UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Universitá degli Studi di Padova
Dipartimento di Matematica

Corso di Dottorato di Ricerca in: Matematica
Curricolo: Matematica Computazionale
Ciclo XXIX

Reconstruction of 3D X-ray tomographic images from sparse data with TV-based methods

---

Coordinatore: Ch.mo Prof. Martino Bardi
Supervisore: Chiar.ma Prof.ssa Elena Loli Piccolomini

Dottoranda: Elena Morotti

Reconstruction of 3D X-ray tomographic images from sparse data with TV-based methods

Elena Morotti
elena.morotti1988@gmail.com

Abstract

This dissertation presents efficient implementations of iterative X-rays image reconstruction methods for the specific case of three-dimensional tomographic imaging from subsampled data. When a complete projection dataset is not available, the linear system describing the so-called Sparse Tomography (SpCT) is underdetermined, hence a Total Variation (TV) regularized model is considered. The resulting optimization problem is solved by a Scaled Gradient Projection algorithm and a Fixed Point method. They both are accelerated by effective strategies, specifically tuned for a SpCT framework where fast reconstructions must be provided in short run time, facing a very large size problem. Good results on digital simulations attest the reliability of the model-based approach and of the proposed schemes. Accurate reconstructions from real medical datasets are also achieved in few iterations, confirming the feasibility of the proposed approaches to sparse tomographic imaging.

Keywords: sparse tomography, Total Variation, iterative solvers, inverse problems, X-ray medical imaging.

Ricostruzione di immagini tomografiche 3D a raggi X da dati sottocampionati con metodi TV-based

Elena Morotti
elena.morotti1988@gmail.com

Sommario

Questa tesi propone l'implementazione efficiente di due metodi iterativi per la ricostruzione di immagini tridimensionali di tomografia a raggi X, nel caso specifico in cui il volume debba essere ottenuto da dati sottocampionati. Quando le proiezioni non possono essere acquisite completamente, la risultante tecnica di Tomografia Computerizzata Sparsa (SpCT) é descritta da un sistema lineare sottodeterminato, quindi ne riformuliamo il modello aggiungendo il termine di Variazione Totale (TV). Definiamo pertanto un problema di ottimizzazione e lo risolviamo con un algoritmo di Gradiente Scalato Proiettato e uno di Punto Fisso. Entrambi i metodi sono stati accelerati con valide strategie, calibrate appositamente per la SpCT. In questo contesto é infatti necessario ricostruire un'immagine in brevissimo tempo, risolvendo un problema di ampie dimensioni. Alcuni test di simulazione forniscono buoni risultati che attestano la validitá sia dell'approccio model-based che dei metodi proposti. Accurate ricostruzioni sono state ottenute a partire da proiezioni mediche reali, in poche iterazioni: ció conferma l'adeguatezza di quanto proposto per la ricostruzione di immagini nel campo della SpCT.

Parole chiave: tomografia sparsa, Variazione Totale, metodi iterativi, problemi inversi, imaging medicale a raggi X

Acknowledgements

My sincere thanks also goes the I.M.S. Giotto Spa for having provided me the opportunity to access to a real medical imaging world and I would like to thank Sara Vecchio for her patience and immense knowledge.
My PhD was supported by Foundation Cariparo (Cassa di Risparmio di Padova e Rovigo): I am grateful for funding my research.

Contents

Abstract **i**

Acknowledgements **iv**

Contents **vii**

Introduction **1**

1 Basis of Computed Tomography and Digital Breast Tomosythesis **5**

1.1 Historical evolution of the tomographic technique 7

1.2 An example of sparse tomographic technique: the Digital Breast Tomosynthesis 10

1.3 Radon transform and its inversion 12

1.3.1 Lambert Beer's law 13

1.3.2 Radon transform 14

1.3.3 Backprojection operator 16

1.4 Modulation transfer steps of the imaging process 17

2 A numerical model based on a priori information for SpCT **19**

2.1 CT discretization into a linear system 19

2.2 The projection matrix operator 22

2.2.1 Ray driven approach for 3D CT imaging 23

2.2.2 Distance driven approach for 3D CT imaging 25

2.2.3 Main features of the system matrix 28

2.2.4 Physical interpretation of the transpose matrix 28

2.3 The model-based formulation for the SpCT 30

2.3.1 Compressive Sensing and medical imaging 30

2.3.2 The regularized model for the SpCT 31

2.3.3 The TV discretization 33

2.3.4 Unconstrained and constrained formulation 34

3 Reconstruction algorithms for Sparse CT problems 37
3.1 Iterative Image Reconstruction state-of-the-art 38
3.2 Scaled Gradient Projection method 39
  3.2.1 Focus on the scaling strategy 42
  3.2.2 Focus on two step-lenght rules 43
3.3 Lagged Diffusivity Fixed Point method 45
  3.3.1 Implementation notes 47

4 Tests on medium size problems in Matlab 51
4.1 Focus on the SGP algorithm 52
  4.1.1 Test problem and setting 52
  4.1.2 Case of Gaussian noise. Results with the LS data function. 54
  4.1.3 Case of Poisson noise. Results with the KL data function. 57
  4.1.4 Algorithms reliability with regard to noise 65
  4.1.5 Comparison of step-length rules 65
4.2 Comparisons on a real 2D SpCT dataset 68
  4.2.1 Test problem and setting 68
  4.2.2 Results from the 120 projections data set 70
  4.2.3 Results from the 20 projections data set 75
4.3 Comparisons on a simulated DBT case 77
  4.3.1 Test problem and setting 77
  4.3.2 Results 79

5 Results with real DBT data 87
5.1 Results on a Quality Control phantom 89
5.2 Results on a breast phantom 93
5.3 Results from real breast projections 98

6 Conclusions 103

Bibliography 105

Introduction

During the 1970s, there was an enormous step towards the advance of diagnostic possibilities in medicine and Computed Tomography (CT) evolved into an indispensable imaging method in clinical routine, because it was the first method to acquire images of the inside of the human body that were not biased by superposition of distinct anatomical structures. However, research in the field of CT is still as exciting as at the beginning of its development during the 1960s and 1970s. Due to its easy use, clear interpretation in terms of phyisical attenuation values, progress in detector technology, reconstruction mathematics, and reduction of the radiation exposure, Computed Tomography will expand its established position in the field of 2D and 3D radiological imaging.
Nowadays, three-dimensional reconstructions would be helpful in many clinical and industrial applications, but a complete projection dataset is not always available. For instance, in mammography and intraoral dental imaging, the projections can be taken only from a restricted angular range which is significantly smaller than 180 degrees: this leads to a limited angle tomographic problem. In applications like surgical imaging, projections are available from all around the body, but the radiation dose per patient is minimized by reducing the number of X-ray projections. The image reconstruction from the above types of incomplete data defines the Sparse Tomography (SpCT) problem. A well-known example is the Digital Breast Tomosynthesis (DBT) technique, that is an emerging alternative to mammography for breast cancer screening. It is characterized by both limited range and few views, nevertheless it allows good quality volumetric reconstructions.
From a mathematical point of view, any CT process can be modeled as a linear system and the imaging reconstruction consists in solving an ill-posed inverse problem. The resolution of this problem is complex in SpCT: the system is underdetermined and has big dimensions, the matrix is sparse and ill-conditioned and it does not have a specific structure. On the other hand, a main driving factor for the development of SpCT imaging systems has been the advent of model-based reconstruction approaches, proposed in compressed sensing for signal processing applications. A multitude of reconstruction formulations exploiting some forms of sparsity on the object have

been proposed and theoretical results promise accurate reconstruction even from heavily undersampled data. At the same time, iterative fast algorithms have been developed for variational imaging in deblurring and denoising. Their potential for successful application to CT problems has been already demonstrated empirically, but not to SpCT cases.
In particular, in SpCT problems the subsampled data do not contain sufficient information to perfectly come back to the scanned object, hence successful reconstructing softwares require additional a priori information, together with denoising tools. In addition, an accurate solution must be provided in short time: commercial systems compute volumetric images in less than one minute, thus a reconstructing algorithm must approach the convergence solution in few iterations.

The contribution of this PhD thesis is the proposal of a model-based framework to severe SpCT imaging problems (like the DBT one) and its resolution by means of two accelerated algorithms, with a commercial-like approach. In particular, the SpCT process is modeled as an optimization problem with a data-fitting term and the Total Variation as prior, to enhance contrasts among the image objects.
The main topic of this work is the efficient implementation of two accurate solvers: the Scaled Gradient Projection (SGP) and the Fixed Point (FP) methods. The former is a gradient descent method, equipped with a specific scaling strategy and a step-length rule to accelerate the descent to an optimal solution in its earlier iterations, without increasing its computational cost: it preserves its general framework and may be applied to a larger set of SpCT formulations. The latter is a quasi-Newton algorithm with a smart approximation of the Hessian of the TV operator: it is fast, not computationally expensive and it requires little memory footprint. Both solvers have a linear convergence rate, but in practical applications they achieve high quality results in very few steps.
The performances of the SGP and FP methods are tested both on synthetic problems and on real datasets, mainly acquired by a DBT device. The iterative reconstruction algorithms are always evaluated on short temporal windows: we aim to compare the proposed solvers with practical needs of commercial softwares, hence we are interested in fast recontructions and not in convergence solutions. Good results on digital simulations attest the reliability of the model-based approach and of the methods. Accurate reconstructions from real medical datasets are also achieved in few iterations, confirming the feasibility of the proposed approaches to sparse tomographic imaging.

The thesis is organized as follows. Chapter 1 concerns the historical evolution of tomographic devices up to the modern example of sparse CT DBT technique. It further introduces physical and mathematical theoretical notions describing the CT data acquisition process, for X-ray imaging. In chapter 2 the SpCT linear system is derived from the discretization of the continuous CT model and a detailed study about the matrix computation is presented. After that, the model-based optimization problem is defined, for the specific case of sparse tomography. In chapter 3 the state-of-the-art of iterative solvers for tomographic imaging is presented and the SGP and FP methods are shown with accelerating strategies, properly tuned for the prefixed model. Both theoretical and implementing remarks are discussed. In chapter 4 medium size SpCT problems are considered. Here we compare the algorithms performances both on 2D real data and on 3D simulations, reflecting different tomographic settings and applications. Then we focus on 3D breast imaging reconstructions from real DBT datasets, in chapter 5. Finally, chapter 6 collects some concluding remarks.

Chapter 1

Basis of Computed Tomography and Digital Breast Tomosythesis

Medical imaging was born not long after Wilhelm Rontgen discoverd X-rays in 1895, when scientists realized X-ray capability of crossing objects: for decades 2D planar images (projection radiographies) have been used to investigate the inner parts of human bodies. This conventional X-ray imaging suffered from the severe drawback that it only produces two-dimensional projections of a three-dimensional object: it results in a reduction in spatial informations, as one projected image represents an averaging.
In 1917, a well-known paper by Radon provided the mathematical foundation for tomographic imaging and, in the 1920s, the desire to undo the averaging process of the conventional X-ray radiography led to the first tomographic concept that was considerably influenced by the physician Grossmann, who was the first to image one single slice of a body in 1934. Only in the 1970s, thanks to the advent of the computers, the Computed Tomography (CT) revolutionized the diagnostic imaging by allowing the three-dimensional orientation of anatomy to be appreciated in *transverse* sections. In fact, as the word *tomography* itself suggests (from the ancient greek, *témnó* (or *tómos*) that means ”to slice” and from *gráphó* that means ”to write” or ”to draw”), a *computed* tomographic system digitally writes a cross-sectional image (made of one or more virtual ”slices”) of a specific object. To performe classical CT, many projections of the same slice must be acquired from many angled views, in a round trajectory, according to the studies of Radon. Such a setting allowed CT technique to provide some important advantages over conventional radiography: for instance, it permitted depth localization and removed the visual chaos due to the anatomical tissue overlapping. In addition, CT improved the constrast of local structures, by means of a dynamic range in the visualization of a single slice; nowadays it also allows volumetric

investigations.
From the primordial rare machineries to the most modern and common gantries that we can now find in any hospital (pictures in figure 1.1), many studies have been done by different research groups, collecting engineers, physicists, mathematicians and computer scientists all together, in order to improve both the technologies and the reconstructing softwares.

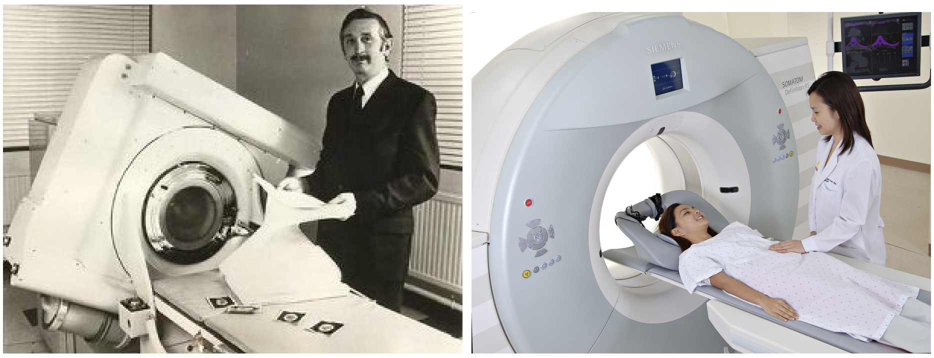


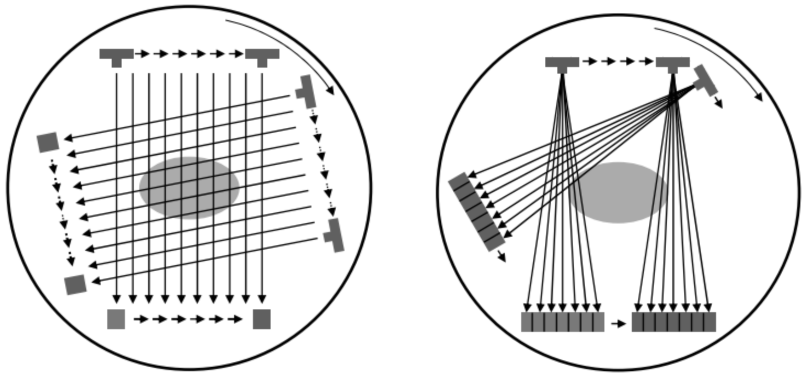
Figure 1.1: On the left: a famous picture of Mr. Hounsfield and his first prototype of head CT machinery. On the right: modern CT device, made of a gantry (a donut-shaped open structure) and a patient-bed that can slowly move through the gantry.

Nowadays, many tomographic devices have been designed to fit different medical needs and, on the other hand, interesting technical, antropomorphic, forensic and archeological as well as paleonotological applications of CT have been developed too [55, 6]. As a consequence, the CT technique evolves into new inquiring forms continuously and all these applications further strengthen the CT methodology as a generic diagnostic tool for non-destructive material testing.
Recently, a growing interest is devoted to the so-called *Sparse* Computed Tomography (SpCT) which aims to produce clear visualizations making use of little radiation dose. In medical field, SpCT allows to apply CT techniques to a wider class of examinations, including screening tests: safer routines can be led without compromising the reliability of their diagnosis.

In this chapter, a fast overview on the historical evolution of the CT devices (in section 1.1) introduces to section 1.2 where the Digital Breast Tomosynthesis technique is presented as a famous example of Sparse Tomography application. In section 1.3 basis physical concepts are shown to describe the tomographic process of X-ray absorbtion and in 1.4 we summarize all the steps of the imaging reconstruction process.

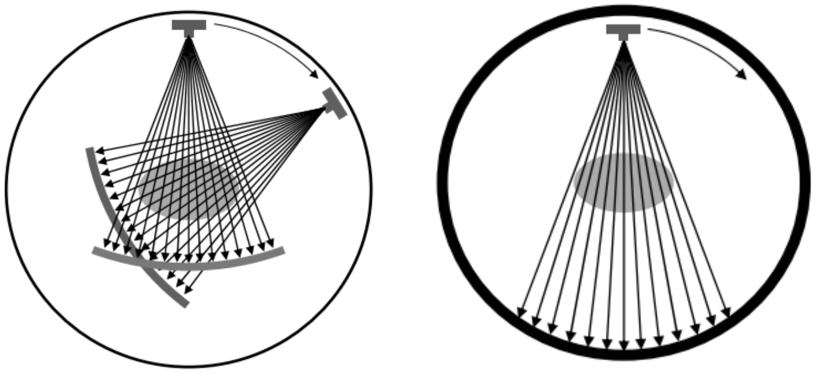
1.1 Historical evolution of the tomographic technique

Figure 1.2 shows the schemes of the four main generations of CT scans, developed since the 1970s for the two-dimensional tomography.



(a) First generation

(b) Second generation



(c) Third generation

(d) Fourth generation

Figure 1.2: Tomographic device sketches in their evolution, from the primordial technology by Cormack and Hounsfield, to the most modern solution for the 2D CT systems.

The first commercial machinery was designed by the physicist Allan Cormack and the engineer Godfrey Hounsfield and it was installed in London in 1971; the two inventors won the Nobel Prize for Medicine in 1979, for such a revolutionary innovation. The first prototype was equipped and based on a naive technology: only a pencil beam was emitted from the X-ray source and detected by one sensor behind the object to scan; then the source-detector couple had to translate in order to project the whole object, as shown in figure 1.2(a). After that, the synchronized couple had to rotate on a circular trajectory and perform a further complete projection, and so on up to a full tomographic scan over 180 degrees. The rotation angle between two subsequent projections was very small, hence the whole scanning process took up to 4-5 minutes, while the first reconstruction of a 13 mm two-dimensional brain slice took around 9 days [26].

The second CT generation (figure 1.2(b)) was powered with a 3-20 degree wide fan beam, whose attenuation was measured on several detector cells simultaneously. Before wheeling the source-detector pair, the translation phase was still necessary to cover all the slices, hence a scan now took 15-30 seconds. However, the measuring field was still small and due to their long ascquisition times, the first- and second-generation scanners use was limited to the imaging of the cranium.

The main goal of further developments was to reduce the acquisition time down to less than 20 seconds, in order to acquire an image of the abdomen with minimal motion error while the patient holds his breath. In the 90s, a third generation of devices was born, equipped with a wide fan beam (40-60 degrees) that can scan all the object in one only shot (figure 1.2(c)). Of course, a larger detector was installed, with up to 800-1000 recording cells. No translation is required and the scanner can rotate continuously, hence the total acquisition became very fast (around 1 second).

Thanks to the technological support, in the fourth generation scans thousands of detector units surround the patient completely, hence only the X-ray source needed to rotate and the complete tomographic process requires less than 1 second (figure 1.2(d)).

On the other hand, when shorter acquistitions are required, the concept of moving sampling systems must be left behind entirely. In 2000s one new approach was introduced on medical markets and it was based on electron beam computerized tomography: this type of CT was developed for cardiac imaging which is characterized by movement artifacts, due to the heartbeats. It represents the fifth generation of scanners. Generally speaking, starting from the third and the fourth generation prototypes, many medical CT scanners have been designed with different configurations and geometries in recent years, in order to fit better with specific areas of the human anatomy or to reconstruct wider human sections.

Devices for 3D imaging

A new target of CT is to reconstruct not only a slice, but a whole volume of human anatomy.

Even if there has not been so far a common definition of the generations for CT machineries, we can identify as a sixth generation those scanners equipped with a slip-ring technology [42]. In this case the energy is provided via sliding contacts, in the so-called *gantry*, and the rotation frequency is of two rotations per second (hence the name ”sub-second scanners”): these devices are commonplace nowadays. In particular, the slip-ring innovation enabled a new acquisition technique: along with a continuous motion of the patient table through the sampling unit, it became possible to measure data in the shape of a spiral and get projections of a larger slice (*spiral CT* scanners). Moreover, this methodology has been evolved into a multislice projection system to perform the *helical CT*: as shown in figure 1.3, the X-ray tube can now emit 4-16 fan beams in a single shot hence, moving the patient towards the gantry, a complete human scan can be performed in olny 40 seconds. This type of machinery has been introduced into clinics in 1998 and supplied first volumetric images.

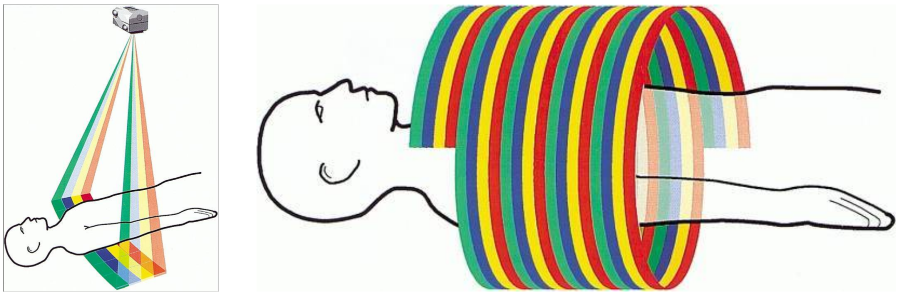


Figure 1.3: Sketches of the multi-slice and the helicoidal CT scanners.

A seventh generation of CT scanners can be identify with the one equipped with a cone-shaped X-ray beam and a two-dimensional detector, as reported in figure 1.4. In this case, the beam is wider than in the multi-slice one and every projection consists in a two-dimensional recorded image: the source-detector couple rotates only once to get all information we need to image a volume.

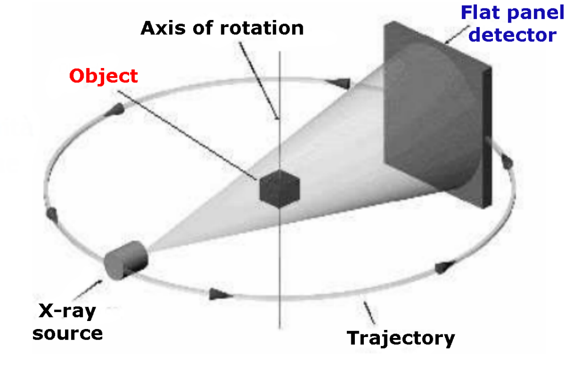


Figure 1.4: A sketch of the seventh generation model for CT scanners, characterized by a conical beam and a two-dimensional recording element.

1.2 An example of sparse tomographic technique: the Digital Breast Tomosynthesis

Motivated by an increasing focus on the potentially harmful effects of CT radiations, a recent trend in CT research is to develop low-dose imaging techniques. In classical CT, the angular step Δ_θ that occurs between two consecutive projection angles is typically lower than one degree. One way to reduce the total amount of radiation consists in increasing Δ_θ: as a consequence, the final image must be computed from a smaller projection dataset and this characterizes the *Sparse Tomography* (SpCT) technique. This approach is widely used not only in medicine but also in materials science, to prevent damage to the subject under study, due to excessive radiations.
When a further dose reduction is required, one possible solution relies on decreasing the scanning angular range, from the classical $[-\pi, \pi]$ to a significantly smaller range $[-\Theta, +\Theta]$ where $0 < \Theta \ll \pi$. In other tomographical applications, the human anatomy does not allow a complete circular motion to the X-ray source, thus the use of a reduced range is mandatory. The resulting technique is called *Tomosynthesis*: it represents a crucial class of SpCT cases because many projection views are missing, with respect to the classical CT case.
By such a reduction of the X-ray exposure for a patient, low-dose imaging has enabled new applications that were prevented by the high-dose levels [56, 78]: nowadays this methodology has many applications in medical imaging such as in vascular, dental, orthopedic, musculoskeletal, chest and mammographic imaging. In particular, the results in chapter 5 are related to the Digital Breast Tomosysnthesis (DBT) technique, which has been recently developed from the seventh CT device generation as a low-dose tomographic alternative to the mammographic screening test [74, 16, 59].

According to American Cancer Society's recent studies, breast cancer is the most common tumor among women today and a serious threat worldwide. In Italy, for instance, every "healthy" woman between 40 and 69 years, is called by the National Health System to performe a breast screening test every two years, because an early detection of the cancer can be crucial for its eradication. Mammographic screening has been proven effective in significantly reducing breast cancers mortality rates in the screened population; nevertheless 11% to 25% of tumors are overlooked by radiologist on the initial mammogram and up to 15% of palpable cancers are missed by mammography [25]. In fact, many important factors affect lesion detectability on X-ray imaging, such as breast positioning, tumor small size and low density, minimal distortion of healthy breast architecture and breast density, too. Moreover, the mammographic detection of a cancer suffers from the obscuring effect of overlapping breast tissue, that is due to the projection

onto a flat final diagnostic image of all the breast volume: the cancer can be masked by sourrounding overlapping structures and not standed out, especially in woman with radiographically dense breasts. Magnetic resonance imaging has been shown to be superior to mammography in dense breasts, but its routine use is often prevented by its high cost, time requirement and limited specificity.
Nowadays, the mammographic test is not recommended as breast cancer screening anymore, while in 2011 the Food and Drug Administration (a federal agency of the United States Department of Health and Human Services) approved the DBT for that in the US [10]: the Digital Breast Tomosynthesis can provide a higher diagnostic accuracy, offering better detection rates, and its dose is reported to be approximately equal to that of only a two view digital mammography [32].

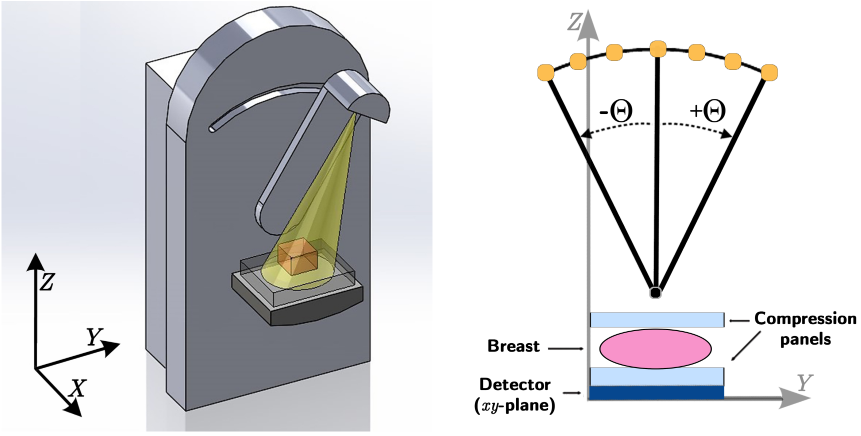


Figure 1.5: Sketches of a breast screening machinery. The volume is scanned from few angled positions while the X-ray source moves over it along an arc; the breast is compressed by two plastic panels and a flat wide detector element lies under, on a plane orthogonal to the vertical $z-$axis.

A modern DBT machinery is presented in figure 1.5. In DBT, like in conventional mammography, breast compression between two plastic panels is used to improve image quality by fixing the breast and preventing its movements. Such impaction is parallel to the detector plane: the detector, in particular, is flat. It is stationary for all the scanning time and almost parallel to the floor. It is convenient to fix the Cartesian axis system by identifing the $xy-$plane with the detector plane and the $z-$axis as the vertical direction. Now, in DBT the X-ray source moves quickly on the $yz-$plane, on an arc whose center of rotation is tipically centered above the breast, as shown in figure 1.5. This arc is 11 to 60 degrees, hence $\Theta \approx 5$ to 30 degrees typically. Then, from equispaced angled points on the arc, the cone-beams draw $N_\theta = 9$-25 low-radiation-dose projection images on the detector. N_θ is an odd number, in order to capture the vertical projection from $\Theta = 0$ (reflecting one mammographic-like view) and to make the DBT scan symmetric. Because of the limited range $[-\Theta, +\Theta]$, DBT is unable to offer the narrow slice thikness that CT offers, hence the z-axis resolution is 1-1.5 mm typically. However, since the acquired data has 85 - 160 micron resolution, most detectors allow very high in-plane resolution (i. e. the resolution on slices parallel to the detector plane is smaller than 0.1 mm).

In comparison to classical medical CT, another important feature of the DBT is its making use of *soft X-rays* with few tens of electron volts: this choice helps to reduce the provided radiations, but it is motivated by the anatomical structure of the breast. In breast imaging, there are no bones nor any other metallic objects, but only adipose and fibro-glandular tissues that have very low attenuating properties: breast materials would not capture many photons from high X-rays. On the contrary, we can get meaningful information from the projection data set with soft X-rays, which have less energy and are more absorbable. Since much more photon scattering occurs, this chioce provides noisier data; nevertheless it also allows to detect the breast objects in a more distinguishable way.

We must highlight that the aim of DBT (and all the screening tests, in general) is finding microcalcifications or suspected masses and try to classify them into benign or malignant tumors. Benign tumors in breast may be cysts or fibroadenomas and they appear like compact ovoidal lumps with net boundaries in X-ray imaging; cancerous masses instead are characterised by branched and spiculated shapes with smoothed edges, because they often occlude glandular ducts and they infiltrate in the sourrounding tissues. Moreover, above all in case of malignant masses, cluster of microcalcifications may be present in all the breast: they are irregular calcium deposits of high density, with diameters smaller than 0.5 mm, hence they appear as tiny contrasting objects among soft tissues in X-ray imaging. The possibility of a 3D imaging reconstruction can help in detecting all these objects and lead to more reliable diagnosis, improving both the sensibility (the capability to find true positive) and the specificity (the capability to find true negative) of the screening tests. For medicians, accurate DBT reconstructions avoid useless biopsies while for scientific community it is a new challenging target.

1.3 Radon transform and its inversion

What is there behind all the X-ray imaging technique? From a physical point of view, the projection data reflect the absorbtion of photons the X-rays are made of, and the visualization of the reconstructed object is a picture of the attenuation coefficient map, of the object itself, in pseudo-colors. In this section, we deepen this topic.

1.3.1 Lambert Beer's law

All physical mechanisms that lead to the attenuation of radiation intensity (i.e. reduction of photons) measured by a detector behind a homogeneous object, are usually to be subsumed by a single *attenuation coefficient* $\mu = \mu(w) \geq 0$ depending on the crossed point w . Within this simple model, the total attenuation of a monochromatic X-ray beam can be calculated in the following way [26]: the radiation intensity (which is propotional to the number of photons) measured after passing a thickness Δw through an object is determined by

$$m(w + \Delta w) = m(w) - \mu(w)m(w)\Delta w \tag{1.1}$$

where $m(w)$ is the intensity of the incoming beam. Reordering this equation, we get

$$\frac{m(w + \Delta w) - m(w)}{\Delta w} = -\mu(w)m(w). \tag{1.2}$$

Taking the limit of the left hand side, we obtain the differential quotient

$$\lim_{\Delta w \to 0} \frac{m(w + \Delta w) - m(w)}{\Delta w} = \frac{dm}{dw} = -\mu(w)m(w). \tag{1.3}$$

By assuming the object to be homogeneous, it can be described by a single attenuation constant, i. e. $\mu(w) = \mu$ along the entire path of penetration Δw. This let us deal with an ordinary and homogeneous, first differential equation with constant coefficient, hence its solution can be derived by separation of variables. In the last equation, we can therefore separate

$$\frac{dm}{m(w)} = -\mu dw. \tag{1.4}$$

By integration of both sides of equation (1.4), we get

$$\int \frac{dm}{m(w)} = -\mu \int dw \tag{1.5}$$

providing

$$\ln |m| = -\mu w + C. \tag{1.6}$$

Because of their physical meaning, all the measured intensities m are positive quantities, so the absolute value can be deleted and we can apply an exponentialization leading to

$$m(w) = e^{-\mu w + C}. \tag{1.7}$$

Since the initial condition $m(0) = m_0$ is the emitted photon count, it is known for any beam of the CT machinery and the special solution of the differential equation (1.3) becomes

$$m(w) = m_0 e^{-\mu w}, \tag{1.8}$$

that is known as *Lambert Beer's law* of attenuation.
Truly, since the linear attenuation coefficient μ is an additive combination of a scatter coefficient and an absorption value, the Beer's law holds for pencil-beam geometry only, because here the scattered radiation is completely removed from the main beam; nevertheless it still remains one basic concept for the X-ray CT.
From a mathematical point of view, the complicated dependency of the attenuation coefficient on the property of the penetrated material means that the differential equation (1.3) cannot be fully integrated as shown in (1.5). In case of spatially varying attenuation $\mu(w)$, in fact, the solution for the intensity measured after a running length W is given by

$$m(W) = m_0 e^{-\int_0^W \mu(w)dw}. \tag{1.9}$$

We can now easily derive the *projection integral*

$$\mathcal{P}(W) = -\ln\left(\frac{m(W)}{m_0}\right) = \int_0^W \mu(w)dw \tag{1.10}$$

that is essentially the negative logarithm of the ratio of the outgoing and incoming numbers of photons for the scanned object.

