École Normale supérieure Paris Saclay

Année ARPE

Département EEA

Analysis and optimization of ray-based 3D ultrasound transmission tomography

Author
Pierre-Antoine Comby

Supervisor
Torsten Hopp
Eric Vourch'

Version of September 9, 2020

https://ftp.comby.xyz/public/ipe/report.pdf

Karlsruher Institut für Technologie

Acknowledgements

Ce travail n'aurai pas été possible sans l'aide de nombreuse personnes, qui doivent êtres remerciées pour le soutien qu'elles m'ont apporté pendant ce stage :

Tout d'abord, je voudrais remercier Torsten Hopp pour m'avoir accompagné tout au long de l'année, que son bureau soit à la porte à coté ou à 650km du mien, il a toujours su apporter un regard critique sur mon travail et me soutenir dans mes pistes de recherches, même dans celles qui ne furent pas couronnées de succès.

Je désire aussi remercier Eric Vour'ch pour son suivi, son écoute et sa bienveillance tout au long de cette année.

Je remercie également Hartmut, Michael et Nicole d'avoir pris le temps de répondre à mes questions et d' en avoir – plus souvent que je ne peux l'avouer – posé une autre bien plus pertinente en retour.

Merci à Julia, Nima, Olga, Sara, Sarah, Roberto, Zewei et tous les membres de l'équipe USCT d'avoir été là pour les pauses café, les brunchs du mardi et toutes les réunion numériques pendant le confinement.

Je tiens particulièrement à remercier tous les membres de la KDStV Schwarzwald, qui m'ont accueillis dans une ville où je ne connaissais personne. Grâce à eux j'ai pu ponctuer cette année de moments riche de camaraderie, qu'ils soient autour d'une botte ou en haut d'une montagne du Tyrol.

Mes derniers remerciements vont à Marine, pour son soutien, sa gentillesse et sa patience durant toute cette année outre Rhin.

Abstracts

Résumé

Au sein du KIT, un nouveau tomographe à ultrasons (USCT) est en cours de développement pour le diagnostic précoce du cancer du sein. Contrairement aux méthodes actuelles telle que la mammographie, cet appareil reconstruit en 3D, n'expose pas la patiente à des rayonnements ioniques, et reste plus abordable que les scanners IRM.

Il existe deux méthodes de tomographie à ultrason pour reconstruire l'intérieur d'objets à partir de l'émission et réceptions d'ondes : En utilisant les informations sur la réflexion d'ondes dans l'objet (similaires à l'échographie classique), où l'on obtient une représentation qualitative de la réflectivité dans les tissus avec une résolution submillimétrique. Ou bien en utilisant les informations issues de la transmission à travers l'objet sondé, qui donnent accès à une représentation quantitative de l'atténuation et de la vitesse du son.

Ce rapport traite principalement de la tomographie par transmission, et plus particulièrement des méthodes de reconstruction d'images utilisées pour obtenir une distribution de la vitesse du son dans les tissus mous.

Le traitement du modèle direct est basé sur une tomographie à rayons, en considérant une fréquence infinie. Ce modèle est ensuite amélioré en considérant des fréquences finies et en introduisant des noyaux de Fréchet. Ce procédé peut s'appuyer sur des données *a priori* (manuelle, ou issues d'une reconstruction précédente). Il est donc nécessaire d'étudier comment résoudre le problème inverse associé, d'autant plus que ce dernier est mal posé, mal conditionné et creux. L'utilisation d'un algorithme de resolution basé sur l'échantillonnage compressif s'avère résoudre partiellement ce problème. Afin d'améliorer et d'accélérer la reconstruction, une réduction automatique de la résolution à une région d'intérêt est également introduite.

Ces différents processus sont évalués dans un cadre de simulations numériques en 2D, avec l'objectif futur de les déployer en 3D puis en essais cliniques. Toutes ces méthodes sont évaluées quantitativement avec des données d'entrée plus ou moins restrictives.

Abstract

Within the KIT, a new ultrasound computer tomograph (USCT) is being developed for the early diagnosis of breast cancer. Unlike current methods such as mammography, this device reconstructs in 3D and does not expose the patient to ionized radiation while remaining more affordable than MRI scanners.

There are two ultrasound tomography methods used to reconstruct images from emitted and received ultrasound wave: By using reflection information (similar to conventional sonography) which gives a qualitative representation of the reflectivity in tissue. Or by using transmission information which gives a quantitative representation of attenuation and speed of sound in tissue.

This report deals mainly with transmission tomography, and more specifically with image reconstruction methods used to obtain a distribution of speed of sound in soft tissues.

The forward model processing, is developed on a ray-based, infinite frequency tomography hypothesis. It is then extended to finite frequency tomography with the introduction of Fréchet kernels. This process can be based on *a priori* data (manually provided or from a previous reconstruction). It is thus needed to study how to solve the associated inverse problem, especially as it is ill-posed, badly conditioned and sparse. The use of a compressive sampling based solver proves to partially address this issue. The introduction of an automated restriction to a region of interest both improves and accelerates the reconstruction.

These different processes are evaluated in a numerical 2D simulation framework, with the aim of deploying them in 3D and clinical trials. The developed methods are quantitatively evaluated with more and less restrictive input data.

Zusammenfassung

Innerhalb des KIT wird ein neuer Ultraschall-Computertomograph (USCT) für die Früherkennung von Brustkrebs entwickelt. Im Gegensatz zu derzeitigen Methoden wie der Mammographie rekonstruiert dieses Gerät in 3D und setzt die Patientin keiner ionisierten Strahlung aus, ist aber dennoch kostengünstiger als MRT-Scanner.

Es gibt zwei Ultraschalltomographieverfahren zur Rekonstruktion von Bildern aus Ultraschall-Rohdaten: Durch die Verwendung von Reflexionsinformationen (ähnlich wie bei der konventionellen Sonographie), die eine qualitative Darstellung des Reflexionsvermögens im Gewebe mit Sub-Millimeter-Auflösung liefern. Oder durch die Verwendung von Transmissionsinformationen, die eine quantitative Darstellung der Schalldämpfung und Schallgeschwindigkeit im Gewebe liefern.

Dieser Bericht befasst sich hauptsächlich mit der Transmissionstomographie und insbesondere mit Bildrekonstruktionsmethoden, mit denen eine Verteilung der Schallgeschwindigkeit in den Weichteilgeweben erzielt werden können.

Die Vorwärtsmodellverarbeitung, die Flugzeitdaten liefert, wird um eine strahlenbasierte, unendliche Frequenztomographie herum entwickelt. Sie wird dann mit der Einführung von Fréchet-Kernel auf die endliche Frequenztomographie erweitert. Dieser Prozess kann auf *a priori* Daten basieren (manuell bereitgestellt oder aus einer früheren Rekonstruktion). Es ist daher notwendig zu untersuchen, wie das verbundene inverse Problem gelöst werden kann, zumal es schlecht gestellt, schlecht konditioniert und spärlich ist. Die Verwendung eines auf kompressiven Stichproben basierenden Lösers erweist sich als eine teilweise Lösung dieses Problems. Im Interesse der Verbesserung und Beschleunigung der Rekonstruktion wird auch eine automatische Reduktion auf eine Region von Interesse eingeführt.

Diese verschiedenen Prozesse werden in einem numerischen 2D-Simulationsrahmen evaluiert, mit dem Ziel, sie in Zukunft in 3D für klinische Studien einzusetzen. Die entwickelten Methoden werden mit mehr und weniger restriktiven Eingabedaten quantitativ ausgewertet.

Contents

List of Figures

Chapter 1

Introduction and objectives

1.1 Motivation

Breast cancer is the most diagnosed cancer for women. It represents 24% of new diagnosed cancers and around 15% of the death of women due to cancer. Around 60% of the deadly cases take place in Occident[1, 2].

In opposition to tumours in other organs, which have already in their early stages impact on patients, breast tumours can stay hidden for a long time, until they build metastases. Consequently, most patients does not die of the breast cancer itself, but rather from the developed metastases and their consequences. The average size of a tumour at its detection is currently 1 cm wide, and the probability of having developed metastases at this stage is around 30%[3]. Those data makes it clear that detecting tumours as early as possible make a huge improvement on the chance of survival. Current devices to detect small tumours are either very expensive (MRI) or use X-ray mammography, which can expose patients to high and repeated ionizing radiations and their consequences. Developing methods to establish a breast cancer diagnosis with same or higher resolution is therefore very interesting. Among the different methods currently in development (low dose computed Tomography CT, photo acoustic devices. . .), ultrasound tomography in 2D or 3D is a serious candidate for this application.

In the Institute for Data Processing and Electronics (IPE) at the Karlsruhe Institute for Technology (KIT) a complete 3D Ultrasound Computer Tomography (3D-USCT) is being developed since 2000. Though 2D systems are also under development around the world, the 3D USCT is the only device that has the capabilities to probe completely the breast, with high isotropic resolution and reproducible 3D volume, while keeping a fast acquisition time. Currently, a new, improved third generation of 3D-USCT is under development and a commercial application is under investigation. Meanwhile, the second generation of the 3D-USCT prototype is also under test in a clinical study at the University hospital Mannheim.

The goal of the 3D-USCT is to be able to detect tumour smaller (and therefore sooner) than the current state of the art. Though other devices can also detect with high resolution and in 3D such small tumours, they are either very expensive (MRT) or use ionizing X-ray (CT). On the other hand, ultrasound tomography is harmless for the human body, is cheaper and have a shorter examination time than MRI. This makes the 3D-USCT a good candidate for performing regular diagnosis. The newest version of the USCT will be able to detect tumour with a resolution of 5 mm. At this stage the probability of development of metastases is only around 5%.[3]

1.2 Principle of Ultrasound Tomography

The groundwork provided by Greenleaf [4] has opened the field of medical transmission ultrasound tomography. The principle of ultrasound tomography can be outlined as follows:

A defined ultrasound wave is emitted by transducers, goes through the measured object and is then received by similar transducers. As the wave goes in the object, its character-

istics (amplitude and phase) are modified. Those modifications carry information about the inside of this object (speed of sound, attenuation, etc.), In the case of probing a breast, this can be used to detect the presence of tumorous cells, as presented in Figure 1.1.

1.2.1 Proof of Concept

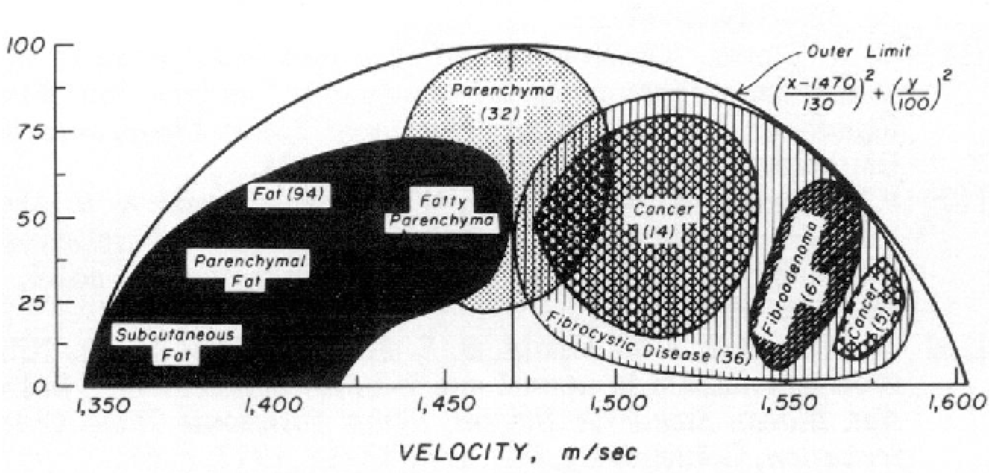


Figure 1.1: Relationship between ultrasonic speed and relative attenuation for several tissues of the breast. Connective tissue and some cancers exhibit high speed and low attenuation and have been found in breasts of young women. Medium-speed cancers with high attenuation have been found in older women.[4]

Ultrasound provides multiple levels of information (unlike X-ray) on the object for imaging: the attenuation of the wave's sound pressure indicates on the object's attenuation coefficient, the time-of-flight of the wave gives speed of sound information, and the scattered wave indicates on the echogenicity of the object (*e.g.* refraction index, surface morphology, etc.).

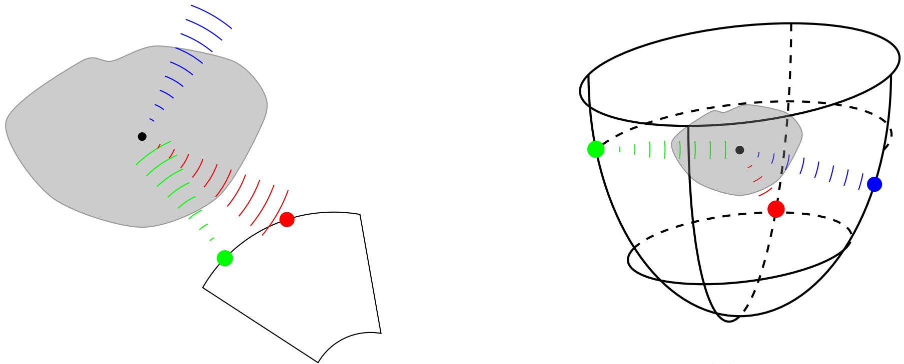
The choice of using Ultrasound frequency (over 20kHz) is driven by two opposite phenomena: the penetration thickness and the resolution power. The penetration thickness decreases with higher frequency, reducing the efficacy of the device to detect deep tumours, and the resolution factor is driven by the wavelength inversely proportional to the emitted frequency: the higher the frequency, the smaller are the detected tumours. As the emitted energy is limited by technical and medical considerations, the choice of an average frequency of 2.5 MHz has been made. More details about the emitted waves characteristics are available in the literature of the project [5, 6].

Unlike conventional ultrasound sonography, which uses phased array technology for beam forming and reflection tomography, the 3D-USCT is based on unfocused spherical waves for imaging, taking a broader picture of the measured media. However, the 3-USCT relies on intensive processing of the acquired data to produce images. The continuous rise of computing power and data storage have made such device now possible.

1.2.2 The 3D-USCT project

Several Ultrasound Computer Tomographs are in development in the world [7, 8, 9], but all of them are 2D or stacks of 2D (2.5D) acquisition and reconstruction devices. This does not able to probe the reflection out of the plane of detection like on Figure 1.2.

The 3D-USCT of IPE is to my knowledge, the only USCT able to probe such out of the plane waves, and have an isotropic resolution.



(a) 2D

(b) 3D

Figure 1.2: Comparison between 2D and 3D tomograph: in the 2D (a) case after passing a scattering object we only get access to in-plane and reflected signals, in a 3D USCT (b) we can catch every response signal, in and out of plan.

1.2.3 General Processing

The 3D-USCT developed at the IPE in KIT is a complex project, where design and financial constraints as well as medical norms defined strong boundaries on the field of possibilities, setting also challenges for the signal processing and image reconstruction, from the hardware and the data storage to the speed of reconstruction. Numerous process are involved, to acquire A-scans (Acquisition of the Amplitude of a wave), process them and use the result data to reconstruct images. These main steps are represented on Figure 1.3. Although the subject of my work is mainly the reconstruction of transmission tomography images, having a global view over the previous signal processing steps will help understanding and improving the reconstructions step, especially for determining uncertainty causes and noise filtering treatment we could apply.

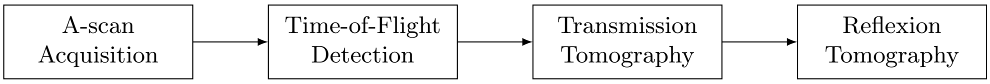


Figure 1.3: General Processing of US Tomography

1.3 Presentation of the IPE

The Institute for Data Processing and Electronics (IPE) is specialized in the development of custom detector, trigger and data acquisition systems for high data rates and in control and monitor systems.

IPE's competences cover the entire signal chain, starting with the physical sensor design, detector assembly through the analogical and digital electronics to the data analysis and archiving. An Electronic Packaging Laboratory is attached to the institute, where the production process is optimized and the detectors and electronic assemblies are cost-efficiently produced.

The USCT project mobilizes the core competences of the IPE with deep interaction in-between. Designing sensors and fast acquisition board, reconstruction algorithms, hardware integration, are key elements driving the project as well as the institute. The diversity of the interacting fields with the USCT is represented on Figure 1.4.

1.4 Objectives

During this year of research at the IPE institute, I will be focusing my work on the medical image reconstruction from simulated data as well as experimental data. Furthermore, my goals will be to analyse and optimize transmission tomography reconstruction algorithms in a simulated framework. Most of my work will consist of 2D reconstruction as the 3D reconstruction algorithms can be deduced from the 2D implementation. Restraining most of the experiment in 2D give us access to faster reconstruction time, and also easier prototyping of new reconstruction methods. The reconstruction will be implemented in MATLAB ®, but also using C function calls for time-critical operations.

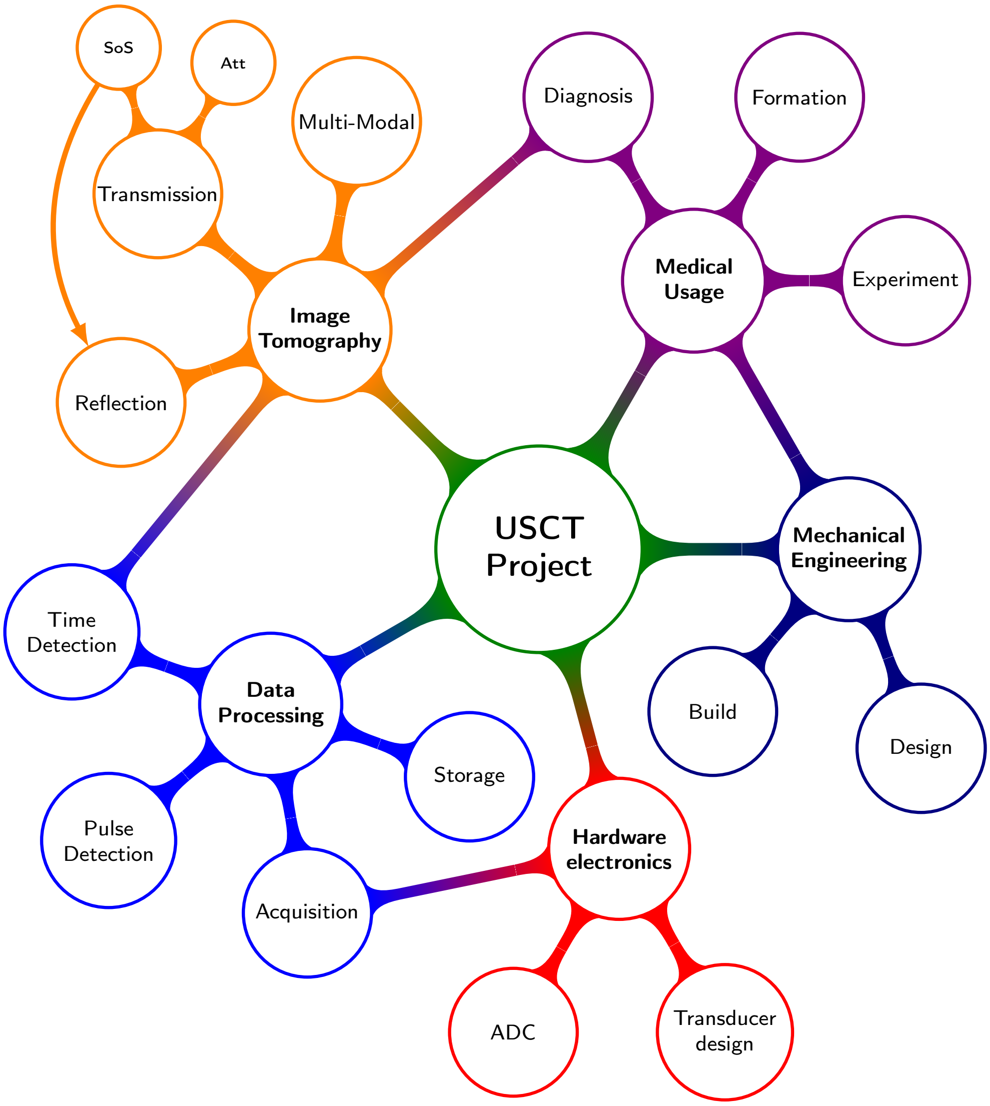


Figure 1.4: Work domains of the USCT Project, my work focuses on the image transmission tomography, but extended knowledge about the data processing is also needed.

Chapter 2

Ultrasound tomography theory

2.1 General description

2.1.1 3D-USCT of IPE

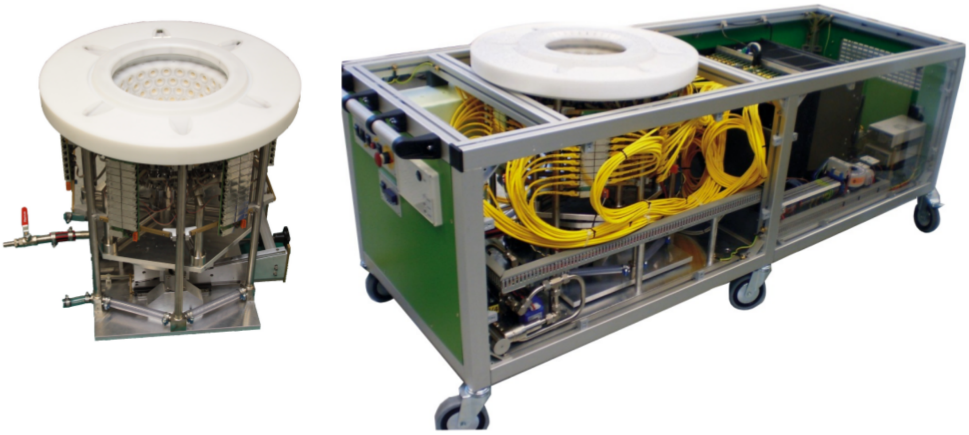


Figure 2.1: Aperture of `3D-USCT-II` (left) and complete view of the system (right).

Currently, a usable prototype (`3D-USCT-II`) is being tested in Mannheim Hospital. This second prototype has benefited from an improved and optimal aperture shape [10], increasing the resolution and contrast in a Region of interest (ROI) and also from a new ordering of the transducers arrays (TAS). This new aperture is build out of Polyoxymethylen in a half ellipsoid form, and contain 157 TAS. Each TAS consists of 4 emitters and 9 receivers. Furthermore, the aperture system can rotate and move up and down, in 47 different positions, giving access to 325 701 681 acquisition points. The data acquisition is done by internally developed ADC cards, which conception and operation are outside the scope of this report. More information can be found in the previous work about USCT[11].

A new prototype, `3D-USCT-III` is currently under development, using the feedback of the medical studies. This new version have a bigger aperture cup, and increases the size of the region of interest. The goal of this version is to make the product ready for commercialization and while improving the quality of the measurements, also respecting the norms established to make such device a medical standard. Some numerical characteristics are gathered in Table 2.1.

Characteristics	`3D-USCT-II` Values
Aperture opening diameter	26 cm
Aperture Volume size	4.1 L
Average Frequency	2.5 MHz
Sample Frequency	20 MHz
Bandwidth	1 MHz
Acquisition time of one A-Scan	300 µs
Number of Emitter	$157 \times 4 = 628$
Number of Receiver	$157 \times 9 = 1413$
Number of Aperture Position	47
Max number of A-Scans	325 701 681
Raw Data Size	5-80 GB
Acquisition time for one position	10 s
Acquisition time for n position	$10n + (n-1)41$s

Table 2.1: Numerical description of the USCT II.

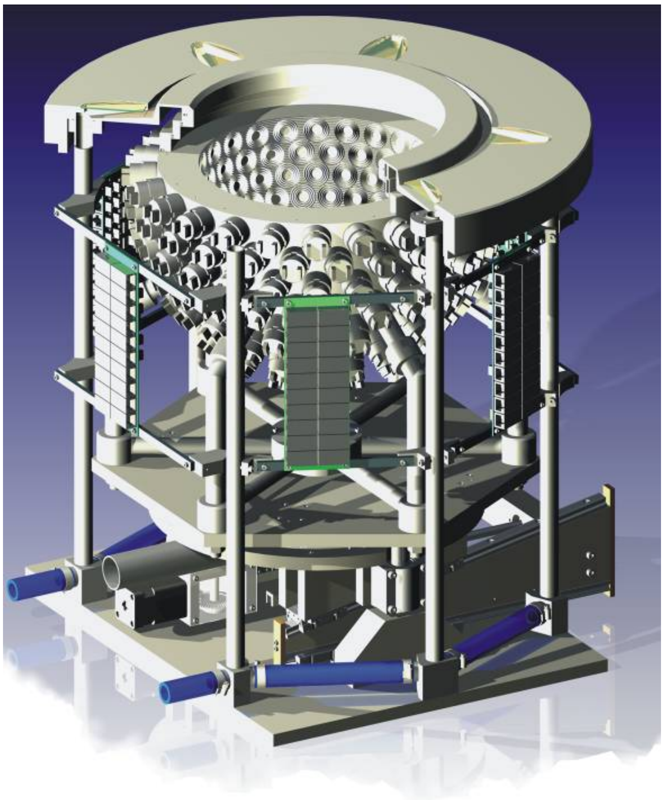
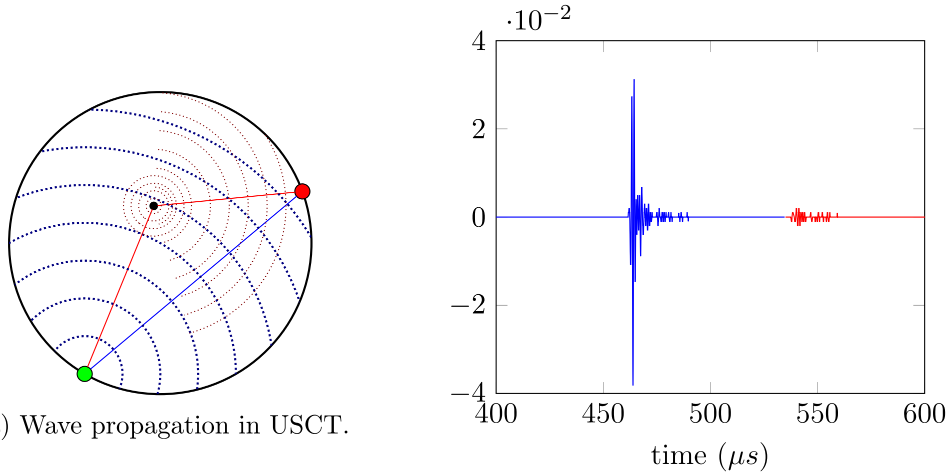


Figure 2.2: Detailed view of the aperture system.



(a) Wave propagation in USCT.

(b) A-Scan acquisition at Receiver.

Figure 2.3: 2D Schematic of the data Acquisition. The received pulse can be seperated into a transmitted one (blue), and a received one (red), representing the single scattering wave field

2.1.2 Physical Principles

The general procedure of the 3D-USCT can be described as follows: An Emitter sends an ultrasonic spherical wave which is then diffused inside the water-filled aperture, represented on Figure 2.2. Inside the aperture the wave will be refracted and/or reflected as it go through a heterogeneous medium. For each receiver, we acquire an Amplitude-scan (A-Scan), representing a pressure-over-time signal at the receiver. As shown on Figure 2.3b, this A-scan can be separated in a transmission pulse and a reflection pulse. The transmission pulse is associated to the fastest path between the emitter and receiver considered. This part of the signal will then be used for the transmission tomography reconstruction, which analysis will give access to information about the speed of sound and the attenuation factor inside the tissues.

What remains from the signal is considered as the reflection pulse, which arrive later at the receiver. The reflection tomography give access to the structural nature of an object, and show the reflective property of the medium (*i.e.* gives qualitative information about the heterogeneity of the medium). These methods are not totally independent, as the transmission tomography can be used to improve the reflection tomography reconstruction by providing a distribution of the speed of sound in the object.

This chapter mainly reviews the physical model developped for the 3D-USCT, based on acoustic and wave theory. The main characteristics of the acquired data is also presented.

2.2 Physical Model

Ultrasound waves with a frequency of 2.5 MHz are used to probe a medium placed in a water-filled aperture. Water is used as transitional medium for the aperture, having an impedance value closer to the breast than simple air. It also suits our medical application goal, as water does not have bad interaction with the human body.

2.2.1 Acoustics Equation

An acoustic wave can be described with 3 basics equations resulting of a mass conservation, dynamics and thermodynamics. From the mass conservation equation we have:

$$\frac{\partial \rho(\boldsymbol{x})}{\partial t} + \nabla \cdot (\rho(\boldsymbol{x})\boldsymbol{v}(\boldsymbol{x})) = \frac{\mathrm{d}\rho(\boldsymbol{x})}{\mathrm{d}t} + \rho(\boldsymbol{x})\nabla \cdot \boldsymbol{v} = 0 \tag{2.1}$$

With $\frac{\mathrm{d}f}{\mathrm{d}t} = \frac{\partial f}{\partial t} + \boldsymbol{v}.\boldsymbol{\nabla} f$ the total derivative of f.

