Abstract

The inverse problem in Diffusive Optical Tomography (DOT) is nonlinear and severely ill-posed, therefore, only low resolution reconstructions are feasible when noise is added to the data nowadays. The purpose of this thesis is to improve image reconstruction in DOT of the main optical properties of tissues with some novel mathematical methods. We have used the Adaptive moment (Adam) optimizer, the Nesterov-Adaptive moment (Nadam) optimizer and its improved AmsGrad optimizer for single image reconstructions of the absorption coefficient. In the first instance, we will compare the behavior of three gradient descent-based optimizers on solving DOT inverse problem by running randomized simulation and analyzing the generated data in order to shed light on any significant difference-if existing at all- in performance among these optimizers in our specific context of DOT. The major practical problems when selecting or using an optimization algorithm in a production context for a DOT system is to be confident that the algorithm will have a high convergence rate to the true solution, reasonably fast speed and high quality of the reconstructed image in terms of good localization of the inclusions and good agreement with the true image. In this work we harnessed a carefully designed randomized simulations to tackle the practical problem of choosing the right optimizer with the right parameters in the context of practical DOT applications, and derived statistical results concerning rate of convergence, speed and quality of image reconstruction. The statistical analysis performed on the generated data and the main results for convergence rate, reconstruction speed and quality between three optimization algorithms are presented. Then, we will explore a different way to construct optimizer algorithms for solving the inverse problem of Diffuse Optical Tomography by using diversification of two stochastic gradient-based algorithms, namely NADAM and AMSGrad. We will study the speed of convergence of the proposed new breed of algorithms, also we will discuss the quality of reconstructed images in both cases of free of noise and noisy measurement data. For analysis and exploration of the potential of the proposed algorithm, we use statistical simulations and analysis approach. Whereas most of the approaches for solving the nonlinear problem of DOT make use of the diffusion approximation (DA) to the radiative transfer equation (RTE) to model the propagation of the light in tissue. Therefore, we have solved the inverse problem in DOT by the more accurate continuous wave Diffusion equation in two dimensions.

Keywords : Inverse problem, Image reconstruction, Adaptive moment algorithms, Optimization, Diffuse optical tomography, Radiative transfer equation

Nada Chakhim

Adaptive Moment Based Algorithms to Solve the Inverse Problem in Diffuse Optical Tomography

Année Universitaire : 2021 – 2022

جامعة الحسن الأول
UNIVERSITÉ HASSAN 1er

Université Hassan 1er
Centre d'Études Doctorales en Sciences et Techniques & Sciences Médicales

كلية العلوم والتقنيات سطات
FACULTÉ DES SCIENCES ET TECHNIQUES SETTAT

Faculté des Sciences et Techniques Settat

THÈSE DE DOCTORAT

Pour l'obtention de grade de Docteur en Sciences et Techniques

Formation Doctorale : Mathématiques, Informatique et Applications

Spécialité : Mathématiques Appliquées

Sous le thème :

Adaptive Moment Based Algorithms to Solve the Inverse Problem in Diffuse Optical Tomography

Présentée par :

Nada Chakhim

Soutenue le : *19/07/2022* à 10h à la Faculté des Sciences et Techniques de Settat devant le jury composé de :

Pr. Mohamed BAHAJ	**PES**	**Faculté des Sciences et Techniques, Settat**	Président
Pr. Abdellah OULD KHAL	**PES**	**Faculté des Sciences Rabat**	Rapporteur
Pr. Ilias CHERTI	**PES**	**Faculté des Sciences et Techniques, Settat**	Rapporteur
Pr. Rachid El JID	**PH**	**Faculté des Sciences et Techniques, Settat**	Rapporteur
Pr. Mohamed HANINI	**PH**	**Faculté des Sciences et Techniques, Settat**	Examinateur
Pr. Mohammed ALAOUI	**PES**	**Faculté des Sciences et Techniques, Settat**	Co-Directeur de thèse
Pr. Mohamed LOUZAR	**PES**	**Faculté des Sciences et Techniques, Settat**	Directeur de thèse

Année Universitaire : 2021/2022

Acknowledgment

This thesis would not have been possible without the guidance and the help of several individuals who in one way or another contributed and extended their valuable assistance in the preparation and completion of this study.

First and foremost, my utmost gratitude to the omnipresent God, for answering my prayers for giving me the strength to plod on.

I am extremely grateful to Pr. LOUZAR MOHAMED, he has been my inspiration as I hurdle all the obstacles in the completion this research work.

Also I would like to thank , Pr. ALAOUI MOHAMMED, for his unfailing support as my thesis co-adviser.

My thanks also goes to Pr.Lamnii abdellah for his valuable advises and availability during this adventure.

I would like to extend my sincere thanks to the members of jury for their availability.

Many thanks goes to the Administrators of the Faculty of science and technology of Settat, for their untiring effort in encouraging the teaching staff to pursue professional growth.

Last but not the least, I would like to thank my family for their encouragement during this journey.

Abstract

The purpose of this thesis is to improve image reconstruction in DOT of the main optical properties of tissues with some novel mathematical methods. We have used the Adaptive moment (Adam) optimizer, the Nesterov-Adaptive moment (Nadam) optimizer and its improved AmsGrad optimizer for single image reconstructions of the absorption coefficient. In the first instance , we will compare the behavior of three gradient descent-based optimizers on solving DOT inverse problem by running randomized simulation and analyzing the generated data in order to shed light on any significant difference in performance among these optimizers in our specific context of DOT. The major practical problems when selecting or using an optimization algorithm in a production context for a DOT system is to be confident that the algorithm will have a high convergence rate to the true solution, reasonably fast speed and high quality of the reconstructed image in terms of good localization of the inclusions and good agreement with the true image. In this work we harnessed a carefully designed randomized simulations to tackle the practical problem of choosing the right optimizer with the right parameters in the context of practical DOT applications, and derived statistical results concerning rate of convergence, speed and quality of image reconstruction. The statistical analysis performed on the generated data and the main results for convergence rate, reconstruction speed and quality between three optimization algorithms are presented. Then,we will explore a different way to construct optimizer algorithms for solving the inverse problem of Diffuse Optical Tomography by using diversification of two stochastic gradient-based algorithms, namely NADAM and AMSGrad. We will study the speed of convergence of the proposed new breed of algorithms, also we will discuss the quality of reconstructed images in both cases of free of noise and noisy measurement data. For analysis and exploration of the potential of the proposed algorithm, we use statistical simulations and analysis approach. Whereas most of the approaches for solving the nonlinear problem of DOT make use of the diffusion approximation (DA) to the radiative transfer equation (RTE) to model the propagation of the light in tissue. Therefore, we have solved the inverse problem in DOT by the more accurate continuous wave Diffusion equation.

Résumé

Le but de cette thèse est d'améliorer la reconstruction d'image en TOD des propriétés optiques du tissu avec de nouvelles méthodes mathématiques. on a utilisé l'algorithme de moment adaptatif (Adam), l'algorithme de moment adaptatif Nesterov (Nadam) et son variant algorithme AmsGrad amélioré pour reconstruire le coefficient d'absorption. Dans un premier temps, on a compareré le comportement de trois algorithmes basés sur la méthode de gradient lors de la résolution du problème inverse TOD en exécutant une simulation aléatoire et en analysant les données générées afin de mettre en lumière toute différence significative de performance entre ces algorithmes dans le cas du TOD. Les principaux problèmes lors de la sélection ou de l'utilisation d'un algorithme d'optimisation pour un système TOD est d'être certain que l'algorithme aura un taux de convergence élevé vers la solution, une vitesse rapide et une haute qualité de l'image reconstruite en termes de bonne localisation des inclusions et bon accord avec l'image réelle. On a exploité des simulations aléatoires soigneusement conçues pour résoudre le problème du TOD en choisissant l'algorithme adéquat avec les bons paramètres, et on a obtenu des résultats statistiques concernant le taux de convergence, la vitesse et la qualité de la reconstruction d'image. L'analyse statistique effectuée sur les données générées et les principaux résultats pour le taux de convergence, la vitesse de reconstruction et la qualité entre les trois algorithmes d'optimisation sont présentés. Ensuite, on a exploré une manière différente de construire des algorithmes d'optimisation pour résoudre le problème inverse de la tomographie optique diffuse en utilisant la diversification de deux algorithmes basés sur le gradient stochastique, à savoir NADAM et AMSGrad. On a étudié également la vitesse de convergence de la nouvelle génération d'algorithmes proposée, et on a discuté la qualité des images reconstruites dans les deux cas de données de mesure sans bruit et bruitées. Pour l'analyse et l'exploration du potentiel de l'algorithme proposé, on a utilisé des simulations statistiques et une approche d'analyse. La plupart des approches pour résoudre le problème non linéaire de TOD utilisent l'approximation de diffusion (AD) à l'équation de transfert radiatif (ETR) pour modéliser la propagation de la lumière dans les tissus. Par conséquent, on a résolu

le problème inverse en TOD en utilisant l'équation de diffusion dans le cas du régime stationnaire.

List of Tables

List of Figures

Contents

Chapter 1

Introduction

1.1 Motivation

One of the earliest medical imaging techniques to find broad popularity was the use of X-rays. Long before the invention of the digital computer, imaging modalities including X-ray Computed Tomography (X-ray CT), Magnetic Resonance Imaging (MRI), and Ultrasound were developed thanks to the widespread use of methods like projection radiography and fluoroscopy. Based on the various susceptibility characteristics of tissue, radio-frequency waves utilized in magnetic resonance imaging (MRI) are used to generate anatomical images. MRI uses non-ionizing radiation instead of the dangerous ionizing radiation used in X-ray imaging, which has the downside of eliminating patients with pacemakers or metal items.

The main conventional imaging methods still in use today are now briefly introduced.

X-rays were first employed in biological imaging in the early 1970s after being discovered by Roentgen in 1895 [1]. In X-ray mammography, high intensity X-rays are utilized to probe biological mediums. When X-rays are focused on biological tissue, they either absorb or scatter the tissue [2, 3, 4]. The performance of biomedical imaging systems is measured by its sensitivity. In biomedical imaging, sensitivity is calculated by dividing the number of positive breast cancer cases detected by the total sum of those that were detected plus those that were missed and then multiplied by 100 [5]. Another important measure in biomedical imaging techniques is specificity. The specificity is given by dividing the number of negative breast cancer cases detected by the total sum of those that were truly negative and were detected as negative breast cancer cases plus those that were incorrectly detected to be positive and then multiplied by 100 [5]. Although X-ray mammography is the current biomedical imaging method of choice for finding breast cancer tumors, its use has been questioned because the specificity of mammography varies

from 90 to 98 percent [2, 3, 4] and because a high percentage of expensive biopsies based on suspicious mammograms result in negative results [6, 7, 8]. In addition, the sensitivity of mammography has been reported to be between 80 and 95 percent [2, 4], whereas the X-ray scan may not be able to identify 15 to 30 percent of tumors. Therefore, its usage leads to false-positive and false-negative diagnoses. The relative lack of contrast between sick and healthy tissues in mammograms is the cause of these inaccurate diagnosis. Ionizing radiation, which is unpleasant and distressing for patients, is another downside of mammography. Another imaging method for tumor diagnosis is X-ray computed tomography (CT). Although X-ray CT has a higher accuracy than mammography, it misdiagnoses early-stage tumors and necessitates large radiation doses [8, 10]. Additionally, mammography is ineffective for finding male breast cancer [11].

Permanent magnets with strengths between 0.25 and 3.9 Tesla are used in MRI scanning. In a 1.5 Tesla magnetic field, hydrogen atoms, which make up the bulk of the atoms in biological tissue, can absorb wave energy at a resonance frequency of 60 MHz [12]. Depending on the magnetic characteristics of the tissue, these hydrogen atoms eventually release the energy they have absorbed. This process is slow since the signal is recorded and transformed one point at a time in a plane's k-space data. Then, the k-space data is subjected to a Fourier transform in order to recreate the magnetic characteristics of tissue [12]. The slowness, high cost, and importability of MRI are all disadvantages [13]. Furthermore, patients who have had metallic components implanted after previous surgical operations are unable to be imaged with MRI. The low specificity of MRI is another problem. The fact that MRI is non-ionizing and has a high sensitivity are two of its advantages [14]. Another advantage is that it has a good prediction of residual disease [15].

In ultrasonic modality imaging methods, sound waves with a frequency of 1 to 10 MHz are employed. Using ultrasonic transducers, sound waves are sent through the tissue, and the reflected sound waves, also known as echoes, are gathered. Resolution and imaging depth can be traded off, and this trade-off is dependent on the frequency of the sound waves being utilized [12]. Non-ionizing US is used to scan pregnant mothers. According to several research, US gives only fair information for classifying particular solid breast masses; its sensitivity and specificity values are 97% and 85%, respectively [15, 5]. Mammography and US are comparatively ineffective at detecting the presence or severity of residual illness [17]. Deep tissue imaging with US has a very low resolution, which is one of its key limitations. Its operator dependence is another. Additionally, due to the closeness of the lungs and ribs, US is unsuitable for the diagnosis of male breast cancer [11]. Over the past few decades, hybrid imaging modalities including photoacoustic tomography and ultrasound modulated optical tomography have become more prevalent.

In many imaging scenarios, it's essential to noninvasively gather some perti-

nent data about an object's interior without cutting or taking sections of an object or organism, which could be dangerous or be more expensive. Instead, we want to gather crucial data on interior anatomical or physiological values. Image reconstruction has a wide range of uses, including the detection and reconstruction of images of objects in luggage at airports, the localization and characterization of oil reservoirs, the satellite recognition of objects in space, the monitoring of industrial processes, and non-invasive biomedical imaging techniques [16]. Breast cancer and brain imaging research in particular have grown in popularity over the past several decades in the field of biomedical imaging. The most common malignancy in women is breast cancer and has one of the worst mortality rates in the world today. In order to reduce unnecessary biopsies and mortality rates, it is crucial to develop better methods for early detection in order to prevent tumors and to obtain an accurate diagnosis of cancer treatment, visualizing small tumors, classifying metabolism illnesses, and differentiating between benign and malignant lesions at a low cost [16, 17]. Medical institutions all across the world are very interested in brain diseases. Anxiety disorders, panic disorders, phobias, addictions, post-traumatic stress disorder, schizophrenia, mood disorders , attention deficit/hyperactivity disorder, insomnia, and novel disorders are examples of mental illnesses. Multiple sclerosis and epilepsy are neurological conditions. Health research has to be done on various levels in order to help with this dilemma. One of these levels relates to getting high-contrast images of the inside of the body, and the technology should be portable, reasonably affordable, and secure.

The models used to describe light propagation and the methods used to reconstruct it are dependent on the application; for example, the Diffusion Approximation is not suggested for small samples because it cannot forecast light propagation on those mediums. Despite the fact that the answer is likely to be difficult and time-consuming, there are some situations where Monte-Carlo models and the Radiative Transfer Equation are advocated. To get beyond the restrictions imposed by each DOI reconstruction technique, hybrid techniques have been developed, such as the radiative-diffusion models shown in [18].

An overview of the mathematical models employed by the software program NIRFAST, which simulates light propagation and provide image reconstructions, is given in [19]. TOAST++, is an alternative program of NIRFAST, which is presented in [20]. Both packages provide with a selection of libraries for Finite Element computation as well as several inverse solvers for image reconstruction. The visualisation of light tissue interactions through scattering, absorption, and fluorescence from endogenous and foreign tissue elements is possible thanks to the great number of current Diffuse Optical Imaging technologies and reconstruction methodologies. The structure, molecular activity, physiology, and biochemistry of tissue are revealed by these optical qualities [21]. It's worth noting that near-

infrared light is greatly dispersed in biological tissue, therefore light does not propagate in a straight path from sources to detectors. As a result, typical computed tomography back-projection algorithms have limited utility. Furthermore, image reconstruction is known to be non-linear and ill-posed, necessitating the use of complex reconstruction methods [22].

1.2 Aims of the thesis

The reconstruction of optical characteristics of biological tissues using information from the exterior boundary was initially stated as an inverse problem in the 1990s [1]. Although the theory and application of regularization methods have advanced quickly since then [23], including the development of hybrid imaging methods in DOT [16, 24], significant advancements must still be made to the current methods as well as the creation of new ones in order to address the significant practical challenges that this modality faces. This is due to a number of factors, including the inverse problem's nonlinearity in DOT, its poor poseability, and its high degree of unpredictability. In order to confirm the inherent potential of DOT as a future imaging modality in several of the above-mentioned applications, this thesis aims to combine some contemporary gradient-based techniques (such as adaptive moment type optimizers) with advanced and precise forward modeling techniques (here, the Diffusion approximation). We are very confident that all of our techniques and numerical simulations generalize with the appropriate modifications to real-world 3D situations because all of the algorithms and ideas provided in these techniques are kept at a very general level using formulations involving PDEs rather than specific discretisations [23, 25]. Such broad generalizations within this thesis.

