right) \left(\nabla \Phi_{\mathcal{S}_{k,1}}(\boldsymbol{x}_k) + \nabla \Phi_{\mathcal{S}_{k,2}}(\boldsymbol{x}_k)\right). \tag{3.2}$$

This combination is appropriate for many PET algorithms where $\boldsymbol{D}_{\mathcal{S}_{k,1}}(\boldsymbol{x}_k)$ and $\boldsymbol{D}_{\mathcal{S}_{k,2}}(\boldsymbol{x}_k)$

[ii] As subset size $|\mathcal{S}|$ is an integer, $n_{k+1} = \max(\tau n_k, n_k + 1)$ to ensure subset size is increased when $S_c < T$.

are parallel, e.g., BSREM and RAMLA.

Full pseudo-code for the AS algorithm is given by Algorithm 3.

Algorithm 3: AutoSubsets Algorithm

Input : $\boldsymbol{x}_0 \in \mathbb{R}^{N_v}$, Φ, $\boldsymbol{D}$ Γ, n_0, $n_{\max}$, T, K, $\tau > 1$, $\{\alpha_k > 0\}_{k=0}^{K}$
Output: $\boldsymbol{x}_K$

for $k = 0, 1, \ldots, K$ **do**
  # Construct two equally sized subsets of size n_k
  $\mathcal{S}_{k,1},\ \mathcal{S}_{k,2} \leftarrow \Gamma(2n_k, 2n_{\max})$
  # Compute two update directions from $\mathcal{S}_{k,1}$ and $\mathcal{S}_{k,2}$
  Compute $\nabla\Phi_{\mathcal{S}_{k,1}}(\boldsymbol{x}_k)$ and $\nabla\Phi_{\mathcal{S}_{k,2}}(\boldsymbol{x}_k)$
  $\boldsymbol{p}_{\mathcal{S}_{k,1}} \leftarrow \boldsymbol{D}_{\mathcal{S}_{k,1}}(\boldsymbol{x}_k)\nabla\Phi_{\mathcal{S}_{k,1}}(\boldsymbol{x}_k)$
  $\boldsymbol{p}_{\mathcal{S}_{k,2}} \leftarrow \boldsymbol{D}_{\mathcal{S}_{k,2}}(\boldsymbol{x}_k)\nabla\Phi_{\mathcal{S}_{k,2}}(\boldsymbol{x}_k)$
  # Cosine check
  if $S_c(\boldsymbol{p}_{\mathcal{S}_{k,1}}, \boldsymbol{p}_{\mathcal{S}_{k,2}}) < T$ **then**
    $n_{k+1} \leftarrow \min(\max(\tau n_k, n_k + 1),\ n_{\max})$
  else
    $n_{k+1} \leftarrow n_k$
  end
  # Combine subset update directions into a single update direction
  $\boldsymbol{p}_k \leftarrow \left(\boldsymbol{D}_{\mathcal{S}_{k,1}}(\boldsymbol{x}_k) + \boldsymbol{D}_{\mathcal{S}_{k,2}}(\boldsymbol{x}_k)\right)\left(\nabla\Phi_{\mathcal{S}_{k,1}}(\boldsymbol{x}_k) + \nabla\Phi_{\mathcal{S}_{k,2}}(\boldsymbol{x}_k)\right)$
  # Update step
  $\boldsymbol{x}_{k+1} \leftarrow P_{\boldsymbol{x}\geq 0}\left[\boldsymbol{x}_k + \alpha_k \boldsymbol{p}_k\right]$
end

3.2.1 Subset Construction

Alternative subset construction methodologies to OS are considered for AutoSubsets. The OS methodology, described in Section 2.3.4.1, bins regularly spaced projection angles into subsets in an attempt to balance the subsets [Hudson et al. 1994]. However, this methodology is limited to creating subsets based on projection angles but also requires an equal number of the N_θ projection angles to be binned into each of the M subsets to maintain balance. Hence, M is limited to a factor of N_θ. Furthermore, the order that the subsets are selected at each update is suggested to be cyclical and subsequent selected subset data should be as orthogonal as possible to the subset data recently applied in previous algorithm iterations [Herman et al. 1993].

Implementing an OS construction/selection methodology to the AS algorithm would enforce restrictions on possible subset sizes and the subset construction would be limited to projection angle binning. Furthermore, OS's subset selection may become non-trivial

in order to avoid introducing bias into the cosine measurements of Equation (3.1). If two similar subset indices were selected with similar data (e.g., $m_1 = 1$ and $m_2 = 2$ when $M = 24$), the S_c values would likely be greater than those computed using dissimilar subset indices and data[iii]. Additionally, maintaining subset balance is not a strict requirement for the AS algorithm. This is because the algorithm aims for $n_k \xrightarrow[k\to\infty]{} n_{\max}$ and, assuming $n_{\max} := |\mathcal{J}|/2$, the algorithm will tend towards full data updates (i.e., $\mathcal{S}_{k,1} \cup \mathcal{S}_{k,2} \xrightarrow[k\to\infty]{} \mathcal{J}$), given the $\mathcal{S}_{k,1} \cap \mathcal{S}_{k,2} = \emptyset$ condition of Equation (2.40) is enforced. As long as the algorithm achieves these full data updates, subset balance is not necessarily required in subset construction because balance is eventually achieved.

The following sections will describe three Γ subset selection functions used in the AutoSubsets algorithm. Two functions are used for geometric projection angle subset construction, RB and Golden Ratio Subsets (GRS) sampling respectively. A third function is used for list mode event based sampling.

3.2.1.1 Geometric Sampling

The number of discrete PET scanner projection angles N_θ is a property of a scanner's design, see Section 2.1.3. Similar to Equation (2.41), the full set of data samples $\mathcal{J}$ may be divided into a set of N_θ subsets ($\Theta = \{\theta_\mu\}_{\mu=1}^{N_\theta}$), where each subset corresponds to a discrete projection angle. In this chapter, each θ_μ may be refered to as a "sample" (set) of the data and sequential θ_μ samples correspond to sequential rows of the sinograms in the projection data. A subset size of $n_k = 1$ corresponds to a set $|\{\theta_\mu\}| = 1$, where $\mu \in \{1, ..., N_\theta\}$.

Random Batches (RB) subset construction function $\Gamma_\Theta^{\text{RB}}(2n_k, 2n_{\max})$ generates a temporary single new subset from Θ by uniform random selection without replacement. This temporary subset has a size of $2n_k$. The function equally divides the elements of this large subset between $\mathcal{S}_{k,1}$ and $\mathcal{S}_{k,2}$. Thus, $\mathcal{S}_{k,1}$ and $\mathcal{S}_{k,2}$ are equal size disjoint subsets and, if $n_k = n_{\max}$, $\mathcal{S}_{k,1} \cup \mathcal{S}_{k,2} = \mathcal{J}$.

Golden Ratio Subsets (GRS) sampling is commonly used in MRI to perform dynamic radial k-space sampling [Feng et al. 2014; Winkelmann et al. 2007]. Yet, GRS sampling was originally implemented for PET image reconstruction [Kohler 2004]. In dynamic MRI nearly uniform sampling of the k-space distribution is required for an unknown acquisition period; GRS radial sampling can be used to achieve this [Feng et al. 2014].

[iii] It is worth noting that this is the motivation for why OS algorithms were designed to select sequential subsets that are as orthogonal as possible to the subsets recently applied. The algorithms attempt to update $\boldsymbol{x}_k$ with "new" data at each iteration.

The golden ratio is computed as

$$\gamma_{\mathrm{GR}} = \frac{\sqrt{5}-1}{2} \approx 0.618, \tag{3.3}$$

The GRS subset selection function $\Gamma_{\Theta}^{\mathrm{GRS}}$ is detailed by Algorithm 4. The algorithm utilises an internal variable γ_c, which is maintained in memory by the function. This is required because successive calls of $\Gamma_{\Theta}^{\mathrm{GRS}}$ should result in varied subset content. The function returns two equally sized disjoint subsets. Note, members of a set must be distinct from one another and therefore line 9 of Algorithm 4 will only increase the size of $|\mathcal{S}|$ if $\theta_\mu \notin \mathcal{S}$.

Algorithm 4: Golden Ratio Subsets Function $\Gamma_{\Theta}^{\mathrm{GRS}}$

Input : $n_{\max}, \Theta$
Output: $(\mathcal{A}_1, \mathcal{A}_2)$
1 Assert $|\Theta| \geq n_{\max} \geq 1$
2 Initialise $\gamma_{\mathrm{GR}} = \frac{\sqrt{5}-1}{2}$
3 Initialise $\mathcal{S} = \mathcal{A}_1 = \mathcal{A}_2 = \emptyset$
4 Maintain γ_c in memory or initialise as $\gamma_c = 0$
5 *#Determine* $\mathcal{S}$
6 **while** $|\mathcal{S}| < n_{max}$ **do**
7 $\quad \gamma_c = \gamma_c + \gamma_{\mathrm{GR}} \mod 1$
8 $\quad \mu = \mathrm{round}_{\mathrm{up}}(|\Theta| \cdot \gamma_c)$
9 $\quad \mathcal{S} = \mathcal{S} \cup \{\theta_\mu\}$
10 **end**
11 *# Divide* $\mathcal{S}$ *into* $\mathcal{A}_1$ and $\mathcal{A}_2$
12 **for** $i = 1, \ldots, N$ **do**
13 $\quad$ *#* $\mathcal{S}_i$ *indicates the* i^{th} *index of* $\mathcal{S}$
14 $\quad$ **if** $i \mod 2 = 1$ **then**
15 $\quad\quad \mathcal{A}_1 = \mathcal{A}_1 \cup \{\mathcal{S}_i\}$
16 $\quad$ **else**
17 $\quad\quad \mathcal{A}_2 = \mathcal{A}_2 \cup \{\mathcal{S}_i\}$
18 $\quad$ **end**
19 **end**
20 **return** $(\mathcal{A}_1, \mathcal{A}_2)$

3.2.1.2 List Mode Event Sampling

The application of the AS reconstruction algorithm to list mode data is also investigated in this work, see Sections 2.1.3.2 and 2.3.4.4. List mode data sets typically contain a large number of individual N_e coincidence events. List mode reconstruction can be performed by taking a set of events from the list mode data and computing an update direction from this

set.

List mode coincidence events are recorded in chronological detection order. Tracer kinetics and radioactive decay modelling were not included in the simulated acquisition in this work. This allows for the list mode data to be split into equally sized "samples" by dividing into equal length time frames. In this study, these samples typically contain $10,000+$ coincidence events.

Most list mode reconstruction algorithms begin at the start of the list mode data and n_k samples are used to compute the update direction and the next iteration selects the next n_k samples. Once the end of the file is reached, reading of the list mode file is reset to the beginning. The AS algorithm uses a similar methodology but with two subsets per update. The $\mathcal{S}_{k,1}$ subset is populated from the first n_k samples from the previous marked position, followed by $\mathcal{S}_{k,2}$.

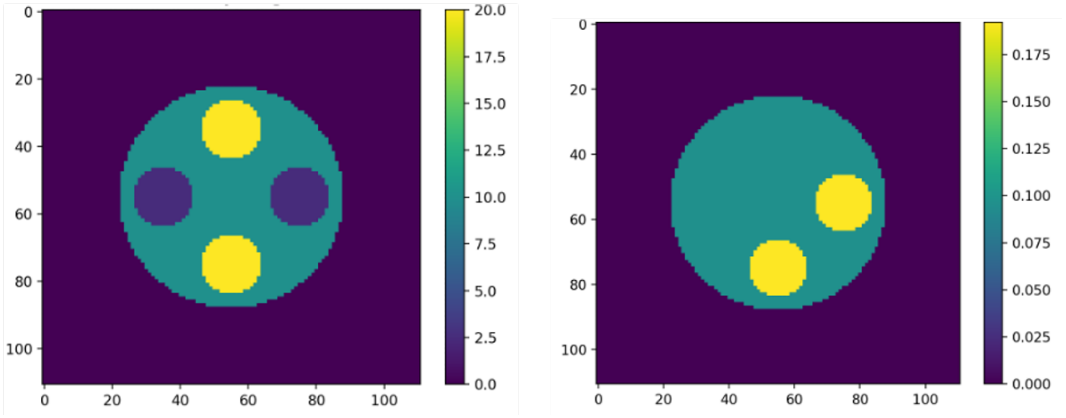
3.3 Experimental Setup

This section describes the experimental setup used to assess AutoSubsets algorithm performance. This includes methods of acquiring simulated phantom data, practicalities of AS algorithm implementation and the metrics used to assess reconstruction performance.

3.3.1 Phantom and Acquisition

A cylindrical phantom of radius 98mm was generated using STIR tools [Thielemans et al. 2012]. The cylinder was designed to have intermediate levels of activity and attenuation values comparable to water. Four smaller cylinders with a radius of 26mm were added with various activity levels to the larger cylinder. Two of the inserted cylinders contained higher activity than the main cylinder, with ratio 4:2, and the other two cylinders contained lower activity, with ratio 1:2. These ratios were taken with respect to the main body of the cylindrical phantom. Additionally, two of these inserts, one hot and the other cold, were assigned an attenuation double that of the main cylinder. The arrangement of these cylinders was to inhibit rotational symmetry in the axial plane. Transaxial slices of both the activity and the attenuation for this phantom are shown in Figure 3.1. This is a modification of the phantom used in Tsai et al. 2016.

Two scanner geometries were used to simulate a PET acquisition of this phantom. The first geometry utilised $N_\theta = 280$ and $N_R = 2$ rings. This geometry was used to assess general applicability of the AS algorithm on a smaller data set, which allowed for faster image



(a) Activity **(b)** Attenuation(cm^{-1})

Figure 3.1: Transaxial slices of the simulated cylindrical phantom with cylindrical inserts of various activity levels.

reconstruction. This data set is referred to as the "toy" data. The acquisition of the toy data was simulated using STIR projectors. Random coincidence events were also simulated, with a 6:1 true-to-background ratio, as low level Poisson noise using STIR functionality.

The second scanner geometry was a model of the GE Discovery 690 PET/CT (D690), with $N_\theta = 288$ (576 detectors) and $N_R = 24$ [Bettinardi et al. 2011]. A 120 second data acquisition of the phantom in this scanner was simulated using GATE and the STIR-GATE-Connection , see Appendix A. GATE is a Monte Carlo simulation software designed to accurately model high energy physics interactions for medical physics applications [Jan et al. 2011]. The list mode data output by GATE was both retained for list mode reconstruction and re-binned into projection data using STIR tools for view based reconstruction analysis, see Appendix A.2. Normalisation, attenuation correction and random and scattered coincidence events were modelled using STIR tools, see Appendix A.3. Approximately 53 million coincidence events were recorded. This data set was significantly noisier than the toy data set.

3.3.2 Algorithm Implementation

As aforementioned, the ML solution is undesirable due to the ill-posedness of the PET problem. Introducing regularisation can significantly improve reconstructed image quality. A PML objective function, with the RDP, was used to assess the AS algorithms applicability to PET reconstruction. The preconditioners implemented into the AS algorithm were $\boldsymbol{D}_{\mathcal{S}_{k,1}}(\boldsymbol{x}_k) = \frac{|\mathcal{J}|}{|\mathcal{S}_{k,1}|}\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_k)$ and $\boldsymbol{D}_{\mathcal{S}_{k,2}}(\boldsymbol{x}_k) = \frac{|\mathcal{J}|}{|\mathcal{S}_{k,2}|}\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_k)$, where $\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_k)$ is given by Equation (2.29). The direction of this diagonal positive preconditioner is subset independent but

the magnitude is scaled proportionally to the subset's size.

The AS algorithm was compared to BSREM reconstructions using various subsets. Typically, BSREM is implemented with a relaxing step size sequence α_k, e.g., Equation (2.35). However, this may result in difficulties when comparing algorithm performances. Firstly, the step size relaxation parameter selections are heuristic and different selections of α_0 and η will result in various performance characteristics. Secondly, α_k is update number dependant. To compare algorithms using step size relaxation, step sizes should be equal after similar computational effort to confirm the performance relates to the algorithm rather than heuristically selected parameters. After $k = M$ updates, BSREM's computational cost is equivalent to one epoch. Compare this to AS where the computational cost associated with computing $k = M$ updates may not be equivalent to one epoch due to n_k varying. Therefore, for simplicity in algorithm analysis, all reconstructions were implemented with a constant α_k.

3.3.3 Reconstruction Performance Analysis

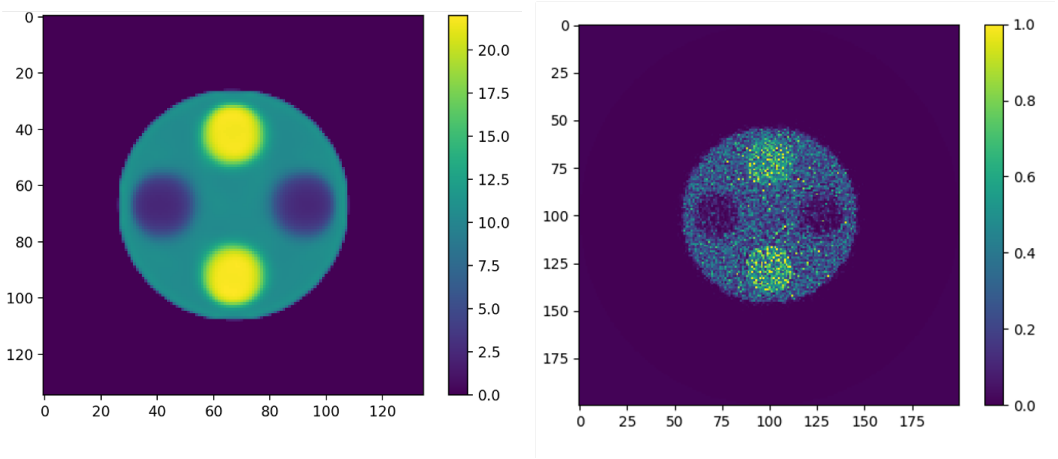
To assess the performance of the AS algorithm, several quantitative measurement metrics were employed at every algorithm iteration k. An image domain assessment was conducted using a distance measure between the k^{th} update and the converged MAP estimate $\hat{\boldsymbol{x}}$. This Δ function is given by

$$\Delta(\boldsymbol{x}_k, \hat{\boldsymbol{x}}) = \frac{\|\boldsymbol{x}_k - \hat{\boldsymbol{x}}\|}{\|\hat{\boldsymbol{x}}\|}, \tag{3.4}$$

where $\|\cdot\|$ is the l_2 norm. The converged image $\hat{\boldsymbol{x}}$ was computed with a single subset modified-EM preconditioned gradient ascent algorithm using an in-exact line search to select optimal α_k values, see Section 2.3.3.3. Once the algorithm reached a region of numerical instability at iteration $k = K$, it was terminated and $\hat{\boldsymbol{x}} = \boldsymbol{x}_K$. A transaxial slice of $\hat{\boldsymbol{x}}$ for the toy and the D690 data sets are shown in Figure 3.2. Additionally, some reconstructions were evaluated with objective function value measurements, computed using all the data. Finally, Regions Of Interest (ROIs) were drawn over the cylindrical inserts and voxel mean and standard deviations values were computed for algorithm iterates.

3.4 Results

This section presents a number of AS algorithm metrics and reconstruction performance results.



(a) MAP Solution of Toy Data Set

(b) MAP Solution of GE Discovery 690 Data Set

Figure 3.2: Transaxial slices of the MAP estimates of (a) the toy scanner data set (2.39, 2.39, 3.27 mm/voxel) and (b) the GE Discovery 690 PET/CT data set (2.13, 2.13, 3.27 mm/voxel).

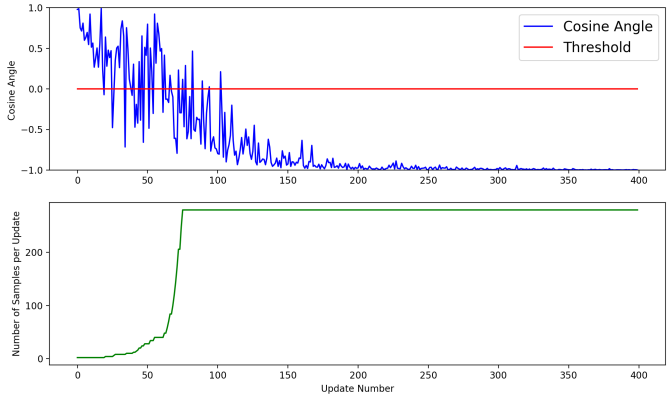
3.4.1 Initial Evaluation

This section assesses the preliminary performance of the AS algorithm and makes comparisons between the two projection angle based sampling methods. All reconstructions were performed using the toy data set, which may be considered to have low noise and complexity. Measurements of the $S_c(\boldsymbol{p}_{\mathcal{S}_{k,1}}, \boldsymbol{p}_{\mathcal{S}_{k,2}})$, given by Equation (3.1), were recorded throughout the reconstructions and algorithm performance was measured using the Δ metric, given by Equation (3.4).

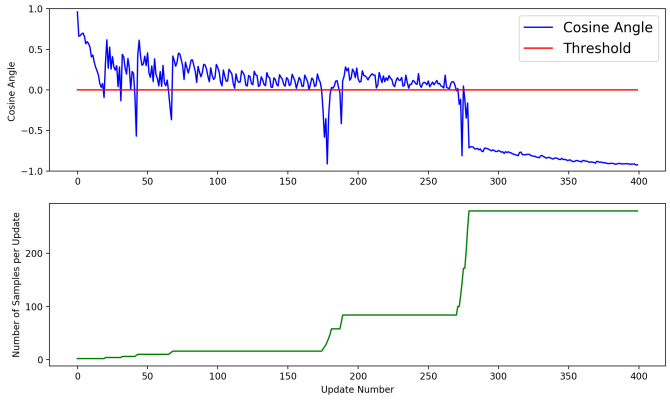
An initial comparison is made between the RB and GRS algorithm performances in this subsection. For this comparison, step size was reduced from the default $\alpha = 1$ to $\alpha = 0.1$ to slow down the reconstruction performance for algorithm evaluation and the two subset independent preconditioners discussed in Section 3.3.2 were implemented. The threshold value for the S_c measure was heuristically set to $T = 0 = \cos(90°)$ and the subset size multiplicative factor $\tau = 1.2$.

3.4.1.1 Cosine Similarity and Subset Sizes

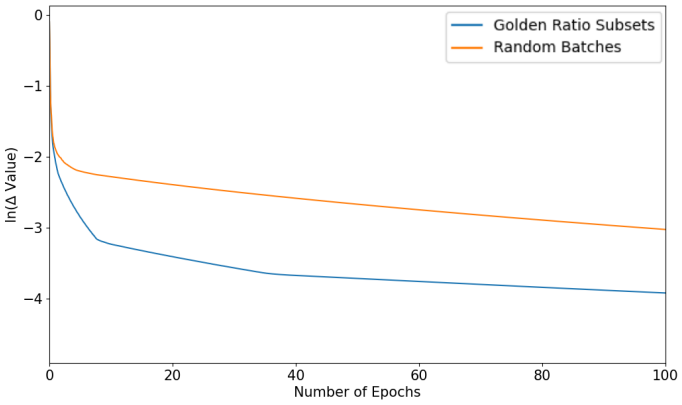
The RB sampling cosine angle values S_c exhibited significant fluctuations between sequential updates throughout the algorithm iterates and quickly converged to full data (280 projection angles) updates, as shown by Figure 3.3a. In comparison, the GRS sampling method displayed a gradual reduction in S_c values during the first few updates. This steady reduction occurred until the S_c value fell below $T = 0.0$ and n_{k+1} was increased. The subsequent iterate measured an increased S_c value. The GRS S_c values continued to be less noisy than



(a) Random Batches Sampling



(b) Golden Ratio Subset Sampling



(c) Δ Value Plot

Figure 3.3: A comparison between two AS reconstructions: (a) RB sampling and (b) using GRS sampling for the toy data set with $\alpha = 0.1$, $\tau = 1.2$ and $T = 0$. The CSI measurements between the two AutoSubsets' subset update directions, given by Equation (3.1), and the number of samples (projection angles) used at each update are compared. (c) Compares the two AS reconstruction sequences performances using the $\ln(\Delta)$ convergence measurement metric, given by Equation (3.4), over the number of epochs.

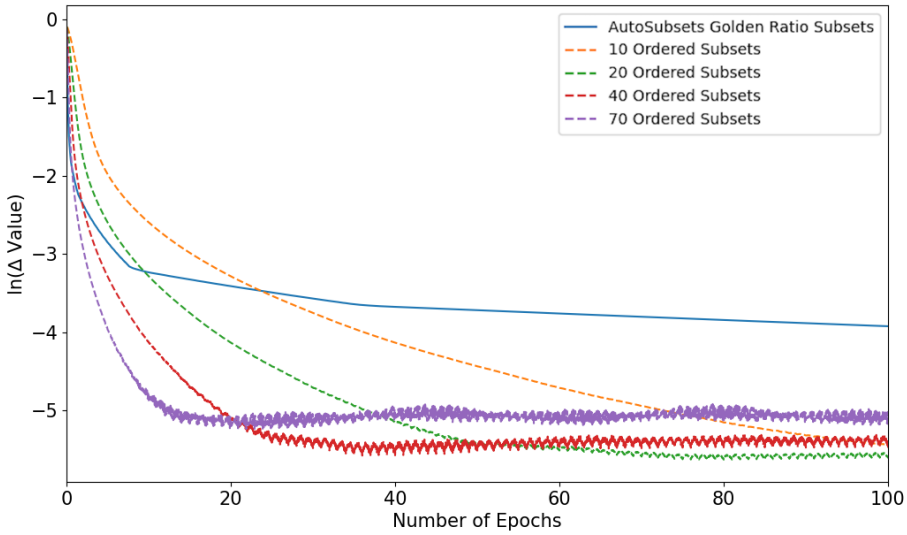


Figure 3.4: A plot comparing the $\ln(\Delta)$ metric performance of a GRS sampling AS reconstruction with different OS reconstructions. AS reconstruction parameters are $\tau = 1.2$ and $T = 0$. All algorithms utilise $\alpha = 0.1$. The number preceding "subsets" in the legend entries indicates the number of subsets used during the reconstruction.

RB's measurements. In addition, the GRS n_k values increased slower but did eventually reach full data updates after $k = 275$.

3.4.1.2 Reconstruction Performance

To assess algorithm performance, every image estimate of the two aforementioned reconstruction sequences was evaluated using the Δ metric, given by Equation (3.4). This metric is plotted in Figure 3.3c against the number of processed data epochs, a scale that may be considered proportional to computational cost. These results demonstrated that GRS sampling reconstruction was capable of converging closer to the solution from very early iterations. The RB sampling reconstruction maximised the subset sizes within 75 updates (approximately 10 epochs), whereas the GRS sampling took advantage of the acceleration achieved via the use of smaller subsets for 275 updates (approximately 40 epochs). This inferior performance of AS using RB sampling was observed to be typical for other data sets (results not shown). Therefore, only AS results using the GRS sampling are considered for the remainder of this chapter.

3.4.1.3 Comparison with Fixed Subset Size Reconstructions

Reconstruction performance comparisons between the GRS AS algorithm and various BSREM reconstructions using $M \in \{10, 20, 40, 70\}$ subsets, with the OS sampling method-

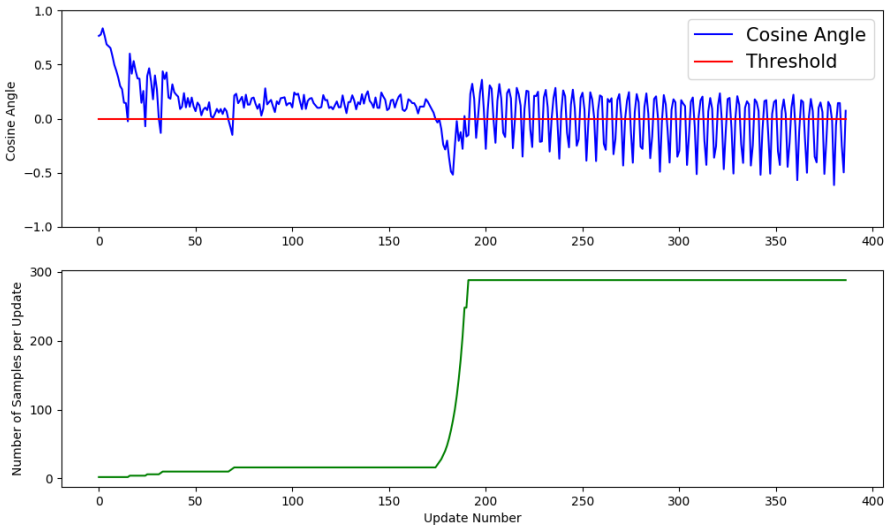


Figure 3.5: A plot illustrating the CSI measurements and number of samples (projection angles) used at each update of an AS reconstruction with GRS sampling using the D690 projection data. The AS algorithm utilised $\tau = 1.2$ and $T = 0$ and $\alpha = 1$

ology, are shown in Figure 3.4. The performance of AS was comparable to that of the 40 and 70 OS reconstructions during the first few epochs and was significantly accelerated compared to $M = 10$ and $M = 20$ OS reconstructions. However, as the AS reconstruction incorporated larger subsets into its updates, this acceleration was reduced. Hence, for this data set, the OS methods continued to enjoy faster objective function value convergence rates.

3.4.2 GE Discovery 690 Projection Data

The CSI and projection angle usage for a GRS AS reconstruction of the GE Discovery 690 projection data is shown in Figure 3.5. The AS algorithm was implemented with $\tau = 1.2$ and $T = 0$ and all reconstructions utilised $\alpha = 1$.

During initial updates, the S_c performance was similar to that shown in Figure 3.3b. However, after $k \approx 70$ until $k \approx 175$, the S_c values varied between 0.1 and 0.3. For this long period of updates, each subset contained 8 projection angles (i.e., 16 projection angles per update and equivalent to $M = 18$ subsets in standard OS methodologies). After this 100+ updates, the $S_c < T$ for a number of updates and n_k increased to $n_{\max}$ in only a few updates.

