

Motion compensation in X-ray tomography

submitted by

Ander Biguri

for the degree of Doctor of Philosophy

of the

University of Bath

Department of Electrical and Electronic Engineering

September 2017

COPYRIGHT

Attention is drawn to the fact that copyright of this thesis rests with its author. This copy of the thesis has been supplied on the condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without the prior written consent of the author.

This thesis may be made available for consultation within the University Library and may be photocopied or lent to other libraries for the purposes of consultation.

Signature of Author

Ander Biguri

Contents

1	Introduction and Motivation	3
1.1	Aim of the thesis	6
1.2	Thesis organization	7
1.3	Publications and contributions	7
2	Image guided radiation therapy and computed tomography	9
2.1	Image Guided Radiation Therapy	9
2.2	CT imaging in other applications	9
2.3	Motion in CBCT	9
3	The image reconstruction problem	12
3.1	Geometry of CBCT	12
3.2	FDK	12
3.3	Iterative reconstruction algorithms	12
3.3.1	Algebraic Reconstruction Techniques	14
3.3.2	Conjugate Gradient Least Squares	16
3.3.3	Total variation minimization with POCS	17
3.3.4	Total variation regularization via Rudin-Osher-Fatemi model	26
3.4	Discussion	29
4	Experiments and applications	30
4.1	Algorithm experiments	31
4.1.1	Convergence rates	31
4.1.2	Total variation minimization	35
4.2	Iterative algorithms in different CT applications	39
4.2.1	Medical head CBCT from The Christie Hospital	39
4.2.2	Cryo soft X-ray tomography at the Diamond Light Source	39
4.3	Discussion	49

Chapter 1

Introduction and Motivation

Lung cancer is the most common cancer in the world in men, both in amount of cases and mortality. In women, is third in cases and second in mortality, after breast cancer[84]. Yearly deaths due to lung cancer go over 1.5 million (see figure 1-1 for incidence), having around 10% of five year survival rate in developed countries, and much lower in developing countries[18]. One over fourteen people has a lifetime risk of developing lung cancer[39], in average between men and women. The high incidence and mortality rates has lead to a big amount of research in research fields from all disciplines, in order to push further the detection and treatment of the disease, with an output of over 23.000 lung cancer related research articles in prestigious journals in the last 10 years[60]. In addition, the actual lung cancer treatment has been transformed from non-existent in the 70s to used worldwide[29].

The treatment of lung cancer varies between different types, but there are four main techniques: Chemotherapy, lobectomy or pneumoectomy, radiotherapy (RT) and palliative care. Generally, in early stages of small cell lung cancer the typical treatment would consist in chemotherapy with radiotherapy, and then brain radiotherapy, as there is chance that the tumour would spread to the head when treated. If the tumour has been detected in a very early stage and has not spread to the lymph nodes a lobectomy may be performed, removing part of the lung. Usually this is followed by radiotherapy and chemotherapy to make sure the tumour is killed. In the case of non-small cell lung cancer, in the first stages a lobectomy or a pneumoectomy (removal of the whole lung) may be performed. Generally radiotherapy and chemotherapy (less likely) are also performed. In the last stages of the cancer, usually the treatment is palliative care i.e. treatments to reduce the symptoms and relief pain[19].

In practically all stages of different lung cancer treatments, radiotherapy is extensively used as above half of the treated patients do undergo the procedure[4]. About 120.000 patients use radiotherapy in the UK every year. Radiotherapy aims to kill malignant cells using ionizing radiation, generally using photons. High energy photons

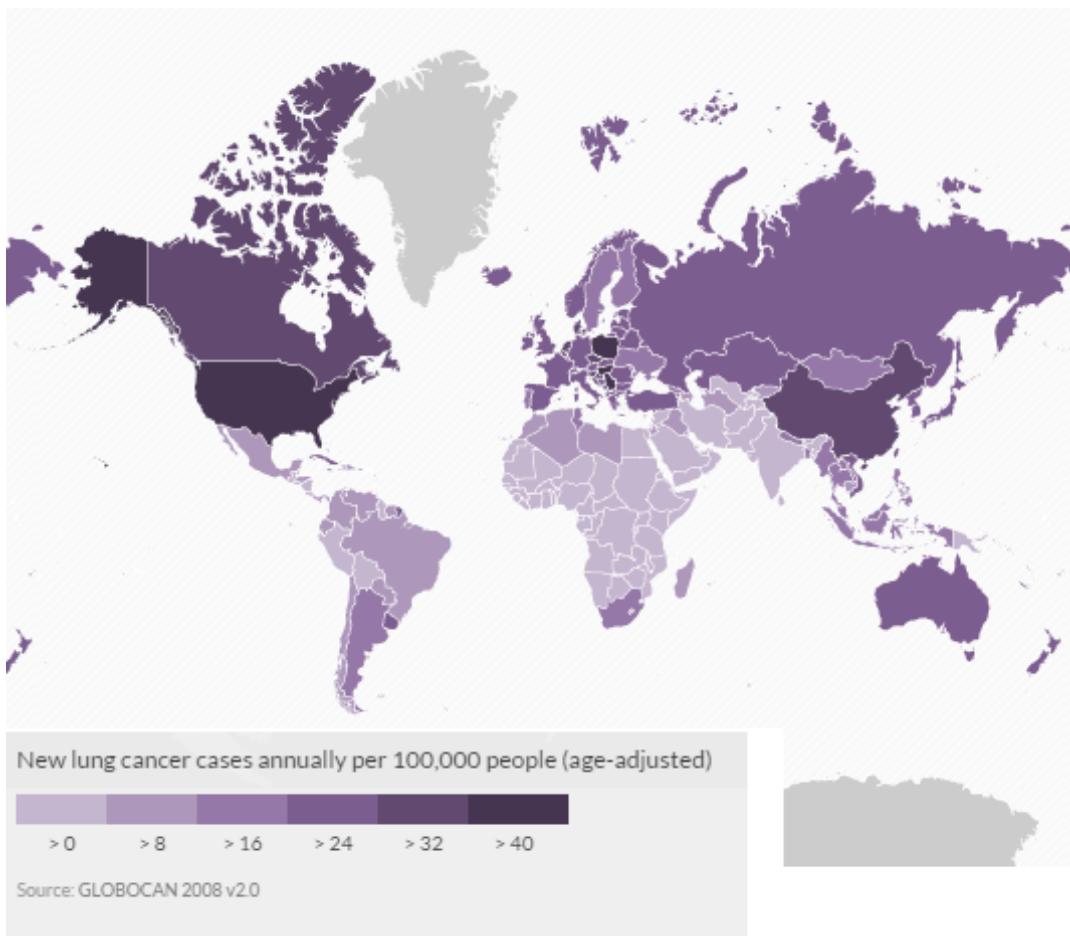


Figure 1-1: Lung cancer incidence per country, age adjusted data. Map and data from GLOBOCAN[32].

(X-rays) ionize the atoms that are part of the DNA chain. In photon therapy, this happens due to the ionization of the water in the cells, that forms free radicals, such as hydroxyl radicals, destroying the DNA of the cells and killing them.

Conventional photon RT is widely used around the world, nowadays generally guided by imaging systems during treatment planning (image guided radiation therapy, IGRT). Imaging systems, such as computed tomography (CT) and magnetic resonance imaging (MRI) are used to carefully tune the X-ray beam to focus in the specific location and shape of the tumour, and monitor the effects during the whole treatment period. However, a different type of radiation therapy exists, particle therapy or hardron therapy, that uses charged particles instead of photons, by accelerating them with circular particle accelerators. These particles (protons and heavy ions) penetrate the tissue with minimal interaction and release almost all the energy before stopping.

Figure 1-2 shows the energy deposition (dose) plotted versus the penetration of the energy beam in tissue. The energy burst that hadron show is referred as the Bragg peak, after its discoverer William Henry Bragg. The Bragg peak allows for a radiation therapy where a larger amount of healthy tissue can be spared, while delivering highly spatially accurate doses to only the tumour areas. While the growth of hadron therapy has been slow in the past due to its cost, it is not being accelerated thanks to international collaboration projects such as ENLIGHT[30], with 100 centres estimated by 2020 around the globe, 30 of them in Europe, 3 of them being already on their final stages in construction in the UK.

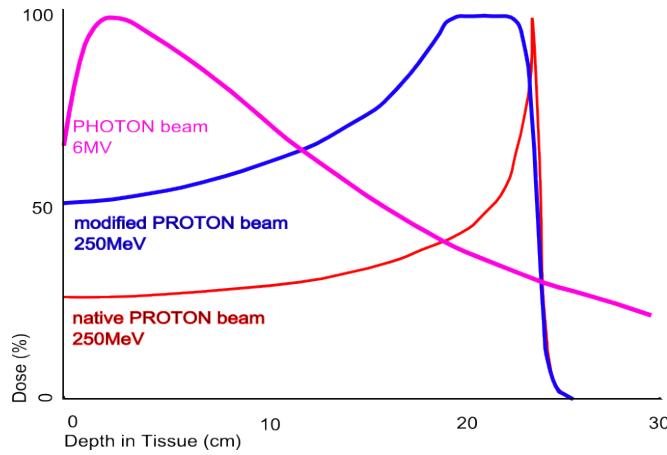


Figure 1-2: The dose produced by a native and by a modified proton beam in passing through tissue, compared to the absorption of a photon or x-ray beam. From Wikimedia[83]

As previously mentioned, for both types of RT but specially for hadron therapy, imaging is generally used and needed for accurate treatment. Tumours not only are very different between patients, but also change considerably during treatment, so the patient does due to the physical toll of cancer treatment. This means that the tumour does change both shape and location, and that if these changes are not known, healthy tissue could be damaged and cancerous tissue spared. Generally patients will be imaged before each treatment, being one of the most common systems for imaging cone beam compute tomography (CBCT). CBCT takes few hundreds of seconds to scan a patient due to mechanical safety limitations. As one can foresee, this is an important limiting factor for tumours that move, such the liver and the lung ones, as the motion during acquisition can generate heave artefacts around the moving parts in image reconstruction. This moving effect is also a key factor to take into account in hadron therapy, as having a moving tumour means the high chance of missing the

treatment target. Providing accurate imaging not only in space, but in time (4D imaging) is a key factor on treatment planning, and thus in cancer treatment.

Interestingly, a motion compensation method for when objects are moving during adquisition was proposed by Hancock *et al*[36][35][1] for monitoring the phase space of high energy particle bunches in particle accelerators at CERN. Phase space tomography is a hybrid algorithm that combines particle tracking in a computer model of a synchrotron with iterative reconstruction algorithms to reconstruct an image of the population of a bunch of particles circulating in the accelerator. The particle motion involves non-linear rotation and is non-cyclic, but a 1D projection of the distribution can be completely acquired as a single snapshot on one turn of the machine. By tracking test particles to gain a knowledge of how the geometry of the 2D image plane (longitudinal phase space) deforms, the information in all the discrete time slices acquired over many turns can be translated back to the same instant and tomographically combined in a single image. Exploring the feasibility of using this tomographic motion compensation technique in medical applications is the objective of this thesis.

1.1 Aim of the Thesis

The CBCT and computed tomography (CT) in general image reconstruction problem is a complex mathematical and computational challenge, even for just 3D spatial reconstruction, without the additional problem of motion. Mathematically CT reconstruction is an ill-posed problem, as the solution may not be unique (as in circular CBCT) and generally the volume to reconstruct is considerably larger than the data obtained, making the problem underdetermined. Often, an analytic approximated solution for the mathematical problem is used, however this solution is considerably sensitive to noise and low amounts of data. As opposed to the analytic approximated solution, algebraic equation solving methods can be used. These generally lead to more robust solutions, especially in noisy and undersampled data. However, they require increased computational times, making them harder to introduce to clinical applications.

There are two main research problems tackled in this work:

- Firstly, this thesis explores the image reconstruction problem, in with focus on implementing accurate iterative algorithms, and accelerating them as much as possible, using GPU technology. The results from this part of the thesis are applicable to any CT application, from the medical one, to industrial or research cases. The work here explores a variety of algorithms for CT reconstruction, with both mathematical and computational focus.

- Secondly, the thesis focuses in translating the motion compensation methods to the medical CBCT, focusing on IGRT applications, focusing also in the computational side of the method, as well as its robustness.

All the research presented here has been made public as part of the TIGRE Toolbox[6] and can be found in a GitHub repository[2] for both MATLAB and Python.

1.2 Thesis Organization

The following chapters of the thesis are organized as following:

Chapter 2: Image Guided Radiation Therapy and Computed Tomography

Chapter 3: The Image Reconstruction Problem

Chapter 4: GPU Methods in Tomography

Chapter 5: Motion Compensation Modelling

Chapter 7: Numerical Study of Motion Compensation

Chapter 8: Discussion and Conclusions

1.3 Publications and Contributions

The work on this thesis was been published either by open source software or publications in conferences and peer reviewed journals. The following publications directly relate to the content of this thesis:

- “GPU based iterative CBCT for prospective motion compensated algorithm for radiation therapy”. Short paper based on the presentation on the conference ICTR-PSE 2016. [CITE]
- “TIGRE: A MATLAB-GPU toolbox for CBCT image reconstruction”[6]. Journal article condensing the research in Chapter 3, Chapter 4 and Chapter 5.
- “A General Method for Motion Compensation in X-ray Computed Tomography”[8]. Journal article on Chapter 6.

This work has been also presented in various conferences and meetings, via posters or presentations talks. Posters have been presented in ToScA 2016 with the title

“TIGRE: Tomographic Iterative GPU-based Reconstruction toolbox”[7], in the ENLIGHT 2016 meeting titled “Motion correction in X-ray tomography using a priori known deformation vector fields and iterative reconstruction methods”[10] and in BIGART 2017 titled “Improvement of image quality in 4D-CBCT respiratory correlated and motion-compensated reconstruction using iterative algorithms and GPU acceleration”[11]. The work has also been presented in various talks and seminars. Finally, a Medical Physics Web article by Tami Freeman is available for wider audiences at <http://medicalphysicsweb.org/cws/article/research/66343>.

The TIGRE Toolbox and specifically some of the total variation based image reconstruction code has been also used in the article “Parameter selection in limited data cone-beam CT reconstruction using edge-preserving total variation algorithms” by Lovithee *et al*[56].

Other Publications

During the length of the PhD, mainly in the early stages, other work was published mainly focused in dual modality electrical impedance tomography (EIT) - CBCT, as some initial work explored the use of EIT for real time tumour tracking. This work is not presented in this thesis, but has been published in few items. The work is summarized in a peer reviewed journal article “Tracking boundary movement and exterior shape modelling in lung EIT imaging”[9], and extended in peer-reviewed conference articles for the EIT 2015 meeting in the works titled “Statistical and deterministic approaches for electrode movement in lung EIT”[15] and “4D FEM models of the human thorax”[12]. Initial work on this field was also presented as a poster in the ENLIGHT 2014 meeting titled “Dual modality EIT-CBCT for lung radiation therapy”[13] and in AIP 2015 as “Electrode movement due to breathing in lung EIT imaging”[14].

Chapter 2

Image Guided Radiation Therapy and Computed Tomography

2.1 Image Guided Radiation Therapy

2.2 CT Imaging in Other Applications

2.3 Motion in CBCT

Due to the lower doses used than in a conventional CT scan and to its slow data acquisition rate, a CBCT image is generally riddled with noise and motion artefacts. Research into the removal of motion artefacts in CBCT is widespread and numerous articles have been published on the subject. The most studied method to deal with motion is phase-correlated CBCT, also called 4D-CBCT[75][64][49][63][78]. In 4D-CBCT, projection data are binned according to respiratory phase and then the data from each bin are reconstructed separately to produce a series of images. This approach has several drawbacks. Even though the amount of data per reconstructed image is smaller than usual, the total number of projections increases which means a longer irradiation time and a higher dose for the patient, limiting its clinical use. In addition, the image quality of each 4D-CBCT reconstruction is inferior to a 3D-CBCT one due to its reduced angular sampling and to small inconsistencies resulting from binning inaccuracies.