Adaptive moment type optimizers, so far, are the methods of choice for performing the image reconstruction in DOT. Recent work has shown that applying these new type of optimizers in DOT based on the diffusion equation shows a promising improvement on denoising and helps to address the lack of sharp features. Therefore, we have implemented Adam-type methods with different cases to deal with this problem. In contrast to most approaches taken so far, our work is based on the Diffusion equation, which is the most basic model for the propagation of light in tissue. Notice that the diffusion equation (DE) provides a reasonable model in medical imaging not only for Diffuse optical tomography (DOT), but also for fluorescence diffuse tomography (FDOT). Therefore, the results of this thesis should also be of interest to those modalities and applications. For more information about FDOT we refer to [1, 26], bioluminiscence tomography see [27,28]. Other applications of inverse transport theory outside of medical imaging include geophysics and atmospheric physics.

Most techniques for solving the nonlinear problem of DOT in biomedical imag-

ing, as well as many of the above-mentioned related techniques, apply the diffusion approximation (DA) to the RTE to represent light propagation in tissue [20]. As a result, in the continuous wave domain, we have solved the inverse problem in DOT using the more accurate (DA), in 2D.

In this thesis we have implemented three types of adaptive moment optimizers to solve the inverse problem of DOT based on the diffusion approximation of the RTE. As already mentioned above, we want to emphasis once again that all techniques proposed and tested in this thesis can be applied with only few changes to other imaging systems.

1.3 Structure of thesis

The thesis contains seven chapters whose content is outlined in the following.

Chapter 2 describes broad methods to medical imaging and provides an overview of the state-of-the-art of conventional biomedical imaging techniques, highlighting some of their advantages and disadvantages. Then, a more detailed introduction of DOT is provided, along with its benefits and drawbacks.

Chapter 3 presents the formulation of the forward problem and outlines an overview of the analytical methods used to model the light propagation in tissue.

Chapter 4 presents and summarizes some mathematical background which is needed in later chapters of this thesis. The fundamentals of the Radiative Transfer Equation and its related diffusion equation to RTE are covered, and then a brief proof of the well possedness of the forward problem in DOT will be presented.

In chapter 5, the inverse and shape reconstruction problems in DOT are formulated. The convergence behavior of Adam, Nadam and AmsGrad optimizers, will be examined, when applied to the problem of DOT. A comparison between these optimizers will be investigated and discussed. We will characterize the performance of these algorithms with respect to the choice of some hyper parameters and the initial guess error. Finally, numerical results are presented using these methods.

In chapter 6, we present a new approach to improve optimizers performance by using a Diversified Algorithm to boost the performance and the accuracy of adaptive moment optimizers in the case of image reconstruction in diffuse optical tomography.

Finally, in chapter 7, some conclusion and directions for future work are given.

Chapter 2

Diffuse Optical Imaging equipments

In this chapter we outline general approaches to the medical imaging and overview of the state-of-the-art of conventional biomedical imaging techniques, highlighting some of their advantages and disadvantages. Then, a more detailed introduction of DOT is provided, along with its benefits and drawbacks.

2.1 Introduction

Richard Bright presented the first clinical investigations utilizing light transillumination in 1831 for hydrocephalus and intraventricular haemorrhage studies. Cutler also conducted transillumination tests in 1929, where he evaluated breast lesions by projecting light from an electric lamp onto the breast [29]. Despite the fact that light imaging has been demonstrated to be an effective tool for clinical diagnosis, this method was eventually abandoned because it was challenging to interpret and overheated the patient's skin [25]. In the 1970s, technical breakthroughs and the use of near-infrared illumination allowed for the creation of light scanning for optical mammography research [30]. Jobsis published a paper in 1977 that described the use of near infrared spectroscopy to examine the hemodynamics and oxygenation of the brain [31], which paved the way for the creation of a number of diffuse optical imaging techniques that covered everything from analysis to whole tissue examinations. During the 1990s, a number of optical devices began to be employed for research and diagnostic purposes [32]. These provided non-invasive, simple inspection, and cost-effective hardware, but they were unable to be used in clinical research because to technological issues, insufficient resolution, and the absence of physical models to define and quantify light propagation. The major areas of interest for the developing research community were mammography

[33] and brain imaging [34], but those restrictions did not deter them from trying to apply optical imaging technology to therapeutic applications. Although clinical applications were severely constrained, advances in both computational and near-infrared (NIR) technology were developed throughout this decade [35]. Improved light sources and detectors enhanced the process of light generation and measurement. Researchers were also able to construct more difficult algorithms because to better computers [36, 37]. Additionally, by understanding the physics governing the path taken by photons, researchers proposed alternate solutions to simplify light propagation models. These researchers proposed a set of coupled diffusion equations to model fluorescence-tissue interaction, which neglected the fluorescence that a highly scattered medium presents [38], thereby reducing the complexity of the light propagation model. The aforementioned advancements in data collection, coupled with computers equipped with cutting-edge mathematics and algorithms, have led to significant advancements in diffuse optical imaging (DOI) techniques like functional NIR spectroscopy (fNIRS), which improves tissue characterization by enhancing the analysis wavelength, diffuse optical tomography (DOT), which enables the study of more complex mediums, and fluorescence diffuse optical tomography (FDOT), which creates new fluorophores.

An illumination source, a detecting unit, an imaging platform, a hardware controller, and a signal processor system are all components of DOI instrumentation. The light source, image platform, and detecting unit are all considered to be separate instruments [39,40,41,42]. The signal processor is a computer-based system that includes software for measurement filtering, calibration, light propagation modeling, and image reconstruction techniques. The hardware controller is either a component of the computer or a stand-alone device [43]. If an external hardware controller or a signal processor is located outside of the main computer, they will still require signals from the computer to perform the required operation. Diffuse Optical Imaging instrumentation can be classified in two groups. The first type, referred as imaging domains, focuses on the type of illumination source and the corresponding detection method. Geometry measurements is the second class. In terms of lighting technology, it considers the arrays of sources and detectors arranged on a target medium [43].

2.2 Types of Diffuse Optical Imaging Systems

Diffuse Optical Imaging instrumentation is grouped into three domains based on the kind of sources and detectors used to make intensity measurements, as well as the type of processing done to the data obtained: Continuous Wave (CW), Time Domain (TD), and Frequency Domain (FD). The obtained instrumentation is known as a hybrid optical system if it is coupled with another medical imag-

ing modality or is the result of the combination of two optical systems. This instrumentation may offer several benefits, but it is also expected to increase the complexity and costs of such implementation.

2.2.1 Continuous Wave systems

Continuous wave systems, also known as steady-state domain (SSD) systems, illuminate a target material with a continuous light source and then measure the transmitted or reflected light intensity. Absorption and scattering processes within the medium reduce the measured light, when compared to the other two modalities.

Among DOI systems, continuous wave systems equipment offers the greatest data recording rate, enabling high temporal resolution real-time investigations of haemodynamic changes in tissue [44]. Due to its simpler electronics, easier construction, and reduced related cost, it is frequently employed in clinical applications [45]. The following are some CW intensity measuring's drawbacks [42]:

- Reconstructing deep tissue is challenging since measurements are more sensitive for tissue close to the surface. For instance, brain image reconstructions are significantly impacted by skin systemic reactions and also affected by the thickness of the skull and cerebrospinal fluid (CSF) [46].

- The presence of hair, variations in skin color, changes in position, and pressure from the optode against the target medium all have an impact on intensity measurements [40]. These two statements are also true for the frequency domain (FD) and time domain (TD), but with more information, it is possible to lessen these effects [47,48].

- Although Pei et al. [49] conducted experiments and discovered that using a normalized-constraint algorithm and a priori knowledge is possible to separate both coefficients using CW measurements, theoretical studies showed that absorption and scattering parameters cannot be decoupled due to cross-talk effects [48].

- The difficulty of doing fluorescence lifetime reconstruction with CW apparatus is a severe constraint [50], however numerous options have been presented, such as normalizing the fluorescence with excitation measurements [51]. Theoretical demonstrations of the cross-talk effect revealed that reconstructions of absorption and scattering characteristics using only intensity data were not possible [48].

Nevertheless, it was illustrated in [49] that a normalized constraint may be used to get around this restriction. Difference method based on the usage of the

normalized relative measures in [52]. Since the suggested technique assumes that the anomalies are small and close to background values, it can only be used for bigger anomalies. However, it had several benefits when a single parameter was taken into account. For instance, the reconstruction is less susceptible to the effects of tissue probes, reduce the effect of erroneous measurements and quality disparities in measurements model-related issues.

Despite the above-mentioned drawbacks, CW systems have developed into a potent and popular imaging technique, thanks to developments in computation and reconstruction algorithms like the use of constrained solutions and the incorporation of a priori data through regularization and normalization techniques [49]. Additionally, systems with several sources and detectors may be designed due to their lower cost and simplicity [53].

2.2.2 Time-Domain systems

The distribution of photons from the temporal point spread function (TPSF) a light pulse with a brief duration delivered through a scattering material [34]. The time-domain (TD) method employs pulsed light given to the medium in the order of milliseconds. picoseconds. The TPSF is expanded to a few nanoseconds after traveling further inside the tissue before being recorded by the detectors [54]. TD systems have a higher spatial resolution and improved depth sensitivity. Compared to CW systems, TD instrumentation has the ability to inhibit early coming photons to allow late arriving photons to be exploited unlike CW instrumentation, which only measures height, depth information is available, and the amount of light that is transmitted or reflected [55]. The TPSF may be transformed into the frequency domain (FD) using a Fourier transform, where the amplitude and phase correspond to the intensity and photon time of flight, respectively. Due to data types and the FD diffusion equation, this conversion is useful in the development of simpler reconstruction approaches [56]. But the limited number of modulation frequencies offered by FD platforms prevents the time-domain system from reaching its full potential. As a result, using information from higher-order moments of the TPSF improves the quality of the instrument over FD instruments [57]. Lasers are commonly employed in TD instrumentation because they emit reliable and narrow-width light pulses, but they are difficult to deploy and expensive. Commercial high-performance LED adoption has recently increased interest for TD systems because to the simplicity and low cost of LED technology in comparison to laser installations [43]. In both cases, the light is measured with a photon multiplier tube or a multi-channel photon multiplier for more sensitive applications. Both of them have great linearity, quick rising times, and high sensitivity, but they also have complicated circuitry, large sizes, and observable aging effects. In order to decrease "black noise" produced by the photon multi-

plier tube's heat and to prolong the functioning and life of the sensing elements, temperature management is necessary in applications that employ photon multiplier tubes. The avalanche photo-diode, often referred to as a solid-state photon multiplier tube, is an alternative to the photon multiplier tube and is typically employed in applications requiring little area and low power consumption at the expense of acquired sensitivity [43, 58]. Both an electron-multiplication and an intensified coupled charge device are often used in applications that need a high spatial resolution [58].

In 1967, Hundley reported about the first use of TD systems to the study of fluorescence [59]. TD systems were used in spectroscopy around the end of the 1980s to calculate concentrations of oxygenated and deoxygenated hemoglobin and hemoglobin [60], as well as in the estimation of myoglobin using the Beer-Lambert equation to fit the measured reflectance decay [61]. A key contribution from TD systems was the proof that the diffusion equation correctly predicts light propagation in a scattering medium [62]. Mammography and cancer studies [33, 63], adult brain imaging [64], research on the functional development of infant brain activity [65], fluorescence DOT aiming to improve signal-to-noise ratio by reconstructing the emitted time of flight of the fluorescent agent [66, 67], and small animal research for the development of new TD hardware and algorithms [68] are some clinical applications for time-domain systems. Analysis of significant tissue constituents including collagen, lipid, and bone is covered in other pilot studies [69]. These applications have to do with the study of oxygenated and hypoxic blood. When compared to the FD and CW domains, TD systems have higher costs and more complicated implementations [70]. Furthermore, TD measurements are noisier and display chromatic aberration effects, which are caused by the refractive index of the lens elements varying with wavelength and result in a failure to focus all of the light from the same location. This effect causes objects close to the boundary to deform even more [37], necessitating significant and meticulous calibrations in order to obtain accurate reconstructions [57]. Additionally, PMTs are complicated and more expensive since they need their own power and cooling systems. Another drawback is the slow measurement speed caused by the need for large-area, low noise, and effective detectors for single-photon counting of photons, which restricts the number of clinical and real-time applications [53].

2.2.3 Frequency-Domain systems

A megahertz amplitude modulated source is used in frequency domain (FD) instrumentation, a subset of temporal domain (TD) technology. The choice of frequency has a considerable impact on the temporal resolution of FD reconstructions, in contrast to TD systems, which get the highest resolution by taking into account all available frequencies [71]. The frequency domain is another way to represent

TD measurements, and it has the benefit of requiring fewer sources and detectors [40]. Numerical models fitted with FD observations have a higher chance of detecting global absorption and scattering patterns than a CW example. FD is less dependable than CW because it has a worse signal-to-noise ratio (SNR) at high frequencies [72]. The reconstruction of FD images is easier than that of TD, sometimes needing just one modulation frequency in the data [73]. The creation of FD equipment was made possible by the advancement of fluorescence spectrometers, sometimes referred to as fluorometers [74].

Following the development of laser technology and signal processing techniques, which enabled the use of the Fast Fourier Transform (FFT) to analyze modulated observations, fluorometers advanced toward tissue spectrometers [75]. Optical properties were calculated by fitting the analytical solution to modulation and phase measurements from tissue-like models (phantoms) and in-vitro experiments after it was established that the diffusion equation accurately described light propagation in tissue [76, 77]. Analytical solutions in various geometries were then obtained. By using LEDs in place of more expensive lasers or arc lamps, FD systems gained notoriety [78]. LEDs provided a wide spectrum range from 500 nm to 900 nm, simple modulation, a constant output, safety, and a lower cost. With only minor differences in their baseline and stability, LEDs and lasers provide outcomes that are equivalent when compared on a performance level [79]. In contrast to laser technology, which needs licenses and certificates, LED technology has changed how light is produced and offers a low-cost solution to DOI applications that can be quickly built in a lab and then put into line production. In most FD equipment, the modulation frequency is 100 MHz, which corresponds to the Fourier spectrum of time-domain observations at one frequency. This technique reduces the quality of image reconstruction by limiting the amount of content information available in FD systems. This constraint can be solved by utilizing multiple modulation frequencies up to 1 GHz, which increases the information richness of FD instrumentation and allows it to compete with TD systems [80].

Planar and tomographic optical mammography, which improves the visibility of breast lesions and vasculature [81], functional brain imaging, which ranges from small animals to adult humans [82, 83], and fluorescence diffuse optical tomography, where FD instrumentation aims to improve the image reconstruction resolution [84, 85] are clinical applications of frequency-domain instrumentation. Researchers and business investors are drawn to frequency domain instrumentation because it offers a trade-off between TD and CW instrumentation, although maintaining such equipment still requires a high level of skill and is more expensive than CW. FD systems are therefore more often used as a reference than a front-end application.

2.2.4 Hybrid systems

Different groups developed multi-modal or hybrid systems to increase the DOI instrumentation's capabilities. These hybrid imaging systems may combine the aforementioned DOI domains with other imaging modalities, such as MRI, X-ray, or ultrasound.

The fluorescent ultrasound device used for prostate guided biopsies [86], fluorescent-MRI equipment developed for in vivo atherosclerosis studies, and in vitro experiments to characterize tissue and carry out small animal studies [87, 88] are examples of hybrid systems coupled to other imaging modalities. The fluorescence-X-ray system presented in [89] employs high spatial sampling of photons throughout whole angle projections to target subcutaneous and lung tumors. Fluorescence DOT has also created a number of hybrid systems. A frequency or continuous wave device used for three-dimensional breast imaging, such as the one presented in [90] that uses a CW system capable of operating as a frequency domain, thus adding the characteristics of FD or TD to the simplicity of a CW instrument [91], is an example of a hybrid system that combines other domains. Another example is the FD-CW system, which allowed researchers to investigate the uterus' optical features [92]. Although more complicated than single optical imaging devices, hybrid systems provide the potential to experimentally utilize the physical limitations of a single apparatus or technique. The most important application for these systems in the research community is when they are used in conjunction with other medical imaging modalities to obtain more information about the tissue under investigation in a single test, rather than performing multiple tests that could change the expected results.

2.3 Conclusion

The variety of diffuse optical imaging equipment currently on the market includes a wide range of imaging domains, geometries, and source-detector configurations that have been used successfully in clinical applications for mammography, brain imaging, tracking the healing of wounds, and also in the development of cancer treatments by studies done on small animals. Despite the fact that a number of these instruments are in in the development stage, several DOI have found use in a variety of commercial applications, from molecular to organ size studies.

There is no definitive answer as to whether system domain—CW, TD, or FD—is superior. Although CW systems are more accessible and less expensive to use, they fall short in terms of sensitivity and spatial resolution when compared to FD and TD systems. However, given to the decreased complexity and cheaper cost that CW provides, this domain has been extensively employed and enhanced

through the employment of models that better characterize light propagation or by the coupling with other imaging modalities. Contrarily, TD instrumentation is the most comprehensive source of data for tissue characterisation, but it is also the most difficult and expensive to use. Although it is not as popular as CW instrumentation, FD might be seen as a trade-off between price and information quality.