The objective function value was computed after each AS algorithm update and performance is compared to BSREM algorithms with various numbers of subsets in Figure 3.6.

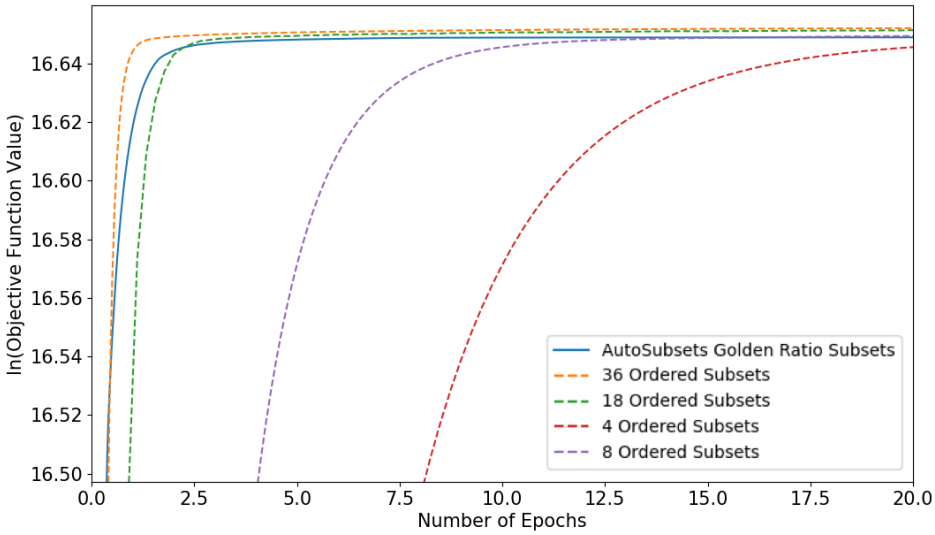


Figure 3.6: A plot comparing the objective function value performance of the GRS subset sampling AS reconstruction with different OS subset size reconstructions for the GE Discovery 690 projection data-set. AS algorithm parameters are the same as those used for Figure 3.5. The number preceding "OS" in the figure legend indicates the number of subsets used for each reconstruction.

As this was a maximisation problem, the greater the objective function value, the better the estimate fit the measured data. Similar to Section 3.4.1, the AS reconstruction initially accelerated the reconstruction and was comparable to the 18 and 36 OS reconstructions and significantly outperformed the 4 and 8 OS reconstructions for several epochs. The objective function plateaued after a number of epochs and, after achieving full data updates, was overtaken by the 8 OS reconstruction.

3.4.3 List Mode Reconstruction

The D690 list mode data-set contained 120 million coincidence events. Rather than treat a single event as a list mode sample, a single sample contained approximately $14,000$ list mode events. The list mode AS reconstruction was initialised with $n_0 = 1$ while $n_{\max} = 4320$. The same $\alpha = 1$, $T = 0.0$ and $\tau = 1.2$ were implemented.

The S_c values and number of samples used at each iteration of a GRS AS reconstruction are plotted in Figure 3.7. The reconstruction duration of this study was limited to 6 epochs. This was due to a technical limitation where reconstructing an epoch of list mode data was significantly slower than an epoch of projection data in STIR. Note that the AS algorithm did not increase to full data updates within the presented 6 epochs.

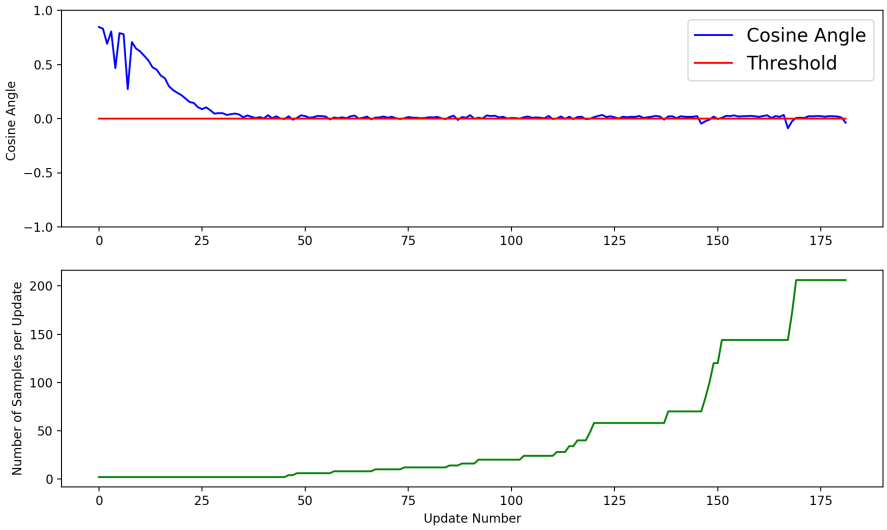


Figure 3.7: A plot illustrating the (a) CSI measurements between AutoSubsets's update directions and (b) number of list mode data samples used at each update of the algorithm. The algorithm configuration used $\alpha = 1$, $T = 0$ and $\tau = 1.2$ and each sample corresponds to approximately 14,000 coincidence events.

The metrics plotted in Figure 3.7 differ from the metrics plotted for the projection data in Figures 3.3b and 3.5. The initial S_c values were significantly slower in their decline towards $T = 0.0$ and the first instance of $S_c < T$ occurred at $k \approx 45$. Rather than a significant increase in S_c value after n_k was increased, the list mode S_c value only slightly increased to $0 < S_c < 0.05$. Instead, over the following iterations, the S_c values remained slightly above T and occasionally n_k increased in size as $S_c < T$. At later iterations, the increases in n_k appeared to occur in frequent succession before a period of updates with no subset size increase.

The objective function values of list mode AS algorithm reconstruction performance is compared to fixed subset size list mode reconstructions in Figure 3.8. The AS algorithm's use of significantly smaller subset sizes granted the algorithm superior performance during early updates. However, after less than half an epoch, the algorithm's objective function values were outperformed by the $M = 480$ fixed subset reconstruction.

Additional analysis, shown in Figure 3.9, illustrates the mean and standard deviations of the voxels in the top cylindrical insert ROI. This analysis was not included in the projection data results. The AS reconstruction ROI mean voxel value almost converged within a fraction of an epoch. Similar performance was noted for the fixed subset size reconstruc-

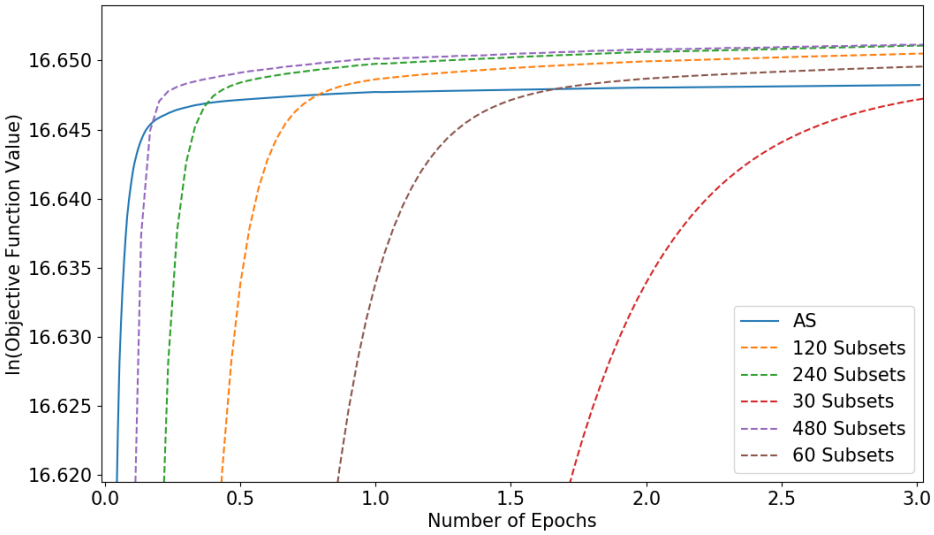


Figure 3.8: A plot comparing the objective function value performance of a list mode AS reconstruction with various fixed subset size reconstructions. The list mode AS reconstruction is initialised with approximately 13,000 coincidence events in each subset. The number preceding OS in the legend indicates the number of subsets used during the reconstruction.

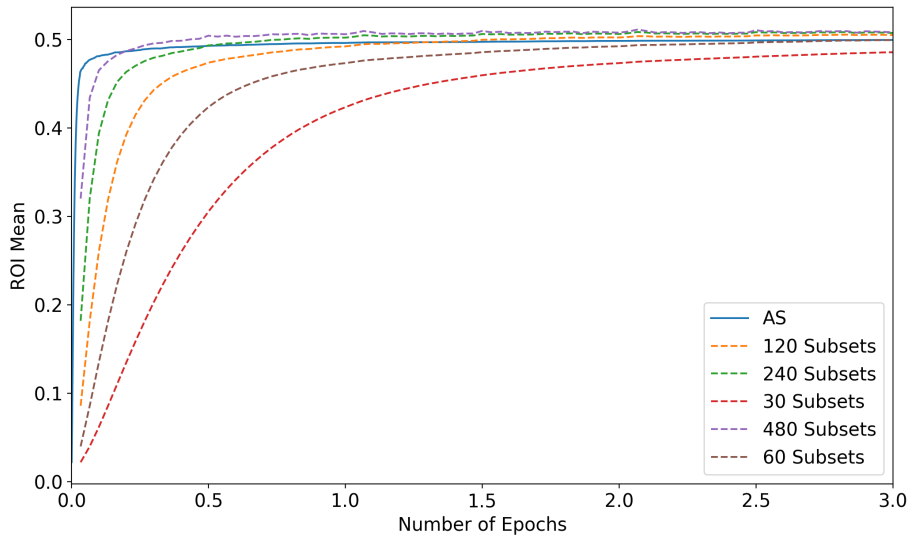
tions (except $M = 30$ subsets). This type of behaviour has been previously observed for list mode data reconstructions in Reader et al. 2002. This mean value converged faster when more subsets were utilised. The ROI's voxel standard deviations continuously increased for all fixed subset size reconstructions and a greater number of subsets resulted in increased voxel standard deviation. Yet, the AS reconstruction standard deviation peaked after a small fraction of an epoch before declining in value.

3.5 Discussion

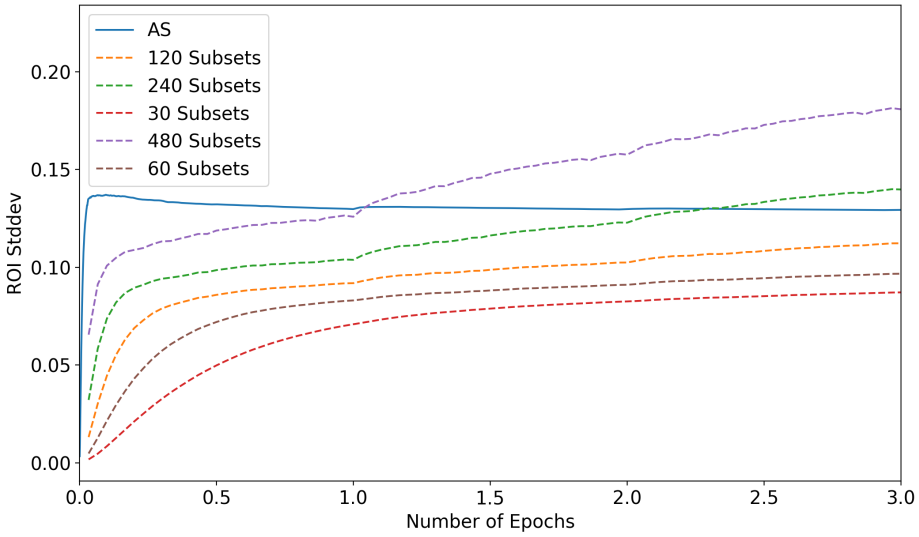
The AS reconstruction algorithm was designed to automatically increase subset size when the direction of two updates, computed using two separate subsets of the data, diverged. In Section 3.4, the AS algorithm was applied to three data sets, two projection data sets and one list mode. Measurements of the algorithms CSI and number of projection angle or sample usage were plotted to assess the methodology. Additionally, reconstruction performance was compared to BSREM.

3.5.1 Subset Sampling Methodology

In Section 3.4.1, the AS algorithm was applied to the toy data set using two subset sampling methodologies: RB and GRS. The results of this study indicated that the GRS sampling



(a) Mean.



(b) Standard Deviation.

Figure 3.9: A plot comparing voxel mean and standard deviations values of the top, high activity, cylinder insert of the phantom (Figure 3.1) over three epochs of the list mode reconstruction using AS and five fixed subset size reconstructions. Sub-figure (a) plots mean voxel values in the ROI while (b) plots the voxel value standard deviation.

method for projection data outperformed the RB sampling method. The S_c measurements were observed to vary significantly between updates with both sampling methods but more-so with RB sampling. Consequently, this may have resulted in the algorithm increasing to full data updates relatively early in the reconstruction, which slowed convergence rates. Similar results were replicated for the D690 projection data set but are not shown in this thesis for clarity.

To explain this poor reconstruction performance associated with RB sampling, consider the subset construction methodologies. Subsets constructed using RB sampling are likely to be more unbalanced than those constructed using OS. This statement is supported by the conclusions of a later conducted study, presented in Section 4.3.1 and discussed in Section 4.4. As aforementioned in section 3.2.1, $S_c(\boldsymbol{p}_{\mathcal{S}_{k,1}}, \boldsymbol{p}_{\mathcal{S}_{k,2}})$ comparisons between two similar subsets will likely result in a higher S_c value than those selected from dissimilar subsets. Hence, the S_c measurements are influenced by the subset selection methodology and this may explain the highly varied S_c values reported in Figure 3.3a.

The GRS sampling method was implemented into the AS algorithm as an attempt to construct balanced subsets based on an arbitrary subset size (n_k). The measured S_c values demonstrated a steadier decline during early updates, compared with the RB AS reconstruction, as shown in Figure 3.3a. As a consequence, the first $S_c < T$ measurement occurred at a later iteration k. Assuming much of the AS algorithm discrepancies in Figures 3.3a and 3.3b are related to their subset selection methodologies, it appears the GRS resulted in improved algorithm performance due to the reduced variations in S_c value measurements.

3.5.2 Premature Increase of Subset Size

As a result of the aforementioned GRS AS reconstruction reduced sensitivity to subset selection methodology, the algorithm was able to increase n_k less frequently. This allowed for the use of smaller subsets for a greater number of iterations, which appeared to translate into improved algorithm performance in Figure 3.3c and Figure 3.6.

In comparison to the AS algorithm, the 40 and 70 OS reconstructions plotted in Figure 3.4 tend towards an asymptote in approximately 20 epochs. Similar reconstruction performance was observed for the D690 projection data set in Figure 3.6 as the $M = 18$ and $M = 36$ reconstructions objective function values plateaued after approximately 2.5 epochs. In both of these figures, the AS reconstruction performed comparably to these reconstructions during the early epochs. However, the AS algorithm increasing subset size may have

resulted in poorer later metric performances.

Some iterations where the AS algorithm increased the subset sizes are visible on the plotted AS metrics in the form of non-smoothness in the plot, which resulted in a shallower plot. This metric behaviour indicates an abrupt decrease in convergence rate that was caused by an increase of n_k at a sub-optimal iteration. Without this increased subset size, it can be assumed that the next algorithm iteration would have measured an improved metric value. Hence, the algorithm increased n_k too early. Therefore, either the $T = 0.0$ threshold is too high or the $S_c(\boldsymbol{p}_{\mathcal{S}_{k,1}}, \boldsymbol{p}_{\mathcal{S}_{k,2}})$ metric is not effective at indicating the optimal iterations to increase subset size.

3.5.3 List Mode AutoSubsets Reconstruction

The list mode reconstruction objective function values plotted in Figure 3.8 demonstrates the potential of list mode reconstruction. Both the AS algorithm and the large M subset reconstructions achieve objective function values after 0.2 epochs that are comparable to reconstruction performance after 1 epoch of projection based subsets in Figure 3.6. Hence, with the results of Figure 3.9a, the list mode reconstructions using a large number of subsets demonstrated excellent performance within a list mode data epoch[iv]. This is likely a consequence of the large redundancy between various event data. During the list mode reconstruction, subsets of samples are expected to represent varied, but unbiased, estimates of the full data. Compare this to projection data where subsets are divided by physical projection angles. These projection angle subset construction methodologies may introduce some bias and almost certainly increase the variance between subset data, which will reduce balance.

The list mode S_c values plotted over algorithm iterations, shown in Figure 3.7, were less noisy than those observed for the projection data in Figure 3.5. Besides some larger variations in S_c values during early epochs, the S_c measurements slowly declined towards values just above $T = 0.0$. However, the $S_c < T = 0.0$ condition was not met until $k \approx 45$. This was likely a consequence of the low variance between the list mode subsets. The increase in subset size occurred much slower for the list mode reconstruction, despite the subsets being relatively small with respect to the full data.

As aforementioned in Section 3.4.3, while the list mode AS reconstruction does not optimise the objective function as fast as the fixed subset size methods, it does approximately

[iv] The definitions of list mode and projection data epochs differ and computational cost comparisons between the data processing are hard to evaluate.

converge in mean ROI value. In comparison, Figure 3.9b plots the stardard deviation voxel values in the ROI. This metric was used to measure voxel noise in the ROI, which was assumed to be a smooth region. For all fixed subset size algorithms, these standard deviation measures increased in value throughout the reconstructions. Yet, after an initial increase in voxel value standard deviation during updates with small n_k values, the AS reconstruction's standard deviation decreased. Therefore, the AS algorithm appeared to suppress noise because of the use of larger subsets at later epochs. Hence, although the $M = 480$ subset reconstruction was quick at optimising the objective function in three epochs, the noise in the reconstructed image continued to increase, whereas the AS reconstruction limited this.

3.5.4 Limitations

This chapter presented a preliminary study of the AutoSubsets algorithm. A number of the limitations of this study are discussed in this sub-section.

3.5.4.1 Cosine Metric

The AS algorithm utilises the $S_c(\boldsymbol{p}_{\mathcal{S}_{k,1}}, \boldsymbol{p}_{\mathcal{S}_{k,2}})$ measurement to quantify how similar subset update directions are to one another and when subset size should be increased to maintain similarity between subset data. However, the AS algorithm appeared to increase subset size too early for both projection data sets. The use of a lower T value could assist in delaying the increase in subset size but, as this is currently a heuristic parameter, reconstruction performance might be sensitive to the value selected. Furthermore, the optimal T value may be data dependant or dependant on a number of algorithm factors, i.e., α_k, the relationship between n_k and $n_{\max}$ and Γ_Θ.

3.5.4.2 Experimental Limitations

As this was a preliminary study of the AS algorithm, a number of parameters were selected heuristically. By extension, a number of other reconstruction parameters were kept constant through the reconstructions, including the preconditioner, step sizes and penalty strengths. Varying any of these algorithm configuration parameters may lead to alternative algorithm results. Additionally, the simulated cylindrical reconstruction object was not a complex phantom. Although two distinct acquisitions were simulated, little is understood from the presented results regarding how data noise and scanner configuration might impact the algorithms performance. Finally, a limited selection of post-processing metrics were employed to assess the performance of this algorithm.

3.5.4.3 Fixed Step Size

In Section 3.1 it was stated that algorithm convergence may be achieved by iterating with a fast non-convergent algorithm for a number of epochs before switching to a convergent algorithm. The example given was OSEM and MLEM. However, because of the use of the prior in the objective function in this chapter's numerical studies, the presented AS algorithm was not guaranteed to converge with $\alpha_k = 1$. Alternative methods, such as step size relaxation, should be employed to guarantee convergence.

3.5.5 Follow-Up Study

A follow up study was conducted by another student as part of an MSc research project that built on the work presented in this chapter [Zeng 2021]. The author of this thesis held a supervisory role for that project. A brief discussion of that study is presented in Appendix D.

3.6 Conclusion

In this chapter the AutoSubsets reconstruction algorithm was presented. This algorithm measures the divergence between two subset update directions in order to automatically adapt their size for future updates. AS accelerated all initial reconstructions by allowing the use of small subsets before automatically increasing subset sizes. The image estimate sequences were found to be reasonable and comparable to the MAP solution, yet at later stages, the convergence of the algorithm was sub-optimal when compared to fixed subset size reconstructions.

This method was evaluated using two simulated projection data sets and a simulated list mode data set. Two projection based subset sampling methods were investigated throughout this work, and the Golden Ratio Subset sampling was found to be superior to the Random Batches sampling. Comparatively, the AS algorithm took advantage of the large list mode data-set, accelerating initial reconstruction before increasing subset size and suppressing image noise at later updates.

Chapter 4

Stochastic Gradient Algorithms

4.1 Introduction

Generally, current clinical PET reconstruction algorithms use deterministic Ordered Subsets (OS) algorithms in a preconditioned gradient-ascent/surrogate framework, see Sections 2.3.2.1 and 2.3.2.2. Notable example algorithms are OSEM and BSREM, see Section 2.3.4.1 and Section 2.3.4.3 respectively. These are fast and reliable algorithms during early epochs, an attribute that encouraged their wide acceptance. However, fast convergence is not always realised at later epochs because of limit cycle behaviour and relaxation parameters.

An alternative subset methodology is the use of randomly selected or constructed subsets. In addition, a number of novel stochastic algorithms have been developed that boast faster convergence rates at later epochs. One example stochastic algorithm that has demonstrated excellent reconstruction behaviour for PET image reconstruction is Stochastic Primal-Dual Hybrid Gradient (SPDHG), as discussed in Section 4.6.8.5 [Chambolle et al. 2018; Ehrhardt et al. 2019].

This chapter evaluates the application of a novel class of Stochastic Variance Reduction (SVR) gradient methods to the iterative PET reconstruction problem. Three algorithms of this class are considered: SAG [Roux et al. 2012], SAGA [Defazio et al. 2014] and SVRG [Johnson et al. 2013]. However, as the PET problem is ill-posed, direct application of SAG, SAGA and SVRG (given by Algorithms 1 and 2) would require small step sizes or the algorithms may be non-convergent. Additionally, these SVR algorithms are unconstrained, allowing $0 > x_i \in \boldsymbol{x}$. The *update step* in the aforementioned algorithms is modified to include

preconditioning $D(\boldsymbol{x})$ and a non-negativity constraint $P_{\boldsymbol{x}\geq\mathbf{0}}[\cdot]$, given by

$$\boldsymbol{x}_{k+1} = P_{\boldsymbol{x}\geq\mathbf{0}}[\boldsymbol{x}_k + \alpha_k D(\boldsymbol{x})\tilde{\nabla}_{k,m_k}]. \tag{4.1}$$

The resulting equation closely resembles Equation (2.44). For simplicity, these preconditioned and non-negative constrained algorithms shall be refereed to with their original abbreviations (SAG, SAGA and SVRG[i]) throughout this thesis.

In this chapter, a comparison between deterministic and stochastic subset sampling is conducted in Section 4.3 for both a preconditioned subset gradient algorithm and the SVR algorithms. The results of this investigation are discussed in Section 4.4. Following this study, the performance of the aforementioned SVR algorithms is assessed with various PET data sets in Section 4.5. Numerical experiments are conducted to determine the impact of algorithm configuration by applying various numbers of subsets, preconditioners, step size methodologies and "warm-starting". These results are discussed in Section 4.6. Note, unlike Chapter 3, only projection angle based subsets are considered in this chapter. The following section details the data sets used throughout this chapter.

4.2 PET Data Sets

A number of PET data sets are investigated to assess the applicability of SVR algorithms for PET reconstruction. These data sets are described in the following sub-sections.

4.2.1 Small XCAT Data

For preliminary studies, an XCAT thorax phantom [Segars et al. 2010] scan was simulated using STIR [Thielemans et al. 2012]. The scanner utilised 280 projection angles and two rings using STIR. Poisson noise was added and scattered events were simulated in the measured data. The respective data corrections were included within the forward model and, to further suppress noise in the images while encouraging edges, the RDP, see Section 2.2.3, was used to penalise the objective function [Nuyts et al. 2002]. The penalty strength β was hand-tuned for a qualitatively appealing converged solution $\hat{\boldsymbol{x}}$, which was computed using the L-BFGS-B-PC reconstruction algorithm, see Section 2.3.2.3 [Tsai et al. 2018].

During preliminary studies of the variance reduction algorithms applied to PET reconstruction (results not shown), significant performance improvements were observed during early epochs if the algorithms were initialised from an image with relatively correct scale

[i]Note, SVRG's full gradient re-computation parameter is fixed at $\gamma = 2$ throughout this work.

and structure of the reconstruction object. For this reason, all reconstructions in this chapter were initialised from an image computed using a single epoch of OSEM, denoted as $\boldsymbol{x}_{\text{OSEM}}$. For this data set, $\boldsymbol{x}_{\text{OSEM}}$ was computed using $M = 20$ subsets.

4.2.2 XCAT GATE Simulation Data

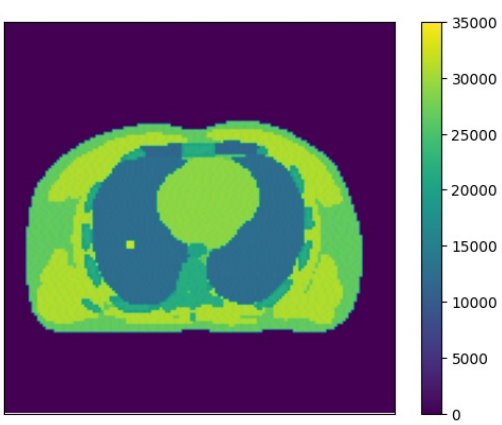
The second class of data sets were acquired by numerical GATE simulations [Jan et al. 2011] modelling the acquisition of the GE Discovery 690 scanner [Bettinardi et al. 2011]. This was conducted with STIR-GATE-Connection , see Appendix A. The photon emission simulation was performed using back-to-back 511 keV photon emissions from a voxelised XCAT thorax phantom [Segars et al. 2010] with activity concentrations representative of an ^{18}F-FDG study (see emission and attenuation distributions in Figure 4.1a and Figure 4.1b, respectively). Cardiac and respiratory motion, along with radioactive decay, were not modelled for experiment simplicity. A 1 cm diameter, 1 cm long, cylindrical hot "lesion", with 2.6:1 lesion to lung contrast, was inserted into the lung of the XCAT emission.

Multiple Region of Interest (ROI) volumes were drawn for image quality analysis. ROIs were drawn around the aforementioned inserted lung lesion and over a large portion of the interior of the liver. Additionally, an ROI was drawn outside of the thorax, where no activity is expected.

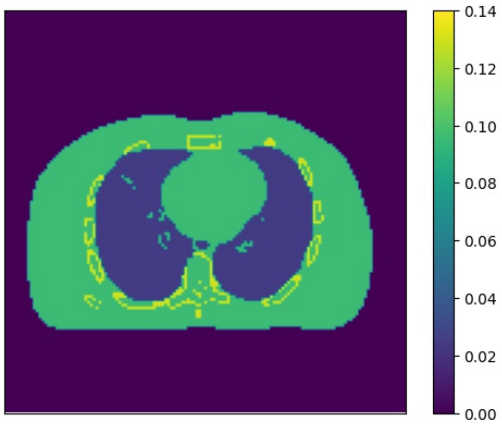
To explore the sensitivity of the SVR algorithms to data noise levels, the list mode file was sub-sampled by event random rejection. Three distinct projection data sets containing 50, 250 and 1200 million coincidence events were acquired. A true-to-background ratio of 0.93:1 was maintained across the three data sets, where background is defined as random plus scattered events. This resulted in data sets with low (50 million events) to high (1200 million events) SNR, which were re-binned into projection data with 288 projection angles. The resulting projection data was span 1 with a maximum ring difference of 45.

As discussed in Section 2.2.1, normalisation factors are required for an accurate system model. The normalisation factors were computed by a comparison between forward projections in STIR and GATE of a cylindrical phantom the size of the Discovery 690's field of view. These factors, as well as attenuation correction factors, randoms and scatter models, were computed for each data set using STIR-GATE-Connection tools, see Appendix A.3 for more details.

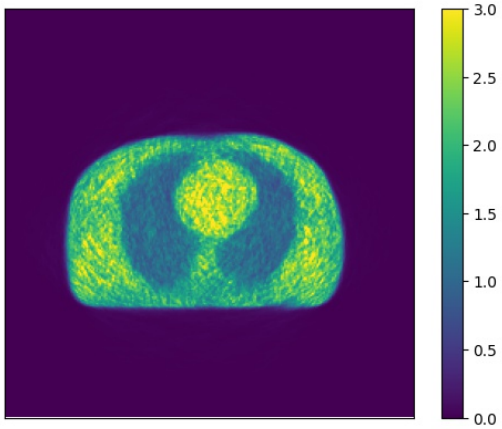
An initial epoch of OSEM using $M = 24$ subsets was performed for each data set resulting in $\boldsymbol{x}_{\text{OSEM}}$. This initial image for the 50 million event data set is depicted by Figure 4.1c.



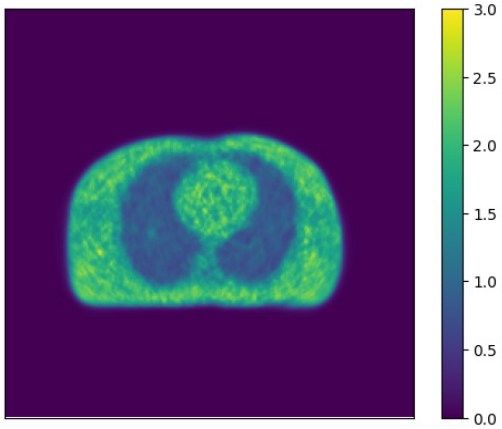
(a) Photon Emission Distribution



(b) Attenuation Distribution



(c) OSEM Reconstruction (1 epoch with 24 subsets)



(d) Converged Solution

Figure 4.1: Transaxial slices of various distributions for the 50 million event simulated XCAT data set. (a) Emission and (b) attenuation (cm^{-1}) consisted of $(141,141,47)$ voxels of size $(3,3,3.27)$ mm. Reconstructed images (c) and (d) contained $(251,251,47)$ voxels of size $(2.13,2.13,3.27)$ mm.