Due to the limitations of standard 4D-CBCT imaging, extensive research has been conducted to improve the quality of the images. This work can be divided into two main groups: algorithmic approaches and deformation vector field (DVF) optimization methods. Methods in the first group rely on regularization and other similar approaches. An example is the work by Jia *et al.*[42], who implemented a non-local means of reconstruction to improve the temporal similarity between images. Total

variation methods (TV)[74], which minimize gradients within an image, have been also proposed with a temporal dimension included in the gradient[69]. Another method based on TV minimization is the so-called PICCS algorithm[23][46][24] (it is actually a regularizer), which minimizes the TV and the difference between the reconstructed image and a prior image. This prior image is generally a CBCT reconstructed with motion artefacts. PICCS can reconstruct 4D-CBCT images from highly undersampled datasets. More complex algorithms have also been proposed, such as ROOSTER[58], where a series of regularizations and minimizations are performed inside a region of interest to create clear 4D images in that area.

The methods of the second group generally (but not always) rely on a previous high-quality 4D-CT treatment planning scan as the basis from which to compute the DVF s . As breathing motion is neither truly periodic nor reproducible in a given patient over time, the DVF s are corrected by matching real projections with simulated ones. Finally, when the best DVF is computed, a synthetic image is generated by deforming the prior high-quality CT scan. Examples include the work of Brock *et al.*[16] and Ren *et al.*[66], who managed to reduce the number of projections required to about 60 using non-linear conjugate-gradient methods. In order to improve robustness and reduce the dimensionality of the problem, DVF principal component analysis methods have also been proposed[87]. Li *et al.*[47][48] demonstrated that good accuracy can be achieved using only a single projection for the DVF optimization.

Hybrids between DVF-based and algorithmic approaches also exist, such as using TV regularization methods to improve convergence by initializing the DVF s [82] or using temporal regularization with DVF s to improve the ROOSTER algorithm[59]. Hybrid methods can lead to highly complex optimization strategies. Examples include segmented mesh-based 4D-CBCT[88] and the separation of static and moving images using TV, tight frame regularization and DVF optimization[33]. In addition, Christoffersen *et al.*[25] have proposed a multi-step algorithm using TV and optical flow for motion estimation.

Finally, some special mathematical algorithms have also been suggested that are unique in their approach. These include the cine-CBCT algorithm[17] and the 5D motion modelling approach[52], which does not use phase-correlated binning.

The literature is full of these and many other approaches, ranging from the computationally and mathematically complex to those that sacrifice accuracy for simplicity and speed. Most have been shown to yield good 4D-CBCT reconstructions, some in clinical scenarios. But they all have drawbacks. CBCT is a severely ill-posed problem where the amount of data is key for a good reconstruction. The simplest methods that rely on binning will always suffer to some extent from a lack of data, even if temporal

coherence is enforced with mathematical norms. Additionally, they involve the reconstruction of several images, which is very expensive both computationally and in terms of memory.

Most DVF-based approaches ultimately use the DVFs to deform a prior image rather than using the acquired data directly to produce a reconstruction. Further, they assume that a DVF can describe every possible anatomical change with respect to that prior image and this does not necessarily hold.

In this work, a modelling method for motion compensation is presented, as first proposed by Hancock *et al.*[36] outside the medical domain and later independently proposed by Rit *et al.*[68][67] for CBCT. Since the publication of their work, computing on graphical processing units (GPUs) has taken a significant leap forward affording more modern techniques that can be used to reconstruct with greater accuracy and computational efficiency. With the use of GPUs even generic motion compensation is possible, without any numerical approximation of the weights in the projection and back projection and using better forward modelling[85]. Such an approach is presented in this work.

This thesis focuses on thorax CBCT, but the method is generalizable to any X-ray absorption CT modality and to arbitrary motion. The method requires no binning, but instead uses all projections to reconstruct an image at any respiratory phase. It does require a sufficiently accurate description of the motion in terms of DVFs, but the approach is a modelling one so it can be used to introduce motion compensation into any iterative reconstruction algorithm.

Chapter 3

The Image Reconstruction Problem

This chapter tries to explain the mathematics behind CT reconstruction, the FDK algorithm and iterative reconstruction algorithms. After the formal proposition of the mathematical problem of integrating over straight lines the FDK algorithm is introduced. Then, the alternative proposal of the iterative algebraic methods is shown, followed by a wide variety of different algorithms that can be used to solve the algebraic problem. These include gradient descent techniques, Krylov subspace methods and compressed sensing techniques. Finally, the challenges that arise from the use iterative algorithms are discussed.

3.1 Geometry of CBCT

3.2 FDK

3.3 Iterative Reconstruction Algorithms

Nowadays the FDK algorithm is the most widely used algorithm, and only until very recently it has been the only algorithm available in any commercial medical or industrial CT device. Companies started using hybrid iterative algorithms for a limited subsection of their software in 2010 earliest[34], for standard CT, and for IGRT/CBCT one of the first has been Varian Medical Systems Inc, patented Apr. 2017[62], published on web Sep. 2017[40]. While using FDK is advantageous in some cases, often the algorithm behaves poorly, especially when errors in the data, or the amount of data is limited. This is because FDK is based on an analytical approximation of straight path integrals in continuous spaces. The reality is far from straight path integrals as, due to X-ray physics, photons do not behave linearly. Photons from CT machines are polychromatic and human tissue is behaves non-linearly in respect to X-ray energy deposition. Additionally, Compton scattering is a common effect, where the photons

get deflected at different angles dependent on their energy. Apart from photon physics related errors, electronic noise is always present in detector technology being the only feasible way of avoiding long exposition times that would be harmful to living tissue. Limited data can additionally impair the image reconstruction, as CT has generally less detector data than the amount of voxels it is desired to reconstruct. All these effects make CT image reconstruction a challenging problem and have a strong effect on the behaviour of FDK. As an alternative to FDK, iterative algebraic reconstruction algorithms try to minimize a functional by updating the image continuously and comparing it to the measured data. These algorithms have been shown to improve reconstruction quality when the data is noisy and/or limited. Additionally, as they are an algebraic tool, they allow careful tuning of the mathematics, enabling them to change their behaviour.

Iterative algorithms in CT generally refer to those algorithms that, as the name says, iterate, but solve the linearized model

$$Ax = b + \tilde{e} \quad (3.1)$$

where $x \in \mathbb{R}^{N_{voxels}}$ is a vector representing the lexicographically ordered voxels of the 3D image, $b \in \mathbb{R}^{N_{pixels}}$ a vector of the detector measured pixels. A is the linearized model matrix, a matrix that describes the behaviour of the CT system. Each row of the matrix A describes the behaviour of the X-rays that affect each single pixel in the detector. However, this matrix is so big that in practice its explicit form is impossible to store, and the matrix product operations Ax (or projection) and A^Tb are implemented instead. The next chapter goes into a bit more detail on how to operate with matrix A and its limitations. Errors from measurement are inevitable in any application, and there are linearization errors, as no model is perfect. In equation 3.1, \tilde{e} represents all those errors.

As an exact solution for x can not be found, the problem in equation 3.1 is minimized as

$$\hat{x} = \arg \min_x \|Ax - b\|^2 + R(x), \quad (3.2)$$

where $R(x)$ is an optional regularization function. This minimization function has been widely studied in mathematics and there are multiple algorithms that can solve it. However not all algorithms that solve the equation can be used in CT reconstruction, due to the nature of the A matrix, as it is very big (approximately $10^8 \times 10^8$ in a standard medical image) and very sparse (approximately 0.0017% of sparsity in a standard medical image). This makes the matrix severely ill-conditioned and impossible to store in memory. Additionally, often the CT problem can be underdetermined,

making the problem ill-posed and further constraining the possible algorithms that can be used. That said, a wide variety of algorithm have been proposed to solve the CT algebraic problem and new ones are still being published. This section discusses a few of the available and most common algorithms that have been studied in this work and implemented in TIGRE. There are numerous other algorithms in the literature, and none of them are necessarily worse than the ones presented here.

3.3.1 Algebraic Reconstruction Techniques

Arguably the most well known iterative algorithm is the method known as the algebraic reconstruction technique (ART)[43], known as the Kaczmarz method outside the CT imaging field. The ART algorithm, for matrix elements $a_{ij} \in R$ is defined as

$$x^{n+1} = x^n + \frac{b_i - \langle a_i, x^n \rangle}{\|a_i\|^2} a_i^T, \quad (3.3)$$

where a_i is the i -th row of matrix A and \langle , \rangle denotes the inner product. The ART method projects the image into the hyperplane described by the equation in row i . Generally the method includes a relaxation parameter λ_n that controls the update size and that decreases with the iteration number n . This generally avoids the cyclical convergence that the method describes when the solution is not unique (the intersection of the hyperplanes is not a single point). By relaxing the update step, the algorithm converges to a single point. Generally the algorithm is also run with some inequality constraints, the most common one being a positivity constraint for x , as negative values are not physically possible.

Studies on the convergence of the ART algorithm show[37] that randomly choosing the order of the rows in each iteration increase the convergence rate, even more if the probabilities of picking rows are different than one (different methods propose different probabilities)[76][51].

However, the ART method has a major disadvantage: the image x^{n+1} needs to be updated i times each iteration. In current CT applications, and specifically in CBCT, the amount of rows in the matrix i.e. the total amount of independent pixel measurements in the detector is a massive number. Following the same definition of standard medical image size from the thesis, a 512×512 detector with 360 projection angles means that the amount of rows is in the order of 10^8 . In order to update the image less, the Simultaneous Iterative Reconstruction Technique (SIRT)[44] can be used, a method that is very similar to Cimmino method[26], that updates the image using simultaneously (instead of sequentially) all data in the measurement b , thus each iteration is a single update. While SIRT generally solves the problem of the high

amount of updates in ART, it suffers from a very slow convergence in comparison, and will generally plateau in a solution that is not as good as what ART provides. The SIRT algorithm can be described in matrix form as

$$x^{n+1} = x^n + \lambda_n V A^T W^{-1} (b - Ax) \quad (3.4)$$

where $V = 1/\text{diag}(\sum_j a_{ij})$ and $W = \text{diag}(\sum_i a_{ij})$.

However, a middle ground has also been proposed. Kak and Andersen proposed[3] the Simultaneous Algebraic Reconstruction Technique (SART) where the image is updated using simultaneously all data from each X-ray projection, but still updating the image multiple times per iteration. Finally, the update can also be done using block-based methods, or Oriented Subsets (OS) with a variety of methods generally described as OS-SART[21] methods.

Similarly as with ART, the order of the subsets in both OS-SART and SART influence the convergence, but with a lower impact than in ART. In this work a three methods have been implemented, a completely ordered method, a randomized ordered method with full sampling (i.e. all projections are ensured to be used once and only once per iteration) and an angular distance based one. This last one orders the subsets by selecting the next one as the subset with largest angular distance from the ones already used. The heuristic rationale is that the projections at larger angular distance update the image by a bigger step than projections angularly near. In this thesis, the default ordering is random unless otherwise explicitly stated.

Relaxation Parameter λ

As previously mentioned, changing the relaxation parameter per iteration can be of advantage, by avoiding cyclical convergence and often by increasing the general convergence rate. One of the commonly used methods for the reduction of lambda is simply reducing it by a reduction factor each iteration as

$$\lambda_{n+1} = \lambda_n * r_\lambda \quad (3.5)$$

where r_λ is some value close to one, such as $r_\lambda = 0.99$ or $r_\lambda = 0.999$. However this method, while useful to avoid cyclical convergence in ART methods, is of less use in simultaneous methods, as it generally slows the convergence rate.

It is worth noticing that this family of algorithms is very closely related to the well-known gradient descent methods, as the gradient of equation 3.2 is proportional to $A^T(Ax - b)$, or in other words $V = I$ and $W = I$ in equation 3.4. The gradient descent methods have been widely studied in the past years[77][70], as the Neural

Networks community tries to find faster converging methods to train the nets they research. Among other methods proposed, Nesterov proposed an accelerated version of the gradient descent[61], that obtains a rate of convergence of $1/n^2$. The proposed update updates the result image in each iteration by pushing it in the current update and previous update direction combined. Nesterovs Accelerated Gradient (NAG) defines

$$\lambda_{n+1} = \frac{1 + \sqrt{1 + 4\lambda_n^2}}{2} \quad (3.6)$$

$$\gamma_n = \frac{1 - \lambda_n}{\lambda_{n+1}} \quad (3.7)$$

$$y^{n+1} = x^n - \frac{1}{\beta} \nabla f(x^n) \quad (3.8)$$

$$x^{n+1} = (1 - \gamma_n)y^{n+1} + \gamma_n y^n \quad (3.9)$$

with $\lambda_0 = 1$ and β being the Lipschitz smoothness of the function f . The line on equation 3.8 can be replaced by the SART/OS-SART/SIRT update on equation 3.4 to obtain an accelerated convergence rate. Some experimental results on the convergence of the algorithms can be found in Chapter 5.

3.3.2 Conjugate Gradient Least Squares

The conjugate gradient for the least squares problems (CGLS) is, at its name says, an algorithm that solves the least squares problems for the normal equations (equation 3.10) using the conjugate gradient method. As a very short description of the method, CGLS is a 2-norm minimization method that iterates through Krylov subspaces. Doing this, each step minimizes the next biggest eigenvector of the residual of the 2-norm of the normal equation, thus updating the solution by the biggest linear step possible. This results in a superior convergence compared to other iterative methods[50], and Krylov subspace methods are the fastest iterative solvers for linear equations.

$$A^T A x = A^T b \quad (3.10)$$

However, these methods have their drawbacks, as when only the approximated transpose of the matrix A is known and when rounding errors appear after several iterations, the update step may be wrongly computed. One solution to avoid divergence is restarting the algorithm, by recomputing the initial steps using as a prior image the current intermediate solution of CGLS.

The full description and implementation of the algorithm can be found in table ???. For further reading about the this method, the paper by Shechuk is suggested[73].

Various different Krylov subspace methods are also available in the literature, such as LSQR[65], GMRES[27], Bi-CGStab[80], among others.

3.3.3 Total Variation Minimization With POCS

Sometimes solving a regularized problem may result in a better final image than just trying to solve the data constraint with the model. This is especially useful in more ill-conditioned problems, such as when the data is very noisy (thus the model does not fit the data accurately) or when few projections are available (the system is more under-determined). In these cases, regularization can add a user constraint in the image domain that pushes the algorithm towards a specific solution among all the multiple possibilities. While a variety of regularization techniques and norms exist, the most suitable for CT imaging is the total variation (TV) norm.

The total variation norm is defined as the sum of the 2-norms of the directional gradients of the variable,

$$\|x\|_{\text{TV}} = \sum_n \left\| \sum_{\alpha} \partial_{\alpha} x_n \right\|_2. \quad (3.11)$$

Applied to CT imaging, the total variation norm is the sum of the total change occurring in the image. An image with less total variation would be an image that would have less change, or more flat, same valued regions. Regularizing with the TV norm as a minimization term will yield an image that is piecewise smooth and it happens that most of the objects imaged in CT scanners are piecewise smooth in linear attenuation, even more in medical CT imaging.

However, solving the minimization problem in equation 3.2 is not trivial with TV regularization. One of the first robust algorithm to do so is the so-called Adaptive Steepest Descent, Projection Onto Convex Subsets, or ASD-POCS algorithm[74]. This algorithm not only minimizes the data constraint with TV regularization but also adaptively controls the TV minimization update, in order to adapt its strength according to the data constraint update. Several adaptations and improvements of this algorithm have been proposed in the literature[55][79][54], all based on the same mathematical approach.

ASD-POCS

The previous algorithms discussed in this chapter were unconstrained minimization methods. While the TV minimization problem can be solved similarly (see section 3.3.4), formalizing the algorithm as a non-linear constrained minimization adds an

advantage in the case where there system is under-determined. In an unconstrained problem such as in equation 3.2, the balance between the data constraint and the regularization constraint can be tuned via a hyperparameter, but in the case of an under-determined system, multiple solutions for the data fidelity term may exist. By reformulating it as shown in the rest of this section, the image with the same data fidelity 2-norm but the lowest TV norm can be chosen.

The minimization will yield an image \vec{x}^* that minimizes

$$\vec{x}^* = \arg \min_x \|\vec{x}\|_{\text{TV}} \quad (3.12)$$

subject to

$$\|A \cdot \vec{x} - \vec{b}\| \leq \epsilon, \quad (3.13)$$

$$\vec{x} \geq 0. \quad (3.14)$$

As previously described in this chapter, the data fidelity in equation 3.13 while desired to be zero, it will never reach zero, due to inconsistencies in the data, model, noise, etcetera. Thus, in this algorithm it is introduced as an inequality constraint, instead of as the minimization problem itself. This introduces the parameter ϵ in the algorithm, the maximum 2-norm allowed for the data inconsistency. The problem in hand is now non-linear, due to the constraints, but convex.