Chapter 3

Light propagation models and imaging

Chapter 3 presents the formulation of the forward problem and outlines an overview of the analytical methods used to model the light propagation in tissue.

3.1 Introduction

Spatial and time scales govern the mathematical approach for describing light propagation. Maxwell equations at the microscopic level, the radiative transport equation at the mesoscopic level, and the diffusion equation, which disregards the interference effects of wave propagation, are all examples of the spatial scale [93]. Exogenous chromophores like photosensitive markers used in diagnostic and therapeutic trials, and chromophores like haemoglobin and melanin pigments are the main factors that affect light propagation and must be taken into account in order to properly model the light absorption and scattering events in biological tissue [94]. The abilities required include those in electronics, signal processing, data management, mathematical/statistical modeling, and some familiarity with biological applications. Another aspect to take into account for light propagation within tissue is the range of temporal responses that range from femto/picosecond, through nanoseconds for diffuse photon waves, to milliseconds that allow for less expensive hardware and where the majority of the physiological responses can be achieved [95]. This chapter discusses the mathematical models that have been used to simulate how light moves through biological tissue. For the so-called forward issue in two and three dimensions, theoretical and numerical approaches are presented. The Diffusion Equation, which is solved using the Finite Element Method and forms the basis for the creation of Reduced Order Models, is used to justify the use of the basic techniques for estimating light propagation in this

chapter.

3.2 Formulation of the forward problem

For a given distribution of optical parameters $u(r)$ and light sources $s(\eta)$, the forward issue $\Psi(r) \in \Omega$ includes calculating the photon fluence function throughout the domain of interest Ω and making the relevant measurements on the boundary $y(\eta) \in \partial\Omega$. Taking into account the locations of the S source s_i , $s_t \in \partial\Omega(i = 1, ;S)$ and the D detector $d_j \in \partial\Omega(j = 1, ..., D)$ that result in a total of $y_{i,j}$ observations. The relationship between optical tissue properties and measurements for each source-detector arrangement is then the forward problem.

$$y(i, j) = P_{i,j}(u(r)) \tag{3.1}$$

where $P(.) : U \rightarrow Y$ is the operator from the space of optical parameters U into the space of measurements Y [12]. Different methods have been proposed to cope with the forward problem given by equations. Eq. (3.1) described in the following sections.

3.3 Modeling light propagation by using analytical methods

The greatest description of how light propagates in a material is provided by this method, which uses Green's function to solve the diffusion equation (DE) [96,97]. However, the majority of its applications are restricted to imaging simple and homogeneous geometries with a single perturbation.

The infinite medium, used for light propagation on phantoms and in vitro experiments, the semi-infinite medium, used for topographic imaging [98], and the slab geometry, used for layered tissue research [99] are geometries that are often used in analytical solutions. The DE, presented in the CW form, has difficulties when it comes to describing light propagation with small source-detector separations. Additionally, early arriving photons in the TD situation are disregarded since DE validity requires photons to have experienced several dispersed events prior to measurement. The limitations of DE have been addressed analytically. Using a path-integral technique for two and four dimensions and an interpolation for three dimensions, [100] provides an exact solution to the time-domain RTE. Although precise, this approach took into account an infinite geometry that was challenging to use in an experimental setting. As an alternative RTE solution, and taking the work in [100] as the starting point, in [101], an heuristic analytical solution to the RTE for practical applications in the biomedical field have been

proposed. This approach allows it to be used to geometries like slabs, but it is only capable of performing accurate measurements when matched liquid is used. Although the formulation of analytical solutions is complicated due to the complexity of biological tissue, theoretically analytical solutions might be thought of as an exact and direct method to acquire light propagation.

3.4 Modeling light propagation by using stochastic methods

These models simulate the distributions of light intensity inside a tissue volume using probabilistic techniques. Given that photon propagation is stochastically dependent on elements like particle size and concentration, wavelength, and index of refraction, absorption and scattering events may be explained using stochastic models and probability density functions. The unpredictable route that photons follow inside a target medium is simulated by stochastic models. Each photon is "traced" as it passes through the medium; it is anticipated that at each site of interaction, the photon will give up some of its probabilistic weight in the form of absorbed energy and will then emerge from the interaction with a new weighting factor. A change in the photon's motion direction symbolizes the scattering action. A photon is either absorbed or exits the medium after multiple encounters, where some of them are detected by a detector. To determine the average dispersion of photons across the medium, it is essential to simulate a large number of photons [102]. Light propagation through biological tissue is one of several physical issues that may be solved using the stochastic Monte-Carlo (MC) simulation technique. The predicted values of a few random variables, which are comparable to the physical quantities to be determined, are used to build the stochastic model. By averaging several independent samples from the introduced random variables, the predicted values are calculated [26]. Monte Carlos simulations are used as a benchmark for other techniques, such as light propagation models based on the diffusion equation and the Radiative Transfer Equation, since they properly depict photon propagation through tissue [103]. A random number method with instructions to travel from one sample point to another point and the Random walk technique, which employs a lattice to discretize the medium, are two further stochastic ways [104, 105]. Using pixels to discretize the medium and the Markov property (probability of going forward, backward, or transitioning out of the pixel), Markov chain models is a stochastic approach[106]. Although stochastic approaches are extremely precise, they have practical limits since they need to simulate millions of photons in order to acquire useful data to characterize light propagation [107]. Research began developing techniques to speed up stochastic

solutions, such as scaling methods that use baseline simulations that are later scaled to fit a different medium, perturbation MC methods that are similar to the previous one but assume that optical properties are close to the baseline, hybrid methods that use a diffuse approximation to reduce computational load, or parallel computation techniques that take advantage of multiple processors[3].

3.5 Describing light propagation by using numerical methods

With an adequate balance between the quality and the time needed to acquire the light distribution inside a medium, numerically based imaging approaches have the ability to overcome the limits that analytical and stochastic models exhibit. Because they are suited to modeling light propagation through complicated, more realistic geometries and diverse media, numerical models are frequently utilized for imaging biological tissue. Additionally, they make it easier to combine near-infrared imaging techniques with other clinical imaging systems like MRI or X-ray, as the latter may be discretized into meshes that can be used to model light propagation.

Finite Difference Method

Partial Differential Equations (PDE) are solved using this numerical technique [108]. The technique entails discretizing the material using a regular grid and shaping complicated designs while taking the internal grid points into account. The absorption value for points outside the intended form is specified to be in the thousands. It has been proven that this strategy produces findings that accurately match other ways like Monte Carlo and analytical solution [109]. However, the FDM is not a method that is frequently utilized in DOT applications since the Finite Element Method is so straightforward when dealing with complicated geometries. Despite having been used to measure light distributions in the rat's skull [110] and the human brain [3].

Finite Element Method

In optical imaging applications, this is one of the recommended ways to solve the diffusion equation (DE). Using a set of basis functions, or interpolation functions, on a mesh, the approach transforms the PDE into a system of differential equations on a finite dimensional space [111]. Both the RTE and DE models have been solved using it [111, 112, 113]. Because it makes handling irregular geometries simple, it is frequently used in DOT and FDOT image reconstructions [114].

Finite Volume Method (FVM)

Partial differential equations are represented and evaluated using this technique. This approach computes values at discrete locations inside a mesh shape, much as

the finite difference and finite element methods. The little volume that surrounds each mesh node is where the approach gets its name [108]. The finite volume approach has been frequently utilized to resolve the RTE in optical tomography reconstructions because it conserves energy in a discrete sense [115]. It is computationally intensive since it gives a high degree of mesh adaptability that is helpful to model complicated geometries.

Boundary Element Method (BEM)

For large scale geometries, it has demonstrated superior performance than FEM [116, 117], but because of the intricacy of the boundaries encountered between tissue surfaces, it is unable to accurately represent light propagation in complex heterogeneous domains [118].

3.6 Conclusion

Light may be modeled using a variety of light propagation models, ranging from the precise but computationally costly radiative transfer equation to the more straightforward diffusion approximation models. Similar to the inverse problem, there are several solutions available today, some of which are more complicated and allow for more accuracy but are often time consuming.

Due to its simplicity and ability to explain tissue light interactions from comparatively large tissue volumes, the diffusion approximation (DA) to the RTE is the light transport model most frequently employed in diffuse optical tomography. The Finite Element Method is a prominent method for solving the DA because it allows for the modeling of light propagation via intricate geometries like those present in biological tissues.

Chapter 4

Mathematical formulation of Diffuse Optical Tomography

In this chapter, we collect some mathematical background that is crucial to this thesis but is difficult to obtain in a single reference. This chapter also aims to standardize notation and present some topics and abbreviations that will be used later in the thesis. To begin, some basic mathematical aspects of the Diffusion Equation (DE) are quickly reviewed because we used this model to model the forward problem in DOT.

4.1 The Radiative Transport Equation

The mathematical description of light propagation in random media is grouped according to spatial length scales. At the microscale, Maxwell Equations or geometric optics are usually utilized, at the mesoscale, the Radiative Transfer Equation (RTE) or transport equation, and at the macroscale, the Diffusion theory. This thesis will primarily focus on the mesoscale model, which deals with turbid media such as tissue, a cloud, or milk, in which light propagation acts as photons pursuing meandering courses through the media rather than a propagating ray or wave. [25] give a derivation of the transport equation using scalar wave models. The transition from the transport equation to the diffusive regime is shown below. According to the degree and frequency of contact between the particles and the medium in which they propagate, there are essentially three regimes of particle propagation in the Radiative Transfer model. The ballistic regime is defined as no interaction between particles and background material. When there is moderate interaction, it is termed the transport regime, and when there is substantial interaction, it is called the diffusive regime. In common optical and fluorescence tomography settings, three types of data acquisition methods exist: the continu-

ous wave system, the frequency domain system, and the time domain system. The frequency domain system monitors the modulation amplitude and phase shift of the light sources, whereas the continuous wave system measures the transmitted intensity at the detectors. In this thesis, we will concentrate on the continuous wave domain system for data acquisition. The continuous wave RTE model is used in this case. The RTE is a linearized version of the more general Boltzmann equation. The goal is to capture photographs of their optical characteristics in the near-infrared (NIR) range (600 - 1000nm) [119]. NIR light is redirected to the body by laser sources, and the amount of light transmitted is measured at detector points along the body's perimeter. It is assumed that photons travel through the medium at constant speed c, implying that the refractive index is constant. The one-speed continuous wave RTE is used in this thesis, with c normalized to 1. Let X be the space described by

$$X = \Omega \times S^{n-1} \tag{4.1}$$

where Ω is the domain of interest which is assumed to be a compact, convex subset of $\mathbb{R}^n$ with boundary $\partial X = (\partial\Omega \times S^{n-1})$, where $\partial X = ((x,\theta) \in \partial X, \theta.v > 0)$, is assumed to be Lipschitz, which implies that the outward normal vector $\hat{n}$ exists almost everywhere on ∂X. The function $u(x,\theta) \in X$ is physically considered to describe the density of the photons traveling in the region Ω at position x with velocity direction θ . The density of the photons u and the source term q are scalar quantities and are assumed non negative functions in $L^1(X)$. The following equation is called the Continuous wave RTE

$$\theta. \bigtriangledown u + (\mu_a(x) + \mu_s(x))u - \mu_s(x) \int_{S^{n-1}} \eta(\theta.\theta^{'})u(x,\theta^{'})\partial\theta^{'} = q(x,\theta) \quad inX; \tag{4.2}$$

with initial condition

$$u(x,\theta^{'}) = 0 \qquad in \quad \Omega \times S^{n-1} \tag{4.3}$$

and boundary condition

$$u(x,\theta) = 0 \qquad on \quad \partial X \tag{4.4}$$

Here, q is the source term given by

$$q(x,\theta) = \begin{cases} q_j(x,\theta) & x \in \bigcup_j^M p_j \\ 0 & x \in \partial\Omega \setminus \bigcup_j p_j \end{cases}$$

with p_j being source positions along $\partial\Omega$. These sources q_j are often picosecond laser pulses which can be described mathematically by the expression

$$q_j(x,\theta) = \delta_x(p_j)\delta_\theta(\theta_j) \tag{4.5}$$

where δ_x, and δ_θ, are Dirac delta distributions corresponding to the position, and angular , respectively; the coefficients μ_a and μ_s are the absorption and scattering coefficients of the medium, respectively, and θ is a unit direction vector pointing into the direction of velocity of the photons. The scattering function $\eta(\theta.\theta^{'})$, also known as the scattering kernel, is assumed to be independent of the position of the scattering event and to depend on the cosine of the angle denoted by ϑ, which is the angle between the initial direction θ and the deviated direction $\theta^{'}$ after the scattering event has occurred, that is, $\cos\vartheta = \theta.\theta^{'}$. Particle conservation in pure scattering events is expressed by the additional requirement

$$\int_{S^{n-1}} \eta(\theta.\theta^{'})d\theta^{'} = 1 \tag{4.6}$$

The Henyey Greenstein phase function is used in DOT in 2D by

$$\eta(\cos\vartheta) = \frac{1-g^2}{2\pi(1+g^2-2g\cos\vartheta)} \tag{4.7}$$

however in 3D it is described by

$$\eta(\cos\vartheta) = \frac{1-g^2}{4\pi(1+g^2-2g\cos\vartheta)^{\frac{3}{2}}} \tag{4.8}$$

where g is the mean cosine of the scattering function η with values between $-1 < g < 1$. In this model, values of g close to 1 indicate that the scattering is primarily forward directed, values close to zero indicate that the scattering is almost isotropic, whereas values near to -1 indicate that the scattering is primarily backward directed. Typical values for g based on animal tissue in Diffuse Optical Tomography are $0.9 \leq g \leq 0.99$ [120]. The mean free path is the mean distance traveled by a photon before changing its velocity direction, and is given in the Henyey-Greenstein model by

$$l = \frac{1}{(1-g)(\mu_a+\mu_s)} \tag{4.9}$$

The total attenuation coefficient an is defined as $a(x) = \mu_a(x) + \mu_s(x)$. The mean free path is thus $l = ((1-g)a) - 1$ using the total attenuation coefficient. The RTE is a conservation rule for number of photons. The simple derivation given below can be found in [25]. To begin, we define the angular photon current density as the vector number $\psi(x,\theta) = u(x,\theta)\theta$. The number of photons flowing out of a $\partial\Omega$ with velocity direction θ and for a differential surface element dS is given by $\psi(x,\theta).\vartheta dS$. The RTE represents the fluctuation of photon density inside a tiny but otherwise arbitrary region $\Upsilon \subseteq \mathbb{R}^n$ around location x, which originally travels with velocity direction θ.

4.2 Diffusion Approximation

We give an overview of the derivation of the diffusive RTE approximation. To begin, we define photon radiance $\phi(x)$ as

$$\phi(x) = \int_{S^{n-1}} u(x,\theta)d\theta \tag{4.10}$$

and the energy density current $J(x)$ by

$$J(x) = \theta u(x,\theta)d\theta \tag{4.11}$$

The diffusive approximation of the RTE may be obtained by expanding the angular component of the density of photons, phase, and source functions in spherical harmonics as follows:

$$u(x,\theta) = \sum_{l=N}^{\infty}\sum_{n=-l}^{l}\sqrt{\frac{2l+1}{4\pi}}\psi_{lm}(x)Y_{lm}(\theta) \tag{4.12}$$

$$q(x,\theta) = \sum_{l=N}^{\infty}\sum_{n=-l}^{l}\sqrt{\frac{2l+1}{4\pi}}q_{lm}(x)Y_{lm}(\theta) \tag{4.13}$$

$$\eta(\theta.\theta^{'}) = \sum_{l=N}^{\infty}\sum_{n=-l}^{l}\sqrt{\frac{2l+1}{4\pi}}\eta_{lm}(\theta.\theta^{'})Y^{*}_{lm}(\theta^{'}) \tag{4.14}$$

A P_N approximation of the RTE is obtained if the first N terms are taken into account of the above expressions, in particular the P_1 approximation is obtained after substituting (4.12) into (4.2) and truncating the series of (4.12) after the $l = 1$ terms, resulting in

$$\mu_a\phi(x) + \nabla J(x) = q_0(x) \tag{4.15}$$

$$\frac{J(x)}{3D(x)} + \frac{1}{3}\nabla\phi(x) = q_1(x) \tag{4.16}$$

Here $D(x)$ is known as the diffusion coefficient given by

$$D(x) = \frac{1}{3(\mu_a + \mu_s^{'})} \tag{4.17}$$

with

$$\mu_s^{'} = (1-g)\mu_s \tag{4.18}$$

where $\mu_s^{'}$ is known as the reduced scattering coefficient.

