

An Investigation of Stochastic Variance Reduction Algorithms for Relative Difference Penalised 3D PET Image Reconstruction

Robert Twyman, *Graduate Student Member, IEEE*, Simon Arridge, Zeljko Kereta, Bangti Jin, Ludovica Brusaferrri, Sangtae Ahn, *Member, IEEE*, Charles W. Stearns, *Fellow, IEEE*, Brian F. Hutton, *Senior Member, IEEE*, Irene A. Burger, Fotis Kotasidis, and Kris Thielemans, *Senior Member, IEEE*

Abstract—Penalised PET image reconstruction algorithms are often accelerated during early iterations with the use of subsets. However, these methods may exhibit limit cycle behaviour at later iterations due to variations between subsets. Desirable converged images can be achieved for a subclass of these algorithms via the implementation of a relaxed step size sequence, but the heuristic selection of parameters will impact the quality of the image sequence and algorithm convergence rates. In this work, we demonstrate the adaption and application of a class of stochastic variance reduction gradient algorithms for PET image reconstruction using the relative difference penalty and numerically compare convergence performance to BSREM. The two investigated algorithms are: SAGA and SVRG. These algorithms require the retention in memory of recently computed subset gradients, which are utilised in subsequent updates. We present several numerical studies based on Monte Carlo simulated data and a patient data set for fully 3D PET acquisitions. The impact of the number of subsets, different preconditioners and step size methods on the convergence of regions of interest values within the reconstructed images is explored. We observe that when using constant preconditioning, SAGA and SVRG demonstrate reduced variations in voxel values between subsequent updates and are less reliant on step size hyper-parameter selection than BSREM reconstructions. Furthermore, SAGA and SVRG can converge significantly faster to the penalised maximum likelihood solution than BSREM, particularly in low count data.

Index Terms—Positron Emission Tomography, Image Reconstruction, Stochastic Optimisation, Variance Reduction Algorithms, Relative Difference Penalty

I. INTRODUCTION

PENALISED maximum likelihood image reconstruction algorithms for Positron Emission Tomography (PET) have demonstrated improved desired image properties due to the inclusion of accurate statistical models and *a priori* information for promoting sharp edges and preserving smoothness

This research is supported by: GE Healthcare, the NIHR UCLH Biomedical Research Centre, the EPSRC Centre for Doctoral Training i4Health (EP/S021930/1) and EPSRC project (EP/T000864/1). Software used for this project is part-maintained with funding by CCP SyneRBI (EPSRC grant EP/T026693/1).

R. Twyman, B. Hutton, and K. Thielemans are with the Institute of Nuclear Medicine, UCL, UK (contact: robert.twyman.18@ucl.ac.uk). S. Arridge, B. Jin, and Z. Kereta are with the Department of Computer Science, UCL, UK. L. Brusaferrri was with the Institute of Nuclear Medicine, UCL, supported by Siemens Healthineers, and is now with Athinoula A. Martinos Center for Biomedical Imaging, Harvard Medical School. S. Ahn is with GE Research, Niskayuna, NY 12309 USA. C. W. Stearns and F. Kotasidis are with GE Healthcare, Waukesha, WI 53188 USA. I. A. Burger is with the Nuclear Medicine Department, University Hospital Zurich, University of Zurich.

in uniform regions [1], [2]. To find the (unique) Penalised Maximum Likelihood (PML) solution for PET reconstruction, there are various classes of optimisation algorithms, e.g., preconditioned gradient ascent [3], primal-dual algorithms [4] and surrogate methods [5]. However, these algorithms remain computationally intensive for clinical data as a consequence of the repeated projection operations between the image and measured data spaces. Hence, there has been interest in accelerating these algorithms.

Iterative optimisation algorithms may be accelerated with the use of subsets, also known as blocks or batches, e.g., Ordered Subset Expectation Maximisation (OSEM) [6], Block Sequential Regularized Expectation Maximisation (BSREM) [7] and Ordered Subset Separable Paraboloidal Surrogates (OSSPS) [5] for PET reconstruction. There are several subset methodologies, e.g., list-mode events subsets [8], [9] and projection bin based subsets [6], [10]. In this study, we consider the latter methodology. In these algorithms at each image update, the projection operations are performed only over a fraction (“subset”) of the data set, which greatly reduces the computational cost per update. In current practice, the aforementioned PET algorithms typically cycle through the subsets in a deterministic Ordered Subset (OS) sequence [6], [10]. Initial acceleration for these methods is almost linear with respect to the number of subsets employed, if subsets are reasonably “balanced” [6]. Nevertheless, the variance existing between subset data leads to image estimate voxel value variations between successive updates, which is commonly referred to as limit cycle behaviour [11].

This variance is especially pronounced when large numbers of subsets are utilised. With a constant step size, after many updates, the iterates may only oscillate in a neighbourhood of the PML solution, without convergence. The size of this neighbourhood depends on the step size [6]. While OSEM is generally stopped before a limit cycle is observed, this behaviour is undesirable for PML, and may incur practical consequences in quantitative imaging [12]. Subset algorithms can converge to the PML solution with the use of a relaxed step size sequence [3], [12]. However, in practice the choice of relaxation parameters may greatly impact convergence behaviour and convergence rate may be impeded if poorly selected [1].

We assess the application of two gradient-based Stochastic Variance Reduction (SVR) algorithms to PET reconstruction to

reduce variations in successive image updates and to allow the image to quickly converge to the PML solution under suitable criteria [13], [14]. These algorithms reduce the variance of the gradient estimate by storing and utilising previously computed subset gradients in the determination of the update direction. A number of SVR algorithms have been proposed in recent years, including SAGA [14], Stochastic Variance Reduced Gradient (SVRG) [13], Stochastic Average Gradient (SAG) [15] and Stochastic Recursive Gradient algorithm (SARAH) [16]; see a recent review for an overview of SVR techniques [17]. We elected to use the Relative Difference Penalty (RDP) in this work due to its smoothness and edge preservation properties coupled with its extensive use in clinical evaluation [12], [18], [19]. However, due to the lack of a convex conjugate or a surrogate function, we limit ourselves to gradient-based reconstruction algorithms and forgo comparison with other non-gradient-based algorithms, e.g., stochastic Primal-Dual Hybrid Gradient (sPDHG) [20] or Stochastic Variance Reduced Expectation Maximisation (SVREM) [21].

We aim to demonstrate that SVR optimisation algorithms are at least as fast and stable as the BSREM algorithm when applied to a 3D non-TOF PET objective function regularised using the RDP [2]. We investigate different algorithm implementations using various step sizes, number of subsets and preconditioner configurations. Aside from our own preliminary work [22], [23], to the best of our knowledge, this is the first study of this kind. In Section II, the problem of iterative PET reconstruction is formulated and properties of gradient-based SVR algorithms are explained. Section III describes the experimental setup used to evaluate the reconstruction algorithms. Section IV contains the results of these experiments and Section V provides a discussion of the key findings. The primary conclusions of this research are presented in Section VI.

II. THEORY

A. Penalised Maximum Likelihood PET Reconstruction

PML PET reconstruction is founded on an affine transform that maps the unknown tracer distribution $\mathbf{x} = \{x_i\}_{i=1}^{N_v}$ to measured projection data $\mathbf{y} = \{y_j\}_{j=1}^{N_b}$, where N_v and N_b are the number of image voxels and detector bins, respectively [1]. This is achieved via a system matrix operator $\mathbf{A} = \{a_{ji}\}_{j=1, i=1}^{N_b, N_v}$, where the entry a_{ji} models the probability of an emission from the i^{th} voxel being detected by the j^{th} detector element. Models of detector normalisation and attenuation may be included in \mathbf{A} . The model is complicated by the presence of random and scattered events, which may be accounted for using an additive expected background $\bar{\mathbf{b}} = \{\bar{b}_j\}_{j=1}^{N_b}$. Hence, the acquisition model, which is also known as the forward model, is given by

$$\bar{y}_j(\mathbf{x}) = \sum_{i=1}^{N_v} a_{ji} x_i + \bar{b}_j, \quad j \in \{1, 2, \dots, N_b\}, \quad (1)$$

where $\bar{\mathbf{y}} = \{\bar{y}_j(\mathbf{x})\}_{j=1}^{N_b}$ is the expected data for a tracer distribution given by \mathbf{x} . In PET, a good approximation for the measured data \mathbf{y} is that y_j are statistically independent and follow a Poisson distribution with mean $\bar{y}_j(\mathbf{x})$.

PML methods optimise an objective function:

$$\Phi(\mathbf{x}) = L(\mathbf{x}) - \beta R(\mathbf{x}), \quad (2)$$

where $L(\mathbf{x})$ measures the goodness of fit with the measured data \mathbf{y} , $R(\mathbf{x})$ imposes a penalty on the image and $\beta \geq 0$ is a global scale factor between the two functions. Given the Poisson assumption, it is natural to use the Poisson log-likelihood function $L(\mathbf{x})$, given by

$$L(\mathbf{x}) = \sum_{j=1}^{N_b} y_j \log(\bar{y}_j(\mathbf{x})) - \bar{y}_j(\mathbf{x}),$$

as the goodness of fit. The function is concave with respect to $\bar{y}_j(\mathbf{x})$ and \mathbf{x} .

PET PML methods typically optimise the objective function $\Phi(\mathbf{x})$ with the tracer distribution \mathbf{x} constrained to be a non-negative vector. For most common penalties $R(\mathbf{x})$, the objective function $\Phi(\mathbf{x}) := L(\mathbf{x}) - \beta R(\mathbf{x})$, $\beta \geq 0$, has a unique maximiser, i.e., the PML solution $\hat{\mathbf{x}}$, under mild conditions [11]. Gradient based algorithms aim to converge to the PML solution $\hat{\mathbf{x}}$ by iteratively adding a weighted objective function gradient $\nabla\Phi(\mathbf{x}_k) = \nabla L(\mathbf{x}_k) - \beta \nabla R(\mathbf{x}_k)$ to the current estimate \mathbf{x}_k , and often take the form

$$\mathbf{x}_{k+1} = P_{\mathbf{x} \geq 0}[\mathbf{x}_k + \alpha_k \mathbf{D}(\mathbf{x}_k) \nabla\Phi(\mathbf{x}_k)], \quad (3)$$

where k is the update number index, $P_{\mathbf{x} \geq 0}[\cdot]$ is a voxel-wise non-negativity projection operation, $\alpha_k > 0$ is a scalar step size and $\mathbf{D}(\mathbf{x}_k) \in \mathbb{R}^{N_v \times N_v}$ is a strictly positive definite preconditioner. One example algorithm of this form is Maximum Likelihood (ML) Expectation Maximisation (EM) (MLEM) [24] written in the additive form [25], when $\alpha_k = 1$, $\beta = 0$ and $\mathbf{D}(\mathbf{x}_k) = \text{diag}[\mathbf{x}_k / \mathbf{A}^T \mathbf{1}]$, where the operator $\text{diag}[\cdot]$ constructs a diagonal matrix from a vector and the vector division is element-wise [26].