Honestly speaking, this law stands only for monoenergetic beams, that is quite a theoretical concept: due to the energy dependence of the attenuation values on the wave length of the emitted X-ray, we should integrate over all the energies of the beam

$$m(W) = \int_0^{E_{max}} m_0(E)\ e^{-\int_0^W \mu(E,w)dw}\ dE. \tag{1.11}$$

and consider a polychromatic X-rays beam properly. The assumption of monochromatic rays and equation (1.9) instead of (1.11) is the origin of what is called the *beam-hardening artifact*.

1.3.2 Radon transform

In 1917, the Austrian mathematicians Johann Radon provided the mathematical foundation for tomographic imaging reconstruction in [86].
Let us set a pencil geometry, in which the X-ray beam is collimated to a pencil shape and moved linearly in the direction parallel to a linear X-ray detector array, for every direction of projection, like it really was in the first generation of CT scanners. Let us fix the bidimensional case, where one single slice of the object is examinated at the time: in this case we can identify every point w with a bi-dimensional couple (x, y) of coordinates and the attenuation coefficient is a real and continuous function $\mu(w) = \mu(x, y)$ over the spatial domain of the slice, mapping what it is made of. Let us identify

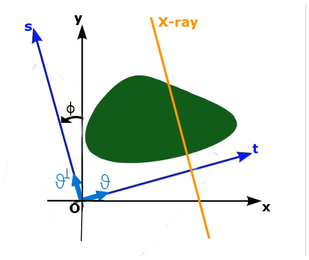


Figure 1.6: Draft of the CT process in both the Cartesian and polar coordinated systems.

one ray through its path L and the emittion angle Φ (with respect to the Cartesian system xOy), as shown in figure 1.6. Now, the projection integral in equation (1.10) represents a line integration along the path L, crossed by the emitted photons, and it is defined by the positions of the X-ray source and the detector unit. The recorded value m is therefore such that

$$-\ln\left(\frac{m}{m_0}\right) = +\int_L \mu(x,y)dw. \tag{1.12}$$

Let us consider the parallel beam in the Cartesian system tOs with versors $\theta = (\cos\Phi, \sin\Phi)$ and $\theta^\perp = (-\sin\Phi, \cos\Phi)$, as shown in figure 1.6. Any X-ray L is a function of Φ (or θ) and t, hence we can reformulate the line integral in (1.12) in these new coordinates [63], getting

$$\int_L \mu(x,y)dw = \int_{-\infty}^{+\infty} \mu(t\theta + s\theta^\perp)ds \ = \mathcal{P}_\theta(t) \tag{1.13}$$

that perfectly mirrows the projection $\mathcal{P}$ of the object along a θ-sloping ray in t. From now on, θ stands for the scanning angle too, because it determines Φ uniquely.

Given the current scanning angle θ, the *Radon transform of* μ is defied as the map $\mathcal{R}_\theta : \mu(x,y) \mapsto P_\theta$, such that

$$(\mathcal{R}_\theta\mu)(t) = \mathcal{P}_\theta(t) = \int_{-\infty}^{+\infty} \mu(t\theta + s\theta^\perp)ds \qquad \forall\, t \in \mathbb{R} \tag{1.14}$$

It means that the Radon transform $\mathcal{R}_\theta$ of an object, described by μ, is the whole projection $\mathcal{P}_\theta$ of the object itself, when it is scanned from the θ angled position entirely.

The circular process defining the CT is based on continuous radial acquisition, meaning that a Radon transform of the slice of interest is measured on the detector from all the angles $\theta \in [-\pi, \pi]$. Of course devices can only performe small angular steps Δ_θ, hence a finite number N_θ of scans are performed from prefixed angles $\theta_k \in \{\theta_1, \ldots, \theta_{N_\theta}\}$, and because of the finiteness

of detector components also projections are recorded in a finite number of points t_i where $i = 1, \ldots, N_p$.
The graphical representation of all the data, in the bi-dimensional case, is called *sinogram* whenever it is organized in the Cartesian system (θ, t).
In case of volumetric tomography, $\mu(w) = \mu(x, y, z)$ is a map of a 3D object and every $\mathcal{R}_\theta(\mu)$ is an X-ray image of projection, made of N_p pixels, accordingly to the detector resolution $n_x \times n_y$ $(N_p = n_x \cdot n_y)$.

1.3.3 Backprojection operator

Once we collect all the projection data, the mathematical inverse problem is defined: how can we come back to the spatial distribution of the attenuation coefficients $\mu(w)$, having measured a certain number N_θ of projection data P_θ? The basic idea is to project backwards every data onto its original ray-path, as schematically shown in figure 1.7.
Several problems arise at this point, in practical implementations. First of all we need to know exactly which points are involved for each data and tracing all the X-rays is an expensive task. Secondly, real data are corrupted by noise, whose propagation must be faced during the reconstruction phase.
The first commercial softwares were based on an *analytical approach*: taking into account the Fourier Slice theorem (also known as Central Slide theorem), it is possible to skip the heavy step of ray-tracing, on behalf of

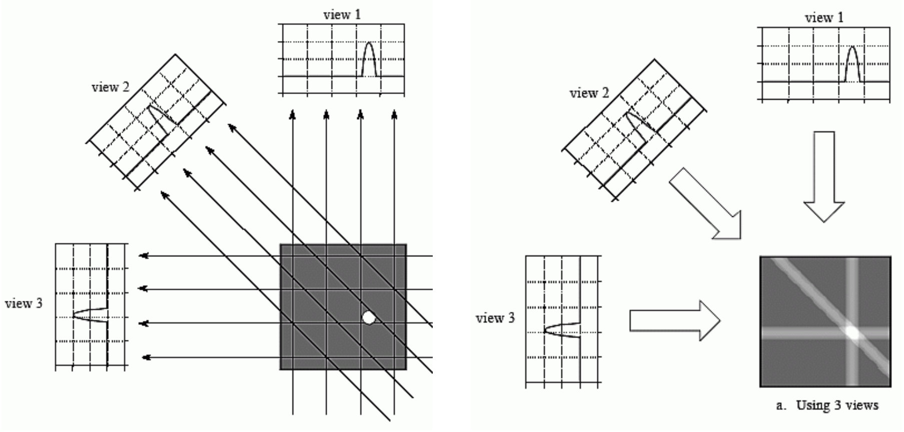


Figure 1.7: Three projections are acquired for a single slice, containing only a small circular object. Then they are back-projected onto an empty image, to show how a Back Projection works, with $N_\theta = 3$. For each Radon transform, the small object provides non-zero values in particular abscissa, according to the scanning angle: by summing the contributions of each back-projected image on a unique slice, we obtain the tomographic reconstruction. From this example, it becomes clear how taking *many* views (from a circular trajectory) is important to achieve high-quality results.

1-dimensional Fourier transform of every Radon transform. Moreover, in the frequency domain it is possible to apply suitable smoothing filters and reduce the high-frequencies that tipically enphasyse noise propagation. These features define the well-known *Filtered Back Projetion algorithm* that has been widely used and developed in many commercial softwares, in the past decades. See [43, 63] for more details.

1.4 Modulation transfer steps of the imaging process

To conclude this first chapter, we want to sum up all the processes that occur before the final visualization of the reconstructed image by a clinician, because it is important to evaluate all the changes the continuous signal μ is subjected to. The signal path can be in fact modeled in five main layers [42, 26]. The first one is the *X-ray imaging* (or physical) layer and it reflects the beam features: the interaction between the radiation and the object generally changes the physical nature of the radiaton itself, producing hardening of the spectrum of polychromatic X-ray. A second layer is the *detector* (or sensor) one: here the signal is subdivided by a grid beam collimator and physically transformed, because the photons are detected with a scintillator and converted into an electric signal. Furthermore, a *digitalization* (or electronic) layer spatially discretizes the incoming electric signal and the values are quantized by an acquisition with an analog-digital converter. The *algorithm* (or reconstruction) step is the fourth one, where we model both the size of the final image and some filter kernels, to impress some special properties to the output. At last, the *display* (or image processing) layer models the type of visualization by mapping, for instance, the physical attenuation values onto a gray-scale range and adjusting it for human perception as well as for the used monitor.

All these steps produce a global deterioration (or better an unavoidable alteration) of the final representation of the object, that must be added to a series of artifacts like motion, sub-sampling and metal ones. For instance, the first layer causes the so-called ”quantum noise” which appears as spots (Poisson noise), while the ”electronic noise” is due to the detector sensibility of the second layer. Moreover, the third layer determines the quantity and influences the quality of measured data. The fourth step shapes the image resolution, can partially avoid some artifacts thanks to the use of a priori information, but it can also introduce some beam-hardening effects and noise propagation. At the end, the last step can hide some structural noise (due to the unavoidable presence of anatomical dense tissues) adjusting the graphic rendering for the medicians, or enforce the contrast between structures making their detection easier.

Of course not all the artifacts and the noise can be deleted, moving from the tomographic setting, through the projection images up to the output; for instance, typical image errors of third-generation scanners are called "ring artifacts" if errors appear like ring-shaped shadows around the objects. They are due to an inaccurate detector calibration. More in general, every specific geometric set up of the scanners leads to some characteristic troubles, as well as reconstruction algorithms influence the final output.

For what concerns the Sparse Tomography and the Tomosynthesis case, it has been studied [74] that:

- increasing the radiation dose, we gain in image denoising immediately;
- wider angluar range $[-\Theta, \Theta]$ provides better depth resolution and gives increased reconstructed slice separation, while a narrow one enhance the in-plane resolution and enlarge the out-of-plane artifacts;
- by fixing a constant dose and angular span, increasing the number of projections N_θ (i.e. reducing the angular step Δ_θ) may not necessary improve the performance too.

Moreover, due to the finite size of the detector, limited angular range and small number of projections, any reconstruction from sparse subsampling will exhibit some artifacts: some of them will obscure some tissue details or interfere with radiologist visual evaluation of subtle anatomical features, hence an active research should match more and more specific reconstruction tools with artifact reduction strategies.

Chapter 2

A numerical model based on a priori information for SpCT

In this thesis we focus on the 3D image reconstruction from low sampling tomographic acquisitions. Since the SpCT system acquires only a reduced number of data sets, the traditional analytical reconstruction methods (such as the Filtered Back-Projection, called Feldkamp-Kress method in 3D [43]) produce images of low quality, with harmful artifacts and high noise. Consequently, Iterative Image Reconstruction (IIR) methods are preferred [81], but their application requires to reformulate the CT process in a new mathematical form.

Dealing with the search of a good model for the Sparse Tomographic imaging, one should care that the mathematics of CT image reconstruction has widely influenced other scientific fields and vice versa. In particular, Compressive Sensing (CS) has attracted much interest in the medical image community because of its great potential to achieve high quality results from sparse data sampling acquitions: in this sense, if the prior knowledge that medical images are compressible is properly exploited, CS may lead to faster and safer image acquisitions and reconstructions, without sacrificing image quality.

In section 2.1 we will derive the reformulation of the CT process into a large linear system and a focus on the forward projection matrix operator is in section 2.2. Because of the subsampling, the SpCT system is underdetermined, hence in section 2.3 the model-based approach is introduced with different regularizing formulations and we will derive the optimization problem we aim to solve.

2.1 CT discretization into a linear system

Recalling equations (1.12) and (1.14), we can derive the linear system which describes any CT acquisition process. Let us fix the notation for a general CT scanner of the seventh generation, i. e. emitting an X-ray cone beam from a source that rotates in a (complete or uncomplete) circular trajectory, together with a bidimensional detector panel, around the object of interest. The derivation of the Sparse CT system follows immediately.

First of all, it is convenient to recall the Cartesian axis system introduced in the previous chapter: let us identify the rotation plane with the yz-coordinate plane and the recording panel lying on the $z = 0$ plane at the beginning. We have to discretize the continuous volume into a certain number N_v of volumetric elements *voxels*, by discretizing it along the three dimensions into N_x, N_y, N_z elements (with step sizes $\Delta_x, \Delta_y, \Delta_z$ respectively): of course $N_v = N_x \cdot N_y \cdot N_z$. The three-dimensional discrete object can be lexicographically reordered into a vector-shaped array, so that we refer to the single element $f_{j_x,j_y,j_z} \forall j_x \in 1, \ldots, N_x,\ j_y \in 1, \ldots, N_y$ and $j_z \in 1, \ldots, N_z$ as the j-th voxel $f_j, \forall j \in 1, \ldots, N_v$, according to the column-ordered reshape.
Moreover, if the whole detector is a grid of $n_x \times n_y$ recording units, each of area $\delta_x \cdot \delta_y$ mm^2 , let us define $N_p = n_x \cdot n_y$ as the number of measurments we get at each X-ray shot, in a single projection view: reordening each projection image into a vector-shape, we will refer to each pixel (i_x, i_y) $\forall i_x \in 1, \ldots, n_x,\ i_y \in 1, \ldots, n_y$ with the corresponding index $i \in 1, \ldots, N_p$. We define N_θ as the number of projections performed by the CT device in the angular range $[-\Phi, +\Phi]$ at equispaced positions , that are defined by the θ_k angles $\forall k = 1, \ldots, N_\theta$. Finally, $N_d = N_\theta \cdot N_p$ is the total number of data: in classical CT, $N_d \gg N_v$, while in case of SpCT N_θ is a small number, thus $N_d < N_v$.
If we now fix the k-th projection image (acquired from the θ_k-th angled position) and the i-th pixel over it (with $i \in 1, \ldots, N_p$), we can read its measured value m_i and, starting from equation (1.12), we define

$$g_i = -\ln\left(\frac{m_i}{m_0}\right) \quad \forall i \in 1, \ldots, N_p \tag{2.1}$$

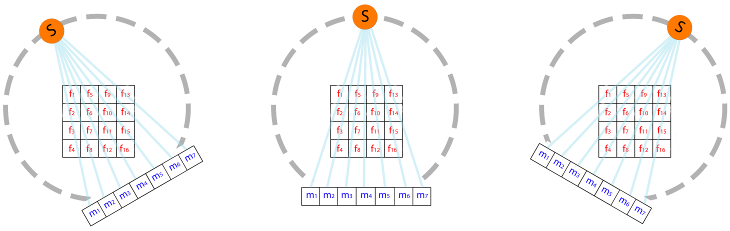
where m_0 is the photon counting of the outcoming beam.
The line integral of equation (1.12) can be discretized into a sum over all the voxels, hence it becomes

$$g_i = \sum_{j=0}^{N_v} M_{i,j} f_j \quad \forall i \in 1, \ldots, N_p \tag{2.2}$$

where $M_{i,j} = 0$ if the $j-$th voxel does not lie on the $i-$th integral path. This formula can be compacted with a system notation into

$$g^{\theta_k} = M^{\theta_k} f \tag{2.3}$$

where M^{θ_k} is a sparse matrix of size $N_p \times N_v$. Taking into account the complete set of data $g^{\theta_k} = \{g_i\}_{i=1,\ldots,N_p}$ obtained from the angle θ_k, we can



(a) (b) (c)

Figure 2.1: 2D scheme of the tomographic rotation process where the slice of interest is scanned from prefixed θ_k angles.

affirm to have discretized the Radon Transform (1.14) into equation (2.3).

Going on and perfoming all the N_θ tomographic scansions as schematically shown in figure 2.1, we collect N_θ projection images: each one defines a vector g^{θ_k} of data. In particular, each X-ray scan can be described with an adequate linear system of form (2.3) and they all have the same size. Gathering together all the N_θ systems, we can construct the following huge linear system

$$\begin{bmatrix} \rule{3cm}{0.4pt} M^{\theta_1} \rule{3cm}{0.4pt} \\ \rule{3cm}{0.4pt} M^{\theta_2} \rule{3cm}{0.4pt} \\ \rule{3cm}{0.4pt} M^{\theta_3} \rule{3cm}{0.4pt} \\ \rule{3cm}{0.4pt} M^{\theta_4} \rule{3cm}{0.4pt} \\ \dots \\ \dots \\ \dots \\ \dots \\ \rule{2.5cm}{0.4pt} M^{\theta_{N_\theta}} \rule{2.5cm}{0.4pt} \end{bmatrix} \cdot \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \\ \dots \\ \dots \\ \dots \\ \dots \\ \dots \\ f_{N_v} \end{bmatrix} = \begin{bmatrix} g^{\theta_1} \\ g^{\theta_2} \\ g^{\theta_3} \\ g^{\theta_4} \\ \dots \\ \dots \\ \dots \\ \dots \\ g^{\theta_{N_\theta}} \end{bmatrix} \tag{2.4}$$

and reducing this heavy notation into a more compact and familiar notation, we finally state the linear system describing the CT process:

$$Mf = g \tag{2.5}$$

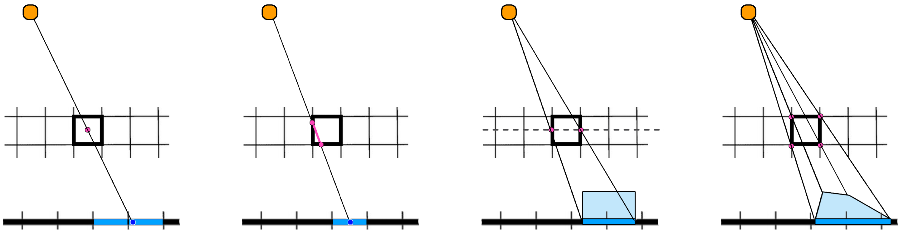
where $M \in \mathbb{R}^{N_d} \times \mathbb{R}^{N_v}$, $f \in \mathbb{R}^{N_v}$ and $g \in \mathbb{R}^{N_d}$.

Once the X-ray scansion has been performed, the right-hand side g is determined, while the dimension of the unknown f is tipically prefixed according to the spatial resolution required by any specific examinations. The most crucial issue in the CT system definition concerns the computation of the matrix coefficients: although very simple in principle, elaborate computer algorithms and a significant amount of computer time is required to evaluate the exact radiological path and determine, somehow, the coefficients $M_{i,j}$ linking each voxel f_j to a specific data g_i. In particular, following this approach, M can be considered the matrix operator $M : \mathbb{R}^{N_v} \to \mathbb{R}^{N_d}$ describing the tomographic projection of the discretized exact volume f^* into the measured dataset g. Many algorithms have been proposed in literature to implement realistic and efficient projection matrix: in section 2.2 we will study this issue more in detail.

2.2 The projection matrix operator

To define an appropriate matrix M, we must realize that it is the mathematical description of the physical process of CT data acquisition, hence M must mirrow the forward projection (FwP) of a volume onto the detector units, for all the scanning views. Because of the big dimensions, M is never stored on commercial systems but it is evaluated run time whenever M or M^T are required by the reconstructing algorithm: the computation of M is required to be a fast routine, for its practical implementation on commercial softwares.

This issue was pointed out quite early: in 1977 two 2D projecting algorithms were already in a library user manual by Huesmanand [54] and their approaches, called pixel-driven and ray-driven, were the basis of following algorithms. We now focus on the most popular approaches for the 2D tomography, then we extend two of them for the 3D case.



(a) Pixel driven (b) Ray driven (c) Distance driven (d) Separable trapezoid footprint

Figure 2.2: 2D schemes of the tomographic forward projections. Even from a first comparison of these images, it is noticeable how every projector describes in a different way the CT scanner. All these approaches are conceptually straightforward to be generalized to the 3D voxel case.

The *pixel-driven* basic algorithm [82] is quite simple and it is schematically shown in the figure 2.2(a): according to the geometry of the device, for each scanning angle every pixel in f is projected from its center (the pink

point on the figure) onto the detector; after its localization onto the detector, its contribution is split among the adjacent measuring units with a linear or more complex interpolation routine [52, 44]. When the spatial resolution of the reconstruction is much bigger than the detector cell size, too few rays are taken into account and it may happen that some detector cells do not receive any values at all (that is, of course, unrealistic). As a result, high-frequency artifacts will appear in the reconstruction: all the tecnhiques introduced to reduce such artefacts (like the splatting one [79]) generally lead to computationally complex alghorithms. This is the main reason why pixel-driven approaches are rarely used in real applications.
In the *ray-driven* (or ray-casting) approach, a straigth line from the source to the center of each detector unit represents the X-ray beam which falls onto the whole recording element: the projection value of that unit is computed as a weighted sum of all the crossed pixels. A simple draw of this methodology is shown in figure 2.2(b) and the subsection 2.2.1 will deal with this routine for the 3D case, more in details. Many developments of these two approaches have been proposed in literature [66, 77, 109].

More recently, different FwP algorithms such as the distance-driven and the separable trapezoid footprints have been introduced, with lower computational costs and reduced high-frequency artifacts.
The *distance-driven* was proposed by De Mann in 2002 for the 2D CT imaging [75] and it has been extended to the 3D case in 2004 [76]. The idea was to project the pixel width onto the detector and then amplify its contribution with a certain weight, that is tipically given by the height of the crossed element. In this case, not only a point but the element in-plane expansion is projected: it provides a linear shadow onto the detector and we tipically talk about a rectangular footprint, if we assume to enlarge the shadow for the pixel height (as shown in figure 2.2(c)). Subsection 2.2.2 will focus on this approach for the 3D case.
The *separable trapezoid footprints* algorithm was introduced in 2010 by Long and Fessler in [72, 71]. In this method, we project all the vertices of the pixel onto the detector and the element footprint is approximated by a trapeze, to shape a more accurate footprint than in the distance driven case, as illustrated in figure 2.2(d).
We now focus on the ray-driven and distance-driven algorithms for the 3D CT imaging, because we use them for some SpCT simulations in chapter 4 and for DBT tests on real data in chapter 5 respectively.

2.2.1 Ray driven approach for 3D CT imaging

As mentioned before, the ray-driven is one of the "historical" approaches for the computation of the matrix coefficients $M_{i,j}$ of equation 2.5. The main idea is to perform a ray tracing, as requestes by the integral equation 1.12. Hence we follow each X-ray from its emission from the source, to its passage through the object, to its final detection in the center of a cell. At the end, we sum over the intersected discretized elements, balancing their influences with certain weights. This technique was introduced between 1982 and 1985 by different works [61, 91, 51], proposing different ideas for computing these weights. In this work, we consider the Siddon approach [91].

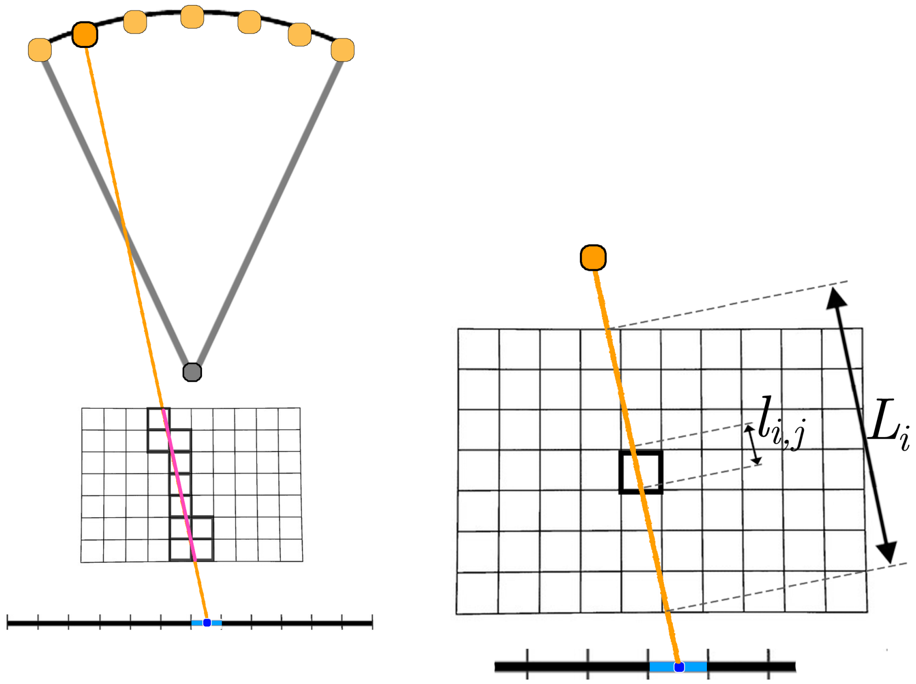


Figure 2.3: On the left: the ray driven approach for the forward projection on a DBT-like device. It is seen on the yz-plane, hence the detector is reduced to a 1D array of n_y elements and the volume to a grid of $N_y \cdot N_z$ voxels. On the right: a dratf explaining how the ray driven approach computes the matrix coefficients, again in the yz-plane.

Figure 2.3 drafts this methodology, projecting it to the bidimensional representation on the yz-plane, for a DBT machinery: the flat detector always lies on the xy-plane, parallel to the slices dividing the volume, and the source rotates over the detector in a C-shaped trajectory. Since each element $M_{i,j}$ of M reflects the influence of a single voxel f_j on a specific pixel g_i (for a fixed θ-angled image), we have to take into account the X-ray emitted from the θ-position and incident to the center of the i-th cell, then compute its path through the volume. It follows that

$$M_{i,j} = \frac{l_{i,j}}{L_i} \tag{2.6}$$

where $l_{i,j}$ is the length of the X-ray inside f_j and L_i is total length of the ray inside the object. By summing over all the voxels for $j = 1, \dots, N_v$ as in equation (2.2), the projection value g_i is determined. In this case, it results $M_{i,j} > 0$ if f_j is crossed by the X-ray, or $M_{i,j} = 0$ otherwise. Moreover, in a reasonable physical setting any row of M is normalized to 1, thanks to the $\frac{1}{L_i}$ factor, and each column has at least one element greater than zero, because each voxel is projected at least once onto the detector.

2.2.2 Distance driven approach for 3D CT imaging

In this subsection we explain in details the distance-driven (DD) method that we use in chapter 5, for the numerical experiments with real DBT data. The classical 3D algorithm in [76] is defined for a general CT process (i. e. suitable for circular detectors too) but we adjusted it to the specific mammographic setting.

The DD key-point consists in computing the overlap between a voxel footprint and the detector-cell "backward footprint". Typically, the comparison between these two areas is made on an arbitrary and intermediate plane, where it is possible to project the squared section of the voxel and backproject the surface of a recording cell (whatever form it has, flat or spherical): this plane is always chosen to be parallel to the volume layers. In the case of DBT, since the detector is a stationary flat panel and it is parallel to the compression planes of the breast (hence to all its layers), it would be enough to directely project any volume slice onto the detector or to backproject the detector grid onto each slice: the resulting implementation avoids several forward or backward footprint computations for every coefficient $M_{i,j}$, making faster the creation of the system matrix. Figures 2.4 and 2.5 help in understanding how these coefficients are indeed computed.

Recalling the notation introduced at the beginning of this chapter, Δ_x, Δ_y and Δ_z are the spatial resolution of the discretization into voxels of the volume, respectivelly along the $x-, y-, z-$axis, while δ_x, δ_y are the dimensions of each detector unit mounted on the DBT machinery. For prefixed scanning angle and pixel g_i, the FwP operator computes the "backward X-ray cone" having as a basis the detector pixel itself and vertex on the X-ray source (figure 2.4); then it determines the backward footprints in any slice, at its middle height $\frac{\Delta_z}{2}$ (figure 2.5(a)). Fixing one slice and calling A_i the area of this backward footprints on the layer and $a_{i,j}$ the intersection with the j-th voxel (with j such that f_j lies on the layer, of course), we compute

$$M_{i,j} = \frac{\Delta_z}{\alpha_i \gamma_i} \frac{a_{i,j}}{A_i} \tag{2.7}$$

for all the voxels on that slice.

In this equation, α_i and γ_i are the in- and out-of-plane angles, i. e. the two angles we can derive to describe the X-ray linking the source to the center of the i-th detector pixel. If we draw the perpendicular from the X-ray source to

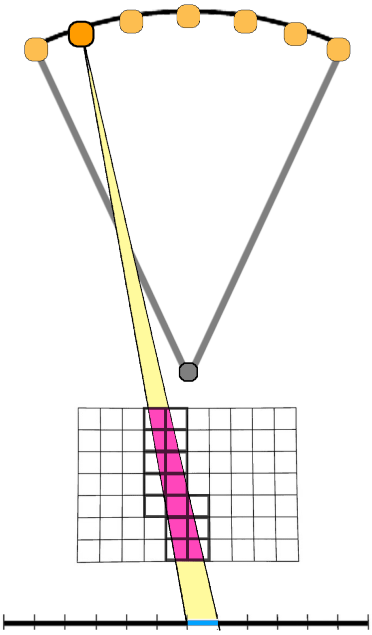
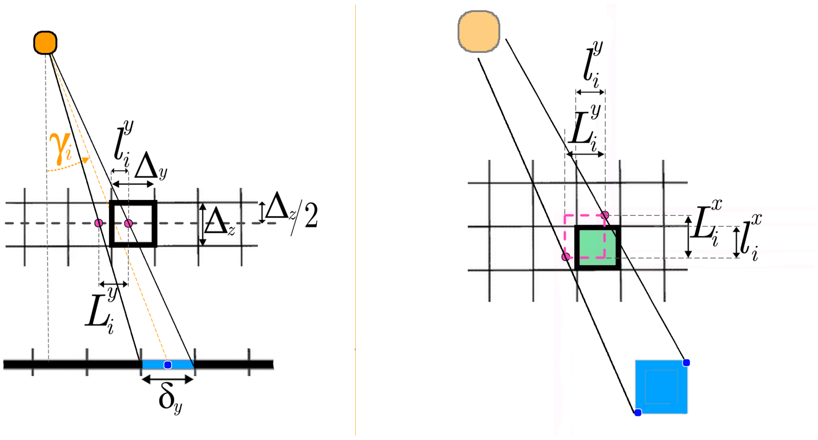


Figure 2.4: The DD approach for the forward projection on a DBT-like device. It is seen on the yz-plane, hence the detector is reduced to a 1D array of n_y elements and the volume to a grid of $N_y \cdot N_z$ voxels.

the detector and we trace the X-ray that comes down to the middle point of the i-th pixel, γ_i is the angle between these two elements on the $yz-$plane, as shown in figure 2.5(a), while α_i is that angle on the $xy-$plane. As a consequence, the weighting factor $\frac{\Delta_z}{\alpha_i \gamma_i}$ can be read as the length of the X-ray portion inside each voxel and it is also the quantity we use to "raise" the rectangular footprint. Moreover, this factor can be interpretated as a sort of normalization by $\frac{1}{r^2}$, which is known as the *inverse-square* physical law stating that a specific physical quantity (like the photon intensity in our case) is inversely proportional to the square of the distance from the source of that physical quantity. On the other hand, the ratio of the two areas stands for a normalization to 1 (if we omit the weighting factor for a while) over all the voxels lying on the same slice, because it reflects the percentage of influence of a certain voxel on the final measurement with respect to the other voxels on its same layer.

Implementation notes on the FwP operator

How does the FwP work in details? Once we have fixed, in this sequence, the scanning angle, the i-th recording unit (to which we aim to assign a value)



(a) (b)

Figure 2.5: On the left: the DD approach seen on the yz-plane. Only one slice is partially reported and one voxel is highlighted together with the values computed by the algorithm. On the right: the DD approach seen on a xy-plane, from above the source. The blue square stands for one recording element of the detector panel, the grid stands for the voxel discretization at height $\frac{\Delta_z}{2}$ inside the layer and the green square is the voxel on it for which we compute the matrix coefficient, the pink dashed square stands for the square backward footprint of the pixel on the grid.

and one of the volume layer, we have to performe the backward projection of the unit surface on the layer. To efficiently implement this last task, let us focus on figure 2.5(b). Starting from the cyan pixel, we can look for the intersections of the two rays connecting its two opposite vertices (the blue dots) to the source, at the middle altitude of the slice: we get therefore the coordinates of two pink points, from which we approximate the backprojected footprint into a rectangle (the pink and dashed one). This simplification allows to directely compute the area A_i as

$$A_i = L_i^x \cdot L_i^y \tag{2.8}$$

where L_i^x and L_i^y are the distances between the pink points, respectively along the x- and y-axes.