And from the Euler Equation(obtained from Newton 2nd law) we also have:

$$\rho(\boldsymbol{x})\frac{\mathrm{d}\boldsymbol{v}(\boldsymbol{x})}{\mathrm{d}t} + \boldsymbol{\nabla} P(\boldsymbol{x}) = \boldsymbol{0} \tag{2.2}$$

With:

- $P(\boldsymbol{x})$ the scalar pressure field.
- $\boldsymbol{v}(\boldsymbol{x})$ the speed of the particle.
- $\rho(\boldsymbol{x})$ the scalar density field.

Furthermore, using 2nd law of thermodynamics we can establish the following identity:

$$\mathrm{d}S = \frac{c_v}{TP\beta}\left[\mathrm{d}P - \frac{1}{\chi_S}\mathrm{d}\rho\right] \tag{2.3}$$

With S the entropy, c_v the heat capacity at constant volume, $\beta = \frac{1}{P}\left(\frac{\partial P}{\partial T}\right)_v$ and χ_s the adiabatic compressibility coefficient. By assuming an adiabatic transformation (no dissipation effect like viscosity or thermal transfers) we have the simple relation:

$$\mathrm{d}P = \frac{1}{\rho(\boldsymbol{x})\chi_S}\mathrm{d}\rho = c(\boldsymbol{x})^2\mathrm{d}\rho \tag{2.4}$$

2.2.2 Wave Equation

In the following of this chapter to simplify he notation, the space dependency of vector and scalar field will be implied (*e.g.* $\boldsymbol{v}(\boldsymbol{x}) = \boldsymbol{v}$)

Model and Hypothesis

By applying the divergence operator on (2.2) and time derivative on (2.3) we get:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{\rho}\frac{\mathrm{d}\rho}{\mathrm{d}t}\right) + \frac{\mathrm{d}}{\mathrm{d}t}\nabla\boldsymbol{v} = 0 \tag{2.5}$$

$$\nabla\frac{\mathrm{d}\boldsymbol{v}}{\mathrm{d}t} + \nabla\left(\frac{1}{\rho}\boldsymbol{\nabla} P\right) = 0 \tag{2.6}$$

the terms of speed are equal and by using (2.4) and assuming $\frac{\mathrm{d}\chi_s}{\mathrm{d}t} = 0$,

$$\begin{aligned}
&\nabla\left(\frac{1}{\rho}\boldsymbol{\nabla}P\right)-\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{\rho}\frac{\mathrm{d}\rho}{\mathrm{d}t}\right)=0\\
\Longleftrightarrow\ &\nabla^2P\frac{1}{\rho}+\boldsymbol{\nabla}P\cdot\boldsymbol{\nabla}\rho\frac{1}{\rho^2}-\frac{\mathrm{d}}{\mathrm{d}t}\left(\chi_S\frac{\mathrm{d}P}{\mathrm{d}t}\right)=0\\
\Longleftrightarrow\ &\nabla^2P+\boldsymbol{\nabla}P\cdot\boldsymbol{\nabla}\rho\frac{1}{\rho}-\rho\chi_S\frac{\mathrm{d}^2P}{\mathrm{d}t^2}=0
\end{aligned}$$

With $c^2(\boldsymbol{x})=\frac{1}{\rho(\boldsymbol{x})\chi_s}$ we have finally the acoustic wave equation:

$$\boxed{\nabla^2P+\boldsymbol{\nabla}P\cdot\boldsymbol{\nabla}\rho\frac{1}{\rho}-\frac{1}{c^2}\frac{\mathrm{d}^2P}{\mathrm{d}t^2}=0} \tag{2.7}$$

if the density ρ is considered constant, equation (2.7) simplifies into the Helmholtz equation:

$$\nabla^2P-\frac{1}{c^2}\frac{\mathrm{d}^2P}{\mathrm{d}t^2}=0 \tag{2.8}$$

Numerical values for speed of sound, density, compressibility and impedance have been measured[12], and reported on Table 2.2. From these results, we can see that in soft tissues like breast, the variance of the density is less than half of the one of compressibility. Greenleaf [4] also reveals that it is not the density itself but its variation that have an impact on the image reconstruction, we may assume that these variations are small, and neglect them. For the rest of the development, the density will therefore be assume constant. *However the speed of sound distribution and the attenuation remain variable in space.*

Tissues	$c/\mathrm{m\,s^{-1}}$	$\rho/\mathrm{kg\,m^{-3}}$	$\chi_s/10^{-10}\mathrm{Pa}^{-1}$	$Z/10^6\,\mathrm{kg\,m^{-2}s^{-1}}$
Water (37 °C)	1483	993	4.0	1.47
Water (20 °C)	1524	993	4.3	1.51
Fat	1420	950	5.2	1.39
Cancer	1580	1100	3.6	1.73
Muscle	1568	1040	3.9	1.63
$\sigma/$ %	4.3	5.6	15	9.8

Table 2.2: Acoustic characteristics of different media: speed of sound c, density ρ, compressibility χ_S, acoustic impedance Z.

We will stay in the Fourier space for the rest of this development, and consider if not stated otherwise a monochromatic pointwise source wave.

$$\nabla^2P+k^2P=0 \tag{2.9}$$

where $k=\frac{2\pi}{\lambda(\boldsymbol{x})}=\frac{\omega}{c(\boldsymbol{x})}$ is the wave number.

Solution of the wave equation

In the case of a heterogeneous medium (in terms of speed of sound), one can use a reference wave number k_0 as follows:

$$k^2 = k_0^2 \frac{c_0^2}{c^2(\boldsymbol{x})} = k_0^2 n^2(\boldsymbol{x}) = k_0^2 + \underbrace{k_0^2(n^2(\boldsymbol{x}) - 1)}_{f(\boldsymbol{x})} \tag{2.10}$$

Where $n(\boldsymbol{x})$ is the refractive index of the medium and c_0 a reference speed of sound (like the one of a water-only filled aperture). The left-hand side of (2.9) can then be expressed as a homogeneous Helmoltz equation:

$$\nabla^2 P + k_0^2 P = -f(\boldsymbol{x})P \tag{2.11}$$

With $n = n_0 + \delta n$ (and $n_0 = 1$) the refractive index of the water background we can express $n(\boldsymbol{x})$ as:

$$\delta n = n - n_0 = \frac{c_0}{c_0 + \delta c} - 1 \simeq -\frac{\delta c}{c_0} \tag{2.12}$$

This separation can be also applied to the pressure field, setting $P = P_i + P_s$, adding the incident field P_i to the scattered field P_S, see Figure 2.4. For instance one could get the incident field by making measurement in a homogeneous background medium, but only the scattered field will be relevant as it contains information about the local variation in the aperture.

According to [12], the solution to (2.11) can then be determined using the Green-function G, assuming a boundary condition like the Sommerfeld radiation condition.

$$G(\boldsymbol{x}, \boldsymbol{y}) = \frac{e^{ik_0(\|\boldsymbol{x}-\boldsymbol{y}\|)}}{4\pi\|\boldsymbol{x} - \boldsymbol{y}\|} \tag{2.13}$$

The Green function is the impulse response for an inhomogeneous linear (partial) differential equation in our situation we can write:

$$\left(\nabla^2 + k_0^2\right) G(\boldsymbol{x}, \boldsymbol{y}) = \delta(\boldsymbol{x} - \boldsymbol{y}) \tag{2.14}$$

therefore, with a source point at $\boldsymbol{x_e}$ with amplitude P_0, the incident pressure field at $\boldsymbol{y}$ is defined by:

$$P_i(\boldsymbol{y}, \boldsymbol{x_e}) = P_0 G(\boldsymbol{y}, \boldsymbol{x_e}) \tag{2.15}$$

From the linearity of our problem and using the property of Dirac's distribution

$$f(\boldsymbol{x}) = \int \delta(\boldsymbol{x} - \boldsymbol{y}) f(\boldsymbol{y}) \mathrm{d}\boldsymbol{y}$$

we can determine the scattered field as a weighted integral of our problem:

$$\boxed{P_s(\boldsymbol{x}) = \int G(\boldsymbol{x}, \boldsymbol{y}) f(\boldsymbol{y}) P(\boldsymbol{y}) \mathrm{d}\boldsymbol{y}} \tag{2.16}$$

and setting f to zero outside our reconstruction area.

This equation can not alone be a solution to the problem, as P_s is implicitly on both sides of the equation. Some approximations are still to be made to solve this equation.

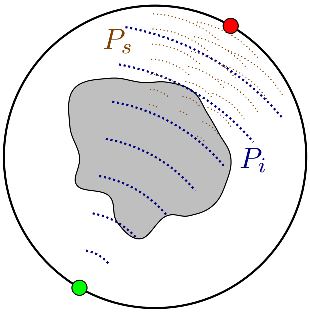


Figure 2.4: Field superposition: The incident Field P_i meet the heterogeneous medium, resulting in multiple and complex scattering yielding the scattered field P_s. This decomposition is the starting point of the Born Approximation. Only the scattered field contains information about the heterogeneity of the medium.

2.3 Acoustic approximations

Among the acoustic approximations, the most used are the *Born* and *Rytov* approximations. Reconstruction methods using these approximations are part of so-called *Diffraction Tomography*. Using the transmitted (resp.) reflected signal with this method would be *transmission tomography* (resp.) *reflection tomography*.

If none of these approximations are made, and one use raw computation, the method would be a *Full-Wave tomography*, but such methods are harder to compute, and are a subject of an on-going PhD at IPE.

2.3.1 Born approximation

The hypothesis of separating the incident from the scattered field presented on Figure 2.4, transforms (2.16) into:

$$P_s(\boldsymbol{x}, \boldsymbol{x_e}) = \int G(\boldsymbol{x}, \boldsymbol{y}) f(\boldsymbol{y}) \left(P_i(\boldsymbol{y}) + P_s(\boldsymbol{y}) \right) \mathrm{d}\boldsymbol{y} \tag{2.17}$$

Using the *Born Approximation* consist of neglecting the scattered field (considered smaller) in respect to the incident field in the right-hand side of (2.17). This also neglects multiple scattering waves, and we get:

$$\boxed{P_{Born}(\boldsymbol{x}, \boldsymbol{x_e}) = \int G(\boldsymbol{x}, \boldsymbol{y}) f(\boldsymbol{y}) P_i(\boldsymbol{y}) \mathrm{d}\boldsymbol{y} = \int G(\boldsymbol{x}, \boldsymbol{y}) f(\boldsymbol{y}) G(\boldsymbol{y}, \boldsymbol{x_e}) P_0 \mathrm{d}\boldsymbol{y}} \tag{2.18}$$

With this approximation the computation of the pressure field can be done, with taking a particular care of respecting the Shannon-Nyquist theorem for time and space sampling. Equation (2.18) can also used to build an iterative scheme, which in case of convergence, will lead theoretically to the true pressure field. However, this system is not converging in most cases, or with high requirements[13].

The Born approximation implies that the phase shift between scattered and incident field should be lower than π [14]. At our frequency this sets a maximal object size of 9.6 mm. In the 3D-USCT an object (*e.g.* breast) has a size of 10 to 20 cm, at this scale we are far outside the field of the Born-approximation.

2.3.2 Rytov approximation

In the case of the Rytov approximation, the phase of the total pressure field is separated between an incident φ_i and scattered phase φ_s.

$$P(\boldsymbol{x}) = P_0 e^{j\varphi(\boldsymbol{x})} = P_0 e^{j\varphi_i(\boldsymbol{x})+j\varphi_s(\boldsymbol{x})} = P_i e^{j\varphi_s(\boldsymbol{x})} \tag{2.19}$$

applying (2.9):

$$\nabla^2\varphi + |\boldsymbol{\nabla}\varphi|^2 + k^2 = 0 \tag{2.20}$$

And by using the phase decomposition:

$$\nabla^2(\varphi_i + \varphi_s) + |\boldsymbol{\nabla}\varphi_i|^2 + 2\boldsymbol{\nabla}\varphi_i \cdot \boldsymbol{\nabla}\varphi_s + |\boldsymbol{\nabla}\varphi_s|^2 + k_0^2 n^2 = 0 \tag{2.21}$$

The incident field cancels itself as a solution to (2.20).

$$\nabla^2\varphi_s + 2\boldsymbol{\nabla}\varphi_i \cdot \boldsymbol{\nabla}\varphi_s + |\boldsymbol{\nabla}\varphi_s|^2 + k_0^2\left(2\delta n + (\delta n)^2\right) = 0 \tag{2.22}$$

The Rytov approximation involves discarding the second order terms. In the remaining we apply the substitution: $\varphi_s = h_s(\boldsymbol{x})\exp(-\varphi_i)$. We get:

$$\left(\nabla^2 + k_0^2\right) h_s = -2k_0^2\delta n \tag{2.23}$$

This is an inhomogeneous Helmoltz equation, and we can then similarly use the Green function, with $f_r \simeq 2k_0^2\delta n(\boldsymbol{x})$ and get:

$$h_s(\boldsymbol{x}, \boldsymbol{x_e}) = \int G(\boldsymbol{x}, \boldsymbol{y}) f_r(\boldsymbol{y}) P_i(\boldsymbol{y}, \boldsymbol{x_e}) \mathrm{d}\boldsymbol{y} = P_{Born} \tag{2.24}$$

Finally, we get the *Rytov Approximation*

$$\boxed{P_{Rytov} = P_i \exp\left(\frac{P_{Born}}{P_i}\right)} \tag{2.25}$$

A condition for the validity of the Rytov Approximatio can be established [14] as:

$$k\delta n \gg |\boldsymbol{\nabla}\varphi_s|^2 \tag{2.26}$$

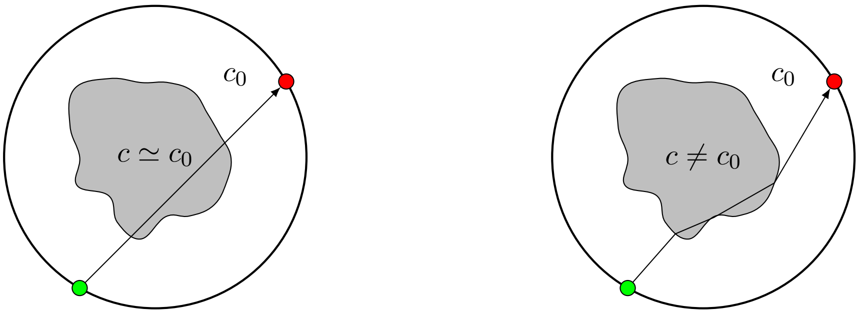
Equation (2.26) has been empirically verify for the 3D-USCT. Unlike the Born approximation the phase shift can be more important, making the Rytov approximation suitable for longer propagation distances. More results and comparison between the Born and Rytov Approximation are available in [14].

2.3.3 Eikonal Ray Approximation

The eikonal approximation is more restrictive than the Born or Rytov approximations, as it also assumes infinite frequency, and neglects scattering. This has as direct consequence that only the refraction of the wave will be considered. In this way, the wave propagation can be considered as a curve or a straight ray, taking the fastest path between an emitter and a receiver, like on Figure 2.5.

Like in the Rytov Approximation, for a plane wave, we linearize the phase, taking: $\varphi(\boldsymbol{x}) = jk_0\boldsymbol{r}\boldsymbol{x}$. And we have:

$$\nabla^2\varphi + |\boldsymbol{\nabla}\varphi|^2 + k^2 = 0 \tag{2.27}$$



(a) Straight rays

(b) Bent rays

Figure 2.5: Example of ray approximation through the Eikonal equation. straight rays (a) are considering a homogeneous medium, bent rays (b) are the consequences of refraction in a heterogeneous medium

taking $P = P_i e^{jk_0\varphi}$ yields:

$$ik_0\nabla^2\varphi - k_0^2|\boldsymbol{\nabla}\varphi|^2 + k_0^2 n^2 = 0 \tag{2.28}$$

By assuming a infinite frequency we have $k_0 \to \infty$ and (2.28) becomes the *Eikonal Equation*[1]:

$$\boxed{|\boldsymbol{\nabla}\varphi|^2 = n^2(\boldsymbol{x})} \tag{2.29}$$

Equation (2.29) can be adapted to a more suitable form: for a given wave front ($\varphi(x)$ is fixed) a *ray* is defined as a field line of $\boldsymbol{\nabla}\varphi$. Any point on a ray $\mathcal{R}$ can be indexed by a curvilinear coordinate l. For a small element $\mathrm{d}l$ of $\mathcal{R}$ and $\boldsymbol{\tau}$ the unit vector tangent to this element:

$$\boldsymbol{\nabla}\varphi = n\boldsymbol{\tau}$$

Using Einstein's notation, the variation of $\boldsymbol{\nabla}\varphi$ along the ray is:

$$\begin{aligned}
\frac{\partial\boldsymbol{\nabla}\varphi}{\partial l} &= \frac{\partial\varphi_{,i}}{\partial l} = \varphi_{,ij}\frac{\partial x_j}{\partial l} = \varphi_{,ij}\frac{1}{n}\varphi_{,j} \\
&= \varphi_{,ji}\frac{1}{n}\varphi_{,j} = \frac{1}{2n}(\varphi_{,j})^2_{,i} \\
&= \frac{1}{2n}n^2_{,i} \\
\frac{\partial\boldsymbol{\nabla}\varphi}{\partial l} &= n_{,i} = \boldsymbol{\nabla}n
\end{aligned} \tag{2.30}$$

Thus in the case of a homogeneous medium , $\boldsymbol{\nabla}n = 0$ and $\boldsymbol{\nabla}\phi = n\boldsymbol{\tau}$: the ray is a straight line. If not (2.30) can be expressed as the Euler-Lagrange's equation:

$$\boxed{\frac{\partial n\boldsymbol{\tau}}{\partial l} = \boldsymbol{\nabla}n} \tag{2.31}$$

---

[1] The Eikonal equation was derived about 150 years ago by Sir William Rowan Hamilton. The word *Eikonal* was introduced in 1895 by H. Burns. It comes from the Greek word "$\epsilon\iota\kappa\omega\nu$":image (and gave the word "icon") The equation's title is descriptive because it controls the formation of images in optical systems.

For a general ray we have :

$$\delta\varphi = \varphi(\boldsymbol{x_r}) - \varphi(\boldsymbol{x_e}) = \int_{\mathcal{R}(\boldsymbol{x_r},\boldsymbol{x_e})} n \mathrm{d}l \tag{2.32}$$

If we know the path of the ray, it should be noticed that we now have a relation between the phase difference and the refractive index. This relation will be used in the image reconstruction. The same can be realised for the attenuation, considering a complex refractive index, the imaginary part handling this attenuation.

The eikonal approximation can be pushed further, by neglecting also the refraction. The shortest path is therefore assumed to be the fastest, see Figure 2.5. This is known as the straight ray approximation. That approximation can hold as long as the variation of speed of sound are locally small, and that, the total travel path is relatively short (the ray does not have the time to bent).

2.4 Available Data

The physical modelling of the ultrasound tomography shows us how emitted waves will behave in the the aperture of the 3D-USCT. Acquiring the resulted transformed waves at multiple receivers (in the form of A-Scans, see Figure 2.6 for a example of A-Scan), and processing them is therefore a crucial step for getting the best of the information available.

One should note that this data pre-processing is completely independent of the reconstruction methods that will be studied in the next chapters.

The emitted signal is a coded-excitation puls. It is formed by a chirp signal with 2.5 MHz mean frequency and a bandwidth of 1 MHz[6].

2.4.1 Time of flight detection

For the speed of sound transmission reconstruction, we need to determined how long the delay between the emitted pulse and received pulse is. In the literature such time if referred as *Time of Flight* (TOF), *Time of Arrival* (TOA), *Time Delay of Arrival* (TDOA), or *Time Delay Estimation* (TDE).

In practice the TOF detection is done by using an optimal filter, making the cross correlation between the received signal and the same received signal but with a water-only sample.

The methods used for the time of flight detection are not detailed here, as they are the same as what Dapp used [15].

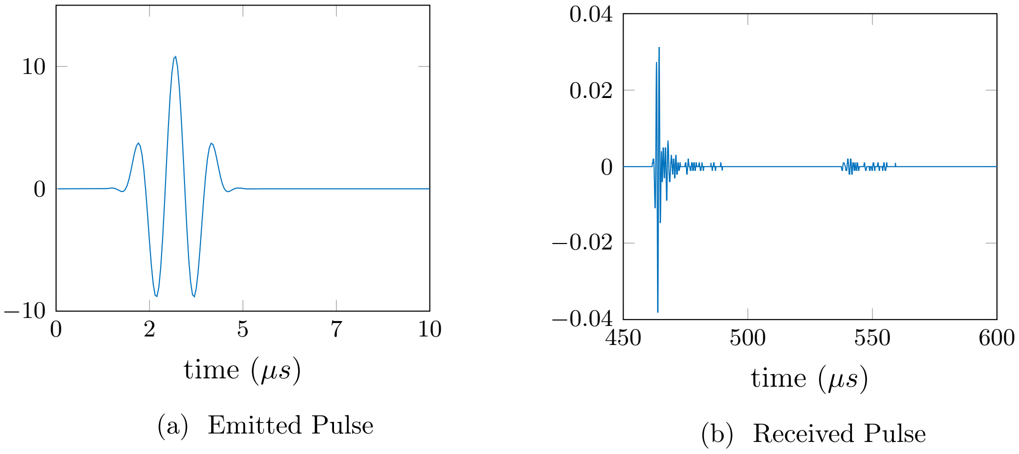
2.4.2 Transducer limitations

The transducers used in the USCT project are not perfect, and optimizing their conception is currently the topic of a PhD at IPE.

Among the different limitations impacting the available data for the reconstruction one can list:

Number of transducers

This parameter defines the amount of useable data. For N transducer we have $N(N-1)$ A-Scan acquired, and as many as time-of flight detection. However a rather larger part



(a) Emitted Pulse

(b) Received Pulse

Figure 2.6: Example of simulated A-Scans

of theses emitter-receiver pair (that one could considered as a ray) will not significantly probe the center of the aperture, *i.e.* where the breast should be.

Angle sensibility

For an incoming wavefront, a transducer sensibility varies with the incoming angle (because our transducer cannot be considered as a single point). An incoming wave in the same axis as a transducer would get a better restitution in the acquisition of the signal. Moreover, this response is not linear and is represented on Figure 2.7.

The main lobe of Figure 2.7 defined the best pairs of emitters-receivers, with the highest signal-to-noise ratio, the others pairs are unfortunately too attenuated to be relevant, and are thus discarded.

The width of this main lobe can be interpreted as an opening angle of the transducer (θ, on Figure 2.7). With the size of the aperture it defines a region of interest (the darkest region on Figure 2.8), where the best information is available. Outside, the lack of redundancy makes the reconstruction less resilient against noise.

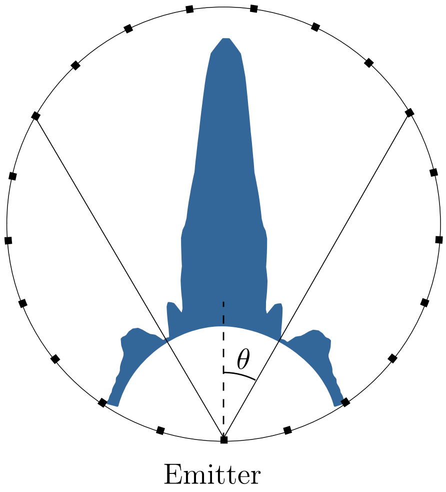


Figure 2.7: The amplitude of the acquired wave signal depends on the arrival of the incoming angle of the wavefront. A good signal to noise ratio is only available inside the main lobe. [15].

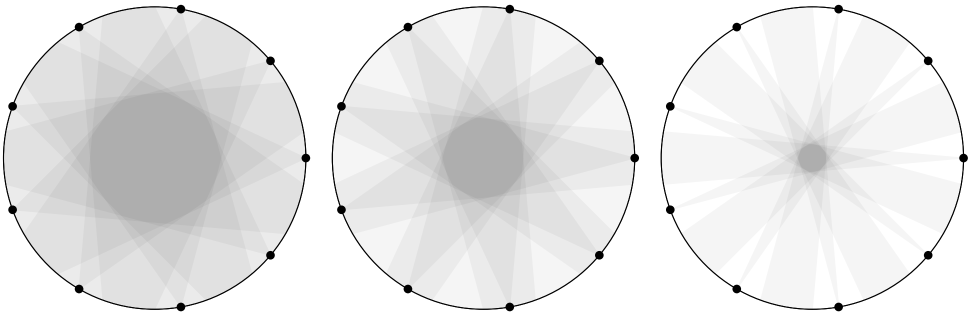
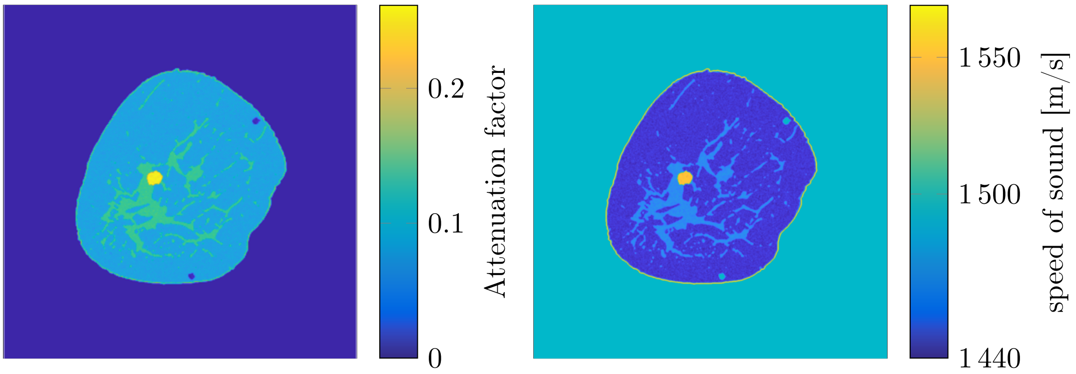

(a) large open angle (b) medium open angle (c) small open angle

Figure 2.8: The majority of the available information is condensed in the center of the reconstruction area (dark gray), and defines a particular region of interest where we could assume a good reconstruction, outside of this region we get a high sensibility to noise and reconstruction artifact (light gray/white)

2.4.3 Simulation

In order to easily test the different reconstruction methods being developed for the USCT-Project in a controlled and reproducible environment we use a forward simulation based on the k-Wave software toolbox [16, 17]. The problem size and the limits of the available computers (not only computational speed but also the available memory) make a complete simulation in 3D nearly impossible, and thus the simulations are run in a 2D space.

The simulation is run on a high resolution phantom (2700 x 2700 px, 1px $= 0.1\,\text{mm} \simeq \lambda/6$), created from the coronal view of a MRI image. This image has been segmented in different regions of similar tissues, which ultrasonic characteristic (density, speed of sound and attenuation) had been assigned with a gaussian distribution. 128 transducers are place in a circle around the breast, creating a maximum of $128 \times 127 = 16\,256$ time of flight acquisition.



(a) Groundtruth for the attenuation factor (b) Groundtruth for the speed of sound map

Figure 2.9: Breast phantom used for the simulation

In the k-Wave software, the emitted pulse is a wavelet represented on Figure 2.6a.

2.5 Summary

The US tomography can be modelled via simple wave equations, yet complex to solve as is. Many approximations are available to reduce the complexity of the problem and yields to different reconstructions methods, which we will study in the next 2 chapters.

In practice, the emission and reception of sound waves going through the aperture is handled by transducers, with their limitations. An extensive analysis of their behaviour and command is outside the scope of this report, but cannot be ignored, as they provide the input data for the reconstruction.

For conceiving and testing reconstruction algorithms in a reproducible way, we resort to the `k-wave` simulation software and a high resolution breast phantom in 2D.

Chapter 3

Infinite frequency tomography

The eikonal equation (2.31) is not linear, and it is *a priori* not possible to determine the refractive index (*i.e.* the speed of sound) from the phase differences. By assuming an infinite frequency wave, we can consider an ideal ray propagation. Only macroscopic effects are relevant, and scattering – happening at a microscopic scale – is neglected.

This chapter presents the actual approach to ray-based transmission tomography, and its limits.

3.1 Determining the ray path

As we are now considering rays, rather than full waves, parallels can be drawn with the geometrical optics. Determining the propagation time or *time of flight*, is dependent of the path taking by the ray, which is assumed to be the fastest one[18]. We can express the eikonal equation (2.31) in terms of a Euler-Lagrange Equation, and solving this equation is equivalent to a variation problem:

$$\frac{\mathrm{d}}{\mathrm{d}l}\left(\frac{1}{c}\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}l}\right) = \nabla\left(\frac{1}{c}\right) \tag{3.1}$$

With l the length of the ray. Solving this equation gives the path of the ray. With this path, we can then compute a *time of flight* defined as:

$$t = \int_{\mathcal{R}} \frac{1}{c(\boldsymbol{x}(l))}\mathrm{d}l = \int_{\mathcal{R}} s(\boldsymbol{x})dl \tag{3.2}$$

and introducing the slowness vector $s(\boldsymbol{x}) = 1/c(\boldsymbol{x})$.