B. Subset Methods

The computation of the log-likelihood gradient $\nabla L(\mathbf{x})$ generally dominates the computational effort for iterative methods, even when utilising an efficient vectorised implementation [1]. This is due to the application of the forward and backward projection operators, \mathbf{A} and \mathbf{A}^T respectively. Image reconstruction may be accelerated by evaluating only sub-likelihood gradients at each update. This is achieved by projecting into and from a subset of the data. The measured projection data \mathbf{y} may be divided into a set of M subsets $\{\mathcal{S}_m\}_{m=1}^M$ that are pairwise disjoint, that is, $\mathcal{S}_m \cap \mathcal{S}_n = \emptyset, \forall m \neq n$, and complete, that is, $\bigcup_{m=1}^M \mathcal{S}_m = \{1, 2, \dots, N_b\}$.

The objective function $\Phi(\mathbf{x})$ may be written as a sum of sub-objective functions $\Phi_m(\mathbf{x})$ as

$$\Phi(\mathbf{x}) = \sum_{m=1}^M \Phi_m(\mathbf{x}) = \sum_{m=1}^M \left(L_m(\mathbf{x}) - \frac{\beta}{M} R(\mathbf{x}) \right), \quad (4)$$

where L_m is the log-likelihood of the m^{th} subset. The gradients are given by

$$\nabla\Phi(\mathbf{x}) = \sum_{m=1}^M \nabla\Phi_m(\mathbf{x}) = \sum_{m=1}^M \left(\nabla L_m(\mathbf{x}) - \frac{\beta}{M} \nabla R(\mathbf{x}) \right). \quad (5)$$

The additive preconditioned subset update may be recast as

$$\mathbf{x}_{k+1} = P_{\mathbf{x} \geq 0} [\mathbf{x}_k + \alpha_k M \mathbf{D}_m(\mathbf{x}_k) \nabla \Phi_m(\mathbf{x}_k)], \quad (6)$$

where $\mathbf{D}_m(\mathbf{x}_k)$ is a preconditioner of the m^{th} subset. Many PET reconstruction algorithms utilise this format, e.g., (additive) OSEM [6], BSREM [7], [11], OSSPS [5] and Row-Action ML Algorithm (RAMLA) [3], each of which differ by the choice of $\mathbf{D}_m(\mathbf{x})$ and α_k . These algorithms generally use a deterministic cyclical sequence to select the subset index m at each image update [6], [10].

Computing only a subset gradient at each update leads to almost linear acceleration during early iterations if the subsets are balanced [6], [11]. The subset balancing amounts to

$$\nabla \Phi_1(\mathbf{x}_k) \approx \nabla \Phi_2(\mathbf{x}_k) \approx \dots \approx \nabla \Phi_M(\mathbf{x}_k). \quad (7)$$

The diagonal preconditioner may be used to further balance the weighted subset gradients, e.g., OSEM's subset dependant EM preconditioners. If the subsets and preconditioners are chosen properly, the subset balancing condition holds when \mathbf{x}_k is far from the solution $\hat{\mathbf{x}}$. However, the assumption of subset balance does not generally hold due to various physical factors, e.g., projection angle geometry, attenuation and data noise. Discrepancies between subset data will lead to variations between subset functions and gradients. Thus, while a subset reconstruction algorithm efficiently optimises $\Phi(\mathbf{x})$ when the iterate \mathbf{x}_k is far from the solution $\hat{\mathbf{x}}$, after a number of passes through the data (epochs), the performance generally suffers when approaching $\hat{\mathbf{x}}$. This is exacerbated when large numbers of subsets are utilised.

As an alternative to deterministic cyclic subset selection, stochastic algorithms select m randomly at each update, usually with uniform probability. Stochastic Gradient Ascent (SGA) is a popular stochastic algorithm used to optimise a finite sum of separable functions, in the form of (4), using an update equation comparable to (6) with $\mathbf{D}(\mathbf{x}) = \mathbf{I} \in \mathbb{R}^{N_v \times N_v}$, the identity matrix [27]. This results in an unbiased estimate of $\Phi(\mathbf{x})$ and $\nabla \Phi(\mathbf{x})$, i.e., $\mathbb{E}[M \Phi_m(\mathbf{x})] = \Phi(\mathbf{x})$ and $\mathbb{E}[M \nabla \Phi_m(\mathbf{x})] = \nabla \Phi(\mathbf{x})$, where the expectation is taken with respect to the random subset index m [27]. Note the inclusion of M in these expectations and (6). These are included to allow for comparable gradient magnitudes between algorithms implemented with varying numbers of subsets.

C. Step Size Relaxation

Subset methods, in the form of (6), may converge to the PML solution $\hat{\mathbf{x}}$ if a relaxed step size sequence $\{\alpha_k\}_{k=0}^{\infty}$ is applied [3], [7], [11]. A common relaxation method is to diminish step sizes over the iterations k so that $\lim_{k \rightarrow \infty} \alpha_k = 0$. The sufficient conditions on relaxation sequences for global convergence are given by [3], [7], [28]

$$\sum_{k=0}^{\infty} \alpha_k = \infty \quad \text{and} \quad \sum_{k=0}^{\infty} \alpha_k^2 < \infty. \quad (8)$$

Deterministic algorithms, e.g., BSREM [3] and RAMLA [7], utilise this methodology. However, the selection of hyperparameters for the relaxation sequence can greatly impact reconstruction performance.

D. Stochastic Variance Reduction Algorithms

Variance between subset gradients may impede reconstruction performance as \mathbf{x}_k approaches $\hat{\mathbf{x}}$. Recently, a novel class of stochastic variance reduction techniques have been proposed to combat this. In this work, we investigated the application of SAGA (Algorithm 1) [14] and SVRG (Algorithm 2) [13] to PET image reconstruction. These algorithms approximate the full gradient $\nabla \Phi(\mathbf{x}_k)$ with a low computational cost gradient estimate $\tilde{\nabla}_{k,m}(\mathbf{x}_k)$, but with a lower variance than the (randomly selected) subset gradient $\nabla \Phi_m(\mathbf{x}_k)$ [17]. This is achieved by computing $\nabla \Phi_m(\mathbf{x}_k)$ for a randomly selected subset index m at each update and utilising M previously computed subset gradients \mathbf{g}_m stored in memory. Under certain convexity assumptions, SVR methods converge linearly in expectation where as traditional subset algorithms, e.g., (6), are sub-linear [17].

Generally, SVR algorithms formulate the update direction $\tilde{\nabla}_{k,m_k}(\mathbf{x}_k)$ as

$$\tilde{\nabla}_{k,m_k}(\mathbf{x}_k) = \xi \left(\nabla \Phi_{m_k}(\mathbf{x}_k) - \mathbf{g}_{m_k} \right) + \sum_{m=1}^M \mathbf{g}_m, \quad (9)$$

where $\xi = M$ for SAGA and SVRG [29], which results in an unbiased estimator of the gradient, i.e., $\nabla \Phi(\mathbf{x}_k) = \mathbb{E}[\tilde{\nabla}_{k,m}(\mathbf{x}_k)]$ [17]. Different choices for ξ result in a biased estimator, e.g., $\xi = 1$ for SAG [15].

Algorithm 1 SAGA algorithm

Given $\mathbf{x}_0 \in \mathbb{R}^{N_v}$, M , K and $\{\alpha_k > 0\}_{k=0}^K$
 Store $\mathbf{g}_m = \mathbf{0}$, $\forall m \in \{1, 2, \dots, M\}$
for $k = 0, 1, \dots, K$ **do**
 Choose m_k uniformly in $\{1, 2, \dots, M\}$
 Compute $\nabla \Phi_{m_k}(\mathbf{x}_k)$
 $\tilde{\nabla}_{k,m_k}^{\text{SAGA}}(\mathbf{x}_k) \leftarrow M \left(\nabla \Phi_{m_k}(\mathbf{x}_k) - \mathbf{g}_{m_k} \right) + \sum_{m=1}^M \mathbf{g}_m$
 Store $\mathbf{g}_{m_k} \leftarrow \nabla \Phi_{m_k}(\mathbf{x}_k)$
 $\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k + \alpha_k \tilde{\nabla}_{k,m_k}^{\text{SAGA}}(\mathbf{x}_k)$ # Update step
end for

Compared with SAGA, SVRG does not store the subset gradient $\mathbf{g}_{m_k} \leftarrow \nabla \Phi_{m_k}(\mathbf{x}_k)$ at each iteration and instead periodically updates each of the \mathbf{g}_m terms from an anchor image $\tilde{\mathbf{x}}$. This periodic computation occurs every γ epochs, i.e., γ passes over the entire data set. For convex optimisation problems, $\gamma = 2$ is heuristically suggested [13]. This periodic computation of M \mathbf{g}_m terms is equivalent to the computation of the full gradient (i.e., an epoch) with a single image update applied. This allows for the evaluation of the objective function without additional projection operations and, to the best of our knowledge, this is a unique feature for a subset algorithm.

Note, the M factor in (9) is not present in the works of [13] and [14] because the objective functions in the said works are formulated as $\Phi(\mathbf{x}) = \frac{1}{M} \sum_{m=1}^M \Phi_m(\mathbf{x})$, whereas this work uses (4).