We need now to locate the two positions inside the layer grid, in order to get the indices of the involved voxels. After that, a cycle over all the touched voxels computes the resulting non-zeros coefficients of the i-th row of M. For these voxels, in particular, the overlapping area $a_{i,j}$ is again rectangular and it is given by

$$a_{i,j} = l_{i,j}^x \cdot l_{i,j}^y \tag{2.9}$$

where $l_{i,j}^x$ and $l_{i,j}^y$ are the overlapping lenghts along the x- and y-axes.

2.2.3 Main features of the system matrix

To conclude this section, we synthesize some important characteristics of the system matrix M, that are independent from the approach we use to define it. Under reasonable 3D tomographic setting, we can state that:

- M is a large $\mathbb{R}^{N_d} \times \mathbb{R}^{N_v}$ matrix, because the detector resolution makes the N_d value quite high and the number N_v of voxels must be large to have precise reconstructions: in some DBT cases, for instance, they could reach values of order of 10^9;

- $M_{i,j} \geq 0 \; \forall i, \forall j$ because negative values are not coherent with the CT process and the negative logarithmic transform of equation (2.1): if a voxel is crossed by the beam reaching one pixel, it influences the corresponding measurment with a "positive", effective absorbtion, otherwise its contribution is null (but never negative as no photons can be added to the beam by the volume elements);

- M is a very sparse matrix, because few voxels are effective for a single pixel value of a projection image, hence each row has mostly zero elements;

- M is underdetermined in case of SpCT, hence no unique solution exists for the CT linear system;

- M can not be stored because of its huge dimensions, neither in sparse form, for most of the real CT imaging: whenever we need a matrix product, M or M^T must be recalculated row by row and this represents a noticeable computational cost;

- M does not have empty columns, because the reconstruction is performed only on a scannable volume, hence at least one projection for every voxel is included in the dataset;

2.2.4 Physical interpretation of the transpose matrix

In CT imaging, many algorithms make use of a "backward projection" step to move from the dataset values back to the voxel intensities. Instead, if we use classical iterative solvers for the solution of a tomographic problem, we only handle with the matrix M and its transpose M^T to converge to the solution. We do not care about the physical meaning of that transposition basically, but the M^T matrix of a tomographic process has a precise role: it performs, in fact, the backward projection of a dataset onto a volume, accordingly to the CT device geometry defined in the FwP operator M.

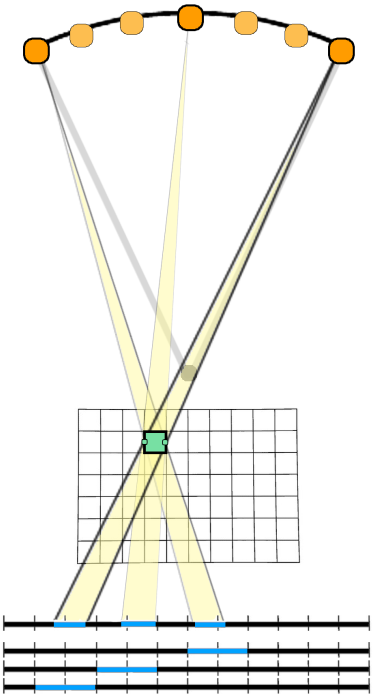


Figure 2.6: The DD approach for the backward projection on a DBT-like device. It is seen on the yz-plane, hence the detector is reduced to a 1D array of n_y elements and the volume to a grid of $N_y \cdot N_z$ voxels.

We describe this process with a distance-driven scheme (for a DBT device), but the following observations may be extended to all the FwP approaches. If we fix the j-th voxel, that belongs to a certain slice of the volume discretization, we can cycle over the N_θ angles and project the voxel onto the detector, as shown in figure 2.6. In this case, the projection plane could be the one the detector panel lies in for semplicity, since we need to identify all the detector elements involved for each projection on the corresponding projected image (similarly to what we have done for the FwP operator). At the end, we cycle on these elements and compute the ratio of overlapping areas on the detector plane and the weighting factors too, with the same approach of the FwP in equation (2.7). Thanks to the ratio between areas, it does not matter in which plane we are computing the areas, hence this procedure provides exactly the same coefficients both for the forward- and the backward-projections. For this reason, the cycle provides exactely the elements of the j-th column of M and the backprojection of a N_d-dimensional vector g into a volume $\tilde{f}$ can be seen as

$$\tilde{f} = M^T g. \tag{2.10}$$

To conclude, we remark that the earlier commercial CT softwares were based on a filtered Back Projection algorithms, where the tomographic dataset was retrospectively projected (and filtered in the Fourier domain) to achieve the reconstruction, hence this approach is still present but in a very different implementation scheme.

2.3 The model-based formulation for the SpCT

The CT reformulation in (2.5) hides several drawbacks. First of all, the presence of noise on real data makes this system inconsistent. In addition, M does not have any standard form but big dimensions, hence any matrix factorization-based method is prohibitive. Moreover, equation (1.12) is a Fredholm integral equation of the first kind, i. e. it is an ill posed problem and the linear system arising from its discretization is usually highly ill-conditioned. As a result, standard solutions will not be stable with respect to small changes in g, hence the CT model should be adjusted exploiting some kind of regularization.
Furthermore, in the case of SpCT where $N_p \cdot N_\theta < N_v$, infinite solutions exist for (2.5) and the introduction of the a priori information into the model is necessary to choose one of the possible infinite solutions. In this thesis, we formulate the reconstruction from limited data as an optimization problem where the Total Variation regularizer balances a data-fitting function. The resulting minimization problem will be solved by some iterative methods that are discussed in the next chapter.

2.3.1 Compressive Sensing and medical imaging

The wide literature about medical imaging and iterative solvers has been largely influenced by the Compressed Sensing (CS) theory.
To understand its approach, it could be useful to start this dissertation by defining some basic CS concepts, as they were precisely cleared up by Graff and Sidky in [50]. First of all, an Iterative Image Reconstruction (IIR) method can be labelled as Sparsity-Exploiting Image Reconstruction (SEIR) when it exploits some form of sparsity of the image (i.e. a sparsity on a certain mathematical domain), as an a priori information. Among all the SEIR methods, Compressed Sensing can be defined as a set of models and related methods that are specifically intended to improve image quality with a reduced sampling. Furthermore, when CS is reinforced by mathematical proofs or empirical accurate studies, we can talk about Compressive Sensing Guarantee (CSG) as a ”security” passport: in 2006, papers by Candés and Donoho [27, 28, 38] communicated these CSG concepts to the CT imaging community. Since then, many papers address the issue of work in SEIR/CS prior as CSGs.
However, one of the main features of the CSG is the idealization of the imaging systems, for which the *exact* image reconstruction can be reached from

a sparse sampling, while for the medical community the mathematical exact image does not provide the most useful visualization necessarily [31, 98]. For this reason, Graff and Sidky have observed that TV-based approaches for sparse tomography have been in use long before the concept of CS guarantees and CS is often confused with optimization problems involving ℓ_1 norms or related solvers. Hence, since CS is also a distinct field of study from medical imaging, CS guarantees should olny be considered as potentially useful IIR design tools, which provide practical intuitions to guide IIR algorithm developments.

2.3.2 The regularized model for the SpCT

The SpCT community has been recently pushed towards the *model-based* optimization where a penalty term containing a priori information on the solution image is introduced to avoid artifacts, due to data inconsistency or poor conditioning of the linear system [93, 95, 92, 106, 103, 53, 73, 105, 108, 40, 104, 58].
As a consequence, due to the underdetermined peculiarity of Sparse Tomography, the imposition of a prior $\mathcal{R}$ on the CT system forces the IIR to look for one precise solution $\bar{f}$, among all the infinite solutions of (2.5). More in detail, the CT problem can be reformulated as a constrained or uncostrained minimization problem, where an objective function is defined by taking into account a data-fitting term $\mathcal{F}$ and/or some form of regularity $\mathcal{R}$. Different formulations are proposed in literature:

$$arg \min_{\mathcal{R}(f)<\epsilon} \mathcal{F}(f; M, g) \tag{2.11}$$

$$arg \min_{\mathcal{F}(f;M,g)<\delta} \mathcal{R}(f) \tag{2.12}$$

$$arg \min_{f} \mathcal{F}(f; M, g) + \lambda\mathcal{R}(f) \tag{2.13}$$

where ϵ, δ and λ are parameters balancing the influence of the two quantities on the output.

It is interesting to highlight that the unconstrained formulation 2.13 can be seen, from a statistical perspective, in an equivalent model stated as

$$\bar{f} = arg \min_{f} \{P(f|g) + \mathcal{R}(f)\}. \tag{2.14}$$

Here, $P(f|g)$ is called likelihood term and quantifies the probability to get a certain recostruction f for a given data g: it represents a data-fidelity function. $\mathcal{R}(f)$ is again the prior term, including the "a priori" information on f. Following this further approach, a class of IIR methods has been developped and labelled as *Statistical Solvers*.

In this thesis, we fix the unconstrained minimization problem of equation 2.13. For what concerns the fit to data function $\mathcal{F}(f; M, g)$ which reflects a distance between the left and the rigth hand side of equation 2.5, its expression is related to the noise on the data. If the noise can be modeled with a Gaussian distribution, the linear Least Squares (LS) function [19]

$$LS(f) = \frac{1}{2}\|Mf - g\|_2^2 \tag{2.15}$$

gives the best fitting, while if the noise has a Poisson distribution, the nonlinear Kullback-Leibler (KL) divergence [65]

$$KL(f) = \frac{1}{2}\sum_{i=1}^{N_d}\left(\sum_{j=1}^{N_v} M_{ij}f_j + bg - g_i - g_i \ln\left(\frac{\sum_{j=1}^{N_v} M_{ij}f_j + bg}{g_i}\right)\right) \tag{2.16}$$

(bg is the background value) is the best estimator.

The noise on the CT data is mixed Poisson (due to the X-rays particles behaviour) and Gaussian (due to the recording digital system) and the dominant one depends on the particular system we consider.

In literature, many choises have been proposed for the function $\mathcal{R}(f)$: it should reduce the noise, regularize the ill-conditioned problem arising from the discretization of an ill-posed Fredholm integral equation and impose some sparsity on the computed solution following the CS theory. Since many medical images are almost uniform inside the organs, much work of the IIR has been inspired by edge-preserving techniques, where the regularization mainly focuses on a noise reduction, but preserving the edges of the image at the expense of the exact reconstruction inside the main objects. Among the most used regularizers, we find the Total Variation (TV) semi-norm [106, 64, 88, 93, 94, 95, 92, 96] and the Huber penalty function [57]. The nonlinear TV was introduced in denoising with the well-known paper of Rudin, Osher and Fatemi in 1992 ([89]): for a smooth continuous function $f = f(x, y, z)$, the TV operator is essentially defined as the ℓ_1 norm of derivatives, that is [100]:

$$\mathcal{TV}(f) = \int_0^1\int_0^1\int_0^1 |\nabla f| dx\ dy\ dz \tag{2.17}$$

where $|w| = |(x, y, z)| = \sqrt{x^2 + y^2 + z^2}$ denotes the Euclidean norm in $\mathbb{R}^3$ and $\nabla f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}\right)$.

Due to the nondifferentiability of $\mathcal{TV}$ in zero, one can take an approximation $|w| \approx \sqrt{|w|^2 + \beta^2}$, where β is a small positive parameter. In this case, the TV function is approximated by the following smoothing differentiable function

$$\mathcal{TV}_\beta(f) = \int_0^1\int_0^1\int_0^1 \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2 + \left(\frac{\partial f}{\partial z}\right)^2 + \beta^2}\ dx\, dy\, dz \tag{2.18}$$

In particular, by fixing a Total Variation regularizer we ask for a sparsity in the gradient domain of the image: this SEIR approach is coherent with the need of reliable detection of anatomical tissues, that is much more important than the reconstruction of the exact "real" anatomy.

2.3.3 The TV discretization

If we consider the discretized volume f in a vector shape, the discretization of equation (2.18) leads to

$$TV_\beta(f) = \sum_{j=1}^{N_v} \sqrt{\|\nabla f_j\|_2^2 + \beta^2}, \tag{2.19}$$

but we must keep in mind that $\|\nabla f_j\|$ has to be computed with respect to the three cartesian directions.

To avoid misunderstanding, we introduce here the discretization of the problem in the 3D setting, by using the notation j_x, j_y, j_z to indicate the coordinates inside the object of the j-th voxel, with respect to the x, y, z axis respectively.
When the TV_β function is discretized by forward differences with boundary periodic conditions in each voxel (j_x, j_y, j_z), the discrete TV_β function of the image is

$$TV_\beta(f) := \sum_{j_x=1}^{N_x} \sum_{j_y=1}^{N_y} \sum_{j_z=1}^{N_z} \sqrt{\frac{(f_{j_x+1,j_y,j_z} - f_{j_x,j_y,j_z})^2}{\Delta_x^2} + ... \atop ... + \frac{(f_{j_x,j_y+1,j_z} - f_{j_x,j_y,j_z})^2}{\Delta_y^2} + \frac{(f_{j_x,j_y,j_z+1} - f_{j_x,j_y,j_z})^2}{\Delta_z^2} + \beta^2} \tag{2.20}$$

where Δ_x, Δ_y and Δ_z are the dimensions of the voxels, as already introduced in section 2.1. The expression (2.20) can be written as

$$TV_\beta(f) := \frac{1}{2} \sum_{j_x=1}^{N_x} \sum_{j_y=1}^{N_y} \sum_{j_z=1}^{N_z} \phi(D^2 f_{j_x,j_y,j_z}) \tag{2.21}$$

where

$$\begin{aligned} D^2 f_{j_x,j_y,j_z} = & \left(\frac{f_{j_x+1,j_y,j_z} - f_{j_x,j_y,j_z}}{\Delta_x}\right)^2 + \left(\frac{f_{j_x,j_y+1,j_z} - f_{j_x,j_y,j_z}}{\Delta_y}\right)^2 + \\ & + \left(\frac{f_{j_x,j_y,j_z+1} - f_{j_x,j_y,j_z}}{\Delta_z}\right)^2 \end{aligned} \tag{2.22}$$

implements the forward differences and

$$\phi(t) = 2\sqrt{t + \beta^2} \tag{2.23}$$

represents the differentiable approximation of the Total Variation.

Whenever we need to compute the gradient of the TV_β function, we can exploit this notation and derive the following formula

$$\nabla TV_\beta(f) = \sum_{j_x=1}^{N_x}\sum_{j_y=1}^{N_y}\sum_{j_z=1}^{N_z} \phi'(D^2 f_{j_x,j_y,j_z}) D f_{j_x,j_y,j_z} \nabla D f_{j_x,j_y,j_z} \tag{2.24}$$

where $\phi'(t) = \frac{1}{\sqrt{t+\beta^2}}$.
It means that for every element of indices (j_x, j_y, j_z), the discrete derivative of the TV_β function is:

$$\begin{aligned} \frac{\partial TV_\beta}{\partial f_{j_x,j_y,j_z}}(f) = {} & \phi'(D^2_{j_x-1,j_y,j_z}) \frac{f_{j_x,j_y,j_z} - f_{j_x-1,j_y,j_z}}{\Delta_x} + \\ & + \phi'(D^2_{j_x,j_y-1,j_z}) \frac{f_{j_x,j_y,j_z} - f_{j_x,j_y-1,j_z}}{\Delta_y} + \\ & + \phi'(D^2_{j_x,j_y,j_z-1}) \frac{f_{j_x,j_y,j_z} - f_{j_x,j_y,j_z-1}}{\Delta_z} + \\ & \phi'(D^2_{j_x,j_y,j_z}) \left(\frac{f_{j_x,j_y,j_z} - f_{j_x+1,j_y,j_z}}{\Delta_x} + \frac{f_{j_x,j_y,j_z} - f_{j_x,j_y+1,j_z}}{\Delta_y} + \frac{f_{j_x,j_y,j_z} - f_{j_x,j_y,j_z+1}}{\Delta_z} \right) \end{aligned} \tag{2.25}$$

2.3.4 Unconstrained and constrained formulation

From now on, we keep fixed the Total Variation regularizer in the differentiable and discrete form of equation (2.20) and the optimization problem we aim to solve is finding $\bar{f}$ such that

$$\bar{f} = arg \min_{f\in\mathbb{R}^{N_v}} \mathcal{J}(f) = arg \min_{f\in\mathbb{R}^{N_v}} \{\mathcal{F}(f; M, g) + \lambda TV_\beta(f)\} \tag{2.26}$$

We can easily show that:

Proposition 2.3.1. *If $\mathcal{F}(f; M, g)$ has the form* (2.15) *or* (2.16)*, then the solution of problem* (2.26) *exists and is unique.*

Proof. For what concerns the existence of the solution of (2.26), the function $\mathcal{J}(f)$ is coercive, hence we refer to the Weierstrass theorem [80]. To prove the uniqueness, we show the strict convexity of the function $\mathcal{J}(f)$ on the domain. Since $Ker(TV_\beta(x))$ contains only the constant images, while $Ker(\nabla^2_f(\mathcal{F}(f; M, g))$ does not contain constant images except the null image, both in the case of LS and KL, then:

$$\begin{aligned} Ker(\nabla^2_f(\mathcal{F}(f; M, g)) + \lambda\nabla^2(TV_{\beta(f)})) &= Ker(\mathcal{F}(g, Mf)) \bigcap Ker(TV_\beta(f)) \\ &= \{0\} \end{aligned} \tag{2.27}$$

and $\mathcal{J}(f)$ is strictly convex. □

On the other hand, as already mentioned, the reconstructed image is a map of the linear attenuation coefficient μ that has always non-negative values. Hence it is quite common to consider a nonnegative constrained model as

$$\bar{\bar{f}} = arg \min_{f \geq 0} \mathcal{J}(f) = arg \min_{f \geq 0} \{\mathcal{F}(f; M, g) + \lambda TV_\beta(f)\} \tag{2.28}$$

In particular, the above proposition still holds for this problem, because the objective function is coercive also on the domain $\Omega = \{f \in R^{N_v} : f \geq 0\}$.

Chapter 3

Reconstruction algorithms for Sparse CT problems

In section 2.1 we have derived the formulation of the SpCT process into a linear system, characterized by a huge sparse matrix (equation(2.5)). As this system is underdetermined, the SpCT problem has been reformulated as an optimization problem with a prior term that guarantees uniqueness and stability of the solution, in paragraph 2.3. When a minimization problem is solved with an Iterative Image Reconstruction (IIR) method, the convergence solution is often achieved quickly but at least two matrix-vector products are required at each iteration: the high computational cost per iteration is the main drawback of an IIR algorithm use in large scale problems. On the other hand, thanks to the dramatic improvement of CPUs speed and the possibility of performing parallel computations at low cost on GPUs, the time per iteration of an IIR algorithm may be incredibly reduced.
Furthermore, reliable reconstructions must be provided almost in real-time in SpCT applications, hence an iterative solver is always stopped before its actual convergence. For this reason, a *good* algorithm must get close to the convergence solution in its very first steps. To do that efficiently, an IIR method must be enforced by no computationally onerous acceleration techniques, of course.

In this chapter, an overview of the state-of-the-art algorithms for sparse tomography is presented in section 3.1, then two solvers are shown in details: the Scaled Gradient Projection algorithm (SGP) in section 3.2 and the Lagged Diffusivity Fixed Point (FP) in section 3.3. The SGP is a gradient descent first-order method and it is applied to the constrained problem (2.26) with both the LS and the KL functions as data-fitting quantity; the FP algorithm, instead, follows a quasi-Newton approach and it solves the unconstrained minimization (2.28) where the objective function is the LS fidelity term. In addition, the FP implementation requires to solve an inner linear system, making its tuning more difficult for practical use. Nevertheless, they both are equipped with accelerating strategies making them competitive in their earlier iterations and suitable for SpCT image reconstructions.

3.1 Iterative Image Reconstruction state-of-the-art

Initial effort to solve (2.5) with IIR for CT took an algebraic approach, where the linear system is solved by sequentially projecting onto hyperplanes, each defined by a row of the matrix M.
In the 1970, the maiden implementations of an *Algebraic Reconstruction Technique* (ART) was introduced by Gordon [49] to the imaging community, as a 3D electron microscopy tool, which is quite similar to the Kaczmarz algorithm for the resolution of a system of linear equations [62]. Two years later, Gilbert [46] proposed a Simultaneous Iterative Reconstruction Technique (SIRT), while in 1984 Anderson [12] introduced a development of the ART, called Simultaneous ART (SART), and later on he further developed it in [11] for sparse tomography. The main idea behind ART and SART is to iteratively update the current image after processing subsets or some elements of the projection data set. In particular, in the running iteration, a forward projection of the current iterate is computed and its comparison with the measured data establishes an updating term which is backprojected afterwards onto the image domain. Such term involves only the elements related to a single ray (i. e. a single measurement) in the ART approach, or all the elements of a specific projection image.
However, the algebraic methods aim to solve the linear system directely, hence the ART approach tends to provide noisy but fast reconstruction, while SIRT tipically results in good denoising but a lower convergence speed. The SART has both good noise results and the ART-like fast convergence property: implementations of SART-based methods are still settled on many commercial softwares.

On the other hand, the novel *model-based* approach was proposed in literature because an increasing interest in Compressive Sensing (CS) has redoubled the effort in improving optimization problem solvers. Two main cathegories of CS-motivated algorithms can be distinguished [50]: the one of approximate solvers, based on sequential algorithms, and the one of accurate solvers for non-smoothed large-scale optimization problems.
The *approximate sequential algorithms*, on the contrary, may provide reliable reconstructions in ten or fewer iterations, but the output quality strongly depends on the parameters of both the optimization problem and the chosen algorithm for the resolution. In particular, all these methods update the current image after processing subsets or some elements of the preprojection

data set: it results in a strong acceleration at the first steps of the convergence towards a good image, even if the convergence to the optimal solution can be slower from a certain iteration on. In this class we can find many methods solving a TV-regularized problem and starting from the well-known POCS algorithm, such as the Adaptive Steepest Descent Projection Onto Convex Sets ASD-POCS algorithm [94] and other [30, 41]. We can include here also some developments of the algebraic methods.

The group of *accurate solvers* containes many standard methods that have been adapted to the hugeness of realistic system size only recently, but they have not been largely applied to real SpCT problems up to now. They use slightly smoothed versions of optimization problems involving TV in one of the formulations presented in equations (2.11), (2.12) and (2.13). In this class, we find algorithms like the gradient descent, the nonlinear conjugate gradients and a limited memory BFGS with the approximate Newton's method [57, 68, 69]. The main drawback of these solvers is that they typically require hundreds of iterations for a good CT image, hence great computational resources, because the convergence rates vary significantly depending on the CT scanning configuration and data quality. However, recent works propose improved versions of some first-order method, suited for realistic big scale problems, such as the split Bregman [13], ADMM [87] and the Chambolle-Pock [29, 83, 92] approaches. At last, accurate solvers eliminates the problem of deeply studing the dependence of the output on the algorithm parameters and such robustness is an appreciable feature for real images, where the parameter setting of the solving software is fixed once at all.

Nowadays, sequential algorithms seems to be favored for CT problems because of their faster convergence, but aim of this thesis is to test the robustness and efficiency of two accurate solvers, both on simulations and on real measurements, when they are applied to the complicated issue of 3D imaging reconstruction from severe subsampling.

3.2 Scaled Gradient Projection method

Aim of this section is to illustrate a first order accurate solver in the class of Gradient Projection (GP) methods [18], for the solution of the constrained formulation (2.28) of the SpCT problem. This approach is accelerated by a scaling matrix which improves the matrix conditioning: the resulting algorithm is thus called *Scaled Gradient Projection* (SGP) method [23]. In addition, two clever rules for the choice of an inner steplength parameter are shown: they both provide a further acceleration to the scaled algorithm and good quality reconstructions may be achieved in few iterations. Often, algorithms acceleration techniques introduce new parameters that must be tuned for each experiment: in our case, the proposed algorithm is robust with respect to the new parameters that are fixed on the basis of the applications considered and they don't change for the different experiments.

SGP algorithms have been successfully proposed in image restoration applications [20, 24, 107], where the problem formulation is very similar to (2.28), but the problem size is much smaller than in 3D CT imaging and the involved matrix had a structure that can be exploited in the matrix-vector product by using a Fast Fourier Transform. Hence the efficiency of the SGP methods in the considered SpCT applications was not predictable.

We recall now a general framework that allows to design first-order descent methods for the resolution of an optimization problem with non-negative constraints. This framework is described in table 3.1 and schematically drawn in figure 3.1, with both scaled gradient directions and effective step-length selection rules.

```
Initialize: $f^{(0)} \geq 0,\quad S_0 \in \mathcal{S}_{\rho_0},\quad \gamma, \sigma \in (0,1),\quad 0 < \alpha_{min} \leq \alpha_{max}$,
            $\alpha_0 \in [\alpha_{min}, \alpha_{max}]$;
for  $k = 0, 1, \dots$
    $d^{(k)} = P_+\left(f^{(k)} - \alpha_k S_k \nabla \mathcal{J}(f^{(k)})\right) - f^{(k)}$; (scaled gradient projection step)
    $\eta_k = 1$;
    while  $\mathcal{J}(f^{(k)} + \eta_k d^{(k)}) > \mathcal{J}(f^{(k)}) + \sigma \eta_k \nabla \mathcal{J}(f^{(k)})^T d^{(k)}$
        $\eta_k = \gamma \eta_k$;                                   (backtracking step)
    end
    $f^{(k+1)} = f^{(k)} + \eta_k d^{(k)}$;
    define the diagonal scaling matrix $S_{k+1} \in \mathcal{S}_{\rho_{k+1}}$; (scaling updating rule)
    define the step-length $\alpha_{k+1} \in [\alpha_{min}, \alpha_{max}]$;  (step-length updating rule)
end
```

Table 3.1: Algorithm SGP (Scaled Gradient Projection).

The following notation is used in the SGP description: $\mathcal{S}_\rho$ denotes the set of diagonal matrices S with entries $s_{j,j},\ j = 1, \dots, N_v$, such that $\frac{1}{\rho} \leq s_{j,j} \leq \rho$, with $\rho > 1$, and $P_+(z)$ is the euclidean projection of the vector $z \in \mathbb{R}^{N_v}$ onto the non-negative orthant $\Omega = \{f \in \mathbb{R}^{N_v} | f_j \geq 0\ \forall j = 1, \dots, N_v\}$.
At the k-th iteration of the SGP algorithm, a matrix $S_k \in \mathcal{S}_{\rho_k}$ is used to define the scaled gradient direction $-S_k \nabla \mathcal{J}(f^{(k)})$ along which we move from the current point $f^{(k)}$ with a step established by the step-length parameter $\alpha_k > 0$. After that, by projecting the resulting vector onto the nonnegative orthant Ω, the descent direction $d^{(k)}$ is obtained. This direction is then exploited in a linesearch procedure to generate the new approximation $f^{(k+1)}$ in which the objective function achieves a sufficient reduction with respect to its current value $\mathcal{J}(f^{(k)})$. Finally, the scaling matrix and the step-length parameter are re-evaluated for a new iteration, by means of suitable adaptive updating rules.

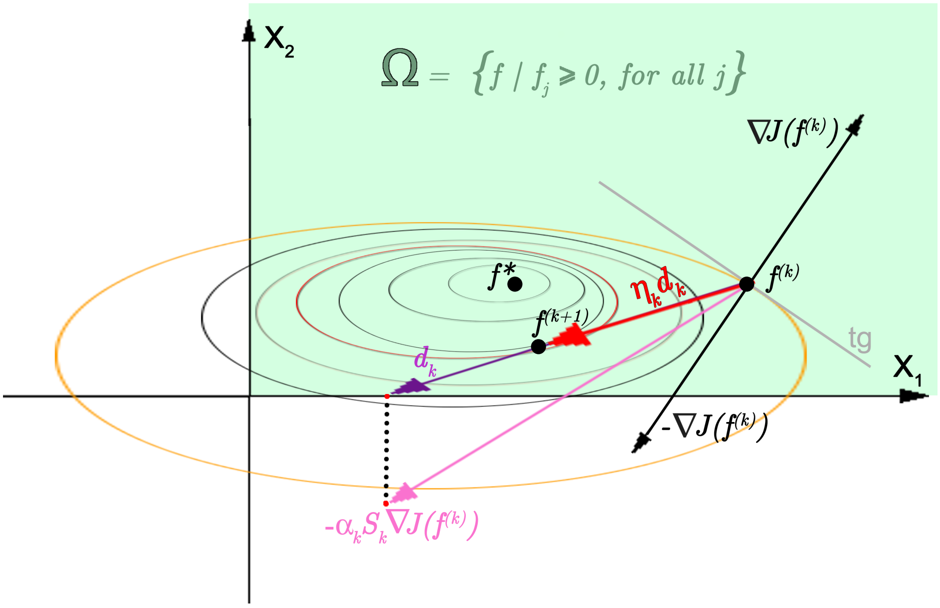


Figure 3.1: Draft of the $k-$th iteration of the SGP algorithm, in $\mathbb{R}^2$.

For an effective application of SGP to the problem (2.28), important theoretical and practical aspects need to be discussed. The main convergence properties of SGP are stated in the following proposition, whose proof can be found in [22].

Proposition 3.2.1. *Let* $\{f^{(k)}\}$ *be the sequence generated by applying SGP to the problem* (2.28)*. The following properties hold:*

i) if $\rho_k^2 = 1 + \zeta_k$, $\zeta_k \geq 0$, $\sum_{k=0}^{\infty} \zeta_k < \infty$, *then the sequence* $\{f^{(k)}\}$ *converges to the solution of the problem.*

ii) if $\mathcal{J}^*$ *denotes the optimal value for the objective function of the optimization problem, then*

$$\mathcal{J}(f^{(k)}) - \mathcal{J}^* = \mathcal{O}(1/k)$$

The proposition ensures that SGP converges without restrictive assumptions on the step-length parameter α_k and the diagonal scaling matrix S_k, whose choices can be directed to accelerate the convergence rate of the scheme. Even if the theoretical convergence rate $\mathcal{O}(1/k)$ on the objective function values is lower than the rate $\mathcal{O}(1/k^2)$ of some optimal first-order methods exploiting extrapolation steps, the practical performance of SGP achievable by suitable selections of S_k and α_k is very well comparable with that of the optimal algorithms. In the following we provide the updating rules for S_k and α_k that allow SGP to efficiently solve problem (2.28).