The first assumption is equivalent to requesting that the medium is scattering dominant, that is, $\mu_a << \mu_s^{'}$. After these assumptions, equation (4.16) converts into Fick's law

$$J(x) = -\zeta_0\nabla\phi(x). \tag{4.19}$$

Finally, substituting Fick's law (4.18) into equation (4.15), one obtains the diffusion approximation

$$-\nabla.\zeta_0\nabla\phi(x) + \mu_a\phi(x) = q_0(x). \tag{4.20}$$

As a result, the diffusion approximation is a subclass of the RTE's $P1$ approximation, in which the radiance function $\phi(x)$ does not rely on the direction velocity vector θ, as opposed to the RTE's density of photons $u(x, \theta)$.

4.3 Well posedness of the forward problem in DOT

The DOT forward problem requires solving an elliptic partial differential equation with Robin boundary conditions in which μ_a and D are known. The photon density of the scattered light arriving at the detectors is described by the solution (u). The whole DOT experiment is presented in the continuous wave domain, using the Robin boundary condition.

$$-\nabla.(D\nabla u) + \mu_a u = 0 \quad in \quad \Omega \tag{4.21}$$

$$\gamma_R u = u + 2D\frac{\partial u}{\partial n} = f \quad on \quad \partial\Omega \tag{4.22}$$

here, D is the diffusion coefficient, μ_a is the absorption coefficient, f is the source, and

$$\gamma_R : H^1(\Omega) \to H^{\frac{-1}{2}}(\partial\Omega) \tag{4.23}$$

is the Robin trace map, with the solution space defined as:

$$H^1(\Omega) = v \in L^2(\Omega)| \int_\Omega (|\nabla v|^2 + v^2)dx < \infty \tag{4.24}$$

and

$$H^{\frac{1}{2}}(\partial\Omega) = (\gamma_D(v)|v \in H^1(\Omega)/H_(0)^1(\Omega)) \tag{4.25}$$

We also consider$\Omega\mathbb{R}^\kappa$, where $n = 2, 3$ a bounded, connected Lipschitz domain. Consider the Robin problem described previously

$$-\nabla.(D\nabla u) + \mu_a u = 0 \quad in \quad \Omega \tag{4.26}$$

$$\gamma_R u = u + 2D\frac{\partial u}{\partial n} = f \quad on \quad \partial\Omega \tag{4.27}$$

where $u \in H^1(\Omega)$, $g \in H^{\frac{1}{2}}(\partial\Omega)$ and $q = (D, \mu_a) \in Q^-$, where

$$Q^- = (D, \mu_a) \in L^\infty(\Omega) \times L^\infty(\Omega) : 0 < D_0 < D < D_1, 0 < \mu_0 < \mu_a < \mu_1 \tag{4.28}$$

To explain the Robin problem's well-posedness, we first derive the weak formulation [121]. We have $vinH_1(\Omega)$ for any test function.

$$-\int \nabla.(D\nabla u)\bar{v}dx + \mu_a u\bar{v})dx = 0 \tag{4.29}$$

which implies that

$$-\int_\Omega D\nabla u.\nabla\bar{v}dx + \int_\Omega \mu_a dx u\bar{v} - \int_{\partial\Omega} D\frac{\partial u}{\partial n}\bar{v}ds = 0 \tag{4.30}$$

where $\bar{v}$ represents the complex conjugate of v. Now using the boundary condition (4.24), we get

$$\int_\Omega D\nabla u.\nabla\bar{v}dx + \int_\Omega \mu_a dx = \frac{1}{2}\int_{\partial\Omega}(g-u)\bar{v}ds \tag{4.31}$$

which implies

$$\int_\Omega D\nabla u.\nabla\bar{v}dx + \int_\Omega \mu_a dx + \frac{1}{2}\int_{\partial\Omega} u\bar{v}ds = \frac{1}{2}\int_{\partial\Omega} g\bar{v}ds \tag{4.32}$$

Defining

$$B(u,v) = \int_\Omega D\nabla u.\nabla\bar{v}dx + \int_\Omega \mu_a u\bar{v}dx + \frac{1}{2}\int_{\partial\Omega} u\bar{v}ds \tag{4.33}$$

and

$$f_R(v) = \frac{1}{2}\int_{\partial\Omega} g\bar{v}ds \tag{4.34}$$

we have the weak formulation of the robin problem as, $B(u,v) = f_R(v)$, for any $u, v \in H^1(\Omega)$.

Definition 1. *A map $b : V \times V \to \mathbb{C}$, is called a sesquilinear form, if it satisfies the following conditions,*

$$b(x+y, z+w) = b(x,z) + b(y,z) + b(x,w) + b(y,w)$$
$$b(c_1x, c_2y) = c_1c_2b(x,y)$$

where $x, y, z, w \in V$ and $c_1, c_2 \in \mathbb{C}$.

Lemma 1. *$B(u,v) = \int_\Omega D\nabla u.\nabla\bar{v}dx + \int_\Omega \mu_a u\bar{v}dx + \frac{1}{2}\int_{\partial\Omega} u\bar{v}ds$ is a sesquilinear form.*

Proof. For any $u, v, w, z \in H^1(\Omega)$, we have

$$\begin{aligned}
B(u+v, z+w) &= \int_\Omega D\nabla(u+v).\nabla(\bar{z}+\bar{w})dx + \int_\Omega \mu_a(u+v)(\bar{z}+\bar{w})dx + \frac{1}{2}\int_{\partial\Omega}(u+v)(\bar{z}+\bar{w})ds \\
&= \int_\Omega D\nabla u.\nabla\bar{z}dx + \int_\Omega \mu_a u\bar{z}dx + \frac{1}{2}\int_{\partial\Omega} u\bar{z}ds + \int_\Omega D\nabla u.\nabla\bar{w}dx + \int_\Omega \mu_a u\bar{w}dx \\
&+ \frac{1}{2}\int_{\partial\Omega} u\bar{w}ds + \int_\Omega D\nabla v.\nabla\bar{z}dx + \int_\Omega \mu_a v\bar{z}dx + \frac{1}{2}\int_{\partial\Omega} v\bar{z}ds \\
&+ \int_\Omega D\nabla v.\nabla\bar{w}dx + \int_\Omega \mu_a v\bar{w}dx + \frac{1}{2}\int_{\partial\Omega} v\bar{w}ds \\
&= B(u,z) + B(u,w) + B(v,z) + B(v,w).
\end{aligned}$$

For any $c_1, c_2 \in \mathbb{C}$, we have

$$\begin{aligned}
B(c_1u, c_2v) &= \int_\Omega D\nabla c_1 u.\nabla \bar{c_2} v dx + \int_\Omega \mu_a c_1 u \bar{c_2} v dx + \frac{1}{2}\int_{\partial\Omega} c_1 u \bar{c_2} v ds \\
&= c_1\bar{c_2}(\int_\Omega D\nabla u.\nabla \bar{v} dx + \int_\Omega \mu_a u\bar{v} dx + \frac{1}{2}\int_{\partial\Omega}\int_\Omega D\nabla c_1 u.\nabla \bar{c_2} v dx + \int_\Omega \mu_a c_1 u \bar{c_2} v dx \\
&+ c_1 u\bar{c_2}\frac{1}{2}\int_{\partial\Omega} u\bar{v} v ds) \\
&= c_1\bar{c_2} B(u, v).
\end{aligned}$$

Hence $B(u, v)$ is a sesquilinear form. □

The Lax-Milgram theorem for sesquilinear form will be used to demonstrate the existence and uniqueness of the solution to the Robin issue. We define the $H_1(\Omega)$ as the inner product for the space.

$$\langle u, v\rangle_{H^1(\Omega)} = \int_\Omega (\nabla u.\nabla \bar{v} + u\bar{v}) dx \tag{4.35}$$

In order to show the boundedness and coerciveness of $B(u, v)$, it is convenient to define an equivalent norm for the space $H^1(\Omega) \times H^1(\Omega)$ as

$$||u||_*^2 = \int_\Omega D|\nabla u|^2 dx + \int_\Omega \mu_a |u|^2 dx + \frac{1}{2}\int_{\partial\Omega} |u|^2 ds \tag{4.36}$$

Lemma 2. *$\|u\|_{H_1}$ and $\|u\|_*$ are equivalent.*

Proof.

$$\begin{aligned}
\|u\|_{H_2} &= \int_\Omega |\nabla u|^2 + |u|^2 dx \\
&\leq \int_\Omega \frac{D(x)}{D_0(x)}|\nabla u|^2 + \frac{\mu_a(x)}{\mu_0(x)}|u|^2 dx \\
&= \frac{1}{D_0(x)}\int_\Omega D|\nabla u|^2 dx + \frac{1}{\mu_0(x)}\int_\Omega \mu_a |u|^2 dx \\
&\leq \max\frac{1}{D_0(x)}, \frac{1}{\mu_0(x)}(\int_\Omega D|\nabla u|^2 + \mu_a|u|^2 dx + \frac{1}{2}\int_{\partial\Omega}|u|^2 ds) \\
&\leq C_1||u||_*^2
\end{aligned}$$

□

Lemma 3. *$B(u, v)$ is bounded and coercive.*

Proof. By using Cauchy-Schwartz inequality we obtain

$$\begin{aligned}
|B(u,v)| &= |\int_\Omega D\nabla u\nabla\bar{v}dx + \int_\Omega \mu_a u\bar{v}dx + \int_{\partial\Omega}\frac{1}{2}u\bar{v}ds \\
&\leq (\int_\Omega |D\nabla u|^2dx)^{\frac{1}{2}}(\int_\Omega |\nabla v|^2dx)^{\frac{1}{2}} + (\int_\Omega |\mu_a u|^2dx)^{\frac{1}{2}}(\int_\Omega |v|^2dx)^{\frac{1}{2}} + \\
&(\int_{\partial\Omega}|u|^2ds)^{\frac{1}{2}}(\int_{\partial\Omega}|v|^2ds)^{\frac{1}{2}} \\
&\leq D_1(\int_\Omega |\nabla u|^2dx)^{\frac{1}{2}}(\int_\Omega |\nabla v|^2dx)^{\frac{1}{2}} + \\
&\mu_1(\int_\Omega |u|^2dx)^{\frac{1}{2}}(\int_\Omega |v|^2dx)^{\frac{1}{2}} + \frac{c_1}{2}(\int_\Omega |\nabla u|^2dx)^{\frac{1}{2}}(\int_\Omega |\nabla v|^2dx)^{\frac{1}{2}} \\
&\leq C((\int_\Omega |\nabla u|^2dx)^{\frac{1}{2}}(\int_\Omega |\nabla v|^2dx)^{\frac{1}{2}} + (\int_\Omega |u|^2dx)^{\frac{1}{2}}(\int_\Omega |v|^2dx)^{\frac{1}{2}}) \\
&\leq C||u||_{H^1(\Omega)}||v||_{H^1(\Omega)}
\end{aligned}$$

where $C = max(D_1 + c_2 + \mu_1)$. Hence $B(u,v)$is bounded. Next We need to show $B(u,v)$ is coercive.

$$\begin{aligned}
|B(u,u)| &> |Re(B(u,u)| \\
&> |\int_\Omega D|\nabla u|^2dx + \int_\Omega \mu_a|u|^2dx + \int_{\partial\Omega}\frac{1}{2}|u|^2ds| \\
&= ||u||_*^2
\end{aligned}$$

Hence $B(u,v)$ is coercive. □

To fulfill all the assumptions of the Lax-Miligram theorem, we need to prove that $f(v) = \frac{1}{2}\int_{\partial\Omega} g\bar{v}ds$ is a bounded linear functional.

Lemma 4. *$f_R(v)$ is a bounded linear functional.*

Proof. Clearly, $f_R(v)$ is linear. We have the duality pairing,

$$f_R(v) =< g, v >_{H^{\frac{-1}{2}}(\partial\Omega)\times H^{\frac{1}{2}}(\partial\Omega)}$$

By Riesz representation theorem, there exists a bounded linear operator $S : H^{\frac{-1}{2}} \to H^{\frac{1}{2}}$, such that

$$\begin{aligned}
| < g, v >_{H^{\frac{-1}{2}}(\partial\Omega)\times H^{\frac{1}{2}}(\partial\Omega)} | &= | < Sg, v >_{H^{\frac{1}{2}}(\partial\Omega)\times H^{\frac{1}{2}}(\partial\Omega)} | \\
&\leq ||Sg||_{H^1(\Omega)}||v||_{H^1(\Omega)} \\
&\leq c||v||_{H^1(\Omega)}
\end{aligned}$$

Thus$f_R(v)$ is a bounded linear operator. □

The hypotheses of the Lax-Miligram theorem are satisfied by Lemma 3-4. As a result, we may deduce that there exist a unique solution to (4.23). (4.24).

Chapter 5

Image reconstruction in diffuse optical tomography using adaptive moment gradient based optimizers: A statistical study

In this chapter, the inverse and shape reconstruction problems in DOT are formulated. we will examine the convergence behavior of Adam, Nadam and AmsGrad optimizers when applied to the problem of DOT. A comparison between these optimizers will be investigated and discussed. We will characterize the performance of these algorithms with respect to the choice of some hyper parameters and the initial guess error. Finally, numerical results are presented using these methods.

5.1 Introduction

In recent years, the problem of DOT is becoming more attractive since it presents many advantages. It is a non invasive, non ionizing and an inexpensive technique compared to other imaging modalities such as Magnetic Resonance Imaging (MRI) and X-ray [122,123,124]. DOT has been applied to detect breast tumors [125,126,127,128], brain injuries [129,130], imaging newborn infant head [131], and provides some important information about tissue metabolism . Solving the DOT problem involves addressing the radiative transfer equation (RTE) that describes the light propagation in biological tissues [132,133]. However, the RTE does not have an analytical close form solution for complex geometries and its numerical alternative is computationally expensive. Since the diffusion approximation (DA) of the RTE is easy to implement we will use it as the forward model throughout this work.

It is a well known fact that the inverse problem in DOT is non linear and severely ill posed. Gradient based methods are commonly used to solve minimization problems in optical tomography [132].

In recent years, a number of new optimizers have been proposed to tackle the problem of convergence when there is insufficient prior knowledge to elect a good learning rate. One of the most popular and practical techniques used to control the distance of each step. The Adaptive moment estimation (Adam)is one of the first adaptive moment optimizers proposed in literature and was presented by Diedriek kingma and Jimmy Ba [134]. It is a combination of adaptive gradient algorithm (AdaGrad) [135] and Root Mean Square propagation with momentum (RMS prop) [136]. Adam is an efficient optimizer that only requires first order gradients and uses square gradients to scale the learning rate implementing momentum by using the moving average of the gradient rather than the gradient itself. To cope with the shortcoming of Adam, mainly the lack of convergence guaranties, a number of variants of Adam algorithm have been derived lately such as Nesterov-accelerated Adaptive Moment Estimation(Nadam) [137] and AmsGrad optimizer [138]. For more details we refer the reader to [139,140].

In this work, we will examine the convergence behavior of Adam, Nadam and AmsGrad optimizers when applied to the problem of DOT. A comparison between these optimizers will be investigated and discussed. We will characterize the performance of these algorithms with respect to the choice of some hyper parameters and the initial guess error. To evaluate the quality of reconstructed images by the algorithms in quantitative manner, we use quality metrics, such as the structural similarity index (SSIM) and the Peak signal-to-noise ratio (PSNR) on the reconstructed images.

The structure of this chapter is as follows: In Section 5.2 we give an overview of the mathematical formulation of the diffusion approximation in continuous wave (CW) case. In Section 5.3 we describe the inverse problem and the algorithms we use to reconstruct the absorption coefficient of DOT. In Section 5.4 we show the results of our statistical analysis of the simulation data.

5.2 Forward problem

In this section we describe the mathematical formulation of the diffusion approximation (DA).
Let $\Omega \subset R^n$,$n = 2, 3$, be our domain of interest, and $\partial\Omega$ the boundary of Ω. Then the DA inside the domain Ω satisfies the partial differential equation

$$-\nabla[.D(r)\nabla\Phi(r)] + \mu_a(r)\Phi(r) = 0 \qquad r \in \Omega \tag{5.1}$$

with the Robin-boundary condition

$$\Phi(r) + 2aD(r)\frac{\partial \Phi(r)}{\partial \hat{n}} = S(r) \qquad r \in \partial\Omega \tag{5.2}$$

where $\Phi(r)$ is the photon density, $D(r)$ is the diffusion coefficient defined by $D(r) = \frac{1}{3(\mu_a+\mu_s')}$. a is the Fresnel reflection coefficient, which depends on the mismatch between the refractive indices, μ_a and μ_s the absorption and scattering coefficient respectively, and μ_s' the reduced scattering coefficient expressed as $\mu_s' = (1-g)\mu_s$, where g is the anisotropic factor. $S(r)$ describes the boundary condition for the incoming radiation and $\hat{n}$ is the outward normal vector to Ω.

We assume that the medium is highly scattering such that, $\mu_a \ll \mu_s$.The forward model (5.1)-(5.2) is solved by using the finite element method as described in [141].