E. Adaptation to PET

The PET problem tends to be ill-conditioned. Thus, direct application of the aforementioned standard SAGA or SVRG

Algorithm 2 SVRG algorithm

Given $\mathbf{x}_0 \in \mathbb{R}^{N_v}$, M , K , $\{\alpha_k > 0\}_{k=0}^K$ and $\gamma > 0$

for $k = 0, 1, \dots, K$ **do**

if $k \bmod \gamma M \equiv 0$ **then**

$\tilde{\mathbf{x}} \leftarrow \mathbf{x}_k$

for $m \in \{1, \dots, M\}$ **do** # Loop over all subsets

 Compute $\nabla \Phi_m(\tilde{\mathbf{x}})$

 Store $\mathbf{g}_m \leftarrow \nabla \Phi_m(\tilde{\mathbf{x}})$

end for

$\tilde{\nabla}_{k,m_k}^{\text{SVRG}}(\mathbf{x}_k) \leftarrow \sum_{m=1}^M \mathbf{g}_m$

else

 Choose m_k uniformly in $\{1, 2, \dots, M\}$

 Compute $\nabla \Phi_{m_k}(\mathbf{x}_k)$

$\tilde{\nabla}_{k,m_k}^{\text{SVRG}}(\mathbf{x}_k) \leftarrow M(\nabla \Phi_{m_k}(\mathbf{x}_k) - \mathbf{g}_{m_k}) + \sum_{m=1}^M \mathbf{g}_m$

end if

$\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k + \alpha_k \tilde{\nabla}_{k,m_k}^{\text{SVRG}}(\mathbf{x}_k)$ # Update step

end for

algorithms may result in slow reconstructions. We replace the update step of Algorithms 1 and 2 with a preconditioned update, given by

$$\mathbf{x}_{k+1} = P_{\mathbf{x} \geq 0}[\mathbf{x}_k + \alpha_k \mathbf{D}_k \tilde{\nabla}_{k,m}(\mathbf{x}_k)], \quad (10)$$

where the positive-definite diagonal preconditioner $\mathbf{D}_k \in \mathbb{R}^{N_v \times N_v}$ can depend on k . Furthermore, this modification also includes the projection operation onto non-negative vectors and thus the optimisation problem is constrained [14]. The resulting update closely resembles (6). We will denote these preconditioned algorithms with their original abbreviations (SAGA and SVRG) in the remainder of this paper.

There are several possible choices for the preconditioner \mathbf{D}_k and step size α_k , which can lead to different convergence behaviour. The simplest choice is to set the preconditioner \mathbf{D}_k to be constant, diagonal and positive definite, and the step size constant as well. Then SAGA and SVRG will converge almost surely to the PML solution $\hat{\mathbf{x}}$ provided that the step size α is upper bounded by a critical value $\alpha_{\mathcal{L}_D}$, which depends on the global Lipschitz constant $\mathcal{L}_D > 0$ (that is assumed to exist) of the preconditioned system [21, Theorem 4.1]¹. The convergence result remains valid if the preconditioner and step size are fixed after a finite number of epochs.

Due to the challenge of determining the optimal step size *a priori*, we also investigated a diminishing step size sequence of the form (8) for these stochastic algorithms. Building upon the convergence proof of [11, III.B], for a diagonally scaled incremental gradient method of the type (6), we expect that the aforementioned almost sure convergence remains valid for relaxed step size sequences satisfying (8) and a constant preconditioner, although a complete proof is still unavailable. Nonetheless, the experiments below confirm the convergence.

¹A modified likelihood function is required if $\bar{b}_j = 0$ for any j [21], but 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 upon $\mathcal{B} \triangleq \{\mathbf{x} \in \mathbb{R}^p : 0 \leq x_i \leq U, \forall i\}$ (where U is an upper bound [11, appx A]) and $\Phi_m(\mathbf{x})$ to be concave over this bounded region.

III. NUMERICAL EXPERIMENTS

A. Synthetic Data Generation and System Modelling

To assess the applicability of SVR algorithms for PET reconstruction, a numerical OpenGATE (GATE) simulation [30] of the GE Discovery 690 scanner [31], with uniform crystal efficiencies, was performed via the STIR-GATE-Connection [32]. A photon emission simulation was performed using back-to-back 511 keV photon emissions from a voxelised XCAT torso phantom [33] with activity concentrations representative of an ¹⁸F-FDG study (see emission and attenuation distributions in Fig. 1a and Fig. 1b, respectively). A 1 cm diameter, 1 cm long, cylindrical hot lesion, with 2.5:1 lesion to lung contrast, was inserted into the lung of the XCAT emission. Cardiac and respiratory motion, along with radioactive decay, were not modelled. The resulting list mode data were re-binned into sinograms with 288 projection angles.

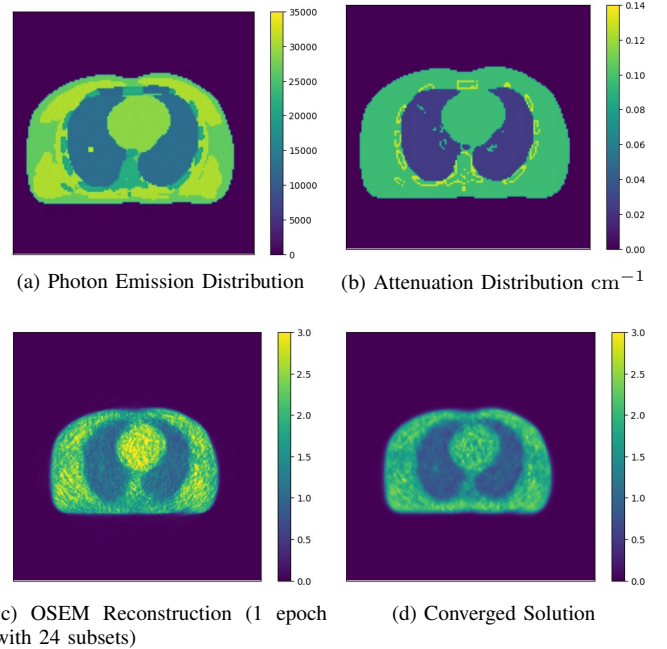


Fig. 1: Transaxial slices of various distributions for the 50 million event simulated XCAT data set. (a) Emission and (b) attenuation consisted of $(141, 141, 47)$ voxels of size $(3, 3, 3.27)$ mm. Reconstructed images (c) and (d) contained $(251, 251, 47)$ voxels of size $(2.13, 2.13, 3.27)$ mm.

Normalisation factors are required for an accurate system model. A cylindrical phantom, the size of the scanner’s Field of View (FOV), was forward projected in both GATE and Software for Tomographic Image Reconstruction (STIR) [34] without attenuation using ray-tracing with 10 rays per bin. The normalisation was computed, as per component-based normalisation factors with crystal efficiencies and “geometric” effects, using maximum likelihood computation with STIR-GATE-Connection tools [32], [35]. However, as the GATE simulation did not contain any variability between blocks, the crystal efficiency aspect of the normalisation may be ignored.

An estimation of random coincidence events was also made with STIR-GATE-Connection tools. Singles were estimated

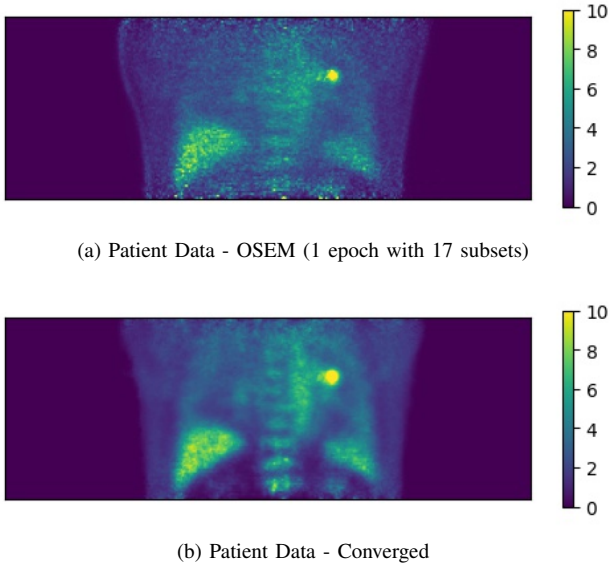


Fig. 2: Coronal slices of the patient data set reconstruction with a bronchial carcinoma (arrow) in the left upper lobe (kBq/mL). Reconstructed images consist of $(323, 323, 89)$ voxels of size $(2.21, 2.21, 2.76)$ mm.

from the delayed coincidence events (recorded by GATE) using a maximum likelihood algorithm and then multiplied to find the randoms rates [36]. However, we observed a discrepancy between the total number of delayed coincidence events and the true number of random coincidence events. Therefore, the estimated randoms were globally scaled to match the true randoms count [32].

An estimation of the scatter contribution was similarly computed using STIR's iterative single scatter simulation and estimation utilities [34], [37]. Image reconstructions were performed using custom reconstruction software, based on STIR.

To assess reconstruction algorithm performance, the list mode file was sampled to acquire sinogram data with 50, 250, and 1200 million events with a true-to-background ratio of 0.93:1, where background is random plus scattered events. This resulted in data sets with low to high Signal-to-Noise Ratio (SNR), respectively. The aforementioned data corrections were computed individually for these data sets. Attenuation factors and normalisation factors were incorporated within the system matrix \mathbf{A} .

B. Patient Data

A 150 second ^{18}F -FDG static patient scan was acquired on a 5 ring GE Discovery MI (272 projection angles) [38] and 143 million events were recorded. Non Time Of Flight (TOF) quantitative data corrections were computed using vendor-provided procedures but image reconstructions were performed similarly to the aforementioned simulated data. 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 patient was diagnosed with a bronchial

carcinoma in the left upper lobe (Fig. 2).

C. Subset Methodology

Equally spaced rows of a sinogram were binned into a subset. Subsequent subsets were constructed similarly but were offset from one another by their respective subset index m . BSREM and OSEM utilise the cyclical and deterministic subset sampling methodology, as discussed in Section II-B. The stochastic algorithms select m uniformly at random at each update.

D. Algorithm Warm Starting

The proposed SVR algorithms have been observed to be sensitive to initial conditions [39]. This is because the initially stored subset gradients, which are computed when \mathbf{x}_k is far from the PML solution $\hat{\mathbf{x}}$, are not representative of the subset gradients when \mathbf{x}_k approaches $\hat{\mathbf{x}}$. These gradients are retained in memory and can lead to undesirable update directions once the image has begun to resolve. Therefore, we initialised the investigated algorithms from an image computed by one epoch of OSEM (\mathbf{x}_{OSEM}). A 24 subset OSEM was performed for the simulated data (Fig. 1c) while 17 subsets were used for the patient data (Fig. 2a).