3.2.1 Focus on the scaling strategy

In order to update the diagonal scaling matrix S_{k+1}, we adapt to the SGP framework the strategy proposed in [67], that in several applications of scaled gradient methods has shown the ability to force convergence acceleration, especially in the first steps of the iterative process [23, 20]. The strategy exploited in [67] leads to define the diagonal scaling matrix by means of special splittings of the gradient of the objective function, which is

$$\nabla \mathcal{J}(f) = V^{\mathcal{J}}(f) - U^{\mathcal{J}}(f), \quad \text{with } V^{\mathcal{J}}(f) > 0,\ U^{\mathcal{J}}(f) \geq 0. \tag{3.1}$$

In particular, $V^{\mathcal{J}}(f)$ and $U^{\mathcal{J}}(f)$ are obtained by splitting both the fit to data function and the regularization term in

$$V^{\mathcal{J}}(f) = V^{\mathcal{F}}(f) + \lambda V^{TV}(f), \quad U^{\mathcal{J}}(f) = U^{\mathcal{F}}(f) + \lambda U^{TV}(f),$$

where

$$\begin{array}{rcll} \nabla \mathcal{F}(f) & = & V^{\mathcal{F}}(f) - U^{\mathcal{F}}(f), & V^{\mathcal{F}}(f) > 0, \quad U^{\mathcal{F}}(f) \geq 0, \\ \nabla TV_{\beta}(f) & = & V^{TV}(f) - U^{TV}(f), & V^{TV}(f) > 0, \quad U^{TV}(f) \geq 0. \end{array} \tag{3.2}$$

Given the splitting (3.1), in [67] the choice $s_{j,j}^{(k+1)} = \frac{f_j^{(k+1)}}{V_j^{\mathcal{J}}(f^{(k+1)})}$, $\quad j = 1, \ldots, N_v$, is suggested for the diagonal entries of the matrix S_{k+1}. Exploiting a similar idea within the SGP scheme, we propose to update the diagonal scaling matrix by setting

$$s_{j,j}^{(k+1)} = \min\left(\rho_{k+1}, \max\left(\frac{1}{\rho_{k+1}}, \frac{f_j^{(k+1)}}{V^{\mathcal{J}}(f^{(k+1)})}\right)\right), \qquad j = 1, \ldots, N_v. \tag{3.3}$$

i. e. by adding an upper and a lower bound to the main splitted entries. Following the suggestions in [21, 35], the parameter ρ_{k+1} is chosen as $\rho_{k+1} = \sqrt{1 + 10^{15}/(k+1)^{2.1}}$, hence the resulting variable bounds slowly reduce the feasible range for the scaling values $s_{j,j}$ and we suit the first point of Prop 3.2.1.

Of course, the vectors $V^{\mathcal{F}}(f^{(k+1)})$ and $V^{TV}(f^{(k+1)})$ are set taking into account the special form of $\nabla \mathcal{F}(f)$ and $\nabla TV_{\beta}(f)$, respectively.
When the fitting function is the LS one (equation (2.15)), the gradient of $\mathcal{F}(f)$ has the form

$$\nabla \mathcal{F}(f) = M^T M f - M^T g;$$

hence we can split it into

$$V^{\mathcal{F}}(f^{(k+1)}) = M^T M f^{(k+1)}, \quad U^{\mathcal{F}}(f^{(k+1)}) = M^T g \ \forall k.$$

On the other hand, when the fit-to-data term is the KL function of equation (2.16), we have

$$\nabla \mathcal{F}(f)) = M^T \mathbf{1} - M^T Y^{-1} g,$$

where $\mathbf{1} \in \mathbb{R}^{N_d}$ is a vector whose components are all equal to one and $Y = \mathrm{diag}(Mf + bg)$ is a diagonal matrix with the entries of $(Mf + bg)$ on the main diagonal. Therefore we can define

$$V^{\mathcal{F}}(f^{(k+1)}) = M^T\mathbf{1}, \;\; U^{\mathcal{F}}(f^{(k+1)}) = M^T Y^{-1} g$$

where the dependance of k is hidden in the Y diagonal matrix.
Focusing on the regularizer and recalling the notation introduced in 2.3.3, from equation (2.25) we derive the splitting functions of $\nabla TV(f^{(k+1)})$ as

$$\begin{aligned} V^{TV}_{j_x,j_y,j_z}(f^{(k+1)}) = &\left[\frac{\phi'(D^2_{j_x,j_y,j_z})}{\Delta_x} + \frac{\phi'(D^2_{j_x,j_y,j_z})}{\Delta_y} + \frac{\phi'(D^2_{j_x,j_y,j_z})}{\Delta_z} + \right. \\ &\left. \frac{\phi'(D^2_{j_x-1,j_y,j_z})}{\Delta_x} + \frac{\phi'(D^2_{j_x,j_y-1,j_z})}{\Delta_y} + \frac{\phi'(D^2_{j_x,j_y,j_z-1})}{\Delta_z} \right] f^{(k+1)}_{j_x,j_y,j_z} \end{aligned} \tag{3.4}$$

and

$$\begin{aligned} U^{TV}_{j_x,j_y,j_z}(f^{(k+1)}) = &\phi'(D^2_{j_x,j_y,j_z}) \left(\frac{(f^{(k+1)})_{j_x+1,j_y,j_z}}{\Delta_x} + \frac{(f^{(k+1)})_{j_x,j_y+1,j_z}}{\Delta_y} + \frac{(f^{(k+1)})_{j_x,j_y,j_z+1}}{\Delta_z} \right) + \\ &\phi'(D^2_{j_x-1,j_y,j_z}) \frac{(f^{(k+1)})_{j_x-1,j_y,j_z}}{\Delta_x} + \\ &\phi'(D^2_{j_x,j_y-1,j_z}) \frac{(f^{(k+1)})_{j_x,j_y-1,j_z}}{\Delta_y} + \\ &\phi'(D^2_{j_x,j_y,j_z-1}) \frac{(f^{(k+1)})_{j_x,j_y,j_z-1}}{\Delta_z} \end{aligned} \tag{3.5}$$

3.2.2 Focus on two step-lenght rules

Once the scaling matrix S_{k+1} has been defined, a new value for the step-length α_{k+1} must be computed with the aim to achieve further acceleration of the iterative process. The wide literature of the last decades identifies the Barzilai-Borwein (BB) rules as effective selection rules for the step-length updating in gradient methods [17]. In particular, recent studies have shown that selection rules based on special adaptive alternations of the two classical BB rules generally provide the best performance [110, 45]. For these reasons, first we derive the two BB rules for the step used by SGP and then we exploit these step-lengths rules within the alternating strategy proposed in [45], that has provided interesting convergence rate acceleration in many imaging algorithms [23, 84]. Due to the use of scaled gradient directions, by applying the quasi-Newton properties on which the classical BB rules are based, the following step-lengths are obtained:

$$\alpha_{k+1}^{\mathrm{BB1}} = \arg\min_{\alpha_k \in \mathbb{R}} \|(\alpha_{k+1} S_{k+1})^{-1} h^{(k)} - z^{(k)}\| = \frac{{h^{(k)}}^T S_{k+1}^{-1} S_{k+1}^{-1} h^{(k)}}{{h^{(k)}}^T S_{k+1}^{-1} z^{(k)}} \tag{3.6}$$

and

$$\alpha_{k+1}^{\mathrm{BB2}} = \arg\min_{\alpha_k \in \mathbb{R}} \|h^{(k)} - (\alpha_{k+1} S_{k+1}) z^{(k-1)}\| = \frac{{h^{(k)}}^T S_{k+1} z^{(k)}}{{z^{(k)}}^T S_{k+1} S_{k+1} z^{(k)}} \tag{3.7}$$

where $h^{(k)} = \left(f^{(k+1)} - f^{(k)}\right)$ and $z^{(k)} = \left(\nabla \mathcal{J}(f^{(k+1)}) - \nabla \mathcal{J}(f^{(k)})\right)$.
The alternating strategy ABB introduced in [45] leads to the following selection rule:

```
if $\alpha_{k+1}^{\mathrm{BB2}} / \alpha_{k+1}^{\mathrm{BB1}} < \tau$
    $\alpha_{k+1}^{\mathrm{ABB}} = \min\left\{\alpha_j^{\mathrm{BB2}} \,|\, j = \max\{1, k+1-m_\alpha\}, \ldots, k+1\right\}$ , $\tau = 0.9\,\tau$,
else
    $\alpha_{k+1}^{\mathrm{ABB}} = \alpha_{k+1}^{\mathrm{BB1}}$,    $\tau = 1.1\,\tau$,
end;
```

where m_α is a nonnegative integer and τ_k is a positive real number.

Other step-length updating rules are currently investigated in literature [14, 48, 36]; in particular, in 2015 Porta [85] proposed a strategy based on the use of a limited number m of "previous" gradients for capturing second-order information.
This rule can be explained as follows. Let us start denoting by G the $N_v \times m$ matrix we can define taking the last m scaled gradients, i. e.

$$G = \left[S_{k-m+1}^{\frac{1}{2}} \widetilde{g}^{(k-m+1)}, \ldots, S_k^{\frac{1}{2}} \widetilde{g}^{(k)} \right], \tag{3.8}$$

where

$$\widetilde{g}_j^{(k)} = \begin{cases} 0 & if\ f_j^{(k)} = 0, \\ \left[\nabla \mathcal{J}(f^{(k)})\right]_j & if\ f_j^{(k)} > 0. \end{cases} \tag{3.9}$$

Let R be the $m \times m$ upper triangular matrix obtained by the Cholesky factorization of $G^T G$ and consider the resulting small linear system

$$R^T r = G^T S_{k+1}^{\frac{1}{2}} \widetilde{g}^{(k+1)}$$

whose solution r is easily computable. We can further define a matrix

$$\tilde{T} = [R \quad r] \Gamma R^{-1}$$

where Γ is an $(m+1)\times m$ bidiagonal matrix with non-zero entries

$$\begin{array}{rcl}\Gamma_{j,j} & = & (\eta_{k-m+j}\alpha_{k-m+j})^{-1} \\ \Gamma_{j+1,j} & = & -(\eta_{k-m+j}\alpha_{k-m+j})^{-1},\ j=1,\dots,m.\end{array} \tag{3.10}$$

Now, from the upper Hessenberg matrix $\tilde{T}$ we define an $m\times m$ symmetric tridiagonal matrix T, by replacing the strictly upper triangle of $\tilde{T}$ with the transpose of its strictly lower triangle. At the end, according to [85], we compute the m eigenvalues of T, i. e. t_j, $j=1,\dots,m$, which are called *Ritz-like values*, and, in the current $k-$th iteration we define the step-lengths for the next m iterations as

$$\alpha_{k+j}=\frac{1}{t_j},\ j=1,\dots,m. \tag{3.11}$$

Since m should be a small number (tipically $m=3$), all these computations handle with small matrices, thus this step-lenght rule is not too much computationally expansive, even if it requires to store in memory m past gradient vectors.

3.3 Lagged Diffusivity Fixed Point method

The FP method has been proposed by Vogel since 1996 [101, 99, 37] for image denoising, to solve a Total Variation-penalized Least Squares minimization problem, with a fixed point method involving the Hessian of the objective function. He further developed the algorithm into a limited memory BFGS scheme for 2D large scale problems [47] and into a fast quasi-Newton solver for deblurring 2D imaging [102]. Starting from this last approach and its implementation proposed in [100], we derived the 3D lagged diffusivity FP algorithm for 3D imaging, to solve the unconstrained problem (2.26) where we fix the fitting term with the LS function (2.15).

If we aim to compute

$$\min_f \frac{1}{2}\|Mf-g\|_2^2+\lambda\int_V\sqrt{|\nabla f|^2+\beta^2} \tag{3.12}$$

where V is the discretized volume, the associated Euler-Lagrange equations are

$$\begin{array}{r}G(f)=M^T(Mf-g)+\lambda L(f)f=0,\quad \text{in } V \\ \dfrac{\partial f}{\partial n}=0,\quad \text{on } \partial V\end{array} \tag{3.13}$$

where L(f) is the differential operator whose action on a function w is

$$L(f)w=-\nabla\cdot\left(\frac{1}{\sqrt{|\nabla f|^2+\beta^2}}\nabla w\right). \tag{3.14}$$

It represents a diffusion operator. We highlight that, in (3.13), $M^T M$ and L(f) are symmetric and positive semidefinite operators, in a nonlinear integro-differential equation of elliptic type. Furthermore, equations in (3.13) may also be expressed [102] as a nonlinear first order system

$$\begin{aligned} M^T M f - \lambda \nabla \cdot \vec{v} = M^T g \\ -\nabla f + \sqrt{|\nabla f|^2 + \beta^2}\; \vec{v} = \vec{0} \end{aligned} \tag{3.15}$$

where $\vec{v}$ is the dual variable

$$\vec{v} = \frac{\nabla f}{\sqrt{|\nabla f|^2 + \beta^2}}. \tag{3.16}$$

If we linearize by fixing $f = f^{(k)}$ in the square root and substitute (3.14) into (3.15), we get

$$[M^T M + \lambda L(f^{(k)})] f^{(k+1)} = M^T g \quad \forall k = 0, 1, \dots \tag{3.17}$$

that is equivalent to the following *Fixed Point iteration* rule:

$$f^{(k+1)} = [M^T M + \lambda L(f^{(k)})]^{-1} M^T g \quad \forall k = 0, 1, \dots. \tag{3.18}$$

This scheme can be seen as a quasi-Newton solver. In fact, the Hessian of our objective function is

$$Hess(\mathcal{J}(f^{(k)})) = M^T M + \lambda L(f^{(k)}) + \lambda L'(f^{(k)}) f^{(k)} \tag{3.19}$$

while the descent direction of this FP scheme is

$$\begin{aligned} s_k &= f^{(k+1)} - f^{(k)} = \\ &= -[M^T M + \lambda L(f^{(k)})]^{-1} \cdot ([M^T M + \lambda L(f^{(k)})] f^{(k)} - M^T g) = \\ &= -[M^T M + \lambda L(f^{(k)})]^{-1} \cdot (\nabla LS(f^{(k)}) + \lambda \nabla TV_\beta(f^{(k)})). \end{aligned} \tag{3.20}$$

Therefore, by omitting the $\lambda L'(f^{(k)}) f^{(k)}$ quantity, we can define the matrix $H_k = M^T M + \lambda L(f^{(k)})$ as an approximation of $Hess(\mathcal{J}(f^{(k)}))$ and equation (3.18) can be written in a quasi-Newton form

$$f^{(k+1)} = f^{(k)} - H_k^{-1} G_k \tag{3.21}$$

where $G_k = G(f^{(k)})$ is exactly the gradient of the objective function $\nabla \mathcal{J}(f^{(k)})$.

To conclude, we remark that $\mathcal{J}(f^{(k)})$ is strictly convex in our prefixed model. In [37], the authors proved the global convergence of this algorithm with a linear convergence rate, under the assumption of strictly convexity for the objective function. In addition, increasing the smoothness parameter β accelerates the convergence as well as with relatively small values of λ.

```
Initialize: $f^{(0)} \geq 0$;
for $k = 0, 1, \dots$
    $G_k = \nabla \mathcal{J}(f^{(k)})$;
    $H_k = M^T M + \lambda L(f^{(k)})$;                (hessian approximation)
    solve the linear system $H_k s_k = -G_k$           (inner solver calling)
    $f^{(k+1)} = f^{(k)} + s_k$;
end
```

Table 3.2: Algorithm FP (Lagged Diffusivity Fixed Point).

3.3.1 Implementation notes

A scheme of the FP algorithm is shown in table 3.2, to focus on some implementation remarks.

The most expensive step is the computation of the descent direction s_k, since it requires the resolution of an inner linear system $H_k s_k = -G_k$ in each k-th iteration: because of its large size, we propose to solve it with a CG algorithm. An internal loop must be handled and every inner iteration requires two matrix-vector products which make remarkably increase the computational cost of the FP solver. On the other hand, we consider CG approximated solutions by stopping it before convergence: asking for less data-fitting is a way of getting more regularized final solutions. Thus, the CG iterations end when the residual 2-norm is smaller than a large tolerance (typically of 10^{-3}) or a maximum number of iterations (of order of tens) is reached.

Once we get s_k, the FP updating rule comes from equation 3.21: we observe that it is executed without line-search, because the convergence of the scheme is demonstrated to be global.

Concerning the Hessian approximation into the H_k matrix, we highlight that it theoretically involves one matrix-matrix product $M^T M$. Of course it is never computed, because its action can be evaluated runtime by the inner solver, if we make use of the CG algorithm.

We now recall the notation introduced in paragraph 2.3.3: the $TV_\beta(f)$ functional of equation (2.20) can be computed as

$$TV_\beta(f) = \frac{1}{2} \sum_{j_x=1}^{N_x} \sum_{j_y=1}^{N_y} \sum_{j_z=1}^{N_z} \phi((D^x_{j_x,j_y,j_z} f)^2 + (D^y_{j_x,j_y,j_z} f)^2 + (D^z_{j_x,j_y,j_z} f)^2) \tag{3.22}$$

where $\phi(t) = 2\sqrt{t + \beta^2}$ and

$$D^x_{j_x,j_y,j_z} f = \frac{f_{j_x+1,j_y,j_z} - f_{j_x,j_y,j_z}}{\Delta_x}, \quad D^y_{j_x,j_y,j_z} f = \frac{f_{j_x,j_y+1,j_z} - f_{j_x,j_y,j_z}}{\Delta_y},$$

$$D^z_{j_x,j_y,j_z} f = \frac{f_{j_x,j_y,j_z+1} - f_{j_x,j_y,j_z}}{\Delta_z}$$

define the forward differences used for the TV evaluation. Starting from the 2D implementation proposed in [100], we have developed the matrix L(f) of the 3-dimensional diffusion operator (3.14) as

$$L(f) = \begin{bmatrix} (D^x)^T & (D^y)^T & (D^z)^T \end{bmatrix} \begin{bmatrix} diag(\phi') & & \\ & diag(\phi') & \\ & & diag(\phi') \end{bmatrix} \begin{bmatrix} D^x \\ D^y \\ D^z \end{bmatrix} \tag{3.23}$$

where $\phi'_{j_x,j_y,j_z} = \phi'((D^x_{j_x,j_y,j_z} f)^2 + (D^y_{j_x,j_y,j_z} f)^2 + (D^z_{j_x,j_y,j_z} f)^2)$.
The resulting $L(f^{(k)})$ is a 7-diagonal matrix, because it derives from the discretization of the 3D forward differences of the TV operator that involves 7 voxels for every point (j_x, j_y, j_z). Figure 3.2 shows the sparsity pattern of $L(f^{(k)})$, for a very small problem: for every row $j \approx (j_x, j_y, j_z)$, the three elements in the upper diagonals reflect the forward voxels along the three directions x, y and z, while the elements in the lower diagonals are related to those points for which the current j-th voxel is the forward point.
As a consequence, it is always enough to store seven vectors of maximum length N_v, instead of a full $N_v \times N_v$ matrix H_k.

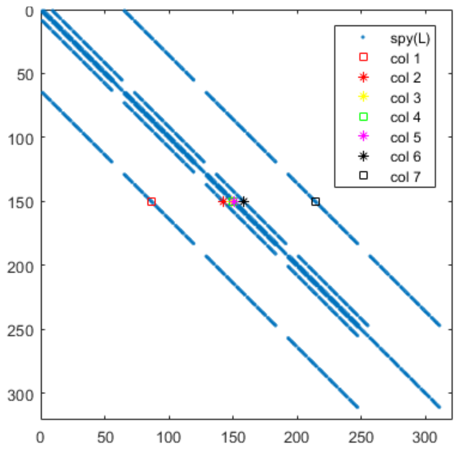


Figure 3.2: Sparsity pattern of the matrix $L(f)$ for a volume f discretized into 8x8x5 voxels: the diagonals are not completely full because of hte boundary conditions.

The FP algorithm is stopped when

$$\|f^{(k+1)} - f^{(k)}\| < \tau \tag{3.24}$$

or a maximum number of iterations is reached

$$k = \texttt{MAXIT}. \tag{3.25}$$

We avoid stopping rules involving the objective function, because its evaluation is not required by the FP scheme and it would increase the computational cost of the algorithm.
We further observe that in every iteration, $\nabla\mathcal{J}(f^{(k)})$ is computed as

$$\nabla\mathcal{J}(f^{(k)}) = M^T(Mf^{(k)} - g) + \lambda L(f^{(k)})f^{(k)},$$

hence it requires two matrix-vector products. We underline that, in the following chapters, the number of iterations performed by the FP algorithm to provide a certain solution in considered as the sum of the FP standard iterations and the total number of CG executions: we believe it could represent a more practical approach to the tomographic imaging problem.

Chapter 4

Tests on medium size problems in Matlab

In this section, we report some tests performed to try out the Scaled Gradient Projection and the Fixed Point methods, on different phantoms and according to very different sparse tomographic settings: the aim is to preliminarly show the efficiency of the proposed algorithms for IIR from sparse subsampling.
These tests have been executed in Matlab environment thanks to the medium sizes of the problems. We start focusing, in section 4.1, on the performances of the scaling strategy and the two step-length rules (shown in the previous chapter) for the SGP algorithm on a 3D version of the well-known Shepp Logan phantom: the proposed strategies seem to indeed accelerate the convergence to the problem solution, in the first iterations. In section 4.2 the comparison of SGP and FP methods is performed on 2D imaging reconstruction problelm, where the dataset comes from a real CT acquisition: here we investigate the robustness of the algorithms with regard to the regularization parameter and to an increasing sparsity level. At the end, in section 4.3, the comparison is made on a Digital Breast Tomosynthesis sham problem, to introduce the results of the following chapter which are performed on real DBT data.

4.1 Focus on the SGP algorithm

This section is centered on the evaluation of the SGP accelerating strategies, proposed in chatper 3.2. To do that, we made use of some functions of the `TVReg` Matlab Toolbox, available at [9]. It provides both a ray-driven forward projector and a useful implementation of the state-of-the-art algorithm UNP [57, 60] for a comparison.
These numerical results have been performed on a MacBook Pro, 3GHz Intel Core i7, 8 Gb of RAM, equipped with Matlab R2015a [34, 33].

4.1.1 Test problem and setting

We start considering as the true object f^* the digital Shepp Logan phantom, having values in [0, 1] and discretized into $N_v = N_x \times N_y \times N_z = 61 \times 61 \times 61 = 226981$ voxels lexicographically ordered in a vector: in figure 4.1 the slices number 24, 31 and 35 in the z direction are shown. To simulate

(a) layer 24 (b) layer 31

(c) layer 35 (d) layer 31

Figure 4.1: Different layers in the z-direction of the original phantom. For the analysis of the results, in (d) some interesting features are highlighted on layer 31: a yellow small crop where the Standard Deviation is computed, an orange line along which we analyse the vertical profile and a red pixel to examine the depth profile.

the tomographic process, the projection data have been created as:

$$g^* = M \cdot f^*$$

where M is the projection matrix, obtained with the functions in the `TVReg` Toolbox for a 3D geometry with random angles over the half of a sphere. The detector is supposed with $N_p = 61 \times 61$ pixels and the number of angles N_θ varies in the set $\{19, 37, 55\}$: in all the cases the problem is underdetermined and it allows us to test the SGP algorithm on different sparsity levels. The projections are corrupted by noise, with both Gaussian and Poisson distribution, as specified in the following subsections.

We describe here the stopping criterion for the SGP algorithm and the setting of its main parameters. If k is the index of the current iteration and $\mathfrak{d}_k^{\mathcal{J}} := \frac{|\mathcal{J}(f^{(k+1)}) - \mathcal{J}(f^{(k)})|}{|\mathcal{J}(f^{(k)})|}$ is the relative distance between successive values of the objective function, we consider the conditions

$$\mathfrak{d}_k^{\mathcal{J}} \leq \epsilon_1, \qquad \frac{1}{p}\sum_{j=0}^{p-1} \mathfrak{d}_{k-j}^{\mathcal{J}} \leq \epsilon_2 \ \text{ if } \ k \geq p-1,$$

where $\epsilon_1 = 10^{-6}$, $p = 20$ and $\epsilon_2 = 10^{-5}$; the SGP stopping criterion consists in satisfying both the conditions or performing a maximum number of $k = 1000$ iterations.

For what concerns the SGP parameters, the setting reported below is used:

- $\gamma = 0.4$ and $\sigma = 10^{-4}$ as backtracking parameters;
- $\alpha_{min} = 10^{-10}$, $\alpha_{max} = 10^5$, $\alpha_0 = 1$, $m_\alpha = 2$ and $\tau_0 = 0.5$ for the step-length selection.

In order to evaluate the reconstruction results, we consider the following parameters: the Relative Error (*RelErr*) between the exact volume f^* and the reconstructed image $\tilde{f}$, that is

$$RelErr = \|f^* - \tilde{f}\|_2 / \|f^*\|_2, \tag{4.1}$$

and the Standard Deviation (*StdDev*) of the image values inside a small crop (represented by the yellow 8×8 pixel square in Figure 4.1(d)) on a uniform region in the central layer. The reconstructed images are also evaluated by plotting the profile of the yellow vertical line in Figure 4.1(d), vertical profile (VP), and the profile over the 61 layers in the z direction of the red pixel in Figure 4.1(d), depth profile (DP).

We show the results obtained by the algorithms at three different temporal windows: at 5 seconds (10-15 iterations), for simulating a real-time execution; at 20 seconds (40-60 iterations), corresponding to an over-time of few minutes in real applications; at the convergence, i.e., when the convergence criterium is satisfied (this is a long execution that can be performed only off-line in a real application). We think that each of these three different outputs may have a practical interest and together they represent the evolution of the methods in time.

4.1.2 Case of Gaussian noise. Results with the LS data function.

In this paragraph we show the results obtained on the simulated data $g = g^* + e$, where e is the vector representing white Gaussian noise with level defined as $\nu = \frac{\|e\|_2}{\|b^*\|_2}$; we consider here $\nu = 0.01$, corresponding to a Signal-to-Noise Ratio SNR $:= 20 \cdot log_{10}(\frac{\|g\|_2}{\|g-g^*\|_2})$ of about 40. We consider the fit-to-data function $J(x)$ as the LS function and we set the TV smoothing parameter β equal to 0.001 in all the experiments. The regularization parameter λ has been heuristically set to 0.09; we have experimented that for this test the model is not very sensitive to the value of λ (similar results have been obtained with different values of λ in the interval $[0.005, 0.5]$).

We compare the results of the SGP implementation with the BB step-length rule, with the non-scaled Gradient Projection method (GP) (with the step-length selection used by SGP, but with $S_{k+1} = I$ in the definitions (3.6) and (3.7) of the BB rules) and with the UPN method proposed in [57], implemented in the `TVReg` toolbox. The UPN method has been equipped with the same stopping criterion and regularization parameter used for SGP, while its own parameters have been set at their best values after a careful tuning.

In Table 4.1 we present the results obtained with different number of views ($N_\theta = 19, 37, 55$) for the GP, SGP and UPN methods. In the columns from left to right we report the Relative Error, the Standard Deviation, the objective function value and the number of performed iterations in the three considered temporal windows: at 5 seconds, at 20 seconds and at convergence. From the table, we see that the SGP method outperforms the others in the first iterations (5 and 20 seconds) for all the considered angles, while they all give very similar results at convergence, mirroring their convergence to the unique solution of the constrained problem (2.28). Comparing, in fact, the decreases of the objective functions for the GP and SGP cases related to the earliest reconstructions, we can appreciate the acceleration of the latter method, even if every scaled iteration takes a little longer time. Focusing on the StdDev values, we can assign good denoising properties to all the tested strategies. In particular, the reconstructions of central layer (layer 31) obtained with the three considered methods in the case $N_\theta = 37$ are shown in figure 4.2: at first glance, it seems that the 5

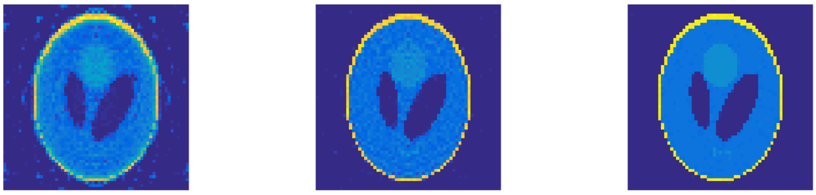
			$RelErr$	$StdDev$	fun	iters
$N_\theta = 19$	GP	5 secs	0.3816	0.0559	5574.36	16
		20 secs	0.1548	0.0312	1596.75	69
		conv	0.0559	0.0052	1498.49	263
	SGP	5 secs	0.2637	0.0463	3396.96	19
		20 secs	0.1178	0.0241	1560.48	71
		conv	0.0543	0.0052	1498.57	198
	UPN	5 secs	0.3785	0.0640	6075.55	11
		20 secs	0.1786	0.0414	1652.62	48
		conv	0.0580	0.0056	1484.46	606
$N_\theta = 37$	GP	5 secs	0.3475	0.0331	11537.90	16
		20 secs	0.0898	0.0241	1795.02	64
		conv	0.0245	0.0028	1645.77	154
	SGP	5 secs	0.1840	0.0261	4335.56	18
		20 secs	0.0477	0.0777	1689.30	66
		conv	0.0247	0.0030	1646.39	194
	UPN	5 secs	0.4001	0.0281	19918.90	8
		20 secs	0.1045	0.0343	1917.49	46
		conv	0.0241	0.0028	1632.07	224
$N_\theta = 55$	GP	5 secs	0.3091	0.0438	14306.80	15
		20 secs	0.0779	0.0276	1997.80	60
		conv	0.0199	0.0044	1783.11	142
	SGP	5 secs	0.2148	0.0306	9662.70	16
		20 secs	0.0277	0.0077	1814.38	60
		conv	0.0199	0.0044	1783.60	147
	UPN	5 secs	0.4315	0.0325	40865.00	6
		20 secs	0.0677	0.0293	2033.66	46
		conv	0.0199	0.0043	1769.47	200

Table 4.1: Results obtained on the test problems with data affected by Gaussian noise.

second outputs only detect the main structures, the 20 second ones make the small objects more visible among the brain background disuniformity and the final reconstructions validate our model with high quality images.

(a) GP method at 5 seconds, 20 seconds, convergence.

(b) SGP method at 5 seconds, 20 seconds, convergence.



(c) UPN method at 5 seconds, 20 seconds, convergence.

Figure 4.2: Reconstructions obtained in case of Gaussian noise on the data. From the left to the right: reconstructions after 5 seconds, after 20 seconds, at convergence.

In Figure 4.3 the errors versus the iterations (on the left) and the objective function values versus the iterations (on the right) are shown in log-log scale. We compare here the GP method (blue line), the SGP method (red line) and the UPN method (green line) up to the convergence of the methods.

The advantage of using the scaling matrix is evident, especially in the first iterations, and is confirmed by analysing some important profiles on the reconstructions. Figure 4.4 displays the VP (on the left) and DP (on the right) after 5 seconds, 20 seconds and at convergence: we compare again the GP reconstruction (blue line), SGP reconstruction (red line) and UPN reconstruction (green line) with the phantom profile (grey line).

The VP plots confirm that after few iterations (5 seconds) we can identify, in the signal reconstructed by the SGP method, all the objects with a good approximation of their intensity; in the DP plot after 20 seconds the SGP method has almost completely eliminated the noise, while the GP and UPN plots show a residual noise yet. In figure 4.5 we show a zoom of figure 4.4 for the pixels between 4 and 27 for the VP plots on the left and between 19 and 43 for the DP plots on the right, in order to clearer represent the behaviour of the method for reconstructing the small object represented by the red dot in Figure 4.1(d). We can see that the SGP profiles are less noisy than the others and in the DP the peak of the SGP line is the closest the exact one.

4.1.3 Case of Poisson noise. Results with the KL data function.

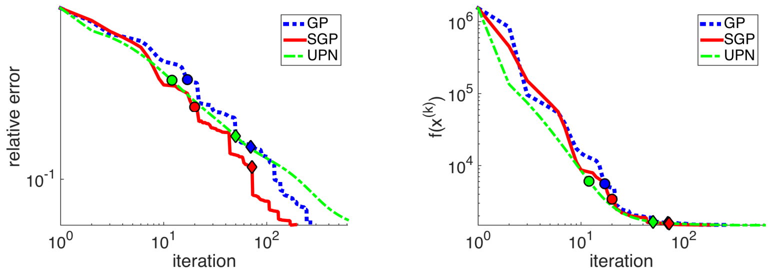
We consider now some tests where the projections are affected by Poisson noise, with SNR $\simeq 40$ and background $bg = 10^{-5}$. The problem is solved by using the KL fit-to-data function in (2.16). In this case the regularization parameter λ has been heuristically set to 0.03 and the TV smoothing parameter $\beta = 0.001$; we have experimented that, as in the case of Gaussian noise, similar results have been obtained with different values of λ in the interval $[0.001, 0.1]$.