The conditions for a constrained minimization to find the optimal solution can be obtained by satisfying the Karush Kuhn-Tucker conditions (a generalization of the Lagrange multiplies for inequality constraints). First, the Lagrangian for the current problems needs to be defined, as

$$\mathbf{L} = \|\vec{x}\|_{\text{TV}} + \lambda_0 \cdot (\|A \cdot \vec{x} - \vec{b}\|^2 - \epsilon^2) - \vec{\lambda} \cdot \vec{x}, \quad (3.15)$$

where $\vec{\lambda}$ is a vector of the same size as the image, but λ_0 is a single value. Two inequality constraints are imposed to the Lagrange multipliers, namely non-negativity

$$\lambda_i \geq 0, \quad (3.16)$$

and complementarity

$$h_i(\vec{x})\lambda_i = 0, \quad (3.17)$$

where $i = 0, 1, \dots, N_{pixels}$, and h_i is an alternative form of the inequality constraints as

$$h_0(\vec{x}) = \|A \cdot \vec{x} - \vec{b}\|^2 - \epsilon^2 \leq 0 \quad (3.18)$$

$$h_i(\vec{x}) = -x_i \leq 0 \quad i \in [1, N_{pixels}] \quad (3.19)$$

Thus, only when the inequalities are violated does h_i turns non-zero, and with the complementarity condition, does the corresponding λ_i turns zero. A solution can be found for \vec{x} when the gradient of the Lagrangian is zero, and if the differential operator is defined as

$$\nabla_{\vec{x}} Q(\vec{x}) = \sum_i \partial_{x_i} Q(\vec{x}) \vec{\delta}_i \quad (3.20)$$

where $\vec{\delta}_i$ is the Kronecker delta. The gradient of the Lagrangian can be then written as

$$\begin{aligned} \nabla_{\vec{x}} \mathbf{L} &= \nabla_{\vec{x}} \|\vec{x}\|_{TV} + \lambda_0 \nabla_{\vec{x}} h_0(\vec{x}) + \sum_{i=1}^{N_{pixels}} \lambda_i \nabla_{\vec{x}} h_i(\vec{x}) = 0 \\ &= \nabla_{\vec{x}} \|\vec{x}\|_{TV} + 2\lambda_0 A^T \cdot (A \cdot \vec{x} - \vec{b}) - \vec{\lambda} = 0 \end{aligned} \quad (3.21)$$

Further simplification can be applied to equation 3.21. As the non-negativity constraints are only active in zero valued voxels, the Lagrange multipliers are zero for strictly positive voxels. Thus, by adding an indicator function

$$\vec{x}_{indic} = \begin{cases} 1 & \vec{x} \neq 0 \\ 0 & \vec{x} = 0 \end{cases} \quad (3.22)$$

the Lagrangian gradient can be shortened to

$$\nabla_{\vec{x}} \mathbf{L} = \text{diag}(\vec{x}_{indic}) (\nabla_{\vec{x}} \|\vec{x}\|_{TV} + \lambda_0 \nabla_{\vec{x}} h_0(\vec{x})) = 0. \quad (3.23)$$

Separating this new equation into two vectors,

$$\begin{aligned} \vec{d}_{TV} &= \text{diag}(\vec{x}_{indic}) (\nabla_{\vec{x}} \|\vec{x}\|_{TV}) \\ \vec{d}_{data} &= \text{diag}(\vec{x}_{indic}) (\nabla_{\vec{x}} h_0(\vec{x})) \end{aligned} \quad (3.24)$$

brings to the Karush Kuhn-Tucker conditions: \vec{x} will be an optimal condition if \vec{d}_{TV} and \vec{d}_{data} are pointing in exactly the opposite direction. In practice the algorithm will only check if the vectors are pulling in opposite direction (by computing the dot

product) and that the inequality constrains are satisfied. By checking the direction of the vectors the algorithm ensures that even if the data constraint is satisfied, only the optimal solution regarding both TV norm and data fidelity is chosen.

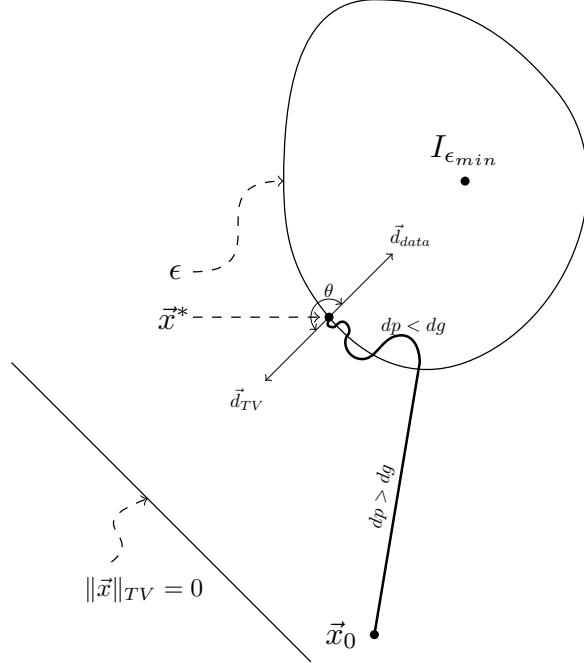


Figure 3-1: Conceptual diagram of the ASD-POCS algorithm path to the solution.

Figure 3-1 shows a conceptual diagram of the ASD-POCS algorithm. There is an area around the image with minimum data constraint, $I_{\epsilon_{min}}$. The solution \vec{x}^* generally lies on the boundary of the area with the user specified ϵ . From an initial image \vec{x}_0 , the algorithm walks towards the area of acceptable ϵ more strongly than towards the area of minimum TV, as the step sizes of the vectors \vec{d}_{TV} and \vec{d}_{data} , dp and dg respectively, are adaptively controlled to be $dp > dg$. Once the image is within acceptable 2 norm, then the step size is changed in order to have stronger \vec{d}_{TV} ($dp < dg$). The optimal solution can be found when both vectors point in opposing direction, or in other words, when the angle between them is 180 degrees, or the cosine of it is -1,

$$\cos \theta = \frac{\vec{d}_{TV} \cdot \vec{d}_{data}}{\|\vec{d}_{TV}\| \|\vec{d}_{data}\|}. \quad (3.25)$$

The pseudocode for ASD-POCS can be seen in algorithm 1. The algorithm is essentially solving the two vector in equation 3.24, the data vector in lines [5-8] and the TV vector in lines [18-22]. Line [9] enforces the positivity constraint. In the algorithm,

dtv is initialized according to α , an user specified TV hyperparameter for TV, together with dp , the step size performed by the data constraint. After the TV minimization is performed, the step size of the TV vector is rechecked. If the TV minimization step is too big (bigger than the data step size), and the desired ϵ is still not achieved, the step size is reduced further. This method of adaptively setting the step size of the TV iteration relating to the data step size is what ensures the optimal condition is achieved. Finally, the stopping criteria relies in either achieving the desired ϵ with a desired $\cos \theta$, or stopping due to reaching a maximum amount of iterations (β decreases with iteration number). In the original proposition of the ASD-POCS algorithm (and shown here), the data constraint is solved using SART, however any other algorithm solving the same minimization problem can be used here (e.g. CGLS or OS-SART).

The algorithm has 7 parameters that need to be set up: β and β_{red} are the initial value and reduction ratio of the SART hyperparameter, similarly α and α_{red} serve as hyperparameter and reduction ratio for the TV minimization. r_{max} controls the maximum allowed ratio of change between the data minimization and the TV minimization, in order to adapt the step sizes. The number of iterations the TV minimization performs per iteration of the data minimization is defined as n_{TV} . Finally, the allowed data error is ϵ , as described before. The initial values of the variables in the algorithm are a key factor on its convergence. Empirical tests show that wrong parametrization of the algorithm can lead to severely noisy reconstructions. A study of the sensitivity of these parameters to changes has been performed by Lohvithee *et al*[56]. The study shows that some parameters can be safely set up to a static value regardless of the data, such as the data hyperparameters, but that ϵ , n_{TV} and α are critical parameters to tune in order to get an usable reconstruction, and they are heavily data dependant. While some algorithms have successfully replaced the initial set of α by some data based heuristics[53]¹ to the best of the authors knowledge there is no mathematical proposal for setting these parameters. The biggest drawback of this method is that several reconstructions may be needed to find the best parameters for an specific application.

Note that this minimization approach, while used for TV minimization in the original article, can be used for a variety of different minimization functions. For example, the TV minimization step could be replaced by a prior image minimization[23], or any other convex minimization function. Similarly, the data minimization step can be replaced by any other minimization algorithm, as long as it minimizes the 2-norm of the data constraint.

¹These algorithms, namely PCSD and Aw-PCSD, are also available in TIGRE, by Manasavee Lohvithee.

Algorithm 1 ASD-POCS

```
1: Set:  $\beta, \beta_{red}, n_{TViter}, \alpha, \alpha_{red}, r_{max}$ 
2:  $\vec{x} = 0;$ 
3: while Stopping criteria not met do
4:    $\vec{x}_{prev} = \vec{x}$ 
5:   for  $n_{angles}$  do
6:      $\vec{x} = \vec{x} + \beta V A^T W^{-1} (\vec{b} - A\vec{x})$                                  $\triangleright$  SART update
7:   end for
8:    $\beta = \beta * \beta_{red}$ 
9:    $\vec{x} = \max(0, \vec{x})$                                                   $\triangleright$  Enforce positivity
10:   $\vec{x}_{out} = \vec{x}$ 
11:   $\epsilon_{now} = \|A\vec{x} - \vec{b}\|$                                           $\triangleright$  Current  $\epsilon$ 
12:   $dp = \|\vec{x} - \vec{x}_{prev}\|$                                           $\triangleright$  Change in  $\vec{d}_{data}$ 
13:  if first iteration then
14:     $dtv = \alpha * dp$                                           $\triangleright$  Initialize TV hyperparameter
15:  end if
16:   $\vec{x}_{prev} = \vec{x}$ 
17:
18:  for  $n_{TViter}$  do                                               $\triangleright$  TV update
19:     $\vec{dx} = \nabla_{\vec{x}} \|\vec{x}\|_{TV}$ 
20:     $\hat{dx} = \vec{dx} / \|\vec{dx}\|$ 
21:     $\vec{x} = \vec{x} - dtv \cdot \hat{dx}$ 
22:  end for
23:   $dg = \|\vec{x} - \vec{x}_{prev}\|$                                           $\triangleright$  Change in  $\vec{d}_{TV}$ 
24:  if  $dg > r_{max} * dp$  and  $\epsilon_{now} > \epsilon$  then
25:     $dtv = dtv * \alpha_{red}$ 
26:  end if                                          $\triangleright$  Check stopping criteria
27:   $\cos \theta = \vec{dp} \cdot \vec{dg} / \|\vec{dp}\| \cdot \|\vec{dg}\|$ 
28:  if ( $\cos \theta < -0.9$  and  $\epsilon_{now} < \epsilon$ ) or  $\beta < 0.005$  then
29:    Stop
30:  end if
31: end while
```

B-ASD-POCS- β

Xue *et al*[86] proposed a faster converging modification of the ASD-POCS algorithm by adding a relaxed Bregman iteration to the overall algorithm. The Bregman iteration updates the projection data with part of the remaining error, thus reducing the residual faster and reaching convergence faster. The top level pseudo-code of the B-ASD-POCS- β can be seen in algorithm 2.

Algorithm 2 B-ASD-POCS- β

```

1: Set:  $\beta, \beta_{red}, n_{TViter}, \alpha, \alpha_{red}, r_{max}$ 
2:  $\beta$  update ratio:  $c < 1$ 
3:  $\beta$  update period:  $T$ 
4:  $\vec{x} = 0;$ 
5: while Stopping criteria not met do
6:   for  $n_{Bregman}$  do
7:     ASD-POCS algorithm
8:   end for
9:    $\vec{b} = \vec{b} + \beta \cdot (\vec{b}^0 - A \cdot \vec{x})$ 
10:  Update  $\beta = c \cdot \beta$  if current iteration is update period  $T$  multiplier
11: end while

```

In the article, the authors show that this approach results in a faster (less amount of iterations) convergence, getting to a solution faster than in the standard ASD-POCS approach. This algorithm however requires 3 extra parameters for tuning.

The Gradient of the TV Norm

In order to minimize the TV norm via gradient descent, the gradient of the TV norm needs to be computed, $\nabla_{\vec{x}}||\vec{x}||_{TV}$, being \vec{x} the vectorized form of a N-dimensional image.

The main challenge with the $\nabla_{\vec{x}}||\vec{x}||_{TV}$ term is that $||\vec{x}||_{TV}$ is not differentiable in the general case. However, in the CT case, \vec{x} can be described as x_{ijk} , a regularly discretized mesh of directional indices i, j, k of maximum value $i_{max}, j_{max}, k_{max}$. The gradient of x has an additional Cartesian index α :

$$g^\alpha = (\nabla x)^\alpha = \partial_\alpha x \quad (3.26)$$

$$g_{ijk}^\alpha = \partial_\alpha x_{ijk}. \quad (3.27)$$

The TV norm can be then defined as sum of the 2-norms of the gradient of x , g , over the Cartesian coordinate, resulting in a scalar.

$$\|x\|_{\text{TV}} = \sum_{ijk} \sqrt{\sum_{\alpha} \left(g_{ijk}^{\alpha} \right)^2} = \sum_{ijk} \sqrt{\sum_{\alpha} (\partial_{\alpha} x_{ijk})^2}, \quad (3.28)$$

This is the term that the total variation regularization algorithm minimizes with a gradient descent. In order to perform this, the gradient of this term with respect to x is needed, now defined in a scalar field

$$(\nabla_x \|x\|_{\text{TV}})_{ijk}. \quad (3.29)$$

This derivative can be expanded to a 3 component value for each x_{ijk} as:

$$\begin{aligned} (\nabla_x \|x\|_{\text{TV}})_{ijk} &= \frac{\partial}{\partial x_{ijk}} \|x\|_{\text{TV}} = \partial_{x_{ijk}} \sum_{i'j'k'} \sqrt{\sum_{\alpha} (\partial_{\alpha} x_{i'j'k'})^2} \\ &= \sum_{i'j'k'} \partial_{x_{ijk}} \sqrt{\sum_{\alpha} (\partial_{\alpha} x_{i'j'k'})^2} \\ &= \sum_{i'j'k'} \frac{\sum_{\alpha} (\partial_{\alpha} x_{i'j'k'}) \partial_{x_{ijk}} (\partial_{\alpha} x_{i'j'k'})}{\sqrt{\sum_{\alpha} (\partial_{\alpha} x_{i'j'k'})^2}}. \end{aligned} \quad (3.30)$$

This term now contains ∂_{α} derivatives, i.e. derivatives in the Cartesian coordinate system $[x, y, z]$. These are defined as

$$\begin{aligned} \partial_x x_{i'j'k'} &= \lim_{h \rightarrow 0} \frac{x_{i'+h,j',k'} - x_{i',j',k'}}{h} \\ \partial_y x_{i'j'k'} &= \lim_{h \rightarrow 0} \frac{x_{i',j'+h,k'} - x_{i',j',k'}}{h} \\ \partial_z x_{i'j'k'} &= \lim_{h \rightarrow 0} \frac{x_{i',j',k'+h} - x_{i',j',k'}}{h}. \end{aligned} \quad (3.31)$$

However, x is discrete, thus the limit definition of the derivative can not be used to numerically compute it, but an approximation of it can. By setting $h = 1$, equation 3.31 becomes the backward finite differences of the first order approximation of a derivative, a very computationally cheap operation. The derivative w.r.t. the Cartesian coordinate can be rewritten as

$$\begin{aligned}\partial_\alpha x_{i'j'k'} &= \delta_{\alpha x} (x_{i',j',k'} - x_{i'-1,j',k'}) + \delta_{\alpha y} (x_{i',j',k'} - x_{i',j'-1,k'}) \\ &\quad + \delta_{\alpha z} (x_{i',j',k'} - x_{i',j',k'-1})\end{aligned}\quad (3.32)$$

where δ_α is a Kronecker delta for the Cartesian axis. The other partial derivative term that appears in equation 3.30 is $\partial_{x_{ijk}} (\partial_\alpha x_{i'j'k'})$. As the derivative is w.r.t. x_{ijk} , each component of x is an independent variable, thus $\partial_{x_{ijk}} (\partial_\alpha x_{i'j'k'})$ is zero everywhere but in indices $i = i' \wedge j = j' \wedge k = k'$, where the derivative is 1. The term then becomes

$$\begin{aligned}\partial_{x_{ijk}} \partial_x x_{i'j'k'} &= \partial_{x_{ijk}} (x_{i',j',k'} - x_{i'-1,j',k'}) = \delta_{i',i} \delta_{j',j} \delta_{k',k} - \delta_{i'-1,i} \delta_{j',j} \delta_{k',k} \\ &= \delta_{i',i} \delta_{j',j} \delta_{k',k} - \delta_{i',i+1} \delta_{j',j} \delta_{k',k} \\ \partial_{x_{ijk}} \partial_y x_{i'j'k'} &= \partial_{x_{ijk}} (x_{i',j',k'} - x_{i',j'-1,k'}) = \delta_{i',i} \delta_{j',j} \delta_{k',k} - \delta_{i',i} \delta_{j'-1,j} \delta_{k',k} \\ &= \delta_{i',i} \delta_{j',j} \delta_{k',k} - \delta_{i',i} \delta_{j',j+1} \delta_{k',k} \\ \partial_{x_{ijk}} \partial_z x_{i'j'k'} &= \partial_{x_{ijk}} (x_{i',j',k'} - f_{i',j',k'-1}) = \delta_{i',i} \delta_{j',j} \delta_{k',k} - \delta_{i',i} \delta_{j',j} \delta_{k'-1,k} \\ &= \delta_{i',i} \delta_{j',j} \delta_{k',k} - \delta_{i',i} \delta_{j',j} \delta_{k',k+1}.\end{aligned}\quad (3.33)$$

These terms are practically a selecting function for i', j', k' , matching only in the indices $i, i+1, j, j+1, k, k+1$ in the sum of the right hand side of equation 3.30. However the indices are limited to $i' \in [1, i_{max}]$, $j' \in [1, j_{max}]$ and $k' \in [1, k_{max}]$. As boundary conditions, Neumann boundary conditions are set to zero. To enforce that, a Kronecker deltas can be introduced for each index, $(1 - \delta_{i,i_{max}})$, with the same approach with the other indices.