The spatially variant penalty strength $\kappa(\boldsymbol{x}_{\text{OSEM}})$ was computed for each of the three data sets using Equation (2.11). The RDP was used as the prior throughout these studies with the edge preservation parameter $\nu = 2$. The penalty strengths for each of the data sets were determined based on qualitative assessment and had values $\beta_{50M} = 0.008$, $\beta_{250M} = 0.04$ and $\beta_{1200M} = 0.2$. The linear relationship was maintained between these β values based on the number of events in each data set. This was to maintain spatial resolution [Tsai et al. 2020].

Unlike the small XCAT data of Section 4.2.1, the converged solution $\hat{\boldsymbol{x}}$ was obtained by reconstructing the data using Equation (2.51) with SVRG and the modified-EM preconditioner, given by Equation (2.29). The reconstruction algorithm's step size α_k was reduced if $\Phi(\boldsymbol{x}_k) < \Phi(\boldsymbol{x}_{(k-\gamma M)})$ when measured periodically at each epoch. An initial step size $\alpha_0 = 1$

was selected and if the inequality was not satisfied, $\alpha_{k+1} = \alpha_k$, otherwise, $\alpha_{k+1} = \nu\alpha_k$, where $\nu = 0.9$ was selected heuristically. The algorithm was terminated and $\hat{\boldsymbol{x}} = \boldsymbol{x}_k$ when the estimate satisfied the Karush-Kuhn-Tucker conditions (within numerical precision), see Equation (2.21) [Nocedal et al. 2006]. For the 50 million event data, this algorithm termination occurred after more than 50,000 epochs with $M = 24$ and a final $\alpha_k \approx 1e-9$, which resulted in updates that were numerically unstable with floating point computations. The reconstructions computing $\hat{\boldsymbol{x}}$ for the 250 and 1200 million event XCAT data sets performed similarly. This methodology was faster at computing $\hat{\boldsymbol{x}}$ for 3D data than a line search preconditioned gradient ascent algorithm (see Section 2.3.3.3), and the line search in L-BFGS-B-PC was found to be sensitive to numerical instability when close to $\hat{\boldsymbol{x}}$.

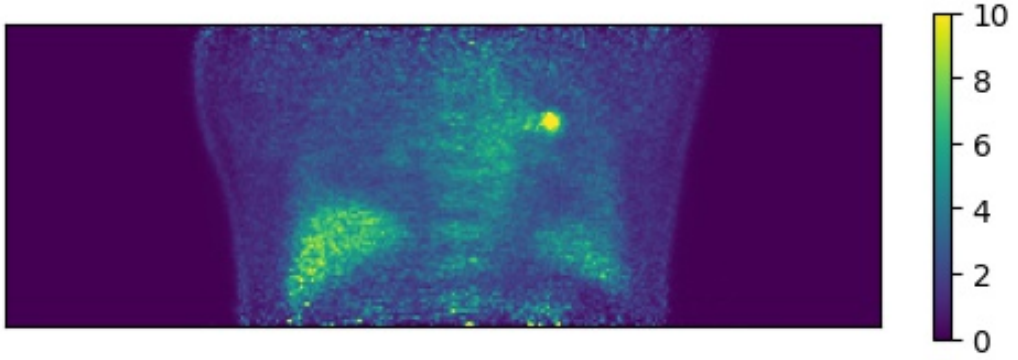
4.2.3 Patient Data

A 150 second ^{18}F-FDG static patient scan was acquired on a 5 ring GE Discovery MI (272 projection angles) [Pan et al. 2019] and 143 million events were recorded. The acquired list mode data was processed into projection data and non-TOF statistical data models were computed with GE's Duetto Toolbox and converted into a STIR compatible interfile format. The patient was part of a prospective study, approved by the ethical committee (BASEC-Nr 2018-01012), investigating whole body dynamic PET. The patient gave informed consent for further use of their data. The data described here represents a fraction of the full dynamic scan. This static scan was acquired at a later time in the full scan duration to minimise the impact of PET tracer kinetics. No modelling or correction of motion is applied to this data.

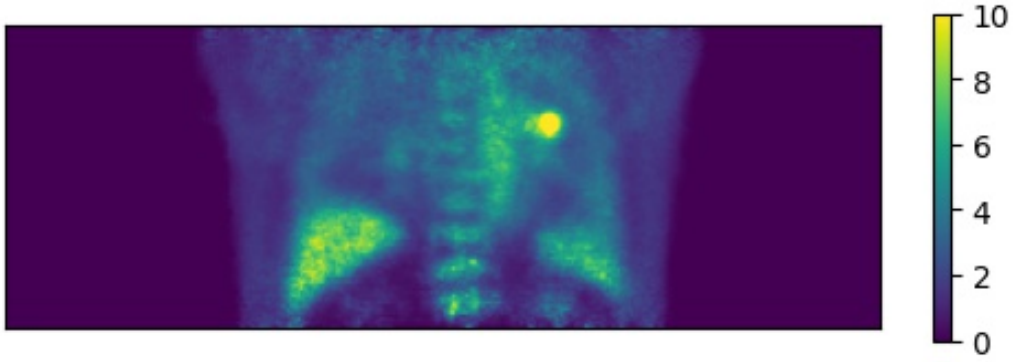
The patient was diagnosed with a bronchial carcinoma and a hilar lymph node metastasis in the left upper lobe of the lung. The bronchial carcinoma is shown in Figure 4.2. A $(20 \times 26 \times 20)$ mm^3 ellipsoidal ROI was drawn over the bronchial carcinoma so the entire carcinoma is contained within the ROI. Additionally, $(12 \times 15 \times 16)$ mm^3 ellipsoidal ROI was drawn over the hilar lymph node metastasis with the same criteria. Finally, a $(50 \times 60 \times 60)$ mm^3 ellipsoidal ROI was drawn within the patient's liver with a minimum 10mm distance maintained from the organ's boundary. Liver ^{18}F-FDG uptake region was expected to be high but uniform allowing for measures of noise to be made [Ahn et al. 2015]. Furthermore, as the injected dose was known, Standardised Uptake Value peak (SUVpeak) and SUV values can be computed [Boellaard et al. 2015].

Similar to Section 4.2.2, $\kappa(\boldsymbol{x}_{\text{OSEM}})$ was computed and β selected based on qualitative reconstructed image assessment. An epoch of OSEM, with $M = 17$, was used to com-

pute $\boldsymbol{x}_{\text{OSEM}}$. The converged solution $\hat{\boldsymbol{x}}$ was computed using the same methodology as Section 4.2.2.



(a) OSEM (1 epoch with 17 subsets)



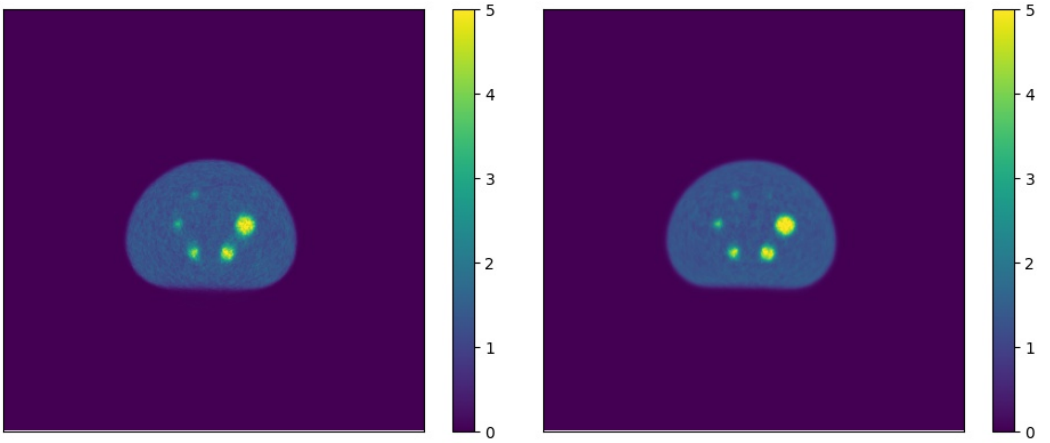
(b) Converged Solution

Figure 4.2: Coronal slices in kBq/mL units of the patient data set reconstruction with a bronchial carcinoma in the left lung upper lobe. Reconstructed images consist of $(323, 323, 89)$ voxels of size $(2.21, 2.21, 2.76)$ mm.

4.2.4 NEMA Phantom

NEMA phantom reconstructions are a benchmark of many works due to their wide adoption throughout the clinical environment. In this work, a Germanium NEMA phantom, with a 4:1 spheres to background ratio, was scanned in a GE Discovery 710 PET/CT to realise 120 million coincidence events. The measured list mode data was processed into non-TOF projection data and statistical data models were computed similarly to the aforementioned patient data. A CT was concurrently acquired for attenuation correction. ROIs were drawn around each of the six spheres to assess reconstruction performance.

Similar to Section 4.2.2, $\kappa(\boldsymbol{x}_{\text{OSEM}})$ was computed and β selected based on qualitative reconstructed image assessment. An epoch of OSEM, with $M = 24$, was used to compute



(a) OSEM **(b)** Converged

Figure 4.3: Transaxial slices of NEMA reconstructions. (a) OSEM reconstruction, performed with 24 subsets, and (b) converged image reconstruction. Reconstructed images contained $(337, 337, 47)$ voxels of size $(2.13, 2.13, 3.27)$ mm in kBq/mL.

$\boldsymbol{x}_{\text{OSEM}}$, which is shown in Figure 4.3a. The converged solution $\hat{\boldsymbol{x}}$ was computed using the same methodology as Section 4.2.2.

4.3 Subset Selection Methodology Investigation

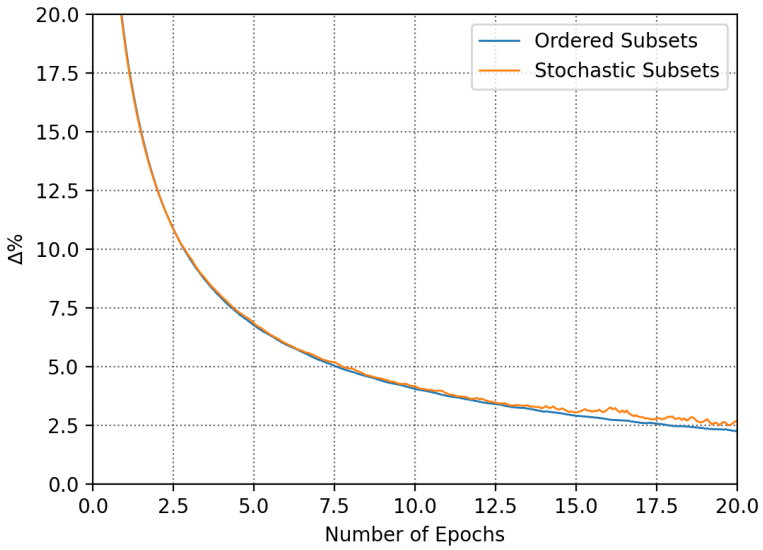
This section introduces two stochastic subset methodologies for PET image reconstruction with the aim of exploring the impact subset selection has on reconstruction performance. For multiple algorithms, performance using these methods is compared to the OS method, as described in Section 2.3.4 and Herman et al. 1993. Four algorithms were investigated: a preconditioned gradient ascent and preconditioned SAG, SAGA and SVRG. This study solely utilises the "small XCAT data", described in Section 4.2.1 and the SVR algorithms are initialised from $\boldsymbol{x}_{\text{OSEM}}$. These results were published in Twyman et al. 2021a.

The first stochastic subset sequence methodology, denoted as "Stochastic Subsets", uses the same subset construction method as OS. However, at each iteration k of the reconstruction algorithm, a subset index is selected at random from a uniform probability distribution. The second stochastic subset methodology is "Randomised Batches" (RB), introduced in Section 3.2.1.1. In this method, a number of projection angles are selected at random (without replacement) and added to a subset. A subset gradient is computed from this set of random projection angles and, at each update, a new subset is selected from all scanner projection angles. This does not enforce disjoint subsets between each update.

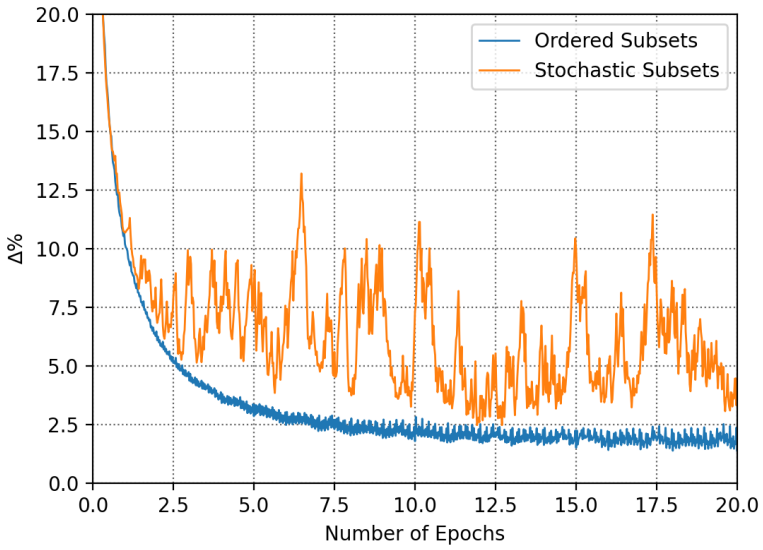
To assess the performance of the subset methodologies, the Δ distance metric was

reused, see Equation (3.4). However, may plots are given as $\Delta\% = 100\% \cdot \Delta$. Furthermore, to isolate the impact of subset selection from other variables, such as variations in the preconditioner, a constant step size of $\alpha_k = 1$ was employed along with a constant $\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_{\text{OSEM}})$ preconditioner, given by Equation (2.29). This preconditioner is "anchored" at $\boldsymbol{x}_{\text{OSEM}}$.

4.3.1 Subset Gradient Ascent



(a) 14 Subsets



(b) 40 Subsets

Figure 4.4: Global convergence performance, measured with the $\Delta\%$ metric, of $\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_{\text{OSEM}})$ preconditioned subset gradient ascent reconstructions using the OS and stochastic subsets methodologies, with (a) 14 subsets and (b) 40 subsets.

The constrained (preconditioned) subset gradient ascent update equation is given by

Equation (2.44) and reconstruction performance of the subset selection methodologies is shown in Figure 4.4. The randomised batches reconstructions are not included in either of these sub-figures as the measured $\Delta\%$ metric values indicated significantly poorer performance and seldom were $\Delta\%$ measurements less than 20.

Reconstruction performance using $M = 14$ subsets with various subset methods is shown in Figure 4.4a. The OS method exhibited slightly improved performance compared to the stochastic subsets method. However, this improvement was only realised after approximately 15 epochs where the impact of the randomly selected subsets began to affect the $\Delta\%$ performance.

Reconstructions using an increased number of subsets ($M = 40$) is shown in Figure 4.4b. Both OS and stochastic subsets demonstrated an improved initial convergence rate during the first few epochs when compared to the $M = 14$ reconstructions. Yet, the stochastic subsets method exhibited larger fluctuations in $\Delta\%$ values and generally performed poorer than the OS method after approximately 3 epochs.

4.3.2 Variance Reduction Methods

The (preconditioned) variance reduction algorithm update step, given by Equation (4.1), utilised the same $\alpha_k = 1$ and $\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_{\text{OSEM}})$ preconditioner as above. These algorithms are capable of utilising more subsets than the subset gradient ascent algorithms, as discussed in Section 4.5.3, and therefore utilise $M = 70$ subsets. For convenience, regardless of subset methodology, SAG, SAGA and SVRG shall be referred to as such.

The $\Delta\%$ metric performance of SAG, shown in Figure 4.5a, indicates that the stochastic subsets method optimised the reconstruction problem at a faster rate than Random Batches and with fewer inter-update variations. During the first four epochs, the OS method outperformed both of the stochastic methods. After this point, the algorithm demonstrated divergent behaviour as the $\Delta\%$ values increased rapidly.

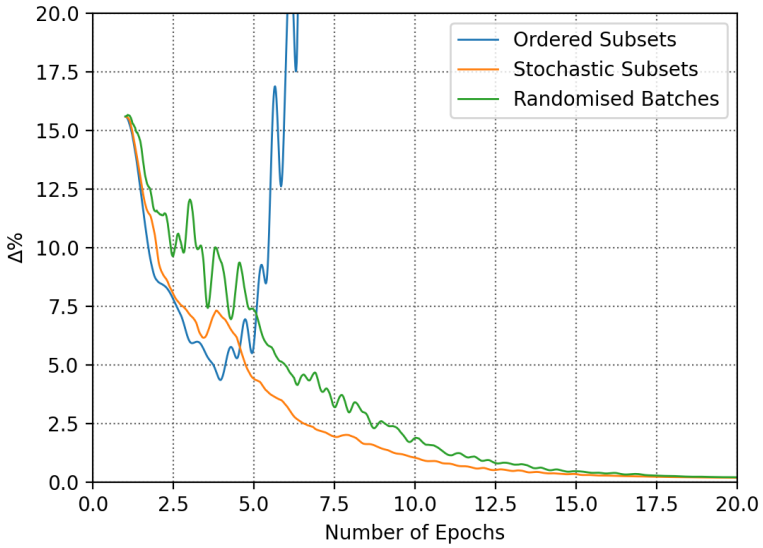
For the various SAGA reconstructions, depicted in Figure 4.5b, the OS method outperformed the two stochastic methods, both of which exhibited significant fluctuations during the first 6 epochs. After this, reconstruction performance became comparable and all plots converged within $1\Delta\%$.

The subset selection variations of SVRG are shown in Figure 4.5c and all three methods reduced the $\Delta\%$ metric quickly and consistently. The OS method reconstruction marginally outperformed both of the stochastic methods during early iterations. At later

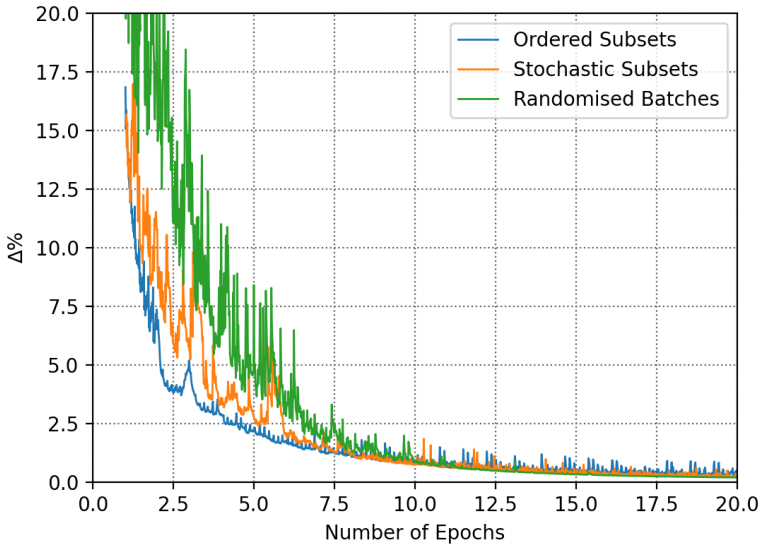
iterations, all three methods performed equally as the reconstructions approached convergence.

4.4 Discussion of Subset Selection Methodology

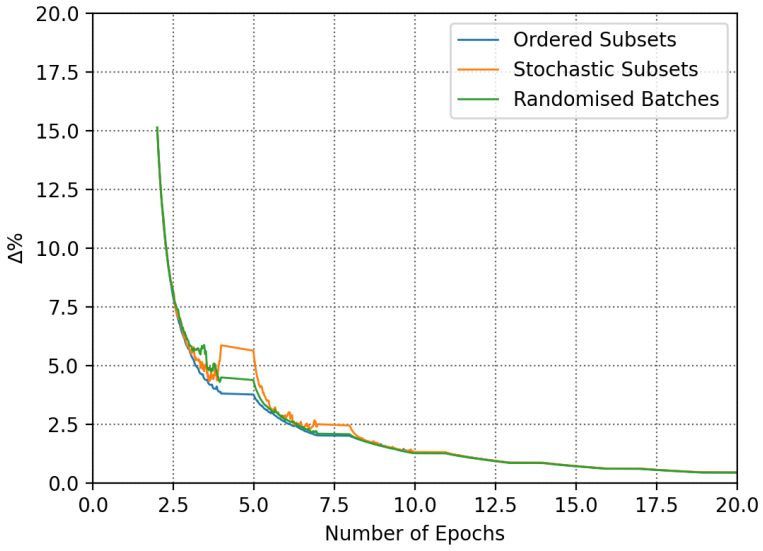
The impact of three subset methodologies was assessed for four reconstruction algorithms in Section 4.3. These subset methods were Ordered Subsets (OS), Stochastic Subsets and Random Batches and reconstruction performance was measured over multiple epochs using the $\Delta\%$ metric. This section discusses the impact of these results.



(a) SAG



(b) SAGA



(c) SVRG

Figure 4.5: Global convergence performance, measured with the $\Delta\%$ metric, of $\boldsymbol{D}_{\mathrm{EM}}(\boldsymbol{x}_{\mathrm{OSEM}})$ preconditioned (a) SAG, (b) SAGA and (c) SVRG reconstructions using the OS and stochastic subsets methodologies, with $M = 70$ subsets.

4.4.1 Stochastic Gradient Ascent

The first study, see Section 4.3.1, applied these subset selection methods to a subset gradient ascent algorithm where it was observed that the OS method outperformed the stochastic subsets method in Figure 4.4. Algorithm performance using stochastic subsets appeared to worsen when a greater number of subsets was used. The randomised batches reconstructions were not included in either of the plots in Figure 4.4 for clarity because a $\Delta\%$ value greater than 20 was generally observed. Thus, this evidence indicates that subset gradient ascent algorithms perform better with a non-stochastic subset sequence. It should be noted, stochastic algorithms are random and there may exist a subset selection sequence $\{m_k\}_{k=1}^{K}$ that allows a reconstruction algorithm to converge faster and closer to the solution than the OS method. Additionally, subset gradient ascent algorithms appear to benefit from structured subsets, which almost certainly relates to balanced subsets.

4.4.2 Variance Reduction Algorithms

The SAGA and SVRG reconstructions in Figure 4.5b and Figure 4.5c were able to converge within $1\Delta\%$ in the first 20 epochs. The OS method applied to these algorithms led to faster convergence during early iterations than the reconstructions using either stochastic method. As the algorithms approached convergence, the two stochastic methods exhibited

either comparable or superior performance to OS. Therefore, one might consider a hybrid subset method configuration whereby the OS method is used during the first few epochs before switching to a stochastic subset selection methodology. However, this suggestion is heuristic and would require additional study as implementation would likely be data and application specific.

The randomised batches method was consistently outperformed by the stochastic subsets method until later epochs. Moreover, implementation of this subset methodology is complex for the variance reduction algorithms because every projection angle required its own $\boldsymbol{g}_\theta$ entry, where θ is the index of each of the 280 projection angles. In some instances, this may become expensive in computer memory. Hence, this subset methodology was not further considered this section (Section 4.3).

The primary outlier result of this study was the OS SAG reconstruction in Figure 4.5a that appeared to diverge after a few epochs. The use of a step-size that is too large might explain this behaviour. OS SAG (or deterministic cyclical subset selection SAG) is equivalent to an algorithm known as Incremental Aggregated Gradient (IAG) [Blatt et al. 2007]. The IAG algorithm is not as robust as SAG because it requires much smaller step sizes in order to converge [Schmidt et al. 2017]. It may be that OS SAG, with a reduced step size, might demonstrate non-divergent behaviour but this would likely significantly decelerate reconstruction.

4.4.3 Limitations

This study had several limitations, including limited experimental configurations, i.e., the use of a single reconstruction object and noise level, a fixed step size $\alpha_k = 1$ with a fixed preconditioner $\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_{\text{OSEM}})$ and a basic simulated PET data acquisition.

An additional limitation of this study is that repeat stochastic reconstructions will result in different reconstruction sequences $\{\boldsymbol{x}_k\}_{k=0}^{K}$, which may lead to different results. The deterministic OS reconstructions do not suffer from this uncertainty. The stochastic reconstruction performance presented in Section 4.3 demonstrated the general behaviour observed during algorithm development. However, to verify these findings, multiple stochastic realisations should be considered for each algorithm configuration.

Finally, only the Δ metric was considered to evaluate algorithm performance. This requires the computation of $\hat{\boldsymbol{x}}$, which is fairly computationally demanding for this small XCAT data set. The determination of $\hat{\boldsymbol{x}}$ for larger PET data sets may require tremendous

computational effort.

4.4.4 Conclusions

The stochastic subset selection methods applied to subset gradient ascent of the PET data did not perform as well as the OS methodology. However, the variance reduction algorithms did not appear as sensitive to subset selection, except the OS SAG reconstruction. Although algorithm performance of OS SAGA and OS SVRG appeared promising and comparable to the stochastic subset method, the application of OS to the variance reduction algorithms was not continued in future studies after this section. This was partially because the OS versions of these algorithms do not enjoy the same convergence proof (i.e., in expectation) as stochastic subset selection methods. However, this study was able to isolate the impact of various subset selection methods to some degree and the conclusions should generally translate to fully 3D PET data reconstructions and other algorithm configurations.

4.5 Applying Stochastic Variance Reduction Algorithms to PET Reconstruction

Section 4.3 and Section 4.4 provided a preliminary demonstration of the SAG, SAGA and SVRG algorithms applied to PET image reconstruction. This section will explore the various configurable parameters of the SAGA and SVRG algorithms and study their efficacy for PET image reconstruction. Additionally, comparisons are made between these algorithms and BSREM. The SAG algorithm is not evaluated in this section because of poor performance in the previous section. It was generally observed to be outperformed by the SAGA and SVRG algorithms in similar configurations throughout this study (results not shown). This may be related to the algorithm's gradient estimator bias, see Appendix C.

This section builds on Sections 4.3 and 4.4, as well as published works (i.e., Twyman et al. 2020 and Twyman et al. 2021a). Section 4.5.1 provides an in-depth discussion of the application of the variance reduction algorithms to PET, with particular focus on the step sizes and preconditioning modifications to the SAGA and SVRG algorithms. Section 4.5.2 details the experimental configuration used in this section, including the various ROI metrics employed and the use of multiple stochastic realisations to assess expected algorithm performance. Sections 4.5.3 to 4.5.5 present a number of numerical studies that investigated the impact of the number of subsets, preconditioners and step sizes. These studies were used to identify a generally well performing algorithm configuration. The impact

of prepopulating the SAGA gradients in memory is investigated in Section 4.5.6 and Section 4.5.7 demonstrates the impact of noise levels on SAGA's and SVRG's reconstruction performance. Finally, Section 4.5.8 demonstrates algorithm reconstruction performance applied to the patient data set and NEMA phantom scan data, which are described in Section 4.2.3 and Section 4.2.4 respectively.

4.5.1 Application to PET Image Reconstruction

SAGA and SVRG will converge, almost surely, to the PML solution $\hat{\boldsymbol{x}}$, provided the step size α is upper bounded by a constant $\alpha_{\mathcal{L}} \propto 1/\mathcal{L}$, see Section 2.3.3.1 and Section 2.3.5.2. It is assumed that $\mathcal{L}$ exists[ii] [Kereta et al. 2021, Theorem 4.1]. However, there are challenges in determining $\mathcal{L}$, which is parameterised by the largest eigenvalue of the system matrix and the measured data [Arridge et al. 2019]. The system matrix cannot be predetermined for a given scanner because of the inclusion of an attenuation model $\boldsymbol{A}_{\text{attenuation}}$ in Equation (2.4). Thus, $\mathcal{L}$ will vary between reconstruction objects (i.e., clinical patients or phantoms).

In line with Section 4.3, preconditioned SVR algorithms are considered in this section to improve the ill-posedness of the reconstruction problem. Algorithm convergence proofs remain valid if the preconditioner is fixed and positive-definite [Kereta et al. 2021]. This is because preconditioning can be considered a scaling of the problem and system matrix [Tsai et al. 2018]. Ideally, this decreases $\mathcal{L}$ to a preconditioned value to $\mathcal{L}_D$, thereby increasing $\alpha_{\mathcal{L}}$ to $\alpha_{\mathcal{L}_D}$ and allowing the SVR algorithms to converge with larger constant step sizes. However, it remains a challenge to compute $\alpha_{\mathcal{L}_D}$. In this section, variations of the modified EM preconditioner $\boldsymbol{D}_{\text{EM}}(\boldsymbol{x})$, given by Equation (2.29), are evaluated.

The SVR reconstructions in Figure 4.5 were able to approach $\hat{\boldsymbol{x}}$ within 1 $\Delta\%$ with the use of the constant strictly positive and diagonal preconditioner $\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_{\text{OSEM}})$ and $\alpha_k = 1$. Yet, convergence is not guaranteed for such an algorithm configuration because it is not known if $\alpha_k = 1 < \alpha_{\mathcal{L}_D}$. Therefore, in this section, diminishing step size sequences are investigated in tandem with preconditioning. The relaxation methodology used in this chapter is a modification of Equation (2.35), given by

$$\alpha_k = \frac{\alpha_0}{\frac{\eta}{M}k + 1}. \tag{4.2}$$

[ii] A modified likelihood function is required if $\bar{b}_j = 0$ for any j [Kereta et al. 2021]. However, for any realistic PET reconstruction (like those included in this work) $\bar{b}_j > 0 \; \forall j$. The proof also requires $\nabla\Phi_m$ to be bounded on $\mathcal{B} \triangleq \{0 \leq \boldsymbol{x} \leq U\}$ (where U is an upper bound [Ahn et al. 2003, Appendix A]) and $\Phi_m(\boldsymbol{x})$ to be concave and smooth over this bounded region.