For this model, we compare the results obtained with the GP and the SGP methods, since the UPN method is provided only for the case of LS fit-to-data function. Table 4.2 reports the results in the case $N_\theta = 19, 37, 55$, with the same information of table 4.1. For the KL model, the performance improvement due to the scaling is more consistent than in the LS model, as it can be seen by the RelErr and StdDev values. If the number of performed iterations in the last column is equal to 1000 an asterisk reminds that the algorithm has stopped after reaching the maximum number of iterations. We want to stress that this happens only for the GP method, confirming its slower convergence rate. In figure 4.6 we plot the Relative Error versus the iterations in the left panel, while the objective function values versus the iterations are displayed in the right panel. In figure 4.7 the reconstructions of the layer 31 obtained with both GP and SGP methods after 5 seconds, 20 seconds and at convergence are represented. Both the plots and the images confirm the convergence acceleration of the scaled algorithm with respect to the non-scaled one.

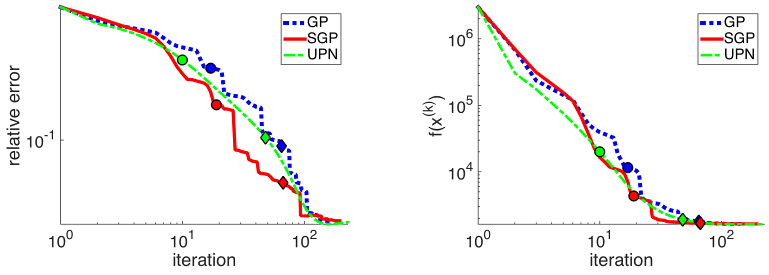
The analysis of VP and DP profiles for $N_\theta = 37$ in figure 4.8 shows that the scaling allows recovering very good profiles in very short time: after 20 seconds the line of the reconstructed image almost overlap the line of the exact phantom and the only exception is the small peak in the center of the DP, whose detection is very hard (due to the sparsity of the scanning views).

			RelErr	*StdDev*	fun	iters
$N_\theta = 19$	GP	5 secs	0.6305	0.0154	16781.30	12
		20 secs	0.5685	0.0204	9031.95	61
		conv	0.4353	0.0290	3085.58	877
	SGP	5 secs	0.2145	0.0332	767.85	19
		20 secs	0.0984	0.0097	524.80	72
		conv	0.0869	0.0063	522.01	172
$N_\theta = 37$	GP	5 secs	0.6914	0.0085	55813.10	11
		20 secs	0.6349	0.0107	31273.40	53
		conv	0.4045	0.0180	4913.71	1000*
	SGP	5 secs	0.1752	0.0201	1127.98	18
		20 secs	0.0798	0.0129	597.74	66
		conv	0.0335	0.0031	545.41	393
$N_\theta = 55$	GP	5 secs	0.6919	0.0105	85188.70	12
		20 secs	0.6495	0.0121	54467.70	52
		conv	0.4191	0.0222	8215.40	1000*
	SGP	5 secs	0.1745	0.0226	1552.37	16
		20 secs	0.0496	0.0107	596.71	60
		conv	0.0387	0.0068	580.17	198

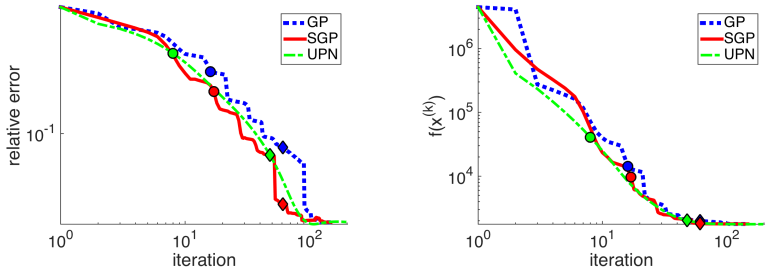
Table 4.2: Results obtained on the test problems with data affected by Poisson noise.



(a) $N_\theta = 19$

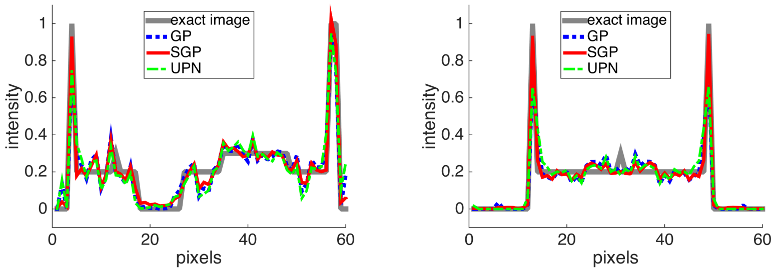


(b) $N_\theta = 37$

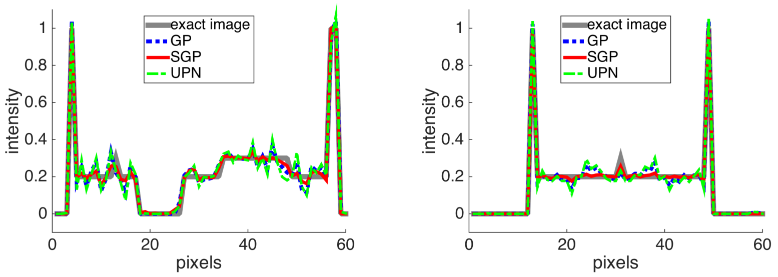


(c) $N_\theta = 55$

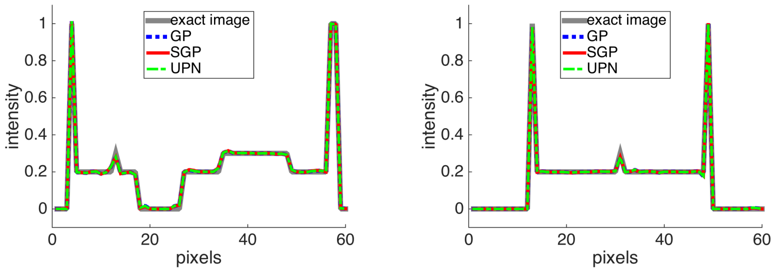
Figure 4.3: Case of Gaussian noise. On the left: errors vs iterations; on the right: function values vs iterations. The circles and the diamonds represent the values at 5 and 20 seconds, respectively.



(a) Profiles after 5 seconds

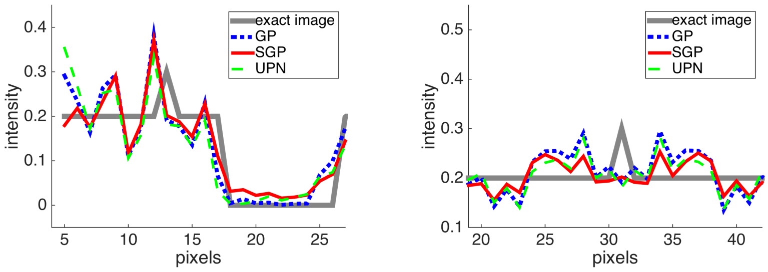


(b) Profiles after 20 seconds

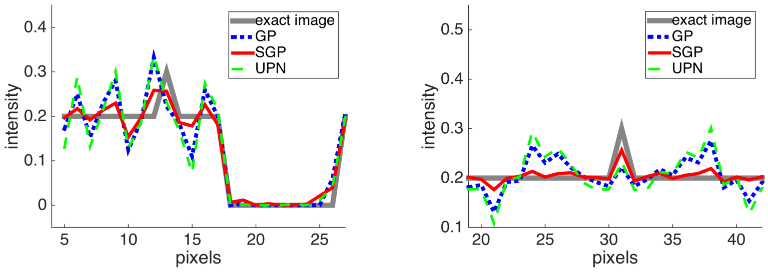


(c) Profiles at convergence

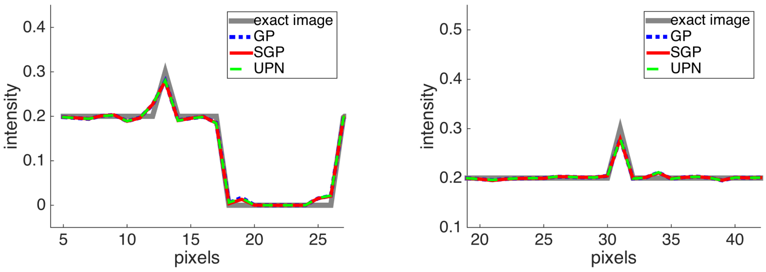
Figure 4.4: Case of Gaussian noise. Profiles for 37 angles: on the left VP plots and on the rigth DP plots at different temporal windows.



(a) Profiles after 5 seconds

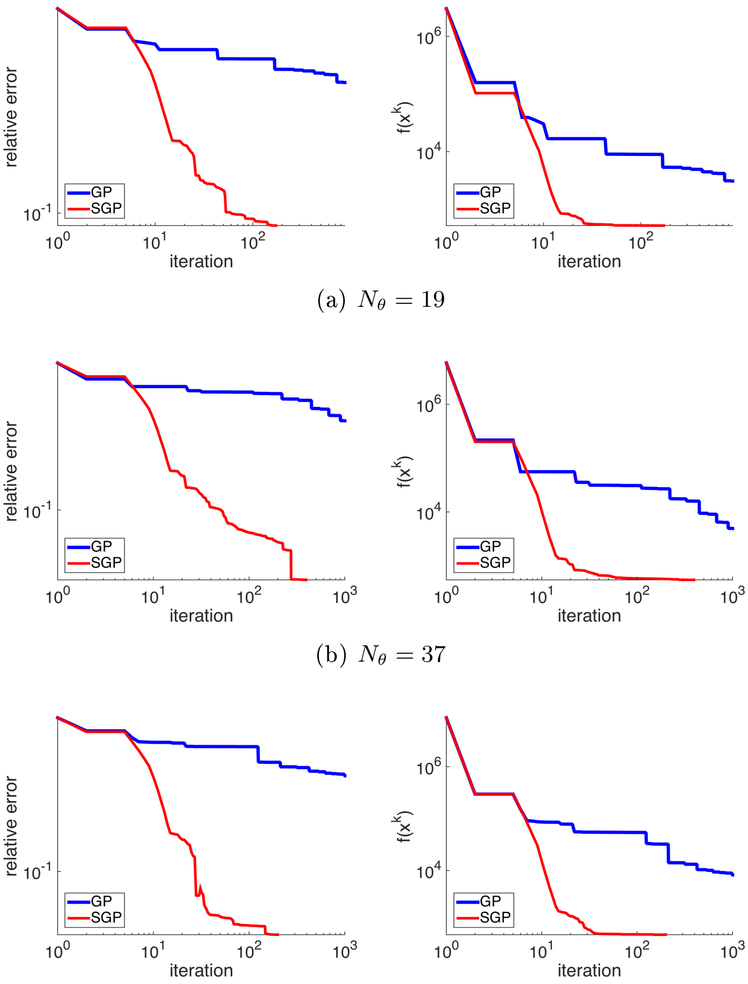


(b) Profiles after 20 seconds



(c) Profiles at convergence

Figure 4.5: Zooms of the plots in figure 4.4.

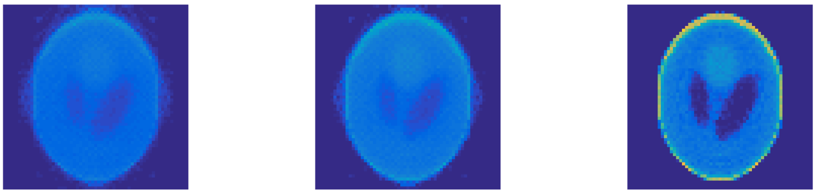


(a) $N_\theta = 19$

(b) $N_\theta = 37$

(c) $N_\theta = 55$

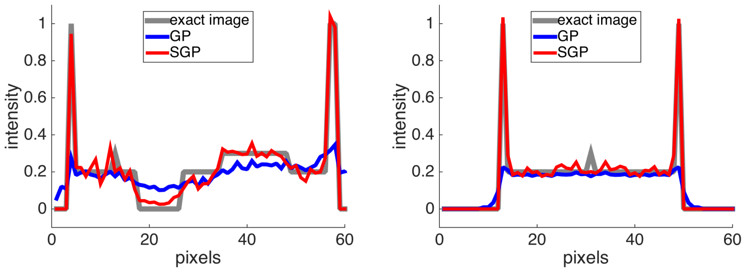
Figure 4.6: Case of Poisson noise. On the left: errors vs iterations; on the right: function values vs iterations.



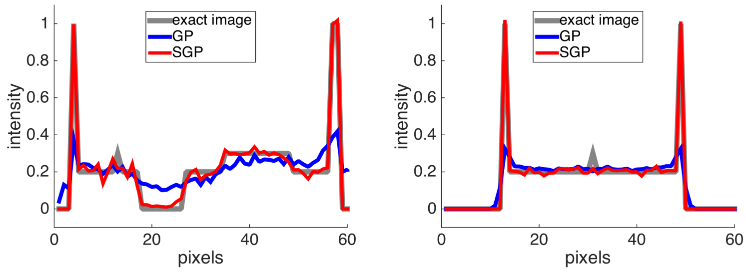
(a) GP method at 5 seconds, 20 seconds, convergence.

(b) SGP method at 5 seconds, 20 seconds, convergence.

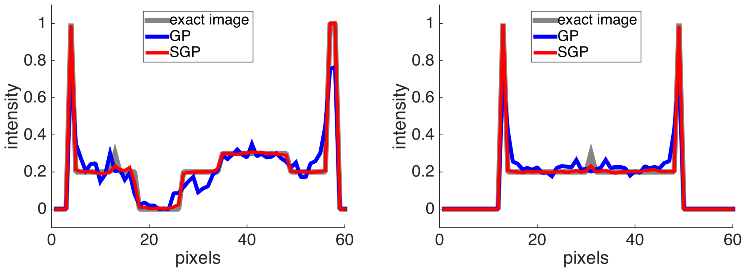
Figure 4.7: Reconstructions obtained in case of Poisson noise on the data. From the left to the right: reconstructions after 5 seconds, after 20 seconds, at convergence.



(a) Profiles after 5 seconds



(b) Profiles after 20 seconds



(c) Profiles at convergence

Figure 4.8: Case of Poisson noise. Profiles for 37 angles: on the left VP plots and on the rigth DP plots at different temporal windows.

4.1.4 Algorithms reliability with regard to noise

To test the reliability of the scaled algorithm on harder problems, in this paragraph we take into acconut higher levels of noise on the projections, which can reflect a CT process with a lower X-ray dose. In particular, we have performed some tests by adding noise of different intensity on the data and solving the reconstruction problem with a suitable regularization parameter in each case: we show in figure 4.9 the plots of the quantitative indeces to analyse the results obtained after 20 seconds. In particular, we plot the relative errors as a function of the SNR in the case of Gaussian noise (left panel) and Poisson noise (right panel): the relative error obviously increases with increasing noise but the SGP method always performs better for all noise levels. We highlight the remarkable difference between the errors obtained with SGP and GP methods in the case of Poisson noise, where the proposed scaling is indeed more effective. On the other hand, we can observe that even for SNR ≈ 30 the scaled implementation provides a good reconstruction in only 20 iterations, with a relative error smaller than 0.1 in the case of Gaussian noise, too.

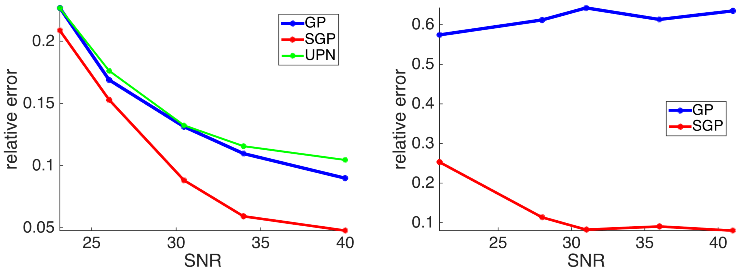


Figure 4.9: Relative errors vs SNR after 20 seconds for all the methods. On the left panel, the results with Gaussian noise on the data; on the right panel, the results with Poisson noise on the data.

4.1.5 Comparison of step-length rules

To conclude the analysis of the SGP algorithm, we now focus on the two step-length selection strategies prososed in paragraph 3.2.2: the alternating Barzilai-Borwein (BB) rule and the one based on the Ritz-like values (R). The exact projections g^* are again corrupted by random noise from Poisson distribution with level 10^9 and bg $= 10^{-5}$ and the regularization parameter λ is heuristically fixed as 0.03, while the smoothing parameter is $\beta = 0.01$. Now we show the results obtained with $N_\theta = 37$ angles, but the algorithms showed a very similar behavior in the case of 19 and 55 angles. The plots in figure 4.10 compare the errors (on the left panel) and the values of the objective function (on the right panel) versus the iterations, for the different SGP implementations: the red and blue lines refer to the scaled methods with the step-length based on the BB rules (SGP_BB) and the Ritz-like values (SGP_R), respectively, while the black and green lines denote the non-scaled methods (GP_BB and GP_R, respectively), that are again the SGP versions with scaling S_k equal to the identity matrix at each iteration. Independently of the step-length rule, the scaling strategy accelerates the GP methods considerably, especially in the first iterations, and it represents a great guarantee for the proper use of an SGP algorithm in SpCT. In figure 4.11 we show the reconstruction of the central layer obtained with the scaled methods, both after 20 and 1000 iterations, and the quality of the SGP_R reconstruction after 20 iterations is indeed noticeable: the lighter object has a sharper edge and the whole brain looks neater. In order to better compare the SGP_BB and SGP_R results after 20 iterations, we analyse again the Vertical and Depth profiles, respectively on the left and on the right hand pictures of figure 4.12. We observe that the noise is suppressed almost everywhere and the small objects represented by narrow peaks are detected, even if the intensity of the signal is somewhere lower than the exact one. Furthermore, it is evident that the SGP_R gets a better reconstruction in few iterations and this is crucial for medical softwares that must provide reliable images almost in real time. To conclude, in table 4.3 we report both the value of the objective function and the relative error for the results obtained by the SGP_BB and SGP_R algorithms at 20 and 1000 iterations, with a different number of source angles: The table confirms that the choice of the step-length based on the Ritz-like values improves the algorithm speed in all the considered underdetermined cases, providing useful reduction of the reconstruction error and a remarkable approach toward the minimum value of the minimizing function.

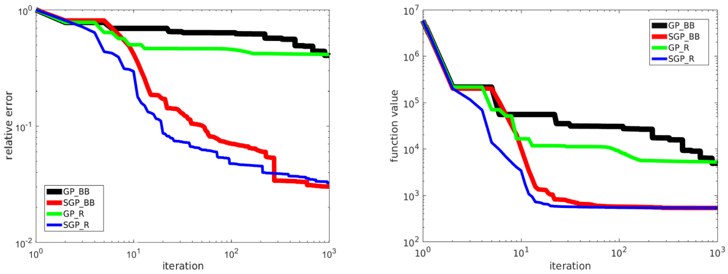
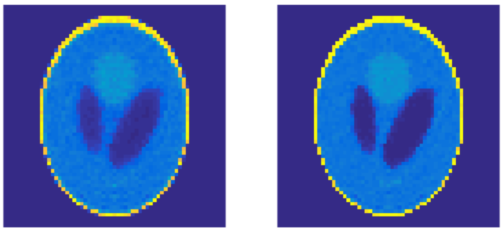
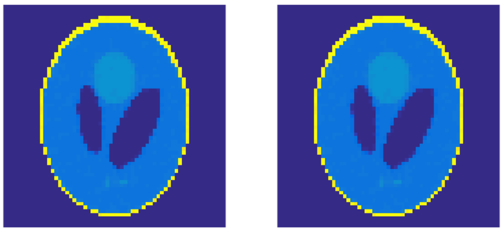


Figure 4.10: Relative errors vs iterations on the left; function values vs iterations on the right. GP_BB is in black, SGP_BB in red, GP_R in green and SGP_R in blue.



(a) Profiles after 20 iterations



(b) Profiles after 1000 iterations

Figure 4.11: The central layer of the scaled reconstructions. The SGP_BB outputs in the left column and SGP_R outputs in the right column.

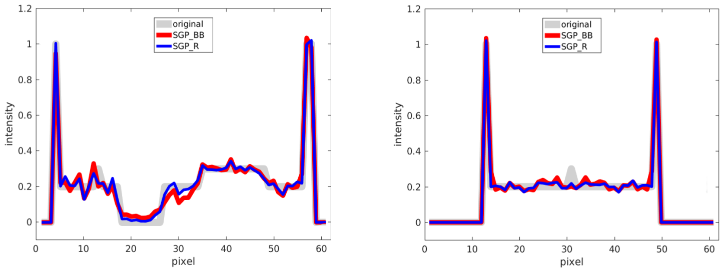


Figure 4.12: Vertical profile (on the left) and depth profile (on the right) over the objects of interest. The gray line is the exact profile, the red line refers to SGP_BB, the blue line denotes SGP_R stopped after 20 iterations.

		$N_\theta = 19$		$N_\theta = 37$		$N_\theta = 55$	
		fun	$RelErr$	fun	$RelErr$	fun	$RelErr$
k = 20	SGP_BB	765.109	0.2140	1055.09	0.1705	1310.77	0.1609
	SGP_R	553.704	0.1522	590.133	0.0856	661.329	0.0894
k = 1000	SGP_BB	520.047	0.0722	543.907	0.0301	569.152	0.0251
	SGP_R	520.813	0.0798	544.602	0.0324	571.04	0.0290

Table 4.3: Behavior of SGP_BB and SGP_R for different numbers of N_θ, at 20 and 1000 iterations.

4.2 Comparisons on a real 2D SpCT dataset

In this section, a first comparison between the SGP and the FP algorithms is made on a 2D example of SpCT, starting from a real dataset. Such dataset is freely downloadable from the web page of the Finnish Inverse Problem Society `www.fips.fi/dataset.php` and the relative documentation can be found in [7]. The object in exam is a lotus root which has been filled with attenuating objects and scanned from 120 positions on a round angle with a fan beam. The reconstruction of one bi-dimensional slice of the lotus root is here achieved both from 120 and 20 projection data, to test the reliability of the SGP and FP methods on both a medium-level SpCT case and on a harder imaging problem.

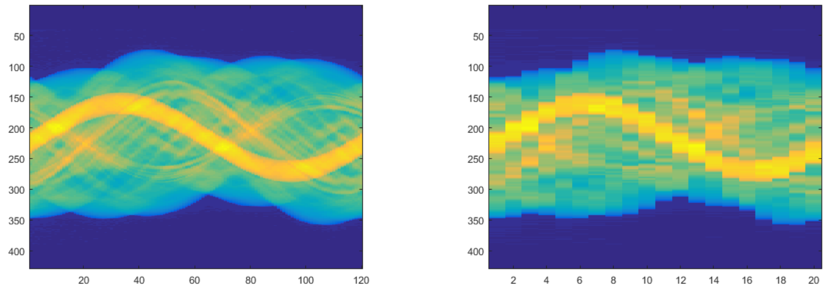
All these tests are implemented in Matlab R2016a and performed on a computer equipped with two processors Intel (R) Xeon(R) CPU E5-2650, 2.30 GHz, 20 cores, with 132 GB RAM.

4.2.1 Test problem and setting

The provided Matlab files allow to test any IIR method to real projection data in 2D. The scanned object is a slice of lotus root, whose holes have been filled with elements made of different chemical elements: if we refer to figure 4.13 and move clockwise on the holes, we find one hole in the upper-left corner filled with some match-heads (made of sulphur), then one filled with a portion of a pencil (made of carbon sourrounded by wood), another one with three rectangular pieces of ceramics (made of calcium) while the last hole contains a thick circular chunk of chalk (made of calcium again). The scanning process consists in 120 projections performed from a circular trajectory, with angular step size $\Delta_\theta = 3$ degrees; each real projection array has been downsampled into 429 recorded values, hence the sinogram is a

Figure 4.13: A picture of the lotus root, filled with different materials.

matrix of size 429 × 120 and it is shown in figure 4.14(a). The dataset also provides the forward projector, as a sparse matrix of size 51480 × 65536, hence the reconstruction will be an image of 256×256 pixels. Because of the variety of the components inside the lotus root, of their shapes and sizes, and because of the tomographic geometry, this dataset can be a good template of a SpCT reconstraction for industrial applications.



(a) (b)

Figure 4.14: The full sinogram of the lotus dataset with 120 projections on the left and the reduced sinogram with only 20 views on the right.

To simulate a more severe sparse tomography, we also considered a reduced dataset made of only 20 projections on the round angle: the resulting CT is thus performed with a wide angular step of 18 degrees and its sinogram is reported in figure 4.14(b).

To solve the two SpCT problems, we set the Least Squares data-fitting function and applied both the SGP method with the BB step-length rule and the FP algorithm. As this 2D case does not lead to a very big dimensional problem, we prefere to stop the executions after having computed a prefixed number of iterations in both the algorithms, but we must remark that in the FP case all the CG iterations compute a couple of matrix-vector product, hence they all must be counted together with every external iteration over k. In details, the inner CG solver always executes 10 iterations: it represents a good compromise between the required accuracy on the nested linear system resolution and the practical need of fast reconstructions. If not specified, the SGP parameters are the same we set in the previous test on the Shepp Logan phantom. The value of β_{TV} is 10^{-3}.

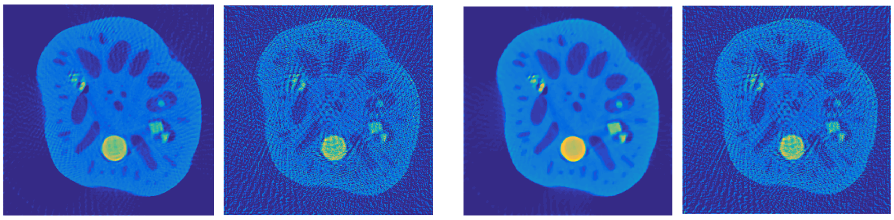
Since we do not have an exact image to compare with, we cannot compute an image error. Neither, we can analyze the algorithms comparing the descent of their objective functions: even if they could make use of the same regularization parameter, the SGP solver includes a projection onto the non-negativity orthant in every iteration that may not be added in the FP formulation. As a consequence, a parallel analysis of the algorithm per-

formances must be done just looking at the outputs.
On the other hand, a post-processing rehash is tipically applied to the output, above all in practical applications, to make the visualization clearer: for example, it is common to cut all the negative values setting them to zeros, in order to keep the physical meaning of the non-negative absorbtion coefficient for every pixel (or voxel). Of course this procedure is not necessary for the SGP reconstructions, but it is for the FP images. Here, only few elements show negative values and they are supposedly due to noisy components: all the figures related to the FP algorithm have been already projected onto non-negative values.

4.2.2 Results from the 120 projections data set

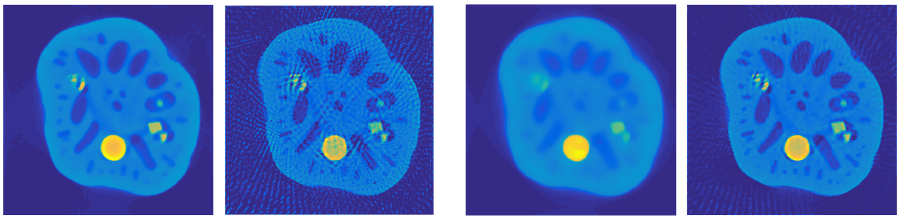
We want to start this dissertation by showing the robustness of both the proposed algorithms with respect to the regularization parameter λ. For what concerns the SpCT case with 120 projections, figure 4.15 shows the outputs of the SGP and the FP methods, after 1000 iterations, for different values of λ spacing from 10^{-2} to 10^{+3}. The FP method never provides an unacceptable reconstraction, even if with the smallest parameters the images are affected by tipical artifacts due to the subsampling (that look like rays, starting from the center of the image and extending to the edges): in this case, we can state that a good regularization parameter should be taken in [10, 100] approximately. On the other hand, it is evident that high values for λ are not adequate to the SGP algorithm, that prefers a parameter in [0.1, 1]: lower values exhibit artifacts, while higher ones blur the objects too much. Anyway, both the suitable ranges are one order large and it confirms the robustness of the two solvers. In addition, the FP requires a stronger regularization than the SGP algorithm: it is to compensate the lack of the additional "a priori" information of the non-negativity of the solution, that is incuded in the constrained formulation (2.28) for the SGP. This feature can also be seen as a positive confirm of the proposed model.

If we performe the same analysis we have just done for the early reconstructions, we probably would set the suitable ranges for λ in a different way: if we are interested in fast reconstructions from sparse tomographic data (and not only in SpCT IIR), we could prefere to fix smaller regularization parameters above all for the FP algorithm. Looking at figure 4.16, displaying the reconstructions achieved in 10 iterations, we realize that the FP outperforms the SGP when $\lambda \leq 1$, while for larger values all the reconstructions are too blurred. This behaviour can be explained by the fact that in this case, the first iterations do not spread noise nor artefacts too much, hence a good reconstruction aims to minimize the data-fitting term at first. On the contrary, a too heavy Total Variation can slow down the object detention in the first steps.



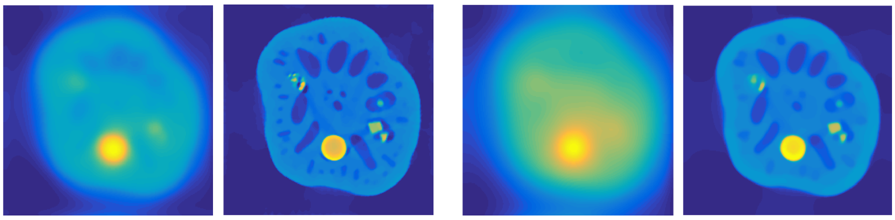
(a) SGP and FP with $\lambda = 0.01$

(b) SGP and FP with $\lambda = 0.1$



(c) SGP and FP with $\lambda = 1$

(d) SGP and FP with $\lambda = 10$



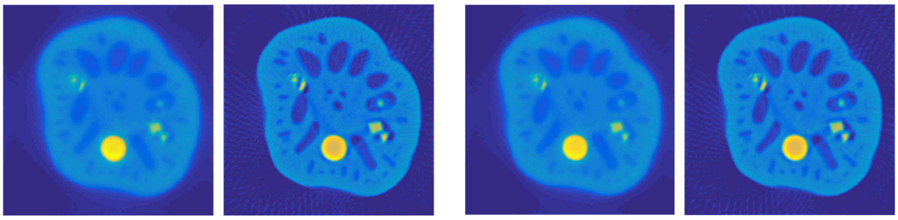
(e) SGP and FP with $\lambda = 100$

(f) SGP and FP with $\lambda = 1000$

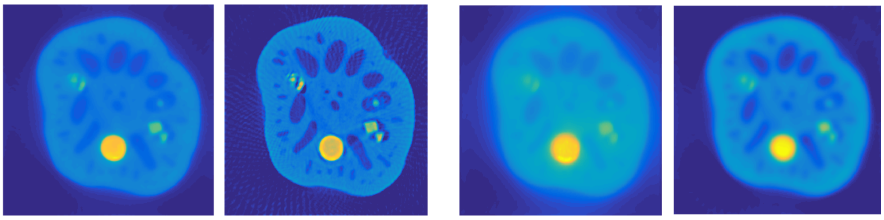
Figure 4.15: Convergence results achieved in 1000 iterations, for varing values of the regularization parameter. In each subfigure, the image on the left is the SGP reconstruction, the one on the right is the projected FP output.

The choise of a proper regularization parameter for SpCT reconstruction may represent an open question for future investigations.