5.3 Inverse problem

The inverse problem we are interested in consists of determining the couple (μ_a, μ_s) from the set of true data y_i such that

$$F_i(\mu_a, \mu_s) = y_i \qquad 1 \leq i \leq m \tag{5.3}$$

where we denote by F_i the forward operator which is assumed to be Fréchet differentiable, and y_i the approximate measured data. In this study, we restrict our attention to the reconstruction of the absorption coefficient, and we assume that the distribution of the scattering coefficient is known. Then, the objective function can be written as follows

$$J(\mu_a) = \frac{1}{2}\sum_{i=1}^{m}(F_i(\mu_a) - y_i)^2 \tag{5.4}$$

Then this problem can be stated in term of an optimization problem

$$\mu_a^* = argminJ(\mu_a) \tag{5.5}$$

Since the inverse problem is ill-posed, it requires regularization. A Total Variation regularization is applied [142]. By adding a regularization term, the cost function is formulated as

$$J_R(\mu_a; \lambda, \mu_a^0) = \frac{1}{2}\sum_{i=1}^{m}(F_i(\mu_a) - y_i)^2 + \lambda R(\mu_a) \tag{5.6}$$

where $R(\mu_a) = \|\mu_a - \mu_a^0\|^2$, is the regularization operator that enforces smoothness conditions in the solution, and λ is the regularization parameter. μ_a^0 denoted the

initial guess error.The forward operator F_i is linearized around some initial guess μ_a^0.

$$F_i(\mu_a) = F_i(\mu_a^0) + F_i'(\mu_a^0)(\mu_a - \mu_a^0) + W(\mu_a^0, i) \tag{5.7}$$

where F_i' is the Fréchet derivative of the forward operator F_i, and W denotes the Taylor remainder for the linearization around μ_a^0.
The gradient of the objective functional can be written as follows

$$\nabla J_R(\mu_a; \lambda, \mu_a^0) = \sum_{i=1}^{m} F_i'(\mu_a)^*(F_i(\mu_a) - y_i) + \lambda R'(\mu_a) \tag{5.8}$$

where $R'(\mu_a)$ is the Fréchet derivative of regularization operator with respect to μ_a.

5.4 Iterative inverse problem solution

We consider an iterative optimization algorithm denoted by Q. the statement of our problem can be reduced to the iterative form:

$$\begin{cases} \mu_0 = \mu_a^0 \\ \mu_{n+1} = Q(\mu_n; J_R) \end{cases} \tag{5.9}$$

Naturally, the promise of the algorithm is to get us closer to the solution after each step in an iterative manner. The proof of convergence of any specific algorithm ensures that

$$\lim_{n \to +\infty} Q(\mu_n; J_R) = \mu_a^* \tag{5.10}$$

and can give even more information on the speed of convergence by deriving a theoretical formula of $\|Q(\mu_n; J_R) - \mu_a^*\|$ as a bounded formula of n.

In general, this is a hard formula to derive, and it is even more difficult when dealing with complex problems like DOT, with many multidimensional parameters. In practice, the convergence speed is influenced by many factors, related to the algorithm itself, and the configuration of the problem (physical reality and constraints). A numerical approach based on simulation and statistical analysis, will prove to be very useful in tackling with this kind of hard situations, and can help us to gain more insight in the choice of optimization algorithm and all other practical purpose. As we consider in our study a family of optimization algorithms based on gradient descent, We can point out the learning rate hyper parameter as the main factor of interest in this context.

From a practical point of view, J_R depends on the structure of the problem, and by consequence, J_R depends on different factors like the nature of inclusions (their number, form, distribution ....), the properties of the medium, and all other

parameters that shape the above forward problem as stated in the previous section. Also, it depends on the choice of the regularization and initial guess. Table 5.1 below give an example of the parameters and hyper parameters that can be of interest in studying the practical optimization problem (including the iterative algorithm hyper parameters).

Table 5.1: Parameters and hyper parameters of interest.

Parameters of the problem	Hyper parameters of the optimization algorithm
n: number of inclusions	β: learning rate.
D: distribution shape of the inclusion.	μ_a^0: initial guess.
N_d : number of detectors.	λ: regularization coefficient.
N_s : number of sources.	

A more focused statement of the iterative optimization algorithm, to the following study in the present work can be formulated as

$$\begin{cases} \mu_0 = \mu_a^0 \\ \mu_{n+1} = Q_{AM}(\mu_n; n, \beta, \Theta) \end{cases} \tag{5.11}$$

where Q_{AM} describes the adaptive moment algorithm, μ_a^0
the initial guess, n the number of inclusions, β denotes the learning rate hyper parameter and Θ represent all the remaining parameters.

Hereafter, we address our attention only to the number of inclusions (n), the learning rate β, and the initial guess μ_a^0. In our implementation of the optimization problem, we used an objective function C defined as

$$C(l; \beta, n, \mu_a^0) = \frac{1}{2}(J_R(Q_{AM}(\mu_l; n, \beta)) + \epsilon_0 + |J_R(Q_{AM}(\mu_l; n, \beta) - \epsilon_0|))) \tag{5.12}$$

We can easily show that

$$C(l; \beta, n, \mu_a^0) = \begin{cases} J_R(Q_{AM}(\mu_l; n, \beta)) & if J_R(Q_{AM}(\mu_l; n, \beta) > \epsilon_0 \\ \epsilon_0 & if J_R(Q_{AM}(\mu_l; n, \beta) \leq \epsilon_0 \end{cases} \tag{5.13}$$

We define the number of iterations to convergence by

$$N_{Q_{AM}} = min(argmax_l(C(l; \beta, n, \mu_a^0)), L_{max}) \tag{5.14}$$

This formulation guarantees that our optimization algorithm will stop whenever $J_R(Q_{AM}(\mu_l; n, \beta)$ is lower than ϵ_0 or l is greater than L_{max}, where $\epsilon_0 > 0$ and $L_{max} \geq 1$, are parameters used in iteration stopping criteria, which is explicitly set

in this study to be either when the cost function is lower than ϵ_0 or the number of iteration exceeds L_{max}.

The aim of our numerical statistical study of convergence speed can then be brought down to the study of properties of the $N_{Q_{AM}}$ probability distribution $P(N_{Q_{AM}}|n, \beta, \mu_a^0)$ using simulation tools. In the following study, we restrict our attention to the comparison of three algorithms based on the adaptive moment procedure. For more details, we refer the reader to the next section.

5.5 Simulation and data generation

As mentioned before, only the absorption coefficient is reconstructed and discussed. The distribution of the scattering coefficient is assumed to be known. To generate synthetic data, we use the Toast++ software [20], which solve the forward problem (5.1)-(5.2) described above, using finite element method. In all the numerical simulations, a circular domain of radius 20 mm which contains different inclusion sizes and shapes is performed. To avoid inverse crime [143], we use different meshes in the forward and inverse problem. In all cases, we use a circular mesh with 22011 nodes and 43400 tetrahedral elements for the forward problem and 15408 nodes and 30308 tetrahedral elements for the inverse problem. 16 sources and 16 detectors are located on the boundary of the domain with equal distance. The location, size and number of anomalies in μ_a are chosen randomly with a background $\mu_a^{bkg} = 0.01mm^{-1}$ and $\mu_s^{'bkg} = 2mm^{-1}$. We consider that there is no change in the anisotropic factor g which is taken to be equal to 0.9. The regularization parameter λ is set to be equal to 10^{-8}. To solve the minimization problem (5.5) we use algorithms described in pseudo codes below, where β is the learning rate, ρ_1 and ρ_2 are the exponential decay rates for the moment estimates. The parameter of stabilization ϵ is set to be equal to 10^{-10}.

To control all the parameters of our simulation, we first control the error of the initial guess of reconstruction μ_a^0 by taking it to be

$$\mu_a^0 = \mu_a^{real} + \alpha \tag{5.15}$$

where μ_a^{real} is the original image matrix used to solve the forward problem and α is a random matrix variable sampled uniformly such as$\|\alpha\|_{\text{inf}} = \delta$, where δ itself is a uniformly random number taken in range [0,0.2]. We define δ as the initial guess error.

The choice of the learning rate is very important. For this purpose, a preliminary study has been conducted where we experimented with the learning rate of all optimizers in the range [0,0.5], and noticed that when the learning rate is out of the interval [0.001,0.3], the minimization of the objective function does not converge or take a very long time to do. For the sake of this simulation, the learning

Algorithm 1: Pseudocode of Adam.

Require: μ_a^0, β, ρ_1, ρ_2, and ϵ with $\rho_1, \rho_2 \in [0, 1)$

Ensure: μ_a^k

while J not converged **do**

$k \leftarrow k + 1$

$g_k \leftarrow \nabla J_{\mu_a}(\mu_a^{k-1})$

$m_k \leftarrow \rho_1.m_{k-1} + (1 - \rho_2).g_k$

$v_k \leftarrow \rho_1.v_{k-1} + (1 - \rho_2).g_k^2$

$\hat{m_k} \leftarrow \frac{m_k}{(1-\rho_1^k)}$

$\hat{v_k} \leftarrow \frac{v_k}{(1-\rho_2^k)}$

$\mu_a^k \leftarrow \mu_a^{k-1} - \beta \frac{\hat{m_k}}{\sqrt{\hat{v_k}}+\epsilon}$

end while

Algorithm 2: Pseudocode of Nadam

Require: μ_a^0, β, ρ_1, ρ_2, and ϵ with $\rho_1, \rho_2 \in [0, 1)$

Ensure: μ_a^k

while J not converged **do**

$k \leftarrow k + 1$

$g_k \leftarrow \nabla J_{\mu_a}(\mu_a^{k-1})$

$m_k \leftarrow \rho_1.m_{k-1} + (1 - \rho_2).g_k$

$v_k \leftarrow \rho_1.v_{k-1} + (1 - \rho_2).g_k^2$

$\hat{m_k} \leftarrow \frac{m_k}{(1-\rho_1^k)}$

$\hat{v_k} \leftarrow \frac{v_k}{(1-\rho_2^k)}$

$\mu_a^k \leftarrow \mu_a^{k-1} - \frac{\beta}{\sqrt{\hat{v_k}}+\epsilon}(\rho_1 \hat{m}_k + \frac{1-\rho_1}{1-\rho_1^k} g_k)$

end while

rate is constrained to be chosen uniformly from the range [0.001,0.3]. We fixed all the other hyper parameters for all optimizers, to the recommended values from the corresponding literature (momentum ρ_1 and ρ_2 are 0.9 and 0.999 respectively). The number of anomalies is taken among 1, 2 and 3 equi-proportionally randomly.

Figure 5.1 shows the resulting distributions (histograms) from running the

Algorithm 3: Pseudocode of AmsGrad

Require: μ_a^0, β, ρ_1, ρ_2, and ϵ with $\rho_1, \rho_2 \in [0, 1)$

Ensure: μ_a^k

while Jnot converged **do**

$k \leftarrow k + 1$

$g_k \leftarrow \nabla J_{\mu_a}(\mu_a^{k-1})$

$m_k \leftarrow \rho_1.m_{k-1} + (1 - \rho_2).g_k$

$v_k \leftarrow \rho_1.v_{k-1} + (1 - \rho_2).g_k^2$

$\hat{v_k} \leftarrow \max(v_k, \hat{v_{k-1}})$

$\mu_a^k \leftarrow \mu_a^{k-1} - \frac{\beta}{\sqrt{\hat{v_k}+\epsilon}} m_k$

end while

simulation, for the three parameters of the study, the learning rate β, the number of anomalies, and the perturbation coefficient δ. To be fair, we use the same parameters for all optimizers in each simulation instance.

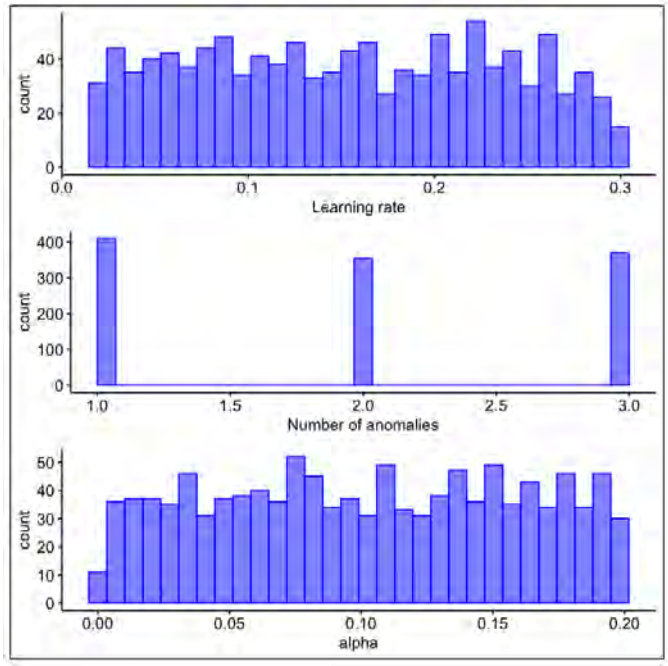


Figure 5.1: Distribution of the learning rate hyper parameter β on the top, number of inclusions n in the middle, and the initial guess error δ on the bottom.

5.6 Results

In this part of our study, we will characterize the convergence rate of the three algorithms ,and compare the convergence/divergence behavior in relation to the parameters of simulation, and finally we will examine the quality of the resulting reconstructions of the three optimizers.

First of all, we choose in the context of this present analysis, the definition of divergence to be the state of the running optimization when the error minimization didn't improve for longer than 200 iterations in total.

The first subject of focus is the convergence rate of each algorithm. Let X_{AD}, X_{NAD}, and X_{AMS} be three random variables representing the state of convergence for Adam, Nadam, and AmsGrad optimizers. This variables take values 0 or 1, depending on either the corresponding algorithm diverges or converges, such that:

$$\begin{cases} X_{AD} = 0 \text{ if Adam diverge} \\ X_{AD} = 1 \text{ if Adam converge} \end{cases}$$

$$\begin{cases} X_{NAD} = 0 \text{ if Nadam diverge} \\ X_{NAD} = 1 \text{ if Nadam converge} \end{cases}$$

$$\begin{cases} X_{AMS} = 0 \text{ if AmsGrad diverge} \\ X_{AMS} = 1 \text{ if AmsGrad converge} \end{cases}$$

The simulation provided us with Three samples of independent and identically distributed $(X_{AD,n})_{n\leq 1340}$, $(X_{NAD,n})_{n\leq 1340}$,and $(X_{AMS,n})_{n\leq 1340}$. To statistically estimate the rates of convergence, namely $P(X_{AD} = 1) = p_{AD}$, $P(X_{NAD} = 1) = p_{NAD}$, and $P(X_{AMS} = 1) = p_{AMS}$
we use the three estimators

$$\begin{cases} \hat{p}_{AD} = \frac{\sum \#(X_{AD}=1)}{N} \\ \hat{p}_{NAD} = \frac{\sum \#(X_{NAD}=1)}{N} \\ \hat{p}_{AMS} = \frac{\sum \#(X_{AMS}=1)}{N} \end{cases}$$

where # denotes the count function, to construct 95% confidence intervals based on the large number normal approximation, as presented in Table 5.2.

Table 5.2: The 95 percent non parametric confidence intervals.

Target proportion	Confidence interval lower bound	Confidence interval upper bound
$\hat{p}_{AD}$	0.87	0.9
$\hat{p}_{NAD}$	0.87	0.91
$\hat{p}_{AMS}$	0.88	0.9

To shed light on the influence of simulation parameters on convergence rates, we run a logistic regression to estimate the conditional distributions $P(X|\beta, n, \delta)$ where $X \in \{X_{AD}, X_{NAD}, X_{AMS}\}$, n denotes the number of anomalies in the image, and the remaining variables β and δ are as mentioned earlier. The result of this procedure is depicted in Table 5.3 presenting p-values for the statistical significance of each regression parameter.

Table 5.3: Logistic regression coefficients and corresponding p-values.

	Adam	Nadam	AmsGrad
β	$(1.17, < 2e^{-16})$	$(1.22, < 2e^{-16})$	$(-0.35, < 1.5e^{-5})$
n	$(-0.005, 0.6)$	$(-0.007, 0.44)$	$(-0.044, < 7.6e^{-16})$
δ	$(0.07, 0.6)$	$(0.1, 0.5)$	$(-0.016, 0.91)$

From Table 5.3, we conclude that the main parameter that same to have significant influence on convergence of these algorithms is the learning rate hyperparameter. Since the logistic coefficient forβ is positive for Adam and Nadam, the larger the learning rate, the more guarantee there is for the algorithm to converge. This statement is reversed for AmsGrad, as we observe a negative coefficient for learning rate. The AmsGrad is also impacted negatively with the number of anomalies in the image.

Running our experiment simulation provided us with 1340 convergent instances in total (That means where $(X_{AD} = 1, X_{NAD} = 1, X_{AMS} = 1)$), to evaluate the comparative performance between optimizers in term of speed of convergence as measured by the number of iterations taken by each optimizer to reach the solution, We will conduct a statistical analysis on the generated data, comparing first the speed globally between optimizers and then relating it to the variables of

simulation such as the initial guess error, the choice of learning rate and the number of anomalies in the image. Also, the influence of this variables on reconstructed image quality will be discussed, PSNR and SSIM score values are calculated for each simulation instance, we kindly refer the reader to later discussions about reconstruction quality in this work for more informations on these scores.