E. Penalty

Ideally, the penalty function $R(\mathbf{x})$ in (2) reduces the impact of noise in the data while promoting structural edges [1]. Due to its use in modern clinical PET scanners, we implemented a spatially-variable [40] RDP [2] to achieve this, given by

$$R(\mathbf{x}) = \sum_{i=1}^{N_v} \sum_{l \in \mathcal{N}_i} \sqrt{\kappa_i \kappa_l} \frac{(x_i - x_l)^2}{(x_i + x_l) + \omega |x_i - x_l|}, \quad (11)$$

where \mathcal{N}_i is a $3 \times 3 \times 3$ neighbourhood about the i^{th} image voxel, and $\omega = 2$ is a parameter used to control edge preservation. κ_i and κ_l are spatially variable penalty strengths with the image $\boldsymbol{\kappa} = \{\kappa_i\}_i^{N_v}$ computed by

$$\boldsymbol{\kappa} = \sqrt{\mathbf{A}^\top \text{diag} \left[\frac{\mathbf{y}}{\bar{\mathbf{y}}(\mathbf{x}_{\text{OSEM}})^2} \right] \mathbf{A} \mathbf{1}}. \quad (12)$$

The computation of $\boldsymbol{\kappa}$ requires two forward projections and one backward projection but this is ignored in our analysis. The RDP encourages smooth cold background (low activity) regions and can improve lesion detectability when compared with regularisers that are less edge-preserving, such as the quadratic penalty [12], [18].

F. Diagonal Preconditioner

The EM preconditioner $\mathbf{D}(\mathbf{x}_k) = \text{diag}[\mathbf{x}_k / \mathbf{A}^\top \mathbf{1}]$ is used in the MLEM and BSREM algorithms. However, this preconditioner is not positive definite and prevents a voxel with zero value ($x_i = 0$) from updating, regardless of the gradient's value. The authors of [11] proposed a modification to the update equation to prevent this occurrence by projecting

$x_i \leq 0$ to a small positive value. In this work, we implemented a different modification of the EM preconditioner, given by

$$D(\mathbf{x}) = \text{diag} \left[\frac{\mathbf{x} + \boldsymbol{\delta}}{\mathbf{A}^T \mathbf{1}} \right], \quad (13)$$

where $\boldsymbol{\delta}$ is a uniform image of heuristic pixel value 10^{-6} , which is included to allow voxels with value 0 to be updated. The $\boldsymbol{\delta}$ is small enough that it does not significantly impact reconstruction performance in this work.

We investigated the impact of different preconditioner inputs on the convergence of SAGA and SVRG. These included $D(\mathbf{x}_k)$, freezing the preconditioner at $D(\mathbf{x}_{\text{OSEM}})$, and freezing the preconditioner after 5 and 10 epochs, $D(\mathbf{x}_{5M})$ and $D(\mathbf{x}_{10M})$ respectively.

G. Step Sizes

As per Section II-E, SAGA and SVRG are guaranteed to converge for constant step sizes (if small enough) but the sharp upper bound on the step size magnitude is difficult to compute. Regardless, we investigated the application of $\alpha_k = 1$ with preconditioner given by (13), a configuration that has been explored in previous studies [22], [23]. We additionally investigated the application of a relaxed step size sequence that satisfies the conditions stated in (8), given by

$$\alpha_k = \frac{1}{\frac{\eta}{M}k + 1}, \quad (14)$$

where the relaxation factor η is scaled by the number of subsets M so that step sizes of reconstructions, using different numbers of subsets, are comparable at similar epochs.

H. Converged Comparison and Performance Metrics

In PML, it is assumed that the penalty term $R(\mathbf{x})$ encourages local smoothness and edge preservation so the PML solution $\hat{\mathbf{x}}$ is desirable [1]. Therefore, we assessed reconstruction performance by comparing each algorithm updates \mathbf{x}_k to the converged image $\hat{\mathbf{x}}$, which was obtained by reconstructing the data using (10) with $\tilde{\nabla}_{k,m_k}^{\text{SVRG}}$ and the modified-EM preconditioner (13). The reconstruction algorithm's step size α_k was reduced if $\Phi(\mathbf{x}_k) \leq \Phi(\mathbf{x}_{(k-\gamma M)})$ when measured periodically at each epoch.

The algorithm was terminated and $\hat{\mathbf{x}} = \mathbf{x}_k$ when the estimate satisfied the Karush-Kuhn-Tucker conditions (within numerical precision) [41].

The following metrics were employed to assess the reconstruction performance. A Region Of Interest (ROI) was drawn exactly over the XCAT data's inserted lung lesion. Similarly, a $(20 \times 26 \times 20)$ mm³ ROI was drawn over the bronchial carcinoma and Standardized Uptake Value peak (SUV_{peak}) values [42] were computed. The ROI contains the entire carcinoma hot spot. Furthermore, a $(50 \times 60 \times 60)$ mm³ ellipsoidal ROI was drawn within the patient's liver with a minimum 10mm distance was maintained from the organ's boundary. ROI percentage error was computed for each update k , with respect to $\hat{\mathbf{x}}$, given by

$$\text{ROI Percentage Error}_k = 100\% \cdot \frac{\theta_k - \hat{\theta}}{\hat{\theta}}, \quad (15)$$

where θ_k and $\hat{\theta}$ are mean ROI voxel values for the k^{th} image estimate and converged image, respectively. Furthermore, the objective value $\Phi(\mathbf{x})$ was computed after every epoch of the reconstructions.

Unlike deterministic algorithms, SVRG and SAGA generate additional noise in the reconstruction performance metrics due to stochastic subset sampling at each update. Thus, a set of 15 independent reconstructions S were run for each stochastic algorithm configuration. The mean and standard deviation values of each metric were computed over the stochastic realisations. Additionally, the general convergence performance of these algorithms was assessed using a percentage Normalised Root Mean Squared Error (NRMSE) metric at each update k , given by

$$\text{NRMSE}\%_k = 100\% \cdot \frac{\sqrt{\frac{1}{|S|} \sum_{s \in S} (\theta_{s,k} - \hat{\theta})^2}}{\hat{\theta}}. \quad (16)$$

IV. RESULTS

Sections IV-A to IV-C, we evaluate the impact of the number of subsets, various preconditioner configurations and various step size relaxation parameters with the SAGA and SVRG gradient estimators. Reconstructions in these sub-sections utilise the noisy XCAT data with 50 million prompt events. The impact of noise levels on SAGA's and SVRG's performance is investigated in Sections IV-D and IV-E demonstrates the algorithms' reconstruction performance when applied to a patient data.

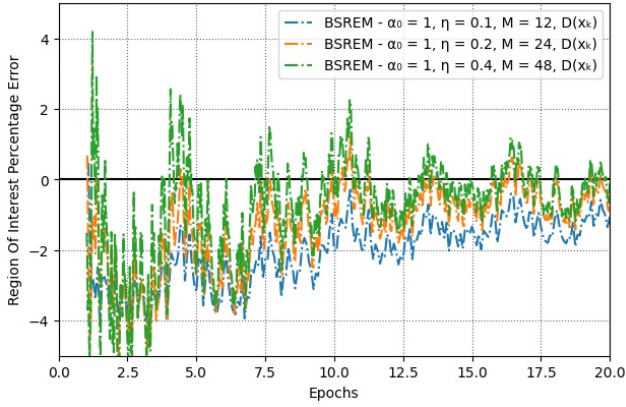
A. Varying the Number of Subsets

Standard PET subset algorithms (e.g., BSREM) are linearly accelerated but, at later iterations, limit cycle behaviour may be observed in the reconstruction sequence. An example of these variations is exhibited by the BSREM profile in Fig. 3a, even with step size relaxation.

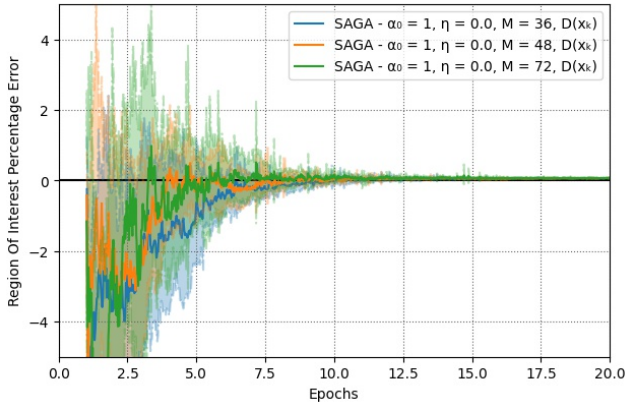
SVRG reconstructions using various numbers of subsets, the $D(\mathbf{x}_k)$ preconditioner (13), and no step size relaxation are shown in Fig. 3c. Note, SVRG's full gradient re-computation parameter γ was fixed at $\gamma = 2$. The usage of fewer subsets resulted in slower convergence of the mean performance but the standard deviations were greatly reduced. The mean values, for each of the configurations, converged to 0% lung lesion error with the deviations reducing dramatically at later epochs. SAGA reconstructions exhibited similar behaviour to SVRG, albeit with larger deviations. Based upon these results, Figs. 4, 5, 6 and 7 plot the performance of SAGA and SVRG using 72 subsets, which demonstrated a reasonable trade-off between convergence rate and stochastic variation.

B. Constant Preconditioner and Step Size

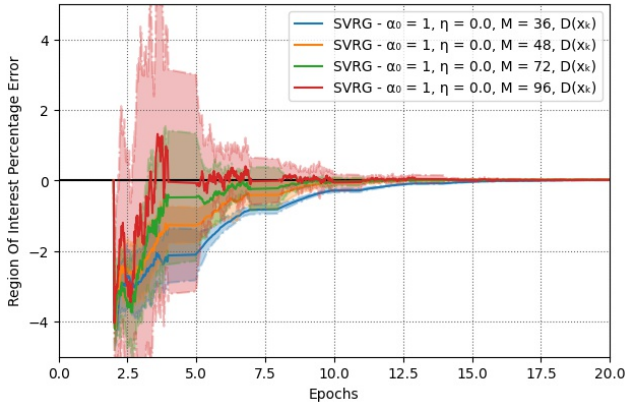
In Section IV-A, the SVR algorithms were implemented with the diagonal preconditioner $D(\mathbf{x}_k)$, which was updated at each iteration k . In Fig. 4 we investigate the impact of anchoring input \mathbf{x} for the preconditioner $D(\mathbf{x})$ after various periods of computation by plotting the NRMSE%.