For the reconstruction, we use the water-filled background as a reference, and so we can only compute a time shift:

$$\delta t = \int_{\mathcal{R}} \left(\frac{1}{c(\boldsymbol{x}(l))} - \frac{1}{c_0}\right)\mathrm{d}l = \int_{\mathcal{R}} s(\boldsymbol{x}) - s_0\mathrm{d}l = \int_{\mathcal{R}} \delta s\mathrm{d}l \tag{3.3}$$

3.2 Straight Rays

If one assume that straight and direct path between an emitter and a receiver is the fastest path, it assumes a homogeneous speed of sound along the ray. The result of the reconstruction based on this approximation will be an average of each pixel-ray value, giving us a first good approximation of the speed of sound distribution, but lacking of dynamic range. Determining the straight path can be done by the well-known Bresenham algorithm [19]. The weighting of the ray can achieves in different ways according to the geometry of Figure 3.1:

1. $m_{k,i} = a$
2. $m_{k,i} = \frac{\|x_r - x_r\|}{n_k}$, the total length is equally spread over the pixel path.
3. $m_{k,i} = l_{k,i}$.

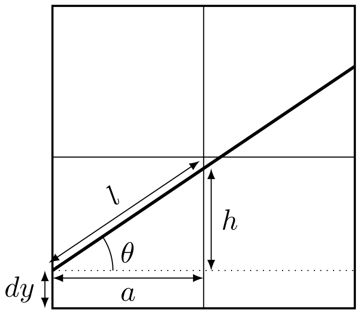


Figure 3.1: Exact distance of a ray in a pixel.

The length computed with the help of Figure 3.1 is more accurate, but computing it requires using heavy computation, and is 10 times slower for the reconstruction. Considering a fraction of the overall length (2.) is a good compromise between speed and quality.

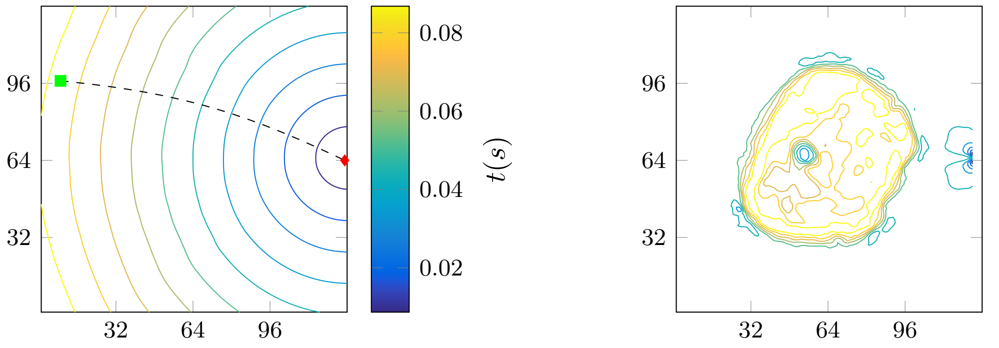
3.3 Bent rays

Bent ray approximation free us from assuming a constant speed of sound on the ray path. The fastest way from an emitter to a receiver is not necessarily a straight path, by passing through pixel with higher speed of sound (and not to far from the straight ray approximation) the ray will arrive faster to destination. Such path are called geodesics, and finding them is done using the fast marching algorithm.

3.3.1 Fast Marching Map

The fast marching method (FMM) is a design to solve the boundary value problem of the Eikonal equation. The idea is to determine the time of arrival of a *front* Γ a close boundary. This front moves at speed $c(\boldsymbol{x})$ so we have: $\nabla T = 1/c(x)$. The initial condition states $T(\Gamma_0) = 0$. Among the different methods, we used the *multistencil Fast Marching (MSFM)* methods [20], as it gives the most precise results.

As the FMM Map is unique for each transducer, we can save a lot of computing time by pre-computing those maps before iterating through the emitter-receiver pairs.



(a) Fast marching map

(b) Gradient of the fast marching map

Figure 3.2: The Fast marching map gives the time of flight between every point and an emittera. The gradient of this map clearly shows the influence of a heterogeneous medium.

3.3.2 Path extraction

For each emitter we thus compute this time of arrival map, and the construction of the geodesic path is done by searching a streamline FMM map. For the weighting of the bent rays different solution have been tested (with r_{ki} the position of the i^{th} pixel of the k^{th} ray and n_k the number of pixel in this ray)

1. $m_{k,i} = L_k/n_k$, same as for straight rays.
2. $m_{k,i} = \frac{1}{n_k}\sum_{i=1}^{n_k-1}\|r_{k,i+1} - r_{k,i}\|_2$
3. $m_{k,i} = \|r_{k,i} - r_{k,i-1}\|$
4. $m_{k,i} = \frac{\hat{t}_k}{c_k n_k}$ where $\hat{t}_k$ is the estimated time of arrival extract from the FMM map.

These different methods have been tested and only the number 1. and 4. of the length approximation have produced analysable results (*i.e.* with a sufficient wide value range in the reconstructed image)

3.4 Limits of Ray-Approximation

In the ray approximations (straight or bent) the maximum resolution is not dependent of the wave interference (diffraction and scattering) as in the Born approximation. However, we can use results from geometrical optics. If we don't rely on the infinite frequency hypothesis, the theoretical resolution power is limited by the so called Fresnel-Zone.

Between the emitter and the receiver, there is some off-axis propagation (not on the line of sight). This can deflect off of objects and then radiates to the receiver, adding to the direct path wave, in or out of phase. The $n-$Fresnel zone is defined as the volume where the phasing between any reflected ray and the direct one is less than n half-wavelength. Such regions are by construction ellipsoid, as shown on Figure 3.3.

In the first Fresnel zone, phase shift will be less than $\pi/2$, and indirect rays will be added constructively. Considering those rays increases the power of the received signal.

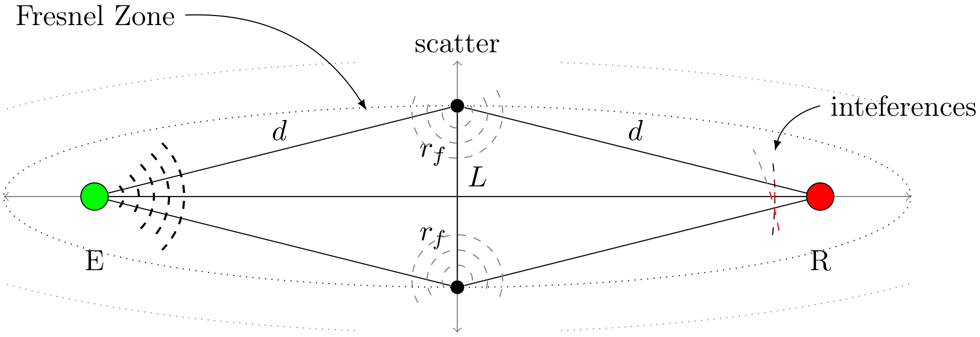


Figure 3.3: Interference in the first Fresnel's zone. The direct path from emitter (E) to receiver (R) receive also interferences from the scatter points.

In a more formal way, for an emitter $\boldsymbol{x_e}$ and a receiver x_r, any point $\boldsymbol{x}$ in the Fresnel zone, we have:

$$\|\boldsymbol{x} - \boldsymbol{x_e}\| + \|\boldsymbol{x_r} - \boldsymbol{x}\| - \|\boldsymbol{x_r} - \boldsymbol{x_e}\| \leq \frac{\lambda}{2} \tag{3.4}$$

Because the Fresnel Zone is an ellipsoid, we can determine its radius as:

$$\begin{aligned} d^2 &= r_f^2 + \left(\frac{L}{2}\right) \\ r_f^2 &= \left(\frac{L}{2} + \frac{\lambda}{4}\right)^2 - \left(\frac{L}{2}\right) \\ d_f &= 2r_f = \sqrt{L\lambda + \frac{\lambda^2}{4}} \overset{\lambda \ll L}{\simeq} \sqrt{L\lambda} \end{aligned}$$

In the case of the USCT, the A-scans have an average frequency of 2.5 MHz and a maximum speed of sound around 1500 m/s, which gives us a wavelength λ of 0.6 mm. The USCT 2 have an aperture diameter of 27 cm, and we then have a width of 12 mm, in the USCT 3 the aperture diameter is 35 cm, having thus a higher region of interest (most of the signal overlap in the center of the aperture, where ideally the breast is fully positioned), in this case we get 14 mm. One should notice that this width is not the limit of the resolution power, but a range where the information about a ray can be spread.[21]

3.5 Forward Model

In our case the goal of the transmission tomography reconstruction is to obtain the speed of sound of each voxel in the breast, as it is a good estimator of the presence of tumorous cells (see Figure 1.1 and [4]). The measuring process and time detection can be expressed as a linear operator (e.g. a matrix):

$$\boldsymbol{t} = \mathcal{M}(\boldsymbol{c}) \tag{3.5}$$

Where $\boldsymbol{t}$ is the measured data, here the time of arrival of a received puls, $\boldsymbol{M}$ is the measurement operator, and $\boldsymbol{c}$ the distribution of the speed of sound. If $\mathcal{M}$ is a linear operator, we can express it as a matrix $\boldsymbol{M}$ calls the measurement-matrix. As we will work with the eikonal approximation, $\boldsymbol{M}$ is also named the *ray-matrix*.

In fact, the Ray matrix is not independent of $\boldsymbol{c}$, as the propagation of the rays are determined by the speed of sound. So we can proceed with an iterative reconstruction method:

1. Compute $\mathcal{M}_k$ from $\hat{\boldsymbol{c}}_k$ with an method from chapter 3.
2. Solve $\boldsymbol{y} = \mathcal{M}_k(\hat{\boldsymbol{c}}_{k+1})$ with an algorithm from chapter 5.
3. Stop reconstruction if $k > n_{max}$ or if the relative error is below tolerance, else increase k.

The initial value of $\hat{\boldsymbol{c}}_0$ without further information is an homogeneous water filled background. With such a scheme, the approximation of the forward model is more precise at each step, and converge to a final estimate $\hat{\boldsymbol{c}}_f$.

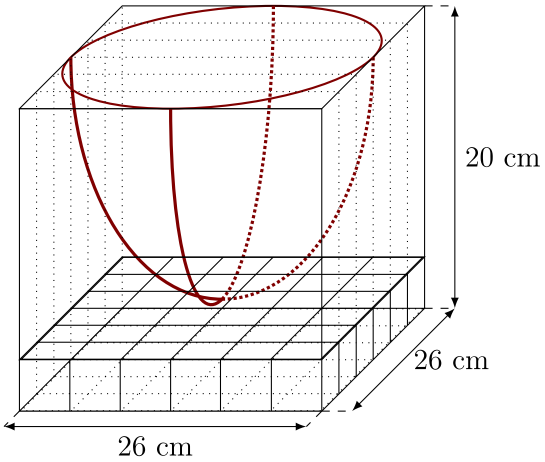
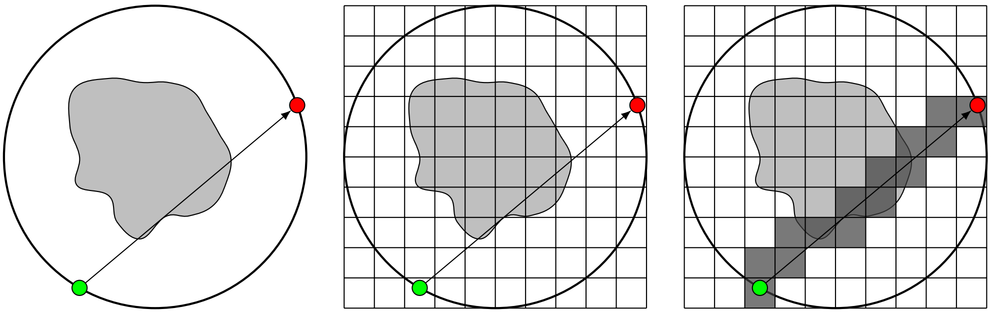


Figure 3.4: Aperture shape and space quantification



(a) Continuous path (b) Discretisation (c) Quantification

Figure 3.5: The different steps of the rays quantification

3.6 Ray Matrix construction: Example of straight ray

In the eikonal approximation, the linearization of equation (3.2) can be done relatively to the water background ($s_0 = \frac{1}{c_0}$) if we consider no dampening across the ray:

$$\delta t = \int_{\mathcal{R}} (s(\boldsymbol{x}) - s_0)\,\mathrm{d}l = \int_{\mathcal{R}} \frac{1}{c(\boldsymbol{x})} - \frac{1}{c_0}\mathrm{d}l \tag{3.6}$$

By considering a discrete space (see Figure 3.4 and Figure 3.5) we have for the k-th ray:

$$\delta t_k = \sum_{i=1}^{n_k} l_{ki}\delta s_i \tag{3.7}$$

Where δs_i is the relative slowness in the i^{th} voxel hit by the ray, and l_{ki} the length of the k^{th} ray in this voxel, and n_k the number of voxels meet by the ray. By considering all the rays and voxel we can build a linear system:

$$\underbrace{\begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_M \end{bmatrix}}_{\boldsymbol{t}} = \underbrace{\begin{bmatrix} l_{11} & l_{12} & \cdots & l_{1N} \\ l_{21} & l_{22} & \cdots & l_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ l_{M1} & l_{M2} & \cdots & l_{MN} \end{bmatrix}}_{\boldsymbol{L}} \cdot \underbrace{\begin{bmatrix} \frac{1}{c_1} \\ \frac{1}{c_2} \\ \vdots \\ \frac{1}{c_N} \end{bmatrix}}_{\boldsymbol{s}} \tag{3.8}$$

From this system $\boldsymbol{t} = \boldsymbol{L}\boldsymbol{s}$ (in the next chapter $\boldsymbol{y} = \boldsymbol{M}\boldsymbol{x}$) we can express the measure operator of the USCT for transmission tomography from equation (3.5):

$$\boldsymbol{t} = \underbrace{\boldsymbol{M}\,\boldsymbol{\mathcal{I}}}_{\boldsymbol{\mathcal{M}}}(\boldsymbol{c}) \tag{3.9}$$

With $\mathcal{I}(c)$ the per-term inverse of c, which is a non-linear operator, but is conveniently an invertible one. In practice the reconstruction software will use the linear decomposition $t_{tot} = \delta t + t_{water}$ $s = \delta s + s_0$; and will only solve for the difference terms $\delta\boldsymbol{t} = \boldsymbol{M}\delta\boldsymbol{s}$, and add the reference speed at the end. Furthermore, the reference speed of sound in the water can be established as a function of the temperature distribution [22], slightly improving the precision of the reconstructed image.

3.7 Problem size

As explained in chapter 2, the eikonal approximation, based on ray propagation is a good compromise between computation power and resolution power. With this method, the smallest pixel size acceptable is 3 to 4 times the wavelength (so around 2 mm wide pixel). This yields to a choice of reconstruction volume of 128 × 128 × 96 voxels, or a slice of 128 × 128 in 2D, we have then 1 572 864 pixels to reconstruct.

As measurement data we got $157 \times 4 \times 157 \times 9 \times 10 = 8\,873\,640$ times of arrival from pair of emitters-receivers (with ten rotation of the aperture). As presented in section 2.4.2 the transducer response is only good in a cone 30° wide.

Only a third of the time picking is usable, we get 2 957 880 times of arrivals in this configuration. Thus, the problem is over-determined, and we also know that noise could also disturb the time of flight picking.

In fact each ray gives only information on the voxel it has met, making $\boldsymbol{M}$ a sparse matrix. In a straight ray with 2D simulation we only get roughly 30% voxels hits in the matrix. This result is much worse in 3D.

The construction of the matrix is done as described in Figure 3.6, where the reconstruction image is flattened into a vector. Using CSC-format for saving it greatly reduces the memory usage, but the sparsity of the matrix also means that we lack information about pixels. Having our rays covering more pixels adds information for the reconstruction. This fact will be the starting point of the different rays approximations we will discuss in chapter 4.

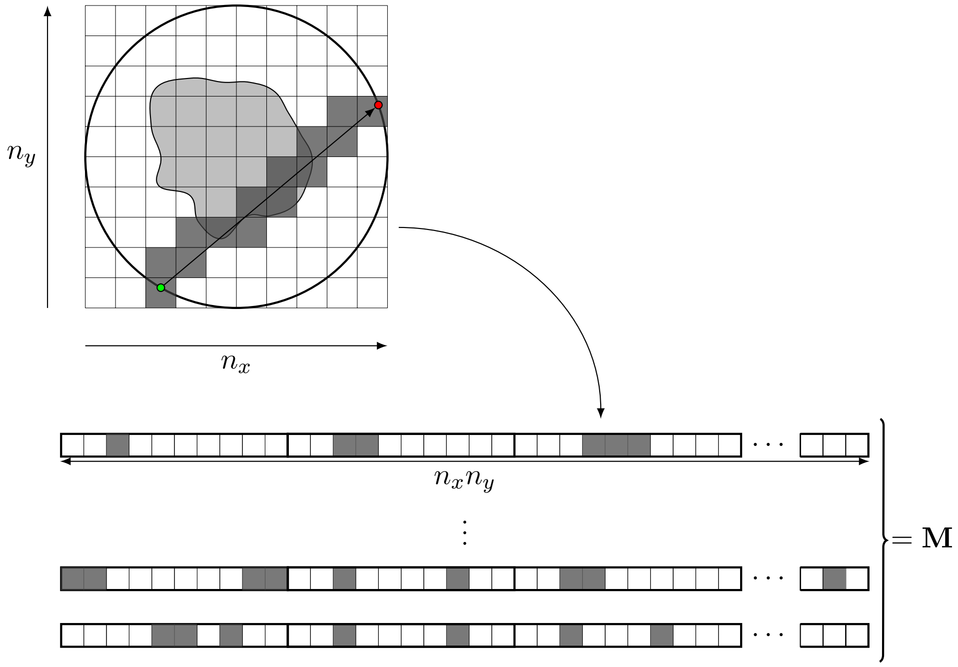


Figure 3.6: 2D example of matrix construction from a straight ray. each row of the matrix corresponds to a specific ray, each column to a voxel

3.8 Reducing the problem size

As we have seen, the problem size is *big*, especially in 3D. But only a part of the reconstructed voxels is worth of interest, *i.e.* the ones where we have breast cells (the rest being considered as water). So it makes sense to only compute the reconstruction in this *region of interest* (ROI). By concentrating the computational power to this region we not only reduce the time needed for the reconstruction, but also discard the background noise, by considering all pixels outside the ROI as water.

To determine this region I tried various image processing techniques to create a binary segmentation of the reconstructed image. The best results (which are the most resilient to noise and a reduced number of transducers) are obtained by computing the gradient of the reconstructed image (via a Sobel Filter) and then used Otsu's Method [23] on the joint histogram of the gradient and a Gaussian filtered one. The gradient work as a good edge detector in our case, as the gradient of the noise background is much smaller as the one between the water and the cell. Otsu's method allows us to compute an automatic and theoretically optimal threshold value for the binary segmentation, and frees us of manually tweaking a threshold for each reconstructed image. To remove the salt or pepper noise which could remain from the binarisation, I applied a median filter, and to reduce the risk of error and false detection, an error margin is fixed and apply through a dilatation filter. Once a close border of the breast is determined, it is filled in, and we get the complete breast include in the ROI The overall process is represented on Figure 3.7.

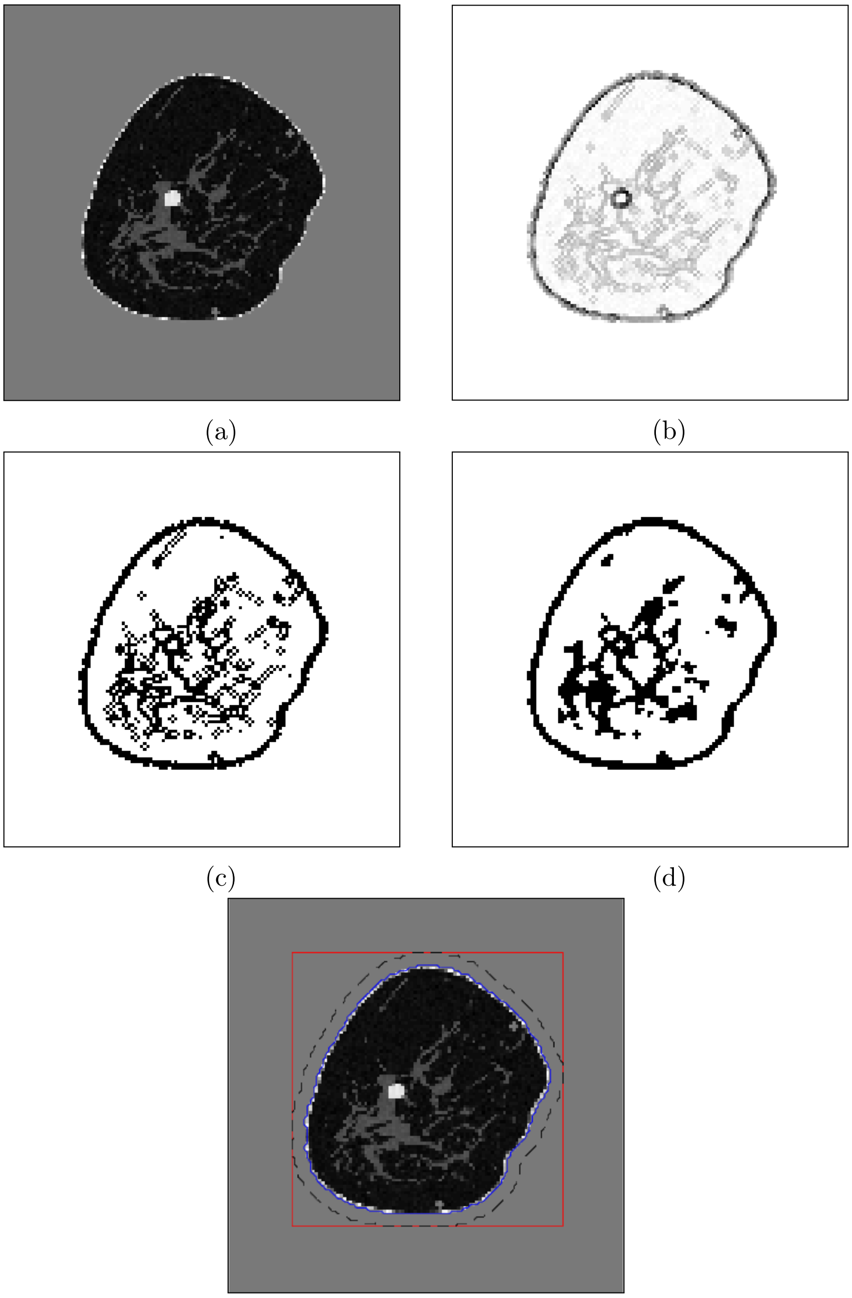
However, more tests would be needed with experimental data, where artifacts could appear in the background and create therefore false detection.

3.9 Summary

The overall process of the Transmission Tomography is represented in the Figure 3.8. From an initial guess on the reconstructed medium and a measured vector of time of arrival , we build a forward model, based on the eikonal equation. This model is represented in the *ray-matrix*, a linear operator between the relative slowness and the difference of time of arrival (with respect to an homogeneous water-filled background).

Using a smart reconstruction by concentrating the solving on a defined region of interest helps us to reduce the influence of the noise on the overall reconstruction. The cost (time) of computing this region of interest is worth it, as it allow us to greatly reduce the problem size (less pixels to reconstruct) and thus makes it faster to solve.

However, the sparse characteristic of the constructed matrix clearly shows that a simple ray tomography does not embrace all of the information available. This implies the need of a large amount of rays (*i.e.* numerous wide-opening-angle transducers) to be able to cover the complete aperture.



(a) (b)

(c) (d)

(e)

Figure 3.7: Segmentations procedure: the original image (a) is differientated through Sobel's filter (b), then binarized with Otsu's method (c), and median filter (d). In the end(e) we get the boundary of the breast (blue), a safe zone to reconstruct (where the real breast should be,dash line) and the bounding box (red) of the reconstruction volume (where the inverse problem will be solved).

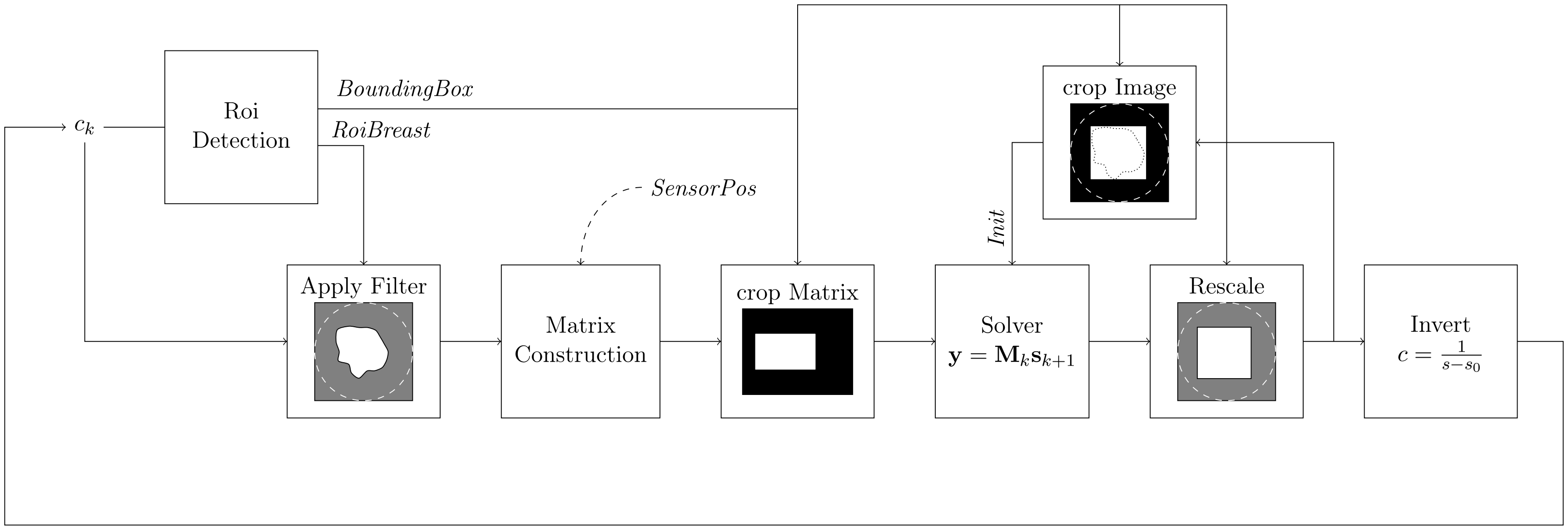


Figure 3.8: Transmission tomography Process, the black part of filter are cut off, the white is conserved and the gray is considered as water.

Chapter 4

Finite frequency tomography

In chapter 2 we saw that the Fresnel Zone defines a volume in which our rays gather information. The Fresnel zone is frequency dependent, and where our previous rays approximation assumed an infinite frequency, the *fat ray* will assume a *finite and fixed frequency.*

Originally used in geophysics and seismic tomography [24, 25, 26], I adapted it to the case of ultrasound transmission tomography . The main idea behind fat ray tomography is to spread the ray among more pixels than with a classic, single-pixel wide ray, by considering also the pixel along the “side” of the ray in the Fresnel Zone, having more entries in the ray matrix, and reducing it sparsity.

Furthermore, the speed of sound distribution can vary inside the Fresnel zone, the use of the fat ray allows us to go beyond this limitation.

4.1 Defining the Fresnel Zone

The $n-$ Fresnel Zone around a straight ray in a homogeneous medium, can be defined as the ellipse which focal point are the emitter and receiver. Odd-numbered Fresnel Zone allows for constructive interference ($< \lambda/2$), even-numbered zone makes destructive interference. In the following will refer only of the first Fresnel Zone, or just Fresnel Zone.

For more complex medium one should return to the original definition of the Fresnel zone: the region between the source and receiver for which the scattered wave field contributes constructively at the receiver position.

$$\delta\tau = \tau(EF) + \tau(FR) - \tau(ER) \leq \frac{1}{2f} \tag{4.1}$$

Where $\tau(EF), \tau(FR), \tau(ER)$, are the time of flight on the according paths.

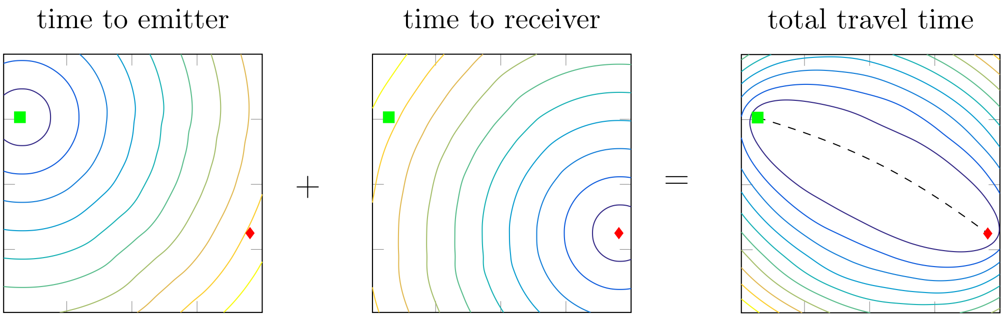


Figure 4.1: Construction of the fat ray using the time of flight between an emitter and a receiver. The dashed line represents the fastest path, passing by the minimum of the map.

4.1.1 Paraxial approximation

The time defined Fresnel volume also has a geometrical definition, that can be used to determining the size of it. Without loss of generality one can consider the straight ray case.