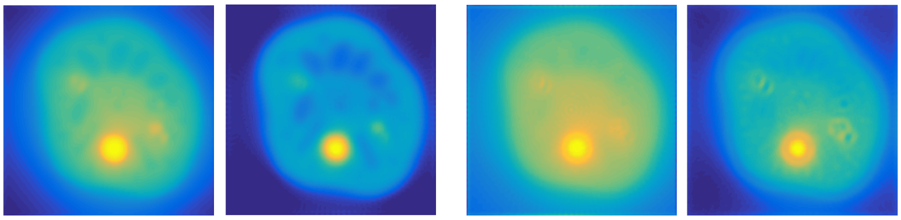
Now, we fix heuristically $\lambda = 1$ for the SGP algorithm and test both $\lambda = 1$ and $\lambda = 100$ for the FP method. We aim to understand how the reconstructions evolve between the two extreme cases already analysed: we think this approach could be of interest also for an industrial SpCT application. Figure 4.17 reports the images we get in 10, 50, 100 and 1000 iteration, while in figure 4.18 two zooms are offered to better appreciate the detection of the match-heads and of the pencil mine together with the ceramic fragments (in the zooms on the left and on the right hand side respectively, in each subfigure) for each reconstruction.
As already stated, in only 10 iterations the best reconstruction is achieved by the FP method with a little regularization (central image in figure 4.17(a) and zoomed crops in 4.18(b)): it is the only one that detects the lotus edges



(a) SGP and FP with $\lambda = 0.01$ (b) SGP and FP with $\lambda = 0.1$



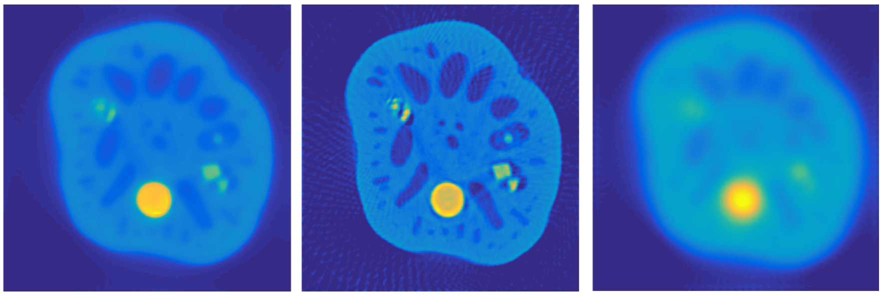
(c) SGP and FP with $\lambda = 1$ (d) SGP and FP with $\lambda = 10$



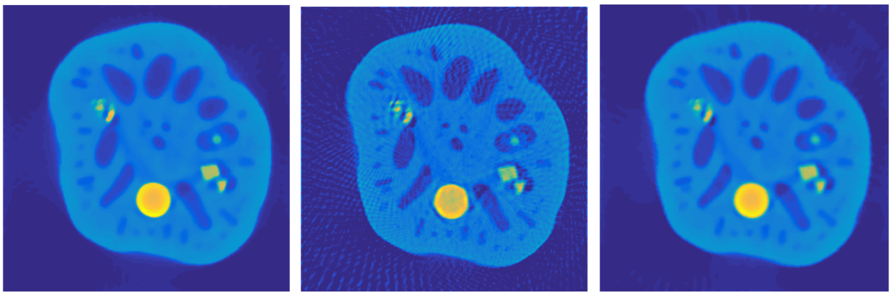
(e) SGP and FP with $\lambda = 100$ (f) SGP and FP with $\lambda = 1000$

Figure 4.16: Results achieved in only 10 iterations, for varing values of the regularization parameter. In each subfigure, the image on the left is the SGP reconstruction, the one on the right is the projected FP output.

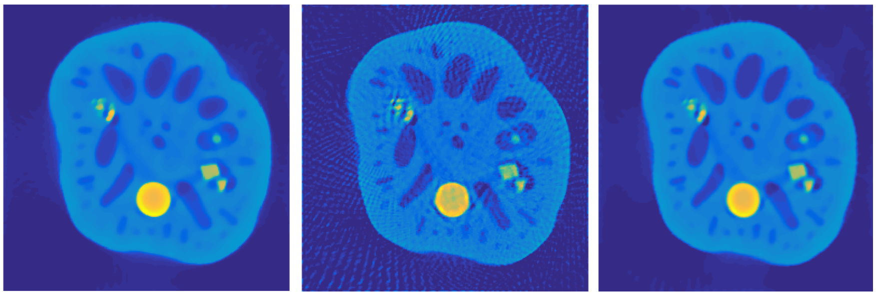
with a certain accuracy and the inserted objects are already well defined, even if many artifact lines are too visible. Going on with the iterations, a turning point is around $k = 50$, when the SGP method and the FP one with $\lambda = 100$ furnish similiar and good reconstructions. Of course these images are not as reliable as the convergence results, but by comparing them to their corresponding later outputs (provided after 100 and 1000 iterations) we can assert that in only 50 executions both the proposed solvers get very close to their targets. Focusing on the details in figure 4.18, we realize that the TV regularization (with its suitable parameter, of course) indeed removes many artifact effects and prevents noise propagation in this SpCT case: all the objects and the lotus background are well detected with their regular shapes, uniform densities and only 100 iterations are enough to reconstruct the pencil mine round section with a good contrast.



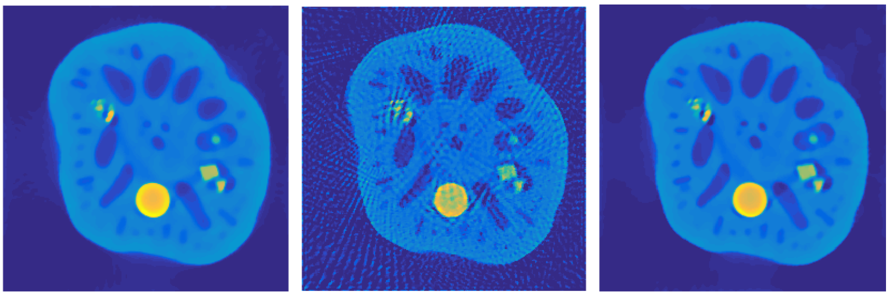
(a) Outputs at 10 iterations



(b) Outputs at 50 iterations

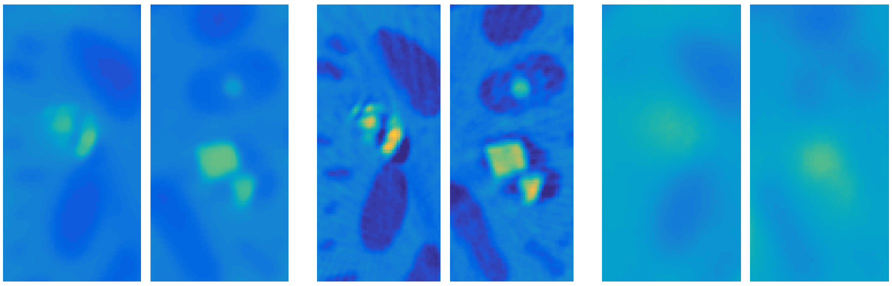


(c) Outputs at 100 iterations

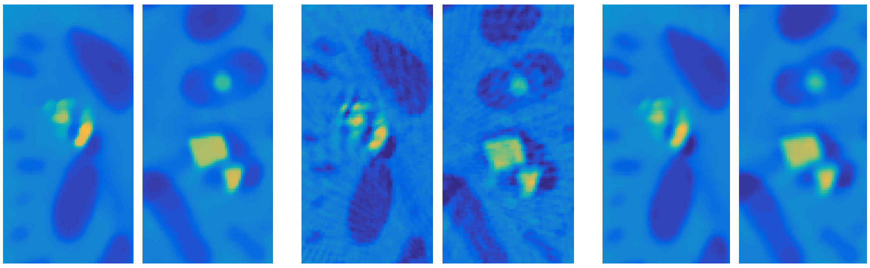


(d) Outputs at 1000 iterations

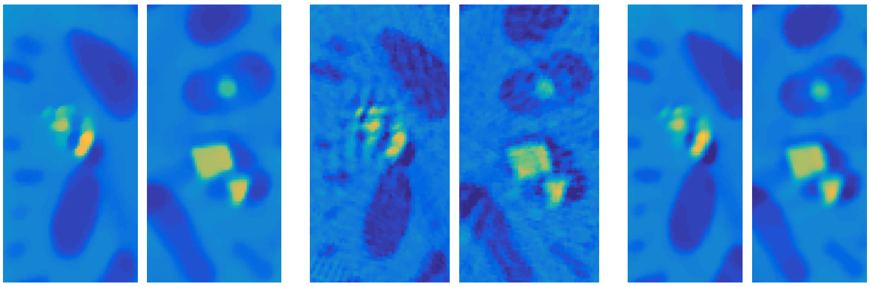
Figure 4.17: Results obtained by the SGP with $\lambda = 1$ in the left column, by the FP with $\lambda = 1$ in the central column, by the FP with $\lambda = 100$ on the right column.



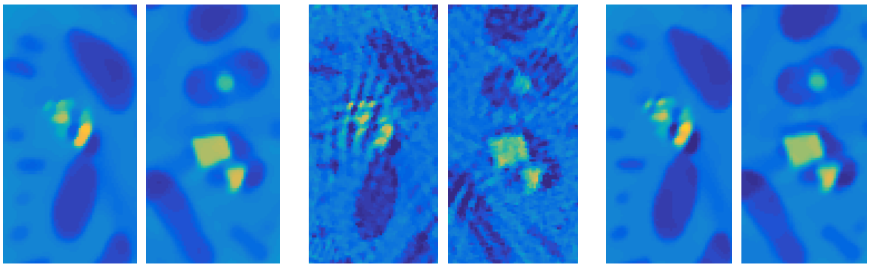
(a) SGP with $\lambda = 1$ (b) FP with $\lambda = 1$ (c) FP with $\lambda = 100$



(d) SGP with $\lambda = 1$ (e) FP with $\lambda = 1$ (f) FP with $\lambda = 100$



(g) SGP with $\lambda = 1$ (h) FP with $\lambda = 1$ (i) FP with $\lambda = 100$

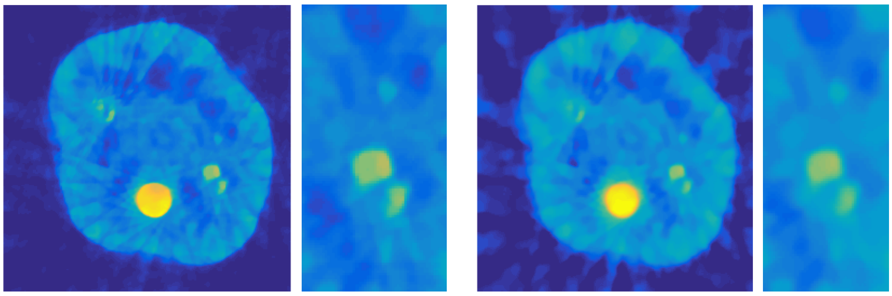


(j) SGP with $\lambda = 1$ (k) FP with $\lambda = 1$ (l) FP with $\lambda = 100$

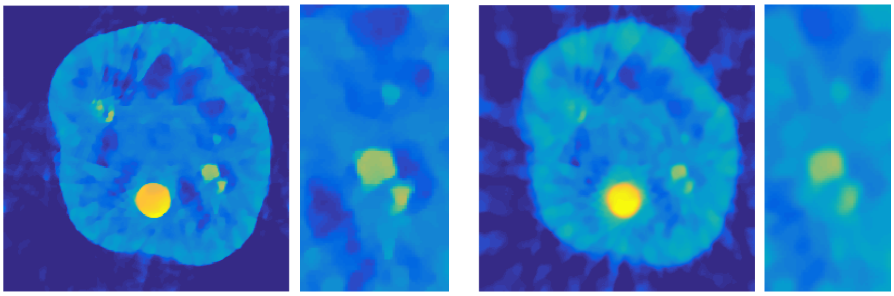
Figure 4.18: Results obtained in 10 iterations on the first row, in 50 iterations on the second row, in 100 iterations on the third row and in 1000 iterations on the last row.

4.2.3 Results from the 20 projections data set

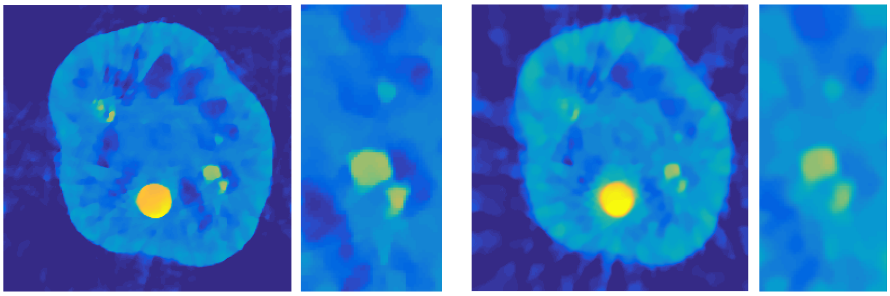
For our studies, it is interesting to compare the SGP and the FP methods also on a more challenging underdetermined CT problem: to do that, we now consider a further subsampling of the CT acquisition where only 20 projections are performed with a wide angular step of 18 degrees. The resulting sinogram is shown in figure 4.14(b).
When so few projections are available, more iterations must be performed to achieve reasonable results and for sure some artifacts will be present in all the reconstructions, whatever solver we use. Figure 4.19 shows some recostructions we get from the SGP algorithm with $\lambda = 0.1$ and the FP method with $\lambda = 100$, together with the relative zooms over the pencil and the ceramics pieces.
In 20 iterations, only the calcium elements are visible, while the mine appears after about 50 iterations but it is still a low contrast object, above all on the FP reconstruction. The match-heads are more difficult to detect: because of the missing views and their small dimensions, their high contrast intensities degenerate into an unreal ray beam, centered on them and spread along many directions. Such artifact sourrounds other objects too, but their more regular shaped reconstruction is favoured by the TV regularization. Enhanced reconstructions are the convergence outputs (i. e. made in 145 iterations by the SGP and 1000 iterations by the FP algorithm), where both the lotus slice and the external background are more uniform.
By the way, from this test we realize the importance of running the reconstructing solvers for longer time, when the CT problem is characterized by a severe subsampling, and it mirrors the difficulty to back-project the data and fit them, in case of SpCT datasets. Furthremore, another difference between the two methods relies in the SGP capability to reach its convergence while the FP can not do that, because it only performs 10 iterations (as we count 1000 iterations taking into accont also the CG inner executions).



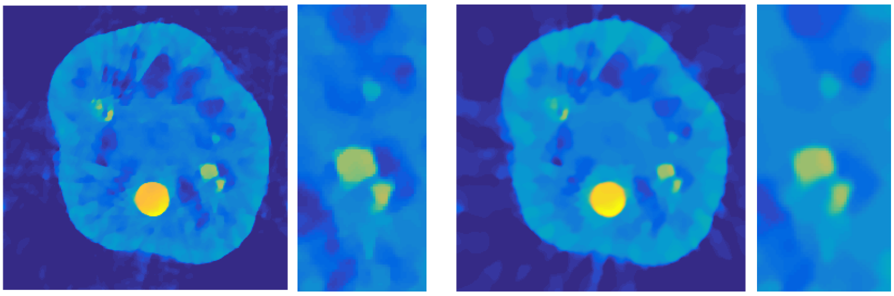
(a) SGP method at 20 iterations (b) FP method at at 20 iterations



(c) SGP method at 50 iterations (d) FP method at 50 iterations



(e) SGP method at 100 iterations (f) FP method at 100 iterations



(g) SGP method at convergence (h) FP method at 1000 iterations

Figure 4.19: Results obtained from the dataset with 20 projections, by the SGP with $\lambda = 10^{-1}$ and by the FP with $\lambda = 100$.

4.3 Comparisons on a simulated DBT case

In this section, the application of the SGP and the FP techniques to SpCT problems goes on comparing them on a 3D simulation test. In this case, the Digital Breast Tomosynthesis geometry is considered, hence projections are virtually acquired from sparse and equispaced positions on a limited angular range. We believe it could be a proper starting point to understand the main issues of such a challenging subsampled tomographic technique and its characteristic artifacts [70].
All these tests are implemented in Matlab R2016a and performed on a computer equipped with two processors Intel (R) Xeon(R) CPU E5-2650, 2.30 GHz, 20 cores, with 132 GB RAM.

4.3.1 Test problem and setting

To better simulate the breast imaging, we have defined a digital version of an accreditation phantom, called *CIRS mod. 015* [1], really used in mammography to quantify the accuracy of a machinery. In our tests, such exact volume f^* is discretized in 128×128×15 voxels with a resolution of 1×1×3 mm^3 that is almost proportional to some real cases.
Figure 4.20 shows the center slices of this phantom: inside a voxel-thin boundary of simulated skin, objects comparable to fibers, microcalcifications and masses are neatly put in a uniform background of adipose-like tissue. Outside the skin-made boundary, air is simulated with null attenuation coefficient and it is part of the volume we want to reconstruct together with the phantom. CIRS mod. 015 is used to check the competency of mammographic systems because the included objects are very important for the early detection of a breast cancer: in particular, they are of different dimensions and thickness, in order to analyse and compare the 3D graphic resolution of different reconstructions, and their attenuation values are performed taking into account a low-dose X-ray scan of 20 keV (in table 4.4 the attenuation coefficients we used to create our phantom).

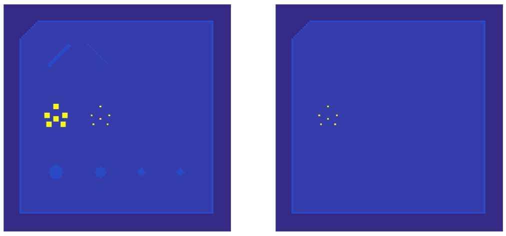
As shown in figure 4.20(c), the phantom has values in [0, 3] and most

Object	Attenuation coefficient
adipose tissue	0.1703
skin	0.24
fibers and masses	0.27
microcalcifications	3

Table 4.4: All breast objects are characterized by different densities. The table reports the attenuation coefficients related to a 20 keV scan for our simulated geometry.

anatomical structures are low contrast objects on the adipose background. On the contrary, microcalcifications are typically of very little size but they have a greater attenuation coefficient than all the other structures. This phantom includes many microcalcifications laying on the central layer and, in particular, one-voxel thin microcalcifications on the right are very useful to analyse the reconstructing algorithms even on the smallest objects.

(a) Layer 6 (b) Layer 7 (c) Layer 8



(d) Layer 9 (e) Layer 10

Figure 4.20: Central layers of the digital mammographic phantom CIRS mod. 015, sliced in 15 digital planes with resolution of 128×128 pixels.

For what concerns the geometry setting, the detector is supposed to have 128×128 pixels and its extension is the same of the volume area on the xy-plane. Of course the detector captures all the projections entirely, even the most angled ones because, thanks to the sourrounding air ring, the Cirs surface is smaller than the detector. The tomosynthesis acquisition is performed from 13 angles, uniformly distributed from -17 and +17 degrees. The central cone beam starts 64 cm above the detector and the X-ray monochrome source wheels along the y-direction, with a 59 cm ray, representing a realistic DBT scanner.
The matrix is computed once, with a ray-driven algorithm, and then stored because of its medium size: M is a 212992×245760 sparse matrix, with only 5494248 nonzeros elements, meaning that only one thousandth of its entries must be stored.
The simulation problem is again defined by computing the projection array as $g = Mf^* + e$, where e is a Gaussian noise of relative noise level

$RNL = \frac{\|e\|}{\|g\|} = 10^{-3}$.
Outputs are quantitative evaluated as we did in section 4.1, i. e. by computing the relative error $RelErr$ of the whole reconstructions and the standard deviation value $StdDev$, that now is computed inside a rectangular area in the upper-right side of the central layer where the actual phantom is uniform (figure 4.21(a)). To better highlight the object detection during the code executions, we analyse some early outputs comparing some profiles of interest: we can take Horizontal Profiles (HP) on the central layer, one over the microcalcifications and one over the masses (in figures 4.21(b) and 4.21(c) respectively), while to analyse the algorithm 3D accuracy we can focus on two specific voxels (figure 4.21(d)) and plot z-directional profiles (Depth Profiles DP) fixing the (x, y) coordinates. We underline that any HP on in-plane slices is interesting because it is parallel to the source motion: due to the reduced scanning range of the DBT, we forecast shadow-like artifacts (called ring-artifacts) along the horizontal direction.

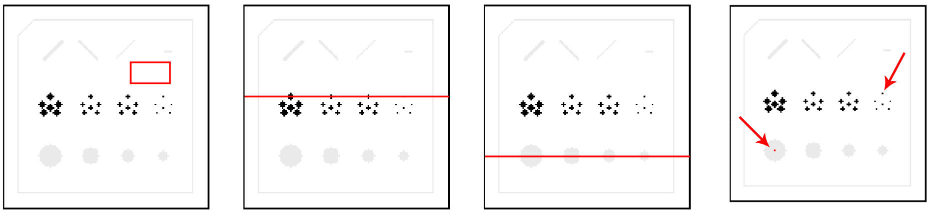

(a) (b) (c) (d)

Figure 4.21: Objects we will focus on to analyze reconstructions deeply.

Furthermore, the TV makes use of $\beta = 10^{-3}$ for its differentiability in both algorithms, while the regularization parameter has been heuristically chosen as $\lambda = 0.09$ for the SGP and $\lambda = 0.5$ for the FP: this choises ensure good results since we carefully evaluated the robustness of both methods. In the SGP code, we made use of the same parameter setting of the previous section, while the maximum number of iteration for the inner CG loop, in the FP algorithm, is 10.
Both the algorithms start from an initial constant vector $f^{(0)} = \frac{1}{N_v}$ and run for a prefixed time.

4.3.2 Results

Table 4.5 shortly analyzes the behaviour of the SGP and the FP methods during their executions, looking at temporal windows of 1, 5, 20, 60 and 120 seconds of run time. The corresponding reconstruction of the eighth slice are shown in figure 4.22 for the 1, 5 and 60 seconds, while three central layers for the 120 outputs are presented in figure 4.23 to better appreciate the accuracy

	RelErr		iters		StdDev	
	SGP	FP	SGP	FP	SGP	FP
1 sec	0.4205	0.3834	4	2	3.5 e-3	9.7 e-4
5 secs	0.2787	0.3057	14	6	2.8 e-3	2.8 e-4
20 sec	0.1655	0.1930	51	21	5.7 e-4	2.6 e-4
60 secs	0.1395	0.1105	147	63	2.1 e-4	8.5 e-5
120 secs	0.0898	0.1067	300	124	1.5 e-4	7.1 e-5

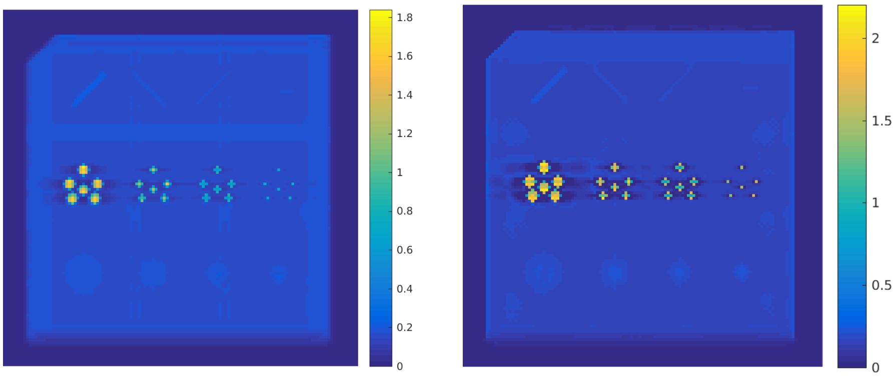
Table 4.5: Results for the SGP and FP methods at different run times.

of advanced reconstructions. We highlight that the FP solutions have been already projected onto non-negative values, as a post-processing step.

First of all, we can appreciate that all the microcalcifications are detected, even the smallest ones and even on the earliest outputs: this is an encouraging remark. In particular, the solutions achieved in only one second are comparable to the fastest reconstructions of commercial softwares provided in less than one minute: the SGP method performed 4 iterations and the FP algorithm only 2 (plus 20 inner executions of CG steps). Focusing on figures 4.22(a) and 4.22(b), we realize that the voxel intensities are not yet in [0, 3] but all the objects are detected. Of course many swiping artifacts are well visible, but they are due to the limited range characterizing the DBT technique and more regularization (hence more iterations) are required to delete them.
After few seconds, the SGP reconstruction is more uniform than the FP one, but its objects are less defined and constant. On the other hand, the FP reconstruction does not fulfill the full range yet and still exhibits strong shadows-like artifacts. They both improve the vertical skin edges, that are more difficult to recover because they are perpendicular to the scanning trajectory.
In one minute long execution, the FP output looks more accurate and in one further minute the outputs seem not to improved significantly.
By looking at the HP shown in figure 4.24, we confirm that both the microcalcifications and the masses have been faster detected by the Fixed Point algorithm (blue lines), but it also provides undesired ring artifacts around the objects borders (which are partially erased by projecting the negative voxels to zero). In 60 second outputs, the FP definitely overcomes the SGP solver avoiding artifacts and providing the same in-plane accuracy of the convergence solution.

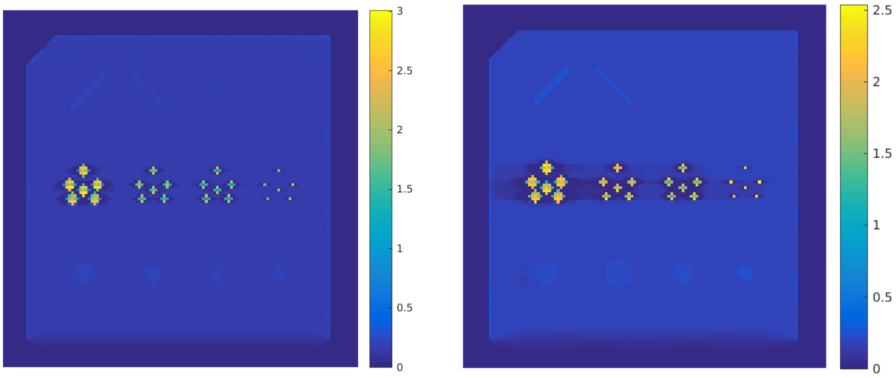
Maybe the depth profiles in figure 4.25 mirrow the most important difference between the proposed algorithms for this simulation. In DBT imaging, in fact, even if we are interessed in 3D reconstructions with anisotropic accuracy, the precision along the $z-$axis provides important measurements in case of biopsies. Looking at the plots related to the thinnest microcalcification, we can thus confirm the excellent capability of both the methods to locate voxel-wise elements of high intensity in their exact positions, even in few iterations. On the contrary, the plots on the right hand side column are related to the central voxel of the biggest mass (that is 3 layer thick): they reveal the inability of the SGP algorithm to detect lighter and plainer objects while the FP method needs short time to detect smoother structures. In particular, the SGP method detects the mass, but with a very lower intensity, and it spreads it into too many slices, causing a long backprojection of the mass measurements. Only with longer executions the SGP approaches the FP depth accuracy.

To conclude, we think that the Scaled Gradient Method and the Fixed Point algorithm may provide useful reconstructions when applied to a DBT problem.



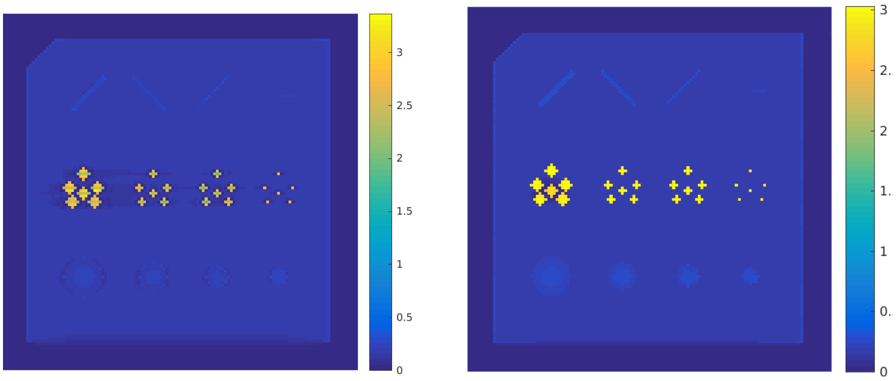
(a) SGP in 1 second

(b) FP in 1 second



(c) SGP in 5 seconds

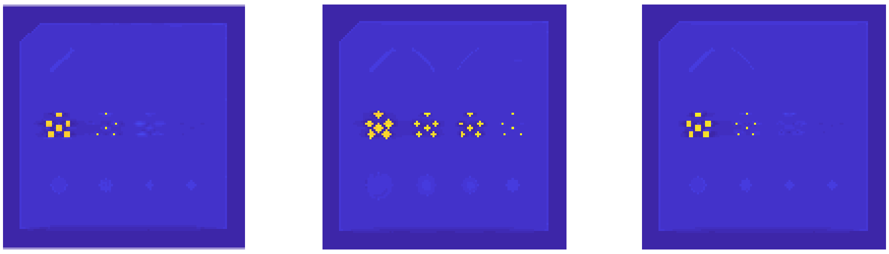
(d) FP in 5 seconds



(e) SGP in 60 seconds

(f) FP in 60 seconds

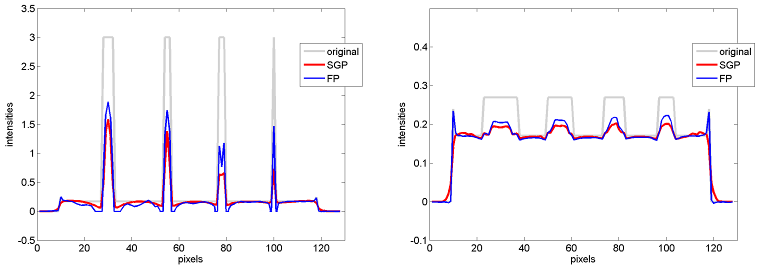
Figure 4.22: Results for the SGP on the left and FP method on the right, at different execution times, with the colorbar that is relative to the whole volume in output and mirrows the



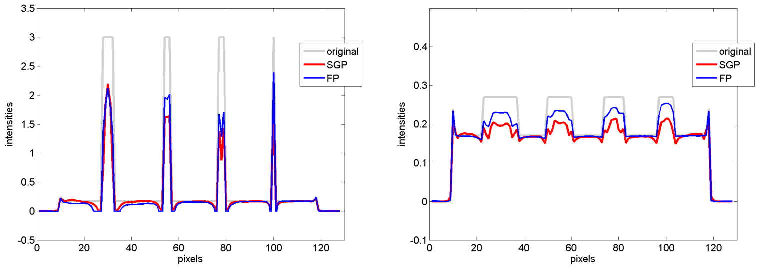
(a) Central layers of the SGP reconstruction, having values in [0, 3.1926]

(b) Central layers of the FP reconstruction, having values in [0, 3.0226]

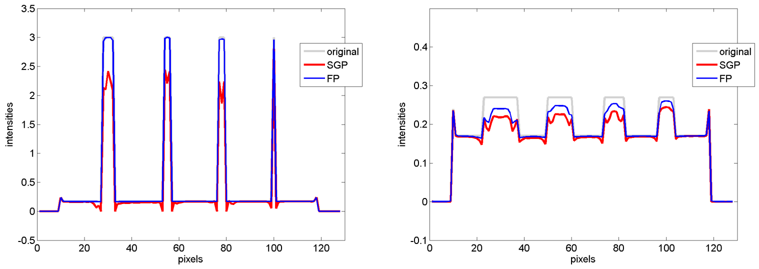
Figure 4.23: Layers 7, 8 and 9 of the SGP and FP reconstruction after 120 seconds of run time.