A number of statistical methods have been applied and results are examined to describe the convergence speed behavior of each algorithm when applied to the inverse problem of DOT.

Image reconstruction in optical tomography is an ill-posed nonlinear inverse problem, the algorithms based on gradient descent present no guaranty to converge to the global minima when there are local minima in the optimization problem at hand, the convergence point depends heavily on the choice of the starting point of the optimization, and generally, these algorithms converge (depending also on the learning rate) to the nearest local minima to the initial starting point.

In this section we address the optimization problem (image reconstruction) from the perspective of the speed of convergence (as one of the very important matters in practical use of DOT in clinical applications) rather than sensitivity of the algorithms to the choice of the initial guess with respect to their efficiency to find global minima (which is the other important practical issue in applying DOT), this last perspective is equally relevant and without doubt needs particular attention and further analysis, but in the scope of our current work remains an open question to follow up, as our randomized simulation design was focused on controlling the factors that influence speed of convergence, we can use the same approach as in this work to quantify (statistically speaking) the efficiency and sensitivity to reach the global minima depending on problem factors, but this needs obviously to redesign the simulation to generate the appropriate data suitable for this substantially different analysis objective.

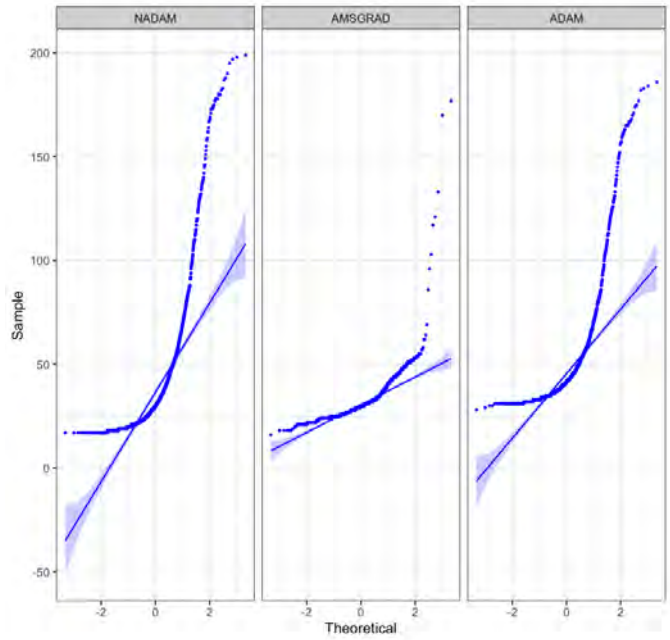
The "blindness" toward the globality/locality character of the reached optimum for the gradient descent-based algorithms is an inherent property because the gradient is a local concept, and by itself carry only local information about the objective function which makes these algorithms very sensitive to the choice of learning rate and initialization. The adaptive moment included features does not add to the image but some amount of "memory" of the recent gradients.

A rough observation that can be mentioned here is the fact that in our generated sample data, most of the time the convergent instances for Nadam and Adam was to the global minima, but we can't really draw any statistical evidence from this naïve observation because our randomized simulation design does not support this analysis.

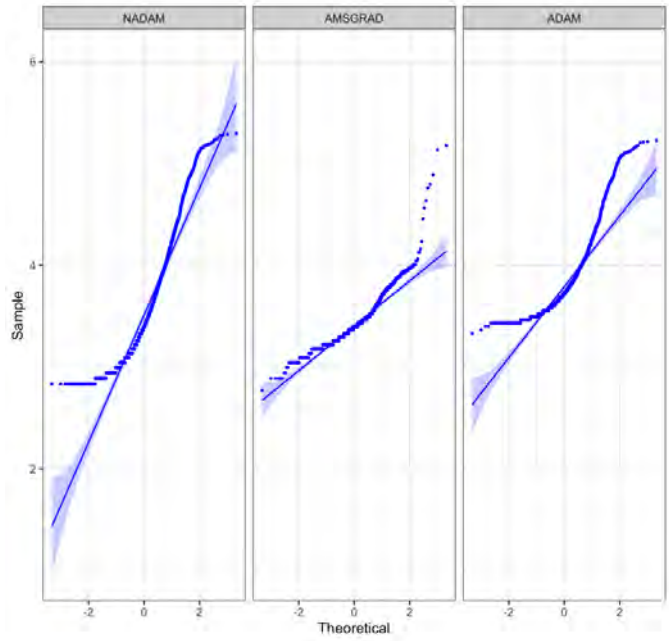
First of all, we check the distributions of number of iterations (speed) for normality, In the hope to be able to harness the large and powerful available para-

metric statistical approaches, from a literature heavily relying on this (workhorse) normal distribution.

Probability distributions of speed of convergence and the log of speed of convergence is shown in the QQ plot described in Figure 5.2(a) and Figure 5.2(b), respectively. From these two graphs, it clearly appears that these distributions are very far from being reasonably considered normally or log-normally distributed. That is not a surprising fact indeed, knowing that these distributions are not symmetric to begin with, and look (strongly) skewed, but we wanted to exclude the possibilities of any approximate (left truncated) normal distributions.



(a)



(b)

Figure 5.2: QQ-plot of speed of convergence and log of speed of convergence , for different optimizers.

Table 5.4: Shapiro-Wilk normality test results.

Optimizers	Shapiro-Wilk [%]	p-value
Nadam	0.71168	$< 2.2e\text{-}16$
Adam	0.68959	$< 2.2e\text{-}16$
AmsGrad	0.63334	$< 2.2e\text{-}16$

Confirming this visual observation, the results of running Shapiro-Wilk normality tests on the three data samples are listed in Table 5.4 using the following formula

$$W = \frac{(\sum_{i=1}^{n} a_i x_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \tag{5.16}$$

where $\bar{x}$ is the sample mean, and x_i are the ordered sample values. The constants a_i are obtained from the covariances, variances, and means of n random variables sampled from the standard normal distribution.

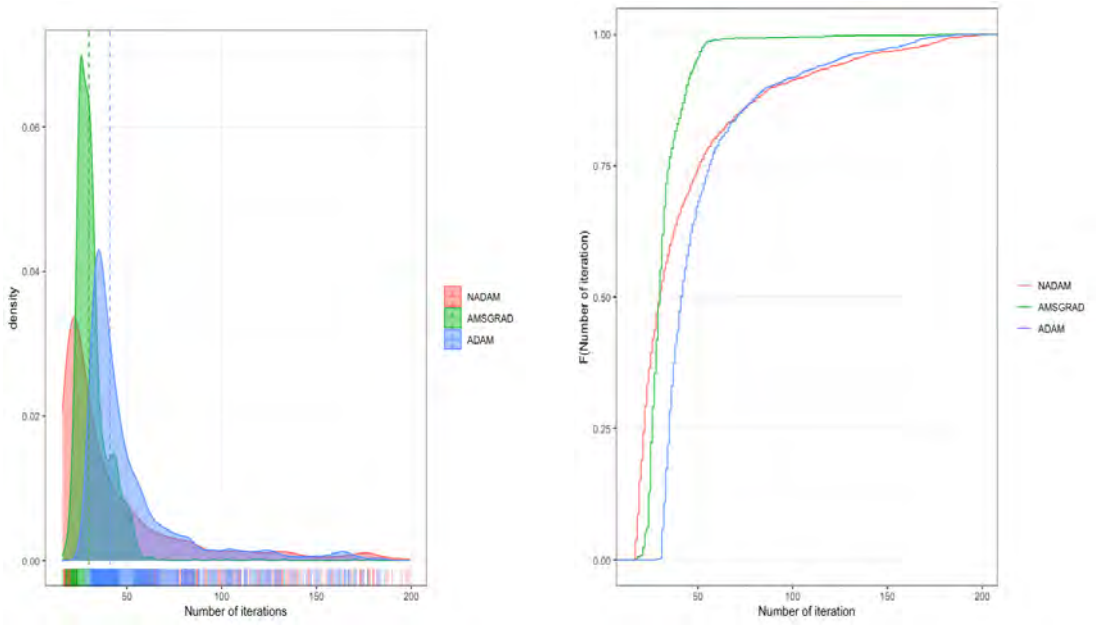
From Table 5.4, we conclude that the number of iteration for different optimizers significantly deviate from being normally distributed and there is very little evidence, if not, that support the normality. We did not test the goodness of fit for other density functions like Gumbel, Fréchet and Weibul, even though the look of the distributions may suggest this family of extreme value distribution (EVD), mainly for two reasons:

First, those EVDs, even if approximately fitted to our empirical distribution will not provide us, following our best judgment, with any advantage, considering the fact that the nature of exact distribution is not our main goal in itself, but rather is the distributions locations, while all the well known available parametric statistical methods for this purpose, are based on the assumption that the samples come from (approximate) normal distribution.

Second, since we stopped the optimization iterations at 200 as mentioned above, we automatically lost informations about the distribution in the extreme left part of the tail (that is almost 10% of the population according to the estimates in Table 5.2, for the three algorithms). This fact would certainly impact (heavily) the estimation of any EVD parameter, and by consequences, would reduce the power of any parametric test based on those inherently biased, and grossly approximate fits, which will minimize the comparative advantage of the eventual parametric over non parametric alternative method.

Following the arguments discussed above, we will use non-parametric statistical approaches to recover further information about the three optimizers performances from data, and since the exact distributions are not well defined, we will use the empirical cumulative distribution as a legitimate approximation.

From the superposition of the three optimizers empirical densities and cumulative densities functions of speed of convergence, as shown in Figure 5.3(a) and Figure 5.3(b) respectively, we note the differences in the central tendencies of the speed of convergence for the three optimizers, and we remark that the minimization of the objective function converges faster in the case of AmsGrad algorithm in comparison to the other two algorithms. To gain more credible evidence about this preliminary raw observations, we conducted a Kruskal-Wallis paired test [144] to elicit any significant difference of means among the three optimizers. Results of the tests are included in box plot shown in Figure 5.4 with p-values. We can conclude with high confidence that there is a significant difference ($p \prec 0.05$) between the speed of convergence for the three optimizers. Comparing the means of number of iterations between each two algorithms individually, and especially between Adam and AmsGrad that look very close (mean wise), we conclude that there is a significant difference between these two groups, too.



(a) (b)

Figure 5.3: a) Densities of number of iteration for each optimizer and b) The emprical cumulative density functions.

To frame those differences in speed between the three algorithms, we generate the 95 percent confidence intervals for the median differences using the bootstrap method with 10 000 replicates each. Normal, Percentile, and pivotal 95 percent confidence intervals have been calculated. Results are summarized in Table 5.5. From this table we spot a clear advantage of Nadam and AmsGrad over Adam

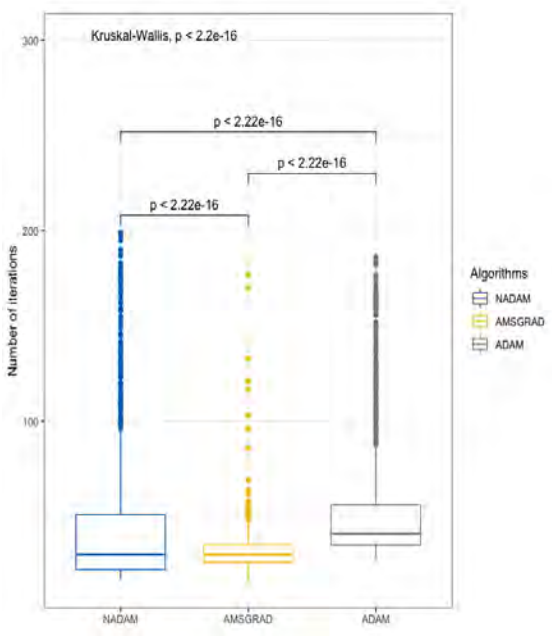


Figure 5.4: The Kruskal-Wallis paired test for number of iterations between different groups of optimizers with the resulting p-value of the test for each group.

in speed of convergence (on average) while the difference between AmsGrad and Nadam is around just 4 steps.

Table 5.5: The 95 percent non parametric confidence intervals based on bootstrap method.

Groups	Point estimation	Standard Error	Normal	Percentile	Pivotal
AmsGrad vs Nadam	4	0.36	(3.29,4.71)	(3,4)	(4,5)
Adam vs Nadam	12	0.5	(11.04,12.96)	(11,12)	(12,13)
AmsGrad vs Adam	15	0.44	(14.13,15.87)	(15,16)	(14,15)

Following the logic of our study, we investigate the relationship between speed of convergence and each of the three factors of the simulation, namely the number

of anomalies in image, the initial guess error and the choice of the learning rate. To verify the impact of number of anomalies on the speed of convergence, Kruskal-Wallis test is applied on each algorithm speed of convergence sample data, as grouped by the number of inclusions. Kruskul-Wallis test results are presented in Figure 5.5, and we can conclude (by failing to reject the Kruskul-Wallis null hypothesis) that the number of anomalies present in the image is not significantly affecting the speed of convergence ($p \succ 0.05$) for different optimizers.

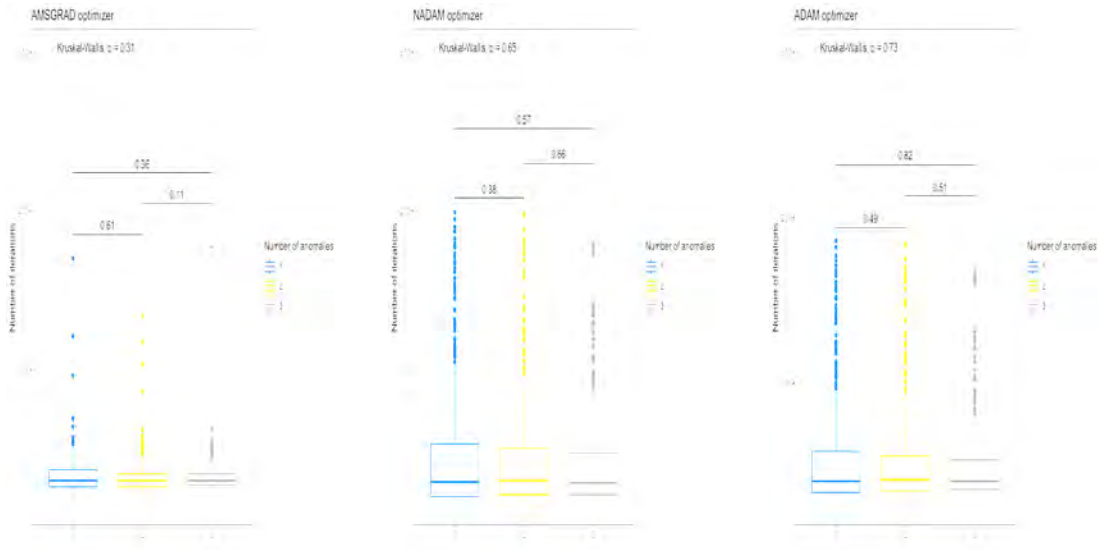


Figure 5.5: Kruskal-Wallis test on number of inclusions for different optimizers with the resulting p-value of the test for each group.