The equation (4.1) is rewritten as:

$$\delta l = l(EF) + l(FR) - l(ER) \leq \frac{\lambda}{2} \tag{4.2}$$

On the boundary of the Fresnel zone its radius can be computed by making the *paraxial approximation* $r \ll x, L - x$:

$$\sqrt{x^2 + r(x)^2} + \sqrt{(L-x)^2 + r(x)^2} - L \simeq \frac{1}{2} r(x)^2 \frac{L}{(L-x)x} = \frac{\lambda}{2}$$

Finally the radius of the Fresnel zone can be approximate by:

$$r(x) = \sqrt{\frac{\lambda(L-x)x}{2L}} \tag{4.3}$$

It is obvious from Figure 4.2 that the paraxial Fresnel volumes are good, sufficiently accurate approximations for the exact Fresnel volumes. Only close to emitter and receiver, particularly for low frequencies, the accuracy is lower. The differences in the vicinity E and R are discussed in greater detail.

4.1.2 Limits of the paraxial approximation

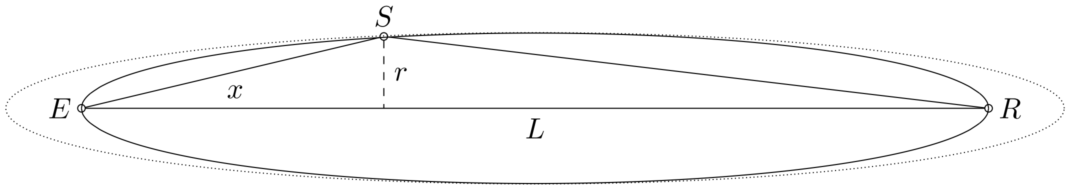


Figure 4.2: 2D projection of the real Fresnel Zone (dotted line) and its paraxial approximation delimitation. The approximation holds only in the middle of the ellipse.

At the focal point (E and R) the exact radius of the fresnel zone is:

$$r_{exact}(E) = r_{exact}(R) = \frac{\lambda}{2} \frac{1 + \lambda/4L}{1 + \lambda/2L} \tag{4.4}$$

the approximation $\lambda \ll L$ leads to $r \simeq \frac{\lambda}{2}$. In the paraxial computation, however, the radius shrinks to zero, as well as for the "overshooting" after the receiver/emitter (it is $\lambda/4$ without the paraxial approximation).

The paraxial approximation holds still in the middle of the ray. It could therefore be used to characterize the Fresnel Zone, and is in some cases [27, 28, 26] also use to compute sensibility Kernel (see section 4.3).

4.2 Linear Interpolation

For a fat ray, the time of an arrival is influenced by all the voxels inside the Fresnel Zone, therefore, equation (3.2) becomes a volume integral:

$$t_k = \int_V K_t(\boldsymbol{x}) \frac{1}{c(\boldsymbol{x})} \mathrm{d}\boldsymbol{x} \tag{4.5}$$

Where $K_t(\boldsymbol{x})$, is the time sensibility Kernel assimilated to the inverse of an area, stating how each voxel is contributing to the fat ray. Without further information, $K(\boldsymbol{x})$ could be considered as a cross section area of the fat ray perpendicular to the central ray. Every point in the ray has hence an equal contribution to the travel time. In a discretized and linearized form it can also be understood as the sensitivity of t_k with respect to change in the slowness distribution, building a Jacobian Matrix. In this case the partial derivative with respect to a change in velocity c_i:

$$\frac{\partial t_k}{\partial c_i} = -\frac{v_i}{V}\frac{t_k}{c_i} \tag{4.6}$$

Where v_k is the volume of the cell, and V the total volume of the fat ray. By using the chain rule the sensitivity term can be expressed in term of slowness, and this form will be used in the reconstruction software.

$$\frac{\partial t_k}{\partial s_i} = \frac{\partial t_k}{\partial c_i}\frac{\partial c_i}{\partial s_i} = \frac{\partial t_k}{\partial v_i}\left(-\frac{1}{s_i^2}\right) = \frac{v_i}{V}\frac{t_k}{s_i} \tag{4.7}$$

A first improvement of this flat kernek it to modulate it by a linear attenuation [29] from the center of the fat ray, even if such adjustment lack of a physical ground.

$$w(\tau) = \begin{cases} 1 - 2f\Delta\tau & (0 \le \Delta t \le 1/2f) \\ 0 & (1/2f \le \Delta t) \end{cases} \tag{4.8}$$

giving the weight in the matrix for the $i-$th ray and the j-th pixel

$$m_{k,i} = \frac{w(\tau_i)}{\sum w(\tau_i)}\frac{v_i}{V}\frac{t_k}{s_i} \tag{4.9}$$

4.3 Wave-based approximation

The Fat Ray theory make the assumption of a finite frequency, and numerous research have been done with this hypothesis, especially in geophysics. The following is a development about the sensitivity kernel [28, 30, 27].

According to the Rytov Approximation (see subsection 2.3.2) the pressure field received at position $\boldsymbol{x}_r$ (emitted from $\boldsymbol{x}_e$ at pulsation ω)

$$P_r(\boldsymbol{x}_r, \boldsymbol{x}_e, \omega) = P_i(\boldsymbol{x}_r, \boldsymbol{x}_e, \omega) \exp\left(\frac{P_B(\boldsymbol{x}_r, \boldsymbol{x}_e, \omega)}{P_i(\boldsymbol{x}_r, \boldsymbol{x}_e, \omega)}\right) \tag{2.19 again}$$

From (2.19) the phase shift between the emitted and received pulse can be determined as:

$$\delta\varphi(\boldsymbol{x}_r, \boldsymbol{x}_e) = \mathrm{Im}\left[\frac{P_B}{P_i}\right] \tag{4.10}$$

For a fixed frequency (4.10) can be expressed as a time difference:

$$\delta t = \frac{1}{\omega}\delta\varphi = \frac{1}{\omega}\mathrm{Im}\left[\frac{P_B}{P_i}\right] \tag{4.11}$$

At the receiver, the Born pressure field (scattered wave) and the initial pressure field are express as:

$$P_B(\boldsymbol{x}_r, \boldsymbol{x}_e) = \int_\Omega G(\boldsymbol{x}_r, \boldsymbol{y}) k_0^2 \delta n(\boldsymbol{y}) P_i(\boldsymbol{y}, \boldsymbol{x}_e, \omega)\mathrm{d}\boldsymbol{y} \tag{2.18 again}$$

$$P_i(y, \boldsymbol{x}_e) = P_0 G(\boldsymbol{y}, \boldsymbol{x}_e) \tag{4.12}$$

By using (4.12) and (2.18) the phase term of (2.19) is computed as:

$$\frac{P_B}{P_i}(\boldsymbol{x}_r, \boldsymbol{x}_e) = \int_\Omega \frac{-2\omega^2 \delta c(\boldsymbol{y})}{c_0^3} \frac{G(\boldsymbol{x_r}, \boldsymbol{y}) G(\boldsymbol{y}, \boldsymbol{x}_e)}{G(\boldsymbol{x}_r, \boldsymbol{x}_e)}\mathrm{d}\boldsymbol{y}$$

And with (4.10) and (4.11):

$$\delta t(\boldsymbol{x}_r, \boldsymbol{x}_e) = \int_\Omega \underbrace{\frac{-2\omega}{c_0^3}\mathrm{Im}\left[\frac{G(\boldsymbol{x}_r, \boldsymbol{y}) G(\boldsymbol{y}, \boldsymbol{x}_e)}{G(\boldsymbol{x}_r, \boldsymbol{x}_e)}\right]}_{K_t(\boldsymbol{y},\omega)} \delta c(\boldsymbol{y})\mathrm{d}\boldsymbol{y} \tag{4.13}$$

From chapter 2:

$$\begin{aligned} c &= c_0 + \delta c \\ s &= \frac{1}{c} = s_0 + \delta s \\ \delta c &= c - c_0 = \frac{1}{s_0 + \delta s} - \frac{1}{s_0} \simeq \frac{-\delta s}{s_0^2} \end{aligned}$$

And (4.13) is rewritten to work with slowness:

$$\delta t(\boldsymbol{x}_r, \boldsymbol{x}_e) = \int_\Omega \underbrace{2\omega s_0 \mathrm{Im}\left[\frac{G(\boldsymbol{x}_r, \boldsymbol{y}) G(\boldsymbol{y}, \boldsymbol{x}_e)}{G(\boldsymbol{x}_r, \boldsymbol{x}_e)}\right]}_{K_t(\boldsymbol{y},\omega)} \delta s(\boldsymbol{y})\mathrm{d}\boldsymbol{y} \tag{4.14}$$

4.3.1 3D Kernel

In 3D the Green function is defined as:

$$G(\boldsymbol{x}, \boldsymbol{y}) = \frac{1}{4\pi} \frac{e^{ik_0 \|\boldsymbol{x} - \boldsymbol{y}\|}}{\|\boldsymbol{x} - \boldsymbol{y}\|}$$

by reusing the notation of (4.2):

$$K_t^{3D}(\boldsymbol{y}, \omega) = \frac{\|\boldsymbol{x}_r - \boldsymbol{x}_e\|}{\|\boldsymbol{x_r} - \boldsymbol{y}\| \|\boldsymbol{y} - \boldsymbol{x}_e\|} s_0 \frac{\omega}{2\pi} \sin(\omega s_0 \delta l(\boldsymbol{y})) \tag{4.15}$$

This formulation of the sensibility Kernel $K(\boldsymbol{y}, \omega)$ only assumes that there is a homogeneous speed of sound for the wave propagation and a mono frequency pressure fields and only one scattering point for each ray in the Fresnel zone.

The mono frequency hypothesis can be dropped by considering the spectrum of the pressure field:

$$\boxed{K_t^{3D}(\boldsymbol{y}) = \frac{s_0}{2\pi} \frac{\|\boldsymbol{x}_r - \boldsymbol{x}_e\|}{\|\boldsymbol{x}_r - \boldsymbol{y}\| \|\boldsymbol{y} - \boldsymbol{x}_e\|} \int_{\omega-\Delta\omega}^{\omega+\Delta\omega} A(\omega)\omega \sin(\omega s_0 \delta l(\boldsymbol{y})) \mathrm{d}\omega} \tag{4.16}$$

With $\int_{\omega-\Delta\omega}^{\omega+\Delta\omega} A(\omega) = 1$ and $A(\omega)$ null outside $\omega \pm \Delta\omega$.

4.3.2 2D Kernel

When restraining us to a 2D reconstruction, the Green function is impacted (use of a cylindrical wave instead of a spherical wave). Using the far field approximation ($\|y - x_e\| \gg \lambda$)

$$G_{2D}(\boldsymbol{x}, \boldsymbol{y}) = \frac{-1}{\sqrt{8\pi \|\boldsymbol{x} - \boldsymbol{y}\|}} \exp\left(ik_0(\|\boldsymbol{x} - \boldsymbol{y}\|) + i\frac{\pi}{4}\right)$$

With this expression of the Green function, the sensibility Kernel become:

$$\boxed{K_t^{2D}(\boldsymbol{y}) = \sqrt{\frac{s_0}{2\pi}} \sqrt{\frac{\|\boldsymbol{x}_r - \boldsymbol{x}_e\|}{\|\boldsymbol{x}_r - \boldsymbol{y}\| \|\boldsymbol{y} - \boldsymbol{x}_e\|}} \int_{\omega-\Delta\omega}^{\omega+\Delta\omega} A(\omega)\sqrt{\omega} \sin\left(\omega s_0 \delta l(\boldsymbol{y}) + \frac{\pi}{4}\right) \mathrm{d}\omega} \tag{4.17}$$

4.3.3 Attenuation Kernel

As with the simple ray theory an attenuation sensibility Kernel can also be computed, using the amplitude variation of the Rytov pressure field A_r with respect to the reference amplitude A_0.

$$\ln\left(\frac{A_r}{A_0}(\boldsymbol{x}_e, \boldsymbol{x}_r)\right) = \mathrm{Re}\left[\frac{P_B}{P_i}\right] \tag{4.18}$$

for small variation $A_r = A_0 + \delta A$ and so :

$$\frac{\delta A}{A_0} \simeq \mathrm{Re}\left[\frac{P_B}{P_i}\right] = \int_\Omega \delta s K_A(\boldsymbol{y}) \mathrm{d}\boldsymbol{y} \tag{4.19}$$

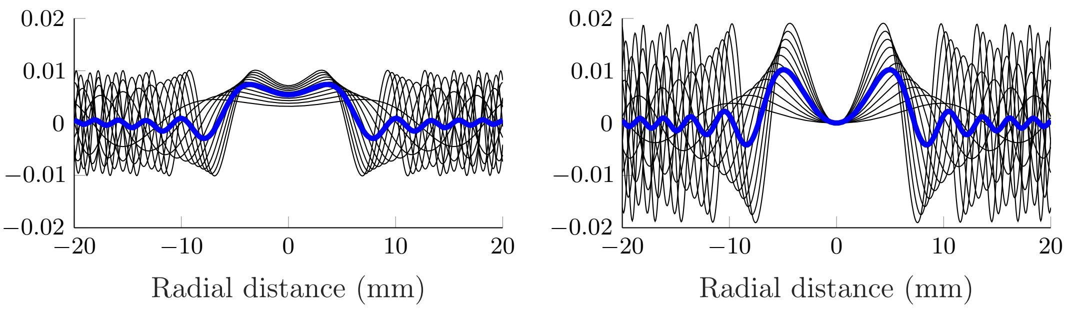
- In 3D:

$$\boxed{K_A^{3D}(\boldsymbol{y}) = \frac{s_0}{2\pi} \frac{\|\boldsymbol{x}_r - \boldsymbol{x}_e\|}{\|\boldsymbol{x}_r - \boldsymbol{y}\| \|\boldsymbol{y} - \boldsymbol{x}_e\|} \int_{\omega-\Delta\omega}^{\omega+\Delta\omega} A(\omega)\omega^2 \cos(\omega s_0 \delta l(\boldsymbol{y})) \mathrm{d}\omega} \tag{4.20}$$

- In 2D:

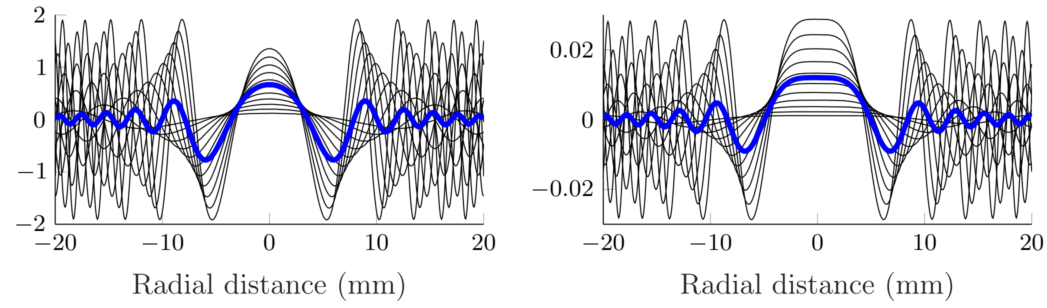
$$\boxed{K_A^{2D}(\boldsymbol{y}) = \sqrt{\frac{s_0}{2\pi}} \sqrt{\frac{\|\boldsymbol{x}_r - \boldsymbol{x}_e\|}{\|\boldsymbol{x_r} - \boldsymbol{y}\| \|\boldsymbol{y} - \boldsymbol{x}_e\|}} \int_{\omega-\Delta\omega}^{\omega+\Delta\omega} A(\omega)\omega^{3/2} \cos(\omega s_0 \delta l(\boldsymbol{y})) \mathrm{d}\omega} \tag{4.21}$$

As shown on Figure 4.3, considering a broad band signal highly decreases the variations inside the sensitivity kernel (due to beat interference). Furthermore, in 3D the kernel is not influenced by the pixels along the direct path (straight ray), as if the ray did not take its own path. This is phenomenon is known as the banana-doghnut paradox in the literature of seismics [31, 27, 26, 32].



(a) Time Kernel 2D

(b) Time Kernel 3D



(c) Attenuation Kernel 2D

(d) Attenuation Kernel 3D

Figure 4.3: Cross section of the presented sensibility kernels (normalized amplitude), the thick blue line represents the averaged Kernel, black line are kernel computed for different frequencies from 0.5 MHz to 2.5 MHz. The higher is the sampling frequency (in space and time), the smoother is the Kernel.

4.4 Discrete Scheme

All the integrals defined previously for the fat ray needs to be discrete, for allowing a numerical approximation.

4.4.1 Numerical Optimisation

Furthermore, the sensibilities kernels are computed for the complete reconstruction space, this full wave approach would be too expensive to compute, and reducing them only to the first Fresnel Zone help to speed the computation process, as well as reducing the influence of noisy data.

4.4.2 Wave-based Kernel

From the continuous Kernel expression:

$$\delta t(\boldsymbol{x}_r, \boldsymbol{x}_e) = \int_\Omega \delta s(\boldsymbol{y}) K(\boldsymbol{y}, \omega) \mathrm{d}\boldsymbol{y} \tag{4.13 again}$$

With $\Delta x, \Delta y, \Delta z$ the resolution step of the different reconstruction space dimensions the elementary volume is $\mathrm{d}y \simeq \Delta x \Delta y \Delta z$ and in the same way $\mathrm{d}\omega = \Delta\omega$

For the k-th ray, in its Fresnel Zone $\mathcal{F}_k$, in 3D:

$$\delta t_k = \Delta x \Delta y \Delta z \Delta\omega \frac{s_0}{2\pi} \sum_{y_i \in \mathcal{F}_k} \sum_j \frac{\boldsymbol{d}_k}{\boldsymbol{E}_{k,i} \boldsymbol{R}_{k,i}} A(\omega_j)\omega_j \sin(\omega_j s_0 \boldsymbol{E}_{k,i} + \boldsymbol{R}_{k,i} - \boldsymbol{d}_k) \tag{4.22}$$

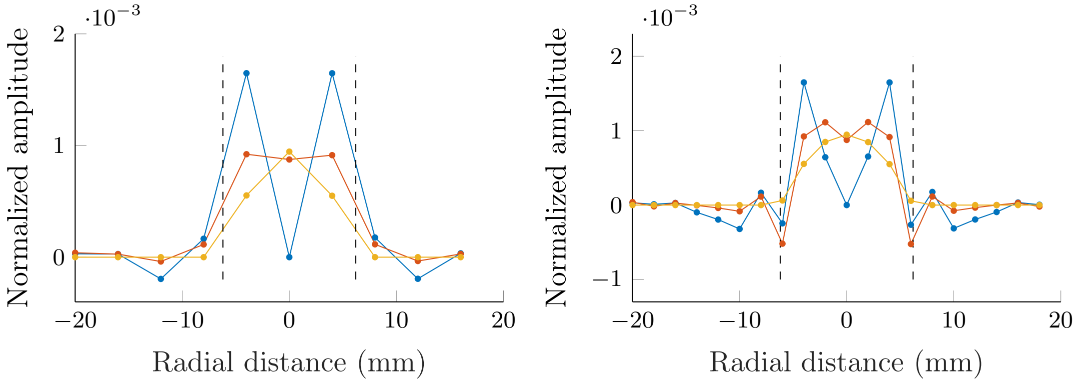
Where $\boldsymbol{E}$ and $\boldsymbol{R}$ are the matrix of distance between every pixel and emitter or receiver; $\boldsymbol{d}_k$ the straight distance between an emitter and a receiver.

These matrices can be pre-computed, as for a fixed emitter (resp. a fixed receiver) the distance between it and every pixel can be reused for every other ray with the same emitter (resp. receiver).

The same expression can similarly be found for the other Kernels.

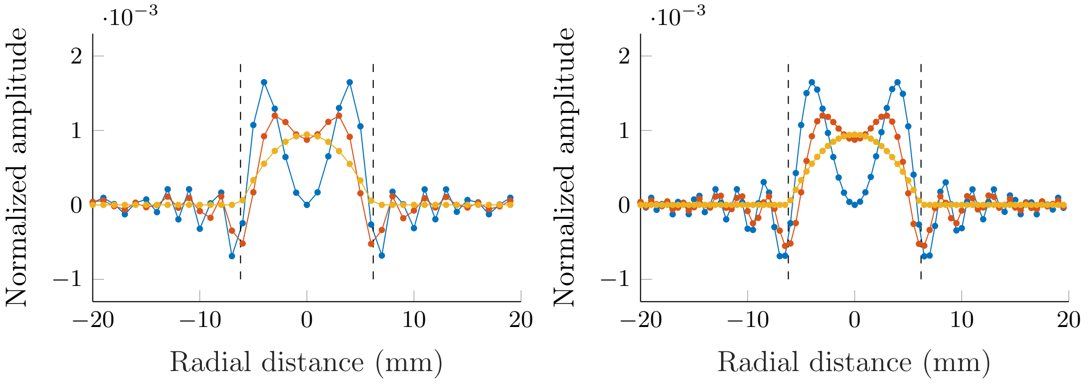
4.4.3 Resolution Step

Another aspect of discretely computing the kernels, is the resolution steps (*i.e.* pixel size). On Figure 4.3 the represented kernels are very smooth, and the high frequency variation are quite visible. By decreasing the resolution, one can see on Figure 4.4 that the sensitivity kernel could be approximated by a simpler, faster to compute function (*e.g.* a trapezoidal shape for the 2D kernel in Figure 4.4a and 4.4b).



(a) 64px, 1px = 4 mm

(b) 128px, 1px = 2 mm



(c) 256px, 1px = 1 mm

(d) 512px, 1px = 0.5 mm

Figure 4.4: The resolution step greatly influence the overall aspect of the kernel

Using a 128 or a 256 pixel wide grid is again a good deal between speed and resolution power, the lobes of the kernel are visible, and the computation on a standard 4-core CPU is below a minute of computation. A MATLAB implementation and a C implementation of the 2D Kernel computation has been realized. The latter being memory efficient it could be extended to a 3D computation. However, it is still a monothreaded computation, and is therefore slower than the MATLAB vectorized version. A Huge improvement would be to used dedicated GPU computation for this task. In fact, the bottleneck of computation has been reduced to the sole computation of the sinus operation of (4.22), which have to be conduct $N_{\mathcal{F}} \times N_\omega \times N_{ray}$ in total.

Another speed and memory improvement could be conduct by using `single` (16 bits float) datatype everywhere, however as of 2020 the current version of MATLAB does not allow for sparse matrix to contain something differents than `double` (32 bits float) datatype.

4.4.4 Extension of the Fat Ray framework

Fourier expression

The expression of the Fréchet Kernel can be rewritten as:

$$\begin{aligned}
K_t^{3D}(\boldsymbol{y}) &= \frac{s_0}{2\pi}\frac{\|\boldsymbol{x}_r-\boldsymbol{x}_e\|}{\|\boldsymbol{x}_r-\boldsymbol{y}\|\|\boldsymbol{y}-\boldsymbol{x}_e\|}\int_{\omega-\Delta\omega}^{\omega+\Delta\omega} A(\omega)\omega\sin(\omega s_0\delta l(\boldsymbol{y}))\mathrm{d}\omega \\
&= \frac{s_0}{2\pi}\frac{\|\boldsymbol{x}_r-\boldsymbol{x}_e\|}{\|\boldsymbol{x}_r-\boldsymbol{y}\|\|\boldsymbol{y}-\boldsymbol{x}_e\|}\mathrm{Re}\left\{\int_{\omega-\Delta\omega}^{\omega+\Delta\omega} A(\omega)\omega\exp(j\omega s_0\delta l(\boldsymbol{y}))\mathrm{d}\omega\right\} \\
&= \frac{s_0}{2\pi}\frac{\|\boldsymbol{x}_r-\boldsymbol{x}_e\|}{\|\boldsymbol{x}_r-\boldsymbol{y}\|\|\boldsymbol{y}-\boldsymbol{x}_e\|}\mathrm{Re}\left[\mathcal{F}^{-1}\left\{A(\omega)\omega\right\}\right](s_0\delta l(\boldsymbol{y}))
\end{aligned} \tag{4.23}$$

Where $\mathcal{F}^{-1}$ denotes the inverse fourier transform. The expression $\mathrm{Re}\left[\mathcal{F}^{-1}\left\{A(\omega)\omega\right\}\right]$ is the same for every kernel. Instead of computing it for each row, a look-up table could be use to determine each pixel value for the kernel. This will likely speed the forward model construction, however, at the time of writing this report, this has not yet been implemented.

Moreover, if one has a an analytical expression for $A(\omega)$ it nmay be possible to acces to an easier way of computing the kernel values.

Higher order approximation

Chapter 2 presents how the Born pressure field could be determined by iterating a recursive scheme:

$$P_B^{(k+1)}(\boldsymbol{x}) = \int G(\boldsymbol{x},\boldsymbol{y})(P_B^k(\boldsymbol{y}) + P_i(\boldsymbol{y}))f(\boldsymbol{y})\mathrm{d}\boldsymbol{y} \tag{4.24}$$

For the kernel development the first order approximation is used *i.e.*, no previous scattering is considered. By adding a second scattering point in the kernel (4.24) becomes:

$$\begin{aligned} P_B^{(k+1)}(\boldsymbol{x}) &= \int G(\boldsymbol{x},\boldsymbol{y})(P_B^k(\boldsymbol{y})+P_i(\boldsymbol{y}))f(\boldsymbol{y})\mathrm{d}\boldsymbol{y} \\ &= \underbrace{\int G(\boldsymbol{x},\boldsymbol{y})(P_B^k(\boldsymbol{y})f(\boldsymbol{y})\mathrm{d}\boldsymbol{y}}_{\tilde{P_B^{(1)}}} + P_B^{(1)}(\boldsymbol{x}) \end{aligned} \tag{4.25}$$

By applying the same development for the first approximation:

$$\delta t = \frac{1}{\omega}\mathrm{Im}\left[\frac{\tilde{P_B^{(1)}}}{P_i}\right] + \int K^{(1)}(\boldsymbol{x}_e,\boldsymbol{x}_r,\omega,\boldsymbol{y})\delta s(\boldsymbol{y})\mathrm{d}\boldsymbol{y} \tag{4.26}$$

With:

$$\frac{1}{\omega}\mathrm{Im}\left[\frac{\tilde{P_B^{(1)}}}{P_i}\right] = \iint \left(\frac{G(\boldsymbol{x}_r,\boldsymbol{y}),G(\boldsymbol{y},\boldsymbol{y}')G(\boldsymbol{y}',\boldsymbol{x}_e)}{G(\boldsymbol{x}_r,\boldsymbol{x}_e)}\right)2\omega^3 s_0^2\delta s(\boldsymbol{y})\delta s(\boldsymbol{y}')\mathrm{d}\boldsymbol{y}\mathrm{d}\boldsymbol{y}' \tag{4.27}$$

Thus, the second order Kernel (for a monochromatic wave) is:

$$K^{(2)}(\boldsymbol{y}) = \frac{\|\boldsymbol{x}_e-\boldsymbol{x}_r\|}{\|\boldsymbol{x}_r-\boldsymbol{y}\|}\int \frac{1}{\|\boldsymbol{y}-\boldsymbol{y}'\|\|\|y'-\boldsymbol{x}_e\|}\sin(\omega s_0\delta l^{(2)}(\boldsymbol{y}'))2\omega^3 s_0^2\delta s(\boldsymbol{y}')\mathrm{d}\boldsymbol{y}' \tag{4.28}$$

So the complete sensitivity kernel is:

$$\delta t = \int K^{(2)}(\boldsymbol{y}) + K^{(1)}(y)\delta s(\boldsymbol{y})\mathrm{d}\boldsymbol{y} \tag{4.29}$$

A test implementation of this second order Kernel has been done, and bring nothing but extra computational time, more study will therefore be needed on this subject to determine the relevance of this higher order approximation. Theoretically such approximation could be pushed further, seeking toward a complete full wave scattering, building a bridge between the Born approximation and the full wave approach.

4.5 Summary

By using the finite frequency tomography, and taking advantages of the Fresnel's zone interference the amount of information collected by a pair of emitter and receiver increases. The rays presented in chapter 3 are now kernels, meeting more pixels on their path. They also put a bridge between the ray based tomography and the full wave approach.

The fat ray developed here does not need a heterogeneous initial guess (like bent rays) to provide good results (see chapter 7). By using a higher order approximation, such hypothesis could be added, however first tests of this methods did not show satisfactory enough results to be shown in this work.

Chapter 5

Tomographic reconstruction methods

This chapter will be a non exhaustive review of the different algorithms that could solve the previously introduced linear inverse problem

$$\boldsymbol{y} = \boldsymbol{M}\boldsymbol{x} \tag{5.1}$$

With as remainder $\boldsymbol{y}$ the time of flight vector, determined by $\boldsymbol{M}$ a $M \times N$ matrix encoding how the M rays (emitter-receiver pairs) are influenced by the slowness vector $\boldsymbol{x}$ representing N pixels of the image.

Hereafter we denote $\mathcal{X}$ the space of the posssible solutions and $\mathcal{Y}$ the space of the possible acquired data.

5.1 Inversion theory

Inverse Problem have been intensely studied in the mathematical field. In particular *well-posed* problem in the sens of Hadamard have the following properties:

- A solution exist $Im(M) = \mathcal{Y}$
- The solution is unique $Ker(M) = 0$
- The solution always depends continuously on the initial conditions (small changes in the data implies small changes in the solution)

The third condition can be described by the condition number of $\boldsymbol{M}$ defined as follow:

$$\xi(\boldsymbol{M}) = \frac{\sigma_{max}(\boldsymbol{M})}{\sigma_{min}(\boldsymbol{M})} \tag{5.2}$$

With $\sigma_{max}(\boldsymbol{M})$ and $\sigma_{min}(\boldsymbol{M})$ the biggest and the smallest singular value of $\boldsymbol{M}$

Problem (5.1) is both a *ill-posed* problem (it is unlikely to have $N = M$) and *ill-conditioned* problem.