Finally, substituting in equation 3.30, the gradient of the TV norm can be described as

$$\begin{aligned}
(\nabla_x \|x\|_{\text{TV}})_{ijk} &= \sum_{i'j'k'} \frac{\sum_{\alpha} (\partial_{\alpha} x_{i'j'k'}) \partial_{x_{ijk}} (\partial_{\alpha} x_{i'j'k'})}{\sqrt{\sum_{\alpha} (\partial_{\alpha} x_{i'j'k'})^2}} \\
&= \sum_{i'j'k'} \frac{\partial_x x_{i'j'k'} \partial_{x_{ijk}} (\partial_x x_{i'j'k'}) + \partial_y x_{i'j'k'} \partial_{x_{ijk}} (\partial_y x_{i'j'k'}) + \partial_z x_{i'j'k'} \partial_{x_{ijk}} (\partial_z x_{i'j'k'})}{\sqrt{\sum_{\alpha} (\partial_{\alpha} x_{i'j'k'})^2}} \\
&= \frac{\partial_x x_{i,j,k} + \partial_y x_{i,j,k} + \partial_z x_{i,j,k}}{\sqrt{\sum_{\alpha} (\partial_{\alpha} x_{i,j,k})^2}} \\
&\quad - \frac{(1 - \delta_{i,i_{max}}) \partial_z x_{i+1,j,k}}{\sqrt{\sum_{\alpha} (\partial_{\alpha} x_{i+1,j,k})^2}} - \frac{(1 - \delta_{j,j_{max}}) \partial_y x_{i,j+1,k}}{\sqrt{\sum_{\alpha} (\partial_{\alpha} x_{i,j+1,k})^2}} - \frac{(1 - \delta_{k,k_{max}}) \partial_z x_{i,j,k+1}}{\sqrt{\sum_{\alpha} (\partial_{\alpha} x_{i,j,k+1})^2}}. \quad (3.34)
\end{aligned}$$

Equation 3.34 is the numerical approximation of the gradient of the total variation norm, and describes scalar field of the same size of the image. The same approach can be used with central and forward differences to obtain a similar equation, however central differences may not correctly minimize the TV norm of the image. As central differences do not take into account the value of the current voxel ijk , a chequerboard pattern would have zero TV norm, and this is the opposite of the purpose of the algorithm, therefore only numerical approximations of derivatives that take immediately adjacent pixel values into account can be used (such as forward or backward finite differences).

3.3.4 Total Variation Regularization via Rudin-Osher-Fatemi Model

A different minimization approach to POCS is the approach proposed by Jia *et al*[41], that uses the Rudin-Osher-Fatemi (ROF) model for total variation minimization, widely used in the denoising literature[71][31][81]. By starting from the same minimization problem, namely

$$\hat{x} = \arg \min_x \|Ax - b\|^2 + \lambda \|x\|_{\text{TV}}, \quad (3.35)$$

a forward-backward splitting algorithm[28] is used to split the minimization into two alternating steps, the TV and the data steps. If the optimality condition is considered to be

$$\frac{\partial}{\partial x_{\alpha}} \|Ax - b\|^2 + \lambda \frac{\partial}{\partial x_{\alpha}} \|x\|_{\text{TV}} = 0, \quad (3.36)$$

being α the set of Cartesian coordinates, then the problem can be split into the following equations, where g is a auxiliary function and $\mu > 0$:

$$\lambda \frac{\partial}{\partial x_\alpha} \|x\|_{TV} = \mu \cdot (x - g) \quad (3.37)$$

$$\frac{\partial}{\partial x_\alpha} \|Ax - b\|^2 = -\mu \cdot (x - g). \quad (3.38)$$

By solving for g , the simplified version of the algorithm can be seen in 3.

Algorithm 3 TV minimization with ROF model

- 1: Solve: $g = x - \frac{\lambda}{\mu} \frac{\partial}{\partial x_\alpha} \|Ax - b\|^2$ ▷ SART
 - 2: Minimize: $x = \arg \min_x \|x\|_{TV} + 0.5 \cdot \mu \|x - g\|^2$
 - 3: Enforce positivity: $x = \max(0, x)$
-

The first line of the algorithm its essentially a gradient descent iteration, which can be replaced with a SART iteration. Note that the this iteration can be replaced by other data-minimization algorithms such as CGLS. The second line is the ROF model, widely researched in image denoising. The ROF model tries to find the image x with minimum total variation subject to having the minimal deviation from its original value g . By changing the value of the hyperparameter μ , the strength of this regularization is controlled. A high μ will ensure that the image is very similar to its original value, while a small μ will be more lax. The advantage of this approach compared to the ASD-POCS algorithm is that it requires no extra projection or backprojection operations. Additionally, minimizing the ROF model is a very well studied problem in the image processing field, and it has lead to finding highly computationally efficient methods.

In the article by Jia *et al*, they solve the ROF model via gradient descent and controlling its step size with Armijo's rule. In this work a different approach is taken, based on the image processing literature.

Primal Dual Formulation of the ROF Model

As previously shown in line 2 of algorithm 3, the ROF model can be formulated as

$$\hat{x}_{ROF} = \arg \min_x \|x\|_{TV} + \frac{\mu}{2} \|x - g\|^2. \quad (3.39)$$

A solution of this problem using a primal-dual (PDU) approach has been proposed in literature[89], by changing the minimization equation to a saddle point optimization problem. While a wide variety of methods have been proposed to minimize the ROF model[71][81][22], the PDU method has the advantage of being very parallelizable,

thus a perfect fit for GPU computing. The dual variable can be proposed by using the TV definition of $\|x\|_{TV} = \|\nabla x\|$ and observing the following consequence of the Cauchy-Schwartz inequality

$$\|\nabla x\| = \arg \max_{\|\mathbf{p}\| \leq 1} \|\mathbf{p} \nabla x\|, \quad (3.40)$$

where $\mathbf{p} = (p^1, p^2, p^3)^T$ (for the 3D case) is the said dual variable. Note that each p^i is the size of the image x . Equation 3.39 can be then rewritten as

$$\hat{x}_{ROF} = \arg \min_x \arg \max_{\|\mathbf{p}\| \leq 1} \|\mathbf{p} \nabla x\| + \frac{\mu}{2} \|x - g\|^2. \quad (3.41)$$

The primal and dual updates can be both obtained from this equation. For the primal update, differentiating the equation according to x results in

$$-\nabla \cdot \mathbf{p} + \mu \cdot (x - g) = 0, \quad (3.42)$$

and one can solve it for x by performing a gradient descent update as

$$x^{n+1} = x^n(1 + \tau_P^n) + \tau_P^n \left(g + \frac{1}{\mu} \nabla \cdot \mathbf{p} \right), \quad (3.43)$$

where τ_P is the primal step size. The dual update can be computed similarly, by differentiating equation 3.41 according to \mathbf{p} , the following equation is obtained:

$$\nabla x + \mathbf{p} \alpha = 0, \quad (3.44)$$

where α is a Lagrange multiplier for the inequality constraint $\|\mathbf{p}\| \leq 1$. This equation can be maximized with a gradient ascend method as

$$\mathbf{p}^{n+1} = \Pi_{B_0}(\mathbf{p}^n + \tau_D^n \nabla x), \quad (3.45)$$

where $\Pi_{B_0}(\mathbf{p}) = \frac{\mathbf{p}}{\max\{1, \|\mathbf{p}\|\}}$ is a projection onto the unit ball centred in the origin.

The PDU algorithm consists in updating \mathbf{p} and x iteratively, by alternating the updates. In [89][45] a step size update is proposed for the primal and dual step sizes:

$$\begin{aligned} \tau_D^n &= 0.3 + 0.02n \\ \tau_P^n &= \frac{1}{\tau_D^n} \left(\frac{1}{6} - \frac{5}{15+n} \right). \end{aligned} \quad (3.46)$$

The same update is used in this work, as the images in their work are structurally

similar to CT images and empirical test showed satisfactory results. The algorithm can be shown to converge as it is shown in [90] that the primal-dual gap decreases with each update of x^n , and the gap is suggested as a control variable for the stopping criteria. In this work the algorithm has been implemented without the stopping criteria check, and an user specified parameter for the number of iterations is passed as an input, with a default value of 50, as it empirically showed good results.

The discretization of the divergence and gradient operators are a key factor when numerically computing the PDU algorithm, as they need to be consequent with each other. Thus, the gradient can be approximated using forward differences, but as the divergence is the adjoint of the gradient, it must be approximated with backward differences.

3.4 Discussion

Chapter 4

Experiments and Applications

In the previous two chapters the mathematical and computational challenges of image reconstruction for CT have been discussed. In chapter 3, a detailed description of a variety of different algorithms has been presented, including the ART family of algorithms, CGLS and a few TV approaches for smooth reconstruction, as well as the classic FDK reconstruction. Additionally in chapter 4, the computational aspect of CT is discussed, where the problems computing the exact adjoint of the projection operation and mainly the computational burden of some of the operations have been mentioned. Considering the variety of available methods and the specifics of the implementation of the software developed, the TIGRE Toolbox, experiments on how these algorithms compare and behave are due. Furthermore, the performance of these algorithms with different experimental datasets is also an important analysis.

This chapter shows experimental analysis on both of the topics. First a variety of convergence analyses with different algorithms using synthetic data is performed, showing the differences not only between algorithms, but also between option on parameter selections. The section tries to illustrate and perhaps help build intuition into all the different parameters and options that each of these algorithms has, both within the algorithms themselves and among the different ones. Additionally some highlights on the practical challenges that the use of the algorithms entail in real applications are given.

In the second section of this chapter, a few examples of some of the algorithms are shown in different CT applications, both cone and parallel beam. Data from various different applications, from medicine to science has been tested using the TIGRE toolbox. While quantitative analysis is not possible with these datasets because the truth is not known, some insight in how the algorithms behave in each case is discussed.

4.1 Algorithm Experiments

This section explores a variety of algorithms and the parameters within them, and shows how they behave with different synthetic data in simulation studies.

4.1.1 Convergence Rates

In chapter 3 the convergence rates of the algorithms has been mentioned, as well as computational times. Different algorithms will reach different residuals at a given iteration and thus understanding which ones can converge faster and theoretically give a better result earlier is important. However, at the scale of the CBCT problem, faster no only means reaching a residual that is smaller in the same number of iterations, as the computational burden of each of the iterations also needs to be considered. And, as the backprojection operator is not exactly the adjoint of the projection operator, an effect that the classic formulation of these algorithms do not take into account can happen: divergence. All the algorithms (at least in this work) are mathematically designed to always reduce the residual each iteration, but that formulation relies on a correct adjoint operator. Thus, sometimes, when the algorithms in TIGRE find a solution very close to the minimum residual solution, they may diverge. The code in the toolbox does generally check for divergence and stop, but one of the effect that can be observed is that some algorithm will always diverge to yield a residual that is larger than others. This means that some algorithm can, regardless of their computational times, reach to a better solution than others.

All tests in this section are performed on the XCAT phantom[72], in a 128^3 voxel size and 256^2 detector. A different number of angles are used, always uniformly distributed around a full circle. Figure 4-1 shows cross sections of the phantom in its mid plane and figure 4-2 shows 3 projections of the phantom as simulated for the following tests.

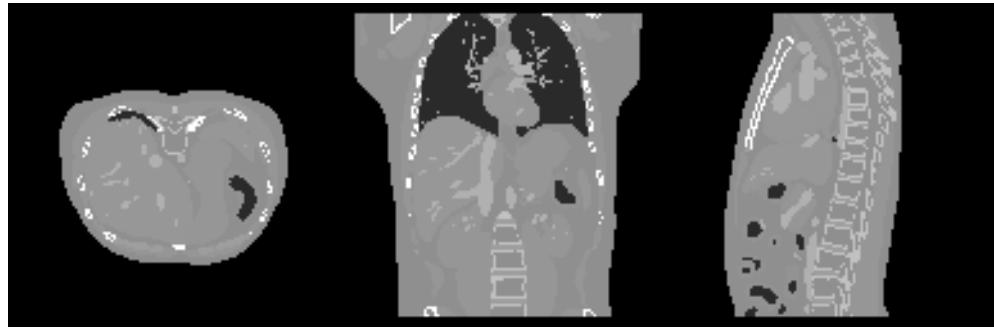


Figure 4-1: Cross section of the XCAT phantom in its mid plane in the three axes, for 128^3 voxels.



Figure 4-2: Simulate projections of the XCAT phantom at three different projection angles for a 256^2 detector.

Update Ordering in SART

An analysis of the different ART-type algorithms is presented in this section. One of the discussed parameters that has an effect in the convergence rate of the ART-type algorithms is the ordering of the projections used. Research has shown that in ART, the angle ordering can have an effect on the residual[38][91], however in the algorithms feasible for big scale tomography, this effect is smaller. Figure 4-3 shows the convergence of SART during 150 iterations using 100 projections as data. The same configuration of SART is run using ordered, randomly ordered and angular distance maximizing ordering schemes for the update order. While minor, the figure shows how random ordering does generally increases the convergence rate of the algorithm, at no computational cost. This is the default value in the software. Note that in this test there is no reduction of the relaxation parameter λ .