This modification is used to allow for comparable step size magnitudes at similar epochs of algorithms with different numbers of subsets. Building on the convergence proof in Ahn et al. 2003, III.B for diagonally scaled incremental gradient methods, e.g., Equation (2.44), it is expected that the aforementioned almost sure convergence remains valid for relaxed step size sequences and a constant preconditioner. However, a complete proof of this is still unavailable. Nevertheless, the following experiments confirm the numerical convergence, within a reasonable tolerance, if step size parameters are suitably selected.

The $\boldsymbol{D}_{\mathrm{EM}}(\boldsymbol{x}_k)$ preconditioner is reused in the following numerical experiments. Yet, it is unknown if the convergence proof for the SVR algorithms with this variable preconditioner is valid. Unlike BSREM and diagonally scaled incremental gradient methods, the SVR algorithms retain information in memory regarding previous updates in the reconstruction sequence, i.e., the $\boldsymbol{g}_m$ variables. As aforementioned, preconditioning may be considered a scaling of the objective function. Thus, variations in this scaling of the objective function between updates may generate issues for the SVR algorithms. That being said, Figure 4.5 and the following numerical results indicate variable preconditioners may demonstrate good reconstruction performance and converge.

4.5.2 Experimental Configuration

XCAT GATE simulation data sets (Section 4.2.2), a NEMA phantom scan (Section 4.2.4) and a static patient acquisition (Section 4.2.3) were used to assess algorithm performance via a number of metrics. Previous studies in this thesis focused on global image convergence by measuring Δ and objective function values. This section transitions into investigating local convergence properties of various objects of interest using ROI values. ROI percentage error is given by

$$\text{ROI Percentage Error}_k = 100\% \cdot \frac{\Omega_k^r - \hat{\Omega}^r}{\hat{\Omega}^r}, \tag{4.3}$$

where Ω_k^r and $\hat{\Omega}^r$ indicate the mean, or standard deviation, of the voxel values in the ROI of $\boldsymbol{x}_k$ and $\hat{\boldsymbol{x}}$ (converged solution) respectively, for the region r. In addition, the objective value $\Phi(\boldsymbol{x})$ was computed after every epoch of the reconstructions.

Repeated stochastic reconstructions will result in different reconstruction sequences $\{\boldsymbol{x}_k\}_{k=0}^K$ that may lead to different $\boldsymbol{x}_K$ estimates. The deterministic subset algorithms do not suffer from this uncertainty. Therefore, a number of stochastic realisations S_R are performed for each algorithm configuration. At each algorithm update, the mean and standard deviation of the metric values were computed.

As an alternative to mean and standard deviations, the general convergence performance of these algorithms was also assessed using a percentage NRMSE (NRMSE%) metric at each update k, given by

$$\text{NRMSE}\%_k = 100\% \cdot \frac{\sqrt{\frac{1}{S_R}\sum_s^{S_R}(\Omega_{k,s} - \hat{\Omega})^2}}{\hat{\Omega}}, \tag{4.4}$$

where s indexes the S_R stochastic reconstruction realisations.

All reconstructions in the following results were computed from an initial $\boldsymbol{x}_{\text{OSEM}}$ reconstruction. This is included in the computational cost for relevant figures with updates beginning from epoch 1. However, the cost of computing $\kappa(\boldsymbol{x}_{\text{OSEM}})$, given by Equation (2.11), is not included as it is considered an optional objective function modification and should not impact reconstruction performance like warm starting from $\boldsymbol{x}_{\text{OSEM}}$ does.

4.5.3 Number of Subsets

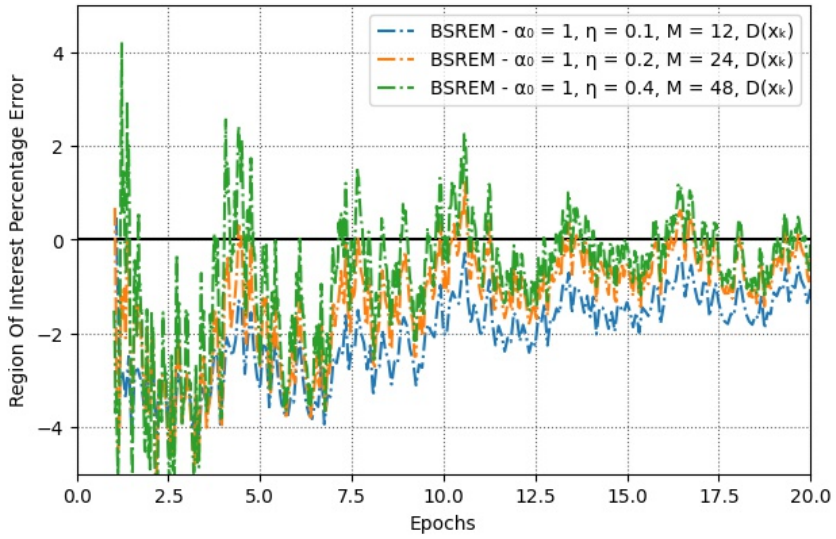
Standard PET subset algorithms (e.g., BSREM) are accelerated but may demonstrate limit cycle behaviour. An example of this behaviour may be observed in Figure 4.4 and in the BSREM reconstructions in Figure 4.6a, even with step size relaxation. Although the reconstruction sequences get close, they rely on $\alpha_k \to 0$ to converge. Additionally, the reconstructions shown in Figure 4.6a were selected from an array of M and η reconstruction configurations with the aim to demonstrate optimal performance in the figure for reconstructions with various M values. Not all results are shown. From this heuristic selection, it can be concluded that reconstructions using more subsets require a greater η value.

SVRG reconstructions using various numbers of subsets, the $\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_k)$ preconditioner and no step size relaxation, are shown in Figure 4.6c. The usage of fewer subsets resulted in slower convergence of the mean ROI performance but the stochastic realisation standard deviations were greatly reduced. The mean values for each of the configurations converged to 0% lung lesion error with respect to the converged solution $\hat{\boldsymbol{x}}$ and the stochastic standard deviation also converged to approximately 0% at later epochs. SAGA reconstructions exhibited similar behaviour to SVRG in Figure 4.6b, albeit with larger stochastic realisation deviations at early epochs.

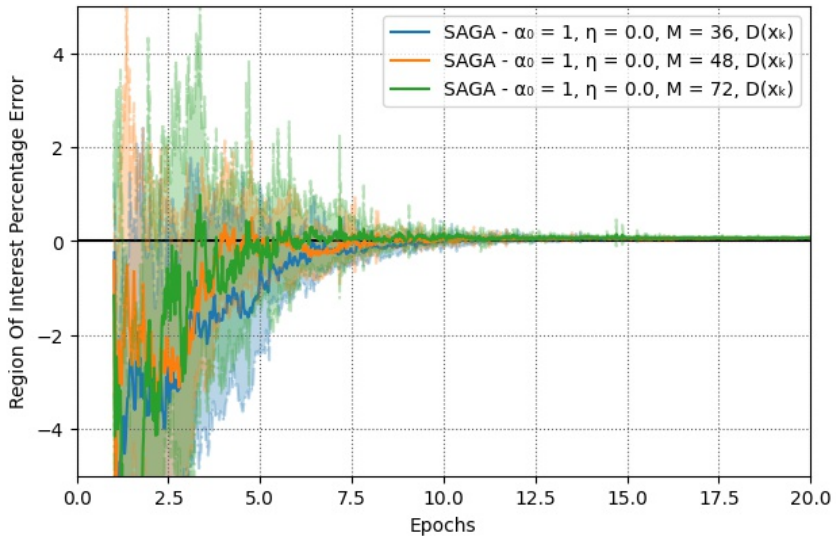
Based on these results, the remaining experiments in this chapter investigated SAGA and SVRG with 72 subsets (or a similar number, depending on the scanner) as a reasonable trade-off between convergence rate and stochastic variation.

4.5.4 Preconditioners

In Section 4.5.3, the SVR algorithms were implemented with the non-constant diagonal preconditioner $\boldsymbol{D}_{\mathrm{EM}}(\boldsymbol{x}_k)$. As discussed in Section 4.5.1, convergence of SAGA and SVRG using such a preconditioner is unknown. Yet, it is apparent that SAGA and SVRG can perform well during the first 20 epochs, as is evident by Figures 4.6b and 4.6c. Therefore, a hybrid methodology was devised to allow the preconditioner to vary for a number of epochs before "anchoring" the $\boldsymbol{x}$ input at a later epoch. Once anchored, the preconditioner becomes

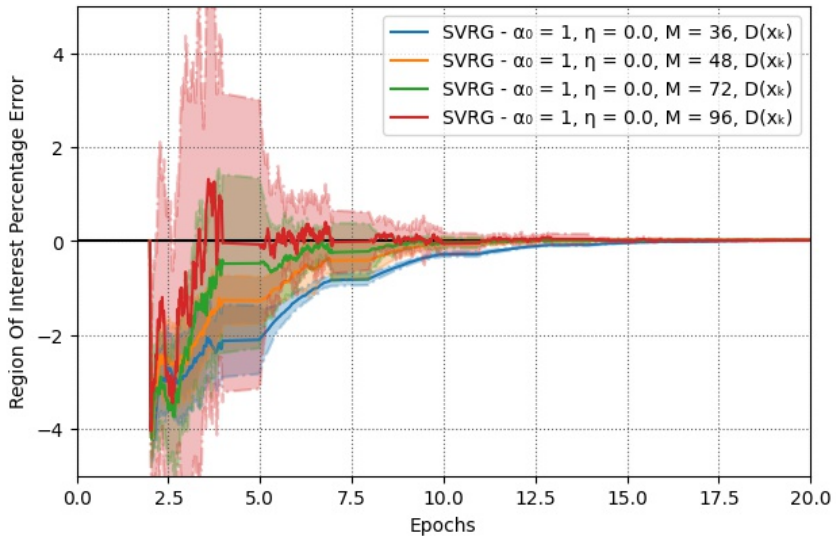


(a) BSREM



(b) SAGA

Figure 4.6: (*Page 1/2*)



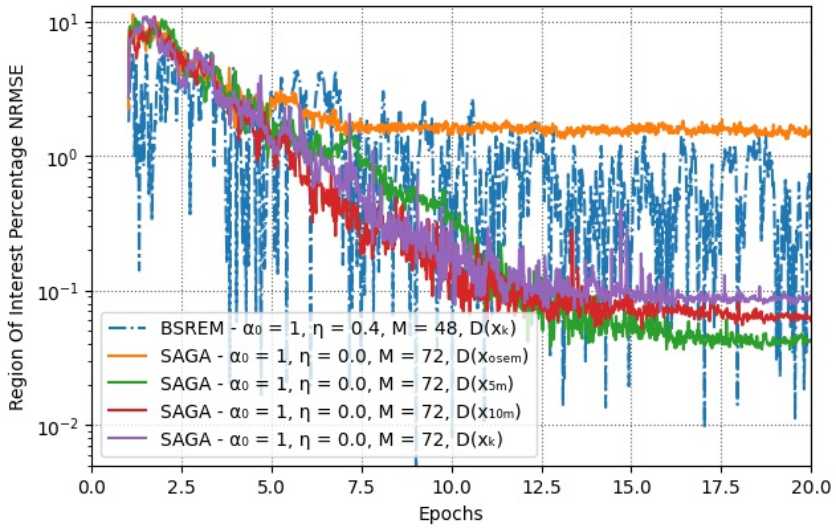
(c) SVRG

Figure 4.6: (*Page 2/2*) The mean lung lesion percentage error of (a) BSREM, (b) SAGA and (c) SVRG reconstructions using various numbers of subsets. The implemented preconditioner was $\boldsymbol{D}(\boldsymbol{x}_k)$. 15 stochastic realisations were used to compute the mean and the shaded region indicates the stochastic realisation standard deviation at each update. BSREM is shown with various η values, which were chosen to demonstrate strong metric performance, but were not optimised. The stochastic algorithms utilised no step size relaxation ($\eta = 0$). The 96 subset SAGA profile is not included for figure clarity.

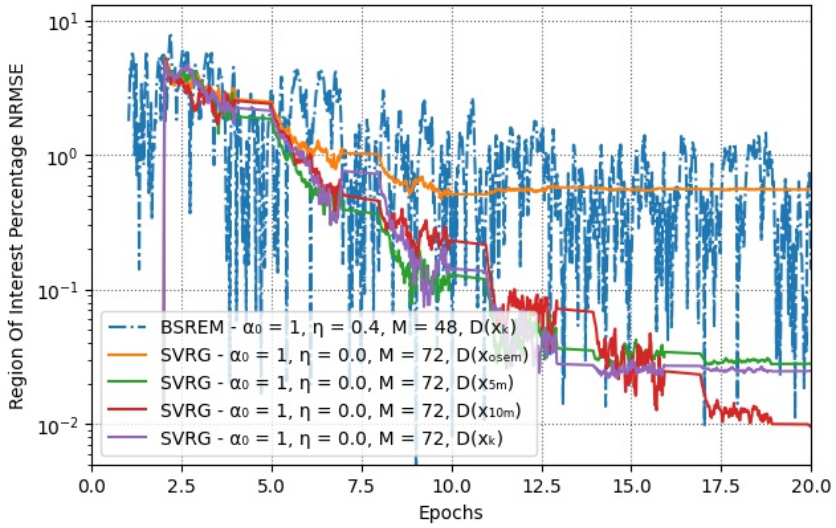
constant, diagonal and positive definite, which allows for algorithm convergence with a suitable step size [Kereta et al. 2021].

The impact of this anchoring after various periods of computation is investigated in Figure 4.7. For this study, a constant step size $\alpha_k = 1$ was implemented, i.e. $\eta = 0$. Four different preconditioners were considered: $\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_{\text{OSEM}})$, $\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_{\text{5M}})$, $\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_{\text{10M}})$ and $\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_k)$. The two hybrid methods, $\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_{\text{5M}})$ and $\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_{\text{10M}})$, vary as $\boldsymbol{x}_k$ for 5 and 10 epochs respectively. After the indicated computational period, i.e., 5 and 10 epochs, $\boldsymbol{x}_{\text{5M}}$ and $\boldsymbol{x}_{\text{10M}}$ are set to a constant input. This constant value is the last iterate $\boldsymbol{x}_k$ before the 5^{th} and 10^{th} epoch. This has the effect of "anchoring" the preconditioner after a period of computation using a variable preconditioner.

Figure 4.7 plots the NRMSE% performance of these preconditioned SAGA and SVRG reconstructions and compares to a BSREM algorithm. During the first 5 epochs, all SVR reconstructions performed comparably by demonstrating a gradual decline on the logarithmic scale. For both SAGA and SVRG, the $\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_{\text{OSEM}})$ preconditioned reconstructions did not converge to the solution but instead the NRMSE% values converged to an error of



(a) SAGA



(b) SVRG

Figure 4.7: The lung lesion percentage NRMSE values of 15 stochastic realisations with preconditioners anchored throughout. (a) SAGA and (b) SVRG reconstructions are plotted separately. The BSREM ROI NRMSE percentage value was computed for a single deterministic realisation with Equation (4.4) and vertical axis are identical and logarithmic.

approximately 1. In comparison, the $\boldsymbol{D}_{\mathrm{EM}}(\boldsymbol{x}_{5\mathrm{M}})$, $\boldsymbol{D}_{\mathrm{EM}}(\boldsymbol{x}_{10\mathrm{M}})$ and $\boldsymbol{D}_{\mathrm{EM}}(\boldsymbol{x}_k)$ preconditioned reconstructions all appeared comparable for the entire 20 epoch reconstruction performance shown. The configurations using SVRG, plotted in Figure 4.7b, appeared to converge closer to the solution than their comparable SAGA configurations.

To investigate why $\boldsymbol{D}_{\mathrm{EM}}(\boldsymbol{x}_{\mathrm{OSEM}})$ with $\eta = 0$ reconstructions failed to converge, ROI values of a region outside of the thorax were computed, which are plotted in Figure 4.8.

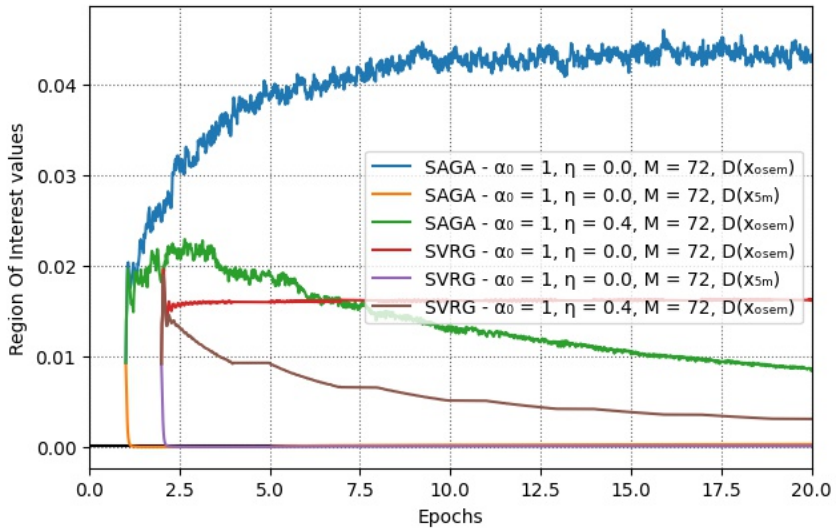


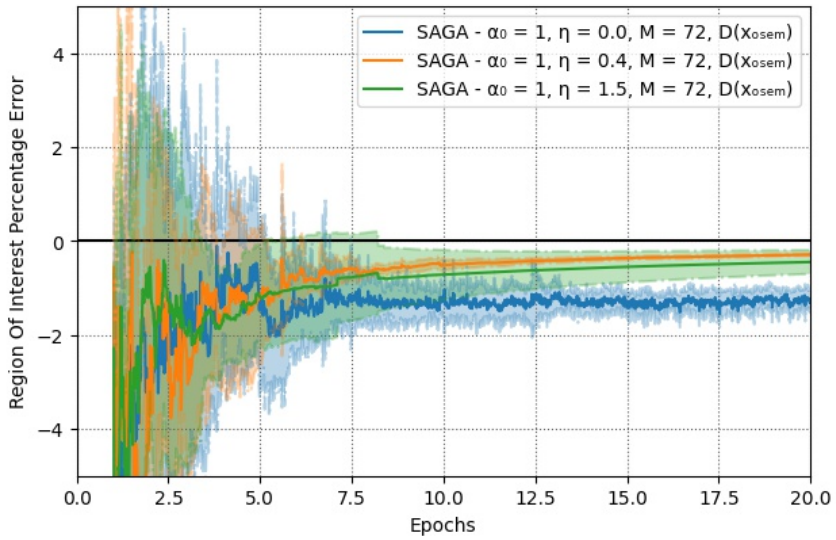
Figure 4.8: Mean image background region values of 50 million event XCAT data reconstructions using SAGA and SVRG across 15 stochastic realisations. Algorithm configurations utilise combinations of the $\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_{\text{OSEM}})$ and $\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_{5\text{M}})$ preconditioners and relaxation parameters $\eta = 0.0$ and $\eta = 0.4$.

The voxels in this region did not converge to zero when $\alpha_k = 1$ and the preconditioner $\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_{\text{OSEM}})$ was implemented. Instead, the voxels in this region retained a positive bias. It is assumed this bias exists for all voxels outside of the object, not only those in this ROI. Allowing the preconditioner to vary as $\boldsymbol{x}_k$ until 5 or 10 epochs permitted the background region to converge to zero. In addition, preconditioning with $\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_{\text{OSEM}})$ and relaxing α_k with $\eta = 0.4 > 0$ reduced the impact of the bias in this region over the course of the reconstruction. The impact of various α_k methodologies is further explored in Figure 4.9.

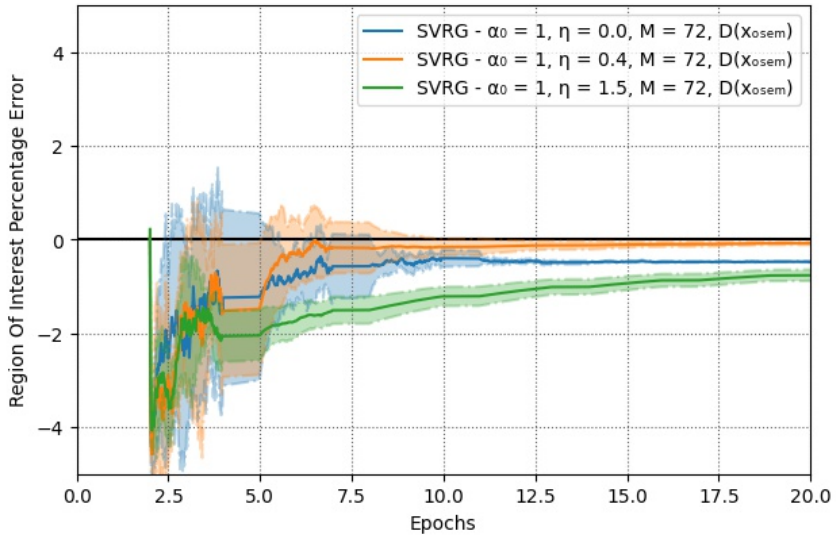
Preconditioners may be considered spatially variant step sizes and the SVR algorithms require $\alpha_k < \alpha_{\mathcal{L}_D}$ to converge. The preconditioned objective function $\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_{\text{OSEM}})\Phi(\boldsymbol{x})$ likely resulted in the non-convergence observed in Figure 4.7 because of the sub-optimal preconditioner. This preconditioned system may not have maximised $\alpha_{\mathcal{L}_D}$, such that $\alpha_{\mathcal{L}_D} > \alpha_k = 1$, as well as $\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_{5\text{M}})\Phi(\boldsymbol{x})$, $\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_{10\text{M}})\Phi(\boldsymbol{x})$ or $\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_k)\Phi(\boldsymbol{x})$.

4.5.5 Step Sizes

Figure 4.9 demonstrates SAGA and SVRG reconstruction performance when implemented with step size relaxation. The poorly performing $\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_{\text{OSEM}})$ preconditioner was used to highlight the impact of relaxation. Similar to the $\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_{\text{OSEM}})$ reconstructions in Figure 4.8, the reconstructions with no relaxation ($\eta = 0$) were non-convergent to $\hat{\boldsymbol{x}}$ within the 20 epochs shown. Yet, reconstructions with $\eta > 0$ appeared to allow the lung lesion ROI



(a) SAGA



(b) SVRG

Figure 4.9: The XCAT 50 million event data set mean lung lesion percentage error of (a) SAGA and (b) SVRG reconstructions using various step size relaxation factors $\eta \in \{0,\ 0.4,\ 1.5\}$, 72 subsets and the $\boldsymbol{D}_{\mathrm{EM}}(\boldsymbol{x}_{\mathrm{OSEM}})$ preconditioner. 15 stochastic realisations were used to compute the mean and the shaded region indicates the stochastic realisation standard deviation at each update.

values to converge to the correct value, although the rate of convergence was impeded when $\eta = 1.5$. The overall trends of step size relaxation are similar to those observed for BSREM in Figure 4.6a although the SVR algorithms appear less susceptible to subset selection as the stochastic realisation standard deviations are small.

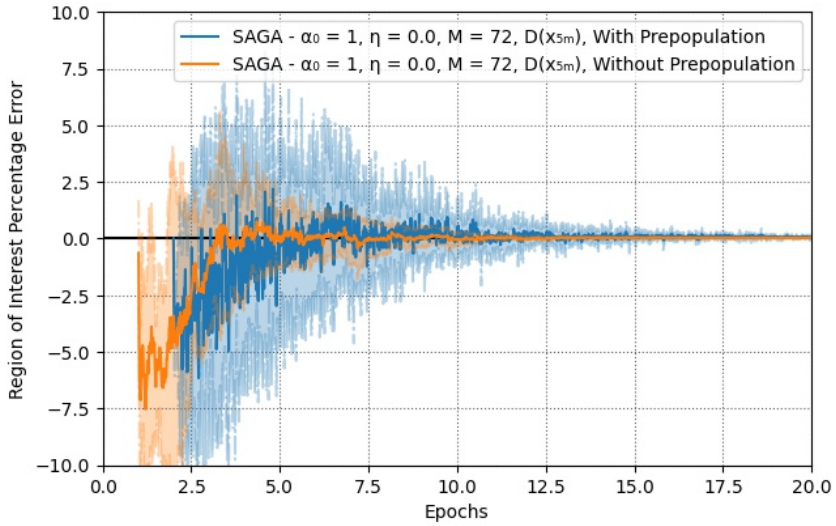


Figure 4.10: The lung lesion percentage error of the 50 million event XCAT data set reconstructions using SAGA with and without prepopulating the gradients in memory. This prepopulating corresponds to an initial epoch of computing $\boldsymbol{g}_m = \nabla\Phi_m(\boldsymbol{x}_{\text{init}}),\ \forall m$ before updates are performed. Without warm starting initialises $\boldsymbol{g}_m = \boldsymbol{0},\ \forall m$. Algorithm configurations use the $\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_{5M})$ preconditioner with $M = 72$ subsets and $\eta = 0.0$.

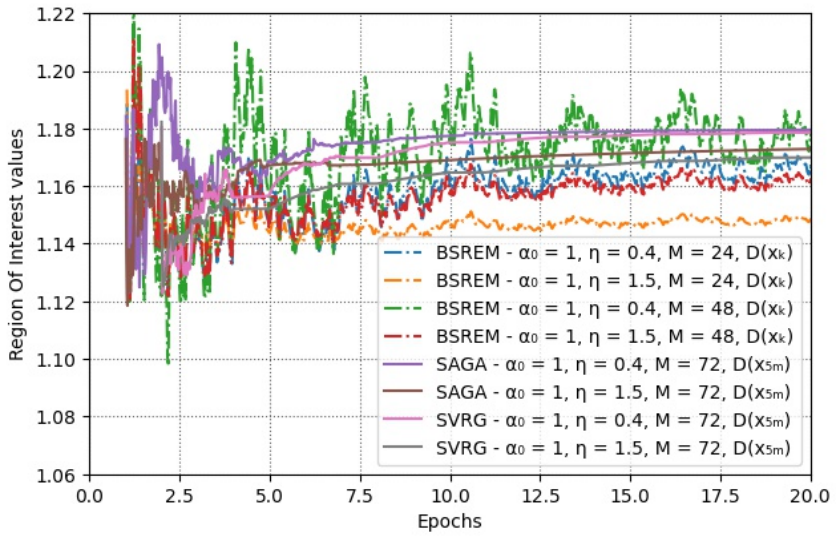
4.5.6 Prepopulating SAGA Gradients

In Figures 4.6b and 4.9a, initial epochs of SAGA reconstructions exhibit large variations between stochastic realisations. A suggestion by Schmidt et al. 2017 indicates that prepopulating the M gradients in memory with $\boldsymbol{g}_m = \nabla\Phi_m(\boldsymbol{x}_{\text{OSEM}}),\ \forall m$ may improve early algorithm performance. This methodology is comparable to SVRG during its first update and reconstruction performance using this methodology is demonstrated in Figure 4.10. These results are compared to similar SAGA configurations without prepopulating and it is observed that reconstruction performance is in fact negatively impacted by prepopulating. The expected mean ROI values fluctuated significantly for a longer reconstruction period and the 15 stochastic realisation standard deviations were significantly larger. The reasoning for this is discussed in Section 4.6.5.

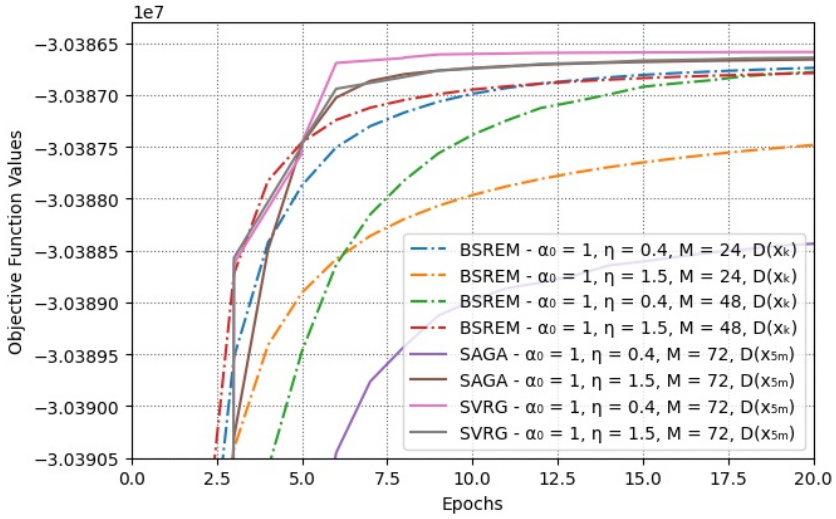
4.5.7 Impact of Noise

The impact data noise has on the stochastic algorithm reconstruction performance is investigated in this sub-section. The implemented stochastic algorithm configurations were selected to be advantageous and are based on the results presented in Sections 4.5.3, 4.5.4 and 4.5.5. Two metric measurements are demonstrated for each reconstruction configuration for each data set in Figure 4.11. On the left, sub-figures (a), (c) and (e), mean lung lesion ROI values throughout the reconstruction are shown (denoted by Ω_k^r) and on the

right, sub-figures (b), (d) and (f), objective function values $\Phi(\boldsymbol{x}_k)$ are presented. In this sub-section, references to (a) - (f) correspond to the sub-figures in Figure 4.11.