Estimation of the conditions number have been done for various combination of pixel resolution and number of rays, yields in USCT case a condition number of $\xi(\boldsymbol{M}) > 10^{10}$. A typical good value should be $\xi(\boldsymbol{M}) < 10$. In other words, it is *very* sensitive to noisy input data.

Solving directly the problem (5.1) if therefore replace with a *surrogate problem* with the following requirements:

- Well-posed
- The solution of the surrogate problem should belong to $\mathcal{X}$ (consistency).
- Computationally feasible.

This problem have often the form:

$$\arg\min_{\boldsymbol{x}} \{f(\boldsymbol{x}, \boldsymbol{y}) + g(\boldsymbol{x})\} \tag{5.3}$$

With $f(\boldsymbol{x}, \boldsymbol{y})$ the data-fidelity term and $g(\boldsymbol{x})$ a prior-knowledge term. Assuming f and g are suitable functions (continuous, convex, possibly differiantiable) and $\mathcal{X}, \mathcal{Y}$ are also convex, the problem will be well-posed.

5.2 Bayesian framework

To compare the various solvers available, let's consider a more realistic problem, where noise is modelled with help of probabilities:

$$\boldsymbol{y} = \boldsymbol{M}\boldsymbol{x} + \boldsymbol{n} \tag{5.4}$$

The noise component $\boldsymbol{n}$ is considered unknown and is thus a random variable: $\boldsymbol{n} \sim p_{\boldsymbol{n}}(\boldsymbol{n})$. In particular $\boldsymbol{n}$ could be choose as an independant white gaussian noise:

$$p_{\boldsymbol{n}}(\boldsymbol{n}) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{n}\|_2^2\right) \tag{5.5}$$

5.2.1 Maximum likelihood

By introducing probabilities, $\boldsymbol{y}$ becomes also a random variable, and is an affine transformation of n:

$$\boldsymbol{y} \sim p_{\boldsymbol{n}}(\boldsymbol{y} - \boldsymbol{M}\boldsymbol{x}|\boldsymbol{x}) \tag{5.6}$$

This is also call the *likelihood* of $\boldsymbol{y}$. Maximizing this likelihood, with respect to $\boldsymbol{x}$ will yields the most probable acquired data. This has for underlying hypothesis that the observation made is not a rare event. The Maximum likelihood problem is then:

$$\boldsymbol{x}_{ML} = \arg\min_{\boldsymbol{x}} \{p_{\boldsymbol{y}}(\boldsymbol{y})\} \iff \arg\min_{\boldsymbol{x}} \left\{\|\boldsymbol{y} - \boldsymbol{M}\boldsymbol{x}\|_2^2\right\} \tag{5.7}$$

Maximizing the Likelihood is equivalent to minimizing the negative log-likelihood. With a Gaussian noise, the Maximum likelihood approach boils down to a classic least square Problem.

Least square solution

This norm minimization problem is then solve (gradient of the solution is null) and give the analytical solution:

$$\boldsymbol{x}_{ML} = (\boldsymbol{M}^T\boldsymbol{M})^{-1}\boldsymbol{M}^T\boldsymbol{y} \tag{5.8}$$

Moreover, the solution $\boldsymbol{x}_{ML}$ is a linear combination of the observation data, and can also be understand as a filtering process, from a tomographic point of view this is understood as the well known *filtered backprojection* (often done in X-ray tomography, and executed in the frequency domain for simplicity).

Correlated Noise

If the noise carry also a correlation between its component, a correlation matrix $\boldsymbol{\Sigma}$ can be added in the mix (matrix of the covariance between each component of $\boldsymbol{n}$). This gives:

$$p_{\boldsymbol{n}}(\boldsymbol{n}) = \frac{1}{\sqrt{|\det(2\pi\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}\boldsymbol{x}^T\boldsymbol{\Sigma}\boldsymbol{x}\right) \tag{5.9}$$

This leads to a solution using a *weighted norm*, hence the weighted least square problem:

$$\arg\min_{\boldsymbol{x}} \{p_{\boldsymbol{y}}(\boldsymbol{y})\} \iff \arg\min_{\boldsymbol{x}} \left\{\|\boldsymbol{y} - \boldsymbol{M}\boldsymbol{x}\|_{\boldsymbol{\Sigma}}^2\right\} \tag{5.10}$$

where $\|\boldsymbol{x}\|_{\Sigma}^2 = \boldsymbol{x}^T\boldsymbol{\Sigma}\boldsymbol{x}$, is the norm weighted by $\boldsymbol{\Sigma}$.

5.2.2 Maximum *a posteriori*

In the maximum likelihood approach, few or nothing is assumed about the properties of $\boldsymbol{x}$. With the maximum *a posteriori*, $\boldsymbol{x}$ is also considered as a random variable, on which one could assume some statistical characteristics, and models it with a probability density function $p_{\boldsymbol{x}}(\boldsymbol{x})$, called the *prior probability* of $\boldsymbol{x}$. This added information can then be used in a bayesian framework, where the likelihood is now dependent of a previously assumed value of $\boldsymbol{x}$, and noted $p_{\boldsymbol{y}}(\boldsymbol{y}|\boldsymbol{x})$. Using the *Bayes Rule*, it gives access to $p_{\boldsymbol{x}}(\boldsymbol{x}|\boldsymbol{y})$:

$$p_{\boldsymbol{x}}(\boldsymbol{x}|\boldsymbol{y}) = \frac{p_{\boldsymbol{x}}(\boldsymbol{y}|\boldsymbol{x})p_{\boldsymbol{x}}(\boldsymbol{x})}{p_{\boldsymbol{y}}(\boldsymbol{y})} \tag{5.11}$$

This quantity is the *posteriori* probability of $\boldsymbol{x}$. In other words, what is the probability of $\boldsymbol{x}$ to yields the measured output data $\boldsymbol{y}$? The most probable value is then the Maximum a posteriori.

The solution to the *Maximum a posteriori* is then:

$$\boldsymbol{x}_{MAP} = \arg\min_{\boldsymbol{x}} \{p_{\boldsymbol{x}}(\boldsymbol{x}|\boldsymbol{y})\} = \arg\min_{\boldsymbol{x}} \{-\log(p_{\boldsymbol{y}}(\boldsymbol{y}|\boldsymbol{x})) - \log(p_{\boldsymbol{x}}(\boldsymbol{x}))\} \tag{5.12}$$

Again, the negative log-likelihood is considered, and the term $p_{\boldsymbol{y}}(\boldsymbol{y})$ is discarded, as it does not depends on $\boldsymbol{x}$.

For example, if $\boldsymbol{x}$ is assumed to be normally distributed:

$$\boldsymbol{x}_{MAP} = \arg\min_{\boldsymbol{x}} \left\{\frac{1}{2\sigma_n^2}\|\boldsymbol{y} - \boldsymbol{M}\boldsymbol{x}\|_2^2 + \frac{1}{2\sigma_x^2}\|\boldsymbol{x}\|_2^2\right\} \tag{5.13}$$

In the same way, if $\boldsymbol{x}$ follow a Laplace law $p_{\boldsymbol{x}}(\boldsymbol{x}) = \frac{\lambda}{2}\exp(-\lambda\|\boldsymbol{x}\|_1)$:

$$\boldsymbol{x}_{MAP} = \arg\min_{\boldsymbol{x}} \left\{\frac{1}{2\sigma_n^2}\|\boldsymbol{y} - \boldsymbol{M}\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|_1\right\} \tag{5.14}$$

5.3 Iterative reconstruction

In the case of the least square problem, the analytical expression of the solution is $\boldsymbol{x}_{ML} = (\boldsymbol{M}^T\boldsymbol{M})^{-1}\boldsymbol{M}^T\boldsymbol{y}$; However, it is not feasible to compute this expression directly. The matrix $\boldsymbol{M}^T\boldsymbol{M}$ is too huge to be stored in memory, and even if it was sparse, its inverse is not. Therefore, an iterative reconstruction is necessary.

There exists many different iteratives algorithms with various names (ART, SART, POCS, EM, CGLS, *etc.*...). Some of them are mostly the same and tweaks to be interpreted as an other one. In such a wide field to explore and test for the USCT project, this report will focused on the Algebraic Reconstruction Technique (ART) familly of algorithms.

5.3.1 Algebraic Reconstruction Technique

The first Algebraic Reconstruction Technique (ART) have been introduced by Kaczmarz [33] and develop by Gordon [34]. The following explanation is based on [35] This method directly try to solve the inverse problem $\boldsymbol{y} = \boldsymbol{Mx}$, without any assumption on the noise. To explain the solving processus of ART let's rewrite (5.1) as:

$$\begin{cases} m_{11}x_1 + m_{12}x_2 + \cdots & +m_{1N}x_N = y_1 \\ m_{21}x_1 + m_{22}x_2 + \cdots & +m_{2N}x_N = y_2 \\ \vdots & \vdots \\ m_{N1}x_1 + m_{N2}x_2 + \cdots & +m_{NN}x_N = y_N \end{cases} \tag{5.15}$$

The image to reconstruct can be considered as a single point in a N-dimensional space. Each equation of (5.15) is thus a hyperplan[1] in the image space.

Knowing an estimate $\boldsymbol{x}^{(k)}$of $\boldsymbol{x}$ (from an initial Guess or a previous reconstruction) one can determine a new estimate by projecting successively on the i_k-th row of $\boldsymbol{M}$, giving a intermediate estimation $\boldsymbol{x}^{(k)}$. With the initialisation: $\boldsymbol{x}^{(k,0)} = x^{(k-1)}$, the process can then be described as follow:

$$\boxed{\boldsymbol{x}^{(k)} = \boldsymbol{x}^{(k)} + \frac{\boldsymbol{m}_{i_k}}{\|\boldsymbol{m}_{i_k}\|_2^2}\left(y_{i_k} - \boldsymbol{x}^{(k)} \cdot \boldsymbol{m}_{i_k}\right)} \tag{5.16}$$

The index of row i_k is made dependent of the iteration. The naive approach would be to use the natural indexing $i_k = k \mod M$, by cycling on the rows of M. The choice of the indexing method can greatly influence the convergence speed of the reconstruction.

5.3.2 Simultaneous Algebraic Reconstruction Technique

The Simultaneous Algebraic Reconstruction Technique (SART) unlike ART does not update $\boldsymbol{x}$ at each row iteration, and its update rule can therefore easily be parallelized. In the Literature the SART algorithm is sometimes also call SIRT (especially in the CT community), where SIRT denote in fact a wider family of algorithms, based on Landweber methods [36] and can also be understood as Richardson's iteration scheme[37]. The `AIRToolBox` [38] provides implementations of this algorithms.

The SART update scheme can be summarize in the following equation:

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \boldsymbol{CM}^T\boldsymbol{R}(\boldsymbol{y} - \boldsymbol{Mx}^{(k)}) \tag{5.17}$$

With $\boldsymbol{C}$ and $\boldsymbol{R}$ diagonal matrices use for the ponderation:

$$\begin{cases} c_{jj} = \left(\sum_{i=1}^{M} |M_{i,j}|\right)^{-1} \\ r_{jj} = \left(\sum_{j=1}^{N} |M_{i,j}|\right)^{-1} \end{cases} \tag{5.18}$$

Furthermore, these expressions explicitly declare that none of the rows are empty (every acquired data y_i as a forward model) and also that none of the columns are empty (every voxel is observable).

In comparison to ART, SART add information about an inherant noise, and its correlation matrix $\boldsymbol{\Sigma} = \boldsymbol{R}^{-1}$, as presented in section 5.2.1. This make the SART

---

[1] NB: the terms hyperplans and equations are use synonymously

method less sensible to noise, but can also reduce the convergence speed. In the same way, $\boldsymbol{C}$ can be understand as a pre-conditioner applied on the system. In practise, this conditioner could also be set to $\boldsymbol{C} = \rho \boldsymbol{I}$. with $\rho = \frac{1}{\|M\|_1}$, this is known as the PSIRT method, and is usefull in the Ordered Subset scheme [39].

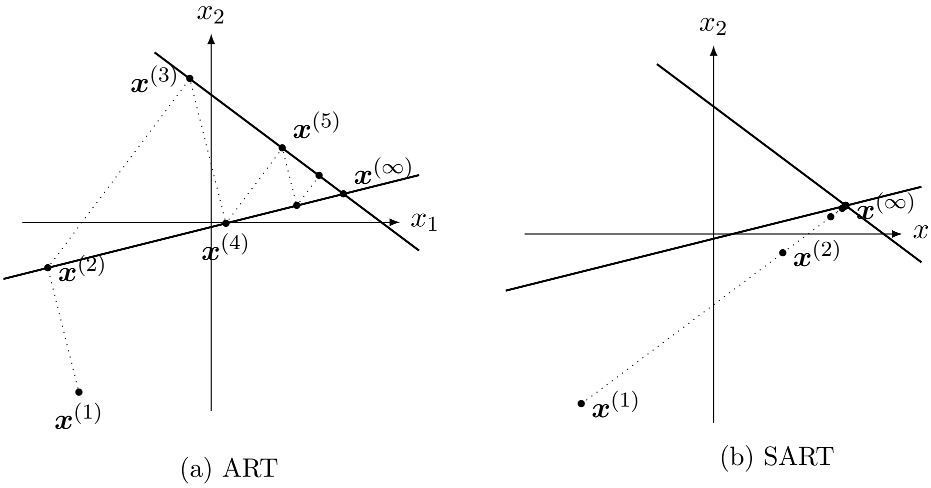
With more equations than dimensions to reconstruct ($M > N$), and the projections $y_1, \cdots, y_M$ are corrupted by noise, no unique solution exists in this case. The estimation of the solution will oscillates in the neighborhood of the intersections of the hyperplans.

With less equations than dimensions to reconstruct ($M < N$), there is no unique solution, and the reconstruction result depends on the initialization. Tanabee [40] has shown that in this case, the solution that the solution converge towards $\boldsymbol{x}^{(s)}$ such that $\|x^{(0)} - x^{(s)}\|$ is minimized.

5.3.3 SART and ART: comparison and improvement

Example If the problem with only 2 dimension, like for (5.19), it easily represent the reconstruction graphically on Figure 5.1a.

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 &= y_1 \\ a_{21}x_1 + a_{22}x_2 &= y_2 \end{aligned} \tag{5.19}$$



(a) ART

(b) SART

Figure 5.1: Algebraic Solver reconstruction with N=2.the SART method clearly converges in less iteration than ART. Furthermore the first projection of the ART (a) could have been on the other hyperplan, where the estimate would have been closer to the solution (roughly to $\boldsymbol{x}^{(5)}$ position), and the convergence would have been faster.

Orthogonality and Convergence speed

In the ART method, If the hyperplans are perpendicular to each other, then the methods converges in M steps, for any initial guess point. That will be one iteration of SART. On the other hand, if hyperplan have only a small angle between them, the number of iterations will be large to attain the solution. One could think of firstly orthogonalizing the system (with the Graham Schmitt procedure for example), but this become too expensive with higher dimensions. Besides, orthogonalization tends to enhance the effect of noise.

However, a few optimisation can be made: Ramakrishnan *et al.* [41] proposed a pair-wise orthogonalization of the equations. A simpler approach is to use equations (ie rays) that have a wider angle between them. Introducing completely new/different information at each step. Indeed, two geometrically close rows are prone to share common information as they probe the same area. Choosing widely separated rays, will then improve the convergence rate of the reconstruction.

Such optimization can also be made for the SART method, and have leads to used Ordered Subset[2] (OS) for speed and quality improvement.

Relaxation

Both ART and SART method can be improved by the introduction of a relaxation parameter:

$$\boldsymbol{x}^{(k+1)} = \boldsymbol{x}^{(k)} + \lambda_k \boldsymbol{C}\boldsymbol{M}^T\boldsymbol{R}(\boldsymbol{y} - \boldsymbol{M}\boldsymbol{x}^{(k)}) \tag{5.20}$$

The relaxation parameter λ_k can be change between iteration, and the search of an optimal one (improving both reconstruction time and quality of the result) is not an easy task. Ensuring convergence leads to the constraint: $0 <$
$lambda_k < 2$. Among the different strategies used to find such relaxation parameters, so called *line search* and *Diminish step-size* presented in the AIR Toolbox [42].

5.4 Regularization

As presented at the beginning of this chapter, the inverse problem $\boldsymbol{yMx}$ is is *ill-posed* and also *ill-conditioned.* If using the surrogate problem gives acces to a well-posed problem, it is still often ill-conditioned. The goal of regularization is thus to tackle this situation by adding extra (*i.e.* prior) information to the problem, and by so computing a *Maximum a posteriori* solution.

5.4.1 Thikhonov Regularisation

The simplest regularization method is known as the *Thikonov* regularization.

$$\arg\min_x \left\{ f(\boldsymbol{x}, \boldsymbol{y}) + \|\boldsymbol{D}\boldsymbol{x}\|^2_{\boldsymbol{\Gamma}} \right\} \tag{5.21}$$

where $\|\boldsymbol{D}\boldsymbol{x}\|^2_{\boldsymbol{\Gamma}}$ is the prior knowledge term, here meaning that within a certain base[3] $\boldsymbol{D}$ the signal $\boldsymbol{D}\boldsymbol{x}$ is normally distributed with a covariance matrix $\boldsymbol{\Gamma}$.

often the regularisation matrix is assumed diagonal ($\boldsymbol{\Gamma} = \lambda\boldsymbol{I}$, and $\lambda > 0$), and the problem becomes :

$$\arg\min_x \left\{ f(\boldsymbol{x}, \boldsymbol{y}) + \lambda\|\boldsymbol{D}\boldsymbol{x}\|^2 \right\} \tag{5.22}$$

In practice the Thikonov Regularization guarantees somme kind of smoothness on the solution, with (5.22) large component of $\boldsymbol{D}\boldsymbol{x}$, also called *outlier* will be penalized, and thus not appear in the final solution.

---

[2] also known as Block Iterative (BI)

[3] or transform, in a analytic framework

5.4.2 Choice of regularization base

The choice of the matrix $\boldsymbol{D}$ is a major step in the modelling of the problem, as it encode the prior information available on the solution.

The simplest regularisation is by setting: $\boldsymbol{D} = \boldsymbol{I}$. This will guarantee a smallest norm $\boldsymbol{x}$, which could be interpreted as the less energetic solution, and thus the most natural one. This is of great advantages with an underdetermined system, where multiple solution exist and only one should be choosen.

An other popular choice is $\boldsymbol{D} = \boldsymbol{\nabla}$ the spatial gradient. In this case, few abrupt variations (or discontinuity) can be enhance, and small but numerous noisy variation can be smeared out. This results in more piece-wise image, and in the case of tomography help to magnify the heterogeneity of the medium.

One could also use Discrete Fourier Transform (DFT) or Discrete Wavelet Transform (DWT) and encode prior information in this domain $\boldsymbol{D} = \boldsymbol{W}$. In particular, Wavelet transform can leads to similar result as a gradient based regularisation, but also give more possibilities of tuning. However the need of simple and fast to implement solution for this topic makes the exhaustive studies of these possibilities out of the scope of this report. A deeper analysis of the impact of regularization onto Iterative reconstruction can be found in [43].

Norm Choice

In addition to the choice of the base of expression of the prior $\boldsymbol{D}$, the choice of the regularization norm is also of great interest. For example the used of the classic ℓ_2 is often associated with a normally distributed associated noise, and have the good taste to be mathematically easy to manipulate, and have homogeneous behaviour. More Recently, the used of the ℓ_1 norm became widespread in the literature. In the bayesian framework this is associated with a Laplace distribution in the $\boldsymbol{D}$ basis. The ℓ_1 regularisation is less sensitive to outlier values (being not squared up), however its non differentiability makes it impossible to gives explicit formula of the solution.

5.4.3 Total Variation

In combination with $\boldsymbol{D} = \boldsymbol{\nabla}$ the use of the ℓ_1 is known as the *Total Variation* (TV) regularization producing the following problem:

$$\arg\min_{x} \left\{ \|\boldsymbol{y} - \boldsymbol{M}\boldsymbol{x}\|_2^2 + \mu \|\boldsymbol{\nabla}\boldsymbol{x}\|_1 \right\} \tag{5.23}$$

Total Variation has been successfully used in the image reconstruction and inverse problem fields. This has been theorized in the *Compress Sensing* framework [44]. In practice, Total Variation (and more generally ℓ_1 regularisation) encourage some kind of sparsity in the solution in basis $\boldsymbol{D}$. In the discrete 2D and 3D cases, the computation of the total variation can be done in two ways:

- The *anisotropic* total variation:

$$TV_{aniso}(\boldsymbol{x}) = \sum_{i,j} |g_{i,j}^{(1)}| + |g_{i,j}^{(2)}| + |g_{i,j}^{(3)}|$$

- The *isotropic* total variation:

$$TV_{iso}(\boldsymbol{x}) = \sum_{i,j} \sqrt{(g_{i,j}^{(1)})^2 + (g_{i,j}^{(2)})^2 + (g_{i,j}^{(3)})^2}$$

The use of an isotropic or anisotropic TV have an impact on the reconstructed image, enforcing (or not) the isotropism of the medium, in particular, isotropic TV promotes more circle/sphere like shape (*i.e.* corner less) whereas the anisotropic TV is prone to yields square like shape, oriented in the gradients axis. The higher the regularisation parameter μ is, the more this effect can be seen. Moreover, the sparsity of the gradient produces a piece-wise reconstruction image, and this property will be even more visible with stronger regularization.

The hypothesis of a piece wised image make sense in the USCT case, as the aperture medium can be separated in different parts: water background, fatty tissues, glandular tissues, tumours, etc... However, assuming that it an absolute truth should be avoided: In the tissues above, there should be some variations (but in a smaller range than between two categories of tissues). A too strong regularization will unfortunately neglect these variations.

5.5 Total Variation Augmented Lagrangian (TVAL)

5.5.1 An Optimisation Problem

Considering that the gradient of $\boldsymbol{x}$ is sparse, the problem to solve can be expressed as:

$$\min \|\boldsymbol{D_x x}\|_1 + \|\boldsymbol{D_y x}\|_1 + \|\boldsymbol{D_z x}\|_1 \text{ with } \boldsymbol{y} = \boldsymbol{Mx} \tag{5.24}$$

Where $\boldsymbol{D_x}, \boldsymbol{D_y}, \boldsymbol{D_z}$ are the discrete gradient operator in the corresponding direction. Hereafter $\boldsymbol{w}_i = \boldsymbol{D}_i \boldsymbol{x}$.

(5.24) is an optimisation problem, that can be solve by introducing the Lagrangian operator:

$$\begin{aligned}\mathscr{L}(\beta, \mu, \boldsymbol{\lambda}, \boldsymbol{\nu}) = & \underbrace{\sum_i \left(\|\boldsymbol{w}_i\| - \nu_i^T (\boldsymbol{D_i x} - \boldsymbol{w}_i) + \frac{\beta}{2} \|D_i \boldsymbol{x} - \boldsymbol{w}_i\|_2^2 \right)}_{\text{w-problem}} \\ & \underbrace{- \boldsymbol{\lambda}^T (\boldsymbol{Mx} - \boldsymbol{b}) + \frac{\mu}{2} \|\boldsymbol{Mx} - \boldsymbol{b}\|_2^2}_{\text{x-problem}}\end{aligned} \tag{5.25}$$

The introduction of $\boldsymbol{w}$ is mandatory as the TV minimisation can not be differentiated. and one needs to solve the w-problem, which is also a optimisation problem. It also guarantees a bounded solution[45].

5.5.2 The algorithm

In an interative way, compute $\boldsymbol{w}_{k+1}$, $\boldsymbol{x}_{k+1}$ and update the Lagrangian multiplicators.

$$\|\boldsymbol{w}_{i,k+1}\| - \nu_i^T \left(D_i \boldsymbol{x}_k - \boldsymbol{w}_{i,k+1} \right) + \frac{\beta}{2} \left\| D_i \boldsymbol{x}_k - \boldsymbol{w}_{i,k+1} \right\|_2^2 \tag{5.26}$$

(5.26) is the case of the l_1 norm (5.26) can be explicitly solve:

$$\boldsymbol{w}_{i,k+1} = \max\left\{\left|D_i\boldsymbol{x}_k - \frac{\nu_i}{\beta_i}\right| - \frac{1}{\beta_i}, 0\right\} \operatorname{sgn}\left(D_i\boldsymbol{x}_k - \frac{\nu_i}{\beta_i}\right) \tag{5.27}$$

the proof and the extended solution in l_2-norm case can be found in [45]. Knowing the value $\boldsymbol{w}_k$ given by (5.25) $\boldsymbol{x}_k$ can be determined, coming from zeroing its gradient.

$$\begin{aligned}\boldsymbol{x}_{k+1} = &\left(\sum_i \beta_i \boldsymbol{D_i}^T \boldsymbol{D_i} + \mu \boldsymbol{M}^T\boldsymbol{M}\right)^{\dagger} \\ &\times \left(\sum_i \left(\boldsymbol{D_i}^T v_i + \beta \boldsymbol{D_i}^T \boldsymbol{w}_{i,k+1}\right) + \boldsymbol{M}^T\lambda + \mu\boldsymbol{M}^T\boldsymbol{b}\right)\end{aligned} \tag{5.28}$$

Where $()^{\dagger}$ is the Moore-Penrose Pseudoinverse. Unfortunately, computing such inverse at each iteration is too costly to be numerically implemented, to solve such problem the one-step steepest descent method is choosen. The first iteration of the steepest descent may not be very close to the real solution, but the alternatively solving of $\boldsymbol{w}$ and $\boldsymbol{x}$ gives a converging solution.

Implementation

A lot of effort has already been put in the implementation of the TVAL3 algorithm, going from a `CPU-MATLAB-double` version (around 80s of compute time) to a `GPU-C++-float` version (<10s of compute time). More information about this optimisation can be found in [46].

5.5.3 A Bayesian Explaination to TVAL

The following try to give a meaning to the regularisation parameters introduced in the TVAL3 algorithm.

Let consider the inverse problem disturbed by $\boldsymbol{n}$ a Gaussian noise of law $\mathcal{N}(0, \sigma_n)$:

$$\boldsymbol{y} = \boldsymbol{M}\boldsymbol{x} + \boldsymbol{n} \tag{5.29}$$

Then with a probabilistic approach to the problem:

$$p(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{(2\pi\sigma)^{N/2}} \exp\left(-\frac{\|\boldsymbol{y} - \boldsymbol{M}\boldsymbol{x}\|}{2\sigma_n}\right) \tag{5.30}$$

With the assumption of the sparsity of the gradient $\boldsymbol{w} = \boldsymbol{D}\boldsymbol{x}$. Where $\boldsymbol{D}$ is the differential operator in 3D, $\boldsymbol{D}$ is a $3N \times N$ linear operator, but not invertible.

In terms of probability the sparsity of $\boldsymbol{w}$ can be express by using an exponential law of parameter $\mu_{\boldsymbol{w}}$.

$$p(\boldsymbol{w}) = \exp(-\mu_{\boldsymbol{w}}\|\boldsymbol{w}\|_1) \tag{5.31}$$

Knowing $\boldsymbol{w}$, $\boldsymbol{x}$ can be determined by using the adjoint operator of $\boldsymbol{D}$.

$$\boldsymbol{x} = \boldsymbol{D}^{\dagger}\boldsymbol{w} + \bar{\boldsymbol{x}} \tag{5.32}$$

With $\bar{\boldsymbol{x}}$ the unknown mean value of x. Without loss of generality let $\bar{\boldsymbol{x}}$ follow a

Gaussian law $\mathcal{N}(\boldsymbol{\mu_x}, \sigma_x)$:

$$\begin{aligned}
p(\boldsymbol{x}|\boldsymbol{w}) &= \frac{1}{(2\pi\sigma)^{3N/2}} \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{D}^\dagger \boldsymbol{w} - \boldsymbol{\mu_x}\|_2^2}{2\sigma_x}\right) \\
&= \frac{1}{(2\pi\sigma)^{3N/2}} \exp\left(-\frac{\|\boldsymbol{Dx} - \boldsymbol{w} - \boldsymbol{D\mu_x}\|_2^2}{2\sigma_x}\right) \\
&= \frac{1}{(2\pi\sigma)^{3N/2}} \exp\left(-\frac{\|\boldsymbol{Dx} - \boldsymbol{w}\|_2^2}{2\sigma_x}\right)
\end{aligned} \tag{5.33}$$

Using (5.30),(5.32) and (5.33):

$$\begin{aligned}
p(\boldsymbol{y}) &= p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x}|\boldsymbol{w})p(\boldsymbol{w}) \\
&\propto \exp\left(-\underbrace{\left(\frac{\|\boldsymbol{y} - \boldsymbol{Mx}\|}{2\sigma_n} + \frac{\|\boldsymbol{Dx} - \boldsymbol{w}\|_2^2}{2\sigma_x} + \mu_{\boldsymbol{w}}\|\boldsymbol{w}\|_1\right)}_{J(\boldsymbol{x})}\right)
\end{aligned} \tag{5.34}$$

5.6 Summary

The inverse problem modelized in the previous chapter cannot be solve by a direct inversion, and needs an interative method. The SART method has been choose for its robustness and its wide use in the literature. To counteract the noise sensibility of this method, some previous knowledge is introduced by using a Total Variation regularisation. In practice the two method are run successively. the USCT project has also been relying on the TVAL3 algorithm, where the TV regularisation is done during the reconstruction.