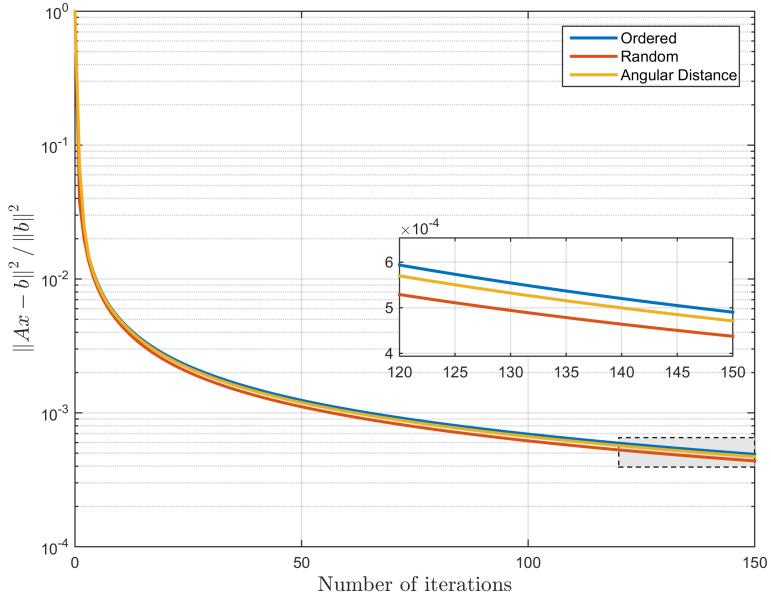


Figure 4-3: Normalized residual versus iteration of SART compared to different angle ordering schemes, using 100 projections and no relaxation parameter reduction

Comparison Between SART, OS-SART and SIRT

These algorithms have very different convergence, as updating the image per-projection angle has the effect of converging faster (in iteration number). However, the computational times are greatly reduced by updating more rows at the same time. This effect can be seen in figure 4-4, where the convergence versus iteration of these three algorithms is plotted. Note the convergence difference between SART and SIRT, where SIRT doesn't reach SART's residual even after 1000 iterations, however, each iteration of SIRT is two orders of magnitude faster than SART. OS-SART provides a middle ground alternative. Due to the specifics of the acceleration procedures for back-projection, OS-SART speeds are closer to SIRT than to SART (i.e., the speed does not change linearly with the image updates per iteration), however it is more prone to divergent behaviour in TIGRE. In the figure, OS-SART stops converging after 48 iterations. Of course, this behaviour is very data-specific, and there are multiple cases where it does not diverge. Figure 4-4 shows the result images of these three algorithms after 150 iterations (48 for OS-SART). Note that this example has limited data, so even in the best case, the images are slightly noisy.

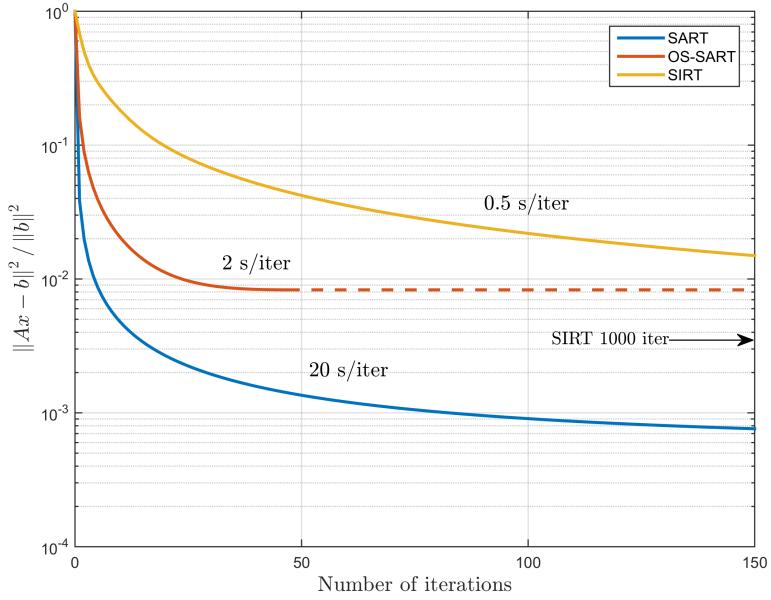


Figure 4-4: Normalized residual vs iteration number for SART, OS-SART and SIRT using 100 projections and no relaxation parameter reduction.

Relaxation Parameter

The choice of a proper relaxation parameter significantly changes the speed which a solution is found, and can avoid infinitely iterating through the same hyperplanes in case of an under determined or noisy solution. In TIGRE, two methods are implemented, as described in chapter 2: multiplying the relaxation parameter by a reduction factor after each iteration, and the Nesterov accelerated update, that does not technically update the relaxation parameter, but updates the image at each iteration using an iteration specific combination ratio of the gradients of the current and previous iterations. It requires more memory as it needs one extra image-sized variable to store the previous update, but the algorithm finds a solution considerably faster, as can be seen in figure 4-5. In the figure each of the SART, OS-SART and SIRT algorithms residuals is plotted, and in each of them three versions are displayed, no relaxation parameter update, reduction with $r_{red} = 0.99$ and the Nesterov update. In the plot it can also be seen that reducing the relaxation parameter by a ratio, while a good approach in SART-based hybrid algorithms such as the TV minimizing ones in TIGRE, leads to slower residual reduction and ultimately to a worse image.

In figure 4-6, the solution found by the three algorithms using reduction of the

relaxation parameter, using a Nesterov update and using a static relaxation parameter of $\lambda = 1$ can be seen side by side . The superior solution found by Nesterov is clear, and both SART and OS-SART reach a minimum in very few iterations. While SART does reach a better image (both in residual and error) without using Nesterov's update, the difference is minimal. It is important to note that using Nesterov's update, likely due to its fast convergence, leads to a faster divergent behaviour by the algorithms, thus the residual needs to be checked in each iteration leading to some computational overhead.

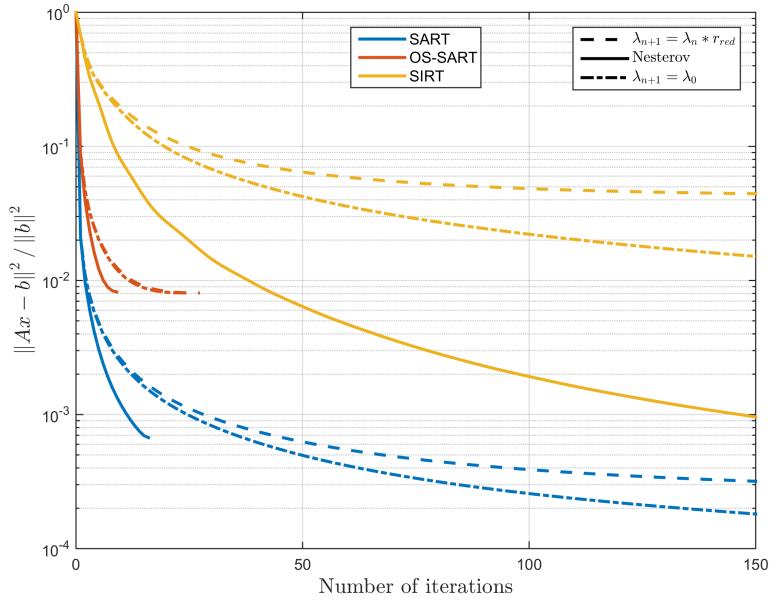


Figure 4-5: Normalized residual vs iteration number for SART, OS-SART and SIRT using 100 projections and different relaxation parameter reduction methods. If the relaxation parameter is reduced by a constant ratio, the residual reduction worsens, and if reduced using Nesterovs update, it converges very fast.

4.1.2 Total Variation Minimization

There are 4 total variation minimizing algorithms in TIGRE, with 2 different minimization functionals. As previously described, ASD-POCS, OS-ASD-POCS and B-ASD-POCS- β minimize the TV using a POCS minimization technique by minimizing the data constraint and TV norm independently using gradient descent. SART-TV however uses the ROF model for the TV-minimization step. The total variation algorithms are designed for applications where the image is piecewise-smooth, as they

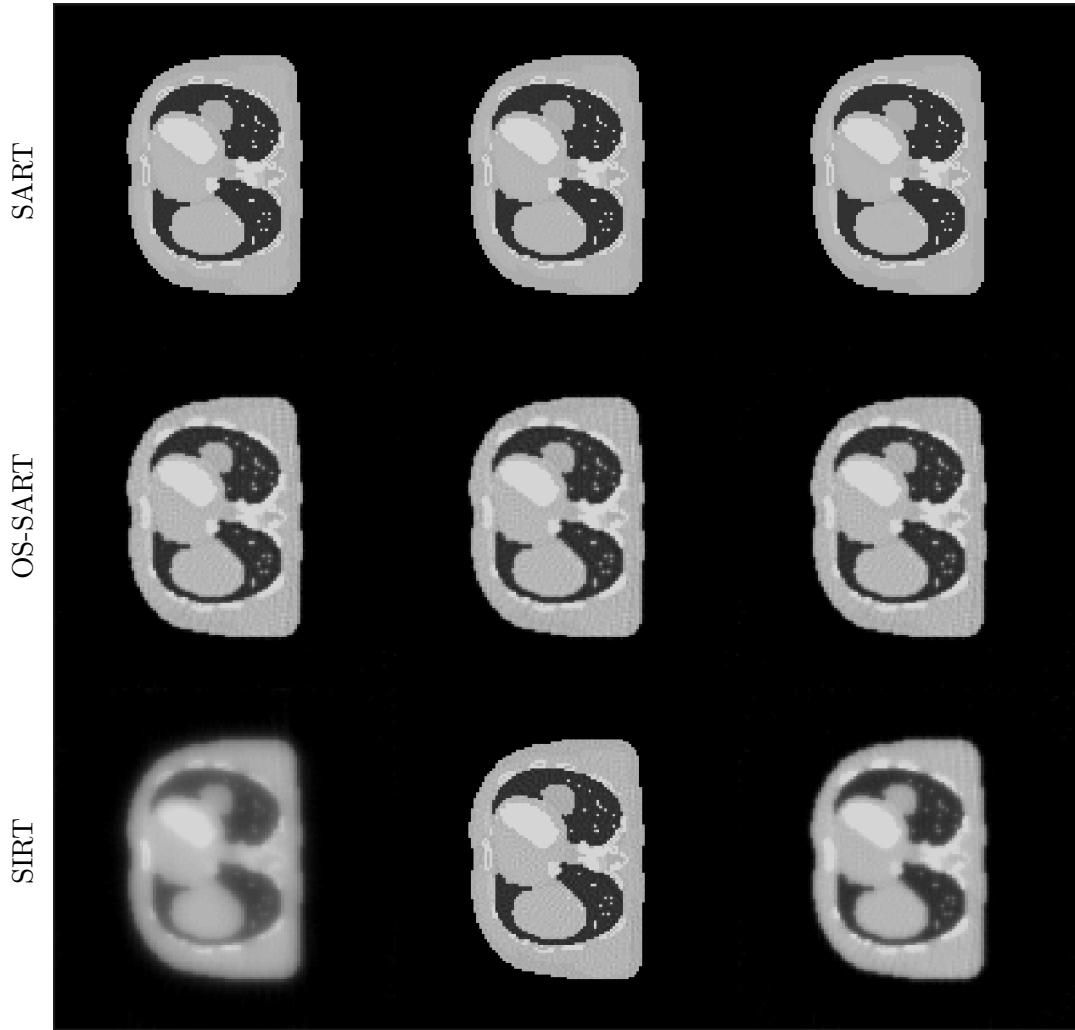


Figure 4-6: Reconstructed images with different relaxation parameter updates. OS-SART stops at iteration 24 in all but the Nesterov case, where it stops at iteration number 9. SART stops at iteration 16 for Nesterov.

Table 4.1: NRMSE for the reconstructed images in figure 4-7

	FDK	OS-SART	B-ASD-POCS- β	SART-TV	ASD-POCS	OS-ASD-POCS
NRMSE	0.1373	0.0678	0.0338	0.0267	0.0304	0.0442

will try to minimize the gradient, by creating single-valued regions in the image. In CT, the most noisy images are reconstructed when either the data is very noisy (generally due to small acquisition times and/or low energy X-rays) or when the data are limited, either due to limited angular range or more importantly a limited number of projections.

An example of the behaviour of the TV algorithms with the same dataset as in figures 4-1 and 4-2 is shown in figure 4-7. In this case, 30 uniformly sampled projections are used perturbed with Poisson and Gaussian noise to simulate photon scattering and electronic noise, respectively. The figure shows FDK and OS-SART reconstructions, and the 4 mentioned TV algorithms. It is clear that the TV algorithms do provide a smoother reconstruction, and with less normalized root mean squared error (NRMSE), as shown in table 4.1. The reconstruction by FDK is plagued with noise. And, while the main structural features can be seen, most of the detail is lost. Even the bones themselves are practically indistinguishable from noise. OS-SART does reconstruct a smoother image as expected, as it minimizes the 2-norm and, while one can see more details in the image, it is still poor. The 4 TV algorithms can be seen to flatten out the attenuation levels to similar values, thus reducing most of the noise. Additionally most of the features get clearly separated from the attenuation levels of the surrounding tissues and some of the algorithms (such as ASD-POCS) are able to reconstruct even single pixel width structures correctly. It is important to note that while the parameters used to tune this specific TV reconstruction (available in demo number 9 in the TIGRE Toolbox), they are far from optimal and very sensitive[56]. When choosing the exact optimal parameters for the TV reconstruction algorithms, the resultant images tend to be significantly better than the ones shown here, but the parameter space is very data dependent and large, thus to the author's knowledge, no parameter selection method has been proposed in the literature.

To illustrate the sensitivity to parameter selection, the algorithm SART-TV is run with 3 different values for the number of TV-iterations per SART iteration for the same data set used in the previous test. The results can be seen in figure 4-8, where one can clearly see how small changes can have a devastating effect in the output image. If a few more TV iterations are added to (b), the image gets a bit smoother and some detail is lost, as expected. However, if few iterations are removed from (b), the image

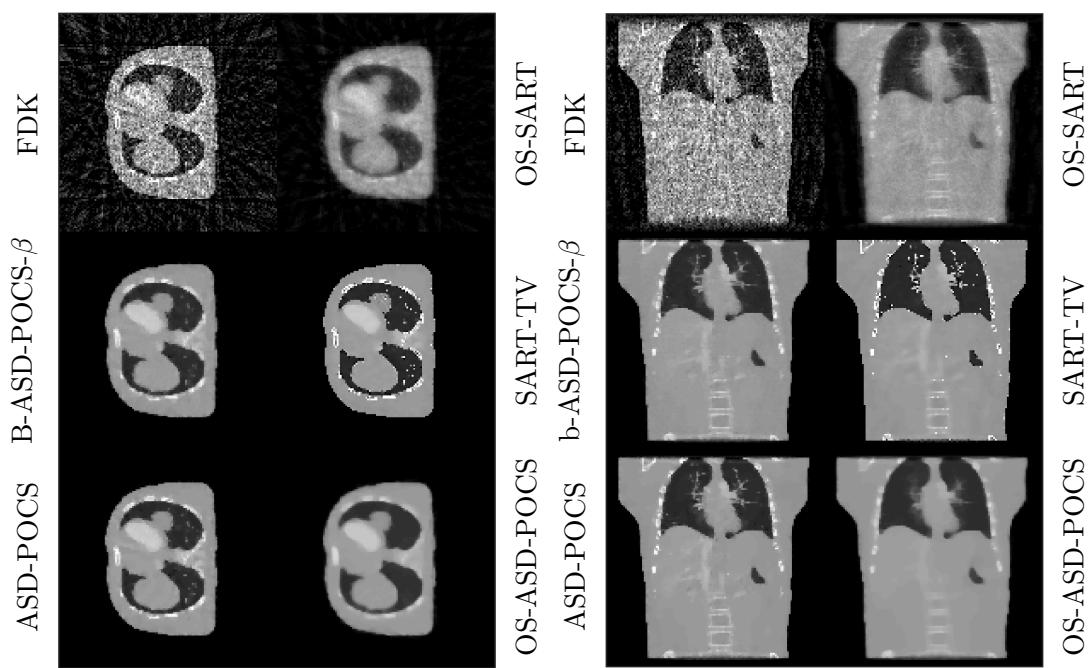


Figure 4-7: Reconstructed images using FDK, OS-SART and the TV algorithms b-ASD-POCS- β , SART-TV, ASD-POCS and OS-ASD-POCS with a limited amount and noisy data. Both figures show the same data and algorithms, but with a different cross-section of the image.

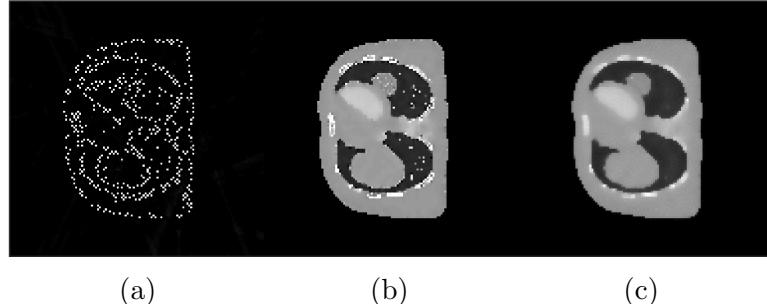


Figure 4-8: SART-TV algorithms with different amount of TV iterations per SART iteration, (a) 32 iterations, (b) 40 iterations, (c) 48 iterations.

can get completely destroyed. Note that these values are only applicable to this image with the exact amount of noise and projections. Different experiments may not show this behaviour or may be more intolerant to parameter change. This is arguably the biggest limitation for the common use of TV algorithms in real applications. As an advantageous point, once the good parameters are found, generally the algorithm will perform similarly for similar images, thus application specific parameters may be an option.