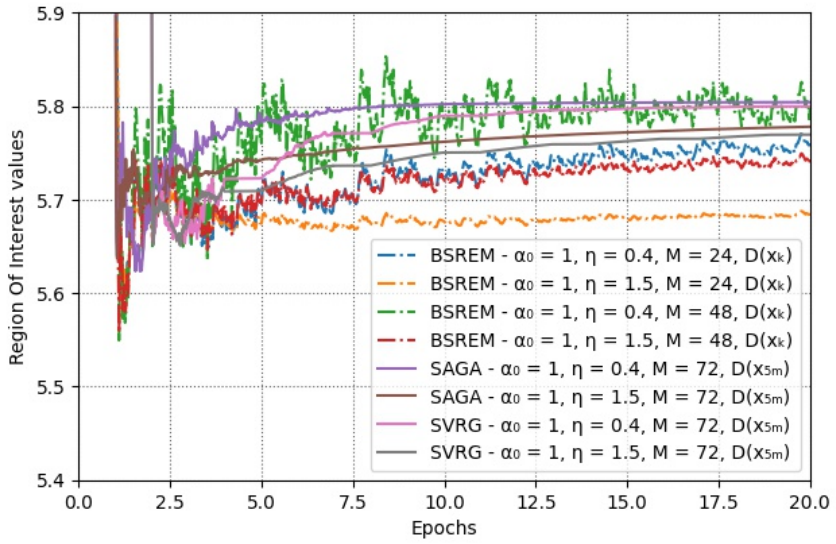


(a) 50 Million Events Ω_k^r

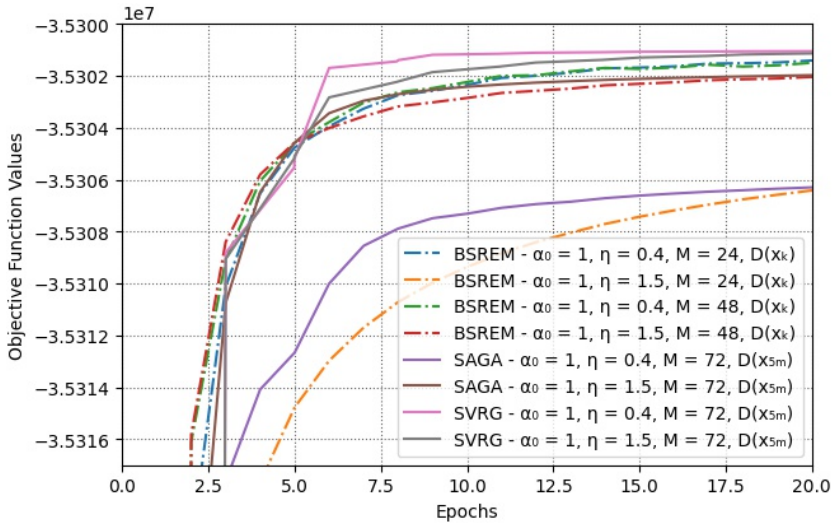


(b) 50 Million Events $\Phi(\boldsymbol{x}_k)$

Figure 4.11: (*Page 1/3*)



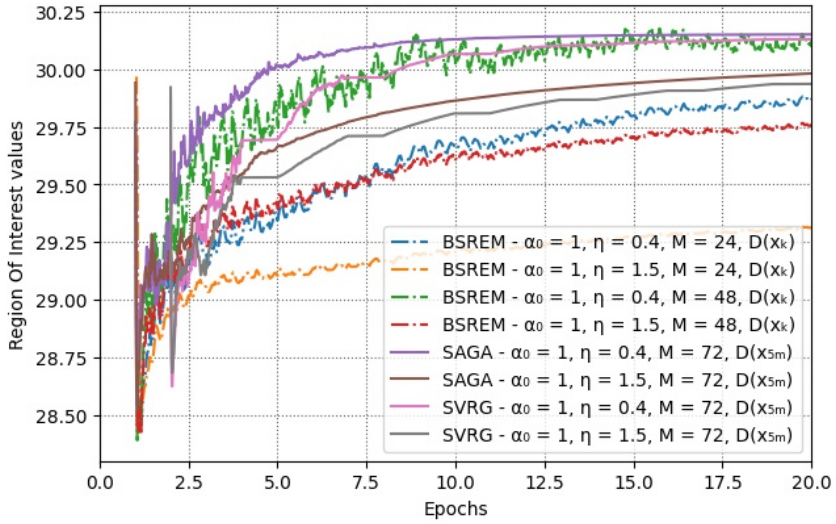
(c) 250 Million Events Ω_k^r



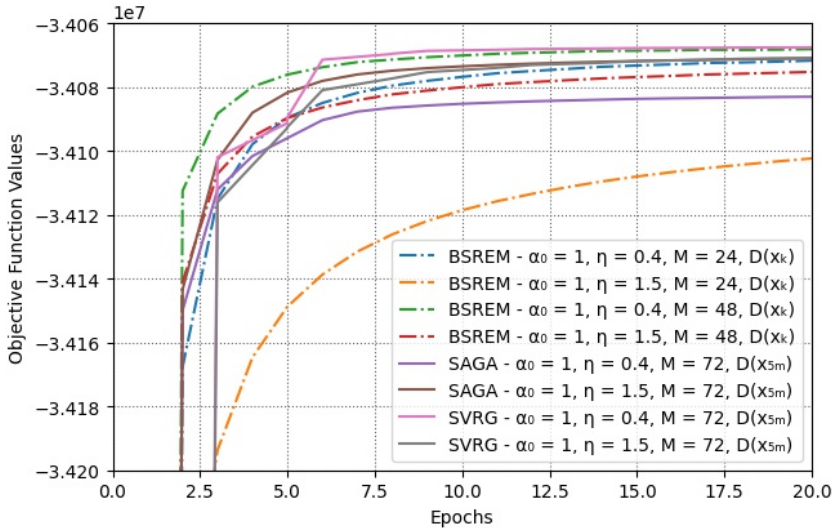
(d) 250 Million Events $\Phi(\boldsymbol{x}_k)$

Figure 4.11: (*Page 2/3*)

For the 50 million event data set, the SVR algorithms demonstrated large fluctuations in Ω_k^r during the first five epochs of (a). These reduced significantly over the remaining epochs and appeared to converge to a value of $\hat{\Omega}^r \approx 1.18$. In comparison, the BSREM reconstructions either exhibited large inter-update variations in Ω_k^r or were slow to converge. This slow convergence was likely a result of the use of fewer subsets ($M = 24$) and a larger $\eta = 1.5$ value. In the adjacent figure (b), the epoch periodic measurements of $\Phi(\boldsymbol{x}_k)$ indicated a larger discrepancy between algorithm performances. However, the



(e) 1200 Million Events Ω_k^r



(f) 1200 Million Events $\Phi(\boldsymbol{x}_k)$

Figure 4.11: (*Page 3/3*) Lung insert mean values across 15 stochastic realisations for XCAT data sets with (a) 50, (c) 250 and (e) 1200 million prompt events. Mean objective function evaluations of 15 stochastic realisations, measured at every epoch, of BSREM, SAGA and SVRG reconstructions for XCAT data sets with (b) 50, (d) 250 and (f) 1200 million prompt events. Standard deviation markers were removed for clarity. SAGA and SVRG reconstructions utilised a $\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_{5\text{M}})$ as a preconditioner and 72 subsets while BSREM utilised the $\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_k)$ preconditioner. Various step size relaxation parameter η values are demonstrated.

other SVR algorithm configurations outperformed the BSREM reconstructions after the 5$^{\text{th}}$ epoch. Note, the fastest converging Ω_k^r SAGA reconstruction in (a), using $\eta = 0.4$, appeared non-competitive in (b).

The 250 million event reconstruction Ω_k^r measurements shown in (c) demonstrated similar performance to the 50 million event data set. The SVR algorithms exhibited some initial variation in mean Ω_k^r measurements before seemingly converging to $\hat{\Omega}^r \approx 5.8$. As with (a), the BSREM reconstructions in (c) suffered from either inter-update variations in Ω_k^r values or slow convergence. The measurements shown in (d) depict low $\Phi(\boldsymbol{x}_k)$ values for the SAGA reconstruction with $\eta = 0.4$ and the $M = 24$ with $\eta = 1.5$ BSREM reconstructions. All other BSREM, SAGA and SVRG reconstructions appeared to similarly optimise $\Phi(\boldsymbol{x}_k)$ over the 20 epochs with the SVRG algorithms slightly outperforming other methods after 5 epochs.

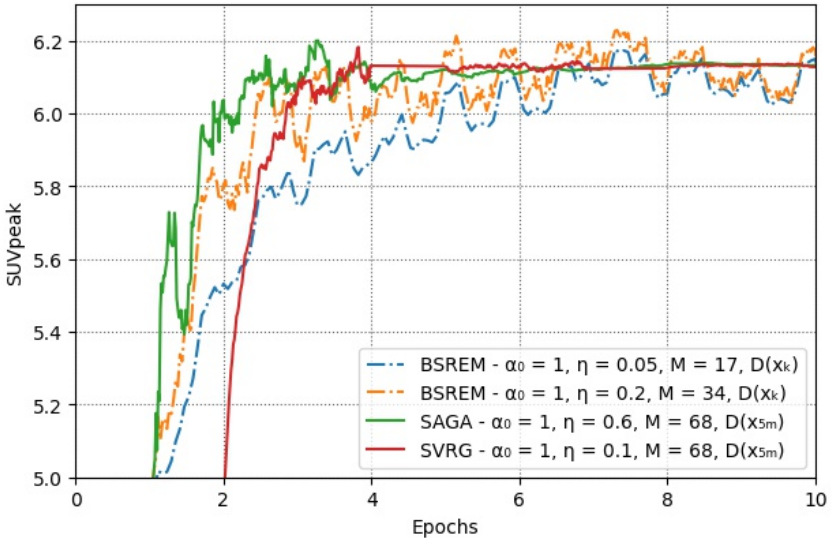
Finally, the 1200M reconstructions indicated greater discrepancies in Ω_k^r performance during early updates in (e), when compared to (a) and (c). Notably, all reconstructions using $\eta = 1.5$ performed poorer than their $\eta = 0.4$ comparisons and, with the exception of $M = 48$ and $\eta = 0.4$ BSREM reconstruction, the SVR algorithms outperformed BSREM. Compare this to the $\Phi(\boldsymbol{x}_k)$ measurements plotted in (f) where all reconstructions, besides BSREM with $M = 24$ with $\eta = 1.5$, were comparable. In both (e) and (f), the BSREM reconstruction using $\eta = 0.4$ and $M = 48$ was fast at optimising the metrics during early epochs and competitive with the SVR algorithms at later epochs, although there were some inter update variations in Ω_k^r when approaching 20 epochs.

4.5.8 Non-Simulated Data Sets

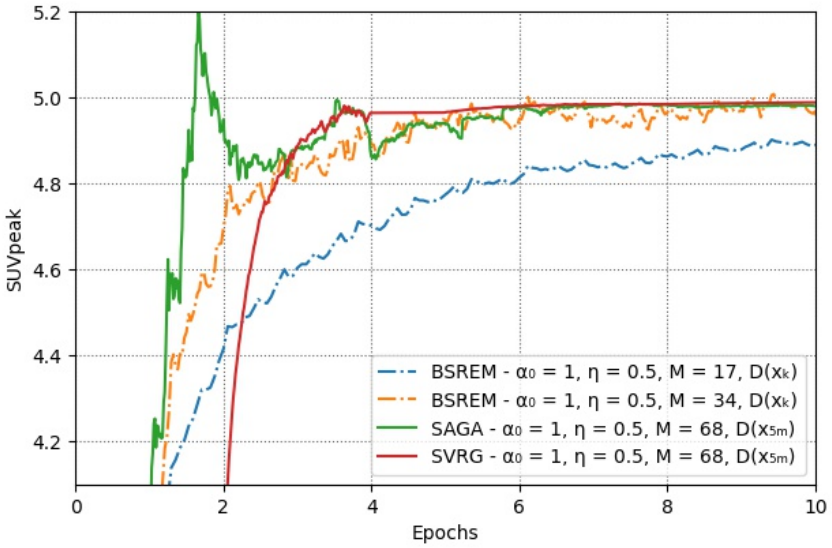
The previous sections assessed various configurations of SAGA and SVRG using the XCAT GATE simulated data set. In this sub-section, the SVR algorithms are applied to non-simulated 3D PET data sets and algorithm performance is compared to BSREM using ROI metrics. SVR algorithms are configured similarly to the algorithms presented in Section 4.5.7, with various step size relaxation parameters η, the $\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_{5\text{M}})$ preconditioner and $M = 68$ or $M = 72$ subsets. However, only a single stochastic realisation is presented for each algorithm configuration.

4.5.8.1 Patient Data

Multiple reconstructions of the patient data set (described in Section 4.2.3) were performed with various algorithm configurations. The presented results in Figure 4.12 correspond to the "optimal" algorithm η value configurations for BSREM (17 and 34 subsets), SAGA and SVRG (68 subsets). Multiple reconstructions were performed with various η values for each algorithm. The presented configuration minimised the ROI percentage error of the mean bronchial carcinoma ROI (with respect to $\hat{\Omega}^r$) over the first 10 epochs. Therefore, these "optimal" algorithms converged fast while minimising inter-update variations.

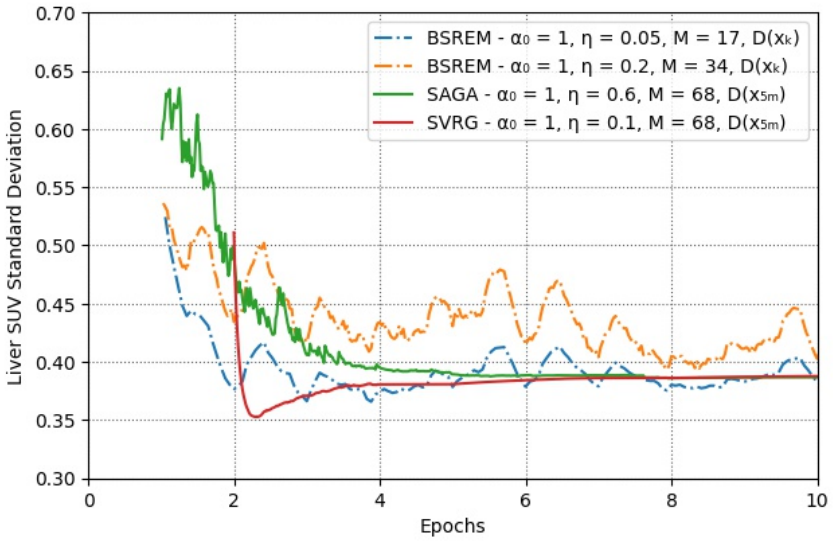


(a) Bronchial Carcinoma



(b) Hilar Lymph Node Metastasis

Figure 4.12: (*Page 1/2*)



(c) Liver Standard Deviation

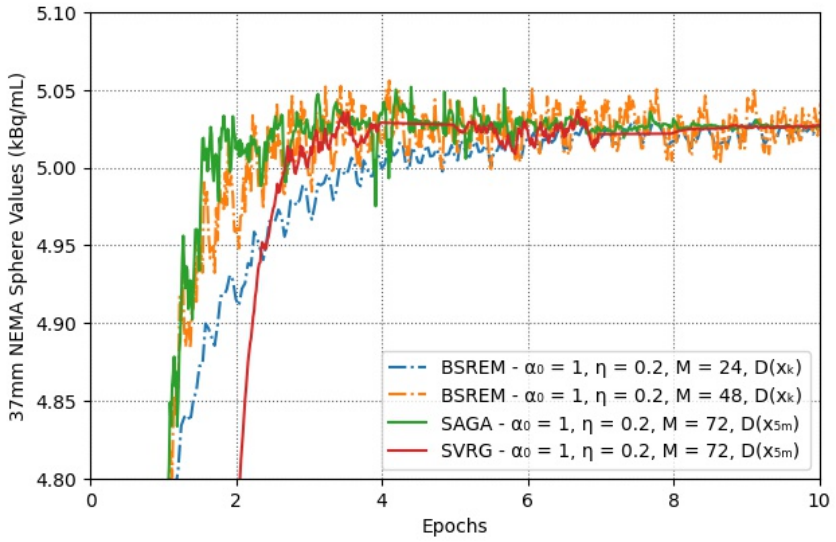
Figure 4.12: (*Page 2/2*) Reconstruction performance of the patient data using BSREM (17 and 34 subsets) and SAGA and SVRG (68 subsets). (a) Bronchial Carcinoma (SUVpeak) and (b) Hilar Lymph Node Metastasis (SUVpeak) and (c) ROI standard deviation in the patient's liver (SUV).

The mean bronchial carcinoma values, shown in Figure 4.12a, converged quickly to an SUV peak value of 6.138 for both SAGA and SVRG with minimal deviations after 3 epochs. The liver mean ROI value converged to 3.68 SUV using SAGA and SVRG within a similar number of epochs (plots not shown). Additionally, the standard deviation of voxel values within the liver ROI converged to a value of 0.39 SUV, see Figure 4.12c.

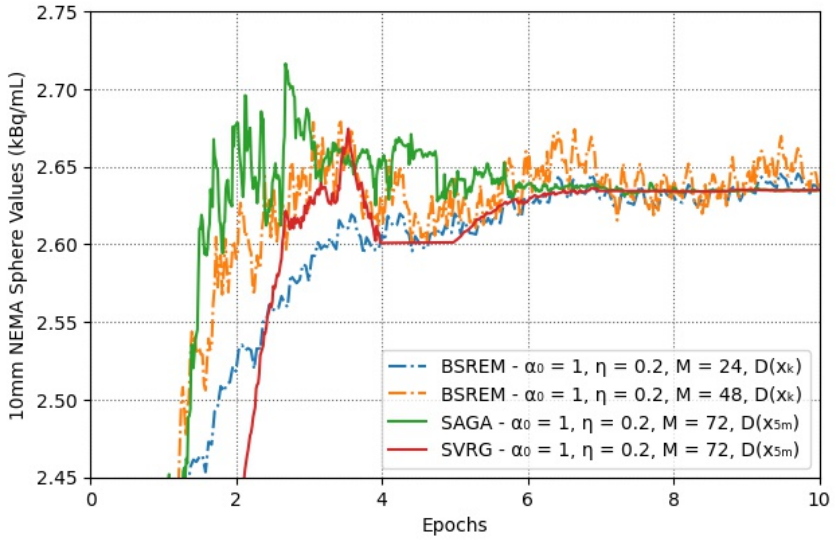
4.5.8.2 NEMA

The NEMA phantom data set (described in Section 4.2.4) was reconstructed using SAGA and SVRG with $\eta = 0.2$, $M = 72$ and the $\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_{5M})$ preconditioner. The mean ROI values of the largest and smallest spheres (37mm and 10mm diameters) over the reconstruction process are plotted in Figures 4.13a and 4.13b, respectively. Although the phantom's spheres were filled with equal activity, the smaller sphere was reconstructed with less intensity. This was likely a consequence of spill-over partial volume effects [Erlandsson et al. 2012], i.e. over-smoothing due to the penalty.

It is evident that the 37mm sphere converges faster than the 10mm sphere for all algorithms. In Figure 4.13a, the single SAGA stochastic realisation quickly achieved the approximately correct ROI scale within 2 epochs. However, a few additional epochs were required for the inter-iteration variations to dissipate. The SVRG reconstruction ROI values



(a) 37mm NEMA sphere



(b) 10mm NEMA sphere

Figure 4.13: Reconstruction profiles of the mean values (kBq/mL) of (a) the 37mm sphere and (b) the 10mm sphere in a NEMA phantom. NEMA phantom details can be found in Section 4.2.4. Algorithms used are BSREM, with 24 and 48 subsets, and SAGA and SVRG, with 72 subsets.

were delayed in achieving converged mean value by the initial full epoch of computation. However, the amplitude of the variations between sequential updates appeared lesser for SVRG than for SAGA at similar epochs. This larger sphere converged quickly for both BSREM reconstructions and, after approximately 6 epochs, BSREM with 24 subsets appeared comparable to the stochastic algorithms.

SAGA's and SVRG's initial performance, shown in Figure 4.13b, was similar to the

results of the previous figure, but, the inter-update variations appeared larger. Within epochs 2 to 4, SVRG over and then under estimated the mean ROI values. After the gradients were recomputed (4^{th} to 5^{th} epoch), the ROI steadily converged with limited inter-iteration variations.

4.6 Discussion of Stochastic Variance Reduction Algorithms Applied to PET

In Section 4.5, the results of the SAGA and SVRG algorithms, used to reconstruct various 3D PET data sets in a variety of configurations, were presented. Throughout this section, these results are discussed.

4.6.1 Number of Subsets Impact

In Figure 4.6b and Figure 4.6c, SAGA's and SVRG's use of a greater number of subsets resulted in faster mean convergence of the ROI. However, large stochastic variations in ROI metric performance, between stochastic realisations, were observed during the initial epochs. These large deviations were a result of stochastic subset selection but reduced after a few epochs. Furthermore, the maximum standard deviation values for the $M = 72$ and $M = 96$ reconstructions were comparable to fluctuations in the comparison BSREM ROI values, shown in Figure 4.6a.

In Section 4.3.2 and the reconstruction results depicted in Figure 4.5, it was observed that the variance reduction algorithms are sensitive to subset selection, particularly during early updates. This sensitivity was demonstrated by the initial epochs of OS SAGA and OS SVRG outperforming the stochastic subset selection methods. From Figure 4.4 it can be predicted that OS methodologies are effective during early updates. Increasing the number of subsets results in less balanced subsets. Hence, because of SAGA's and SVRG's initial subset selection sensitivity and a greater number of subsets being employed, the algorithms demonstrated larger variations in initial algorithm performance between stochastic realisations.

The $M = 72$ SAGA and SVRG reconstructions in Figure 4.6 demonstrated a balance between fast initial mean performance and smaller standard deviations. The standard deviations between stochastic realisations converged to less than 0.1% within 10 epochs. This is likely explained by all $\boldsymbol{g}_m \approx \nabla\Phi_m(\boldsymbol{x}_k)$ for all stochastic realisations after approximately 10 epochs, regardless of the number of subsets. Hence, the approximation that

$\tilde{\nabla}_{k,m_k} \approx \sum_{\mu=1}^{M} \boldsymbol{g}_\mu \approx \nabla\Phi(\boldsymbol{x}_k)$ becomes true and Equation (2.51) becomes (approximately) invariant to m_k.

Therefore, the primary impact the number of subsets has on the SAGA and SVRG algorithms occurs during the first few epochs. Note, even though BSREM utilised a step size relaxation sequence, while SAGA and SVRG used $\alpha_k = 1$, BSREM does not suppress the inter-update variations at later epochs as much as SAGA and SVRG do.

4.6.2 Differences Between SAGA and SVRG

When compared with SAGA, SVRG demonstrated reduced variations between stochastic realisations during early updates in Figure 4.6 and 4.7. This stability is likely a consequence of $\sum_{\mu=1}^{M} \boldsymbol{g}_\mu$ remaining constant for the $\gamma = 2$ epochs of stochastic subset selection. Alternatively, consider the fact that although it is expected that a subset index $m \in \{1, 2, \ldots, M\}$ is selected once every epoch, the expected number of epochs required to select every m once grows as $\mathcal{O}(\log M)$ [Flajolet et al. 1992]. Hence, a number of SAGA's stored subset gradients may not have been computed from recent $\boldsymbol{x}_k$ iterates. Thus, for various m indices, $\boldsymbol{g}_m \not\approx \nabla\Phi_m(\boldsymbol{x}_k)$, which may detrimentally impact the update direction. This may further increase the standard deviation of metric performance between stochastic realisations. In contrast, because SVRG recomputes all $\boldsymbol{g}_m$ terms periodically from $\tilde{\boldsymbol{x}}$, it is expected that $\boldsymbol{g}_m \approx \nabla\Phi_m(\boldsymbol{x}_k)\ \forall m$, as long as γ is small enough and $\boldsymbol{x}_k \approx \tilde{\boldsymbol{x}}$. However, SVRG also exhibited slower convergence, likely because of this periodic recomputation that resulted in significant additional computational cost while not updating $\boldsymbol{x}_k$.

4.6.3 Convergence

In order to assist with convergence to the PML solution, in this work the SVR algorithms were implemented with a modified version of the EM preconditioner. This preconditioner was anchored after various epochs of computation. When using a constant step size, a systematic error was observed in both the lung lesion (Figures 4.7 and 4.9) and the region outside of the thorax (Figure 4.8) at later epochs of reconstructions utilising the $\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_{\text{OSEM}})$ preconditioner with $\eta = 0$. It is hypothesised that this bias relates to an excessive effective step size in this surrounding background region due to the preconditioner, $\alpha_k = 1$ and the non-negativity constraint. This positive background bias appears to have led to a (small) negative bias inside the torso due to data count preservation. By deferring the anchoring of the preconditioner until later epochs, e.g., $\boldsymbol{D}(\boldsymbol{x}_{5e})$, the effective step size in this region was significantly reduced because voxel values tended towards zero. This allowed the recon-

structions to converge closer to the solution in the investigated ROIs.

Additionally, several step size relaxation parameters η were investigated in Figure 4.9. Both the XCAT lung lesion insert and the region surrounding the torso (Figure 4.8) converged closer to the solution with $\eta > 0$ than $\eta = 0$. However, similar to BSREM, over relaxation may have impeded reconstruction performance and reduced the convergence rate. Thus, heuristic tuning of η is still required for this methodology, although it seems apparent that the SVR algorithms are less sensitive to the parameters selection.

The experimental results indicate that the implementation of a moderately relaxed step size sequence, along with a delayed anchored preconditioner, e.g., $\boldsymbol{D}(\boldsymbol{x}_{5e})$, allows for fast numerical convergence.

4.6.4 Impact of Data Noise

Multiple data sets were used in this work to evaluate SVR algorithms. Reconstruction performance of various algorithm configurations at multiple noise/count level reconstructions of the XCAT GATE simulated data, discussed in Section 4.2.2, were presented in Figure 4.11. In these plots, the SAGA and SVRG reconstructions generally achieved higher objective function values faster than BSREM with the 50 million event data set, see Figure 4.11b. Yet, as the number of events increased, BSREM reconstructions became comparable in optimising the objective function. Moreover, as the data set SNR increased, all reconstruction algorithms appeared less sensitive to relaxation parameter η. In comparison, the ROI plots in Figure 4.11 demonstrated that the SVR algorithms reduce the impact of subset selection, which is notable for the 50 million event data set. Additionally, although this impact is reduced for higher count data, likely due to subset balance improving with higher SNR data sets because of relatively reduced noise contributions, the SVR algorithm lung lesion ROI converged faster with the use of more subsets.

4.6.5 SAGA Prepopulating

As observed in Section 4.5.6 and Figure 4.10, prepopulating SAGA's $\boldsymbol{g}_m$ vectors with $\boldsymbol{g}_m = \nabla\Phi_m(\boldsymbol{x}_{\text{OSEM}}),\ \forall m$ did not improve the convergence rate of the XCAT lung insert. This contradicts the observations in Schmidt et al. 2017. To attempt to explain this, consider the SAGA reconstruction without prepopulation. This algorithm's initial $\tilde{\nabla}_{k,m_k}$ evaluations will have smaller magnitude than the prepopulated $\boldsymbol{g}_m$ SAGA algorithm as few non-zero $\boldsymbol{g}_m$ variables exist. This will reduce the effective step size of the update direction during these early updates. Thus the algorithm gradually increases the step size as more $\boldsymbol{g}_m$ variables

become populated as k increases.

Furthermore, during these early updates, with minimal information being provided by the $\boldsymbol{g}_m$ variables, the algorithm likely converges closer to $\hat{\boldsymbol{x}}$ as the SAGA algorithm remains akin to a stochastic preconditioned gradient ascent algorithm. Hence, during this early period, many of the newly assigned $\boldsymbol{g}_m = \nabla\Phi_m(\boldsymbol{x}_k)$ evaluations better represent $\nabla\Phi_m(\hat{\boldsymbol{x}})$, compared to $\nabla\Phi_m(\boldsymbol{x}_{\text{OSEM}})$.

Next, consider the fact that the SAGA algorithm with prepopulation and the SVRG algorithm are indistinguishable for the first update. Yet, the SVRG reconstructions in similar configurations, e.g., $M = 72$ in Figure 4.6c did not suffer the same large stochastic realisation standard deviations between epochs 2 and 4. As the discrepancies between these algorithms occurred over multiple stochastic realisations for both data sets, it may be concluded that the cause relates to the differences in the algorithm methodology. The primary discrepancy is the storage of $\boldsymbol{g}_m = \nabla\Phi_m(\boldsymbol{x}_k)$ after each iteration of SAGA, whereas SVRG retains constant $\boldsymbol{g}_m$, $\forall m$ variables for the γM updates. At the second update a random subset index m_k is selected and the $\boldsymbol{g}_{m_k} = \nabla\Phi_{m_k}(\boldsymbol{x}_k)$ variable will be stored. Subsequent updates will continue this trend and additional $\boldsymbol{g}_m$ variables will be reassigned and will not contain the initial prepopulated $\nabla\Phi_m(\boldsymbol{x}_{\text{OSEM}})$ values. Thus, the stochastic variability will be increased for prepopulated SAGA compared to SVRG.

4.6.6 Image Noise Convergence

Reconstruction performance of a clinical patient data set was evaluated in Section 4.5.8.1. Noise and bias trade-offs are commonly evaluated in PML reconstruction analysis [Ahn et al. 2015; Teoh et al. 2015]. However, this evaluation is sensitive to the reconstruction algorithm, epoch number and objective/penalty function, which was somewhat demonstrated in Figure 4.12a. The SUVpeak of the bronchial carcinoma converged quickly with BSREM in Figure 4.12a, although some inter-update variations were present. Yet, the BSREM liver standard deviations did not converge within 10 epochs in Figure 4.12c. Both of the BSREM reconstructions exhibited standard deviation values that were generally greater than the value obtained after 10 epochs by the SVR algorithms, with significant variations between epochs.

The SUVpeak for the bronchial carcinoma converged quickly for SAGA and SVRG reconstructions, similar to BSREM. However, the SVR algorithms were also able to converge quickly in the liver. Therefore, these SVR algorithms removed a source of potential error in

noise and bias comparisons due to the impact of the number of subsets and reconstruction duration (after 4 epochs). Yet, the reliance on objective function remains.

4.6.7 Practical Implementation for PET Reconstruction

This sub-section discusses some practical aspects of implementing SVR algorithms to PET reconstruction.