SART and an isolated TV regularisation are both well known methods in the literature, and there already exists efficient implementation of them, and strategies to fine tune their parameters

Chapter 6

Numerical Analysis

In chapter 3 we investigated the use of a forward linear operator (our ray Matrix) and how to fill it. The overall reconstruction process can be viewed as an iterative scheme, summarized by Figure 3.8.

This Chapter focused on a numerical analysis of the elaborated forward model , and how it is influenced by external parameters. These insights will later help to choose a suitable solver for the reconstruction, and also shows the need for a regularisation procedure in the solver.

Without further information, reconstruction presented here will be done with the standart SART solver, and the reconstruction will be compare in term of SNR (see (6.1)) to a groundtruth obtained by scaling down the breast phantom used for the simulation to the reconstruction resolution (128x128)

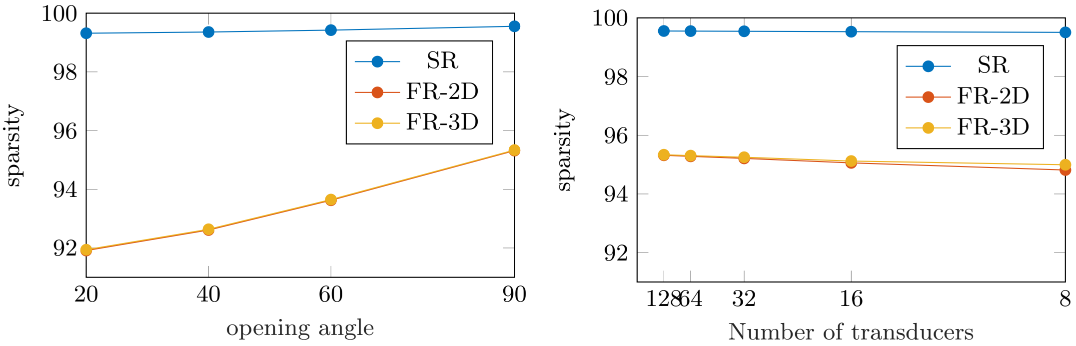
$$SNR = \sqrt{\frac{\sum_k A_k^2}{\sum_k (A_k - B_k)^2}} \tag{6.1}$$

where A is the ground-truth image and B the reconstructed one.

6.1 Matrix construction

6.1.1 Sparsity

A good reconstruction via a solver (SART or TVAL) is only possible if the matrix provided is a "good" matrix. For characterizing it different estimators can be used about the information encoded in this matrix. Straight ray and bent ray have similar sparsity, and only the case of straight ray are represented on Figure 6.1 to 6.3.



(a) Influence of the opening angle (b) Influence of the number of transducer

Figure 6.1: Sparsity of Matrix in straight ray (SR) and fat ray from Fréchet Kernel in 2D (FR-2D) and 3D (FR-3D). Sparsity is defined as the percentage of null value inside the matrix.

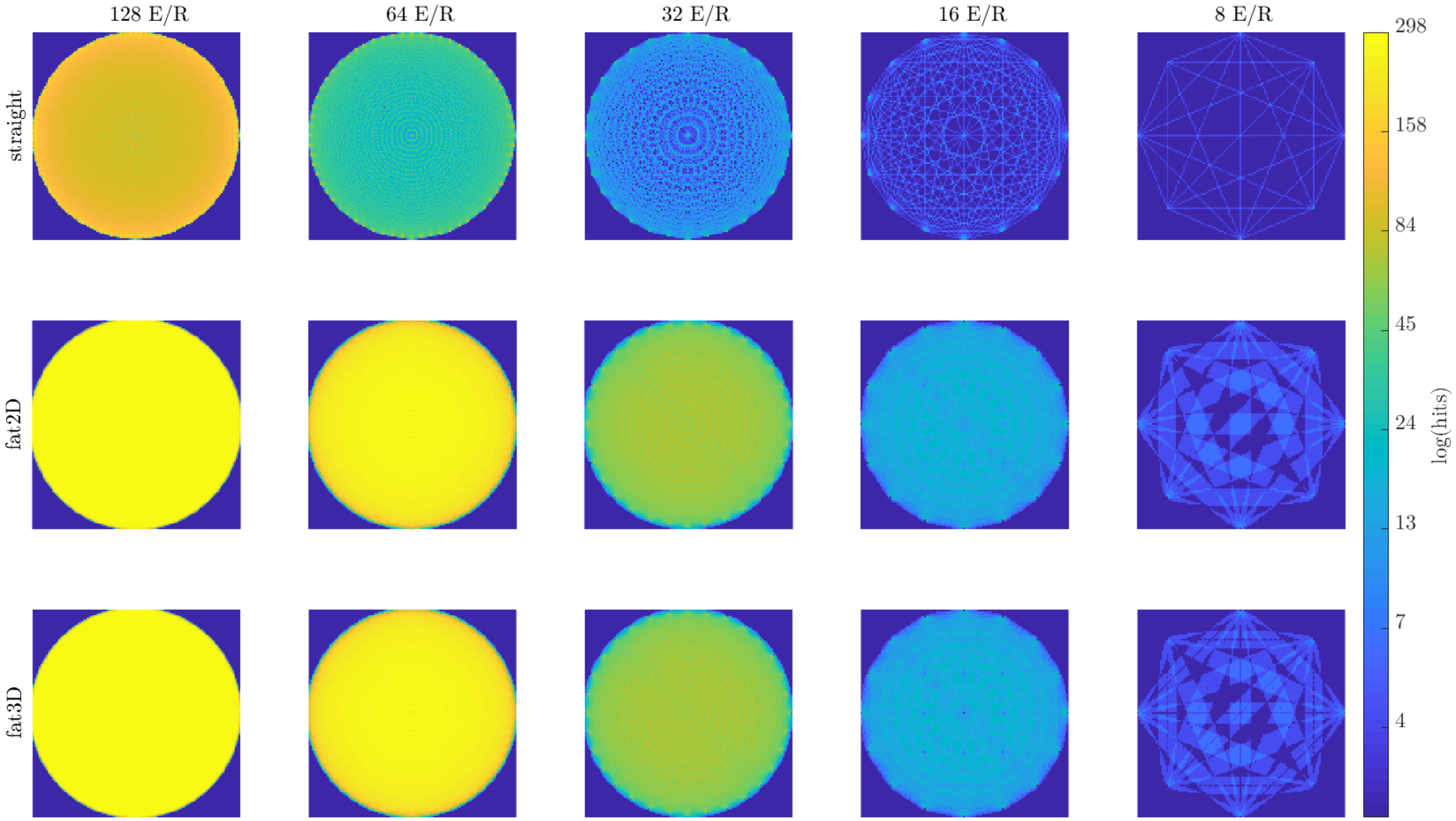


Figure 6.2: The use of fat ray greatly reduces the need of emitter, and increases the amount of information gathered per pixels. The highest hits per ray are achieved by the 2D Kernel. The reconstruction resolution is 128x128 px.

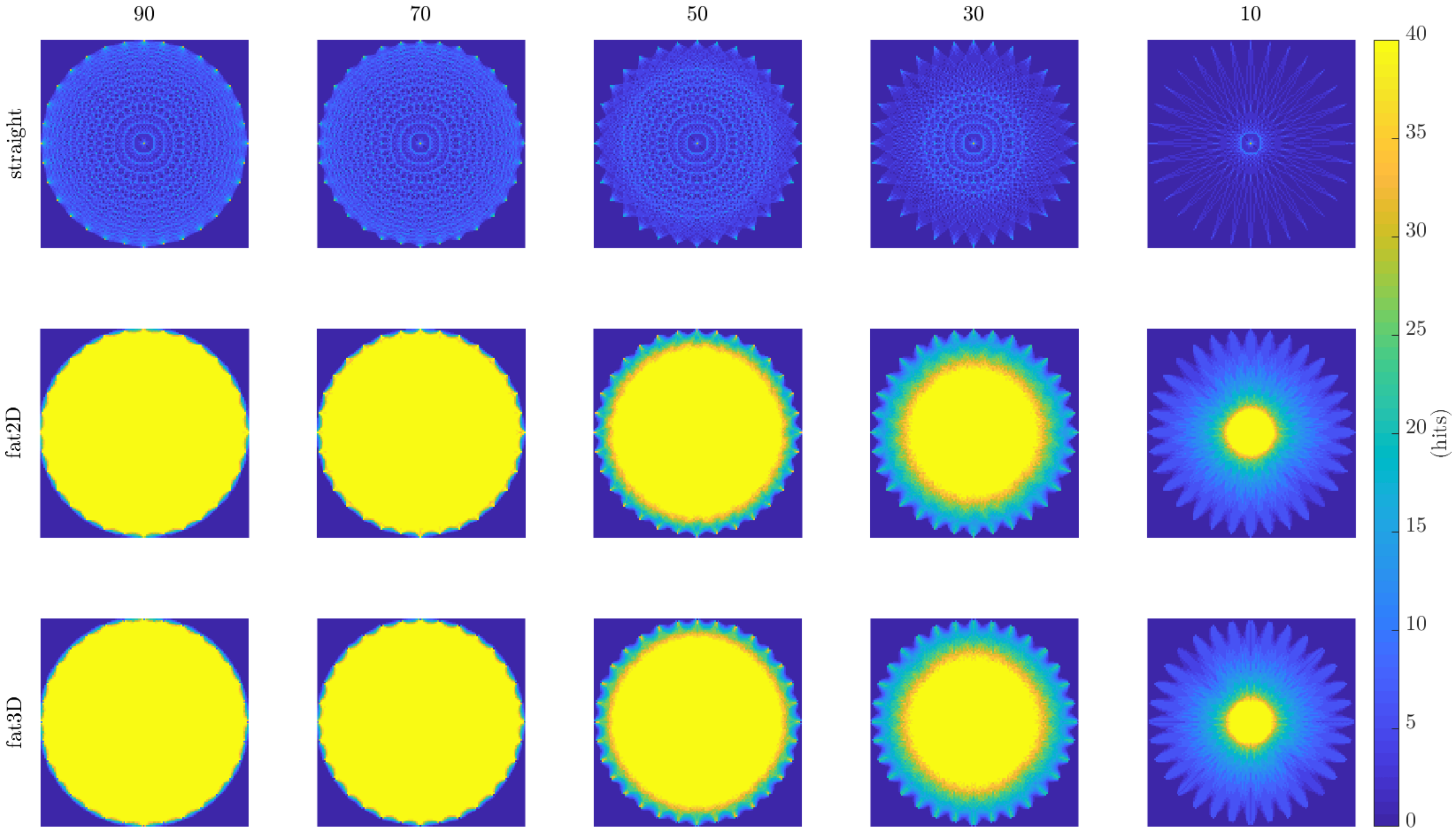


Figure 6.3: The use of fat ray greatly reduces the impact of the opening angle, and increases the amount of information gathered per pixels. For small opening angle, the overlapping area for fat ray is significantly broader.

As it could be foreseen, the use of fat ray kernel (derived from a linear approximation or Fréchet based) greatly reduces the sparsity of the matrix. 5 times more pixels are hit by a fat ray (Figure 6.1).

Both the opening angle and the number of transducers reduce the number of rays considered probing the reconstruction area. On Figure 6.1a, the opening angle has a significant impact on the sparsity of fat ray based matrix, this effect also happen with straight rays, but is less visible, as each straight ray does not bring as much pixel hits as for. With a small opening angle, the number of emitter-receiver pair is reduced, and thus, every fat ray account proportionally for a larger part of the matrix. the same result on a small scale is visible on Figure 6.1b.

Ultimately it can be say that the opening angle has an important influence on the sparsity of the matrix, and will thus have after-effect on the reconstructed image.

6.1.2 Conditioning of M

The dimension of M and its sparsity makes it difficult to evaluate its condition number (computing all its non zeros singular values can take several hours on a computing server). In addition the singular value are slowly decaying after the first hundreds ones, making the use of a classic SVD solver questionable. We can not early cut the decomposition, as it will remove too much energy (or information) for giving a suitable results, see Figure 6.4. Roughly the last half of the singular values are very close to zero (or equal to, with numerical approximation). The use of SVD-based solver will likely not produced good results.

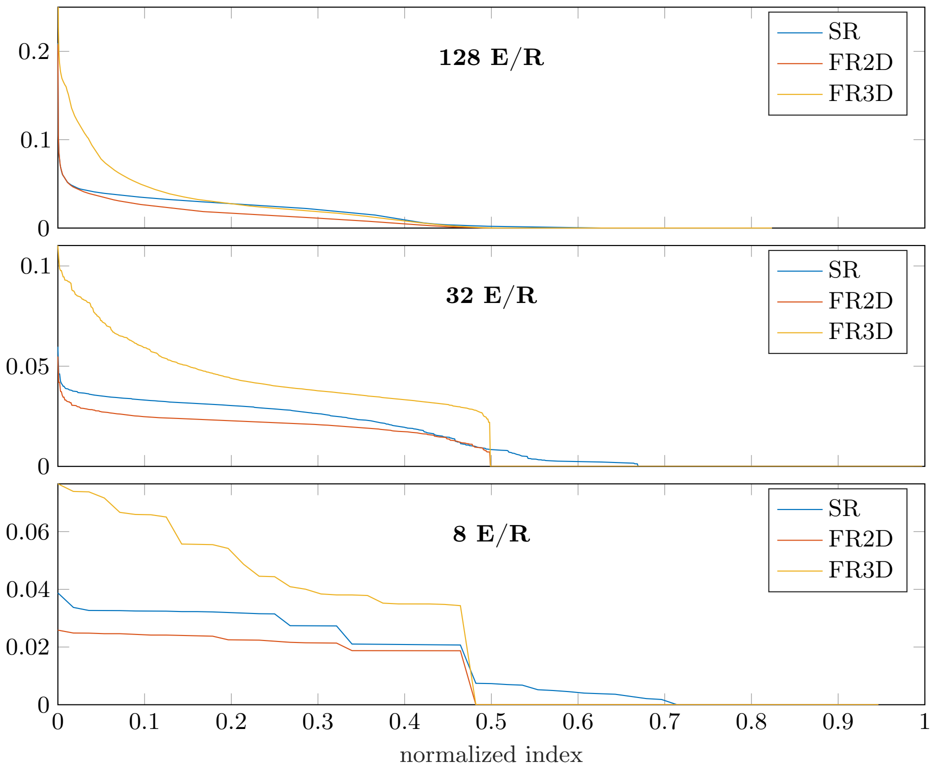


Figure 6.4: Singular Value for different number of transducers (reduce the number of row of the matrix). The normalized index is the index of the value divide by the total number of row.

6.2 Transducer limitations

The use of the SART solver allows to easily compare the influence of the main transducer limitations: their opening angle and their number.

The choice of the regularisation parameters is left to the `AIR II` toolbox, yielding $\lambda \simeq 1.9$ in most cases. No ROI limitation of the computation have been used.

For each situation, solutions with straight rays, bent rays (based on the straight ray approximation), fat rays (independent of previous iteration) have been computed. The use of the signal-to-noise ratio (SNR) provides the largest range of value over the simulation, making it a good candidate to characterise theses solutions.

6.2.1 Opening angle

On Figure 6.5 the Signal-to-noise ratio highly increases for opening angle larger than 30°. With such angles, the breast is completely covered in an area where most of the ray passed, increasing the amount of information gathered on it, and so the quality of the reconstructed image. But having a too wide opening angle also increases reception of noisy data, which will decrease the overall SNR.

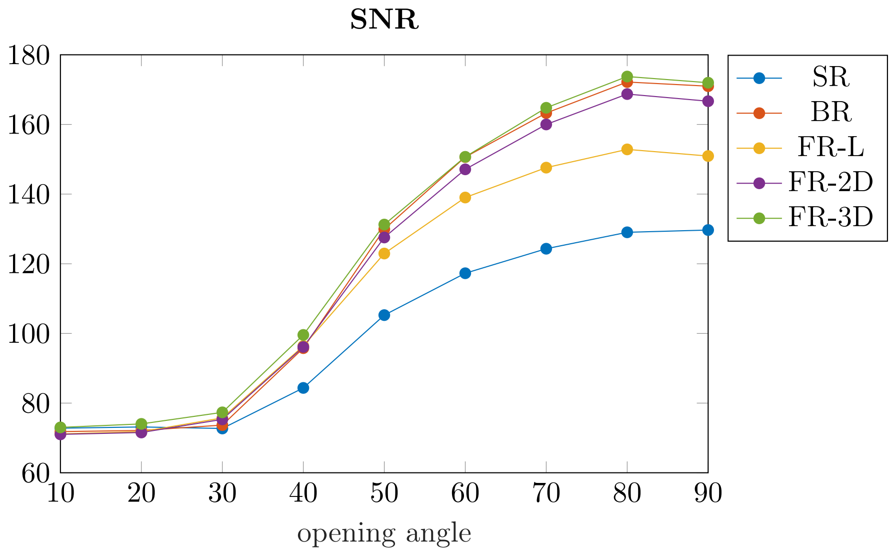


Figure 6.5: Having a wider opening angle create a wider area of overlapping rays, when the full breast is inside this area (starting at 40°). Having a very wide opening angle (over 60°) nearly improve the image, as it adds more information (potentially noisy) about the background area. BR: Bent ray, FR-L: linear interpolated fat ray

6.2.2 Number of transducers

The effect of different number of transducers on the various rays methods used is shown on Figure 6.6, the number of ray is proportional to the square of the number of transducers, and explains the shape of these plots. However, the use of fat ray tones down this effect, and this highlight the main advantage of fat ray, where the impact of a few transducers is decrease, because a single ray carry more information about pixels, as shown on Figure 6.2.

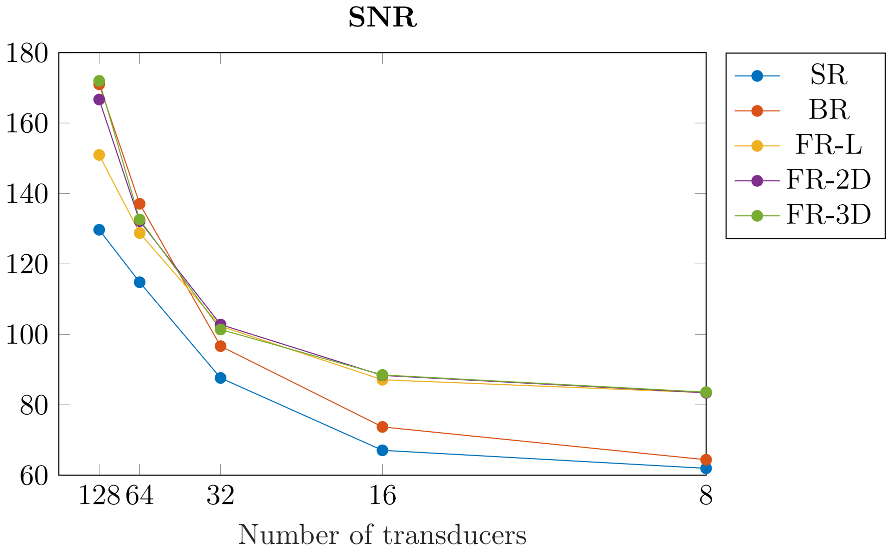


Figure 6.6: Influence of the number of transducer use for the reconstruction, with a totally open angle (90°).

6.3 Resolution Size

Along with the Number of rays (*i.e.* number of transducer squared), the number of pixel to reconstruct is the other main parameter influencing the problem size and the quality of the results.

Furthermore, the conception of the forward model, raised awareness on the smallest pixel size possible, while respecting Shannon's Theorem. In practice, we should not try to reconstruct too small pixel, which will increased the sensitivity to noise and the overall computing time. On the other hand, the goal of USCT is to detect tumors in their early stages, and so a small grid is mandatory.

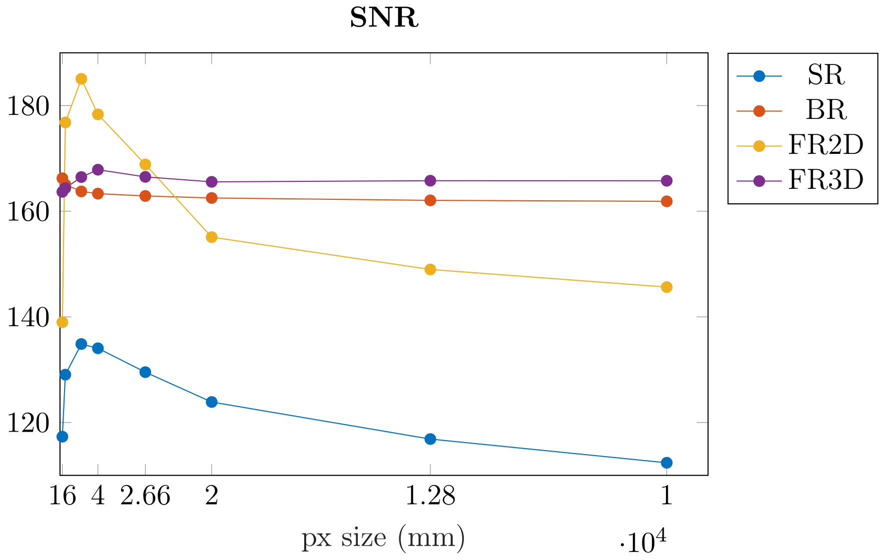


Figure 6.7: Influence of the pixel size used for the reconstruction for different number rays.

Figure 6.7 gives insight on two aspect of the reconstruction: Firstly, by using big pixels, the SNR resulting from the comparison with the downsampled version of the groundtruth (or respectively by upsampling the reconstruction) is higher for small resolution grid. Indeed, the implicit averaging remove noises, however, this also mean we don't have access to fine details of the image.

Then, after a pixel size smaller than 1 or 2 mm the SNR remains constant. Thus there is no need to look after the smallest pixel size computationable. In practice, the standard 128 × 128 grid (2 mm pixel width) is good enough and provides a good compromise between quality of the result and computational time.

6.4 Construction time

If the algorithms developed here are not really time critical, the future use of its in a medical device and the will of not waiting too long during the development process drive the search of a time efficient algorithm, with a low order complexity, and making good use of array parallelisation. Furthermore, in a more business approach, reducing the computing power and memory usage will result into a cheaper device. With such constrains, the goal is still to achieve the best reconstruction possible.

As shown in chapter 3, for each reconstruction based on a ray model (Straight, bent, fat, *etc.*) the overall process can be divide two parts: the matrix construction (forward modelling) and the solver reconstruction (inversion).

Moreover, the matrix reconstruction time is clearly dependent on the problem size (*i.e.* number of ray and number of pixels to reconstruct).

6.4.1 Resolution grid

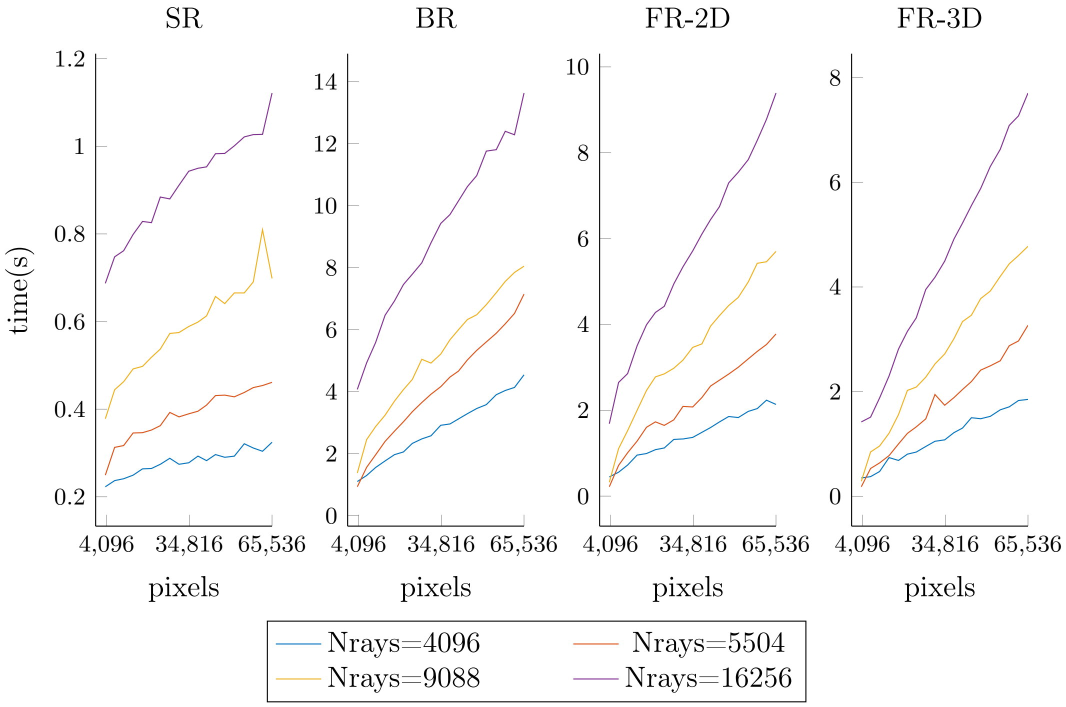


Figure 6.8: Time need to complete the forward modelling (matrix building) for different amount of pixels.

Figure 6.8 indicates a time proportional to the total number of pixel for every ray type. However, the straight rays construction is more than 10 faster. This is result is the consequence of the simplicity of the method, and its impletementation, relying on the bresenham line's algorithm, done in `C` and compiled as a `mex` function. The sole reconstruction of bent ray (without considering the previous necessary straight ray reconstruction) is slower than the fatrays construction. Furthermore, the 3D fat ray kernel, having a slighly simpler expression than the 2D one is faster.

By implementing the fatray methods in `C` and by using a multi-threaded algorithm (each ray Kernel is independent), one could hope to achieve similar construction time than the straight ray one.

6.4.2 Number of ray

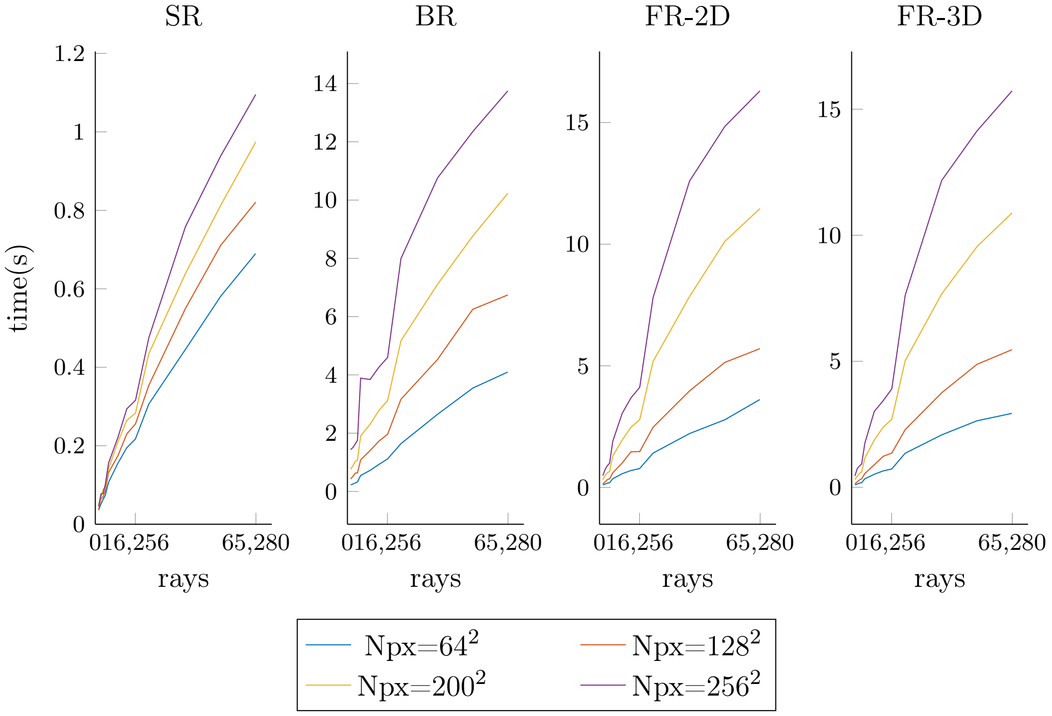


Figure 6.9: Time need to complete the forward modelling (matrix building).

Similarly to the resolution grid, the total number of rays use for the reconstruction also drives the matrix construction time. However, we can see on Figure 6.9 that a sub-linear time complexity is achieved for all non-straight rays. This effect comes in fact in hands with a slightly increased memory usage, as every ray that have the same emitter share some information: for bent rays, it is the fast marching map; for fat rays, the map of distance between this emitter and every point of the aperture.

6.4.3 Use of a region of interest

A more subtile accelerator for the forward modelling, is the usage of a Region of interest, with almost no overhead cost the computation is reduce both for pixels and rays (by ignoring all rays that does not cross this region of interest).