(a) BSREM



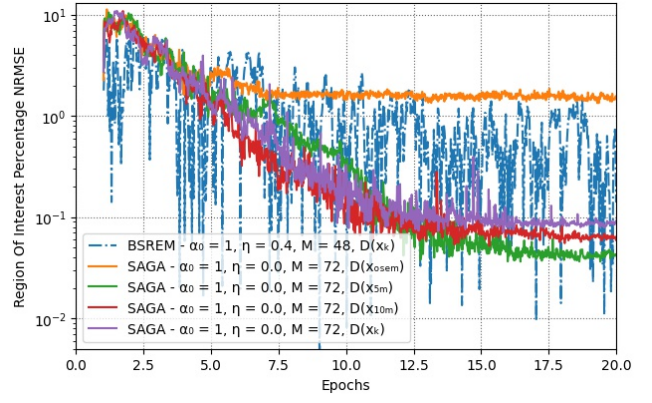
(b) SAGA



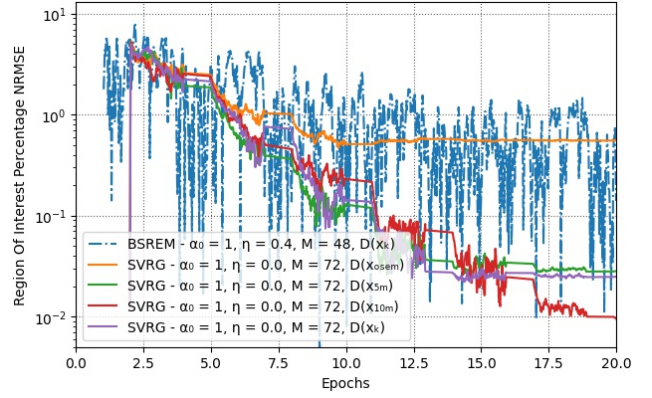
(c) SVRG

Fig. 3: The lung lesion percentage error of (a) BSREM, (b) SAGA and (c) SVRG reconstructions using various numbers of subsets. The implemented preconditioner is $D(x_k)$. 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 ($\eta = 0$) with 15 stochastic realisations of each configuration. The 96 subset SAGA profile is not included for figure clarity.

The SVR reconstruction NRMSE% performances are comparable for the $D(x_k)$, $D(x_{5M})$ and $D(x_{10M})$ precondition-



(a) SAGA



(b) SVRG

Fig. 4: The lung lesion percentage NRMSE over multiple stochastic realisations with preconditioners anchored at different epochs. The BSREM NRMSE value was computed for a single deterministic realisation with (16).

ers. However, the $D(x_{OSEM})$ lesion values do not converge to the solution \hat{x} . SVRG appears to converge closer to the PML solution \hat{x} than SAGA for similar configurations.

In Fig. 5 we plot ROI values of a region outside of the torso. The voxels in this region do not converge to zero when the preconditioner $D(x_{OSEM})$ is implemented without step size relaxation.

C. Relaxed Step Size and Constant Preconditioner

The global convergence of the SVR algorithms is not guaranteed when using a constant step size that is too large. Here we investigate the impact of using step size relaxation when a poor performing anchored preconditioner is utilised.

Fig. 6 indicates that the lesion ROI value converges to the PML solution \hat{x} with relatively little relaxation, even when the preconditioner is fixed at an un-optimised one.

D. Data Noise

Impact of data noise on the performance of the investigated algorithms is shown in Fig. 7. The stochastic algorithms were implemented with a combination of previously observed advantageous configurations. The SVRG reconstructions demonstrate similar performance when optimising $\Phi(x)$, regardless

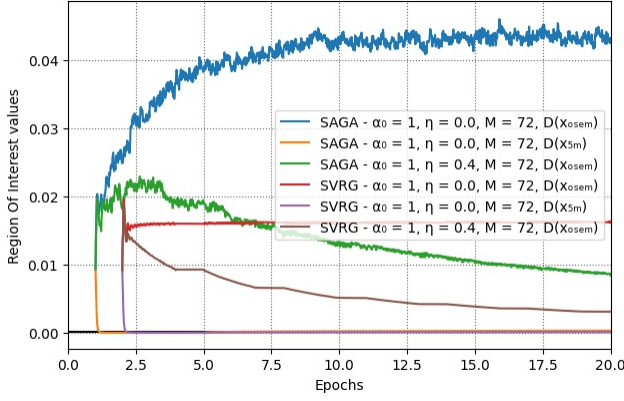
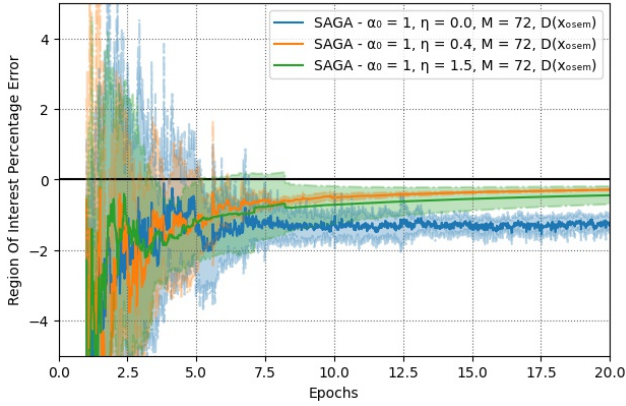
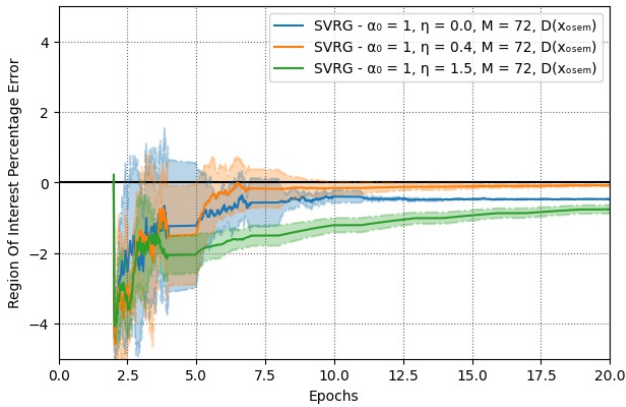


Fig. 5: Mean XCAT reconstruction (50 million event) image background region values across multiple stochastic realisations of SAGA and SVRG with preconditioners anchored at different epochs and relaxation parameters $\eta = 0.0$ and 0.4 .



(a) SAGA

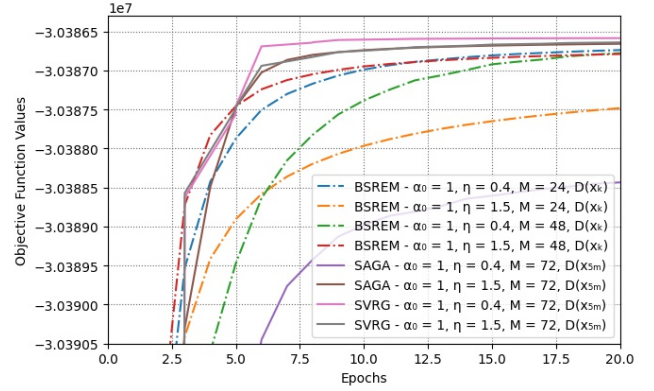


(b) SVRG

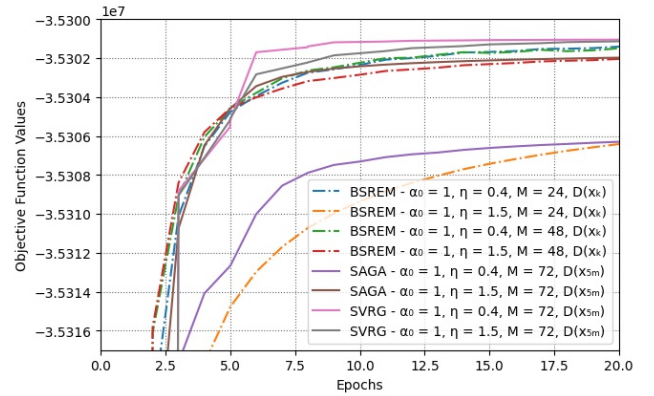
Fig. 6: The XCAT 50 million event data set lung lesion percentage error of (a) SAGA and (b) SVRG reconstructions using various step size relaxation factors η , 72 subsets and the $D(x_{OSEM})$ preconditioner.

of step size relaxation, for all data sets. The SAGA reconstruction profiles with $\eta = 1.5$ result in performance comparable to that of SVRG but the algorithm is sensitive to the selected

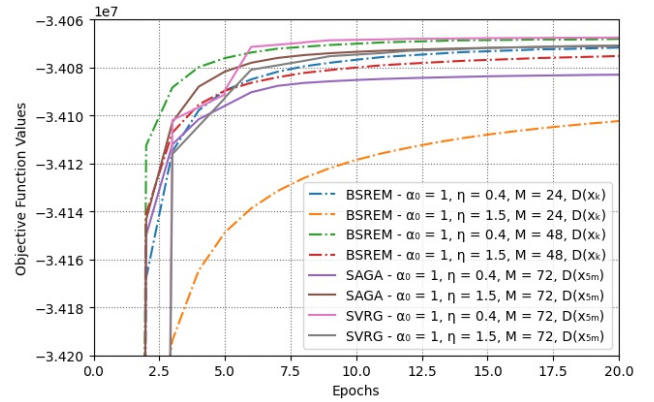
step size parameters.