4.2 Iterative Algorithms in Different CT Applications

This section tries to illustrate the effect of different algorithms within the TIGRE Toolbox for a series of datasets.

4.2.1 Medical Head CBCT from The Christie Hospital

4.2.2 Cryo Soft X-Ray Tomography at the Diamond Light Source

Cryo soft X-ray tomography (Cryo-SXT) is a relatively new[20] technology to image micron size biological samples in full 3D. Generally, cell-imaging is performed with electron microscopy (EM) and all its variants (transmission electron microscopy, scanning electron microscopy, cryo-electron microscopy, electron tomography, etcetera), however these techniques have very limited penetration (less than $1\mu\text{m}$) and thus often require slicing of the samples for volumetric imaging. Cryo-SXT uses the so called water window for X-ray energies around the 500eV energy range. Unlike at higher energies, where everything is invisible, water becomes transparent but carbon-based tissues are clearly visible in that range. Thus, while with lower resolution than most EM, Cryo-SXT allows full volumetric visualization of the cells without damaging the samples. In

order to be able to image with an extremely accurate setup in both sample handling and X-ray parameters, these Cryo-SXT images are captured in synchrotron facilities. The data used in this work is from the B24 beam-line at the Diamond Light Source.

However, Cryo-SXT data has several sources of errors that make its reconstructed images significantly noisy. The typical penetration depth of soft X-rays is around $10\mu\text{m}$, while the samples are generally an order of magnitude bigger than that in height and width. Thus Cryo-SXT is a limited angle problem, where most of the datasets are sampled over a 120 degree arc. In the extrema of this range, the images in the detector tend to have little or no information for some parts of the sample due to photons not reaching the detector. Additionally, the low intensity and small sample size do mean that the detector data are very noisy, as photons spread out more (at the scale of the pixel dimension) and fewer photons reach the detector. The size of the sample also comes with errors in the mechanical systems of the imaging set up. When working on a scale of microns, any small vibration is visible and considerably perturbs the measured data. Generally these types of errors are removed by pre-processing using alignment techniques, but the algorithms involved are often not fault proof and the data used in reconstruction ends up having some misalignment errors. Figure 4-9 shows two of the sinograms of the datasets, where the noisy nature of the data can be intermediately appreciated. In the top figure, attenuation artefacts are visible. In the bottom figure, one can see the darkening of the areas at high angles (upper and lower parts of the figure, at 25% of distance from the left) and areas that have been filled by the alignment algorithm with a single value (mid-left edge and bottom right edge). These last errors do have no influence in filtered backprojection (FBP), as the high-pass filtering of the data sets their values to zero, but they are a source of artefacts in iterative algorithms. Both sinograms show a considerably high amount of random noise.

A few datasets have been reconstructed from this imaging modality using various algorithms. Objective evaluation of the quality of the reconstructed image with each algorithm is not possible as, due to the noisy nature of the images, classifying some of the visual artefacts as data or noise is hard. Thus this section does not intend to claim that any of the algorithms perform better than FBP, just highlight the differences.

The “2017_0207_Trypanosoma_33” Dataset.

This dataset contains a section of an image containing a few Trypanosoma, a unicellular parasitic protozoa that cause different illnesses, such as the sleeping sickness. In the images, the big blob within each of them of similar attenuation level as the rest of the cell is the nucleus, while the smaller circular features are organelles of the cell.

Several algorithms have been used to test the effect of iterative algorithms. Initially

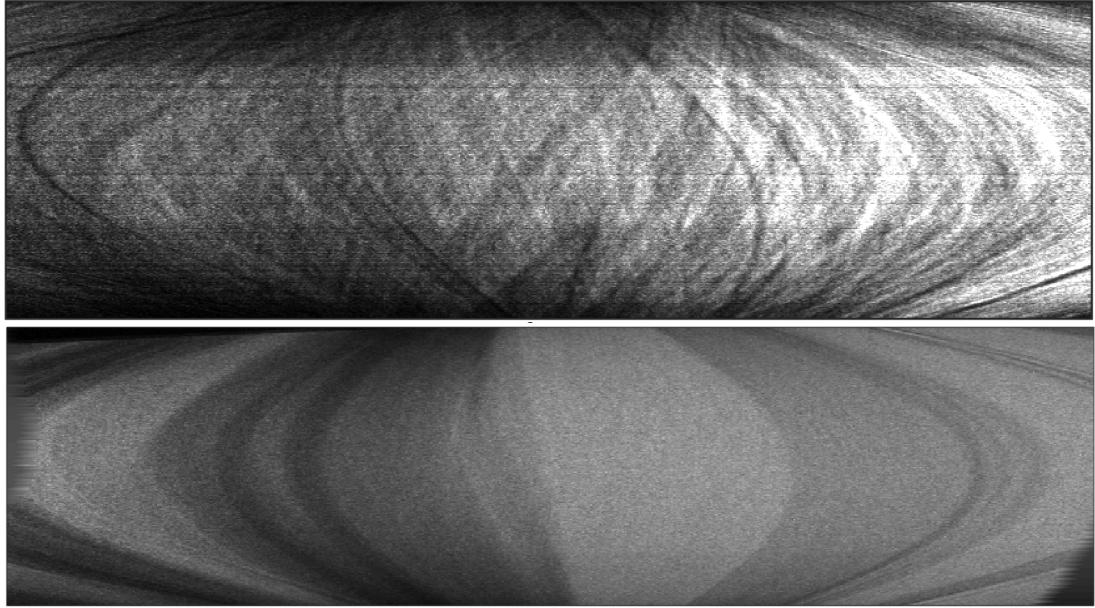


Figure 4-9: Sinograms of two different datasets of the data from a Cryo-SXT. The top figure shows attenuation artefacts over different individual measurements of the data and significant random noise over the whole sinogram. The bottom figure shows strong attenuation at high angles and artefacts generated by alignment.

SIRT and CGLS have been chosen. Both of these algorithms are expected to generate images with very little noise, but perhaps lose the most detailed information, or at least create smoother boundaries than FBP, as they minimize the L₂ norm using all data in one go (per iteration). The result of these, compared to FBP can be seen in figure 4-10. SIRT generates a very smooth image without barely any noise, however the details are very smoothed also, specially the boundaries of the objects. However, very few iterations of SIRT have been performed in this dataset. CGLS however seems to separate data from noise better, while also creating some smooth (not as much as SIRT) boundaries. However it is unclear how much of this is caused by misalignments within the data and how much by the algorithms themselves.

As shown earlier in this chapter, OS-SART can improve the convergence speed of SIRT sacrificing some computational time. SART may improve further the convergence, but it becomes a very computational expensive algorithm at this image sizes. Additionally, total variation minimization can be applied to remove the noise that the images have. Figure 4-11 shows FBP, OS-SART, and ASD-POCS with 20 total variation iterations and using OS-SART instead of SART as data fidelity update (OS-ASD-POCS).

Note that the images have a darker vertical “band”, not as obvious in SIRT and CGLS. This is caused by some errors in the projections that with a proper preprocessing step could be removed. OS-SART reconstructs a similar result to FBP, with slightly lower noise levels and extreme values. Fewer iterations of OS-SART would probably generate a less noisy image, however they will also likely show less contrast and features (similar to SIRT before). The total variation version of OS-SART generates a cleaner image, but it has a slight “watercolour” texture. The strength of the total variation can be controlled by the number of iterations, increasing them enhances this effect. Figure 4-12 shows 0, 20 and 5 TV iterations. While the tuning of this value can not (yet) be done automatically, once the desired one is found it generally works for all similar images. The actual range of values of the reconstructed voxels is different in each algorithm. This can be explained by the nature of the data, as for example, half of the values are negative, which makes no sense physically. Thus, the visualization range has been adjusted to match histograms of attenuation vale in the figures.

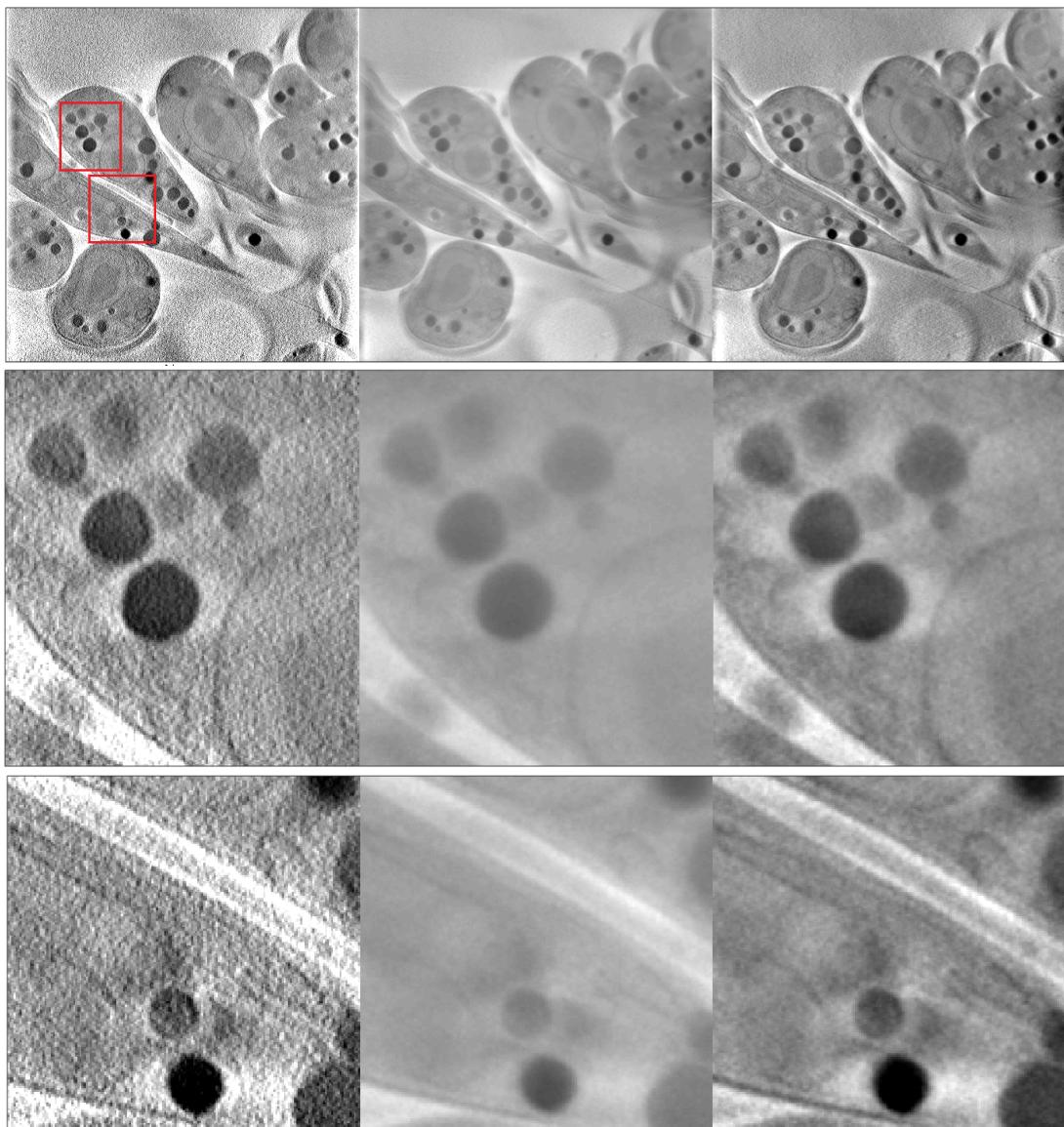


Figure 4-10: Columns: FBP, SIRT (20 iterations) and CGLS (7 iterations). The red squares in the first row show the location of the zoomed-in areas from the second and third row.

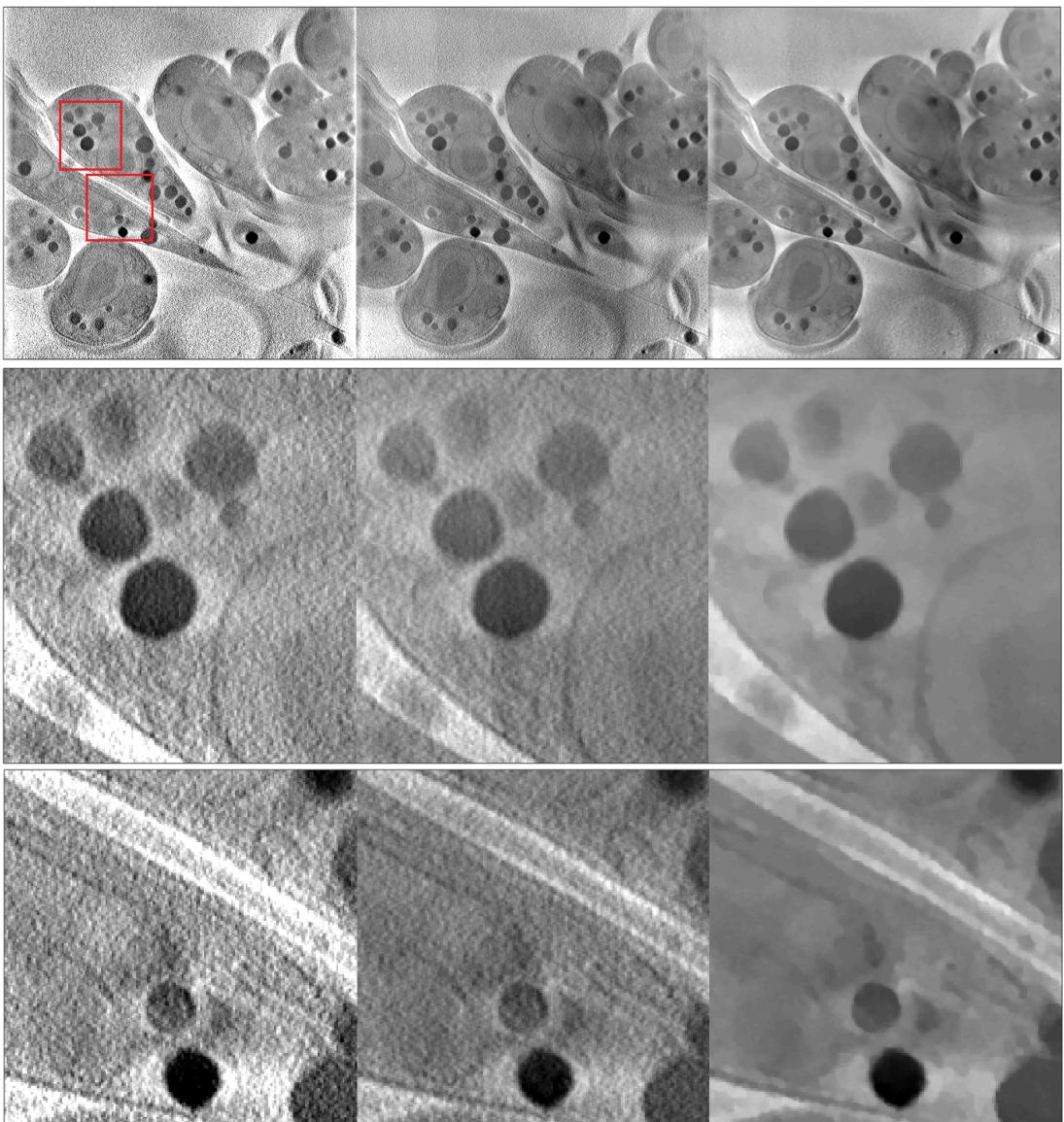


Figure 4-11: Columns: FBP, OS-SART (20 iterations) and OS-ASD-POCS (20 iterations, 20 TV iterations each). The red squares in the first row show the location of the zoomed-in areas from the second and third row.