4.6.7.1 Computation Cost Comparison

Throughout this thesis, it has been assumed that an epoch of each of the SVR algorithms is equivalent to a BSREM epoch. However, the SAGA and SVRG algorithms incur additional computational cost in the evaluation of $\tilde{\nabla}_{k,m}(\boldsymbol{x}_k)$ due to the additional image/vector manipulations associated with Equation (2.51). In an efficient software implementation, the impact of these basic image-space operations should be negligible with respect to the computational cost of evaluating $\nabla\Phi_m(\boldsymbol{x})$, which requires subset forward and backward projection operations. This is because the number of voxels in an image N_v is considered small for modern computing systems, especially when compared to the number of projection data bins N_b in 3D PET data sets, e.g., for the non-TOF GE Discovery MI (5 ring): $N_v = 9.29 \times 10^6$ and $N_b = 2.23 \times 10^8$. For TOF reconstruction, N_b may increase by an order of magnitude. Therefore, the computational effort for evaluating an epoch of SAGA or SVRG is expected to be little more than the computational effort to evaluate an epoch of BSREM.

4.6.7.2 Reconstruction Duration and Static Objective Function

The SVR algorithms demonstrated potential to PET reconstruction in this work. However, because of the population of the M $\boldsymbol{g}_m$ variables, these algorithms require a minimum of an epochs worth of computation to be as effective as designed. This is clearly evident for SVRG with its initial population of the $\boldsymbol{g}_m$ variables before the first algorithm update is applied. Following this, the algorithm iterates for γ epochs before it effectively restarts from a new $\tilde{\boldsymbol{x}}$. Therefore, an SVRG reconstruction duration should iterate for longer than a single epoch and an integer number of $\gamma+1$ epochs is recommended.

Several advanced image reconstruction methods alternate between "model estimation" and an "image reconstruction", e.g., Motion-Compensated Image Reconstruction (MCIR) [Dikaios et al. 2011] and scatter estimation [C. Tsoumpas et al. 2004; Thielemans et al. 2012]. As the model estimation step modifies the objective function, algorithms that keep history of previous updates may retain information that does not reflect the newly modified objective function. The SVR algorithms rely on the approximation that $\nabla\Phi_m(\boldsymbol{x}_k) \approx \boldsymbol{g}_m$

for the evaluation of $\tilde{\nabla}_{k,m_k}$ in Equation (2.51). Hence, the SVR algorithms should only be considered for optimising static objective functions. For use with these advanced image reconstruction methods, these algorithms should be applied with care because significant changes in $\Phi(\boldsymbol{x})$ may result in poor update steps. Alternatively, SVRG could recompute the M $\boldsymbol{g}_m$ variables after every change in $\Phi(\boldsymbol{x})$ but this may be inexpedient.

4.6.7.3 Memory Requirements

The memory requirement for storing M subset gradient variables $\boldsymbol{g}_m$ is insignificant on a modern computing system for current PET images. For example, assuming each $\boldsymbol{g}_m$ requires approximately 20MB in memory (as is the case for the GE Discovery 690), with $M = 72$, the memory requirements are 1440MB. However, with more complex objective functions, e.g., those optimising dynamic data sets, this memory requirement may become restrictive with a large M. Additionally, with the advent of longer PET scanners increasing the FOV, these memory requirements may become more relevant [Spencer et al. 2021].

4.6.8 Alternative Stochastic Algorithms

There exist a number of comparable algorithms to those presented in this work. This includes a recently proposed stochastic algorithm for PET reconstruction (SPDHG) and alternative stochastic variance reduction algorithms.

4.6.8.1 SAG

Based on poor performance during preliminary studies (Twyman et al. 2020, Twyman et al. 2021a and Section 4.3.2), SAG reconstruction performance results were not presented using 3D data sets in Section 4.5, however, they were computed. Some of these SAG algorithm configurations (using the stochastic subsets method) diverged in a similar manner to the OS SAG reconstruction in shown Figure 4.5a. As aforementioned in Section 2.3.5.2 and Appendix C, SAG is a biased algorithm. The SAG $\tilde{\nabla}_{k,m_k}$ computation relies more on the $\sum_{\mu=1}^{M} \boldsymbol{g}_\mu$ term of Equation (2.51) than SAGA because $\xi = 1$. This likely makes the SAG algorithm more sensitive to instances where $\boldsymbol{g}_m \not\approx \nabla\Phi_m(\boldsymbol{x}_k)$ for any m. This may have induced the poor performance observed with some SAG reconstructions.

In Driggs et al. 2020b, various $\xi \in (1, M)$ values were investigated for both SAGA and SVRG in Equation (2.51). This study indicated some algorithm bias can accelerate the optimisation of the objective function.

4.6.8.2 Alternative SVRG

An alternative SVRG implementation that benefits from reduced memory requirements can also be employed. Only the full gradient $\boldsymbol{G} = \sum_{\mu=1}^{M} \boldsymbol{g}_\mu$ and anchored position $\tilde{\boldsymbol{x}}$ are stored in memory. At each update, $\nabla\Phi_m(\boldsymbol{x}_k)$ and $\nabla\Phi_m(\tilde{\boldsymbol{x}})$ are computed. We note that this implementation should generally be unnecessary for the PET reconstruction problem due to the relatively small memory requirements for PET distribution volumes, as discussed in Section 4.6.7.3, and the additional, but significant, computational cost of evaluating $\nabla\Phi_m(\tilde{\boldsymbol{x}})$ at every update.

4.6.8.3 SARAH

The StochAstic Recursive grAdient algoritHm (SARAH) algorithm [Nguyen et al. 2017] is closely related to the previously described alternative SVRG algorithm in Section 4.6.8.2. SARAH also periodically recomputes all $\boldsymbol{g}_m$ terms from $\tilde{\boldsymbol{x}}$ but only stores $\tilde{\boldsymbol{x}}$ and $\boldsymbol{G} = \sum_{m=1}^{M} \boldsymbol{g}_m$ in memory. The algorithm evaluates two subset gradients at each update in its computation of $\tilde{\nabla}_k = \nabla\Phi_{m_k}(\boldsymbol{x}_k) - \nabla\Phi_{m_k}(\boldsymbol{x}_{k-1}) + \boldsymbol{G}$. This additional computation is the reason that SARAH is not included for comparison in this thesis.

4.6.8.4 SVREM

Another SVR algorithm is Stochastic Variance Reduction EM (SVREM), which was applied to penalised PET image reconstruction in a preliminary study [Kereta et al. 2021]. This algorithm was developed in parallel with the research presented in this thesis. SVREM extends OSEM into a stochastic framework and incorporates insights from variance reduction techniques for gradient based methods into the EM framework. The methodology involves closed-form update equations based on parabolic surrogates and features non-negativity preservation. The SVREM algorithm demonstrated similar performance to SAGA and SVRG.

Currently the SVREM is incompatible with the RDP (which was used throughout this thesis) because of the algorithms requirement to construct a surrogate of the penalty function. This requires a penalty energy function in the form of $\rho(x)$, where (generally) $x = x_i - x_l$, see Section 2.2.3.

4.6.8.5 SPDHG

The SPDHG algorithm has been recently developed and applied to numerous PET image reconstruction problems [Chambolle et al. 2018; Ehrhardt et al. 2017; Delplancke et al. 2021]. SPDHG is a subset/stochastic extension of the Primal Dual Hybrid Gradient (PDHG) algo-

rithm that solves a saddle point problem, which is separable in the dual variable allowing for the use of subsets [Chambolle et al. 2018]. Similar to the variance reduction optimisation algorithms, SPDHG retains information from previous iterations, which reduces variance. This information is stored in projection space that may result in computer memory complications for high dimensional data [Schramm et al. 2022]. This algorithm guarantees (almost sure) convergence to the PML solution $\hat{\boldsymbol{x}}$. When combined with preconditioning, SPDHG reconstructions can utilise a large number of subsets and it has shown promising application for 3D PET reconstruction [Ehrhardt et al. 2019]. However, step size selection is generally heuristic for SPDHG and may significantly impact algorithm performance if poorly selected.

Similar to SVREM, the reconstruction performance of SPDHG was not compared to SAGA and SVRG in this work due to difficulties deriving the convex conjugate of the RDP.

4.6.9 Limitations of this Study

There are a variety of limitations regarding the presented studies applying variance reduction algorithms to PET reconstruction in this chapter.

A primary limitation of this research is the lack of comparison between the algorithms presented and other PET algorithms, such as those discussed in Section 4.6.8, in addition to other popular optimisation algorithms in the literature, e.g., OS-SPS [Ahn et al. 2003], Nesterov and momentum methods [Nesterov 2013; Ruder 2016], and L-BFGS-B(-PC) [Tsai et al. 2018]. As aforementioned in Section 4.6.8.5, SPDHG and SVREM comparisons are not currently possible with the objective function used in this work because of the RDP. Nesterov and momentum modifications of the SAGA and SVRG algorithms have been presented in the optimisation literature [Allen-Zhu 2016; Shang et al. 2017; Driggs et al. 2020a], however, they have not been applied to PET. The benchmark comparisons between these novel stochastic algorithms remains a possible avenue of future work.

Many of the figures shown in Section 4.5 plotted the mean and standard deviations of 15 stochastic realisations of each algorithm configuration. This number of realisations was selected heuristically. A large number of stochastic realisations is required to accurately demonstrate the expected performance of each algorithm configuration and its expected deviations. This was balanced against the practical experimental feasibility of running a significant number of PET reconstructions. More reconstructions would improve the depicted accuracy of the expected performance, but may not be necessary.

The presented results illustrated the expected performance of a subset of the possible algorithm configurations in the proposed setup. This included a single preconditioner formulation, albeit with various inputs, and a single step size relaxation methodology. Alternative formulations and methodologies of configuration factors may result in alternative reconstruction performance.

Throughout this thesis, a number of metrics have been employed to quantitatively assess algorithm performance over the iterates. Particular focus was given to quantification of ROI convergence in Section 4.5, although objective function values are shown in Section 4.5.7. Additionally, the $\Delta\%$ metric was used in Chapter 3 and Section 4.3. In some instances, e.g., Figure 4.11, an evaluation using one metric may have indicated superior algorithm performance but an alternative metric did not indicate this was the case. Thus, many of the statements made were metric specific and more numerous or advanced metrics are likely required for a full analysis of algorithm performance.

Finally, although the stochastic algorithms were applied to several data sets of various noise levels, only static acquisitions were investigated using non-TOF data.

4.6.10 Conclusion

To conclude this section, SAGA and SVRG were successfully applied to PET image reconstruction in a preconditioned form. It was found that SAGA and SVRG are less sensitive to the number of subsets than the comparator, BSREM, and thus, using a larger number of subsets, the algorithms converged faster to the MAP solution. However, the amplitude of expected deviations between stochastic realisations increased during early epochs as the number of subsets increased. SAGA generally allowed the investigated metrics to converge faster but SVRG appeared more stable in metric values and in most cases steadily approached convergence rather than fluctuating around the value. Furthermore, allowing the algorithms' preconditioner to vary for a number of epochs improved the conditioning of the reconstruction problem and step size relaxation could assist in accelerating convergence (if optimal parameters were selected). Furthermore, it was observed that the prepopulation of the SAGA gradients in memory did not assist in improving algorithm performance.

The two stochastic variance reduction algorithms demonstrated notably improved convergence properties of low count noisy data sets compared to BSREM. Finally, reconstruction performance was assessed across multiple simulated and non-simulated data sets, which resulted in similar conclusions illustrating SAGA and SVRG capabilities for reliable

PET image reconstruction.

Chapter 5

Conclusion

5.1 Main Conclusions

As discussed in Chapter 2, iterative optimisation is used almost ubiquitously throughout modern clinical PET reconstruction and penalised methods are becoming more commonplace as a result of the Q.Clear algorithm's inclusion into General Electric PET/CT systems [Ross 2014]. Iterative algorithms are no longer required to terminate early to prevent noisy distribution estimates because PML algorithms are able to converge to desirable MAP solutions when regularisation is balanced. Therefore, faster PET reconstruction algorithms are sought after to reduce the computational requirements to numerically converge to the MAP solution.

The current standard of subset PET reconstruction algorithms demonstrate limit cycle behaviour as a consequence of discrepancies between sequentially applied subset data used to update the estimate. Convergence is guaranteed with the use of relaxed step size sequences but performance may be impacted by heuristic selection of parameters. Hence, with practical limitations on reconstruction durations, algorithms may not converge.

The primary motivation of the studies in this thesis were to accelerate PML PET image reconstruction and to improve algorithm stability by limiting the impact of subset variance. To achieve these goals, adaptive subset size and stochastic optimisation algorithms were applied to PET reconstruction.

In Chapter 3, the AutoSubsets algorithm was presented. The algorithm was fast at optimising the objective function during the first few epochs because of its use of small subset sizes. However, the algorithm appeared to increase subset size too quickly and convergence rates were impeded after little computation.

In tandem with the AS algorithm, list mode event-by-event subset reconstruction was

presented. It was noted that reconstruction of this data could be achieved with significantly more subsets than could be used in projection data subset algorithms. However, the reconstructed image's voxel noise increased with more subsets and the AS algorithm was able to suppress this noise by using larger subsets at later updates.

The application of stochastic variance reduction algorithms to the PET reconstruction problem was presented in Chapter 4. These algorithms utilise previously computed subset gradients to better approximate the full data gradient used to update the estimated distribution. The utilisation of this variance reduced gradient approximation allows for more subsets to be used to update the distribution and thus permits more updates within an epoch of computation.

A preliminary study in Chapter 4 investigated the impact of various subset methodologies on iterative PET image reconstruction performance for a number of reconstruction algorithms, see Section 4.3. Two of the investigated subset methods attempted to balance subsets by binning regularly space rows of the projection data into each subset. It was observed that these subset construction methods significantly improved preconditioned gradient ascent performance. However, stochastic subset sampling methods result in poorer reconstruction sequences than those using the OS sampling method. As expected, this result was amplified for reconstructions using more subsets. The variance reduction algorithms demonstrated comparable performance between all subset sampling methods, except the OS SAG configuration that diverged. Thus, the variance reduction algorithms are less sensitive to subset selection than preconditioned subset gradient ascent.

Various configurations of the SAGA and SVRG algorithms were applied to PET image reconstruction in Section 4.5 and compared to BSREM reconstructions across multiple non-TOF PET data sets. This investigation conducted multiple stochastic realisations of each algorithm configuration to assess the expected performance. It was observed that preconditioned SAGA and SVRG are expected to converge within 20 epochs for a number of configurations, outperforming BSREM across multiple metrics. Therefore, the SAGA and SVRG algorithms, in a preconditioned formulation with moderate step size relaxation, have significant potential for PET image reconstruction with minimal modification to current reconstruction methodologies. These algorithms accelerated convergence with the use of a larger number of subsets and reduced the impact of the variance between subsets.

5.2 Algorithm and Software Development

In addition to the iterative reconstruction algorithm research, a large volume of software was developed to support these studies. This included the implementation of the presented BSREM and stochastic variance reduction algorithms along with various experimental methodologies. The primary software contributions provided by the author are summarised below.

5.2.1 Iterative PET Algorithms

A python project was developed to implement and configure the various iterative reconstruction algorithms in a modular framework. The implemented modules were:

(a) The AutoSubsets algorithm.

(b) Various preconditioned subset gradient ascent algorithms e.g., BSREM and OSEM.

(c) Various preconditioner methodologies.

(d) Step size methods, including line searches (back-tracing and golden-section search).

(e) Stochastic variance algorithms (SAG, SAGA, SVRG, SARAH and SVREM).

This software was built to interface with STIR for objective function computations and data set handling.

5.2.2 Cluster Distribution

For the majority of this project, algorithm analysis was primarily performed on a distributed High Performance Computing (HPC) cluster. A number of tools were developed to automate much of the interfacing to allow for the generation, extraction and analysis of the data. This included:

(a) Tools to generate, submit and control (including error state checks) large batch cluster jobs.

(b) Automatic computation of several image quality and algorithm performance metrics at each update during HPC job run-time. This significantly reduced hard disk storage requirements as reconstruction sequences did not need to be saved.

(c) Plotting tools to extract data after multiple jobs, filter specific configuration variables and plot stochastic realisations.

5.2.3 Other Software

Various contributions were made to STIR to support this work, see Appendix B for a detailed list. Additionally, the STIR-GATE-Connection project was formalised, see Appendix A for a full summary. Finally, a number of scripts were produced to convert the data output of GE's Duetto PET Toolbox software to STIR compatible interfiles.

5.3 Suggested Future Research

This thesis has demonstrated a prototype version of the AutoSubsets algorithm and a further in-depth application of SAGA and SVRG to iterative PET reconstruction. This section explores the possible directions of future research that relate to this thesis. These main suggested topics are: improvements to the AutoSubsets algorithm, alternative research to improve SVR algorithm configurations, second order stochastic algorithms, and the application of SVR to more complex clinical PET data set applications.

5.3.1 AutoSubsets

The preliminary study of the AS algorithm presented in this work yielded moderate results. Zeng 2021 presented a number of modifications to the algorithm and an additional investigation, which is discussed in Appendix D. However, further directions of future work might exist for the algorithm and these are discussed in this sub-section.

5.3.1.1 Alternative Methodologies

As aforementioned in Section 3.5.4, the CSI metric, given by Equation (3.1), may not be the best metric to determine when subset size should increase. Future work with this algorithm could investigate alternative metrics, e.g. the use of subset update magnitudes as well as directions. Additionally, the current CSI measurement does not account for the non-negativity constraint on $\boldsymbol{x}$ as it only takes update direction arguments $\boldsymbol{p}_{k,1}$ and $\boldsymbol{p}_{k,2}$. The negativity constraint might be taken into account by a modification to the CSI measurement inputs, given by

$$S_c^\dagger = S_c(\boldsymbol{p}_{\mathcal{S}_{k,1}}^\dagger, \boldsymbol{p}_{\mathcal{S}_{k,2}}^\dagger) = \frac{\boldsymbol{p}_{\mathcal{S}_{k,1}}^\dagger \cdot \boldsymbol{p}_{\mathcal{S}_{k,2}}^\dagger}{\|\boldsymbol{p}_{\mathcal{S}_{k,1}}^\dagger\|_2 \|\boldsymbol{p}_{\mathcal{S}_{k,2}}^\dagger\|_2}, \tag{5.1}$$

where $\boldsymbol{p}_{\mathcal{S}_{k,1}}^\dagger = P_{\boldsymbol{x}\geq 0}\left[\boldsymbol{x}_k + \alpha_k \boldsymbol{p}_{\mathcal{S}_{k,1}}\right] - \boldsymbol{x}_k$ and $\boldsymbol{p}_{\mathcal{S}_{k,2}}^\dagger = P_{\boldsymbol{x}\geq 0}\left[\boldsymbol{x}_k + \alpha_k \boldsymbol{p}_{\mathcal{S}_{k,2}}\right] - \boldsymbol{x}_k$. This modification would reduce the influence of update directions into negative space on the cosine measurement.

An alternative adaptive subset size methodology was proposed by Bollapragada et al.

2018. This method utilises an *inner product test* to determine when subset size should increase by ensuring the search direction is an ascent direction with high probability. This is achieved by evaluating the variance between every sample's update direction in a subset. If the variance exceeds a variable threshold value (based on the norm of the subset update) the subset size is increased. Future work might implement the algorithm proposed in Bollapragada et al. 2018 as a comparison to AS.

5.3.1.2 Relaxed Step Sizes

In this study, a fixed step-size was used in order to allow for easier comparisons. However, the implementation of a small relaxing step size sequence may aid in the convergence of the AS algorithm. As the step sizes are relaxed, the size of the limit set reduces, which will assist in the AS algorithm's ability to converge closer to the MAP solution. A joint implementation step size relaxation and subset size increases may aid in algorithm performance.

5.3.2 Alternative Stochastic Variance Reduction Configurations

A number of alternatives to the stochastic variance reduction algorithms might be investigated in future studies. Three of these possible alternative configuration changes are discussed in this sub-section.

5.3.2.1 Warm Starting

Many authors of complex optimisation algorithms suggest initialising from an approximation of the parameters, known as a warm start [Schmidt et al. 2017; Tsai et al. 2018]. Therefore, all the SVR algorithms in Chapter 4 were always initialised from the output of one epoch of OSEM reconstruction. This methodology generally performed well and it was observed in initial testing that this could accelerate the algorithms. However, a single epoch of OSEM was selected heuristically. Future studies should investigate alternative warm starting methodologies that could result in lower initial computational cost or an improved initialisation estimate.

Several potential methods for computing approximate estimates exist with various levels of associated computational cost. These methods should approximate the general structure of the reconstruction object and the global image scale but do not need to resolve small details. One example method is filtered back projection of the measured data and heavy post smoothing. This will provide general structural information and is relatively cheap to compute [Qi et al. 2006]. Alternatively, an estimate computed from a fraction of an epoch of OSEM may further reduce the computational requirements. Finally, the AS algorithm

was found to perform particularly well during early epochs. A limited duration initial reconstruction using this algorithm may be effective at initialising the SVR algorithms.

5.3.2.2 Penalty Functions

Chapter 4 results demonstrated algorithm performance by optimising an objective function with the RDP. Yet, it can be expected that the results observed will translate for alternative priors. This was demonstrated in a parallel study to this work where SVREM (Section 5.3.3) and preconditioned SAGA and SVRG algorithms optimised an objective function using the log-cosh prior [Kereta et al. 2021]. The conclusions of SAGA and SVRG algorithm performances were similar between the results demonstrated in this thesis and that work.

The SPDHG algorithm is often implemented with non-smooth penalties, e.g., TV, see Section 2.2.3.2 [Ehrhardt et al. 2019]. This is achieved with the use of the proximal operator [Parikh et al. 2014]. Many stochastic variance reduction algorithms have been extended to the proximal setting, including SAGA, SVRG and those discussed in Section 4.6.8 [Defazio et al. 2014; Xiao et al. 2014]. Therefore, the most commonly used priors in PET image reconstruction are applicable for use with these variance reduction algorithms. Future work should investigate the impact of various regularisation parameters have on SAGA and SVRG convergence rates and hyper-parameters.

Alternative applications of the subset/regularisation methodology could be considered for the SVR algorithms. One method, commonly used in SPDHG, is to consider the prior as another subset. This is discussed in Section 2.3.4. However, because subsets are constructed over the data-fit metric (log-likelihood) one might consider the SVR algorithms for just that data and modify Equation (2.51) to $\tilde{\nabla}_{k,m_k} = \left(\nabla L_{m_k}(\boldsymbol{x}_k) - \boldsymbol{g}_{m_k} + \sum_{\mu}^{M} \boldsymbol{g}_{\mu}\right) + \beta \nabla R(\boldsymbol{x}_k)$, where $\boldsymbol{g}_m$ is updated from $\nabla L_m(\boldsymbol{x})$ measurements. Future work may investigate this removal of the penalty gradient from the variance reduction for a simplified methodology.

5.3.2.3 Preconditioners

This work drew conclusions entirely from the application of the SVR algorithms utilising diagonal, positive definite, EM-like preconditioners, given by Equation (2.29). Allowing the EM preconditioner to vary over a few epochs before anchoring $\boldsymbol{x}$ was observed to improve the convergence of SAGA and SVRG. However, it should be noted that the $\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_k)$ preconditioned SAGA and SVRG reconstructions appeared to numerically converge in both Figure 4.6 and Figure 4.7. Future research may wish to investigate the convergence proof of $\boldsymbol{D}_{\text{EM}}(\boldsymbol{x}_k)$ preconditioned SAGA and SVRG reconstructions using a $\eta > 0$ in Equation (4.2).

Additionally, the application of alternative diagonal preconditioners, that do not iteratively change, remains to be investigated. Examples of these preconditioners are the OS-SPS "precomputed denominator" [Ahn et al. 2003] and other related Hessian row-sum preconditioners [Tsai et al. 2016].

5.3.3 Comparisons with Other Stochastic Algorithms

In recent years, a number of stochastic algorithms have been presented in the PET reconstruction literature. These include SAGA, SVRG and some of those discussed in Section 4.6.8, e.g., SVREM [Kereta et al. 2021] and SPHDG [Ehrhardt et al. 2019]. Thus far, no direct comparison has been made between these algorithms. In future, a study should aim to incorporate these algorithms into a common framework and compare algorithm performance across numerous metrics for multiple data sets with the aim of determining which is most applicable for the next generation of clinical PET reconstruction algorithms.

5.3.4 Second Order Stochastic Optimisation

As an alternative to diagonal preconditioners, further work may explore stochastic second order methods. The L-BFGS-B algorithm, discussed in section 2.3.2.3, applies a local approximation of the inverse of the Hessian $(\nabla^2\Phi_k)^{-1}$ as a preconditioner [Nocedal et al. 2006; Tsai et al. 2018]. The algorithm has been shown to improve the convergence properties of iterative PET reconstruction [Tsai et al. 2018]. However, this algorithm requires full gradient evaluations to construct a suitable approximation of $(\nabla^2\Phi_k)^{-1}$ and does not naturally lend itself to stochastic or subset optimisation due to data consistency requirements. However, two novel stochastic-L-BFGS algorithms have been proposed. The first combines SGA with L-BFGS (SQN) [Byrd et al. 2016] and the other SVRG with L-BFGS (SL-BFGS) [Moritz et al. 2015]. These algorithms nest SGA or SVRG into an external loop that computes the L-BFGS correction pairs from subsets of the data. This step is subtly different to L-BFGS's methodology because of the use of subsets.

While these stochastic second order methods may result in faster convergence in theory, it is unlikely that a significant performance improvement would be realised within 10 epochs, compared to the EM preconditioned SVR algorithms in Chapter 4. The second order methods require a number of epochs to compute the correction pairs used to approximate $(\nabla^2\Phi_k)^{-1}$. However, while these stochastic L-BFGS algorithms may not be clinically relevant, fast converging image reconstruction algorithms are often desired for research purposes.

5.3.5 Alternative PET Applications

In recent years, TOF data reconstructions have become standard in clinical practice. These reconstructions generally require fewer updates for desirable images as additional local LOR information is present [Efthimiou et al. 2019]. However, TOF projection operations are often more computationally expensive. Future work should investigate the application of the algorithms presented in this work to TOF data set reconstructions.

Two unmentioned well researched applications of PET reconstruction relate to the measurement and correction for motion present in PET data [Kyme et al. 2021] and parametric image reconstruction [Gallezot et al. 2019; Zuo et al. 2018].

Reconstructed images of the thorax often suffer from significant resolution loss and degradation due to respiratory and cardiac motion [Nehmeh et al. 2002]. Measurements of the motion can be recorded during a PET scan with an external device or measured from the PET data itself [Thielemans et al. 2011]. These measurements allow for the splitting of the PET data into separate gates, where each gate represents a distinct state of the breathing or cardiac cycle. Thus, the data dimensionality increases, which results in additional reconstruction computational cost.

One may reconstruct each of the gated data separately and register them to a reference gate [Bai et al. 2009]. However, the gated data generally suffers from low SNR because the gate frames occupy a fraction of the full scan duration. Another method is to perform MCIR. MCIR algorithms embed motion correction into the reconstruction sequence and this has been accelerated by constructing subsets of projection angles (as was done in this thesis) and motion gates [Dikaios et al. 2011].

Parametric imaging is the process of creating fully quantitative maps of pharmacokinetic parameters by acquiring a time series of images of the tracer concentration [Gallezot et al. 2019; Charalampos Tsoumpas et al. 2008]. Similar to motion gated data, these time frames may be short in duration leading to low SNR. Parametric imaging can also be combined with motion correction to reduce image artefacts.

In both parametric and motion imaging, the data sets may suffer from high noise and hence reconstructions may be more sensitive to subset algorithms. As shown in Figure 4.11, the SVR algorithms can improve reconstruction convergence rates of low SNR data sets compared to BSREM. This is achieved because of the SVR algorithms reduced sensitivity to unbalanced subsets. The application of stochastic methods to these higher dimensional

data sets may be beneficial in reducing reconstruction computation time and reduce the impact of limit cycle variations. Moreover, subsets could be constructed over these additional data dimensions. This has been previously demonstrated using the SPDHG algorithm [Delplancke et al. 2021].

5.3.6 List Mode Stochastic Variance Reduction

As presented in Section 3.4.3, event-by-event subset list mode reconstruction appears to be an effective methodology for reconstructing PET data. Future work might consider the application of SVR algorithms to list mode data. As observed, this type of data allows for a large number of low variance subsets to be constructed. However, multiple epoch list mode reconstruction is rarely needed for large list mode data sets [Reader et al. 2002]; yet, the SVR algorithms require this to perform as designed. Hence, the SVR algorithms may only be applicable to small list mode data sets.

5.4 List of Publications and Presentations

This section details publications and presentations that relate to this research project, which the author was involved in.

5.4.1 Peer Reviewed Journals

5.4.1.1 Published

- Zeljko Kereta, **Robert Twyman**, Simon Arridge, Kris Thielemans, and Bangti Jin, "Stochastic EM methods with variance reduction for penalised PET reconstructions", in Inverse Problems, 2021.

- Ludovica Brusaferri, Elise C. Emond, Alexandre Bousse, **Robert Twyman**, Alexander C. Whitehead, David Atkinson, Sebastien Ourselin, Brian F. Hutton, Simon Arridge and Kris Thielemans, "Detection Efficiency Modelling and Joint Activity and Attenuation Reconstruction in Non-TOF 3-D PET from Multiple-Energy Window Data", in IEEE Transactions on Radiation and Plasma Medical Sciences, 2022.

5.4.1.2 Submitted

- **Robert Twyman**, Simon Arridge, Zeljko Kereta, Bangti Jin, Ludovica Brusaferri, Sangtae Ahn, Charles W. Stearns, Brian F. Hutton, Irene A. Burger, Fotis Kotasidis and Kris Thielemans, "An Investigation of Stochastic Variance Reduction Algorithms for 3D Penalised PET Image Reconstruction", to IEEE Transactions on Medical

Imaging, 2022. (*Resubmitted April 2022 following revisions.*)

5.4.2 Conference Proceedings

5.4.2.1 Published

- (*Oral Presentation*) **Robert Twyman**, Simon Arridge, and Kris Thielemans, "Investigation of Subset Methodologies Applied to Penalised Iterative PET Reconstruction", in Proceedings of the 16th Virtual International Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine, 2021.