(a) 50 Million Events



(b) 250 Million Events



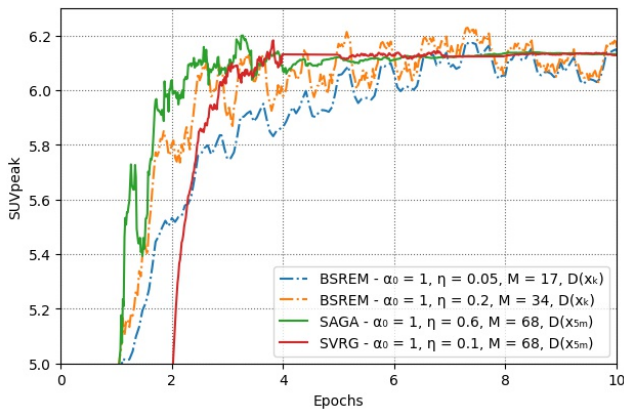
(c) 1200 Million Events

Fig. 7: Mean objective function evaluations of 15 stochastic realisations measured at every epoch of BSREM, SAGA and SVRG reconstructions for XCAT data sets with (a) 50, (b) 250 and (c) 1200 million prompt events. SVR algorithms utilised a $D(x_{5M})$ as a preconditioner and 72 subsets. Various step size methodologies are demonstrated. Standard deviation markers are removed for clarity.

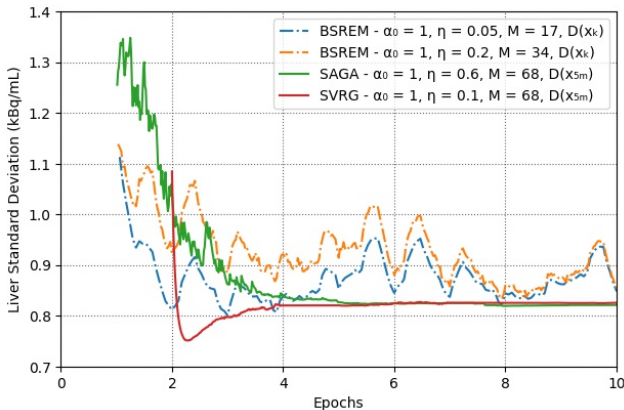
E. Patient Data

Multiple reconstructions of the selected patient data set were performed with various step size relaxation parameters η . AI-

gorithm configuration performance was evaluated in terms of the average MSE obtained over the first 10 epochs ($10M$), i.e., $\frac{1}{M} \sum_{k=0}^{10M} (\theta_k - \hat{\theta})^2$, where θ is the mean bronchial carcinoma ROI value for this data. This was done in an attempt to identify optimal η values with fast SUV convergence within 10 epochs. Using this grid search, the best performing BSREM (17 and 34 subsets) and SAGA and SVRG (68 subsets) reconstructions are plotted in Fig. 8. The mean bronchial carcinoma values (Fig. 8a) 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 7.82 kBq/mL 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 0.83 kBq/mL, see Fig. 8b.



(a) Bronchial Carcinoma



(b) Liver Standard Deviation

Fig. 8: Reconstruction performance of the patient data for BSREM, SAGA and SVRG reconstruction algorithms. (a) Bronchial Carcinoma ROI-mean profile and (b) ROI standard deviation in the patient’s liver.

V. DISCUSSION

A. Impact of Number of Subsets

As observed in Fig. 3, increasing the number of subsets of the SAGA and SVRG algorithms results in greater variations in metric performance between stochastic realisations during

early epochs. However, the magnitude of these variations are still comparable to the BSREM reconstruction inter-update variations, which utilised a step size relaxation sequence, while SAGA and SVRG used a constant $\alpha_k = 1$. SVRG demonstrated reduced variations between stochastic realisations but does exhibit slower convergence. This is likely due to periodically recomputing the \mathbf{g} terms from $\tilde{\mathbf{x}}$ and keeping much of the $\tilde{\nabla}_{k,m}(\mathbf{x}_k)$ computation constant, improving stability, but incurring the periodic additional computational cost.

The stochastic variance observed during early epochs of Fig. 3 and Fig. 4 appears greater for SAGA than SVRG. For SAGA, although it is expected that a subset index $m \in \{1, 2, \dots, M\}$ is selected once every epoch, the expected number of epochs required to select each m once grows as $\mathcal{O}(\log(M))$ [43, Section 4]. Hence, a number of SAGA’s stored subset gradients may not have been computed from recent iterates. Thus, for some m , $\mathbf{g}_m \not\approx \nabla\Phi_m(\mathbf{x}_k)$. This will result in increased variance between stochastic realisations. This issue is exacerbated by SAGA’s initialisation of $\mathbf{g}_m = \mathbf{0}$, $\forall m$ because the variance reduction properties of these algorithms are obtained when \mathbf{g}_m approaches $\nabla\Phi_m(\mathbf{x}_k)$ as k grows [17]. In contrast, SVRG recomputes all \mathbf{g}_m terms periodically and therefore $\mathbf{g}_m \approx \nabla\Phi_m(\mathbf{x}_k)$.

The 72 subset SAGA and SVRG reconstructions demonstrated a good balance between faster initial performance than BSREM and reduced stochastic influence after about 5 epochs (Fig. 3). fast initial performance (in mean) and reduced stochastic influence after about 5 epochs. This evidence suggests that while a larger number of subsets can be used for with SAGA and SVRG than BSREM, too great a number may result in issues, particularly for SAGA.

B. Convergence

In order to ensure convergence to the PML solution, we applied a modified version of the EM preconditioner, frozen at different epochs and investigated several step size relaxation parameters η . We observed a systematic error in the later epochs of reconstructions utilising $\mathbf{D}(\mathbf{x}_{\text{OSEM}})$ without step size relaxation, both in the lung lesion (Figs. 4a, 6b and 6a) and the region surrounding the patient (Fig. 5). We hypothesise that is related to a step size that is too large coupled with the non-negativity constraint. A large step size induces oscillations between updates of the voxel values. However, in the background region this leads to a positive overall bias due to the truncation to non-negative values. This positive bias then leads to a (small) negative bias inside the torso. Our results indicate that this is resolved by using a diminishing step size sequence (e.g., (14)) as well as by deferring the anchoring of the preconditioner until later epochs (e.g., $\mathbf{D}(\mathbf{x}_{5e})$). Note that the latter achieves a smaller effective step size in the background, reducing the need for relaxation. The implementation of a delayed anchoring of the preconditioner might therefore allow for significantly improved performance. Our experiments indicate that the implementation of a moderately relaxed step size sequence, along with an anchored preconditioner, allows for fast and (almost sure) convergence, but this remains to be theoretically confirmed. We note that these observations

are specific to EM-like preconditioners. Performance of these algorithms with a preconditioner that does not iteratively reduce effective step size near zero values, e.g. the OS-SPS “precomputed denominator” [11] and related row-sum of the Hessian preconditioners [26], remains to be investigated but would likely suffer due to the non-negativity constraint.

C. Impact of Data Noise

Fig. 7 indicates that SVR algorithms are particularly effective in reducing the impact of subset imbalance for noisier data. The SAGA and SVRG reconstructions generally achieve higher objective function values faster than BSREM with the 50 million event data set. SVRG and SAGA reconstructions resulting in generally higher objective function values much faster than BSREM for 50 million events. Yet, as the number of events increased, BSREM reconstructions became competitive in optimising the objective $\Phi(x)$. Moreover, as the data set SNR increases, all reconstruction algorithms appear less sensitive to relaxation parameter η .

D. Patient Data

Commonly, noise and bias trade-offs are evaluated in PML reconstruction analysis [12], [44]. This evaluation is sensitive to the reconstruction algorithm, epoch number and objective/penalty function. Due to the criteria used to optimise η in Section IV-E, the bronchial carcinoma converged quickly with BSREM, although some inter-update variations were present (Fig. 8a). However, the liver standard deviation did not converge within 10 epochs in Fig. 8b. Both of the BSREM reconstructions exhibited standard deviation values that were consistently greater than the converged value and significant inter-update variations. Therefore, usage of SAGA and SVRG effectively removes a source of potential error due to selection of the reconstruction algorithm, number of subsets or reconstruction length (after 4 epochs).

SAGA’s performance appears more sensitive to step size selection than SVRG’s. This is evidenced by the discrepancy between SAGA’s and SVRG’s optimal η values and SAGA’s larger standard deviations in Fig. 6. A plausible explanation for this phenomenon is the initial high frequency fluctuations in the SAGA reconstructions, which are not present for SVRG, that are likely caused by the lack of pre-set \mathbf{g}_m terms. Yet, even with the larger η value, SAGA converges with similar speed to SVRG and this indicates that the SVR algorithms are less sensitive to step sizes than compared to BSREM.

E. Practical Implementation for PET

For this study, we found that the memory requirement for storing M subset gradients \mathbf{g}_m terms was insignificant on a modern computing system with each \mathbf{g}_m corresponding to approximately 20MB in memory. Furthermore, the SAGA and SVRG algorithms do incur some additional computational cost in the computation of $\tilde{\nabla}_{k,m}(\mathbf{x}_k)$, compared to BSREM due to the additional image manipulations associated with (9). However, in an efficient software implementation, the impact of this should be negligible with respect to the computation

of $\nabla\Phi_m(\mathbf{x})$, which requires subset forward and backward projection operations. Therefore, an epoch of the SAGA or SVRG algorithms is expected to require slightly increased, but comparable, computational effort to gradient-based algorithms (e.g., BSREM) [21].

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

F. Comparisons to Other Work

While the studied SVR algorithms were designed to utilise stochastic subset selection, results from a preliminary study indicated that the application of an OS selection methodology resulted in superior performance during early iterations but the convergence rate was diminished after 10 or so epochs [23]. In these works, the SAG algorithm [15] was also investigated. SAG utilises a similar algorithm methodology as SAGA but, as aforementioned in Section II, sets $\xi = 1$ in (9). This makes $\tilde{\nabla}_{k,m}^{\text{SAG}}(\mathbf{x}_k)$ a biased estimator of $\nabla\Phi(\mathbf{x}_k)$. SAG was not evaluated in this work because it was observed to be less stable than SAGA under the same conditions [22], [23].

The SARAH algorithm [16] is closely related to the previously described alternative SVRG algorithm (Section V-E). SARAH also periodically recomputes all \mathbf{g}_m terms from a common iterate $\tilde{\mathbf{x}}$ but only stores $\tilde{\mathbf{x}}$ and \mathbf{G} in memory. The algorithm evaluates two subset gradients at each update, $\nabla\Phi_m(\mathbf{x}_k)$ and $\nabla\Phi_m(\mathbf{x}_{k-1})$. This additional computation is the reason that SARAH is not included in this work.