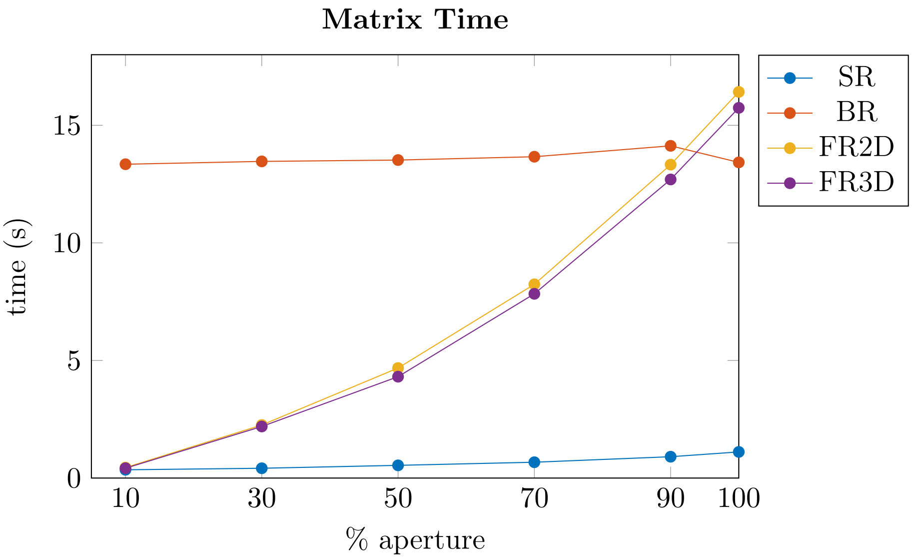


Figure 6.10: Time need to complete the forward modelling (matrix building) with restriction to a ROI, it consists of a disc center in the 2D full aperture: ($256 \times 256 grid$,16 256 rays) with a variable radius

The rays computation react differently to a restriction of the computation to the ROI. On Figure 6.10, the bent rays are almost not affected, by a reduction of the ROI, due to the need of computing the full fast marching map for each ray only after that can the path of a ray be restricted to the part where it crosses the ROI and add into the matrix. This is the same for the straight rays, but the `C` implementation reduces this effect.

On the other hand, the add of a ROI greatly reduces the computation for fat rays. The combined effect on pixels and rays can be seen on the shape of the curve for these methods, which is more quadratic than linear.

6.5 Summary

The USCT transmission reconstruction methods, built on a solid theory shows also some challenges in the numerical applications, the high sparsity of the measurement and their sensibility to various configuration of the device has been studied in 2D, and can be easily extrapolate to 3D, with the matrix being even more sparse. Furthermore, the construction of the matrix is in itself an heavy task, both in time and memory.

Empirically, the time of reconstruction is proportional to $n_{rays} \times n_{pixels}$. The use of Region of interest (ROI) is of great help for reducing the time needed for computation, and it also brings a small improvement on the reconstructed images.

Chapter 7

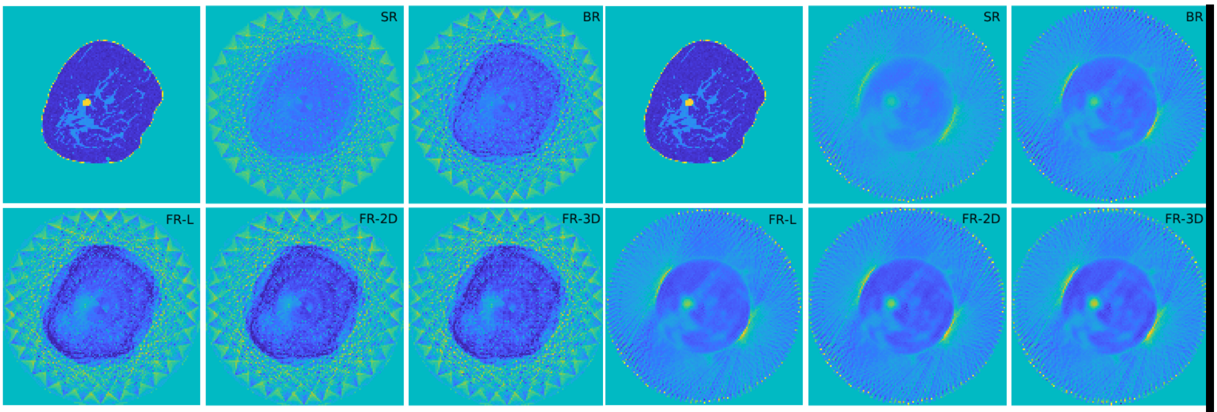
Reconstruction Results

In the last chapter, properties of the forward model matrix has been studied in a quantitavive way. However, As well as the choice of the ray method used, the solver that undergoes the reconstruction is a key feature of the overall process. As a medical imaging devices, the USCT needs to produce also qualitative results and the choice of a solver and the tunig of its parameters are studied in this chapter for this purpose.

7.1 SART and General results

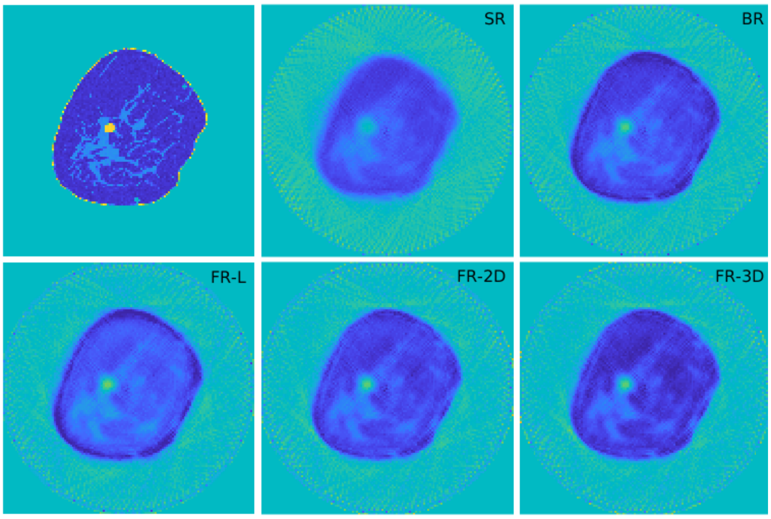
This results in terms of SNR is also visible on the reconstructed images, see Figure 7.1. Reducing the number of transducer increases the presence of geometrical artifacts (Figure 7.1a) and a small opening angle does not allow us to fully reconstruct the breast (Figure 7.1b). Simple straight rays are overcame by bent ray or Fréchet kernels, the linear interpolated fat ray does provide results similar to bent rays. However, their utility can be questioned: to compute them one needs previous iterations (at least one), and increasing the time need to compute them. In comparison, Fréchet kernel give similary, if not better results and does not require previous iterations.

In general, with a big number of transducer and wide opening angle we get far more better results(Figure 7.2a and Figure 7.2), which makes completly sens in term of information processing: the more rays we have, the better the reconstruction, put this also means that the reconstruction will take more time.



(a) 32 transducer and 90° opening angle. (b) 128 transducer and 30° opening angle.

Figure 7.1: SART reconstruction, The use of fat ray reduced the influence of a sparse emitter distribution or small opening angle.



(a) 128 transducer and 90° opening angle.

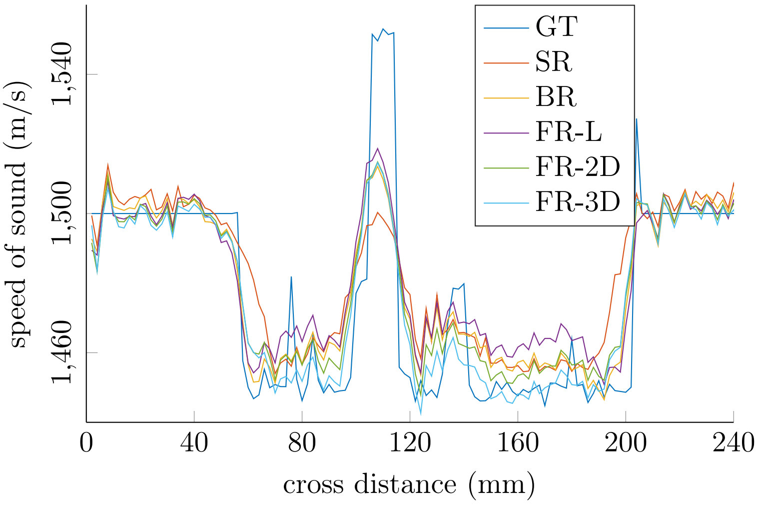


Figure 7.2: cross section of (a) in the middle horizontal line, passing through the tumour.

7.2 TVAL3 Algorithm choice of parameters

The TVAL algorithm (see section 5.5) can be tuned with two regularisation parameter β (regularizing the gradient estimation) and μ (regularizing the reconstructed image).

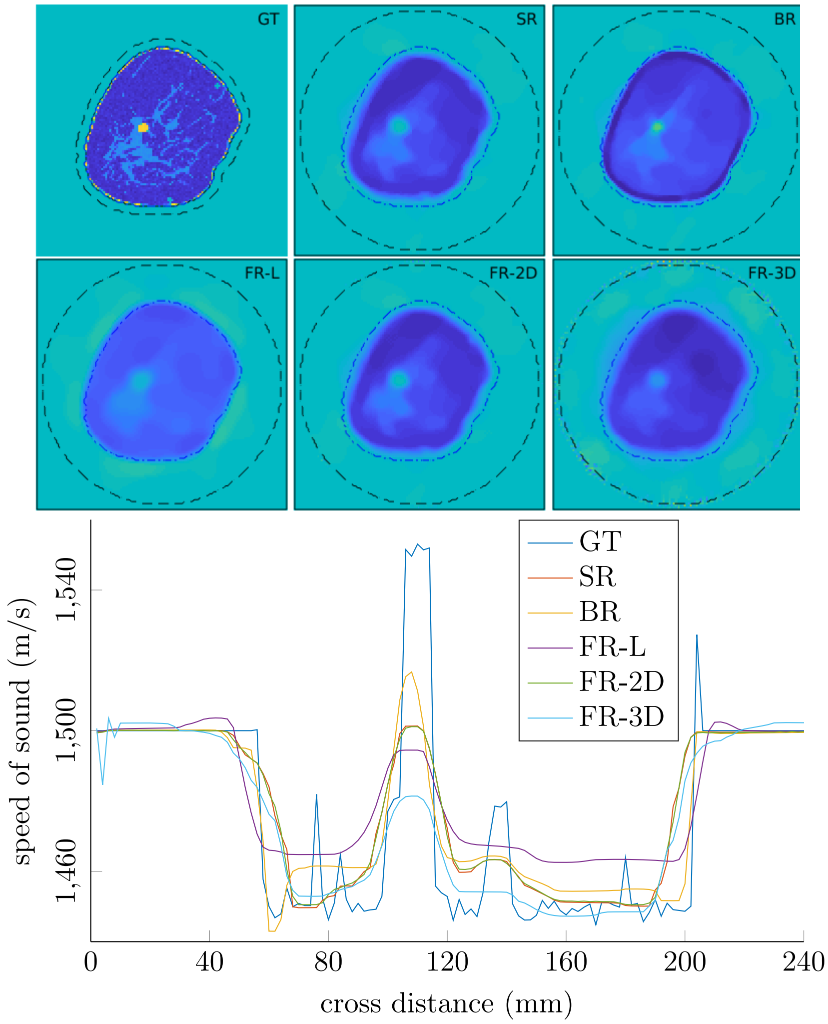
7.2.1 Regularisation vs sparsity

The choice of this parameters can be interpreted as a level of confidence in the provided data. Furthermore we can see on Figure 7.4 and Figure 7.5 that there is no universal optimal value for β and μ,even for the same phantom reconstruction.

In term of SNR the bent rays clearly outcomes the other methods for the reconstruction, in the range of value tested for the regularisation parameters.

However, we can see that with TVAL we get better results in terms of SNR than with SART. The regularisation makes up the difficulties introduce by sparsier data. In practice, the quality of the image seems also better, notice the smoothness of the background.

In comparison to the SART-based reconstruction, we get much smoother results, and the size of the breast is well estimated for bent ray and linear fat rays.



(a) cross section of the middle horizontal line, passing through the tumour.

Figure 7.3: TVAL reconstruction, optimal parameters for bent rays ($\beta = 400, \mu = 10^3$). 128 emitters, full opening angle.

In practice however, the background value does not really matter, and using smaller regularization factor as on Figure 7.6. The peak value representing the tumor is well detected, as well as other stronger variations (at 80 mm and 130 mm for example). also some of the skins(on the border of the breast) tissues starts to appear. This skin tissues are also a good discriminator regarding the resolution size, at 128px (2 mm wide pixels) the skin is only one pixel wide. On strong regularized image, or simple rays methods we cannot detect such tissues.

The linear fat ray show also it limits in this case, based on the bent ray reconstruction, it amplifies the noise, and we get a oscillation of values, with a period of roughtly the width of the Fresnel-Zones use to build these rays. The other version remains stable and provides to the human eye very close results in this case,especially using a 3D based kernel , even if we conducted only 2D simulations.

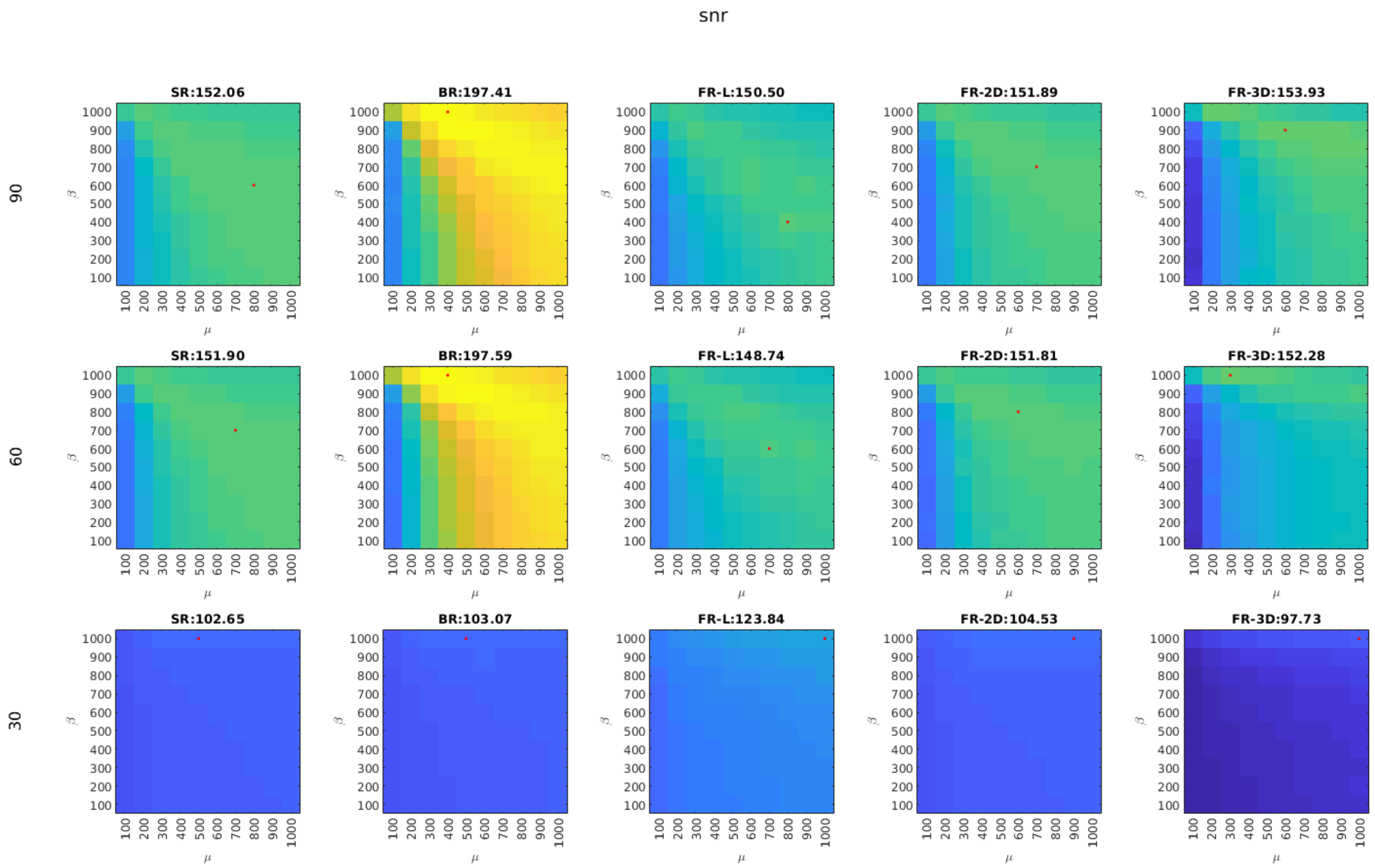


Figure 7.4: SNR for differents regularization parameters, methods and opening angle. The red dot indicates the maximal value given in the title for each type of ray.

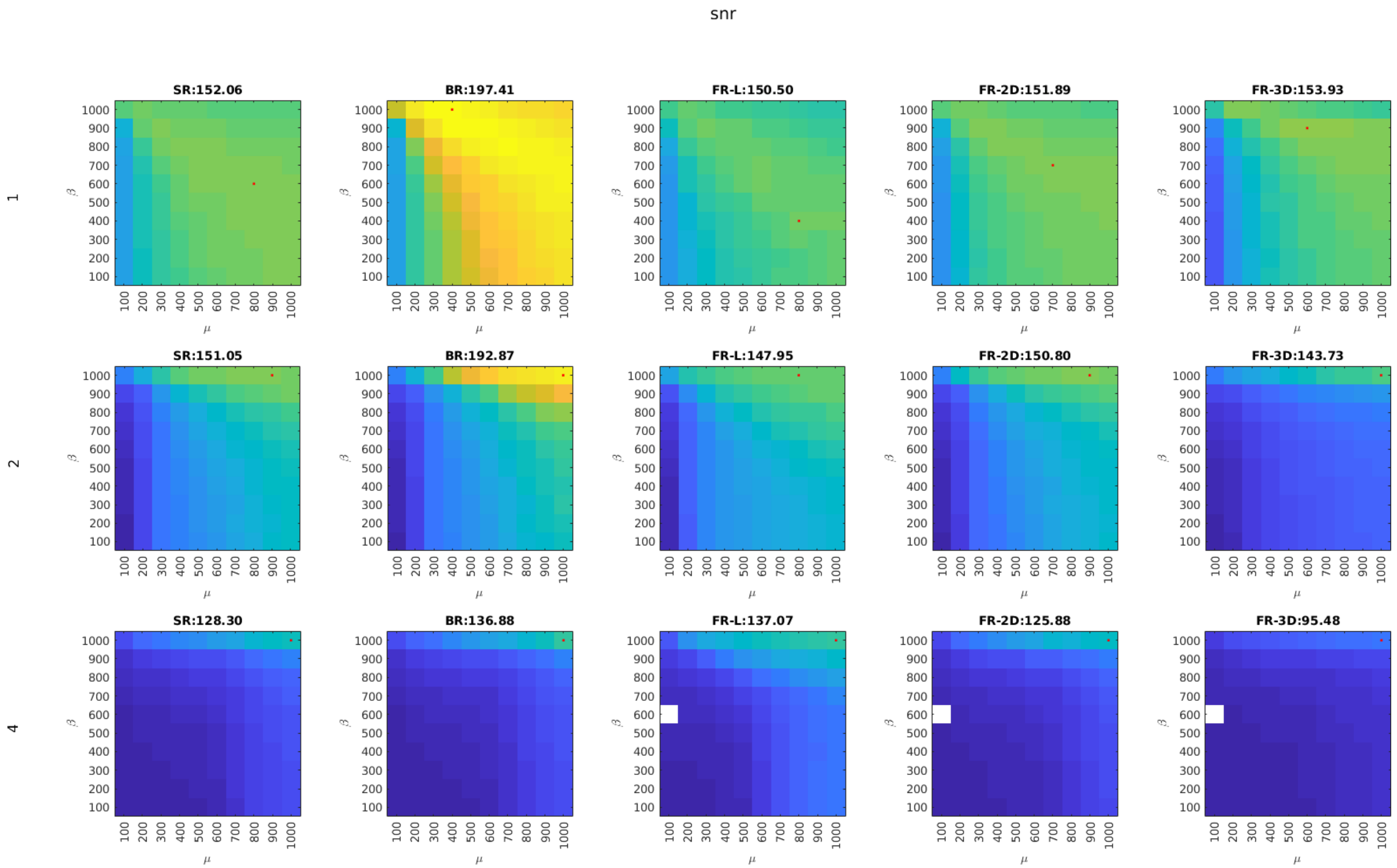
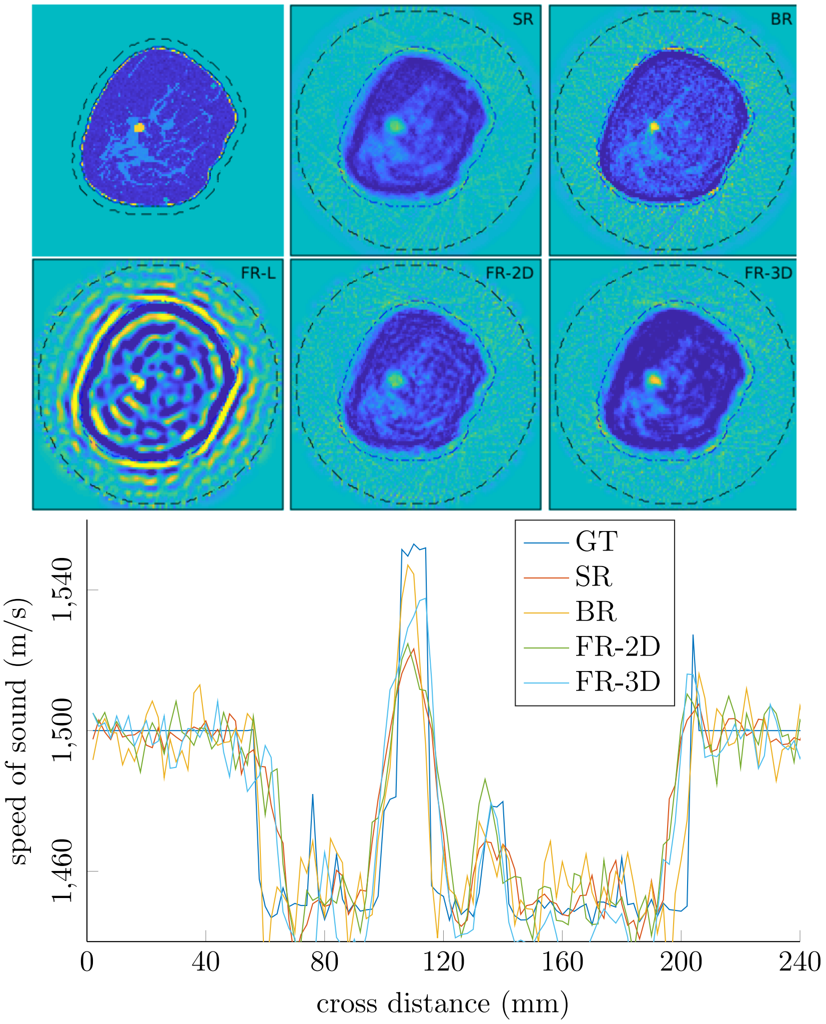


Figure 7.5: SNR for differents regularization parameters, methods and step size. The red dot indicates the maximal value given in the title for each type of ray. Higher steps in emitter size don't converge in every cases.



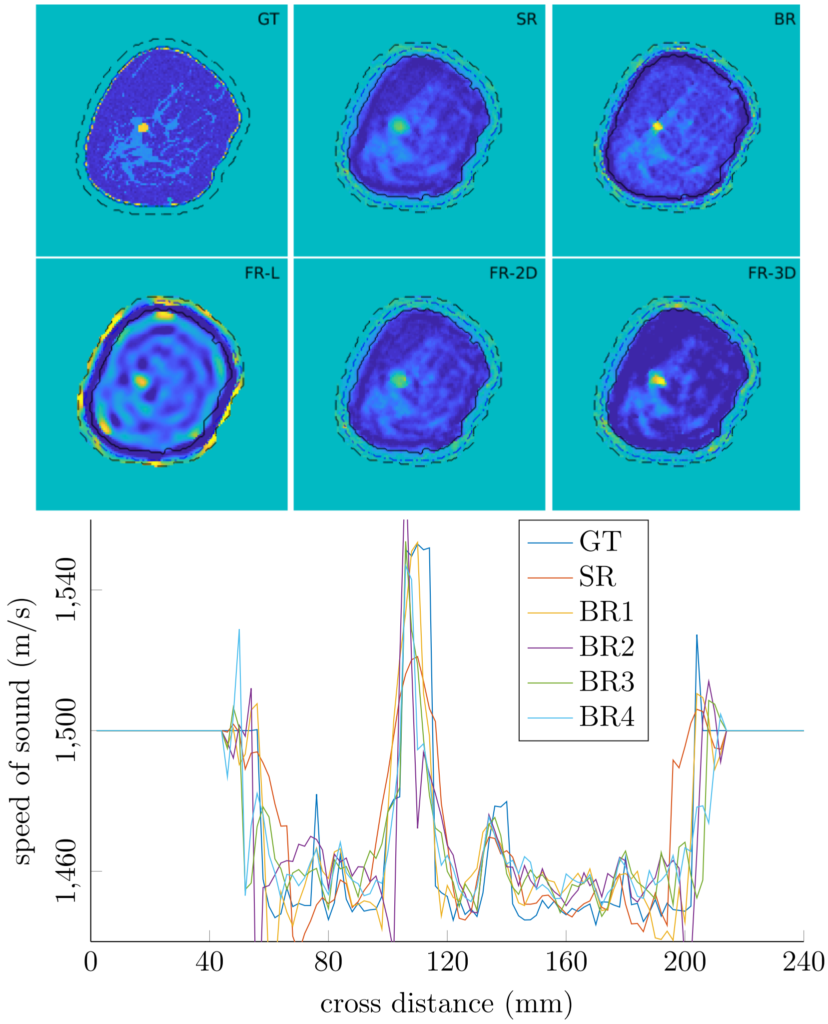
(a) cross section of the middle horizontal line, passing through the tumour. The interpolated fat ray has been ommited for clarity.

Figure 7.6: TVAL reconstruction, weak regularization on gradient ($\beta = 10^{-2}, \mu = 10^{3}$). 128 emitters, full opening angle.

7.3 Use of a region of interest

Considering a reconstruction of the image only inside a region of interest has two main advantage: removing the background noise in water (and thus increase the SNR of the reconstructed image) and accelerates the inversion process.

On Figure 7.7, we can see the importance of a safety margin for the ROI (in dashed black line) if the detection of the estimated breast is done well, the size detected is always smaller than the ground truth phantom (dot-dashed blue line). The use of this dilated ROI also smoothed the edge of the detection, giving a much more regular shape to reconstruct. However, the presented ROI detection has been done in an ideal case, and the presence of artefacts in water (appearing with a small opening angle, or a reduced number of transducer) can yield to false-positive for the ROI detection.



(a) cross section of the middle horizontal line, passing through the tumour.

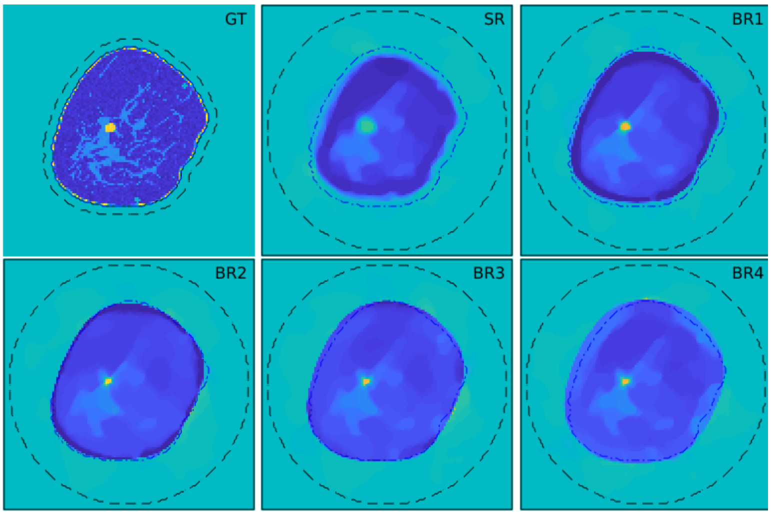
Figure 7.7: TVAL reconstruction, weak regularization on gradient ($\beta = 10^{-2}, \mu = 10^{3}$). 128 emitters, full opening angle. Using a Region of interest.