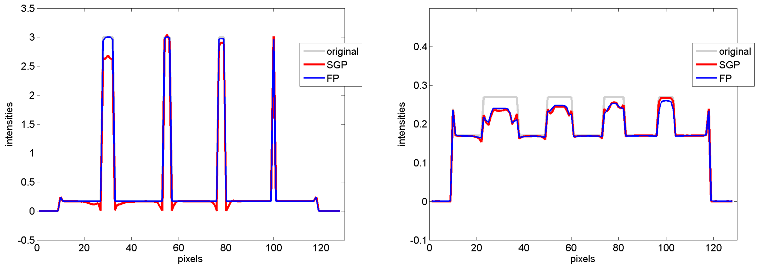
(a) HP in 1 second



(b) HP in 5 seconds

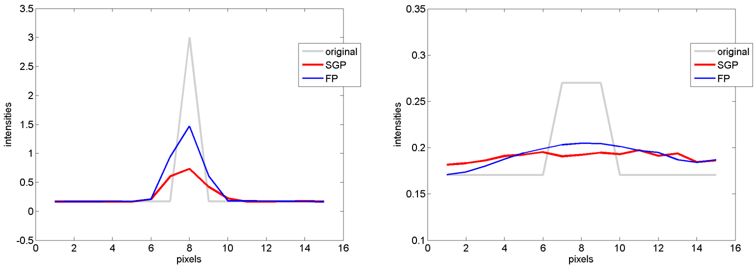


(c) HP in 60 second

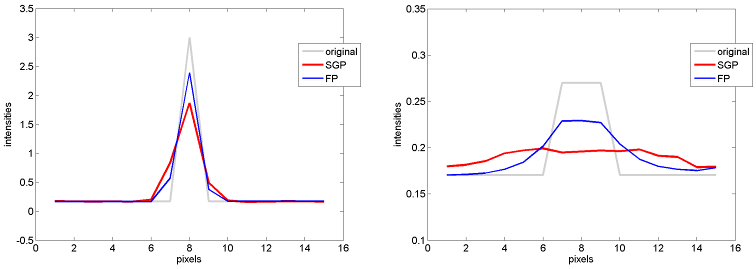


(d) HP in 120 seconds

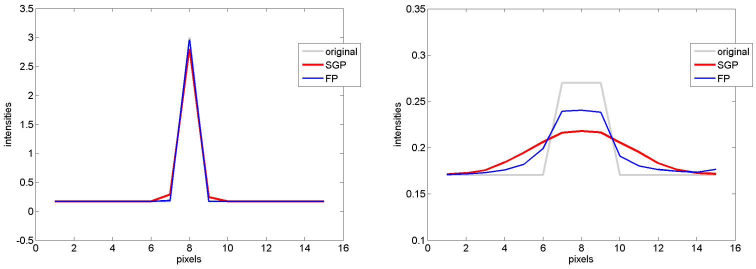
Figure 4.24: Horizontal Profiles (HP) taken on the reconstructions at 1, 5, 60 and 120 seconds, both on the microcalcifications (left column) and on the masses (right column).



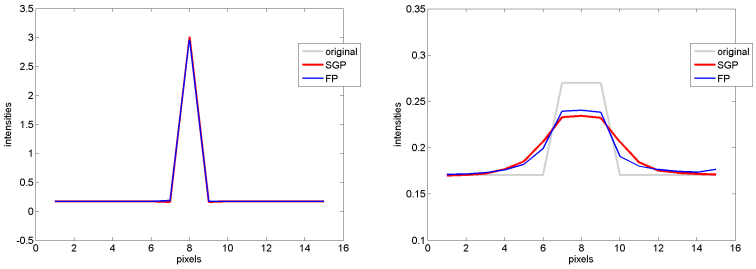
(a) DP in 1 second



(b) DP in 5 seconds



(c) DP in 60 seconds



(d) DP in 120 seconds

Figure 4.25: Depth Profiles (DP) taken on the reconstructions at 1, 5, 60 and 120 seconds, both on the microcalcifications (left column) and on the masses (right column).

Chapter 5

Results with real DBT data

In this chapter we finally test the SGP and the FP solvers on 2D real projection data, that have been acquired from a DBT device.

We used mainly the digital system *Giotto Class* of the IMS Giotto Spa company in Bologna, Italy [5, 3], that is reported in figure 5.1. When in vertical position (θ=0 degrees), the source is approximately 70 cm over the detector, then it rotates on a 30° angular range over the $y-$axis and the unit acquires 11 projection views with a low dose protocol. The stationary digital detector has a sensitive area of 24x30 cm^2 and squared pixel pitch $\delta_x = \delta_y = 0.085\ mm$; the reconstructed voxel sizes are $\Delta_x = \Delta_y = 0.090\ mm$ and $\Delta_z = 1\ mm$.

Figure 5.1: Giotto Class device, by the IMS Giotto company.

To handle with the resulting optimization problems, a C-based code has been implemented. Here the projection matrix M is computed with the Distance Driven algorithm, discussed in 2.2.2. The resulting sparse matrix is never stored, but it is re-evaluated row by row at each matrix-involving product:

looking at the run time, it becomes evident that this is the most onerous tasks in each step of both the algorithms. For this reason, in the following dissertation the SGP and FP solvers are always stopped after a pre-fixed maximum number of iterations. In particular, the primary recostructions are the earlier outputs we get after 4 iterations of the SGP solver or after only 1 iteration of the FP scheme, where 3 CG executions solve the inner linear system, to sum up the same number of matrix-vector products. We set such stopping criterion to fit better with commercial timing: the breast volume must be displayed in less than one minute. The "advance" reconstructions are achieved with 12 global iterations (that become 3 FP executions plus 9 CG total iterations): we believe they could provide improvements on the images and highlight important features to analyse the trend of the algorithms. Nevertheless, performing 12 iterations may be a feasible target, especially in case of uncertain diagnosis where further iterations may be launched offline.

Dealing with reconstructions from real data also forces to face many troubles which do not emerge with simulations. First of all, it is difficult to state quantitative indeces to analyse the results and only the radiographer trained eyes can provide reliable qualitative feedbacks. In addition, we do not know the exact range for the image values and any solver reconstructs the volume in a certain interval: it makes their comparison harder than on digital problems. For sure, the most important feature is to make the suspicious objects as distinguishable as possible among the sourranding tissues, hence we can focus on their relative appearance in the slices, and not on their reconstructed values, as suggested in the Technical Evaluation of a DBT system by the English Public Health [39]. As a consequence, the following in-plane profiles are all normalized in [0, 1] with respect to their original proportions. Furthermore, each in-plane and depth profile represents a mean over adjacent voxels, because one-voxel thin profile would have been a too noisy plot and a disloyal visualization of the object detectability inside the image.

In particular, in both the European Protocol [97] and the American approach [90] to the Quality Control (QC) of DBT imaging, the Depth Profile analysis is made through the study of an artifact spread along the depth-direction. We have thus defined an *Artifact Spread Function* (ASF) with values in [0, 1], in the following way: if an object is centered in the $\bar{z}$ slice, let $m_{obj}(\bar{z})$ be the mean of the intensities of the voxels in a ROI inside the object and $m_{bg}(\bar{z})$ the mean of voxels of a ROI in the same layer but on a background area. If we compute $m_{obj}(z_{j_z})$ and $m_{bg}(z_{j_z})$ in the same ROIs but over all the slices z_{j_z} where $j_z \in 1, \dots, N_z$, we can define the ASF values as

$$ASF(j_z) = \frac{|m_{obj}(z_{j_z}) - m_{bg}(z_{j_z})|}{m_{obj}(\bar{z}) - m_{bg}(\bar{z})} \quad \forall j_z \in 1, \dots, N_z. \tag{5.1}$$

In the following, we report the most noticeable results we get from two

breast quality control phantoms and, at last, from a real medical examination. All the images are rotated, as usual habit for right cranio-caudal views, with the $y-$axis in vertical orientation on the right hand side.
All the reconstruction analysis have been performed in ImageJ environment [4]. ImageJ is a public domain Java-based image processing program developed at the United States National Institutes of Health. It can read many image formats including TIFF, DICOM and "raw" and it supports "stacks" (i.e. series of images that share a single window), hence it is a useful platform for 3D CT imaging. It also includes standard image processing functions and further analysis or processing plugins can be developed using ImageJ's built-in editor and a Java compiler.

5.1 Results on a Quality Control phantom

The TomoPhan (TP) phantom TSP004 by PhantomLab [8] is designed to perform Quality Control tests to the reconstruction system on specific metallic objects and it is reported in figure 5.2.
The TP phantom is 17 cm width and 4,2 cm thick; it is mainly made of

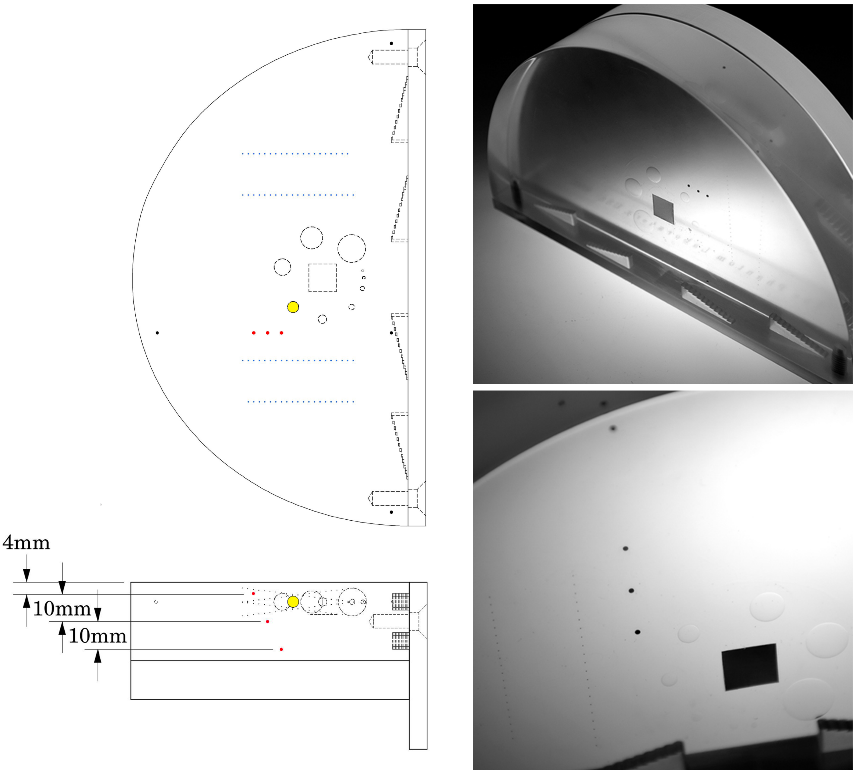
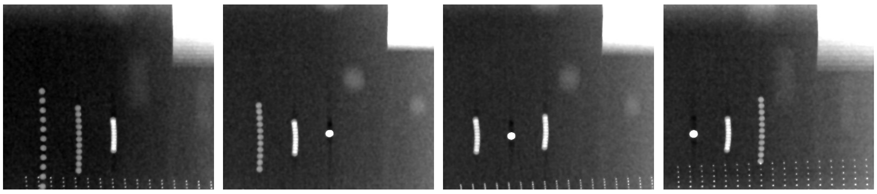
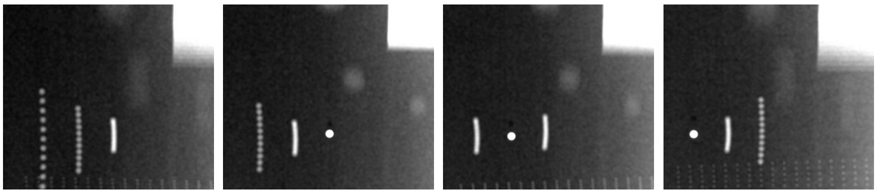


Figure 5.2: Technical drawing of the TP phantom with coloured objects of interest on the left and pictures of the phantom on the right.

uniform breast equivalent material and it has a semi-circular shape to simulate the breast. It contains three Aluminium beads (the pink ones on the TP technical drawing) to test the $z-$axis spacing: they are 0.5 mm spheres and are 10 mm spaced in depth, hence they will lie on different layers on the reconstructions. Since metallic elements cause many more shadows than anatomical tissues in the sourranding layers, it is interesting to see the artifacts they produce, which are further observable because of the homogeneous background. Inside the TP phantom there are also some acrylic bigger spheres simulating breast masses: we focus on the yellow one, in the following discussion. It is centered in layer 24, on the reconstructions.
The dataset contains 11 projection views and the regularization parameters have been set to provide the best solutions in 4 iterations: they are $\lambda = 0.5$ for the SGP and $\lambda = 10$ for the FP. All the other parameters are equal to those used in the Matlab simulations.



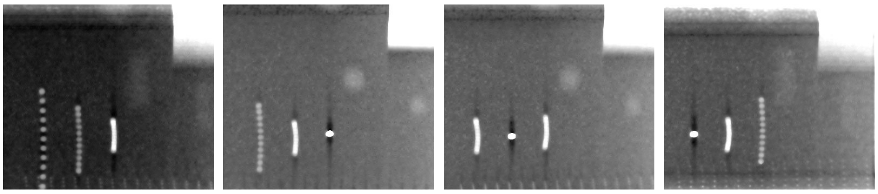
(a) SGP reconstrution in 4 iterations



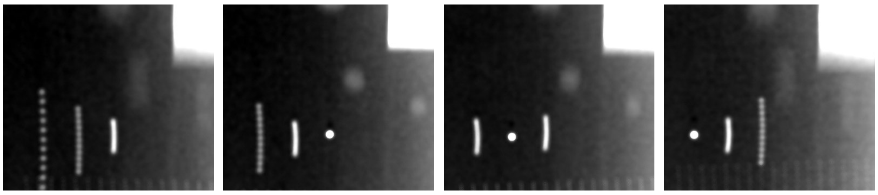
(b) FP reconstrution in 4 iterations

Figure 5.3: Slices 9, 19, 29 and 39 of the reconstructed volumes in 4 iterations.

Images in figure 5.3 exhibit layers 9, 19, 29 and 39 of the volumes reconstructed by the SGP and the FP algorithms in only 4 iterations: these reconstructions are comparable to the commercial ones. We remark that the ranges of the gray-level scale are automatically set by ImageJ, to better visualize the cropped portion of each single slice. If we focus on the three beads we can appreciate their proper shape with net boundaries and the uniformity of the background, that does not suffer too much for the sparse view artifacts in any in-plane visualization. On the other hand, we can perfectly see the shadows of these objects in all the other layers, because they degenerate in a double-fan artifact, above and below its central slice. In particular, we can recognise exactly 11 small points on the farther layers, reflecting the 11



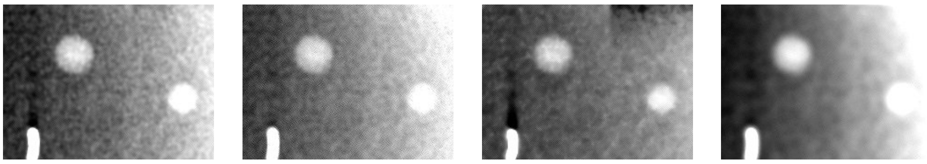
(a) SGP reconstrution in 12 iterations



(b) FP reconstrution in 12 iterations

Figure 5.4: Slices 9, 19, 29 and 39 of the reconstructed volumes in 12 iterations.

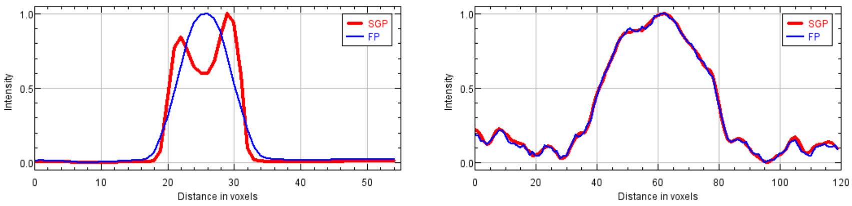
performed projections: this is a typical artifact of all the limited angle CT techniques. Their precise circular shape with clear edges suggests that the bead spread is quite limitated.
Performing more iterations (figure 5.4), the apparence of the reconstructions seems not to change a lot and it represents an important confirm of the capability of both the solvers to go close to the solution in their earlier iterations in SpCT cases. On the other hand, if we look at figure 5.5, we notice that the detection of the acrylic masses improves with longer executions, above all for the FP algorithm that provides more uniform objects and a more effective denoising. This is coherent with the results of the DBT simulation of the previous chapter.



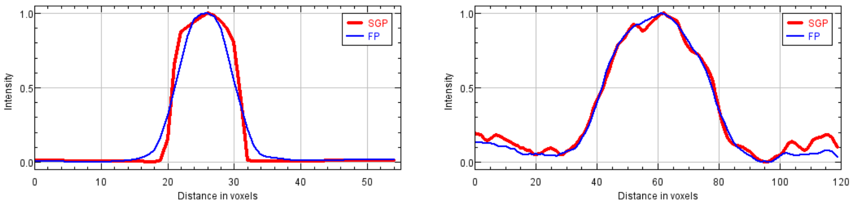
(a) SGP in 4 iters. (b) FP in 4 iters. (c) SGP in 12 iters. (d) FP in 12 iters.

Figure 5.5: Focus on the acrylic spheres.

To better compare the reconstructions, in figure 5.6 we see the Vertical Profiles (VP) on the right bead and on the mass, both in 4 and in 12 iterations. The SGP algorithm (in red lines) can not make the Aluminium bead uniform at the beginning, but in 12 iterations it achieves more precise and stepper edges. The FP solver (in blue lines) gets a smoother object but also more uniform tails out of the mass. They both do not show ringing artifacts, meaning that a good regularization may delete them.



(a) VP in 4 iterations



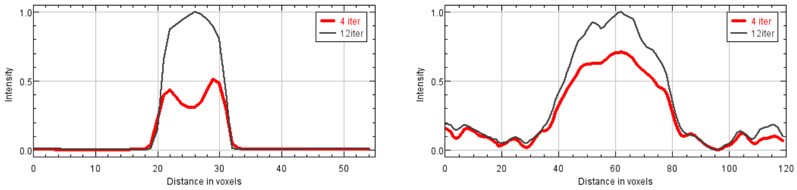
(b) VP in 12 iterations

Figure 5.6: On the left, profiles over a small Aluminium beads; on the right, profiles over the mass-like sphere.

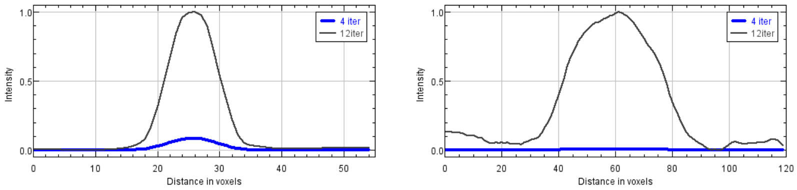
The profiles in figure 5.6 can not reproduce the algorithm executions faithfully, because we lose some details when we scale the plots in the unitary interval. In figure 5.7, the same VP profiles are reported, putting together the two reconstructions of each solver. The black line is always the "advanced" solution, while the red and blue curves are related to the SGP and FP outputs in 4 iterations respectively. The SGP fast reconstruction has values quite close to the advance ones and the acrylic sphere has almost equal intensities, while the FP algorithm shows a slower convergence. In 4 iterations, in fact, it provides very low values to all the volume and the image range changes a lot when a longer execution is allowed.

The Depth Profiles in figure 5.8 report the ASF graphics for the two objects of interest. The SGP is very precise in detecting the Aluminium bead in its actual quote, because the corresponding red lines are indeed very thin Guassian-like bells, while the blue larger bells reflect a spreading effect on the $z-$axis direction for the FP solutions. The mass detection is harder: the object is correctly located around the 24-th layer but we have high tails in all the four reconstructions.

As suggested in [15], two repeated scans were acquired with the TP phantom in exactely the same position and under equal geometry setting, hence the two outputs (computed by the same algorithm with the same



(a) VP of the SGP



(b) VP of the FP

Figure 5.7: In-plane VP over a small Aluminium beads and the mass-like sphere, taken on the 4- and 12- iteration outputs.

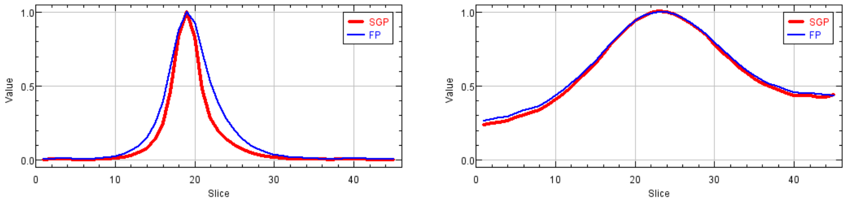
	SGP		FP	
	4 iter	12 iter	4 iter	12 iter
mean	5.96 e-6	1.6 e-4	7.3 e-6	2.22 e-5
StdDev	2.5 e-4	6 e-3	2.7 e-5	2.38 e-5

Table 5.1: Mean and Standard Deviation values of the 3D difference images presenting stochastic noise components.

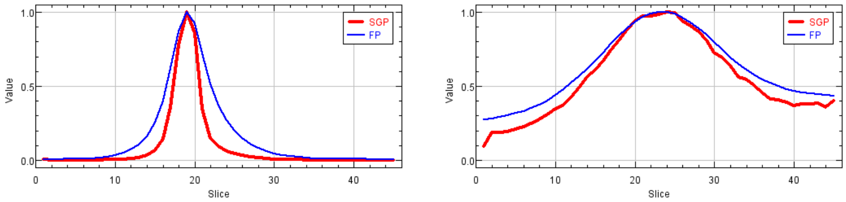
parameters) may be subtracted and the resulting 3D image is a volume of noise. Because the "anatomical" structure and its consequent noise are supposed to be the same in every couple of reconstructions, the difference between the two outputs are related to quantum and electronic noises (that are indipendent at each scan) and to their amplification in the iterative scheme. Table 5.1 reports the mean and the Standard Deviation (StdDev) of the four 3D difference images: the FP algorithm has a powerful denoising effect with longer run times, as already seen for this problem, while the standard reconstructions are comparable.

5.2 Results on a breast phantom

We now consider the projection dataset of a phantom which better simulates the breast anatomy: it is the "BR3D" model 020 phantom of the CIRS company [2]. It is characterized by a heterogeneous background, where adipose-like and gland-like tissues are mixed in approximate 50/50 ratio. It is made of six slabs (as shown in figure 5.9) and they may be arranged to



(a) DP in 4 iterations



(b) DP in 12 iterations

Figure 5.8: On the left, profiles over a small Aluminium beads; on the right, profiles over the mass-like sphere.

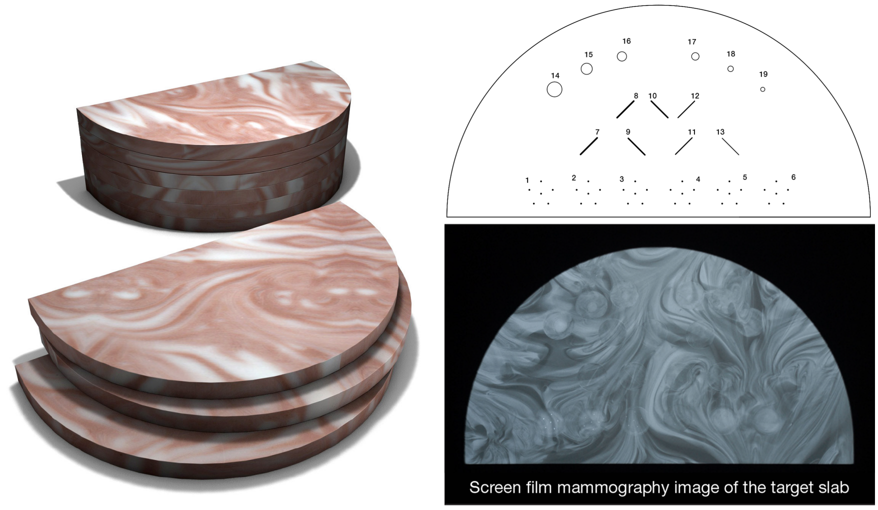
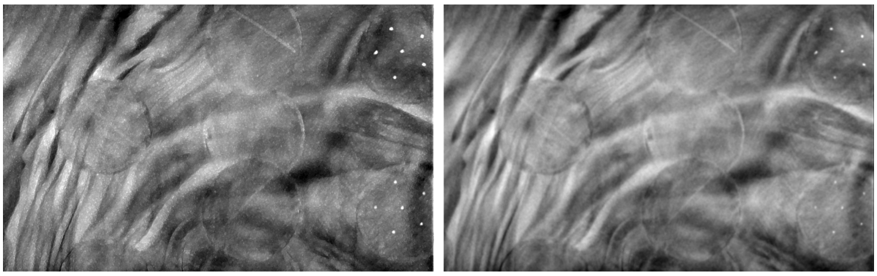


Figure 5.9: A picture of the CIRS phantom "BR3D" model 020, its technical drawing and a mammographic image.

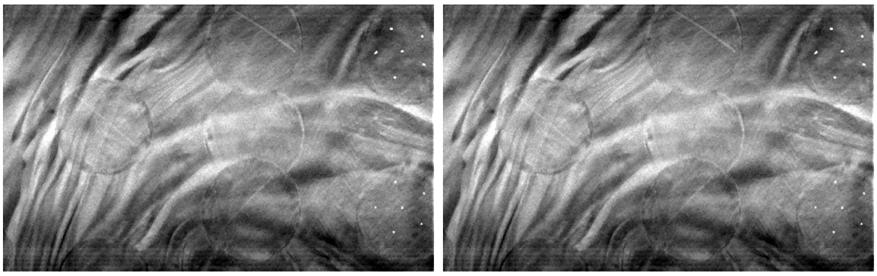
create multiple anatomical backgrounds. Inside one of them, in paricular, we find acrylic spheres simulating breast masses, 1 cm length fibers and many clusters of Calcium carbonate specks simulating microcalcifications. Each slab has a semicircular shape, 1 cm thickness and it is 10 cm x 18 cm.
We scanned the phantom in 11 projections from -14.5 to +14.5 degrees, with two different dose levels, to analyze the influence of the quantum statistics on the algorithm performances.

Reconstructions with high radiation dose

For this IIR problem, the regularization parameters are $\lambda = 0.005$ for the SGP and $\lambda = 0.001$ for the FP: the corresponding results are shown in figure 5.10. Here we focus on a cropped portion over the 5-th slices of the volumes, where we find the fibers number 10 on the left, number 9 and 11 in the middle and the microcalcification-like clusters number 3 and 4 on the right hand side.
Like in the DBT simulation of paragraph 4.3, the small microcalcification appearance is noticeable even on the earlier reconstructions, while the "anatomical" objects may have smaller constrasts in the in-plane visualization too. These images, actually, have been graphically processed to adjust their intensity ranges to the visualized portions and that is the reason why



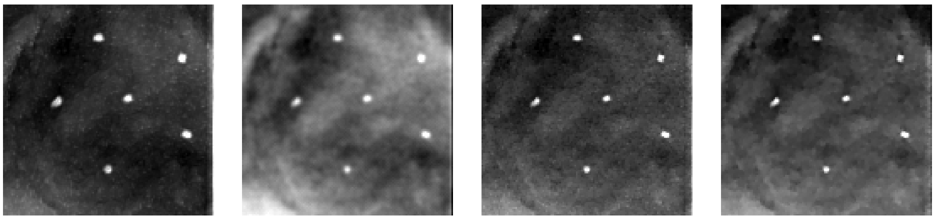
(a) SGP and FP reconstructions in 4 iterations



(b) SGP and FP reconstructions in 12 iterations

Figure 5.10: Crops of the slices 5 of the volumes, reconstructed by the SGP on the left column and by the FP on the right column, in 4 or 12 iterations.

the edges of the adipose and gland tissues are put so much in contrast.

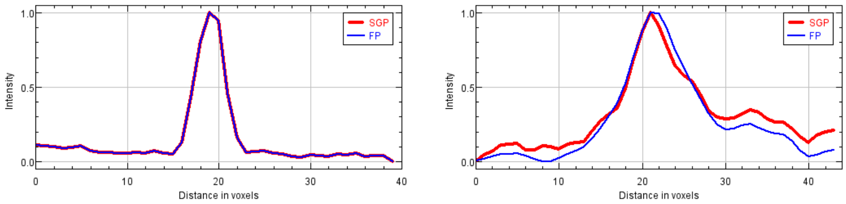


(a) SGP in 4 iters. (b) FP in 4 iters. (c) SGP in 12 iters. (d) FP in 12 iters.

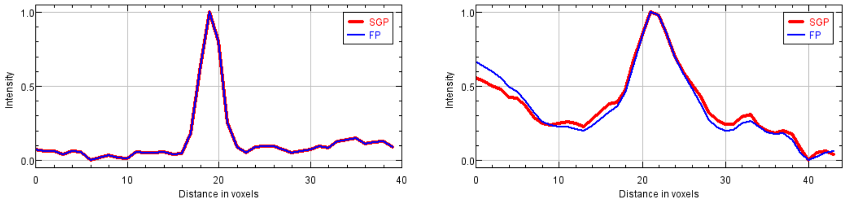
Figure 5.11: Focus on the upper cluster of microcalcifications.

Looking at the zoomed crops over the upper set of microcalcifications (in figure 5.11), we appreciate the deblurring enhancement after longer run time and the clear detection of both the algorithms after only 4 iterations. Most of the specks, in fact, already have a circular net shape and no ringing artifact or sparsity-caused shadows are present. In particular, the Vertical Profiles in figure 5.12 analyze the in-plane reconstructions of the central microcalcification and of the bottom fiber that is not too much evident on the slice. Figure 5.13 shows the same micro in the depth direction with the Artifact Spread Functions over the layers. The VP over the fiber is taken as an average line over 70 voxels and it is orthogonal to the fiber orientation.

The VP plots on the specks confirm the reliability of the reconstructions:



(a) VP in 4 iterations



(b) VP in 12 iterations

Figure 5.12: On the left, the Vertical Profiles on the central microcalcification of the third cluster; on the right, the Vertical Profiles over a thin fiber.

these kind of breast objects are immediately detected and with further iterations we achieve higher precision. The same holds for the DP: the exact in-depth position is found and the ASF becomes smoother with 12 iterations. On the other hand, the VP graphics over the fiber reveal that not all the objects may be better detected with longer executions.

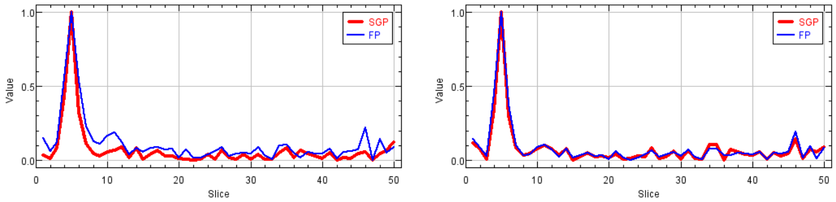


Figure 5.13: Depth Profiles over the microcalcificaion in slice 5, after 4 iterations on the left and after 12 iterations on the right hand side.

Reconstructions with low radiation dose

We now consider the same DBT acquisition but performed with a smaller dose of radiations in each projection. In particular, we have almost halved the mAs (milliampere-second), that is the measure of electric charge to describe the exposure setting of a CT device, hence it determines the density of the X-ray images. As a consequence, the projection images are much noisier then in the previous dataset. By the way, after a careful tuning, we set regularization parameters ten times smaller than in the previous reconstructions ($\lambda = 0.0005$ for the SGP and $\lambda = 0.0001$ for the FP).
Figure 5.14 shows some details to compare with figure 5.11: now we indeed have noisier backgrounds, but the microcalcifications are detached equally, with precise bounds and no shadows. This is confirmed by the VP plots in the left column of figure 5.15. The advanced solutions are not more accurate, because of the noise amplification. In addition, the values of λ have been set to let the solvers perform at their best in 4 iterations: we have already highlighted in section 4.2 that the regularization parameters may be set accordingly to the prefixed number of iterations, for SpCT reconstractions from

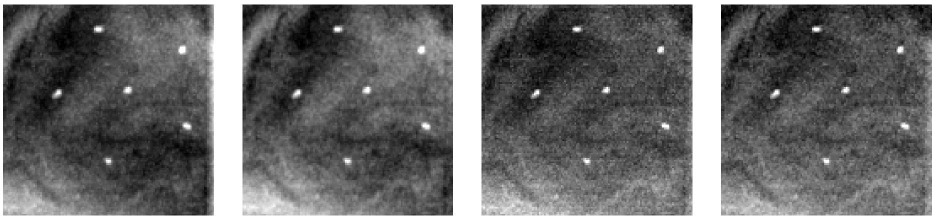
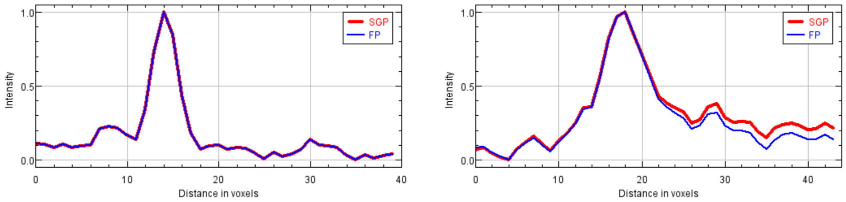
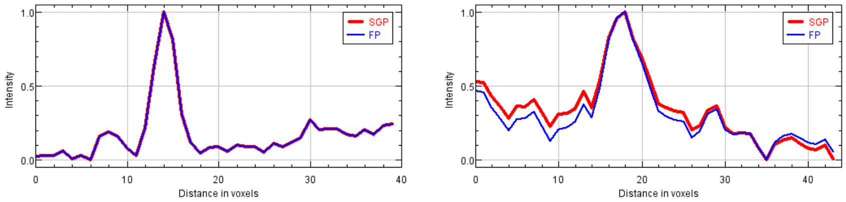

(a) SGP in 4 iters. (b) FP in 4 iters. (c) SGP in 12 iters. (d) FP in 12 iters.