- (*Oral Presentation*) **Robert Twyman**, Simon Arridge, Bangti Jin, Brian F. Hutton, Ludovica Brusaferri, and Kris Thielemans, "Stochastic Variance Reduction Optimisation Algorithms Applied to Iterative PET Reconstruction", in 2020 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2020.

- Ludovica Brusaferri, Elise C. Emond, Alexandre Bousse, **Robert Twyman**, David Atkinson, Brian F. Hutton, Simon Arridge and Kris Thielemans, "Normalisation Factor Estimation in non-TOF 3D PET from Multiple-Energy Window Data", in 2020 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2020.

- (*Oral Presentation*) **Robert Twyman**, Simon Arridge, Brian F. Hutton, Elise C. Emond, Ludovica Brusaferri, Sangtae Ahn and Kris Thielemans, "Iterative PET Image Reconstruction using Adaptive Adjustment of Subset Size and Random Subset Sampling", in 2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2019.

- Zeljko Kereta, **Robert Twyman**, Simon Arridge, Kris Thielemans, and Bangti Jin, "On Stochastic Expectation Maximisation for PET", in Proceedings of the 16th Virtual International Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine, 2021.

5.4.2.2 Accepted for Publication

- (*Poster Presentation*) **Robert Twyman**, Ludovica Brusaferri, Elise C Emond, Francesca Leek, Simon Arridge, Brian F. Hutton, Vesna Cuplov and Kris Thielemans, "A Demonstration of STIR-GATE-Connection", in 2021 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2021.

- Ander Biguri, Palak Wadhwa, Daniel Deidda, Georg Schramm, Kuan-Hao Su, Charles W. Stearns, **Robert Twyman**, Evgueni Ovtchinnikov and Kris Thielemans, "Recent progress in STIR 5.0" in 2021 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2021.

5.4.2.3 To be Presented

- (*Poster Presentation*) **Robert Twyman**, Simon Arridge, Kuan-Hao Su, Fotis Kotasidis, Irene Burger, Catherine Scott, Kris Thielemans, "Application of Stochastic Variance Reduction Algorithms to non-TOF PET Reconstruction", SNMMI, 2022.

5.4.2.4 Submitted

These works are submitted to conferences but not yet accepted.

- Daniel Deidda, Ana M. Denis-Bacelar, Andrew J. Fenwick, Kelley M. Ferreira, Warda Heetun, Brian F. Hutton, Daniel R. McGowan, Andrew P. Robinson, James Scuffham, Kris Thielemans and **Robert Twyman**, "Clinical assessment of Triple Modality Image Reconstruction using Yttrium-90 PET-SPECT-CT", to 2022 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2022.

- Imraj RD. Singh, Riccardo Barbano, **Robert Twyman**, Zeljko Kereta, Bangti Jin, Simon Arridge and Kris Thielemans, "Deep Image Prior PET Reconstruction using a SIRF-Based Objective", to 2022 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2022.

5.4.3 Other Presentations

- (*Oral Presentation*) **Robert Twyman**, "Stochastic Variance Reduction Optimisation Algorithms Applied to Iterative PET Reconstruction", CCP SyneRBI Hackathon, 2021.

- (*Oral Presentation*) **Robert Twyman**, "Image Quality Analysis for Stochastic Algorithms", CCP SyneRBI Hackathon, 2022.

- (*Oral Presentation*) **Robert Twyman**, "STIR-GATE-Connection v1.0", STIR Users Meeting, Dec 2020.

Appendix A

The STIR-GATE-Connection Project

Monte Carlo simulation and image reconstruction software are powerful tools for the development of: new scanner designs, algorithms for image reconstruction, intrinsic resolution modelling, and partial volume modelling. However, the development of Monte Carlo simulation and reconstruction libraries are often separate, resulting in different standards and conventions. Thus, it is important to link these efforts for easier usage. STIR-GATE-Connection aims to provide a simplified pipeline to progress from Monte Carlo simulation using GATE [Jan et al. 2011] to quantitative image reconstruction of a known true tracer distribution using STIR [Thielemans et al. 2012]. STIR-GATE-Connection is a collection of scripts that aid with the: setup of a realistic GATE simulation of a voxelised phantom using a user selected scanner configuration, conversion of the output list mode data into STIR compatible projection data, and computation of additive and multiplicative data corrections for Poisson image reconstruction using STIR. A public release of STIR-GATE-Connection , licensed under the Apache 2.0 License, can be downloaded at: `http://www.github.com/UCL/STIR-GATE-Connection` and has been previously described in Twyman et al. 2021b.

Similar projects exist to link Monte Carlo simulation software to reconstruction software. Examples of these are Customizable and Advanced Software for Tomographic Reconstruction (CaSTOR) [Merlin et al. 2018] and SIMPET [Paredes-Pacheco et al. 2021]. CaSTOR demonstrates the ability to process GATE data into a CASToR datafile format for image reconstruction with a variate of algorithms, along with many other operations. SIMPET is a cloud-based platform that uses SimSET [Baum et al. 2007] for Monte Carlo simulation and STIR for reconstruction.

A.1 GATE Simulations

GATE Monte Carlo simulations are configured via text macro files. STIR-GATE-Connection incorporates a modular approach with these files, which allows various scanners, digitises, phantoms, and more to be substituted into the GATE configuration with ease. The STIR-GATE-Connection project was developed to concatenate collections of GATE macro files and methodologies from a number of previous works, including: Efthimiou et al. 2017, Efthimiou et al. 2019, Emond et al. 2019 and Brusaferri et al. 2022. Initial project development began after the investigation of the AutoSubsets algorithm (Chapter 3) and was used for XCAT simulated data generation for the SVR algorithms studies, see Section 4.2.2.

While GATE offers the ability to simulate emission from voxelised phantoms, the generation of multiple configurations can be tedious and prone to user mistakes due to duplication of parameters. STIR-GATE-Connection simplifies this process with an example GATE simulation script to setup and run voxelised phantom simulations. Given a voxelised phantom, or STIR image generation parameter file, a setup script loads the desired scanner geometry[i] and a density map is generated from the attenuation file. Additionally, a number of GATE macro variables are computed from the voxelised phantom's header files and automatically passed to GATE. These include the origin position, voxel size and number of voxels. Previously, these values were manually set in the macro files by the user. Once setup, the Monte Carlo simulation is run.

STIR-GATE-Connection also supports the running of GATE in batch-array cluster jobs. The initial setup script must be run prior to launching jobs. The simulated scan duration is split across each job of the batch-array with each node computing a small time frame GATE simulation. Once complete, the output files can be concatenated. Example scripts are provided for the Sun Grid Engine cluster system.

A.2 Conversion to Projection Data

The GATE configuration in STIR-GATE-Connection outputs two list mode ROOT files per simulation [Brun et al. 1997]. One file contains coincidence events and the other contains delayed coincidence events. An unlisting script is included in the project that converts these root files into STIR compatible interfile projection data for the relevant scanner. This unlisting can be filtered to exclude various event types.

[i] GE Discovery 690 PET/CT [Bettinardi et al. 2011] and Siemens mMR [Delso et al. 2011] example scanner templates for GATE and STIR are provided in the STIR-GATE-Connection project files.

A.3 Data Modelling Computation via STIR

Iterative reconstruction methods incorporate the estimated contributions of both scattered and random events, as well as physical and geometric factors of the scanner and attenuation factors, into the system model, see Section 2.2.1. STIR-GATE-Connection includes an example script demonstrating how to compute the component-based normalisation factors with crystal efficiencies and "geometric" effects. This normalisation term is represented as $\boldsymbol{A}_{det.sens.}$ in Equation (2.4). Symmetries are imposed on the latter to reduce noise. This is achieved by unlisting only the true coincidence events from a GATE simulation of a uniform cylindrical activity distribution as large as the scanner field-of-view and then comparing the resulting data to a STIR forward projection for the scanner using maximum likelihood computation [Niknejad et al. 2016]. An example, computed using STIR-GATE-Connection , of $\boldsymbol{A}_{det.sens.}$ for the GE Discovery 690 PET/CT is shown in the bottom left of Figure A.1. Additionally, attenuation correction factors, $\boldsymbol{A}_{attenuation}$ in Equation (2.4), may be computed with STIR for inclusion into the system model. An example slice of $\boldsymbol{A}_{det.sens.}\boldsymbol{A}_{attenuation}$ for the XCAT simulated data used in Section 4.2.2 is shown in the bottom right of Figure A.1.

An estimation of random coincidence events may be made with STIR-GATE-Connection tools. Singles are estimated from the delayed coincidence events (recorded by GATE) using a maximum likelihood algorithm and then multiplied to find the randoms rates [Hogg et al. 2002]. However, a discrepancy was observed between the total number of delayed coincidence events and the number of random events in the coincidence data. This was determined to stem from an approximately 9% overestimation of random events in the delayed data (output by GATE) over the randoms events in coincidence data [Twyman et al. 2021b].

An estimation of the scatter contribution may be computed using STIR's iterative single scatter simulation and estimation utilities [C. Tsoumpas et al. 2004; Thielemans et al. 2012]. Three 1D line profiles, across a sinogram projection angle, are shown in Figure A.1. These profiles show the measured coincidence data as well as the estimated and measured random plus scattered coincidence events for the 250 million event XCAT data set described in Section 4.2.2.

Once these models of normalisation, randoms and scatter are computed, example STIR files demonstrate how to reconstruct the data using iterative algorithms. An example OSEM reconstruction is shown in Figure A.2.

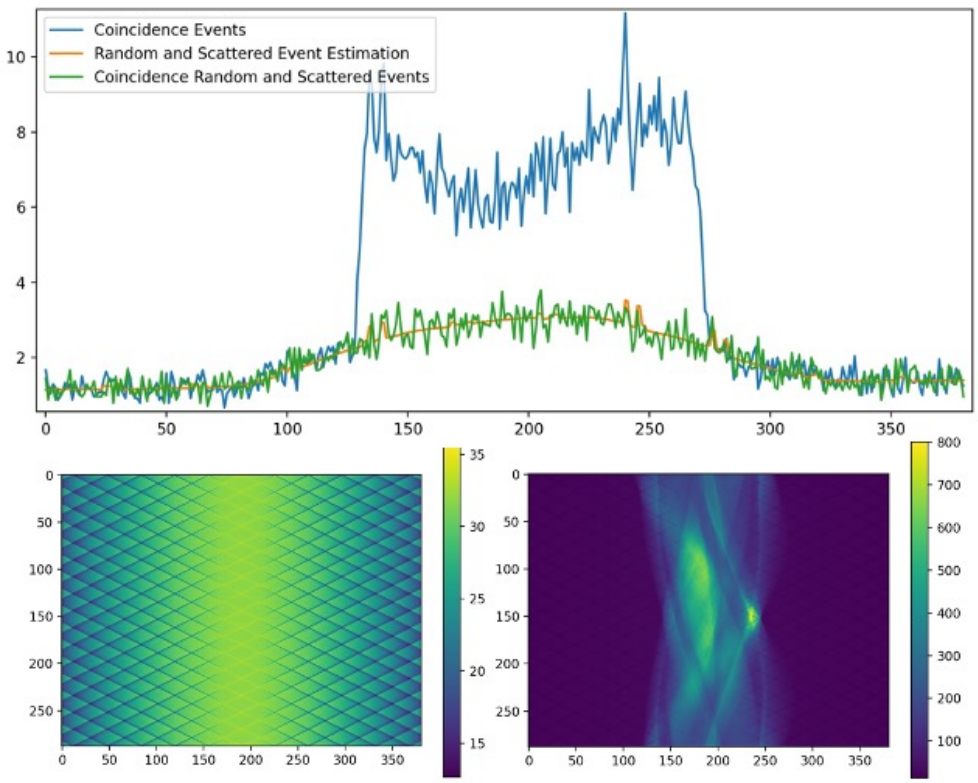


Figure A.1: Top: line profiles over a single sinogram projection angle of the coincidence event data, unlisted random and scattered events data, and the random and scattered event estimation data. Bottom Left: a normalisation sinogram. Bottom Right: a multiplicative correction factors sinogram.

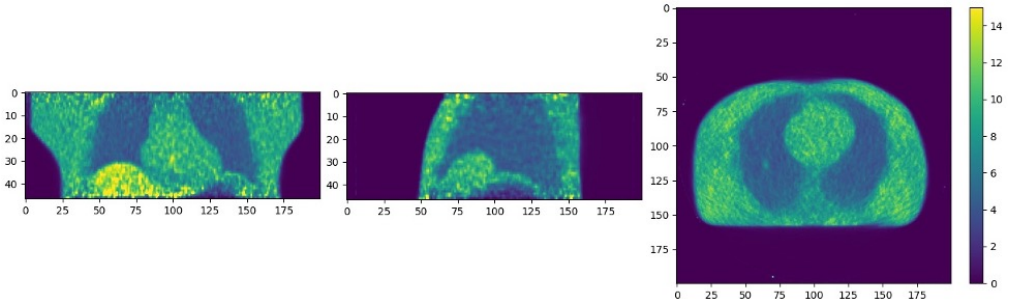


Figure A.2: An OSEM reconstruction of the measured XCAT torso phantom data, simulated in GATE, using 8 subsets and 40 sub-iterations: coronal view (Left), sagittal view (Middle), and transverse view (Right) .

A.4 STIR-GATE-Connection Conclusion

This appendix described the STIR-GATE-Connection project and detailed the easy to use pipeline. STIR-GATE-Connection simplified GATE simulation configuration of PET, provides conversion tools of list mode data into STIR projection data and provides examples for computing system modelling terms. This project allows for iterative reconstruction to be performed on GATE acquired data. Future development of this project could add additional example scanner geometries.

Appendix B

Contributions to STIR

STIR is an open source object-oriented framework for tomographic imaging [Thielemans et al. 2012]. Much of the research presented in this thesis was conducted using STIR tools and utilities. A number of contributions were made to the STIR library that aided the research in this thesis and improve the quality of the software. These changes will assist others in the research community who utilise STIR to some degree. A list of the major and minor contributions, including relevant GitHub links, are provided in this appendix.

B.1 Major Contributions

(a) A number of improvements were made to penalty function classes in STIR. These include:

- Addition of the Relative Difference Prior (RDP). Methods were added to compute the functions value and gradient. `https://github.com/UCL/STIR/pull/335`
- Addition of the Log-cosh prior. Methods were added to compute the functions value, gradient, parabolic surrogate curvature and Hessian vector product. `https://github.com/UCL/STIR/pull/716`
- As aforementioned, Hessian vector products (i.e., `accumulate_Hessian_times_input`) were added to the existing and added prior classes. `https://github.com/UCL/STIR/pull/719/files`
- Consistency tests for penalty function gradients and Hessian methods were added, see: `https://github.com/UCL/STIR/pull/695` and `https://github.com/UCL/STIR/pull/902`

(b) A utility to compute spatially-variant penalty strengths was implemented as per Tsai et al. 2020. In addition to the prior methods, this required the computation of the Hessian vector product for the log-likelihood function in STIR. Previously, `add_multiplication_with_approximate_Hessian` (and its submethods, existed to compute Equation (2.31) [Ahn et al. 2003]. An additional method (`accumulate_Hessian_times_input`) was implemented to compute Equation (2.11) and similar Equation (2.30). `https://github.com/UCL/STIR/pull/733`

(c) A number of modifications were made to STIR's interface with ROOT data, output by GATE. These changes were made in conjunction with STIR-GATE-Connection , see Section A.

- Improvements were made to ROOT I/O by only accessing the needed branch information from a ROOT event when required. This led to (approximately) a factor $2 \times 2 = 4$ speed-up in reading ROOT data from file. `https://github.com/UCL/STIR/pull/813` and `https://github.com/UCL/STIR/pull/814`
- The ability to exclude different types of coincidence events when reading ROOT data was added. `https://github.com/UCL/STIR/pull/850/files`
- Alignment and validation of GATE and STIR crystal location, via consistency tests `https://github.com/UCL/STIR/pull/991`. Similarly, general improvements to STIR's ROOT interface for x-y plane rotations `https://github.com/UCL/STIR/pull/1016`.

(d) A consistency issue was fixed between subset log-likelihood evaluations and subset gradient computation when various sensitivity methods were used. This required a redesign of STIR's log-likelihood gradient computation $\nabla L_m(\boldsymbol{x})$, which was previously formulated as

$$\nabla L_m(\boldsymbol{x}) = \boldsymbol{A}_m^\top \left(\frac{\boldsymbol{y}_m}{\boldsymbol{A}_m(\boldsymbol{x}) + \boldsymbol{b}_m} \right) - \boldsymbol{S}_m. \tag{B.1}$$

This formula was correct when using "subset-sensitivities", i.e., $\boldsymbol{S}_m = \boldsymbol{A}_m^\top \mathbf{1}$. However, an option in STIR exists to use a total sensitivity in the preconditioner. Utilising total

sensitivity in the preconditioner resulted in $\boldsymbol{S}_m = \frac{1}{M}\boldsymbol{S} = \frac{1}{M}\boldsymbol{A}^\top \mathbf{1}$. As these $\boldsymbol{S}_m$ variables were used in the computation of $\nabla L_m(\boldsymbol{x})$, the formula was incorrect. Thus, $\nabla L_m(\boldsymbol{x})$ was modified to

$$\nabla L_m(\boldsymbol{x}) = \boldsymbol{A}_m^T\left(\frac{\boldsymbol{y}_m}{\boldsymbol{A}_m(\boldsymbol{x}) + \boldsymbol{b}_m} - \mathbf{1}\right), \tag{B.2}$$

which resolved this issue. `https://github.com/UCL/STIR/pull/893`

(e) A development branch of STIR exists that allows for the reconstruction and processing of TOF data. Although not used in this thesis, a number of software improvements were made to TOF code.

- Optimised computation of TOF kernel location `https://github.com/NikEfth/STIR/pull/20`.
- Improved projector multithreading over TOF bins `https://github.com/NikEfth/STIR/pull/22`.
- Reduced computational cost of applying TOF kernels by generating a class to interpolate values based upon pre-processed values `https://github.com/NikEfth/STIR/pull/24`.
- Improvements to TOF symmetry and caching of LORs `https://github.com/NikEfth/STIR/pull/25`.
- Re-engineered TOF consistency tests with GATE simulated data `https://github.com/NikEfth/STIR/pull/30`.

B.2 Minor Contributions

(a) Improved documentation and mapping between STIR's and GATE's geometry keywords. `https://github.com/UCL/STIR/pull/569`

(b) Demonstration examples of objective function value computation, gradient ascent (no preconditioning) and a line searches in C++ `https://github.com/UCL/STIR/pull/477`

(c) Allow a Single Slice Re-Binning (SSRB) [Daube-Witherspoon et al. 1987] utility to utilise template projection data geometry for output map. `https://github.com/UCL/STIR/pull/708`

(d) ROI utility improvements `https://github.com/UCL/STIR/pull/913` and `https://github.com/UCL/STIR/pull/915`

(e) Exposed numerous C++ classes and methods to python via swig, e.g:

- Compute total ROI values `https://github.com/UCL/STIR/pull/913`
- Separable gaussian image filter `https://github.com/UCL/STIR/pull/946`

(f) Python script examples `https://github.com/UCL/STIR/pull/934`

(g) Projector OpenMP improvements `https://github.com/UCL/STIR/pull/937`

(h) GE singles to randoms utility fix and investigation python script. `https://github.com/UCL/STIR/pull/961` and `https://github.com/UCL/STIR/pull/941`

(i) Added GE Discovery MI 5 ring scanner geometry `https://github.com/UCL/STIR/pull/981`

(j) Fixed some LOR symmetry operations `https://github.com/UCL/STIR/pull/1000` and `https://github.com/UCL/STIR/pull/1001`

Appendix C

Stochastic Variance Reduction Bias Analysis

A generalised variance reduction analysis is considered to assess the bias estimation of the SVR algorithm gradient approximations [Defazio et al. 2014]. Statistical variance reduction methods estimate a variable $\mathbb{E}[X]$ given the easily computable $\mathbb{E}[Y]$. It is assumed that these two variables are highly correlated, i.e., $X = \nabla\Phi_m(\boldsymbol{x})$ and $Y = \boldsymbol{g}_m$. An estimator ω_ι is introduced to approximate $\mathbb{E}[X]$, which is given by

$$\omega_\iota := \iota(X - Y) + \mathbb{E}[Y], \tag{C.1}$$

where $\iota \in [0,1]$ is a step size. Therefore, $\mathbb{E}[\omega_\iota]$ is given by

$$\mathbb{E}[\omega_\iota] = \iota\mathbb{E}[X] + (1-\iota)\mathbb{E}[Y]. \tag{C.2}$$

Standard variance reduction use $\iota = 1$ such that $\mathbb{E}[\omega_\iota] = \mathbb{E}[X]$ and the estimator is unbiased. The variance of ω_ι is given by

$$\mathrm{Var}(\omega_\iota) = \iota^2\Big[\mathrm{Var}(X) + \mathrm{Var}(Y) - 2\mathrm{Cov}(X,Y)\Big], \tag{C.3}$$

where $\mathrm{Cov}(X,Y)$ is the covariance. If $\mathrm{Cov}(X,Y)$ is large enough then the variance of ω_ι is reduced with respect to X. Varying ι from 1 to 0 reduces the variance towards a minimum but increases the bias towards $\mathbb{E}[Y]$.

Equation (C.1) and Equation (2.51) subtly differ. This is related to the use of $\Phi(\boldsymbol{x}) = \sum_m^M \Phi_m(\boldsymbol{x})$ in this thesis, see Equation (2.41), rather than the $\Phi(\boldsymbol{x}) = 1/M\sum_m^M \Phi_m(\boldsymbol{x})$ applied in Defazio et al. 2014. Thus $\tilde{\nabla}_{k,m_k} = M\omega_\iota$, as ω_ι is a reduced variance estimate of $\nabla\Phi_m(\boldsymbol{x})$, and $\iota = \xi/M$. Hence, in this thesis, the SAGA algorithm uses $\xi = M$ while SAG uses $\xi = \iota$.

This indicates that SAG has greatly reduced variance in its sampling, by a factor $1/M^2$, but at the expense of non-zero bias [Defazio et al. 2014]. Using the same arguments, it is clear that the SVRG algorithm also provides an unbiased estimate of the gradient, like SAGA.

Appendix D

AutoSubsets: A Follow-Up MSc Project

Following the investigations conducted in Chapter 3, AutoSubsets algorithm research was continued by another UCL student, Yiying Zeng, during an MSc project under the co-supervision of this Thesis' author and Prof. Kris Thielemans. This MSc project was presented in [Zeng 2021] and the primary research outcomes are summarised in this appendix.

During the preliminary study of the AutoSubsets (AS) algorithm, discussed in Chapter 3, the algorithm was observed to be sensitive to the subset selection methodology. It was found that the GRS sampling method reduced this sensitivity. However, at later updates, the AS algorithm would increase n_k in sequential updates until $n_k = n_{\max}$. This was a consequence of the S_c values, given by Equation (3.1), maintaining negative values, i.e., $S_c < T = 0.0$. It was theorised that this rapid increase in n_k resulted in a significant reduction in convergence rate.

Zeng 2021 implemented dynamic threshold functions $T(n_k, n_{\max})$ that reduced $T = 0$ to $T = -1$ as $n_k \to n_{\max}$. When implemented into the AS algorithm, reconstruction objective function value performance over the iterates improved. When compared to OS-MAP reconstruction algorithms using various numbers of subsets and subset sampling methods, the AS algorithm (with this dynamic threshold) demonstrated better algorithm performance with reduced limit cycle behaviour.

Additionally, Zeng 2021 compared algorithm performance using different objective function configurations by utilising different β values. It was observed that the AS algorithms performance improved when the stronger penalisation was used.

In summary, the MSc project extended the AutoSubsets algorithm and the use of a dynamic threshold prevented the algorithm from increasing subset size too quickly and improved convergence rates.

Bibliography

Ahn, Sangtae and J.A. Fessler (May 2003). "Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms". In: *IEEE Transactions on Medical Imaging* 22.5, pp. 613–626. DOI: 10.1109/TMI.2003.812251.

Ahn, Sangtae, Steven G. Ross, Evren Asma, Jun Miao, Xiao Jin, Lishui Cheng, Scott D. Wollenweber, and Ravindra M. Manjeshwar (Aug. 2015). "Quantitative comparison of OSEM and penalized likelihood image reconstruction using relative difference penalties for clinical PET". In: *Physics in Medicine and Biology* 60.15, pp. 5733–5751. DOI: 10.1088/0031-9155/60/15/5733.

Allen-Zhu, Zeyuan (Mar. 2016). "Katyusha: The First Direct Acceleration of Stochastic Gradient Methods". In: *Journal of Machine Learning Research* 18, pp. 1–51.

Allisy-Roberts, Penelope and Jerry Williams (2008). *Farr's Physics for Medical Imaging*. Elsevier. DOI: 10.1016/C2009-0-34335-4.

Anand, S. S., H. Singh, and A. K. Dash (2009). "Clinical applications of PET and PET-CT". In: *Medical Journal Armed Forces India* 65.4, pp. 353–358. DOI: 10.1016/S0377-1237(09)80099-3.

Arridge, Simon, Peter Maass, Ozan Öktem, and Carola-Bibiane Schönlieb (May 2019). "Solving inverse problems using data-driven models". In: *Acta Numerica* 28.2019, pp. 1–174. DOI: 10.1017/S0962492919000059.

Asma, Evren, Sangtae Ahn, Hua Qian, Girishankar Gopalakrishnan, Kris Thielemans, Steven G. Ross, Ravindra M. Manjeshwar, and Alexander Ganin (2012a). "Quantitatively accurate image reconstruction for clinical whole-body PET imaging". In: *2012 Conference Handbook - Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, APSIPA ASC 2012*.

Asma, Evren, Sangtae Ahn, Steven G. Ross, Anthony Chen, and Ravindra M. Manjeshwar (Oct. 2012b). "Accurate and consistent lesion quantitation with clinically acceptable penalized likelihood images". In: *2012 IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC)*. x. IEEE, pp. 4062–4066. DOI: 10.1109/NSSMIC.2012.6551928.

Bai, Wenjia and Michael Brady (2009). "Regularized B-spline deformable registration for respiratory motion correction in PET images". In: *Physics in Medicine and Biology* 54.9, pp. 2719–2736. DOI: 10.1088/0031-9155/54/9/008.

Bailey, D.L., J.L. Humm, A. Todd-Pokropek, and A. van Aswegen (2014). *Nuclear Medicine Physics. A Handbook for Teachers and Students.* Vol. 38. 8. IAEA, p. 766. DOI: 10.1118/1.3611048.

Barrett, Harrison H., Timothy White, and Lucas C. Parra (Nov. 1997). "List-mode likelihood". In: *Journal of the Optical Society of America A* 14.11, p. 2914. DOI: 10.1364/JOSAA.14.002914.

Baum, Karl G. and María Helguera (Nov. 2007). "Execution of the SimSET Monte Carlo PET/SPECT Simulator in the Condor Distributed Computing Environment". In: *Journal of Digital Imaging* 20.S1, pp. 72–82. DOI: 10.1007/s10278-007-9058-z.

Bertsekas, Dimitri P. (July 2011). "Incremental Gradient, Subgradient, and Proximal Methods for Convex Optimization: A Survey". In: *Optimization for Machine Learning*. The MIT Press. DOI: 10.7551/mitpress/8996.003.0006.

Bettinardi, V, L Presotto, E Rapisarda, M Picchio, L Gianolli, and M C Gilardi (Sept. 2011). "Physical Performance of the new hybrid PET/CT Discovery-690". In: *Medical Physics* 38.10, pp. 5394–5411. DOI: 10.1118/1.3635220.

Blatt, Doron, Alfred O. Hero, and Hillel Gauchman (2007). "A convergent incremental gradient method with a constant step size". In: *SIAM Journal on Optimization* 18.1, pp. 29–51. DOI: 10.1137/040615961.

Boellaard, Ronald, Roberto Delgado-Bolton, et al. (Feb. 2015). "FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0". In: *European Journal of Nuclear Medicine and Molecular Imaging* 42.2, pp. 328–354. DOI: 10.1007/s00259-014-2961-x.

Bollapragada, Raghu, Richard Byrd, and Jorge Nocedal (2018). "Adaptive Sampling Strategies for Stochastic Optimization". In: *SIAM* 28.4, pp. 3312–3343.

Bottou, Léon, Frank E. Curtis, and Jorge Nocedal (Jan. 2018). "Optimization Methods for Large-Scale Machine Learning". In: *SIAM Review* 60.2, pp. 223–311. DOI: 10.1137/16M1080173.

Bottou, Léon and Yann Le Cun (2004). "Large scale online learning". In: *Advances in Neural Information Processing Systems*. MIT Press.

Browne, Jolyon and A.B. De Pierro (Oct. 1996). "A row-action alternative to the EM algorithm for maximizing likelihood in emission tomography". In: *IEEE Transactions on Medical Imaging* 15.5, pp. 687–699. DOI: 10.1109/42.538946.

Brun, Rene and Fons Rademakers (Apr. 1997). "ROOT - An object oriented data analysis framework". In: *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 389.1-2, pp. 81–86. DOI: 10.1016/S0168-9002(97)00048-X.

Brusaferri, Ludovica, Elise C. Emond, Alexandre Bousse, Robert Twyman, Alexander C. Whitehead, David Atkinson, Sebastien Ourselin, Brian F. Hutton, Simon Arridge, and Kris Thielemans (2022). "Detection Efficiency Modeling and Joint Activity and Attenuation Reconstruction in Non-TOF 3-D PET from Multiple-Energy Window Data". In: *IEEE Transactions on Radiation and Plasma Medical Sciences* 6.1, pp. 87–97. DOI: 10.1109/TRPMS.2021.3064239.

Byrd, R. H., S. L. Hansen, Jorge Nocedal, and Y. Singer (Jan. 2016). "A Stochastic Quasi-Newton Method for Large-Scale Optimization". In: *SIAM Journal on Optimization* 26.2, pp. 1008–1031. DOI: 10.1137/140954362.