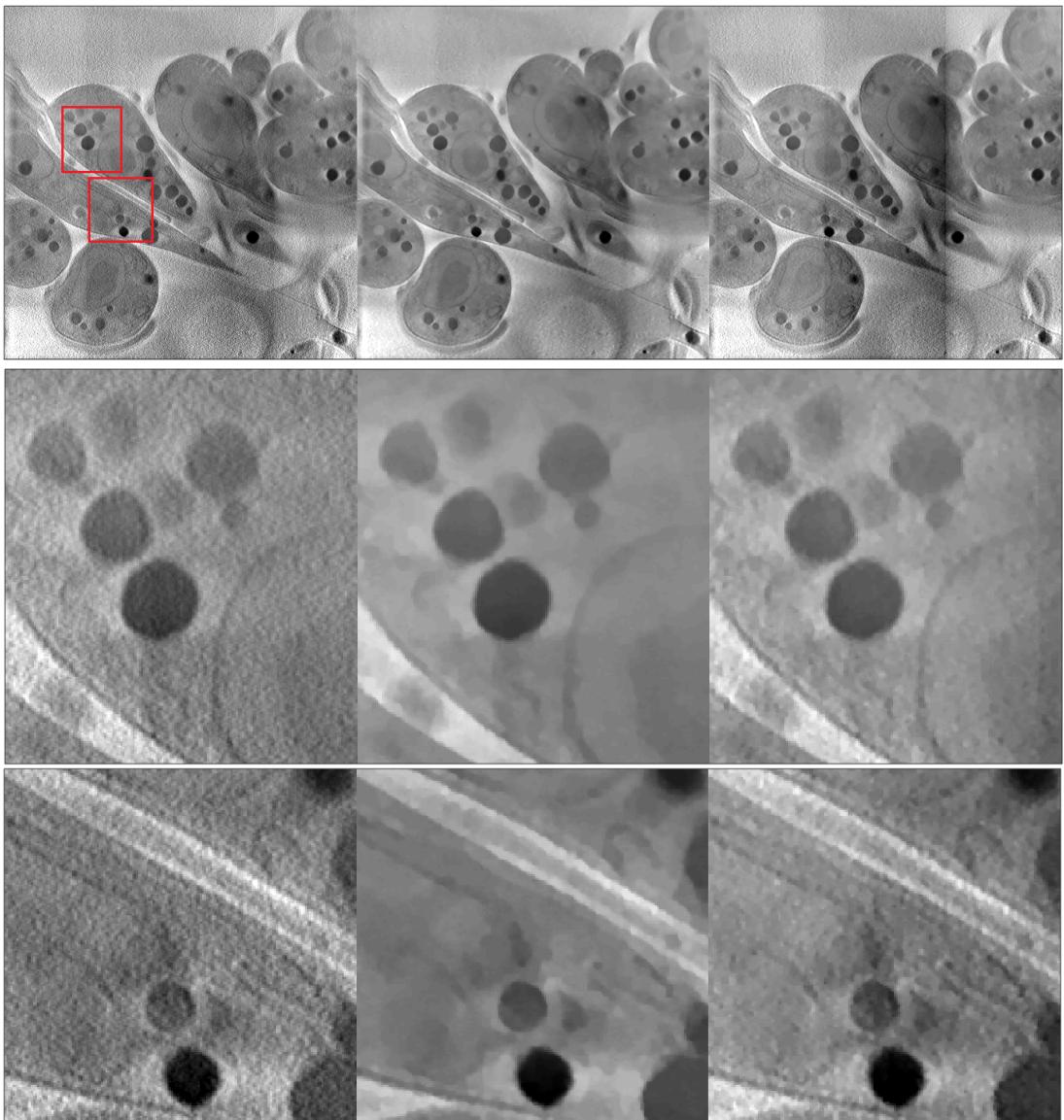


Figure 4-12: Columns: OS-SART (20 iterations, 0 TV iterations), OS-ASD-POCS (20 iterations, 20 TV iterations each), OS-ASD-POCS (20 iterations, 5 TV iterations each). The red squares in the first row show the location of the zoomed-in areas from the second and third row.

The “3_20160218_tomo_65t55_p5_area_2MB1_Export” Dataset.

This dataset contains, as described by Luengo *et al*[57] the zoomed area of a “neuronal-like mammalian cell line (PC-12[5]).” The article has more information on the preparation of the samples.

The big smooth area in the top left side is the nucleus of the cell, while the rest are organelles on the cytoplasm of the cell. In figure 4-13 the reconstructed image can be seen, on where the columns show FBP, OS-SART, CGLS and OS-ASD-POCS algorithms, and the rows different zoomed areas of the image. Due to hinger noise in the projections, the iterative algorithms to have a strong influence in the removal of the noise in this dataset. This is clearly apparent in the second row of the figure 4-13, on where the three iterative algorithms, specially the TV based one, remove significantly the noise of the organelles both in the left and right side of the image.

The “Grid1_Area2_Cell2_tomo3-All_60t60_p5d_6s_mb1_Export” Dataset

This images also show a PC-12 cell, with part nucleus and part organelles, as in the previous dataset. The same algorithms have been used, and the results can be seen in figure 4-14.

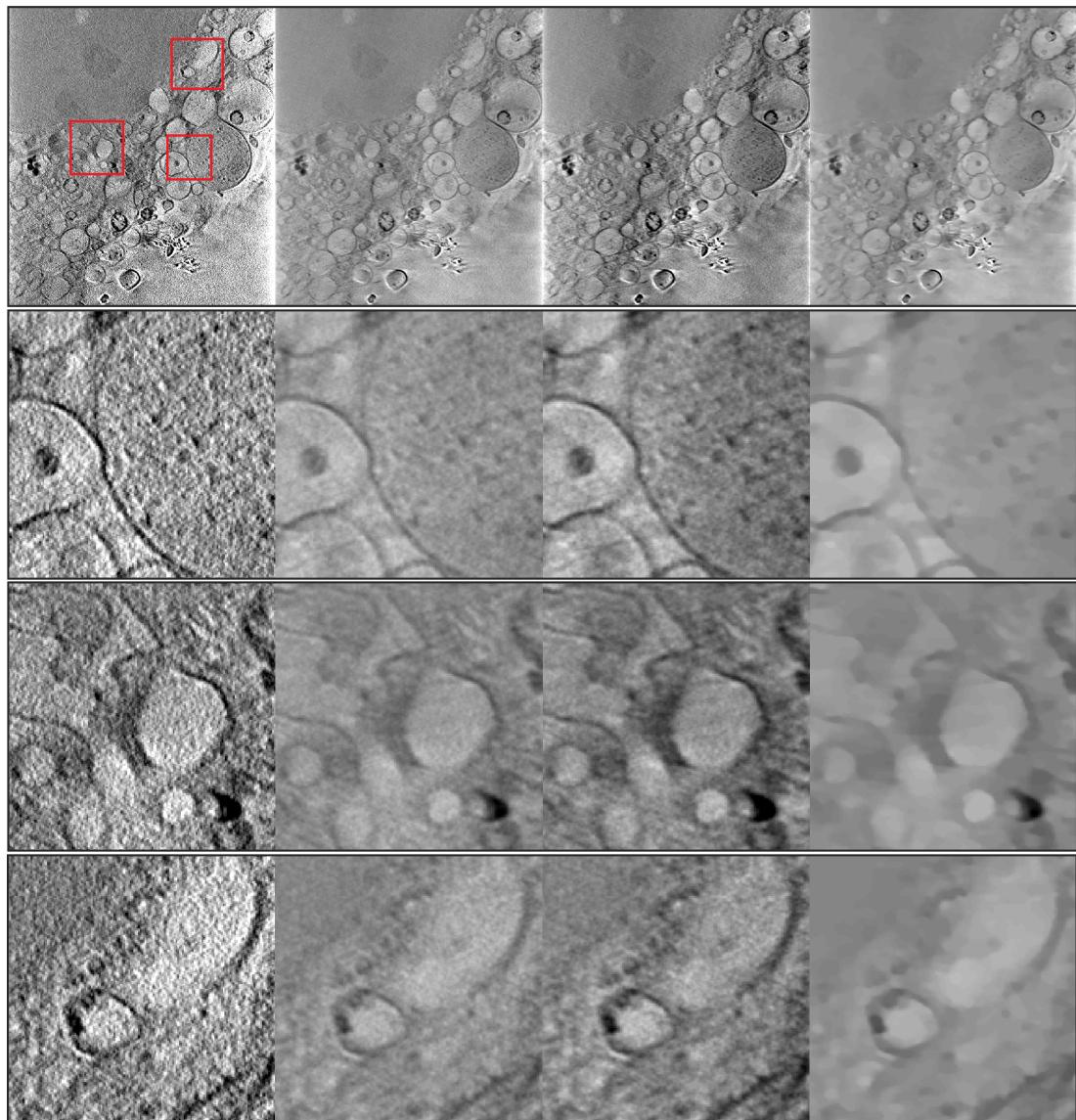


Figure 4-13: Columns: FBP, OS-SART (30 iterations), CGLS (7 iterations) and OS-ASD-POCS (30 iterations, 10 TV iterations each). The zoomed areas are highlighted in the FBP image. The red squares in the first row show the location of the zoomed-in areas from the second third and fourth row.

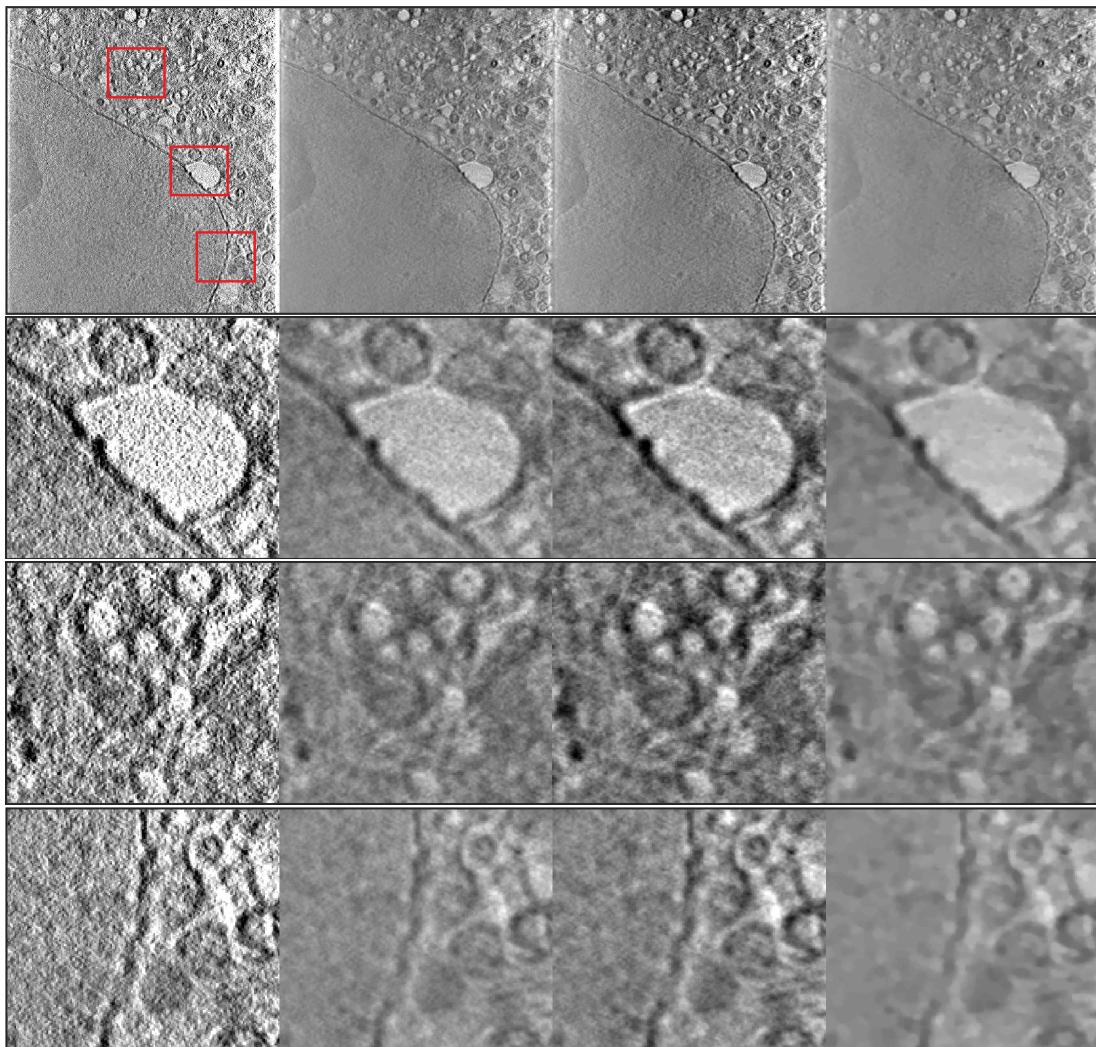


Figure 4-14: Columns: FBP, OS-SART (30 iterations), CGLS (7 iterations) and OS-ASD-POCS (30 iterations, 10 TV iterations each). The zoomed areas are highlighted in the FBP image. The red squares in the first row show the location of the zoomed-in areas from the second third and fourth row.

4.3 Discussion

Chapter 5

Conclusions and Future Work

Bibliography

- [1] CERN phase space tomography. <http://cern.ch/tomography>. Accessed: 2017-10-19.
- [2] TIGRE Github repository. <https://github.com/CERN/TIGRE>. Accessed: 2016-0912-19.
- [3] A.H. Andersen and A.C. Kak. Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm. *Ultrasonic imaging*, 6(1):81–94, 1984.
- [4] Carol DeSantis ans Rebecca Siegel and Ahmedin Jemal. Cancer Treatment and Survivorship Facts and Figures: 2014-2015 . Available from: <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/cancer-treatment-and-survivorship-facts-and-figures/cancer-treatment-and-survivorship-facts-and-figures-2014-2015.pdf>, accessed on 31/10/2017., 2015. Cancer.org.
- [5] Barbara L Apostol, Alexsey Kazantsev, Simona Raffioni, Katalin Illes, Judit Pallos, Laszlo Bodai, Natalia Slepko, James E Bear, Frank B Gertler, Steven Hersch, et al. A cell-based assay for aggregation inhibitors as therapeutics of polyglutamine-repeat disease and validation in Drosophila. *Proceedings of the National Academy of Sciences*, 100(10):5950–5955, 2003.
- [6] Ander Biguri, Manjit Dosanjh, Steven Hancock, and Manuchehr Soleimani. TIGRE: a MATLAB-GPU toolbox for CBCT image reconstruction. *Biomedical Physics & Engineering Express*, 2(5):055010, 2016.
- [7] Ander Biguri, Manjit Dosanjh, Steven Hancock, and Manuchehr Soleimani. TIGRE: Tomographic Iterative GPU-based Reconstruction toolbox, September 2016.

- [8] Ander Biguri, Manjit Dosanjh, Steven Hancock, and Manuchehr Soleimani. A general method for motion compensation in x-ray computed tomography. *Physics in Medicine and Biology*, 2017.
- [9] Ander Biguri, Bartlomij Grychtol, Andy Adler, and Manuchehr Soleimani. Tracking boundary movement and exterior shape modelling in lung EIT imaging. *Physiological Measurement*, 36(6):1119, 2015.
- [10] Ander Biguri, S Hancock, M Dosanjh, and Manuchehr Soleimani. Motion correction in x-ray tomography using a priori known deformation vector fields and iterative reconstruction methods. In *Enlight Annual Meeting 2016 and Training Event*. University of Bath, 2016.
- [11] Ander Biguri, S Hancock, M Dosanjh, and Manuchehr Soleimani. Improvement of image quality in 4d-cbct respiratory correlated and motion-compensated reconstruction using iterative algorithms and gpu acceleration. In *BIGART 2017*, 2017.
- [12] Ander Biguri and Manuchehr Soleimani. 4D FEM models of the human thorax. In *EIT2015*. University of Bath, 2015.
- [13] Ander Biguri and Manuchehr Soleimani. Dual modality EIT-CBCT for lung radiation therapy. In *ENLIGHT 2014*. University of Bath, 2015.
- [14] Ander Biguri and Manuchehr Soleimani. Electrode movement due to breathing in lung eit imaging. In *Applied Inverse Problem AIP 2015*. University of Bath, 2015.
- [15] Ander Biguri and Manuchehr Soleimani. Statistical and deterministic approaches for electrode movement in lung EIT. In *EIT2015*. University of Bath, 2015.
- [16] R Scott Brock, Alen Docef, and Martin J Murphy. Reconstruction of a cone-beam ct image via forward iterative projection matching. *Medical physics*, 37(12):6212–6220, 2010.
- [17] J. F. Cai, X. Jia, H. Gao, S. B. Jiang, Z. Shen, and H. Zhao. Cine cone beam ct reconstruction using low-rank matrix factorization: Algorithm and a proof-of-principle study. *IEEE Transactions on Medical Imaging*, 33(8):1581–1591, Aug 2014.
- [18] Cancer research UK. Lung cancer survival statistics. Cancer Research UK web page: survival statistics, 2014.