7.3.1 Iteration of Bent ray

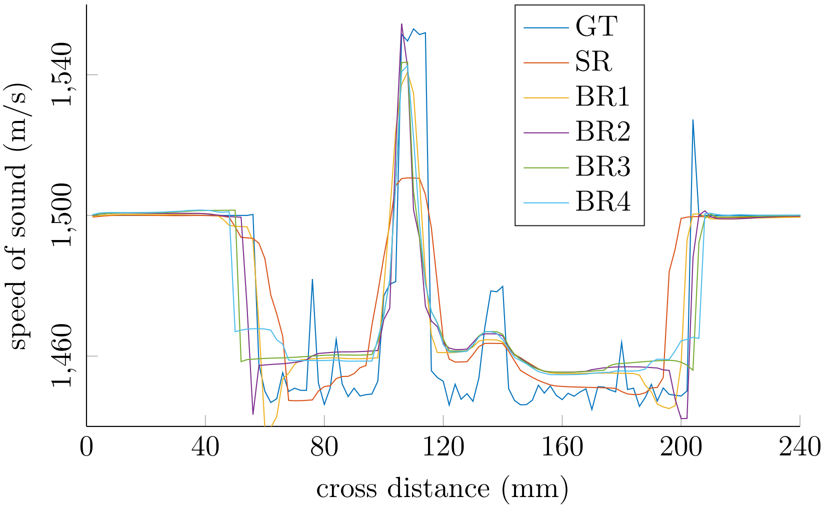
The bent ray method relies on a initial distribution of the speed of sound, from this it will determine ray paths and after inversion this can be repeted as first presented in chapter 3.

This iteration of the bent ray reconstruction is represented on Figure 7.8. This process is however not ideal, and multiplying these iterations does not improve the quality of the image. The main effects are the slowly expansion of the breast size, the reduction of the tumour size estimation, and the apparition of homogeneous area. The latter can be counteracted by reducing the regularization of the solver, but this phenomenon will still appear on the long run.

The expansion of the breast size is also problematic for an automated ROI based approximation. Over the iterations, the ROI will grows along with the breast, increasing the problem size and thus the space for the estimated breast to expand.



(a) Reconstructed images with strong regularization.



(b) Cross section of the middle horizontal line.

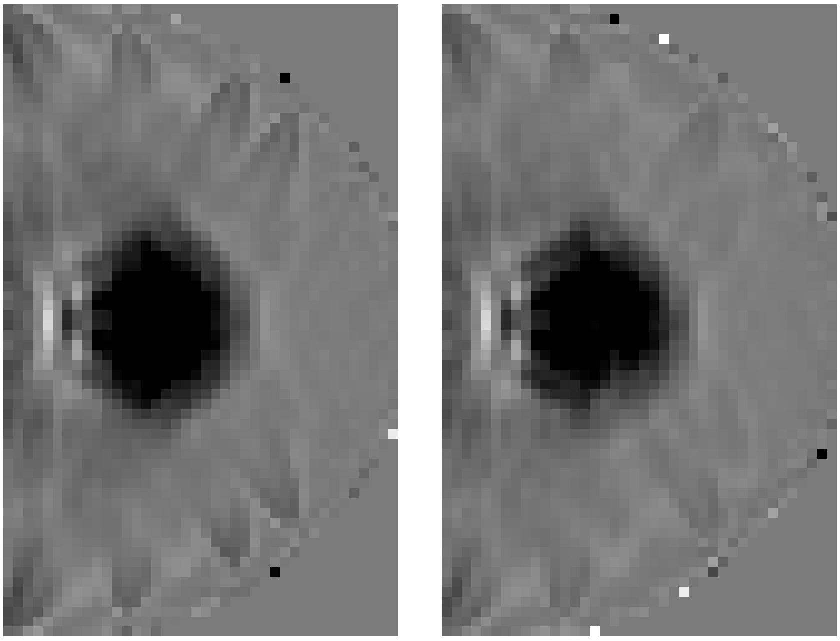
Figure 7.8: Bent rays iterations, The estimated breast size grows, and the tumour size estimation, on the oppisite tends to reduce. TVAL ($\beta = 400, \mu = 10^3$).

7.4 First 3D simulations

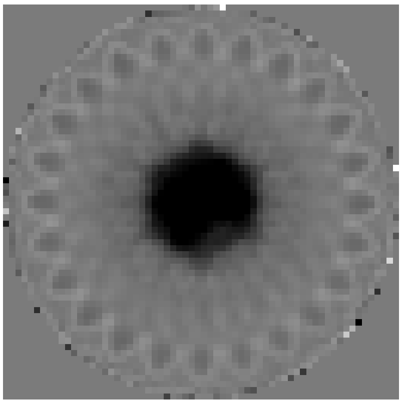
Making a forward simulation of a 3D phantom is not an easy task. Firstly, there is yet no clinical phantom of 3D breast, then simulating a complete USCT acquisition will take several weeks, even on a multi-GPU server. To still test the limits of the reconstruction software however, a very small aperture (having roughly the same amount of voxels as the 2D one) and with fewer emitter (reducing the number of forward simulation waves required) has been used. It consists of a 3D sphere with slow speed of sound, placed in a down-sized aperture of USCT (ten times smaller than the one use in the real USCT-II. On Figure 7.9 a first attempt to reconstruct this phantom has been realized.

To allow reasonable reconstruction time, a ROI is systematically used, thus we only reconstruct 70% of the complete aperture, centered on the phantom.

On Figure 7.9 we can see some geometrical artifact, coming from the fat ray Kernel,



(a) slice x=32 (b) slice y=32



(c) slice z=15

Figure 7.9: 3D phantom (64x64x50) reconstruction. 1 emitter for 4 receiver

it also show how few transducer were used for the simulation. One could hope, that with a higher number of emitter, this result will hold better.

7.5 Time consideration

As well as for the matrix building procedure, the time needed to produced a image is determined by the number of pixels to reconstruct and the number of rays used to do so.

First of, the solver (here SART) is not indifferent to sparsity. Rays methods who produces very sparse matrix (straight rays and bent rays) are ten time faster to reconstruct than the fat rays based matrix, despite only a less than 5% change in sparsity.

As for the matrix reconstruction a sub-linear time complexity can be observed, and may also be the consequences of the vectorization of the computation done by MATLAB.

The TVAL solver, with its rather unpredictible results for a pair of choosen parameter did not undergo the same time profiling, but globally, the reconstruction time is of the same order for the same input problem size. An another drawback of the TVAL solver regarding time optimisation is the impossibility to have a arbitrarily shaped ROI (the implementation of the algorithm requires a rectangular or brick shape).

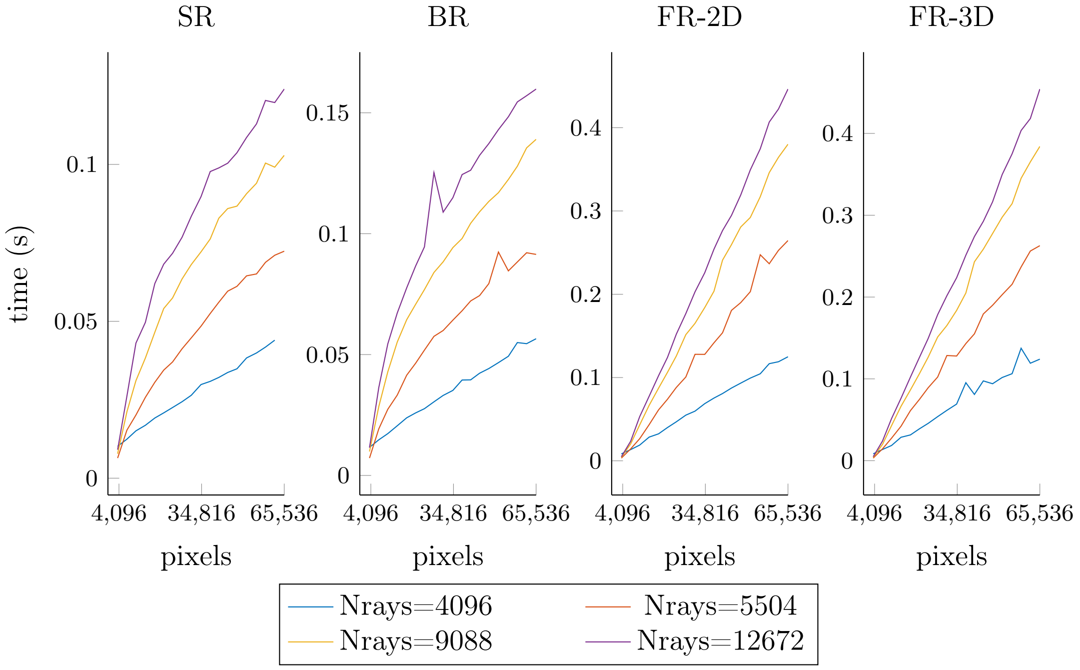


Figure 7.10: Time need to complete the solver reconstruction for different amount of pixels.

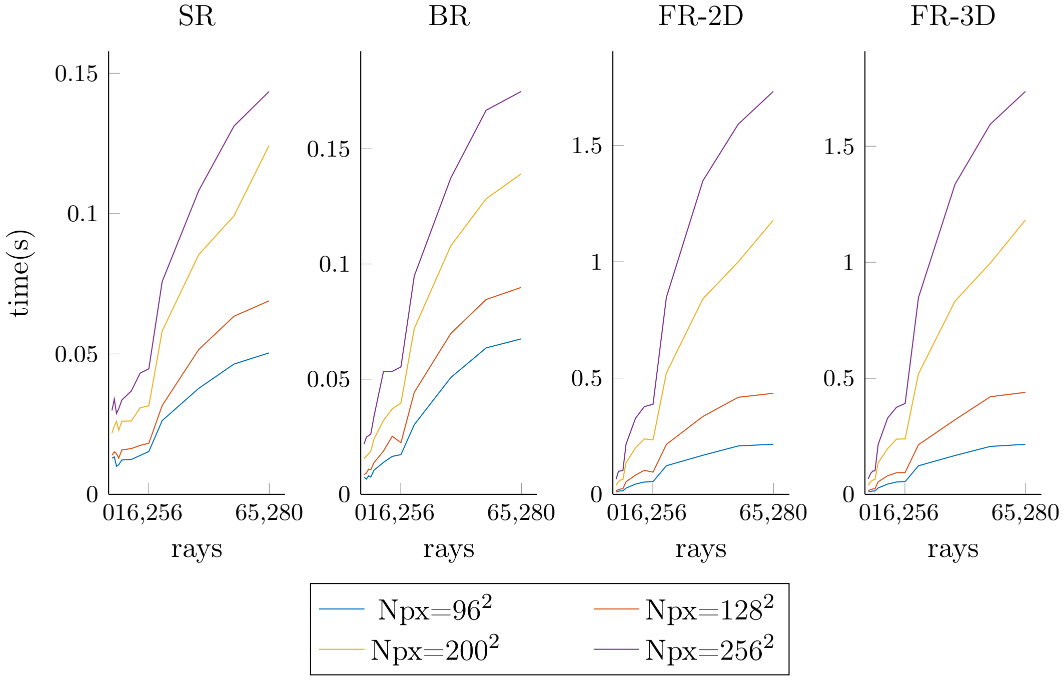


Figure 7.11: Time need to complete the solver reconstruction (matrix building) for different amount of pixels.

7.6 Summary

The data simulated from the `k-Wave` software is used as input for the reconstruction. The two steps of forward modelling (building the matrix) and reconstruction (solving the inverse problem) are not independent.

Providing a large amount of input data, reduce the sparsity of the ray matrix (*i.e.* increase the amount of information collected). It has been shown that Fréchet Kernel based fat ray are a very good solution for it. This effect is also found in the reconstruction step, where a strong regularisation is less needed with fat ray, if not counterproductive.

The SART solver is easy to use but its results can present geometric artifacts, and lack of a wider value range. On the other hand, the TVAL solver can give better results, but no clear rules for the tuning of its parameters has been established.

Finally, using a ROI for limiting the space of reconstruction proves to be working very well in easy cases. Further tuning may be required for noisier configurations.

Chapter 8

Discussion and future work

8.1 Summary

With this year of work in the USCT team at the IPE, further steps have been achieved to improve the transmission tomography reconstruction process. First the development of a more formal theoretical field for the TVAL algorithm and the introduction of the new Fréchet Kernels. Moreover, the introduction of a ROI detection algorithm gives access to faster computation times, which impact is even more significant in 3D reconstruction. The source code of the reconstruction software has also be sanitized, making it closer to the required standard of medical devices.

8.1.1 Ray tomography

The ultrasound tomography is based on the well-studied wave equation and its approximations, yielding to transmission tomography theory. A ray based, linear reconstruction, assuming a infinite frequency emission, is firstly introduced. From this first method, two improvements have been made: the use of a heterogeneous background leads to the concept of straight and then bent ray. By taking into account the influence of the Fresnel-Zone (*i.e.* by assuming a finite frequency emission) I applied the Fréchet Kernel commonly use in seismic to the case of ultrasound.

Both of these methods increase the number and the quality of the information gathered by the acquisition process and gives access to better results, or at least more robust to sparse data.

8.1.2 Solvers

Different solvers have been tested for the constructed inverse problem. The badly conditioned matrix, the sparsity of the problem, and the high dimension of reconstruction, are not easing the reconstruction. The use of a smart solver regularization process partly overcomes these issues.

Among the tested solvers, the total variation based TVAL3 solver gives the best results, and is resilient to an increasing sparsity and a lack of data. The two regularisation parameters allows in theory for a greater tuning of the reconstructed images, but choosing optimal parameters is not an easy task.

On the other and the SART solver, simpler than TVAL, can be easily tuned by various method (*L-curve* search) to find optimal parameter for the reconstruction. At first sight it does not give a wide range of values, and the results seem always blurred, and too regular. This aspect can be overcome by adding a dedicated regularisation step, based on total variation, this extra step will render a sharper image, more piece-wised and reduce the bakcground noise.

Nevertheless, the SART solver by itself is also great tool to easily evaluate how reducing the input data affects the reconstructed image.

8.1.3 Regularization

A Regularization step is essential to produce better results, by adding the *a priori* information that the reconstructed image should be piece-wised like the best regularization procedure seems to be Total Variation, making the assumption that the gradient of the results image should be sparse. The use of a isotropic regularisation provides better results than the anisotropic one in the USCT case.

A seperate regularization step, added after the SART reconstruction provides solid results,

8.1.4 Reconstruction

The aperture volume used in the simulation, – based on the USCT II – is best reconstructed in a 128×128image, making a pixel 2 mm wide. This size gives good detection of details and a fair computational time for the transmission tomography. Higher resolution (and detection of tumours) would be achieved with reflection tomography. Smaller resolutions are possible, but increase the sparsity and halving the resolution makes the problem size 4 times bigger (8 times in 3D) , making it more expensive to compute. Using a ROI detection is clearly a huge improvement, and both reduces the problem size and produces better results by ignoring the background water.

8.2 Discussion

Multiple aspects of the US transmission tomography are prone to discussion, some of them are outside the field of this report like the pre-processing signal-chain (the acquisition of A-Scan, the characteristics of transducer, the time of flight detection algorithm), several ongoing PhD are tackling these topics. However, inside my work various features deserve some enlightenment.

The forward modelling using rays have proven to be a simple yet efficient approach. Straight rays are very simple, and very fast but does not provide a good range of values, compares to ground-truth. A single iteration of bent ray gives promising result, but using multiple iterations of bent rays estimation does not provide a better final image, especially as the tumour size tends to be under evaluated. Fat Ray, in their expression via Fréchet kernel are very interesting. They have the great advantage of hitting more pixels per ray. However, their computation is more complex than the bent-ray, and they remain sensitive to change in the spectrum of the emitted wave, and the pixel size. To unleash their complete potential, they will need a higher resolution, but also more computational power (in term of time and memory).

The actual use of Fréchet kernel is done without any *a priori* on the reconstruction area. Taking into account an initial heterogeneous medium will likely improve the reconstruction results. Using yet banana-shaped kernel, one could take advantage of the bent effect to be more sensible to local variations.

The simulation framework, `k-Waves`, provides a highly idealized models of transducers, and does not yet take into account the specificity of the 3D-USCT transducer. Moreover The lack of a real dataset of breast phantom in 2D, and the yet-to-hard difficulty of produced 3D versions, clearly impede me to deeply analyse the different methods presented.

The forward model matrix is stored as a sparse Matrix in MATLAB, however the lastest version of MATLAB® (R2018 and R2020) does not support single precision sparse matrix, And force us to use almost everywhere double precision array (stored on

64 bits) , whereas single float precision (32 bits) could be precise enough. Achieving such single float implementation (or even with 16bit float, often use in GPUs) will drastically reduce the memory usage, and by having less data to handle, also speed up the overall computation. This will also open the door for more hardware opitmization, such as using GPUs parallelization or implementation on FPGA or ASICs (which as already been done for the acquisition and pre-processing of the A-Scans).

Last, but not least, the region of interest detection seemed to be a direct road to a faster, more accurate reconstruction. Nonetheless, it quickly shows its limits, especially when the first reconstruction does not provide a good estimation of the breast shape. On top of that, every ray method have a tendency to under-estimate the size of the breast.

8.3 Improvement and possible future work

All along this work, multiple research track have been left to explore. Mostly a faster implementation of the reconstruction program (especially the forward modelling)is needed.

The SART algorithm provide smooth and consistent results, yet provides values under evaluated in respect to the ground truth, and presents often geometrical artefacts. The use of Total variation partly overcome this effect, and a extensive tuning of the TV procedure will be required.

Then, the test of attenuation reconstruction. Even if of less interest in the reflection tomography improvement, the attenuation is also a good detector of tumorous cells, and form with the two former, a multi-modal imaging and is thus a great asset for the diagnosis of breast cancer.

The TVAL solver, even if set aside at the end of this work could also deserve some interest, and a automatic search of suitable regularization parameters could be considered, analogous to the L-Curve fitting of the classical Thikonov regularization.

Besides, the next obvious test will be the use of experimental data will also be a new topic of research. A more fine-tuned simulation (taking limitation of the transducer into account) could also be considered.

On a more hypothetical way, the overall process may be enhance by considering as a filter (not particularly linear) and it characterization could be also done by considering its impulse response, which could then be inverted and apply to the reconstruction to come closer to a groundtruth. A machine learning approach to this task could be foresee.

Bibliography

[1] Freddie Bray et al. "Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries". In: *CA: A Cancer Journal for Clinicians* 68.6 (2018), pp. 394–424. ISSN: 1542-4863.

[2] Carol E. DeSantis et al. "Breast Cancer Statistics, 2017, Racial Disparity in Mortality by State". In: *CA: A Cancer Journal for Clinicians* 67.6 (2017), pp. 439–448. ISSN: 1542-4863.

[3] Radhika Sivaramakrishna and Richard Gordon. "Detection of Breast Cancer at a Smaller Size Can Reduce the Likelihood of Metastatic Spread: A Quantitative Analysis". In: *Academic Radiology* 4.1 (Jan. 1, 1997), pp. 8–12. ISSN: 1076-6332.

[4] James F. Greenleaf and Robert C. Bahn. "Clinical Imaging with Transmissive Ultrasonic Computerized Tomography". In: *IEEE Transactions on Biomedical Engineering* BME-28.2 (Feb. 1981), pp. 177–185. ISSN: 0018-9294, 1558-2531.

[5] Hartmut Gemmeke et al. "3D Ultrasound Computer Tomography: Hardware Setup, Reconstruction Methods and First Clinical Results". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 873 (2017), pp. 59–65.

[6] Derouich. *Vergleich Zwischen Chirps Und Binären Codes Für Einen Experimentellen 3D UltraschallComputertomographen.* Forschungszentrum Karlsruhe Institut fur Prozessdatenverarbeitung und Elektronik, 2008.

[7] P. Pellegretti et al. "A Clinical Experience of a Prototype Automated Breast Ultrasound System Combining Transmission and Reflection 3D Imaging". In: *2011 IEEE International Ultrasonics Symposium.* 2011 IEEE International Ultrasonics Symposium. Oct. 2011, pp. 1407–1410.

[8] Neb Duric et al. "Breast Imaging with the SoftVue Imaging System: First Results". In: *Medical Imaging 2013: Ultrasonic Imaging, Tomography, and Therapy.* Medical Imaging 2013: Ultrasonic Imaging, Tomography, and Therapy. Vol. 8675. International Society for Optics and Photonics, Mar. 29, 2013, 86750K.

[9] Radovan Jirik et al. "Sound-Speed Image Reconstruction in Sparse-Aperture 3-D Ultrasound Transmission Tomography". In: *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 59.2 (Feb. 2012), pp. 254–264. ISSN: 0885-3010, 1525-8955.

[10] GF Schwarzenberg, M Zapf, and NV Ruiter. "P3D-5 Aperture Optimization for 3D Ultrasound Computer Tomography". In: *2007 IEEE Ultrasonics Symposium Proceedings.* IEEE, 2007, pp. 1820–1823.

[11] J. Becker et al. *First Evaluation of FPGA Reconfiguration for 3D Ultrasound Computer Tomography.* 2010. URL: `https://publikationen.bibliothek.kit.edu/1000018653` (visited on 10/18/2019).

[12] Marcus Hardt. "Distributed Simulations for 3D Ultrasound Computer Tomography". 2012.

[13] Adrianus T. de Hoop. "Convergence Criterion for the Time-Domain Iterative Born Approximation to Scattering by an Inhomogeneous, Dispersive Object". In: *J. Opt. Soc. Am. A, JOSAA* 8.8 (Aug. 1, 1991), pp. 1256–1260. ISSN: 1520-8532.

[14] Avinash C Kak, Malcolm Slaney, and Ge Wang. "Algebraic Reconstruction Algorithms". In: *Principles of Computerized Tomographic Imaging.* Vol. 29. Wiley Online Library, 2002.

[15] Robin Dapp. "Abbildungsmethoden für die Brust mit einem 3D-Ultraschall-Computertomographen". July 16, 2013.

[16] M.D. Verweij et al. "Simulation of Ultrasound Fields". In: *Comprehensive Biomedical Physics.* Elsevier, 2014, pp. 465–500. ISBN: 978-0-444-53633-4.

[17] Bradley E. Treeby and B. T. Cox. "K-Wave: MATLAB Toolbox for the Simulation and Reconstruction of Photoacoustic Wave Fields". In: *J. Biomed. Opt.* 15.2 (2010), p. 021314. ISSN: 10833668.

[18] Roel Snieder and Malcolm Sambridge. "Ray Perturbation Theory for Traveltimes and Ray Paths in 3-D Heterogeneous Media". In: *Geophys J Int* 109.2 (May 1, 1992), pp. 294–322. ISSN: 0956-540X.

[19] J. E. Bresenham. "Algorithm for Computer Control of a Digital Plotter". In: *IBM Systems Journal* 4.1 (1965), pp. 25–30. ISSN: 0018-8670.

[20] M. Sabry Hassouna and A. A. Farag. "MultiStencils Fast Marching Methods: A Highly Accurate Solution to the Eikonal Equation on Cartesian Domains". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29.9 (Sept. 2007), pp. 1563–1574. ISSN: 1939-3539.

[21] Cuiping Li and Neb Duric. "Resolution Limitation of Travel Time Tomography: Beyond the First Fresnel Zone". In: *Medical Imaging 2013: Ultrasonic Imaging, Tomography, and Therapy.* Medical Imaging 2013: Ultrasonic Imaging, Tomography, and Therapy. Vol. 8675. International Society for Optics and Photonics, Mar. 29, 2013, p. 86751D.

[22] Wojciech Marczak. "Water as a Standard in the Measurements of Speed of Sound in Liquids". In: *The Journal of the Acoustical Society of America* 102.5 (Nov. 1, 1997), pp. 2776–2779. ISSN: 0001-4966.

[23] Jun Zhang and Jinglu Hu. "Image Segmentation Based on 2D Otsu Method with Histogram Analysis". In: *2008 International Conference on Computer Science and Software Engineering.* 2008 International Conference on Computer Science and Software Engineering. Vol. 6. Dec. 2008, pp. 105–108.

[24] Vlastislav Červený and José Eduardo P. Soares. "Fresnel Volume Ray Tracing". In: *GEOPHYSICS* 57.7 (July 1, 1992), pp. 902–915. ISSN: 0016-8033.

[25] Stephan Husen and Edi Kissling. "Local Earthquake Tomography between Rays and Waves: Fat Ray Tomography". In: *Physics of the Earth and Planetary Interiors.* Tomographic Imagining of 3-D Velocity Structure and Accurate Earthquake Location, 123.2 (Apr. 1, 2001), pp. 127–147. ISSN: 0031-9201.

[26] Jesper Spetzler and Roel Snieder. "The Fresnel Volume and Transmitted Waves". In: *GEOPHYSICS* 69.3 (May 2004), pp. 653–663. ISSN: 0016-8033, 1942-2156.

[27] Jeroen Jocker et al. "Validation of First-Order Diffraction Theory for the Traveltimes and Amplitudes of Propagating Waves". In: *GEOPHYSICS* 71.6 (Nov. 1, 2006), T167–T177. ISSN: 0016-8033.

[28] Roel Snieder and Anthony Lomax. "Wavefield Smoothing and the Effect of Rough Velocity Perturbations on Arrival Times and Amplitudes". In: *Geophys J Int* 125.3 (June 1, 1996), pp. 796–812. ISSN: 0956-540X.

[29] Toshiki Watanabe, Toshifumi Matsuoka, and Yuzuru Ashida. “Seismic Traveltime Tomography Using Fresnel Volume Approach”. In: *SEG Technical Program Expanded Abstracts 1999*. 0 vols. SEG Technical Program Expanded Abstracts. Society of Exploration Geophysicists, Jan. 1, 1999, pp. 1402–1405.

[30] Jesper Spetzler and Roel Snieder. “The Effect of Small-Scale Heterogeneity on the Arrival Time of Waves”. In: *Geophysical Journal International* 145.3 (June 2001), pp. 786–796. ISSN: 0956540X, 1365246X.

[31] Yue Tian et al. “Dynamic Ray Tracing and Traveltime Corrections for Global Seismic Tomography”. In: *Journal of Computational Physics* 226.1 (Sept. 10, 2007), pp. 672–687. ISSN: 0021-9991.

[32] Jeroen Tromp, Carl Tape, and Qinya Liu. “Seismic Tomography, Adjoint Methods, Time Reversal and Banana-Doughnut Kernels”. In: *Geophys J Int* 160.1 (Jan. 1, 2005), pp. 195–216. ISSN: 0956-540X.

[33] S Kaczmarz. “Angenäherte Auflösung von Systemen Linearer Gleichungen (English Translation by Jason Stockmann): Bulletin International de l’Académie Polonaise Des Sciences et Des Lettres”. In: (1937).

[34] Richard Gordon, Robert Bender, and Gabor T. Herman. “Algebraic Reconstruction Techniques (ART) for Three-Dimensional Electron Microscopy and X-Ray Photography”. In: *Journal of Theoretical Biology* 29.3 (Dec. 1, 1970), pp. 471–481. ISSN: 0022-5193.

[35] A. H. Andersen and A. C. Kak. “Simultaneous Algebraic Reconstruction Technique (SART): A Superior Implementation of the ART Algorithm”. In: *Ultrasonic Imaging* 6.1 (Jan. 1, 1984), pp. 81–94. ISSN: 0161-7346.

[36] L. Landweber. “An Iteration Formula for Fredholm Integral Equations of the First Kind”. In: *American Journal of Mathematics* 73.3 (1951), pp. 615–624. ISSN: 0002-9327. JSTOR: 2372313.

[37] Yousef Saad and Henk A. van der Vorst. “Iterative Solution of Linear Systems in the 20th Century”. In: *Journal of Computational and Applied Mathematics*. Numerical Analysis 2000. Vol. III: Linear Algebra 123.1 (Nov. 1, 2000), pp. 1–33. ISSN: 0377-0427.

[38] Per Christian Hansen and Maria Saxild-Hansen. “AIR Tools — A MATLAB Package of Algebraic Iterative Reconstruction Methods”. In: *Journal of Computational and Applied Mathematics*. Inverse Problems: Computation and Applications 236.8 (Feb. 1, 2012), pp. 2167–2178. ISSN: 0377-0427.

[39] Jens Gregor and Thomas Benson. “Computational Analysis and Improvement of SIRT”. In: *IEEE Trans Med Imaging* 27.7 (2008), pp. 918–924. ISSN: 1558-254X. pmid: 18599397.

[40] Kunio Tanabe. “Projection Method for Solving a Singular System of Linear Equations and Its Applications”. In: *Numer. Math.* 17.3 (June 1, 1971), pp. 203–214. ISSN: 0945-3245.

[41] R. S. Ramakrishnam et al. “Orthogonalization, Bernstein Polynomials, and Image Restoration”. In: *Appl. Opt., AO* 18.4 (Feb. 15, 1979), pp. 464–468. ISSN: 2155-3165.

[42] Per Christian Hansen and Jakob Sauer Jørgensen. “AIR Tools II: Algebraic Iterative Reconstruction Methods, Improved Implementation”. In: *Numer Algor* 79.1 (Sept. 1, 2018), pp. 107–137. ISSN: 1572-9265.

[43] Pierre Paleo. "Iterative Methods in Regularized Tomographic Reconstruction". PhD thesis. Université Grenoble Alpes, Nov. 13, 2017.

[44] Simon Foucart and Holger Rauhut. *A Mathematical Introduction to Compressive Sensing*. Applied and Numerical Harmonic Analysis. New York, NY: Springer New York, 2013. ISBN: 978-0-8176-4947-0 978-0-8176-4948-7.

[45] Chengbo Li. "An Efficient Algorithm for Total Variation Regularization with Applications to the Single Pixel Camera and Compressive Sensing". 2010.

[46] Matthias Birk et al. "GPU-Based Iterative Transmission Reconstruction in 3D Ultrasound Computer Tomography". In: *Journal of Parallel and Distributed Computing* 74 (Jan. 31, 2014), pp. 1730–1743.