Byrd, R. H., Peihuang Lu, Jorge Nocedal, and Ciyou Zhu (Sept. 1995). "A Limited Memory Algorithm for Bound Constrained Optimization". In: *SIAM Journal on Scientific Computing* 16.5, pp. 1190–1208. DOI: 10.1137/0916069.

Byrne, C.L. (1998). "Accelerating the EMML algorithm and related iterative algorithms by rescaled block-iterative methods". In: *IEEE Transactions on Image Processing* 7.1, pp. 100–109. DOI: 10.1109/83.650854.

Caribé, Paulo R R V, M Koole, Yves D'Asseler, B. Van Den Broeck, and S Vandenberghe (Dec. 2019). "Noise reduction using a Bayesian penalized-likelihood reconstruction algorithm on a time-of-flight PET-CT scanner". In: *EJNMMI Physics* 6.22, pp. 1–14. DOI: 10.1186/s40658-019-0264-9.

Chambolle, Antonin, Matthias J. Ehrhardt, Peter Richtárik, and Carola-Bibiane Schönlieb (Jan. 2018). "Stochastic Primal-Dual Hybrid Gradient Algorithm with Arbitrary Sampling and Imaging Applications". In: *SIAM Journal on Optimization* 28.4, pp. 2783–2808. DOI: 10.1137/17M1134834.

Chilra, P., S. Gnesin, G. Allenbach, M. Monteiro, J. O. Prior, L. Vieira, and J. A. Pires Jorge (2017). "Cardiac PET/CT with Rb-82: optimization of image acquisition and reconstruction parameters". In: *EJNMMI Physics* 4.1. DOI: 10.1186/s40658-017-0178-3.

Conti, Maurizio and Lars Eriksson (2016). "Physics of pure and non-pure positron emitters for PET: A review and a discussion". In: *EJNMMI Physics* 3.22. DOI: 10.1186/s40658-016-0144-5.

Daube-Witherspoon, M. E. and G. Muehllehner (1987). "Treatment of axial data in three-dimensional PET". In: *Journal of Nuclear Medicine* 28.11, pp. 1717–1724.

De Pierro, Alvaro R. (1995). "A Modified Expectation Maximization Algorithm for Penalized Likelihood Estimation in Emission Tomography". In: *IEEE Transactions on Medical Imaging* 14.1, pp. 132–137. DOI: 10.1109/42.370409.

De Pierro, Alvaro R. and Michel Eduardo Beleza Yamagishi (2001). "Fast EM-like methods for maximum "a posteriori" estimates in emission tomography". In: *IEEE Transactions on Medical Imaging* 20.4, pp. 280–288. DOI: 10.1109/42.921477.

Defazio, Aaron, Francis Bach, and Simon Lacoste-Julien (July 2014). "SAGA: A Fast Incremental Gradient Method With Support for Non-Strongly Convex Composite Objectives". In: *Advances in Neural Information Processing Systems*. Vol. 27. Curran Associates, Inc.

Delplancke, Claire, Kris Thielemans, and Matthias J. Ehrhardt (2021). "Accelerated Convergent Motion Compensated Image Reconstruction". In: *2021 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*.

Delso, Gaspar, S. Furst, Björn Jakoby, Ralf Ladebeck, Carl Ganter, Stephan G. Nekolla, Markus Schwaiger, and Sibylle I. Ziegler (Dec. 2011). "Performance Measurements of the Siemens mMR Integrated Whole-Body PET/MR Scanner". In: *Journal of Nuclear Medicine* 52.12, pp. 1914–1922. DOI: 10.2967/jnumed.111.092726.

Dikaios, N. and T. D. Fryer (2011). "Acceleration of motion-compensated PET reconstruction: Ordered subsets-gates em algorithms and a priori reference gate information". In: *Physics in Medicine and Biology*. DOI: 10.1088/0031-9155/56/6/011.

Driggs, Derek, Matthias J. Ehrhardt, and Carola Bibiane Schönlieb (2020a). "Accelerating variance-reduced stochastic gradient methods". In: *Mathematical Programming*. DOI: 10.1007/s10107-020-01566-2.

Driggs, Derek, Jingwei Liang, and Carola-Bibiane Schönlieb (June 2020b). "On Biased Stochastic Gradient Estimation". In: *Journal of Machine Learning Research* 23.

Durie, Brian G.M., Alan D. Waxman, Allesandro D'Agnolo, and Cindy M. Williams (2002). "Whole-body 18F-FDG pet identifies high-risk myeloma". In: *Journal of Nuclear Medicine*.

Efthimiou, Nikos, Elise Emond, Chris Cawthorne, Charalampos Tsoumpas, and Kris Thielemans (Oct. 2017). "Reconstruction of Time-of-Flight Projection Data with the STIR reconstruction framework". In: *2017 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*. IEEE, pp. 1–3. DOI: 10.1109/NSSMIC.2017.8533081.

Efthimiou, Nikos, Elise Emond, Palak Wadhwa, Christopher Cawthorne, Charalampos Tsoumpas, and Kris Thielemans (2019). "Implementation and validation of time-of-flight PET image reconstruction module for listmode and sinogram projection data in the STIR library". In: *Physics in Medicine and Biology* 64.3. DOI: 10.1088/1361-6560/aaf9b9.

Ehrhardt, Matthias J., Pawel Markiewicz, et al. (2016). "PET Reconstruction with an Anatomical MRI Prior Using Parallel Level Sets". In: *IEEE Transactions on Medical Imaging* 35.9, pp. 2189–2199. DOI: 10.1109/TMI.2016.2549601.

Ehrhardt, Matthias J., Pawel J Markiewicz, and Carola-Bibiane Schönlieb (Aug. 2019). "Faster PET reconstruction with non-smooth priors by randomization and precondi-

tioning". In: *Physics in Medicine and Biology*. DOI: 10.1088/1361-6560/ab3d07.

Ehrhardt, Matthias J., Pawel J. Markiewicz, Peter Richtárik, Jonathan Schott, Antonin Chambolle, and Carola-Bibiane Schoenlieb (Aug. 2017). "Faster PET reconstruction with a stochastic primal-dual hybrid gradient method". In: *Wavelets and Sparsity XVII*. August 2017. SPIE, p. 58. DOI: 10.1117/12.2272946.

Emond, Elise C., Ashley M. Groves, Brian F. Hutton, and Kris Thielemans (2019). "Effect of positron range on PET quantification in diseased and normal lungs". In: *Physics in Medicine and Biology* 64.20. DOI: 10.1088/1361-6560/ab469d.

Erdoğdu, Ferruh (2019). *Optimization: An introduction*. ISBN: 978-0367269692.

Erlandsson, Kjell, Irene Buvat, P. Hendrik Pretorius, Benjamin A. Thomas, and Brian F. Hutton (2012). "A review of partial volume correction techniques for emission tomography and their applications in neurology, cardiology and oncology". In: *Physics in Medicine and Biology* 57.21. DOI: 10.1088/0031-9155/57/21/R119.

Feng, Li, Robert Grimm, Kai T.obias Block, Hersh Chandarana, Sungheon Kim, Jian Xu, Leon Axel, Daniel K. Sodickson, and Ricardo Otazo (Sept. 2014). "Golden-angle radial sparse parallel MRI: Combination of compressed sensing, parallel imaging, and golden-angle radial sampling for fast and flexible dynamic volumetric MRI". In: *Magnetic Resonance in Medicine* 72.3, pp. 707–717. DOI: 10.1002/mrm.24980.

Fessler, Jeffrey A. and Hakan Erdogan (1999). "Paraboloidal surrogates algorithm for convergent penalized-likelihood emission image reconstruction". In: *IEEE Nuclear Science Symposium and Medical Imaging Conference* 2, pp. 1132–1135. DOI: 10.1109/nssmic.1998.774361.

Fiechter, Michael, Jelena R. Ghadri, Cathérine Gebhard, Tobias A. Fuchs, Aju P. Pazhenkottil, Rene N. Nkoulou, Bernhard A. Herzog, Christophe A. Wyss, Oliver Gaemperli, and Philipp A. Kaufmann (2012). "Diagnostic value of 13N-ammonia myocardial perfusion PET: Added value of myocardial flow reserve". In: *Journal of Nuclear Medicine*. DOI: 10.2967/jnumed.111.101840.

Flajolet, Philippe, Danièle Gardy, and Loÿs Thimonier (1992). "Birthday paradox, coupon collectors, caching algorithms and self-organizing search". In: *Discrete Applied Mathematics* 39.3, pp. 207–229. DOI: 10.1016/0166-218X(92)90177-C.

Gallezot, Jean-Dominique, Yihuan Lu, Mika Naganawa, and Richard E. Carson (Apr. 2019). "Parametric Imaging with PET and SPECT". In: *IEEE Transactions on Radiation and Plasma Medical Sciences*, pp. 1–1. DOI: 10.1109/trpms.2019.2908633.

Geman, Stuart and Donald Geman (1984). "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*. DOI: 10.1109/TPAMI.1984.4767596.

Gower, Robert M., Mark Schmidt, Francis Bach, and Peter Richtárik (Oct. 2020). "Variance-Reduced Methods for Machine Learning". In: *Proceedings of the IEEE* 108.11, pp. 1968–1983. DOI: 10.1109/JPROC.2020.3028013.

Green, P.J. (Mar. 1990). "Bayesian reconstructions from emission tomography data using a modified EM algorithm". In: *IEEE Transactions on Medical Imaging* 9.1, pp. 84–93. DOI: 10.1109/42.52985.

Herman, Gabor T. and Lorraine B. Meyer (1993). "Algebraic Reconstruction Techniques Can Be Made Computationally Efficient". In: *IEEE Transactions on Medical Imaging* 12.3, pp. 600–609. DOI: 10.1109/42.241889.

Hogg, D., K. Thielemans, S. Mustafovic, and T.J. Spinks (2002). "A study of bias for various iterative reconstruction methods in PET". In: *2002 IEEE Nuclear Science Symposium Conference Record*. Vol. 3. IEEE, pp. 1519–1523. DOI: 10.1109/NSSMIC.2002.1239610.

Howard, Brandon A., Rustain Morgan, Matthew P. Thorpe, Timothy G. Turkington, Jorge Oldan, Olga G. James, and Salvador Borges-Neto (2017). "Comparison of Bayesian penalized likelihood reconstruction versus OS-EM for characterization of small pulmonary nodules in oncologic PET/CT". In: *Annals of Nuclear Medicine* 31.8, pp. 623–628. DOI: 10.1007/s12149-017-1192-1.

Hudson, H.M. and R.S. Larkin (1994). "Accelerated image reconstruction using ordered subsets of projection data". In: *IEEE Transactions on Medical Imaging* 13.4, pp. 601–609. DOI: 10.1109/42.363108.

Iriarte, A., R. Marabini, S. Matej, C. O.S. Sorzano, and R. M. Lewitt (2016). "System models for PET statistical iterative reconstruction: A review". In: *Computerized Medical Imaging and Graphics* 48, pp. 30–48. DOI: 10.1016/j.compmedimag.2015.12.003.

Jan, S., D. Benoit, et al. (2011). "GATE V6: A major enhancement of the GATE simulation platform enabling modelling of CT and radiotherapy". In: *Physics in Medicine and Biology* 56.4, pp. 881–901. DOI: 10.1088/0031-9155/56/4/001.

Johnson, Rie and Tong Zhang (2013). "Accelerating stochastic gradient descent using predictive variance reduction". In: *Advances in Neural Information Processing Systems*. Vol. 26. Curran Associates, Inc. DOI: 10.5555/2999611.2999647.

Kereta, Zeljko, Robert Twyman, Simon Arridge, Kris Thielemans, and Bangti Jin (Nov. 2021). "Stochastic EM methods with variance reduction for penalised PET reconstructions". In: *Inverse Problems* 37.11, p. 115006. DOI: 10.1088/1361-6420/ac2d74.

Kohler, Thomas (2004). "A projection access scheme for iterative reconstruction based on the golden section". In: *IEEE Nuclear Science Symposium Conference Record* 6.C, pp. 3961–3965. DOI: 10.1109/NSSMIC.2004.1466745.

Kumar, R, D Halanaik, and A Malhotra (2010). "Clinical applications of positron emission tomography-computed tomography in oncology". In: *Indian Journal of Cancer* 47.2, p. 100. DOI: 10.4103/0019-509X.62997.

Kyme, Andre Z. and Roger R. Fulton (2021). "Motion estimation and correction in SPECT, PET and CT". In: *Physics in Medicine and Biology* 66.18. DOI: 10.1088/1361-6560/ac093b.

Lange, K. and R. Carson (1984). "EM reconstruction algorithms for emission and transmission tomography". In: *Journal of Computer Assisted Tomography* 8.2, pp. 306–316.

Lantos, Judit, Erik Mittra, Craig Levin, and Andrei Iagaru (2015). "Standard OSEM vs. regularized PET image reconstruction: qualitative and semi-quantitative comparison". In: *Journal of Nuclear Medicine* 56.supplement 3, p. 1805. DOI: DOI10.1093/heapol/czi050.

Leahy, Richard M. and Jinyi Qi (2000). "Statistical approaches in quantitative positron emission tomography". In: *Statistics and Computing* 10.2, pp. 147–165. DOI: 10.1023/A:1008946426658.

Lewitt, Robert M. and Gerd Muehllehner (1986). "Accelerated Iterative Reconstruction for Positron Emission Tomography Based on the EM Algorithm for Maximum Likelihood Estimation". In: *IEEE Transactions on Medical Imaging* 5.1, pp. 16–22. DOI: 10.1109/TMI.1986.4307734.

Meikle, Steven R., Vesna Sossi, et al. (Mar. 2021). "Quantitative PET in the 2020s: a roadmap". In: *Physics in Medicine & Biology* 66.6, 06RM01. DOI: 10.1088/1361-6560/abd4f7.

Merlin, Thibaut, Simon Stute, Didier Benoit, Julien Bert, Thomas Carlier, Claude Comtat, Marina Filipovic, Frédéric Lamare, and Dimitris Visvikis (Sept. 2018). "CASToR: a generic data organization and processing code framework for multi-modal and multi-dimensional tomographic reconstruction". In: *Physics in Medicine & Biology* 63.18, p. 185005. DOI: 10.1088/1361-6560/aadac1.

Moritz, Philipp, Robert Nishihara, and Michael I Jordan (Aug. 2015). "A Linearly-Convergent Stochastic L-BFGS Algorithm". In: *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics, AISTATS 2016*, pp. 249–258.

Nehmeh, S. A., Y. E. Erdi, et al. (Feb. 2002). "Effect of respiratory gating on reducing lung motion artifacts in PET imaging of lung cancer". In: *Medical Physics* 29.3, pp. 366–371. DOI: 10.1118/1.1448824.

Nesterov, Yurii (2004). *Introductory Lectures on Convex Optimization*. Vol. 87. Applied Optimization. Boston, MA: Springer US. DOI: 10.1007/978-1-4419-8853-9.

— (2013). *Gradient methods for minimizing composite functions*. Vol. 140. 1, pp. 125–161. DOI: 10.1007/s10107-012-0629-5.

Nguyen, Lam M., Jie Liu, Katya Scheinberg, and Martin Takáč (Feb. 2017). "SARAH: A Novel Method for Machine Learning Problems Using Stochastic Recursive Gradient". In: *34th International Conference on Machine Learning, ICML*.

Niknejad, Tahereh, Stefaan Tavernier, Joao Varela, and Kris Thielemans (Oct. 2016). "Validation of 3D model-based maximum-likelihood estimation of normalisation factors for partial ring positron emission tomography". In: *2016 IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop (NSS/MIC/RTSD)*. IEEE, pp. 1–5. DOI: 10.1109/NSSMIC.2016.8069577.

Nikolova, M. and M. Ng (2001). "Fast image reconstruction algorithms combining half-quadratic regularization and preconditioning". In: *Proceedings 2001 International Conference on Image Processing*. Vol. 1. 4. IEEE, pp. 277–280. DOI: 10.1109/ICIP.2001.959007.

Nocedal, Jorge and Stephen J Wright (2006). *Numerical Optimization*. Springer Series in Operations Research and Financial Engineering. Springer New York. DOI: 10.1007/978-0-387-40065-5.

Nuyts, J, D. Beque, P Dupont, and L Mortelmans (Feb. 2002). "A concave prior penalizing relative differences for maximum-a-posteriori reconstruction in emission tomography". In: *IEEE Transactions on Nuclear Science* 49.1, pp. 56–60. DOI: 10.1109/TNS.2002.998681.

Pan, Tinsu, Samuel A. Einstein, Srinivas Cheenu Kappadath, Kira S. Grogg, Cristina Lois Gomez, Adam M. Alessio, William C. Hunter, Georges El Fakhri, Paul E. Kinahan, and Osama R. Mawlawi (July 2019). "Performance evaluation of the 5-Ring GE Discovery MI PET/CT system using the national electrical manufacturers association NU 2-2012 Standard". In: *Medical Physics* 46.7, pp. 3025–3033. DOI: 10.1002/mp.13576.

Paredes-Pacheco, José, Francisco Javier López-González, Jesús Silva-Rodríguez, Nikos Efthimiou, Aida Niñerola-Baizán, Álvaro Ruibal, Núria Roé-Vellvé, and Pablo Aguiar (Mar. 2021). "SimPET—An open online platform for the Monte Carlo simulation of realistic brain PET data. Validation for 18 F-FDG scans". In: *Medical Physics*. DOI: 10.1002/mp.14838.

Parikh, Neal and Stephen Boyd (2014). "Proximal Algorithms". In: *Foundations and Trends in Optimization* 1.3, pp. 127–239. DOI: 10.1561/2400000003.

Parra, Lucas and Harrison H. Barrett (1998). "List-mode likelihood: EM algorithm and image quality estimation demonstrated on 2-D PET". In: *IEEE Transactions on Medical Imaging* 17.2, pp. 228–235. DOI: 10.1109/42.700734.

Qi, Jinyi and Richard M. Leahy (Aug. 2006). "Iterative reconstruction techniques in emission computed tomography". In: *Physics in Medicine and Biology* 51.15, R541–R578. DOI: 10.1088/0031-9155/51/15/R01.

Reader, Andrew J., Stijn Ally, Filippos Bakatselos, Roido Manavaki, Richard J. Walledge, Alan P. Jeavons, Peter J. Julyan, Sha Zhao, David L. Hastings, and Jamal Zweit (2002). "One-pass list-mode EM algorithm for high-resolution 3-D PET image reconstruction into large arrays". In: *IEEE Transactions on Nuclear Science* 49 I.3, pp. 693–699. DOI: 10.1109/TNS.2002.1039550.

Rockmore, A. J. and Albert Macovski (1976). "A Maximum Likelihood Approach to Emission Image Reconstruction from Projections". In: *IEEE Transactions on Nuclear Science* 23.4, pp. 1428–1432. DOI: 10.1109/TNS.1976.4328496.

Ross, Steve (2014). "Q. Clear". In: *GE Healthcare, White Paper*, pp. 1–9.

Roux, Nicolas Le, Mark Schmidt, and Francis Bach (Feb. 2012). "A Stochastic Gradient Method with an Exponential Convergence Rate for Finite Training Sets". In: *Nips* 1.2, pp. 1–34.

Ruder, Sebastian (2016). "An overview of gradient descent optimization algorithms". In: *ArXiv* abs/1609.04747.

Rudin, Leonid I, Stanley Osher, and Emad Fatemi (Nov. 1992). "Nonlinear total variation based noise removal algorithms". In: *Physica D: Nonlinear Phenomena* 60.1-4, pp. 259–268. DOI: 10.1016/0167-2789(92)90242-F.

Sah, Bert Ram, Paul Stolzmann, et al. (2017). "Clinical evaluation of a block sequential regularized expectation maximization reconstruction algorithm in18F-FDG PET/CT studies". In: *Nuclear Medicine Communications* 38.1, pp. 57–66. DOI: 10.1097/MNM.0000000000000604.

Schmidt, Mark, Nicolas Le Roux, and Francis Bach (2017). "Minimizing finite sums with the stochastic average gradient". In: *Mathematical Programming* 162.1-2, pp. 83–112. DOI: 10.1007/s10107-016-1030-6.

Schramm, Georg and Martin Holler (Jan. 2022). "Fast and memory-efficient reconstruction of sparse Poisson data in listmode with non-smooth priors with application to time-of-flight PET". In: *Physics in Medicine and Biology* 2.1. DOI: 10.1088/1361-6560/ac71f1.

Segars, W. P., G. Sturgeon, S. Mendonca, Jason Grimes, and B. M. W. Tsui (Aug. 2010). "4D XCAT phantom for multimodality imaging research". In: *Medical Physics* 37.9, pp. 4902–4915. DOI: 10.1118/1.3480985.

Shang, Fanhua, Yuanyuan Liu, James Cheng, and Jiacheng Zhuo (2017). "Fast Stochastic Variance Reduced Gradient Method with Momentum Acceleration for Machine Learning". In.

Shepp, L A and Benjamin F Logan (June 1974). "FOURIER RECONSTRUCTION OF A HEAD SECTION." In: *IEEE Transactions on Nuclear Science* NS-21.3, pp. 21–43. DOI: 10.1109/tns.1974.6499235.

Shepp, L. A. and Y. Vardi (Oct. 1982). "Maximum Likelihood Reconstruction for Emission Tomography". In: *IEEE Transactions on Medical Imaging* 1.2, pp. 113–122. DOI: 10.1109/TMI.1982.4307558.

Somayajula, Sangeetha, Evren Asma, and Richard M. Leahy (2005). "PET image reconstruction using anatomical information through mutual information based priors". In: *IEEE Nuclear Science Symposium Conference Record* 5.5, pp. 2722–2726. DOI: 10.1109/NSSMIC.2005.1596899.

Spencer, Benjamin A., Eric Berg, et al. (June 2021). "Performance Evaluation of the uEXPLORER Total-Body PET/CT Scanner Based on NEMA NU 2-2018 with Additional Tests to Characterize PET Scanners with a Long Axial Field of View". In: *Journal of Nuclear Medicine* 62.6, pp. 861–870. DOI: 10.2967/jnumed.120.250597.

Teoh, Eugene J, Daniel R McGowan, Ruth E Macpherson, Kevin M Bradley, and Fergus V Gleeson (Sept. 2015). "Phantom and Clinical Evaluation of the Bayesian Penalized Likelihood Reconstruction Algorithm Q.Clear on an LYSO PET/CT System". In:

Journal of Nuclear Medicine 56.9, pp. 1447–1452. DOI: 10.2967/jnumed.115.159301.

Thielemans, Kris and Simon Arridge (Oct. 2015). "Adaptive adjustment of the number of subsets during iterative image reconstruction". In: *2015 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*. IEEE, pp. 1–2. DOI: 10.1109/NSSMIC.2015.7582225.

Thielemans, Kris, Shailendra Rathore, Fredrik Engbrant, and Pasha Razifar (2011). "Device-less gating for PET/CT using PCA". In: *IEEE Nuclear Science Symposium Conference Record*. Institute of Electrical and Electronics Engineers Inc., pp. 3904–3910. DOI: 10.1109/NSSMIC.2011.6153742.

Thielemans, Kris, Charalampos Tsoumpas, Sanida Mustafovic, Tobias Beisel, Pablo Aguiar, Nikolaos Dikaios, and Matthew W. Jacobson (Feb. 2012). "STIR: software for tomographic image reconstruction release 2". In: *Physics in Medicine and Biology* 57.4, pp. 867–883. DOI: 10.1088/0031-9155/57/4/867.

Tsai, Yu-Jung, Alexandre Bousse, Matthias J. Ehrhardt, Charles W. Stearns, Sangtae Ahn, Brian F. Hutton, Simon Arridge, and Kris Thielemans (Apr. 2018). "Fast Quasi-Newton Algorithms for Penalized Reconstruction in Emission Tomography and Further Improvements via Preconditioning". In: *IEEE Transactions on Medical Imaging* 37.4, pp. 1000–1010. DOI: 10.1109/TMI.2017.2786865.

Tsai, Yu-Jung, Alexandre Bousse, Charles W. Stearns, Sangtae Ahn, Brian F. Hutton, Simon Arridge, and Kris Thielemans (Oct. 2016). "Performance improvement and validation of a new MAP reconstruction algorithm". In: *2016 IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop (NSS/MIC/RTSD)*. Vol. 2017-Janua. IEEE, pp. 1–3. DOI: 10.1109/NSSMIC.2016.8069458.

Tsai, Yu-Jung, Georg Schramm, Sangtae Ahn, Alexandre Bousse, Simon Arridge, Johan Nuyts, Brian F. Hutton, Charles W. Stearns, and Kris Thielemans (Jan. 2020). "Benefits of Using a Spatially-Variant Penalty Strength With Anatomical Priors in PET Reconstruction". In: *IEEE Transactions on Medical Imaging* 39.1, pp. 11–22. DOI: 10.1109/TMI.2019.2913889.

Tsoumpas, C., P. Aguiar, K.S. Nikita, D. Ros, and K. Thielemans (2004). "Evaluation of the Single Scatter Simulation Algorithm Implemented in the STIR Library". In: *IEEE Symposium Conference Record Nuclear Science 2004.* Vol. 6. June. IEEE, pp. 3361–3365. DOI: 10.1109/NSSMIC.2004.1466455.

Tsoumpas, Charalampos, Federico E. Turkheimer, and Kris Thielemans (2008). "A survey of approaches for direct parametric image reconstruction in emission tomography". In: *Medical Physics* 35.9, pp. 3963–3971. DOI: 10.1118/1.2966349.

Twyman, Robert, Simon Arridge, Brian F Hutton, Elise C. Emond, Ludovica Brusaferri, Sangtae Ahn, and Kris Thielemans (Oct. 2019). "Iterative PET Image Reconstruction using Adaptive Adjustment of Subset Size and Random Subset Sampling". In: *2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*. IEEE, pp. 1–3. DOI: 10.1109/NSS/MIC42101.2019.9059799.

Twyman, Robert, Simon Arridge, Bangti Jin, Brian F. Hutton, Ludovica Brusaferri, and Kris Thielemans (Oct. 2020). "Stochastic Variance Reduction Optimisation Algorithms Applied to Iterative PET Reconstruction". In: *2020 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*. IEEE, pp. 1–2. DOI: 10.1109/NSS/MIC42677.2020.9508105.

Twyman, Robert, Simon Arridge, and Kris Thielemans (2021a). "Investigation of Subset Methodologies Applied to Penalised Iterative PET Reconstruction". In: *Proceedings of the 16th Virtual International Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine* 14.July, pp. 242–245.

Twyman, Robert, Ludovica Brusaferri, Elise C Emond, Francesca Leek, Simon Arridge, Brian F. Hutton, Vesna Cuplov, and Kris Thielemans (2021b). "A Demonstration of STIR-GATE-Connection". In: *2021 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*.

Vos, Charlotte S. van der, Daniëlle Koopman, Sjoerd Rijnsdorp, Albert J. Arends, Ronald Boellaard, Jorn A. van Dalen, Mark Lubberink, Antoon T.M. Willemsen, and Eric P. Visser (Aug. 2017). "Quantification, improvement, and harmonization of small lesion detection with state-of-the-art PET". In: *European Journal of Nuclear Medicine and Molecular Imaging* 44.S1, pp. 4–16. DOI: 10.1007/s00259-017-3727-z.

Wadhwa, Palak, Kris Thielemans, et al. (2018). “Implementation of Image Reconstruction for GE SIGNA PET/MR PET Data in the STIR Library”. In: *2018 IEEE Nuclear Science Symposium and Medical Imaging Conference, NSS/MIC 2018 - Proceedings*, pp. 35–37. DOI: 10.1109/NSSMIC.2018.8824341.

Wadhwa, Palak, Kris Thielemans, et al. (2021). “PET image reconstruction using physical and mathematical modelling for time of flight PET-MR scanners in the STIR library”. In: 185, pp. 110–119. DOI: 10.1016/j.ymeth.2020.01.005.

Wang, Guobao and Jinyi Qi (Jan. 2015). “PET Image Reconstruction Using Kernel Method”. In: *IEEE Transactions on Medical Imaging* 34.1, pp. 61–71. DOI: 10.1109/TMI.2014.2343916.

Webster Stayman, J. and Jeffrey A. Fessler (2000). “Regularization for uniform spatial resolution properties in penalized-likelihood image reconstruction”. In: *IEEE Transactions on Medical Imaging* 19.6, pp. 601–615. DOI: 10.1109/42.870666.

Winkelmann, Stefanie, Tobias Schaeffter, Thomas Koehler, Holger Eggers, and Olaf Doessel (Jan. 2007). “An Optimal Radial Profile Order Based on the Golden Ratio for Time-Resolved MRI”. In: *IEEE Transactions on Medical Imaging* 26.1, pp. 68–76. DOI: 10.1109/TMI.2006.885337.

Wolk, David A., Zheng Zhang, Sanaa Boudhar, Christopher M. Clark, Michael J. Pontecorvo, and Steven E. Arnold (2012). “Amyloid imaging in Alzheimer’s disease: Comparison of florbetapir and Pittsburgh compound-B positron emission tomography”. In: *Journal of Neurology, Neurosurgery and Psychiatry*. DOI: 10.1136/jnnp-2012-302548.

Xiao, Lin and Tong Zhang (2014). “A proximal stochastic gradient method with progressive variance reduction”. In: *SIAM Journal on Optimization* 24.4, pp. 2057–2075. DOI: 10.1137/140961791.

Zeng, Yiying (2021). “Adaptive Adjustment of Subset Size for Iterative PET Reconstruction”. University College London.

Zhu, Ciyou, Richard H Byrd, Peihuang Lu, and Jorge Nocedal (Dec. 1997). “Algorithm 778: L-BFGS-B”. In: *ACM Transactions on Mathematical Software* 23.4, pp. 550–560. DOI: 10.1145/279232.279236.

Zuo, Yang, Jinyi Qi, and Guobao Wang (2018). “Relative Patlak plot for dynamic PET parametric imaging without the need for early-time input function”. In: *Physics in Medicine and Biology* 63.16. DOI: 10.1088/1361-6560/aad444.