Figure 5.14: Focus on the upper cluster of microcalcifications.

real data. Nevertheless, the microcalcifications are still well detected, with acute profiles.
These results further validate the capability of the proposed methods to provide accurate solutions from subsampled tomographic data, in very short run times and under different dose settings.



(a) VP in 4 iterations



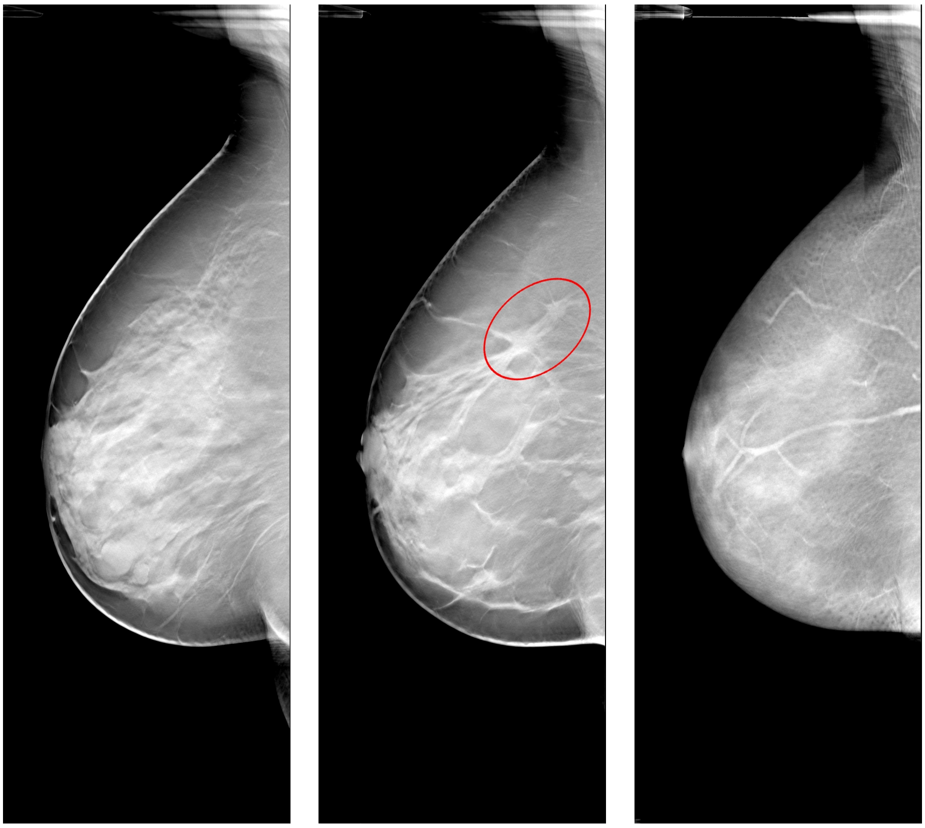
(b) VP in 12 iterations

Figure 5.15: On the left, the Vertical Profiles on the central microcalcification of the third cluster; on the right, the Vertical Profiles over a thin fiber.

5.3 Results from real breast projections

To conclude, we look at the SGP and FP performances on a real breast dataset. It has been acquired by the Giotto TOMO system, performing 13 projections in a range of 40 degrees. It also elaborates the volumetric reconstruction in less than one minute, with an approach very different from the SGP or the FP ones: the layers 30, 41 and 55 are shown in figure 5.16. We remark that the reported images are taken from the mere reconstruction, hence they are stripped of all the optimization steps usually adopted for clinical presentation. By the way, adipose and glandular tissues are well discerned and two tumoral masses are occluding some upper ductals on slice 41. We now focus on these two objects.

We have run the SGP with $\lambda = 0.001$ and the FP algorithm with $\lambda = 0.01$. Figure 5.17 compares the commercial image with our reconstructions: they all show the suspect anatomical feature. The FP solution is smooth but the tumors are prominent, while the SGP image looks much closer to our target with clearer masses. The advanced reconstructions in figure 5.18 exhibit tumors with sharper edges but the backgrounds are nois-

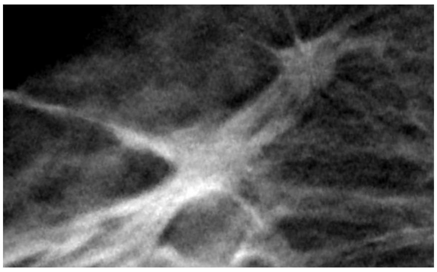


(a) Layer 30 (b) Layer 41 (c) Layer 55

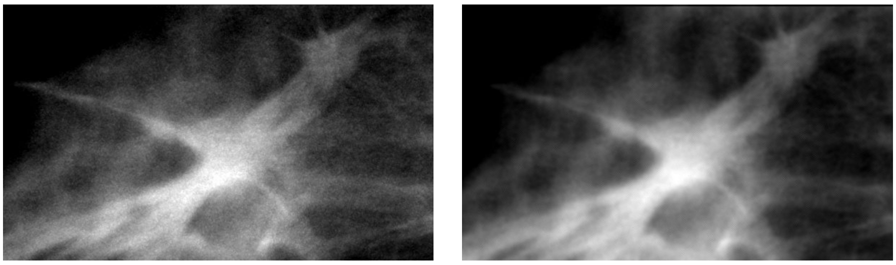
Figure 5.16: Three layers of a real DBT reconstruction, without all the grey level optimization usually performed before clinical review. Two spiculated masses are pointed out in layer 41: their in-depth position is well determined and there are no interferences in the above and below layers.

ier, especially in the SGP result. Looking at the difference images between the volumes achieved in 12 and 4 iterations by each solver (in figure 5.19), we understand the further iterations have worked on the tumoral borders but without particular upsetting improvements on the earlier reconstructions.

On the other hand, if we can run the algorithms for longer executions, we can also change their parameters to obtain their best results. For instance, figure 5.20(a) shows the 12-iterations reconstruction, achieved by the SGP method with $\lambda = 0.1$: a stronger regularization cleans away noise components and the dense masses have well bordered branches. In the FP algorithm, we can also exploit the added iterations to refine the inner linear system resolution. The picture in 5.20(b) is the FP output we get by computing 5 CG iterations in each main step, with the same regularization parameter $\lambda = 0.01$. As a consequence, only two iterations have been per-



(a) Commercial reconstruction

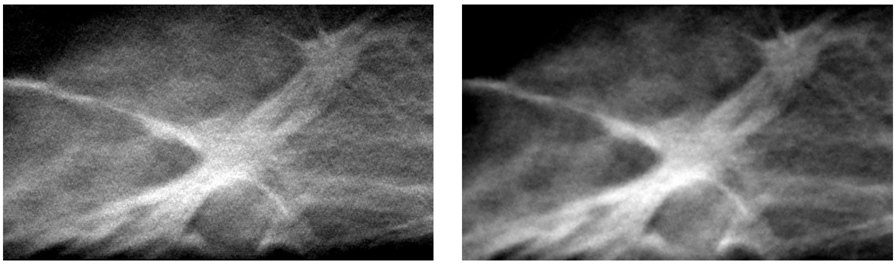


(b) SGP reconstruction

(c) FP reconstruction

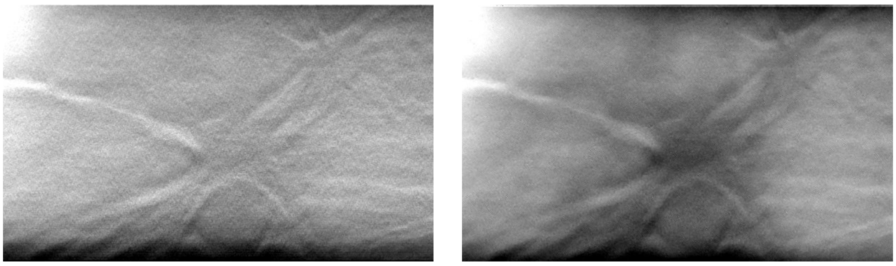
Figure 5.17: Zooms of the reconstructions in layer 41, with tumoral masses.

formed in the FP scheme; nevertheless the resulting solution refines improves the visualization of the spicules.
All these results validate the capability of getting detailed reconstructions in short run times, for the proposed methods.



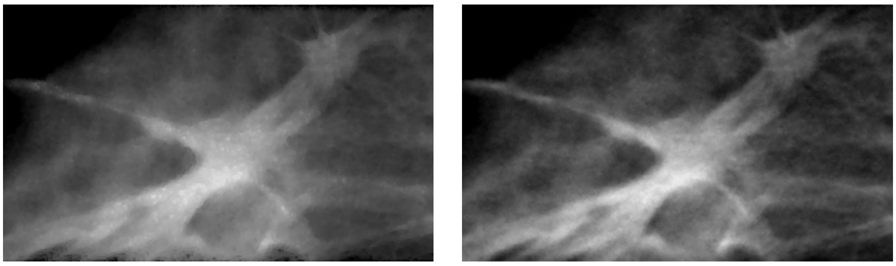
(a) SGP reconstruction (b) FP reconstruction

Figure 5.18: Zooms of the reconstructions achieved in 12 iterations.



(a) SGP (b) FP

Figure 5.19: Differences between the advanced and the standard reconstructions, for the SGP and FP solvers.



(a) SGP reconstruction (b) FP reconstruction

Figure 5.20: Zooms of the reconstructions achieved in 12 iterations.

Chapter 6

Conclusions

This PhD thesis has addressed the iterative reconstruction problem for three-dimensional CT imaging from sparse subsampling, introducing a model-based approach. In chapter 1 a general overview of the CT devices and process has been shown. In chapter 2 the tomographic linear system has been derived and numerical issues have been introduced for the Sparse CT (SpCT) case. In particular, the problem has been formulated as a regularized optimization problem in a general framework which may be adapted to a wider class of model-based approaches. In this dissertation, we have set both the Least Squares function and the Kullback-Leibler divergence to impose a data-fitting, and we have fixed the Total Variation (TV) operator as prior term to force the convergence of the solvers to a desired solution, with preserved edges and little noise.

Up to now, most of the CT devices are equipped with algebraic solvers but the results obtained in this thesis reveal that other approaches may be considered, especially in case of sparse data. Two algorithms have been proposed for the minimization problem, in chapter 3: they both belong to the class of the so-called *accurate solvers*. The first one is a Scaled Gradient Projection (SGP) solver, which has been accelerated by both a scaling strategy, based on a gradient splitting tuned for the prefixed model, and a step-length rule, defined for a more general objective function. The second algorithm is the Lagged Diffusivity Fixed Point (FP), which has been presented with an efficient implementation involving both a not expensive approximation of the TV Hessian matrix and an early-stopped inner solver, to adjust the FP scheme to tomographic problems.

The suitability of the two methods to reconstruct SpCT images has been verified on several test problems. In chapter 4, with medium-sized digital simulations we investigated the algorithms performances up to a ”mathematical” convergence for 3D reconstructions, while results from a real dataset for 2D SpCT imaging have highlighted the robustness of the solvers with respect to the regularization parameter, which is an appreciated feature for practical

applications. All the analysis validate the applicability of a model-based approach to imaging reconstructions from sparse views: accurate solutions are accessible even when many tomographic informations are missing and the reconstructions of the proposed algorithms achieve reliable solutions very early.
This result has been further strengthened by the succesful application of the SGP and FP solvers to real SpCT datasets for the 3D reconstruction of breast images, in chapter 5. To simulate their actual application, the algorithms have been tuned to perform commercial-like reconstructions and provide one volume image in less than one minute on GPUs devices: as a consequence, very few iterations have been executed. The accuracy of the reconstructions has been shown on Quality Control phantoms, widely used to evaluate commercial devices: the objects of interest are detected with sharp boundaries both on the in-plane and on the in-depth visualizations. A real breast reconstruction has been computed too: the good quality images provided by both the algorithms reflect the high reliability of the proposed approaches for their practical use in SpCT.

Bibliography

[1] Computerized Imaging Reference Systems. `http://www.cirsinc.com/products/all/47/mammographic-accreditation-phantom/`. Mammographic accreditation phantom, Model 015.

[2] Computerized Imaging Reference Systems. `http://www.cirsinc.com/products/all/51/br3d-breast-imaging-phantom/`. BR3D Breast Imaging Phantom, Model 020.

[3] Giotto Class Tomosynthesis. `http://www.tomosynthesis-class.com/`.

[4] ImageJ software: Image Processing and Analysis in Java. `https://imagej.nih.gov/ij/`.

[5] IMS Giotto. `http://www.imsgiotto.com/`.

[6] Industrial CT Scanning. `http://www.csarcheometria.it/Tomografia.html`. Accessed: 2017-09-15.

[7] Tomographic X-ray data of a lotus root filled with attenuating objects. `https://arxiv.org/pdf/1609.07299.pdf`. T. Bubba, 2016.

[8] TomoPhan phantom by PhantomLab. `https://www.phantomlab.com/tomophan-phantom/`. The Phantom Laboratory.

[9] TVReg, Version 1.1: Software for 3D Total Variation regularization. `http://www.imm.dtu.dk/~pcha/TVReg/`.

[10] US Food and Drug Administration (p080003). `https://www.accessdata.fda.gov/scrIpts/cdrh/cfdocs/cfpma/pma.cfm?id=P080003`. Accessed: 2017-09-12.

[11] A. H. Andersen. Algebraic reconstruction in CT from limited views. *IEEE Trans. Med. Imag.*, 8(1):50–55, 1989.

[12] A. H. Andersen and A. C. Kak. Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm. *Ultrason. Imaging*, 6(1):81–94, 1984.

[13] W. Yin ands. Osher et al. Bregman iterative algorithms for l1-minimization with application to compressed sensing. *SIAM J. Imaging Sci.*, 1:143–168, 2008.

[14] R. De Asmundis, D. di Serafino, W. Hager, G. Toraldo, and H. Zhang. An efficient gradient method using the Yuan steplength. *Comput. Optim. Appl.*, 59(3):541–563, 2014.

[15] G. J. Gang at al. Anatomical background and generalized detectability in tomosynthesis and cone-beam CT. *Med. Phys.*, 37(5):1948–1965, 2010.

[16] Lo JY Baker JA. Breast Tomosynthesis: state-of-the-art and review of the literature. *Acad Radiol.*, 18(10):1298–310, 2011.

[17] J. Barzilai and J.M. Borwein. Two-point step size gradient methods. *IMA J. Numer. Anal.*, 8:141–148, 1988.

[18] E.G. Birgin, J.M. Martinez, and M. Raydan. Inexact spectral projected gradient methods on convex sets. *IMA J. Numer. Anal.*, 23, 2003.

[19] A. Bjorck. *Numerical methods for least squares problems.* SIAM, Philadelphia, PA, USA, 1996.

[20] S. Bonettini, G. Landi, E. Loli Piccolomini, and L. Zanni. Scaling techniques for gradient projection-type methods in astronomical image deblurring. *Int. J. Comput. Math.*, 90(1):9–29, 2013.

[21] S. Bonettini, F. Porta, and V. Ruggiero. A variable metric inertial method for convex optimization. *SIAM J. Sci. Comput.*, 31(4):A2558–A2584, 2016.

[22] S. Bonettini and M. Prato. New convergence results for the scaled gradient projection method. *Inv. Probl.*, 31(9):1196–1211, 2015.

[23] S. Bonettini, R. Zanella, and L. Zanni. A scaled gradient projection method for constrained image deblurring. *Inv. Probl.*, 25(1), 2009.

[24] S. Bonettini, R. Zanella, and L. Zanni. A scaled gradient projection method for constrained image deblurring. *Inv. Probl.*, 25(1), 2009.

[25] W. Buchberger and A. Niehoff et al. Clinically and mammographically occult breast lesions: detection and classification with high-resolution sonography. *Seminars in Ultrasound, CT and MRI*, 21(4):325–336, 2000.

[26] T. M. Buzug. *Computed Tomography, from photon statistics to modern cone-beam CT.* Springer, 2010.

[27] E. Candes and J. Romberg et al. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory*, 52:489–509, 2006.

[28] E. Candes and T. Tao et al. Near-optimal signal recovery from randon projections: universal encoding strategies? *IEEE Trans. Inf. Theory*, 52:5406–5425, 2006.

[29] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problmens with applications to imaging. *J. Math. Imaging Vision*, 40:120–145, 2011.

[30] G. H. Chen and J. Tang et al. Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projectiondata sets. *Med. Phys.*, 35:660–663, 2008.

[31] K. Choi, J. Wang, and al. Compressed sensing based cone-beam computed tomography reconstruction with a first-order method. *Med. Phys,*, 37(9):5113–5121, 2010.

[32] P. Clauser and A. De Nicolø' et al. Appearance of malignant breast lesions identifyed with Digital Breast Tomosynthesis. *EPOS*, 2013.

[33] V.L. Coli, E. Loli Piccolomini, E. Morotti, and L. Zanni. A fast gradient projection method for 3D image reconstruction from limited tomographic data. *Journal of Physics : Conference Series. IOP Publishing*, 2017.

[34] V.L. Coli, E. Loli Piccolomini, E. Morotti, and L. Zanni. Reconstruction of 3D X-ray CT images from reduced sampling by a scaled gradient projection algorithm. *Computational Optimization and Applications*, accepted for pubblication.

[35] V.L. Coli, V. Ruggiero, and L. Zanni. Scaled first-order methods for a class of large-scale constrained least squares problems. *Numerical Computations: Theory and Algorithms (NUMTA-2016)*, AIP Publishing:040002, 2016.

[36] D. di Serafino, V. Ruggiero, G. Toraldo, and L. Zanni. On the steplength selection in gradient methods for unconstrained optimization. *Appl. Math. Comput.*, 2017.

[37] D. Dobson and C.R. Vogel. Convergence of an iterative method for Total Variation denoising. *SIAM J. Numer. Anal.*, 34:1779–1791, 1997.

[38] D. L. Donoho. Compressed sensing. *IEEE Trans. Inf. Theory*, 52:1289–1306, 2006.

[39] Public Health England. Routine quality control tests for breast tomosynthesis (physicists). NHS Breast Screening Programme equipment report 1407. *Cancer Screening Programmes*, 2015.

[40] M. Ertas, I. Yildirim, M. Kamasak, and A. Akan. An iterative tomosynthesis reconstruction using total variation combined with non-local means filtering. *BioMedical Engeneering online*, 13(65), 2014.

[41] J. C. Ramirez-Giraldo et al. Nonconvex prior image constrained compressed sensing (NCPICCS): theory and simulations on perfusion ct. *Med. Phys.*, 38:2157–2167, 2011.

[42] J. T. Bushberg et al. *The essential physics of medical imaging.* Philadelphia, PA, USA, 2 edition, 2002.

[43] L. Feldkamp, L. Davis, and J. Kress. Practical cone-beam algorithm. *J. Opt. Soc. Am.*, 1:612–619, 1984.

[44] J. Fessler. Equivalence of pixel-driven and rotation-based backprojectors for tomographic image reconstruction. *online access*, 1997.

[45] G. Frassoldati, L. Zanni, and G. Zanghirati. New adaptive stepsize selections in gradient methods. *J. Ind. Manag. Optim.*, 4(2):299–312, 2008.

[46] P. Gilbert. Iterative methods for three-dimensional reconstruction of an object from projections. *J. Theor Biol.*, 36(1):105–117, 1972.

[47] L. Gilles, C. R. Vogel, and J. M. Bardsley. Computationa methods for a large-scale inverse problem arising in atmospheric optics. *Inv. Probl.*, 18(1), 2002.

[48] C. Gonzaga and R.M. Schneider. On the steepest descent algorithm for quadratic functions. *Comput. Optim. Appl.*, 63(2):523–542, 2016.

[49] R. Gordon, R. Bender, and G. T. Herman. Algebraic reconstruction techniques (ART) for 3-dimensional electron microscopy and X-ray photography. *J. Theor. Biol.*, 29(3):478–481, 2011.

[50] Christian G. Graff and Emil Y. Sidky. Compressive sensing in medical imaging. *Applied Optics*, 54(8):C23–C44, 2015.

[51] G. T. Gullberg and R. H. Huesmanand et al. An attenuated projector backprojector for iterative SPECT reconstruction. *Phys. Med. Biol.*, 30(8):799–816, 1985.

[52] G. Harauz and F. P. Ottensmeyer. Interpolation in computing forward projections in direct three-dimensional reconstruction. *Phys. Med. Biol.*, 28(12):1419–1427, 1983.

[53] U. Heil, S. Frankel, and al. Total Variation regularization in Digital Breast Tomosynthesis: Regularization parameter determination based on small structures segmentation rates. In *12th Fully 3D Image Reconstruction in Radiology and Nuclear Medicins*, pages 485–488, 2013.

[54] R. H. Huesmanand and G. T. Gullberg et al. RECLBL library users manual donner algorithms for reconstruction tomography. *online access*, 1977.

[55] S. Hughes. *CT Scanning in Archaeology, Computed Tomography - Special Applications.* InTech, 2011.

[56] J.T. Dobbins III and D. J. Godfrey. Digital X-ray tomosynthesis: current state of the art and clinical potential. *Physics in Medicine and Biology*, 48:R65–R106, 2003.

[57] T. L. Jensen, J. H. Jorgensen, P. C. Hansen, and S. H. Jensen. Implementation of an optimal first-order method for strongly convex total variation regularization. *BIT Numer Math*, 52:329–356, 2012.

[58] A. Jerebko, M. Kowarschik, and T. Mertelmeier. Regularization parameter selection in maximum a posteriori iterative reconstruction for digital breast tomosynthesis. *LNCS*, 6136:548–555, 2010.

[59] Park JM and Franken EA Jr et al. Breast Tomosynthesis: present considerations and future applications. *Radiographics*, 27:S231–40, 2007.

[60] J. H. Jørgensen, T. L. Jensen, P. C. Hansen, S. H. Jensen, E. Y. Sidky, and X. Pan. Accelerated gradient methods for Total-Variation-based CT image reconstruction. In *11th Fully 3D Image Reconstruction in Radiology and Nuclear Medicins*, pages 435–438, 2011.

[61] P. M. Joseph. An improved algorithm for reprojecting rays through pixel images. *IEEE Trans. Med. Imag.*, MI-1(3):192–196, 1982.

[62] S. Kaczmarz. Angenäherte Auflösung von systemen linearer gleichungen. *Bull. Internat Acad. Polon. Sci. Lettres A*, 35:355–357, 1937.

[63] A. C. Kak and M. Slaney. *Principles of computerized tomographic imaging.* IEEE Press, New York, 1988.

[64] D. Kim, S. Ramani, and J. A. Fessler. Accelerating X-ray CT ordered subsets image reconstruction with Nesterov's first-order methods. In *12th Fully 3D Image Reconstruction in Radiology and Nuclear Medicins*, pages 22–25, 2013.

[65] S. Kullback and R. A. Leibler. On information and sufficiency. *Ann. Math. Statist.*, 22(1):79–86, 1951.

[66] P. Lacroute and M. Levoy. Fast volume rendering using a shearwarp factorization of the viewing transformation. *Proc. SIGGRAPH*, pages 451–458, 1994.

[67] H. Lantéri, M. Roche, and C. Aime. Penalized maximum likelihood image restoration with positivity constraints: multiplicative algorithms. *Inv. Probl.*, 18(5):1397–1419, 2002.

[68] P. T. Lauzier and J. Tang et al. Prior image constrained compressive sensins: implementation and performance evaluation. *Med. Phys.*, 39:66–80, 2012.

[69] T. Li and X. Li et al. Simultaneous reduction of radiation dose and scatter for CBCT by using collimators. *Med. Phys.*, 40, 2013.

[70] E. Loli Piccolomini and E. Morotti. A fast TV–based iterative algorithm for digital breast tomosynthesis image reconstruction. *J. Algor. and Computat. Techn.*, 10(4):277–289, 2016.

[71] Y. Long and J. A. Fessler. 3D forward and back-projection for X-ray CT using separable footprints with trapezoid functions. In *Proceedings of The First International Conference on Image Formation in X-Ray Computed Tomography*, pages 216–219, 2010.

[72] Y. Long, J. A. Fessler, and J. M. Balter. 3D forward and backprojection for X-ray CT using separable footprints. *IEEE Trans. Med. Imag.*, 29(11):1839–1850, 2010.

[73] Y. Lu, H.P. Chan, and al. Selective-diffusion regularization for enhancement of microcalcifications in digital breast tomosynthesis reconstruction. *Medical Physics*, 37(11):6003–6014, 2010.

[74] M. Males, D. Mileta, and M. Grgic. Digital Breast Tomosynthesis: a technological reviews. In *53th International Symposium ELMAR*, pages 41–45, 2011.

[75] B. De Man and S. Basu. Distance-driven projection and backprojection. *Nuclear Science Symposium Conference Record IEEE*, 3:1477–1480, 2002.

[76] B. De Man and S. Basu. Distance-driven projection and backprojection in three dimensions. *Phys. Med. Biol.*, 7:2463–2475, 2004.

[77] S. Matej, J. A. Fessler, and I. G. Kazantsev. Iterative tomographic image reconstruction using Fourier-based forward and backprojectors. *IEEE Trans. Med. Imaging*, 23(4):401–412, 2004.

[78] J. L. Mueller and S. Siltanen. *Linear and Nonlinear Inverse Problems with practical applications.* SIAM, 2012.

[79] K. Mueller, R. Yagel, and J. J. Wheller. A fast and accurate projection algorithm for 3D cone-beam reconstruction with the algebraic reconstruction technique (ART). *Proc. SPIE*, 1999.

[80] J. Nocedal and S. J. Wright. *Numerical optimization.* Springer-Verlag New York, Inc., 1999.

[81] J. Nuyts, B. De Man, and J.A. Fessler et al. Modeling the physics in the iterative reconstruction for Transmission Computed Tomography. *Phys Med Biol.*, 58(12):R63–96, 2013.

[82] T. M. Peters. Algorithms or fast back and re-projection in computed tomography. *IEEE Trans. Nucl. Sci.*, 28(4):3641–3647, 1981.

[83] T. Pock and A. Chambolle. Diagonal preconditioning for first order primal-dual algorithms in convex optimization. *International Conference on Computer Vision*, pages 1726–1769, 2011.

[84] F. Porta, M. Prato, and L. Zanni. A new steplength selection for scaled gradient methods with application to image deblurring. *J. Sci. Comput.*, 65(d):895–919, 2015.

[85] F. Porta, M. Prato, and L. Zanni. A new steplength selection for scaled gradient methods with application to image deblurring. *J. Sci. Comp.*, 65:895–919, 2015.

[86] J. Radon. *Ü*ber die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten. *Berichte über die Verhandlungen der Königlich-Slächsischen Gesellschaft der Wissenschaften zu Leipzig. Mathematisch-Physische Klasse*, 69:262–277, 1917.

[87] S. Ramani and J. Fessler. A splitting-based iterative algorithm for accelerated statistical X-ray CT reconstruction. *IEE Trans. Med. Imaging*, 31:677–688, 2012.

[88] S. Rose, M. Andersen, E. Sidky, and X. Pan. Noise properties of CT images reconstructed by use of constrained Total-Variation, data-discrepancy minimization. *Med. Phys.*, 42(5):2690–2698, 2015.

[89] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear Total Variation based noise removal algorithms. *Phys. D*, 60:259–268, 1992.

[90] D. A. Scaduto and M. Goodsitt et al. WE-DE-207B-05: measuring spatial resolution in digital breast tomosynthesis: Update of AAPM Task Group 245. *Med. Phys.*, 43(6), 2016.

[91] R. L. Siddon. Fast calculation of the exact radiological path for a three-dimensional CT array. *Medical Physics*, 12(2):252–255, 1985.

[92] E. Y. Sidky, J. H. Jørgensen, and X. Pan. Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle-Pock algorithm. *Physics in Medicine and Biology*, 57(10):3065–3091, 2012.

[93] E. Y. Sidky, C. M. Kao, and X. Pan. Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT. *Journal of X-Ray Science and Technology*, 14(2):119–139, 2006.

[94] E. Y. Sidky and X. Pan. Image reconstruction in circular cone-beam Computed Tomography by constrained, Total-Variation minimization. *Physics in Medicine and Biology*, 53:4777–4807, 2008.

[95] E. Y. Sidky, X. Pan, I. S. Reiser, and R. M. Nishikawa. Enhanced imaging of microcalcifications in Digital Breast Tomosynthesis through improved image-reconstruction algorithms. *Medical Physics*, 36(11):4920–4932, 2009.

[96] E.Y. Sidky, J.H. Jørgensen, and X. Pan. First-order convex feasibility for X-ray CT. *Med. Phys.*, 40(3):3115–1–15, 2013.

[97] R. E. van Engen et al. Protocol for the Quality Conrtol of the physical and technical aspects of digital breast tomosynthesis systems. Version 1.01. 2016.

[98] Julia Velikina, Shuai Leng, and Guang-Hong Chen. Limited view angle tomographic image reconstruction via total variation minimization. In *3rd Workshop on High Perfonrmance Image Reconstruction*, volume 6510. Proc.SPIE, 2007.

[99] C. R. Vogel. Nonsmooth regularization. In H.W. Engl, A.K. Louis, and W. Rundell, editors, *Inverse Problems in Geophysical Applications.* SIAM, 1997.

[100] C. R. Vogel. *Computational Methods for Inverse Problems.* SIAM, Philadelphia, PA, USA, 2002.

[101] C. R. Vogel and M. E. Oman. Iterative methods for total variation denoising. *SIAM J. Sci. Comput.*, 17:227–238, 1996.

[102] C. R. Vogel and M. E. Oman. Fast, robust Total Variation–based reconstruction of noisy, blurred images. *IEEE Trans. on Image Proc.*, 7:813–824, 1998.

[103] J. Wu, X. Mou, and Y. Zhang. A fast iterative soft-thresholding algorithm for few-view ct reconstruction. In *3rd Workshop on High Perfonrmance Image Reconstruction*, pages 72–75, 2011.

[104] T. Wu and al. Tomographic mammography using a limited number of low-dose cone-beam projection images. *Medical Physics*, 30:365–380, 2003.

[105] T. Wu, R. H. Moore, and al. A comparison of reconstruction algorithms for breast tomosyntesis. *Medical Physics*, 31(9):2636–2647, 2004.

[106] H. Yu and G. Wang. A soft-threshold filtering approach for reconstruction from a limited number of projections. *Physics in Medicine and Biology*, 55:3905–3916, 2010.

[107] R. Zanella, P. Boccacci, L. Zanni, and M. Bertero. Efficient gradient projection methods for edge-preserving removals of Poisson noise. *Inv. Probl.*, 25(5):045010, 2009.

[108] Y. Zhang, H.P. Chan, and al. A comparative study of limited-angle cone-beam reconstruction methods for breast tomosynthesis. *Medical Physics*, 33(10):3781–3795, 2006.

[109] Y. Zhang-O'Connor and J. A. Fessler. Fourier-based forward and back-projectors in iterative fan-beam tomographic image reconstruction. *IEEE Trans. Med. Imaging*, 25(5):582–589, 2006.

[110] B. Zhou, L. Gao, and Y.H. Dai. Gradient methods with adaptive step-sizes. *Comput. Optim. Appl.*, 35(1):69–86, 2006.