Another SVR algorithm is SVREM, which was developed and applied to PET reconstruction in a preliminary study [21]. This algorithm is a member of the EM family involving a closed-form formula for iterate updates based on parabolic surrogates and features non-negativity preservation.

Furthermore, the sPDHG algorithm has been recently developed [20] and applied to PET image reconstruction [45]. This algorithm guarantees (almost sure) convergence to the PML solution $\hat{\mathbf{x}}$ by solving a saddle point problem and has been applied to PET image reconstruction [45]. When combined with preconditioning, sPDHG reconstructions can utilise a large M and have shown great promise for 3D PET reconstruction [46]. However, the selection of sPDHG step sizes is generally heuristic and may impact performance.

This study focused exclusively on optimising $\Phi(x)$ with the RDP, which enjoys differentiability but the associated proximal operator and surrogate function are currently unknown. Consequently, neither SVREM or sPDHG were compared to SAGA and SVRG in this work. In the future, we aim to compare each of these recently proposed stochastic algorithms for clinical PET reconstruction, although an alternative regulariser may be required.

G. Future Work

Clinical reconstructions often only require a few epochs of computation before the algorithm is terminated. In this work, an epoch of the OSEM algorithm preceded all the PML reconstructions, which provided a warm start for SAGA and SVRG (see Section III-D). During this first epoch, when initialising from $x_0 = \mathbf{1}$, standard subset gradient algorithms (e.g., OSEM and BSREM) are highly effective and demonstrate almost linear acceleration [1]. However, in future work, it may be beneficial to investigate alternative strategies for warm starting, e.g., performing a fraction of an epoch of initial OSEM computation. Alternatively, one may consider pre-populating g_m terms for the SAGA algorithm to improve its stability, akin to SVRG's initial computation. This has been discussed in [27].

Regarding alternative step size relaxation methodologies, an adaptive step size selection methodology based on the periodic evaluation of objective function values could be beneficial. Such a methodology would be particularly suitable for SVRG-like algorithms that can compute the objective function value, almost for free, every few epochs.

In this work, we observed that the performance of stochastic algorithms was impacted by the choice of step size and preconditioner. The optimal preconditioner is the inverse Hessian of the objective. However, computation of this is infeasible for the 3D PET problem [1]. Yet, a number of limited memory stochastic second order methods have been proposed in the literature that are combinations of the Limited-memory-Broyden-Fletcher-Goldfarb-Shanno (LBFGS) algorithm [47], [48] and SGA [49] or SVRG [50]. These algorithms approximate the inverse Hessian based on the gradient information in the last few epochs. Alternatively, one might consider the use of nonuniform step sizes [51], the Barzilai-Borwein method [52] or momentum methods [53]. Using these methods, one may be able to further improve the convergence rate.

While in this work we focused on an azimuthal projection angle subset methodology, other methods of dividing the data exist and could be considered for SVR algorithms. List mode event based sampling and subset construction may yield additional reconstruction benefits, e.g., improved subset balance with large numbers of subsets by constructing subsets that are not limited by the geometry of the scanner [8], [9] or as an alternative to sparse projection data [54]. Therefore, construction and usage of subsets over alternative sinogram data dimensions (e.g., TOF bins, oblique angles ("segments"), or axial positions) may be viable with the presented SVR algorithms. In addition, in the context of motion compensated reconstruction, motion gates can also be used as subsets [55]. An initial investigation using sPDHG demonstrated promising results for stochastic algorithms [56]. Moreover, the acceleration that these SVR algorithms realise and the reduced sensitivity to subset selection indicates the potential benefit of their application to other low event reconstructions, such as and motion compensated parametric reconstructions. However, the impact of sparsity or greater subset imbalance may lead to additional challenges.

VI. CONCLUSION

We have shown that SVR algorithms SAGA and SVRG can be successfully adapted (10) for 3D non-TOF PET reconstruction. The investigated lung lesion, inserted into the XCAT volume, and patient data bronchial carcinoma converge to the respective PML values within 20 epochs in the majority of tested configurations cases. We observed that increasing the number of subsets accelerates convergence. While this increases stochastic variability during early epochs, this variability reduces to virtually zero within 10 epochs in most cases. Step size and preconditioner selection have dramatic impact algorithm performance, particularly in the region surrounding the emission object. Best results were obtained using moderate step size relaxation and allowing the image in the EM preconditioner to vary for a short period before anchoring.

In the tested configurations, SVRG generally optimises the objective function faster than SAGA and, when compared to the BSREM reconstructions, the investigated algorithms demonstrated superior convergence properties by every investigated metric.

Implementation of these algorithms in future studies, involving the variation of objective function parameters (e.g., priors and penalty strength), may realise improved quantification results, at lower computational cost, especially for low count data reconstructions.

REFERENCES

- [1] J. Qi and R. M. Leahy, "Iterative reconstruction techniques in emission computed tomography," *Physics in Medicine and Biology*, vol. 51, no. 15, R541–R578, 2006. DOI: 10.1088/0031-9155/51/15/R01.
- [2] J. Nuyts, D. Beque, P. Dupont, and L. Mortelmans, "A concave prior penalizing relative differences for maximum-a-posteriori reconstruction in emission tomography," *IEEE Transactions on Nuclear Science*, vol. 49, no. 1, pp. 56–60, 2002. DOI: 10.1109/TNS.2002.998681.
- [3] J. Browne and A. De Pierro, "A row-action alternative to the EM algorithm for maximizing likelihood in emission tomography," *IEEE Transactions on Medical Imaging*, vol. 15, no. 5, pp. 687–699, 1996. DOI: 10.1109/42.538946.
- [4] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2011. DOI: 10.1007/s10851-010-0251-1.
- [5] H. Erdoğan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," *Physics in Medicine and Biology*, 1999. DOI: 10.1088/0031-9155/44/11/311.
- [6] H. Hudson and R. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Transactions on Medical Imaging*, vol. 13, no. 4, pp. 601–609, 1994. DOI: 10.1109/42.363108.
- [7] A. R. De Pierro and M. E. B. Yamagishi, "Fast EM-like methods for maximum "a posteriori" estimates in emission tomography," *IEEE Transactions on Medical Imaging*, vol. 20, no. 4, pp. 280–288, 2001. DOI: 10.1109/42.921477.
- [8] A. J. Reader, S. Ally, F. Bakatselos, R. Manavaki, R. J. Walledge, A. P. Jeavons, *et al.*, "One-pass list-mode EM algorithm for high-resolution 3-D PET image reconstruction into large arrays," *IEEE Transactions on Nuclear Science*, vol. 49 I, no. 3, pp. 693–699, 2002. DOI: 10.1109/TNS.2002.1039550.