To fulfill our investigation, we discuss the impact of initial guess error and learning rate parameter over number of iterations as shown in Figure 5.6 and Figure 5.6, respectively. The Spearman's coefficient of correlation is used due to its robustness against outliers which appears in data. The Spearman's coefficient of correlation is given by

$$R = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \tag{5.17}$$

Here,

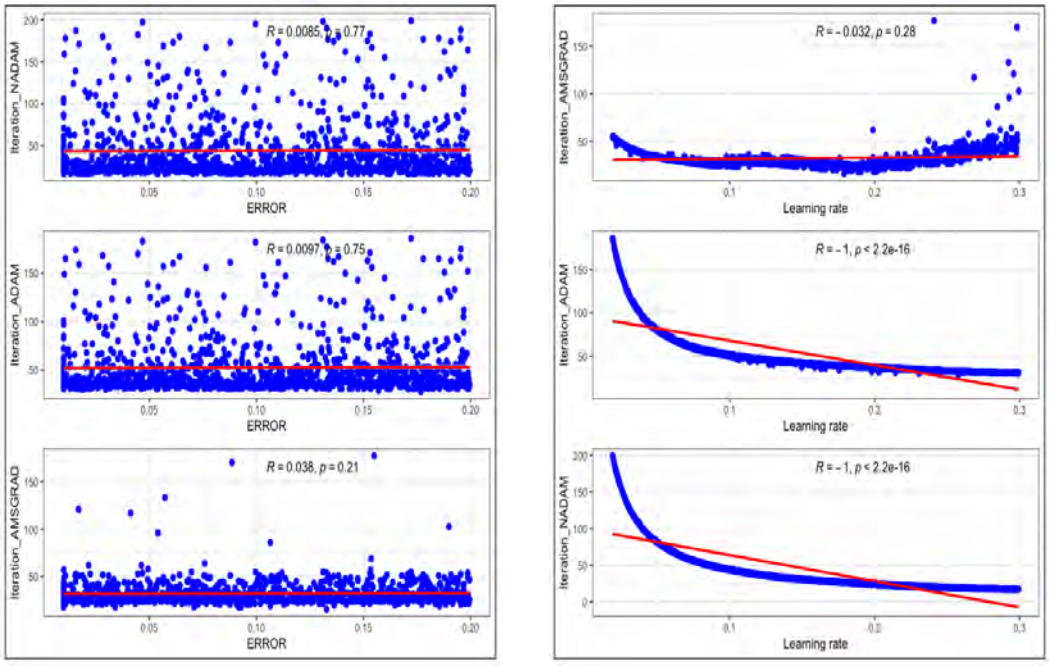
R is the strength of the rank correlation between variables

d_i is the difference between the x rank and the y rank for each pair of data

$\sum d_i^2$ is the sum of the squared differences between x and y variable ranks

n is the sample size

Scatter plots in Figure 5.6(a) and Figure 5.6(b) show the relationship between the initial guess error and the learning rate parameter on the speed of convergence,



(a) (b)

Figure 5.6: Relationship between number of iteration of each optimizer and (a) initial guess error δ (b) learning rate hyper parameter β.

respectively. Spearman's coefficient of correlation R and p-value are mentioned on the top of each graph. From Figure 5.6(b), we notice that when the learning rate ranges in [0.001,0.2], Nadam and Adam algorithms take more iterations than the AmsGrad algorithm. Also,we note that the AmsGrad algorithm present some robustness toward the learning rate in this range and presents some outliers in the range [0.001,0.2]. According to the Spearman's correlation coefficient, we observe a very strong correlation (R=-1) between learning rate parameter and number of iterations for Adam and Nadam optimizers and a negligible correlation for the case of AmsGrad optimizer and presents some outliers when learning rate ranges in [0.2,0.3] . On the other hand, Figure 5.6(a), shows the relationship between the initial guess error and number of iterations taken by each optimizer to reach convergence of cost functional. We remark that error has the same impact on Adam and Nadam algorithms, when comparing their p value and coefficient of correlation. However, we observe that the AmsGrad is more efficient then other two optimizers even if the error is far from the real image.

To assess the quality performance in reconstructed images between these optimizers, we performed statistical tests for differences of means on PSNR and SSIM as measured for reconstructed images, between the optimizers. These two scores

are defined as follows:

$$PSNR = 10\log_{10}(\frac{max^2(\mu_a^{true})}{\frac{1}{N}\sum_{i=1}^{N}(\mu_a^{recon}(i) - \mu_a^{true}(i))^2}),$$

$$SSIM(\mu_a^{true}, \mu_a^{recon}) = [l(\mu_a^{true}, \mu_a^{recon})]^x + [c(\mu_a^{true}, \mu_a^{recon})]^y + [s(\mu_a^{true}, \mu_a^{recon}]^z,$$

where

$$l(\mu_a^{true}, \mu_a^{recon}) = \frac{(2\bar{m}_{\mu_a^{true}}\bar{m}_{\mu_a^{recon}} + C_1)}{(\bar{m}^2_{\mu_a^{true}} + \bar{m}^2_{\mu_a^{recon}} + C_1)}$$

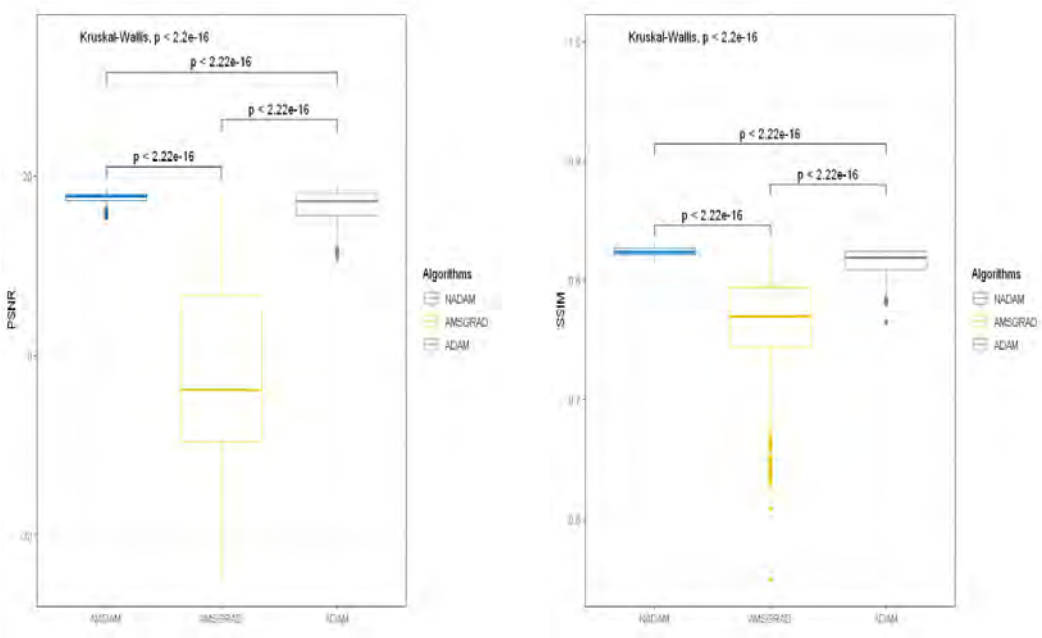
$$c(\mu_a^{true}, \mu_a^{recon}) = \frac{(2\sigma_{\mu_a^{true}}\sigma_{\mu_a^{recon}} + C_1)}{(\sigma^2_{\mu_a^{true}} + \sigma^2_{\mu_a^{recon}} + C_2)}$$

$$s(\mu_a^{true}, \mu_a^{recon}) = \frac{(\sigma_{\mu_a^{true}\mu_a^{recon}} + C_3)}{(\sigma_{\mu_a^{true}}\sigma_{\mu_a^{recon}} + C_3)}$$

where $l(\mu_a^{true}, \mu_a^{recon})$, $c(\mu_a^{true}, \mu_a^{recon})$, and $s(\mu_a^{true}, \mu_a^{recon})$ are the luminance, contrast and structure variations between the true image μ_a^{true} and reconstructed image μ_a^{recon} , respectively, and $x > 0$, $y > 0$, and $z > 0$ are three parameters used to adjust relative importance of the three components of the similarity measure. $\bar{m}_{\mu_a^{true}}$ and $\bar{m}_{\mu_a^{recon}}$ are the means of pixel values of μ_a^{true} and μ_a^{recon}, respectively. We denote by $\sigma_{\mu_a^{true}}$, $\sigma_{\mu_a^{recon}}$, and $\sigma_{\mu_a^{true}\mu_a^{recon}}$ the standard deviation of μ_a^{true} and μ_a^{recon}, and the covariance of image μ_a^{true} and μ_a^{recon}, respectively. C_1, C_2, and C_3 are constants.

The global comparison of quality of reconstructed images are shown in Figure 5.7(a), and Figure 5.7(b), as the result of running the Kruskal-Wallis for PSNR (we eliminated AmsGrad outliers where $PSNR < -25db$) and SSIM, grouping each score sample by optimizer. From Figure 5.7, it appears that the PSNR and the SSIM of AmsGrad are much lower (worse quality) than those of Nadam and Adam. Also, we can observe that the means of PSNR and SSIM for Adam and Nadam are very close.

To evaluate the influence of number of inclusions on image quality, we conduct a Wilcoxon test [145]. The test was applied according different groups of number of inclusions. The resulting p-values of this test are summarized in Table 5.6. The



(a) (b)

Figure 5.7: Kruskal-Wallis test for a) PSNR b) SSIM for different groups of optimizers, with the statistical significance p-value for each group. The simulation cases with $PSNR < -25db$ are excluded.

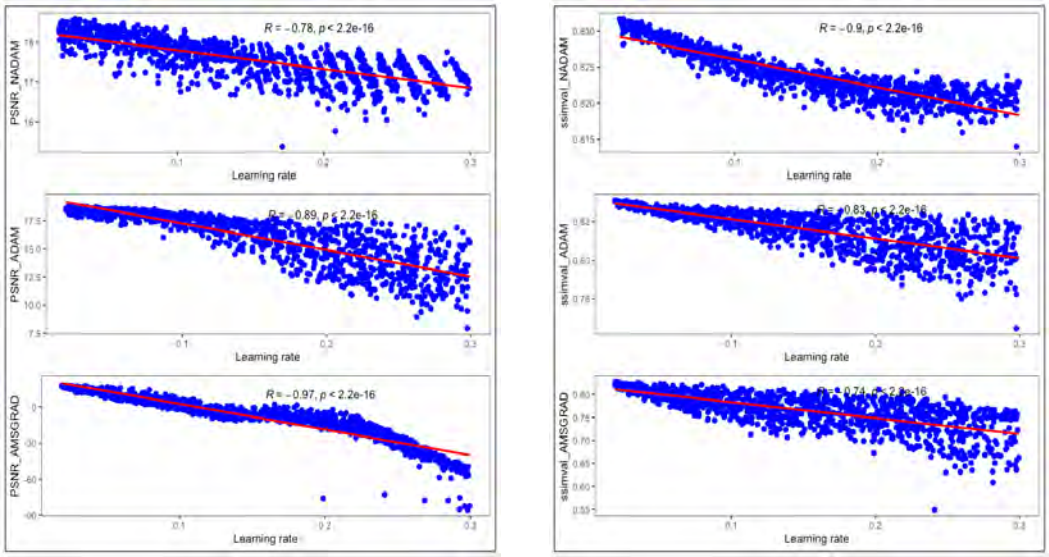
results analysis shows that there is a significant statistical difference between means due to the difference in number of inclusion present in images ($p - value < 0.05$).

Table 5.6: Summary of Wilcoxon test between different number of anomalies for each optimizer.

	Adam			Nadam			AmsGrad		
	1vs2	2vs3	1vs3	1vs2	2vs3	1vs3	1vs2	2vs3	1vs3
PSNR	0.02	2.5e-4	1.8e-6	0.002	0.004	7.8e-11	0.003	0.031	0.001
SSIM	6.4e-5	1.2e-4	1.2e-14	0.008	0.02	0.03	< 2.2e-16	< 2.2e-16	< 2.2e-16

A similar conclusion is deduced about the influence of learning rate on PSNR and SSIM. Scatter plots in Figure 5.8(a) and 5.8(b) show clearly this strong influence of learning rate on PSNR and SSIM, respectively.The resulting Spearman's correlation coefficient by optimizer (and the corresponding p-value) for PSNR and

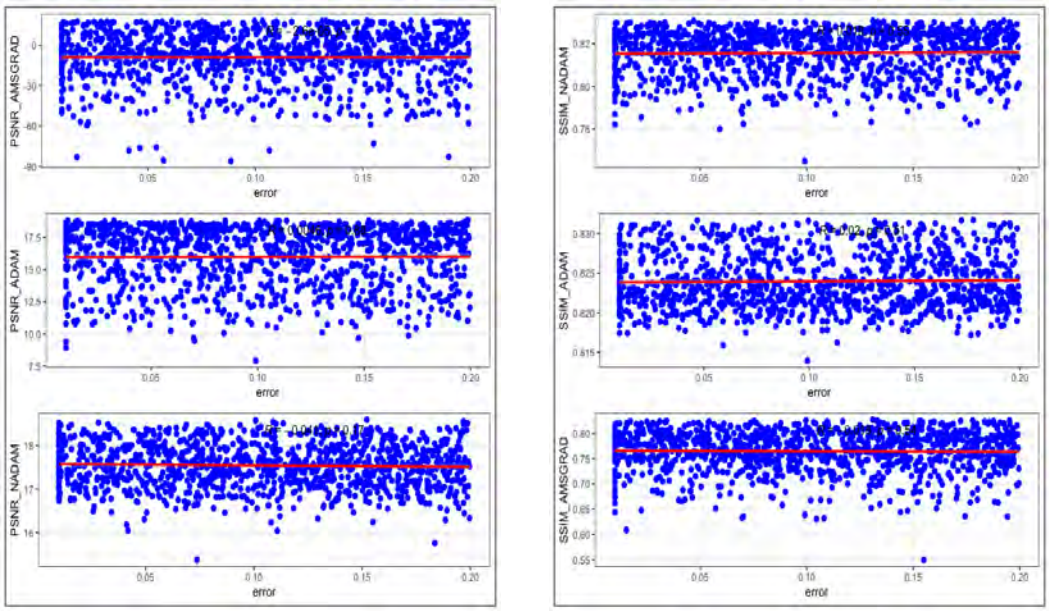
SSIM, are mentioned on the top of each graph. As shown in Figure 5.8(a), we notice that there is a strong negative correlation between learning rate hyper-parameter and PSNR for the case of Adam and AmsGrad. For the case of Nadam , we note a moderate negative correlation between the choice of learning rate and PSNR of reconstructed images. From Figure 5.8(b), we observe that there is a strong negative correlation between learning rate parameter and SSIM for the case of Adam and Nadam. Also, there is a moderate negative correlation between learning rate and SSIM in the case of AmsGrad. The resulting p-value mentioned on the top of each graph, indicate that these correlations are statistically significant, and by consequence, we can conclude the same about the significance of the influence of learning rate choice on the quality of the resulting reconstructed image. Thus, a small value of learning rate that ranges between 0.001 and 0.2 is recommended.



(a) (b)

Figure 5.8: Relationship between learning rate and (a) PSNR (b) SSIM for different optimizers. The resulting Spearman's coefficient R and the p-value are shown on the top of each graph.

Concerning initial guess error, scatter plots in Figure 5.9 demonstrate the influence of initial guess error on reconstructed image quality. From Figure 5.9(a) and Figure 5.9(b), the obtained results show that there is no significant statistical differences between the initial guess error and resulting quality (PSNR/SSIM). Thus, we can conclude with high confidence ($p \succ 0.05$) that the image quality is only influenced by the number of anomalies in the image and the choice of the learning rate.



(a) (b)

Figure 5.9: Relationship between initial guess error and (a) PSNR (b) SSIM for different optimizers. The resulting Spearman's coefficient R and the p-value are shown on the top of each graph.

We illustrate some cases from our simulation. Figure 5.10 shows the reconstructed absorption coefficient μ_a for the case of one inclusion for an initial guess error equal to $\delta = 0.2$. Different values of learning rate are used. The background of true image are taken equal to $\mu_a^{bck} = 0.01mm^{-1}$ and $\mu_s^{'bck} = 2mm^{-1}$. The reconstruction using Nadam and Adam showed a good localization of inclusion. Also, its size is the same compared to the true image with optical properties close to those of true image values. Some artifacts are observed in the borders close to sources and detectors region when the learning rate is higher than 0.1. For the case of AmsGrad reconstruction, we observe that the size of reconstructed image matches those for the true image with some artifacts in the center when learning rate is lower than 0.1. However, when the learning rate is greater than 0.1, we remark that AmsGrad can localize the inclusion, but with some artifacts in the borders. The size and the shape of inclusion does not match with those in the true image. Figure 5.11 shows the reconstructed absorption coefficient μ_a for the case of two inclusions with different shapes for the same values of initial guess error and optical properties used in the first case of one inclusion. From Figure 5.11, we notice that we obtain a good localization of both inclusions for the case of Nadam and Adam for different values of learning rate. However, when the learning rate

is higher than 0.01, we observe some artifacts near the borders. For the case of AmsGrad reconstruction, it is clear that, when the learning rate exceeds 0.1, the size and the shape of inclusion does not matches with those figuring in the true image.