- [19] Cancer research UK. Types of treatment for lung cancer. Cancer Research UK web page: treatment, 2014.
- [20] Raffaella Carzaniga, Marie-Charlotte Domart, Lucy M Collinson, and Elizabeth Duke. Cryo-soft x-ray tomography: a journey into the world of the native-state cell. *Protoplasma*, 251(2):449–458, 2014.
- [21] Y. Censor and T. Elfving. Block-iterative algorithms with diagonally scaled oblique projections for the linear feasibility problem. *SIAM Journal on Matrix Analysis and Applications*, 24(1):40–58, 2002.
- [22] Antonin Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical imaging and vision*, 20(1):89–97, 2004.
- [23] Guang-Hong Chen, Jie Tang, and Shuai Leng. Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets. *Medical physics*, 35(2):660–663, 2008.
- [24] Guang-Hong Chen, Pascal Thériault-Lauzier, Jie Tang, Brian Nett, Shuai Leng, Joseph Zambelli, Zhihua Qi, Nicholas Bevins, Amish Raval, Scott Reeder, et al. Time-resolved interventional cardiac c-arm cone-beam ct: an application of the piccs algorithm. *IEEE transactions on medical imaging*, 31(4):907–923, 2012.
- [25] Christian PV Christoffersen, David Hansen, Per Poulsen, and Thomas Sangild Sørensen. Registration-based reconstruction of four-dimensional cone beam computed tomography. *IEEE transactions on medical imaging*, 32(11):2064–2077, 2013.
- [26] Gianfranco Cimmino. Calcolo approssimato per le soluzioni dei sistemi di equazioni lineari. *La Ricerca Scientifica (Roma)*, 1:326–333, 1938.
- [27] Sophia B. Coban and William R.B. Lionheart. Regularised GMRES-type methods for X-ray computed tomography. In *The third international conference on image formation in X-ray computed tomography*, page 194, 2014.
- [28] Patrick L Combettes and Valérie R Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Modeling & Simulation*, 4(4):1168–1200, 2005.
- [29] Robert L. Comis. A brief history of the research and treatment of lung cancer from 1970 to 2000. *International Journal of Clinical Oncology*, 8(4):230–233, 2003.
- [30] Manjit Dosanjh. From particle physics to medical applications.

- [31] Joan Duran, Bartomeu Coll, and Catalina Sbert. Chambolle’s projection algorithm for total variation denoising. *Image processing on Line*, 2013:311–331, 2013.
- [32] J Ferlay, I Soerjomataram, M Ervik, R Dikshit, S Eser, C Mathers, M Rebelo, DM Parkin, D Forman, and F Bray. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 . Available from: <http://globocan.iarc.fr>, accessed on day/month/year., 2010. Lyon, France: International Agency for Research on Cancer.
- [33] Hao Gao, Jian-Feng Cai, Zuwei Shen, and Hongkai Zhao. Robust principal component analysis-based four-dimensional computed tomography. *Physics in Medicine and Biology*, 56(11):3181, 2011.
- [34] Lucas L Geyer, U Joseph Schoepf, Felix G Meinel, John W Nance Jr, Gorka Bastarrika, Jonathon A Leipsic, Narinder S Paul, Marco Rengo, Andrea Laghi, and Carlo N De Cecco. State of the art: iterative CT reconstruction techniques. *Radiology*, 276(2):339–357, 2015.
- [35] S Hancock, M Lindroos, and S Koscielniak. Longitudinal phase space tomography with space charge. *Physical Review Special Topics-Accelerators and Beams*, 3(12):124202, 2000.
- [36] S Hancock, M Lindroos, E McIntosh, and Mike Metcalf. Tomographic measurements of longitudinal phase space density. *Computer Physics Communications*, 118(1):61–70, 1999.
- [37] G. T. Herman and L. B. Meyer. Algebraic reconstruction techniques can be made computationally efficient [positron emission tomography application]. *IEEE Transactions on Medical Imaging*, 12(3):600–609, Sep 1993.
- [38] Gabor T Herman and Lorraine B Meyer. Algebraic reconstruction techniques can be made computationally efficient (positron emission tomography application). *IEEE transactions on medical imaging*, 12(3):600–609, 1993.
- [39] L Horn, W. Pao, and D.H. Johnson. *Harrison’s Principles of Internal Medicine “Chapter 89”*. McGraw-Hill, 18 edition, 2012.
- [40] Varian Medical Systems Inc. Iterative cone beam CT, 2017.
- [41] Xun Jia, Yifei Lou, John Lewis, Ruijiang Li, Xuejun Gu, Chunhua Men, William Y Song, and Steve B Jiang. Gpu-based fast low-dose cone beam ct reconstruction via total variation. *Journal of X-ray science and technology*, 19(2):139–154, 2011.

- [42] Xun Jia, Zhen Tian, Yifei Lou, Jan-Jakob Sonke, and Steve B Jiang. Four-dimensional cone beam CT reconstruction and enhancement using a temporal nonlocal means method. *Medical physics*, 39(9):5592–5602, 2012.
- [43] S. Kaczmarz. Angenäherte Auflösung von Systemen linearer Gleichungen. *Bulletin International de l'Académie Polonaise des Sciences et des Lettres*, 35:355–357, 1937.
- [44] Avinash C. Kak, Malcolm Slaney, IEEE Engineering in Medicine, and Biology Society. *Principles of computerized tomographic imaging*. IEEE Press, New York, 1988. Published under the sponsorship of the IEEE Engineering in Medicine and Biology Society.
- [45] Florian Knoll, Markus Unger, Clemens Diwoky, Christian Clason, Thomas Pock, and Rudolf Stollberger. Fast reduction of undersampling artifacts in radial MR angiography with 3D total variation on graphics hardware. *Magnetic resonance materials in physics, biology and medicine*, 23(2):103–114, 2010.
- [46] Shuai Leng, Jie Tang, Joseph Zambelli, Brian Nett, Ranjini Tolakanahalli, and Guang-Hong Chen. High temporal resolution and streak-free four-dimensional cone-beam computed tomography. *Physics in Medicine and Biology*, 53(20):5653, 2008.
- [47] Ruijiang Li, Xun Jia, John H. Lewis, Xuejun Gu, Michael Folkerts, Chunhua Men, and Steve B. Jiang. Real-time volumetric image reconstruction and 3d tumor localization based on a single x-ray projection image for lung cancer radiotherapy. *Medical Physics*, 37(6), 2010.
- [48] Ruijiang Li, John H. Lewis, Xun Jia, Xuejun Gu, Michael Folkerts, Chunhua Men, William Y. Song, and Steve B. Jiang. 3d tumor localization through real-time volumetric x-ray imaging for lung cancer radiotherapy. *Medical Physics*, 38(5), 2011.
- [49] Tianfang Li, Lei Xing, Peter Munro, Christopher McGuinness, Ming Chao, Yong Yang, Bill Loo, and Albert Koong. Four-dimensional cone-beam computed tomography using an on-board imager. *Medical physics*, 33(10):3825–3833, 2006.
- [50] Jörg Liesen and Petr Tichý. Convergence analysis of krylov subspace methods. *GAMM-Mitteilungen*, 27(2):153–173, 2004.
- [51] Ji Liu and Stephen Wright. An accelerated randomized kaczmarz algorithm. *Mathematics of Computation*, 85(297):153–178, 2016.

- [52] Jiulong Liu, Xue Zhang, Xiaoqun Zhang, Hongkai Zhao, Yu Gao, David Thomas, Daniel A Low, and Hao Gao. 5d respiratory motion model based image reconstruction algorithm for 4d cone-beam computed tomography. *Inverse Problems*, 31(11):115007, 2015.
- [53] Li Liu, Weikai Lin, and Mingwu Jin. Reconstruction of sparse-view x-ray computed tomography using adaptive iterative algorithms. *Computers in biology and medicine*, 56:97–106, 2015.
- [54] Li Liu, Zhaofang Yin, and Xueyun Ma. Nonparametric optimization of constrained total variation for tomography reconstruction. *Computers in biology and medicine*, 43(12):2163–2176, 2013.
- [55] Yan Liu, Jianhua Ma, Yi Fan, and Zhengrong Liang. Adaptive-weighted total variation minimization for sparse data toward low-dose x-ray computed tomography image reconstruction. *Physics in medicine and biology*, 57(23):7923, 2012.
- [56] Manasavee Lohvithee, Ander Biguri, and Manuchehr Soleimani. Parameter selection in limited data cone-beam ct reconstruction using edge-preserving total variation algorithms. *Physics in Medicine and Biology*, 2017.
- [57] Imanol Luengo, Michele C. Darrow, Matthew C. Spink, Ying Sun, Wei Dai, Cynthia Y. He, Wah Chiu, Tony Pridmore, Alun W. Ashton, Elizabeth M.H. Duke, Mark Basham, and Andrew P. French. SuRVoS: Super-region volume segmentation workbench. *Journal of Structural Biology*, 198(1):43 – 53, 2017.
- [58] Cyril Mory, Vincent Auvray, Bo Zhang, Michael Grass, Dirk Schfer, S. James Chen, John D. Carroll, Simon Rit, Franoise Peyrin, Philippe Douek, and Loc Boussel. Cardiac c-arm computed tomography using a 3d + time roi reconstruction method with spatial and temporal regularization. *Medical Physics*, 41(2), 2014.
- [59] Cyril Mory, Guillaume Janssens, and Simon Rit. Motion-aware temporal regularization for improved 4d cone-beam computed tomography. *Physics in Medicine and Biology*, 61(18):6856, 2016.
- [60] Nature. www.nature.com/Search: lung cancer, 2015.
- [61] Yurii Nesterov. A method of solving a convex programming problem with convergence rate $O(1/k^2)$. In *Soviet Mathematics Doklady*, volume 27, pages 372–376, 1983.

- [62] P. Paysan, M. BREHM, A. Wang, D. Seghers, and J. Star-Lack. Iterative image reconstruction in image-guided radiation therapy, April 20 2017. WO Patent App. PCT/US2016/056,539.
- [63] T. Pengpan, W. Qiu, N.D. Smith, and M. Soleimani. Cone beam {CT} using motion-compensated algebraic reconstruction methods with limited data. *Computer Methods and Programs in Biomedicine*, 105(3):246 – 256, 2012.
- [64] Thomas G. Purdie, Douglas J. Moseley, Jean-Pierre Bissonnette, Michael B. Sharpe, Kevin Franks, Andrea Bezjak, and David A. Jaffray. Respiration correlated cone-beam computed tomography and 4DCT for evaluating target motion in stereotactic lung radiation therapy. *Acta Oncologica*, 45(7):915–922, 2006.
- [65] Wei Qiu, David Titley-Péloquin, and Manuchehr Soleimani. Blockwise conjugate gradient methods for image reconstruction in volumetric ct. *Computer methods and programs in biomedicine*, 108(2):669–678, 2012.
- [66] Lei Ren, Indrin J. Chetty, Junan Zhang, Jian-Yue Jin, Q. Jackie Wu, Hui Yan, David M. Brizel, W. Robert Lee, Benjamin Movsas, and Fang-Fang Yin. Development and clinical evaluation of a three-dimensional cone-beam computed tomography estimation method using a deformation field map. *International Journal of Radiation Oncology*Biology*Physics*, 82(5):1584 – 1593, 2012.
- [67] S. Rit, D. Sarut, and L. Desbat. Comparison of analytic and algebraic methods for motion-compensated cone-beam CT reconstruction of the thorax. *IEEE Transactions on Medical Imaging*, 28(10):1513–1525, Oct 2009.
- [68] Simon Rit, Jochem W. H. Wolthaus, Marcel van Herk, and Jan-Jakob Sonke. On-the-fly motion-compensated cone-beam CT using an a priori model of the respiratory motion. *Medical Physics*, 36(6):2283–2296, 2009.
- [69] Ludwig Ritschl, Stefan Sawall, Michael Knaup, Andreas Hess, and Marc Kachelrie. Iterative 4d cardiac micro-ct image reconstruction using an adaptive spatio-temporal sparsity prior. *Physics in Medicine and Biology*, 57(6):1517, 2012.
- [70] Sebastian Ruder. An overview of gradient descent optimization algorithms. *CoRR*, abs/1609.04747, 2016.
- [71] Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259 – 268, 1992.

- [72] WP Segars, G Sturgeon, S Mendonca, Jason Grimes, and Benjamin MW Tsui. 4D XCAT phantom for multimodality imaging research. *Medical physics*, 37(9):4902–4915, 2010.
- [73] Jonathan Richard Shewchuk et al. An introduction to the conjugate gradient method without the agonizing pain, 1994.
- [74] E.Y. Sidky and X. Pan. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Physics in Medicine and Biology*, 53(17):4777, 2008.
- [75] Jan-Jakob Sonke, Lambert Zijp, Peter Remeijer, and Marcel van Herk. Respiratory correlated cone beam ct. *Medical physics*, 32(4):1176–1186, 2005.
- [76] Thomas Strohmer and Roman Vershynin. A randomized kaczmarz algorithm with exponential convergence. *Journal of Fourier Analysis and Applications*, 15(2):262–278, 2009.
- [77] Ilya Sutskever, James Martens, George Dahl, and Geoffrey Hinton. On the importance of initialization and momentum in deep learning. In *International conference on machine learning*, pages 1139–1147, 2013.
- [78] Ricky T O’Brien, Benjamin J Cooper, Chun-Chien Shieh, Uros Stankovic, Paul J Keall, and Jan-Jakob Sonke. The first implementation of respiratory triggered 4DCBCT on a linear accelerator. *Physics in medicine and biology*, 61(9):3488, 2016.
- [79] Zhen Tian, Xun Jia, Kehong Yuan, Tinsu Pan, and Steve B Jiang. Low-dose CT reconstruction via edge-preserving total variation regularization. *Physics in medicine and biology*, 56(18):5949, 2011.
- [80] Henk A Van der Vorst. Bi-CGSTAB: A fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems. *SIAM Journal on scientific and Statistical Computing*, 13(2):631–644, 1992.
- [81] Curtis R Vogel and Mary E Oman. Iterative methods for total variation denoising. *SIAM Journal on Scientific Computing*, 17(1):227–238, 1996.
- [82] Jing Wang and Xuejun Gu. High-quality four-dimensional cone-beam ct by deforming prior images. *Physics in medicine and biology*, 58(2):231, 2012.
- [83] Wikipedia. Bragg peak, 2017. [Online; accessed 31-October-2017].

- [84] World Health Organization. *World Cancer Report 2014*. Lyon, France: International Agency for Research on Cancer, 2014.
- [85] Fang Xu and K. Mueller. A comparative study of popular interpolation and integration methods for use in computed tomography. In *3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006.*, pages 1252–1255, April 2006.
- [86] H. Xue, L. Zhang, Z. Cheng, Y. Xing, and Y. Xiao. An improved TV minimization algorithm for incomplete data problem in computer tomography. In *Nuclear Science Symposium Conference Record (NSS/MIC), 2010 IEEE*, pages 2621–2624, Oct 2010.
- [87] Qinghui Zhang, Yu-Chi Hu, Fenghong Liu, Karyn Goodman, Kenneth E Rosenzweig, and Gig S Mageras. Correction of motion artifacts in cone-beam ct using a patient-specific respiratory motion model. *Medical physics*, 37(6):2901–2909, 2010.
- [88] Zichun Zhong, Xuejun Gu, Weihua Mao, and Jing Wang. 4d cone-beam ct reconstruction using multi-organ meshes for sliding motion modeling. *Physics in Medicine and Biology*, 61(3):996, 2016.
- [89] Mingqiang Zhu and Tony Chan. An efficient primal-dual hybrid gradient algorithm for total variation image restoration.
- [90] Mingqiang Zhu, Stephen J Wright, and Tony F Chan. Duality-based algorithms for total-variation-regularized image restoration. *Computational Optimization and Applications*, 47(3):377–400, 2010.
- [91] Anastasios Zouzias and Nikolaos M Freris. Randomized extended kaczmarz for solving least squares. *SIAM Journal on Matrix Analysis and Applications*, 34(2):773–793, 2013.