- [9] R. Twyman, S. Arridge, B. F. Hutton, E. C. Emond, L. Brusaferrri, S. Ahn, and K. 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)*, IEEE, 2019, pp. 1–3. DOI: 10.1109/NSS/MIC42101.2019.9059799.
- [10] G. T. Herman and L. B. Meyer, "Algebraic Reconstruction Techniques Can Be Made Computationally Efficient," *IEEE Transactions on Medical Imaging*, vol. 12, no. 3, pp. 600–609, 1993. DOI: 10.1109/42.241889.
- [11] S. Ahn and J. Fessler, "Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms," *IEEE Transactions on Medical Imaging*, vol. 22, no. 5, pp. 613–626, 2003. DOI: 10.1109/TMI.2003.812251.
- [12] S. Ahn, S. G. Ross, E. Asma, J. Miao, X. Jin, L. Cheng, S. D. Wollenweber, and R. M. Manjeshwar, "Quantitative comparison of OSEM and penalized likelihood image reconstruction using relative difference penalties for clinical PET," *Physics in Medicine and Biology*, vol. 60, no. 15, pp. 5733–5751, 2015. DOI: 10.1088/0031-9155/60/15/5733.
- [13] R. Johnson and T. Zhang, "Accelerating stochastic gradient descent using predictive variance reduction," in *Advances in Neural Information Processing Systems*, vol. 26, Curran Associates, Inc., 2013. DOI: 10.5555/2999611.2999647.
- [14] A. Defazio, F. Bach, and S. Lacoste-Julien, "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., 2014.
- [15] N. L. Roux, M. Schmidt, and F. Bach, "A Stochastic Gradient Method with an Exponential Convergence Rate for Finite Training Sets," *Nips*, vol. 1, no. 2, pp. 1–34, 2012.
- [16] L. M. Nguyen, J. Liu, K. Scheinberg, and M. Takáč, "SARAH: A Novel Method for Machine Learning Problems Using Stochastic Recursive Gradient," *34th International Conference on Machine Learning, ICML, 2017*.
- [17] R. M. Gower, M. Schmidt, F. Bach, and P. Richtarik, "Variance-Reduced Methods for Machine Learning," *Proceedings of the IEEE*, vol. 108, no. 11, pp. 1968–1983, 2020. DOI: 10.1109/JPROC.2020.3028013.
- [18] E. Asma, S. Ahn, H. Qian, G. Gopalakrishnan, K. Thielemans, S. G. Ross, R. M. Manjeshwar, and A. Ganin, "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*, 2012.
- [19] M. Wyrzykowski, N. Siminiak, M. Kaźmierczak, M. Ruchała, and R. Czepczyński, "Impact of the q.clear reconstruction algorithm on the interpretation of pet/ct images in patients with lymphoma," *EJNMMI Research*, vol. 10, p. 99, 1 2020. DOI: 10.1186/s13550-020-00690-6.
- [20] A. Chambolle, M. J. Ehrhardt, P. Richtárik, and C.-B. Schönlieb, "Stochastic Primal-Dual Hybrid Gradient Algorithm with Arbitrary Sampling and Imaging Applications," *SIAM Journal on Optimization*, vol. 28, no. 4, pp. 2783–2808, 2018. DOI: 10.1137/17M1134834.
- [21] Z. Kereta, R. Twyman, S. Arridge, K. Thielemans, and B. Jin, "Stochastic EM methods with variance reduction for penalised PET reconstructions," *Inverse Problems*, vol. 37, no. 11, p. 115 006, 2021. DOI: 10.1088/1361-6420/ac2d74.
- [22] R. Twyman, S. Arridge, B. Jin, B. F. Hutton, L. Brusaferrri, and K. Thielemans, "Stochastic Variance Reduction Optimisation Algorithms Applied to Iterative PET Reconstruction," in *2020 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, IEEE, 2020, pp. 1–2. DOI: 10.1109/NSS/MIC42677.2020.9508105.
- [23] R. Twyman, S. Arridge, and K. Thielemans, "Investigation of Subset Methodologies Applied to Penalised Iterative PET Reconstruction," *Proceedings of the 16th Virtual International Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine*, vol. 14, no. July, pp. 242–245, 2021.
- [24] L. A. Shepp and Y. Vardi, "Maximum Likelihood Reconstruction for Emission Tomography," *IEEE Transactions on Medical Imaging*, vol. 1, no. 2, pp. 113–122, 1982. DOI: 10.1109/TMI.1982.4307558.
- [25] R. M. Lewitt and G. Muehlllehner, "Accelerated Iterative Reconstruction for Positron Emission Tomography Based on the EM Algorithm for Maximum Likelihood Estimation," *IEEE Transactions on Medical Imaging*, vol. 5, no. 1, pp. 16–22, 1986. DOI: 10.1109/TMI.1986.4307734.
- [26] Y.-J. Tsai, A. Bousse, C. W. Stearns, S. Ahn, B. F. Hutton, S. Arridge, and K. Thielemans, "Performance improvement and validation of a new MAP reconstruction algorithm," in *2016 IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop (NSS/MIC/RTSD)*, vol. 2017-Janua, IEEE, 2016, pp. 1–3. DOI: 10.1109/NSSMIC.2016.8069458.
- [27] L. Bottou, F. E. Curtis, and J. Nocedal, "Optimization Methods for Large-Scale Machine Learning," *SIAM Review*, vol. 60, no. 2, pp. 223–311, 2018. DOI: 10.1137/16M1080173.
- [28] L. Bottou and Y. Le Cun, "Large scale online learning," in *Advances in Neural Information Processing Systems*, MIT Press, 2004.
- [29] D. Driggs, J. Liang, and C.-B. Schönlieb, "On Biased Stochastic Gradient Estimation," *Journal of Machine Learning Research*, vol. 23, 2020.
- [30] S. Jan, D. Benoit, E. Becheva, T. Carlier, F. Cassol, P. Descourt, *et al.*, "GATE V6: A major enhancement of the GATE simulation platform enabling modelling of CT and radiotherapy," *Physics in Medicine and Biology*, vol. 56, no. 4, pp. 881–901, 2011. DOI: 10.1088/0031-9155/56/4/001.
- [31] V. Bettinardi, L. Presotto, E. Rapisarda, M. Picchio, L. Gianolli, and M. C. Gilardi, "Physical Performance of the new hybrid PET/CT Discovery-690," *Medical Physics*, vol. 38, no. 10, pp. 5394–5411, 2011. DOI: 10.1118/1.3635220.
- [32] R. Twyman, L. Brusaferrri, E. C. Emond, F. Leek, S. Arridge, B. F. Hutton, V. Cuplov, and K. Thielemans, "A Demonstration of STIR-GATE-Connection," *2021 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, 2021.
- [33] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. W. Tsui, "4D XCAT phantom for multimodality imaging research," *Medical Physics*, vol. 37, no. 9, pp. 4902–4915, 2010. DOI: 10.1118/1.3480985.
- [34] K. Thielemans, C. Tsoumpas, S. Mustafovic, T. Beisel, P. Aguiar, N. Dikaos, and M. W. Jacobson, "STIR: software for tomographic image reconstruction release 2," *Physics in Medicine and Biology*, vol. 57, no. 4, pp. 867–883, 2012. DOI: 10.1088/0031-9155/57/4/867.
- [35] T. Niknejad, S. Tavernier, J. Varela, and K. Thielemans, "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)*, vol. 2017-Janua, IEEE, 2016, pp. 1–5. DOI: 10.1109/NSSMIC.2016.8069577.
- [36] D. Hogg, K. Thielemans, S. Mustafovic, and T. Spinks, "A study of bias for various iterative reconstruction methods in PET," in *2002 IEEE Nuclear Science Symposium Conference Record*, vol. 3, IEEE, 2002, pp. 1519–1523. DOI: 10.1109/NSSMIC.2002.1239610.
- [37] C. Tsoumpas, P. Aguiar, K. Nikita, D. Ros, and K. Thielemans, "Evaluation of the Single Scatter Simulation Algorithm Implemented in the STIR Library," in *IEEE Symposium Conference Record Nuclear Science 2004.*, vol. 6, IEEE, 2004, pp. 3361–3365. DOI: 10.1109/NSSMIC.2004.1466455.

- [38] T. Pan, S. A. Einstein, S. C. Kappadath, K. S. Grogg, C. Lois Gomez, A. M. Alessio, *et al.*, “Performance evaluation of the 5-Ring GE Discovery MI PET/CT system using the national electrical manufacturers association NU 2-2012 Standard,” *Medical Physics*, vol. 46, no. 7, pp. 3025–3033, 2019. DOI: 10.1002/mp.13576.
- [39] M. Schmidt, N. Le Roux, and F. Bach, “Minimizing finite sums with the stochastic average gradient,” *Mathematical Programming*, vol. 162, no. 1-2, pp. 83–112, 2017. DOI: 10.1007/s10107-016-1030-6.
- [40] Y.-J. Tsai, G. Schramm, S. Ahn, A. Bousse, S. Arridge, J. Nuyts, B. F. Hutton, C. W. Stearns, and K. Thielemans, “Benefits of Using a Spatially-Variant Penalty Strength With Anatomical Priors in PET Reconstruction,” *IEEE Transactions on Medical Imaging*, vol. 39, no. 1, pp. 11–22, 2020. DOI: 10.1109/TMI.2019.2913889.
- [41] J. Nocedal and S. J. Wright, *Numerical Optimization*. Springer New York, 2006. DOI: 10.1007/978-0-387-40065-5.
- [42] R. Boellaard, R. Delgado-Bolton, W. J. G. Oyen, F. Giannarile, K. Tatsch, W. Eschner, *et al.*, “FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0,” *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 42, no. 2, pp. 328–354, 2015. DOI: 10.1007/s00259-014-2961-x.
- [43] P. Flajolet, D. Gardy, and L. Thimonier, “Birthday paradox, coupon collectors, caching algorithms and self-organizing search,” *Discrete Applied Mathematics*, vol. 39, no. 3, pp. 207–229, 1992. DOI: 10.1016/0166-218X(92)90177-C.
- [44] E. J. Teoh, D. R. McGowan, R. E. Macpherson, K. M. Bradley, and F. V. Gleeson, “Phantom and Clinical Evaluation of the Bayesian Penalized Likelihood Reconstruction Algorithm Q.Clear on an LYSO PET/CT System,” *Journal of Nuclear Medicine*, vol. 56, no. 9, pp. 1447–1452, 2015. DOI: 10.2967/jnumed.115.159301.
- [45] M. J. Ehrhardt, P. J. Markiewicz, P. Richtárik, J. Schott, A. Chambolle, and C.-B. Schönlieb, “Faster PET reconstruction with a stochastic primal-dual hybrid gradient method,” in *Wavelets and Sparsity XVII*, SPIE, 2017, p. 58. DOI: 10.1117/12.2272946.
- [46] M. J. Ehrhardt, P. J. Markiewicz, and C.-B. Schönlieb, “Faster PET reconstruction with non-smooth priors by randomization and preconditioning,” *Physics in Medicine and Biology*, 2019. DOI: 10.1088/1361-6560/ab3d07.
- [47] J. Nocedal, “Updating Quasi-Newton Matrices with Limited Storage,” *Mathematics of Computation*, vol. 35, no. 151, p. 773, 1980. DOI: 10.2307/2006193.
- [48] Y.-J. Tsai, A. Bousse, M. J. Ehrhardt, C. W. Stearns, S. Ahn, B. F. Hutton, S. Arridge, and K. Thielemans, “Fast Quasi-Newton Algorithms for Penalized Reconstruction in Emission Tomography and Further Improvements via Preconditioning,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 4, pp. 1000–1010, 2018. DOI: 10.1109/TMI.2017.2786865.
- [49] R. H. Byrd, S. L. Hansen, J. Nocedal, and Y. Singer, “A Stochastic Quasi-Newton Method for Large-Scale Optimization,” *SIAM Journal on Optimization*, vol. 26, no. 2, pp. 1008–1031, 2016. DOI: 10.1137/140954362.
- [50] P. Moritz, R. Nishihara, and M. I. Jordan, “A Linearly-Convergent Stochastic L-BFGS Algorithm,” *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics, AISTATS 2016*, pp. 249–258, 2015.
- [51] K. Kim, D. Kim, J. Yang, G. E. Fakhri, Y. Seo, J. A. Fessler, and Q. Li, “Time of flight pet reconstruction using nonuniform update for regional recovery uniformity,” *Medical Physics*, vol. 46, pp. 649–664, 2 2019. DOI: 10.1002/mp.13321.
- [52] C. Tan, S. Ma, Y. H. Dai, and Y. Qian, “Barzilai-borwein step size for stochastic gradient descent,” *Advances in Neural Information Processing Systems*, vol. 0, no. May, pp. 685–693, 2016.
- [53] D. Kim, S. Ramani, and J. A. Fessler, “Combining ordered subsets and momentum for accelerated x-ray ct image reconstruction,” *IEEE Transactions on Medical Imaging*, vol. 34, pp. 167–178, 1 2015. DOI: 10.1109/TMI.2014.2350962.
- [54] G. Schramm and M. Holler, “Fast and memory-efficient reconstruction of sparse poisson data in listmode with non-smooth priors with application to time-of-flight pet,” *Physics in Medicine & Biology*, vol. 67, p. 155 020, 15 2022. DOI: 10.1088/1361-6560/ac71f1.
- [55] N. Dikaio and T. D. Fryer, “Acceleration of motion-compensated PET reconstruction: Ordered subsets-gates em algorithms and a priori reference gate information,” *Physics in Medicine and Biology*, 2011. DOI: 10.1088/0031-9155/56/6/011.
- [56] C. Delplancke, K. Thielemans, and M. J. Ehrhardt, “Accelerated Convergent Motion Compensated Image Reconstruction,” in *2021 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, 2021.