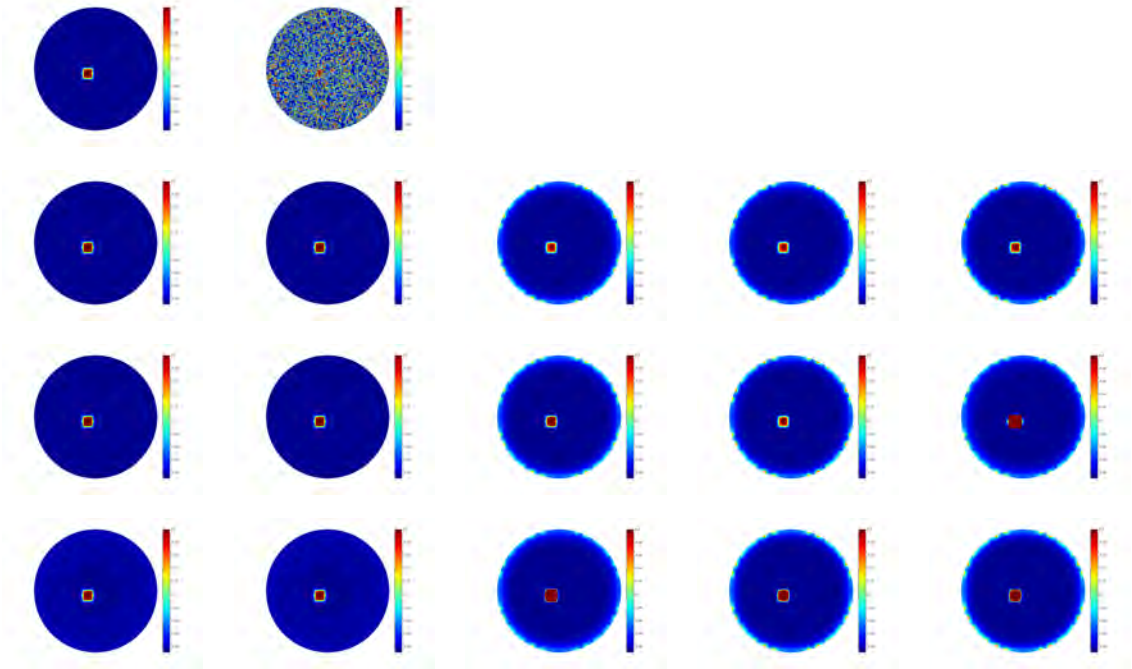


Figure 5.10: Reconstruction of the absorption coefficient μ_a with one inclusion. The first row presents the true image (left) and initial guess image with an initial guess error $\delta = 0.2$ (right). The second, third and fourth rows present the reconstruction images using Nadam, Adam, and AmsGrad, respectively. With learning rate β values equal to 0.001, 0.01, 0.1, 0.2, and 0.3 from (left to right).

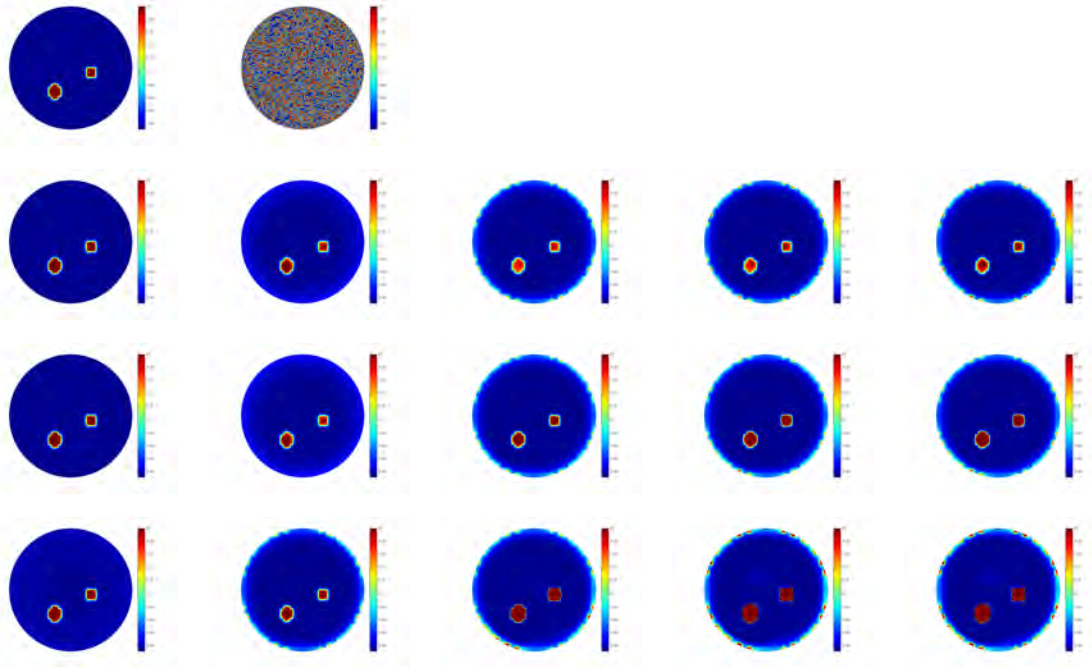


Figure 5.11: Reconstruction of the absorption coefficient μ_a with two inclusions. The first row presents the true image (left) and initial guess image with an initial guess error $\delta = 0.2$ (right). The second, third and fourth rows present the reconstruction images using Nadam, Adam, and AmsGrad, respectively. With learning rate β values equal to 0.001, 0.01, 0.1, 0.2, and 0.3 from (left to right).

5.7 Discussion and conclusion

This research work analyzed the behavior of three optimizers when applied to the inverse problem of DOT regarding the speed of convergence and quality of reconstruction. The three optimizers under study, namely Nadam, Adam and AmsGrad, are enhanced versions of the simple gradient descent algorithm, and had proved to be of very good performance in solving optimization problems in other areas of applications, especially in Deep Learning model search. Our study as we performed is based on a carefully designed randomized numerical simulation that aimed to gain credible statistical evidence on the actual performance of these optimizers when applied to solving the DOT inverse problem. We addressed our attention on the impact of number of anomalies, the learning rate choice, and initial guess error on the speed of convergence. We inquired also the impact of these same parameters on the quality of image reconstruction. The results derived using mainly non-parametric statistical approaches, provides a scientifically credible

quantification of the actual performance of these optimizers, with respect to the choice of learning rate, and under the constraint of the true numbers of anomalies and the arbitrariness of the initial starting point of the optimization.

The study provided valuable guidelines in terms of statistical evidence of the importance of the good choice of the learning rate for the three algorithms, and statistically proved the robustness of Nadam and Adam to the initial guess and the number of anomalies, these results can help improve and promote more the application of DOT in practical medical applications. However, we did not study the impact of these parameters on the simultaneous reconstruction of the absorption and scattering coefficients.

Chapter 6

A New Approach To Improve Optimizer Performance Through Algorithms Diversification For Image Reconstruction In Diffuse Optical Tomography

In this chapter, we present a new approach to improve optimizer performance by using an Algorithms Diversification that switch back and forth between AmsGrad and Nadam in a controlled fashion, where AmsGrad optimizer will be used to accelerate the speed of convergence and Nadam to obtain a good quality of reconstructed image.

6.1 Introduction

The fairly interesting technique for medical imaging suffers nevertheless from the complexity of the related inverse problem as mentioned before, which hinders its broad adoption in medical imaging practice, compared to other readily available procedures like ultrasound and mammography, due to the time consuming numerical optimization solution. The most crucial factor that impacts the acceptance of this new promising procedure of DOT is the quality of the generated image of the biological tissue under-diagnosis and the speed at which we can get feedback from it in real-life configuration [146, 147]. Both of these issues have been studied in previous work [148] that aimed to enhance the solution of DOT problem practically by acting on the optimizer algorithm used to solve the underlying inverse problem. The conclusion reached can be briefly stated that adaptive mo-

ment gradient descent algorithms show statistically significant higher performance in solving numerically the inverse problem of DOT, which represents a significant step toward making DOT procedure more ready for practical application. However, with a caveat, that is the assessed algorithms, namely NADAM [149] and AMSGrad [138], performed individually with respect to only one of the previously cited criteria of interest, namely quality of reconstructed image and speed of convergence while achieving inferior performance concerning the other. In this situation, NADAM is very efficient algorithm in providing superior quality of reconstructed images, but consumes in the way much larger number of iterations than AMSGrad. Conversely, this latest algorithm is very fast in reaching convergence, yet the reconstructed images are far less clear and of low quality. In the present work, we devised a procedure to mix these two optimizers into a new brand of algorithms. This procedure of combining the two algorithms that we will refer to in this work by "diversification" will receive the proper description and explanation in Section 6.3 hereafter. Statistical simulation results will show that, interestingly and significantly enough, the proposed brand of algorithms inherit the best of both worlds by accelerating the rate of convergence and achieving a good quality of the reconstructed images, which is a significant step toward accelerating and enhancing DOT medical imaging techniques to be adopted and used broadly in the practical medical context.

The rest of this chapter is organized as follows: In Section 6.2, we give a brief overview of the mathematical formulation of the diffusion approximation in the continuous wave (CW) case and describe the inverse problem. In section 6.3, we describe the proposed new algorithm that we used to reconstruct the absorption coefficient of DOT. In Section 6.4, we discuss the results obtained by our algorithm in the case of free noise and noisy measurement data.

6.2 Forward and inverse problem

6.2.1 Forward problem

In this section we describe the mathematical formulation of the diffusion approximation (DA).
Let $\Omega \subset R^n$,$n = 2, 3$, be our domain of interest, and $\partial\Omega$ the boundary of Ω. Then the DA inside the domain Ω satisfies the partial differential equation

$$-\nabla[.D(r)\nabla\Psi(r)] + \mu_a(r)\Psi(r) = 0 \qquad r \in \Omega \tag{6.1}$$

with the Robin-boundary condition

$$\Psi(r) + 2aD(r)\frac{\partial\Psi(r)}{\partial\hat{n}} = S(r) \qquad r \in \partial\Omega \tag{6.2}$$

where $\Psi(r)$ is the photon density, $D(r)$ is the diffusion coefficient defined by $D(r) = \frac{1}{3(\mu_a+\mu'_s)}$. a is the Fresnel reflection coefficient, which depends on the mismatch between the refractive indices, μ_a and μ_s the absorption and scattering coefficient respectively, and μ'_s the reduced scattering coefficient expressed as $\mu'_s = (1-g)\mu_s$, where g is the anisotropic factor. $S(r)$ describes the boundary condition for the incoming radiation and $\hat{n}$ is the outward normal vector to Ω.

6.2.2 Inverse problem

The main goal of the inverse problem of DOT, is to determine the optical parameters μ_a and μ_s based on the boundary measured data y_i such that

$$F_i(\mu_a, \mu_s) = y_i \qquad 1 \leq i \leq s \tag{6.3}$$

where F_i is the forward operator which is assumed to be Fréchet differentiable, and y_i the approximate measured data. through out this work, we will focus our attention on the reconstruction of the absorption coefficient μ_a, and we will consider that the scattering parameter μ_s is known. Then, the inverse problem of DOT can be written as follows

$$J(\mu_a) = \frac{1}{2}\sum_{i=1}^{s}(F_i(\mu_a) - y_i)^2 \tag{6.4}$$

Then this problem can be stated in term of an optimization problem with an additive regularization term

$$\mu_a^* = argminJ(\mu_a) + \lambda R(\mu_a) \tag{6.5}$$

where $R(\mu_a)$ is the regularization operator that enforces smoothness conditions in the solution, and λ is the regularization parameter.

6.3 Proposed algorithm

This section will describe our proposed approach to construct the new brand of algorithms to solve the inverse problem of DOT, using the method that we denominate by " Algorithms Diversification ". The main idea behind this technique of algorithm construction procedure is to benefit from the adversarial advantages of Nadam and AMSGrad. In our previous work [17], statistical numerical results have shown that when the learning rate is in the range $[0.001, 0.2]$ NADAM algorithm takes more iterations to converge than AMSGrad. However, always in the same learning rate range, the NADAM algorithm achieves a significantly

better-reconstructed image quality relative to AMSGrad. To exploit the adversarial advantages of NADAM and AMSGrad, we propose a diversified algorithm, called NADA-p (Nadam-AMSGrad Diversified Algorithm with ratio p), that combines NADAM and AMSGrad algorithms in a stochastic fashion according to a probability ratio p. Literally speaking, we combine the two gradient-based algorithms with incompatible advantages (speed and reconstruction quality), hoping to design an algorithm that inherits the benefits of both, since, as statistically established in the previous article, NADAM is significantly more efficient in terms of reconstruction quality. However, the better quality it achieves, the poorer it performs in terms of the rate of convergence. In contrast, AMSGrad evolves in the opposite direction, giving a better convergence speed than NADAM, but a low quality of reconstructed images. For performance assessment, we will compare the proposed algorithm with AMSGrad and NADAM. For more details about NADAM and AMSGrad algorithms, we refer the reader to [139]. The implementation of the NADA-p algorithm is outlined below in algorithm 1. Also, we empirically note that the NADA-p algorithm inherits the same guaranties of convergence from NADAM and AMSGrad under the assumptions that they both converge. The case where NADAM or AMSGrad or both fail to converge is not examined in this work.

In the following outlined implementation of the NADA-p algorithm, p is a decision parameter taken in the range $[0, 1]$, and b is the value of the random variable that follow Bernoulli distribution with parameter p at each iteration. α_{Nadam} and $\alpha_{AMSGrad}$ are the learning rate parameters for Nadam and AMSGrad algorithms, respectively, ρ_1 and ρ_2 are the exponential decay rates for the moment estimates. We denote by ϵ the stabilization parameter, and μ_a^0 is the initial guess estimation.

Algorithm 4: Pseudocode of NADAM-AMSGrad Diversified Algorithm with ratio p (NADA-p)

Require: p, μ_a^0, α_{Nadam},$\alpha_{AMSGrad}$, ρ_1, ρ_2, and ϵ

Ensure: μ_a^n

while J not converged **do**

$b \leftarrow bernoulli(p)$

$k \leftarrow k+1$

$g_k \leftarrow \nabla J_{\mu_a}(\mu_a^{k-1})$

$m_k \leftarrow \rho_1.m_{k-1} + (1-\rho_2).g_k$

$v_k \leftarrow \rho_1.v_{k-1} + (1-\rho_2).g_k^2$

$\hat{m_k} \leftarrow \frac{m_k}{(1-\rho_1^k)}$

$\hat{v_k} \leftarrow \frac{v_k}{(1-\rho_2^k)}$

$\hat{x_k} \leftarrow \max(v_k, \hat{x_{k-1}})$

if $b == 0$ **then**

$\mu_a^k \leftarrow \mu_a^{k-1} - \frac{\alpha_{NADAM}}{\sqrt{\hat{v_k}+\epsilon}}(\rho_1 \hat{m_k} + \frac{1-\rho_1}{1-\rho_1^k} g_k)$

if $b == 1$ **then**

$\mu_a^k \leftarrow \mu_a^{k-1} - \frac{\alpha_{AMSGrad}}{\sqrt{\hat{x_k}}+\epsilon} m_k$

end while

6.4 Performance evaluation

To evaluate the reconstructed image quality of the proposed algorithm for the problem of DOT, peak noise ratio (PSNR) and Structural similarity (SSIM) [150,151], were calculated as expressed below

$$PSNR = 10\log_{10}\left(\frac{max^2(\mu_a^{true})}{\frac{1}{N}\sum_{i=1}^{N}(\mu_a^{recon}(i) - \mu_a^{true}(i))^2}\right),$$

$$SSIM(\mu_a^{true}, \mu_a^{recon}) = [l(\mu_a^{true}, \mu_a^{recon})]^x + [c(\mu_a^{true}, \mu_a^{recon})]^y + [s(\mu_a^{true}, \mu_a^{recon}]^z,$$

where

$$l(\mu_a^{true}, \mu_a^{recon}) = \frac{(2\bar{m}_{\mu_a^{true}}\bar{m}_{\mu_a^{recon}} + C_1)}{(\bar{m}^2_{\mu_a^{true}} + \bar{m}^2_{\mu_a^{recon}} + C_1)}$$

$$c(\mu_a^{true}, \mu_a^{recon}) = \frac{(2\sigma_{\mu_a^{true}}\sigma_{\mu_a^{recon}} + C_1)}{(\sigma^2_{\mu_a^{true}} + \sigma^2_{\mu_a^{recon}} + C_2)}$$

$$s(\mu_a^{true}, \mu_a^{recon}) = \frac{(\sigma_{\mu_a^{true}\mu_a^{recon}} + C_3)}{(\sigma_{\mu_a^{true}}\sigma_{\mu_a^{recon}} + C_3)}$$

here we denote by $l(\mu_a^{true}, \mu_a^{recon})$, $c(\mu_a^{true}, \mu_a^{recon})$, and $s(\mu_a^{true}, \mu_a^{recon})$ the luminance, contrast and structure variations between the true image μ_a^{true} and reconstructed image μ_a^{recon} , respectively, and $x > 0$, $y > 0$, and $z > 0$ are three parameters used to adjust relative importance of the three components of the similarity measure. $\bar{m}_{\mu_a^{true}}$ and $\bar{m}_{\mu_a^{recon}}$ are the means of pixel values of μ_a^{true} and μ_a^{recon}, respectively. We denote by $\sigma_{\mu_a^{true}}$, $\sigma_{\mu_a^{recon}}$, and $\sigma_{\mu_a^{true}\mu_a^{recon}}$ the standard deviation of μ_a^{true} and μ_a^{recon}, and the covariance of image μ_a^{true} and μ_a^{recon}, respectively. C_1, C_2, and C_3 are constants.

6.5 Simulation and data processing

we executed a two stages plan for our simulation experiments. First, we run simulations for free noise data to compare the performance of the NADA-p with NADAM and AMSGrad, where we sampled the probability ration p uniformly randomly in the range $[0, 1]$. We run a similar simulation with the measured data contaminated with 5% white Gaussian noise in the second stage.

To generate a set of data, we use the Toast++ software [143], which solves the forward problem described above. We assume that the medium is highly scattering such that, $\mu_a \ll \mu_s$. 16 sources and 16 detectors are located on the boundary of the domain with equal distance. The location, size and number of anomalies in μ_a are chosen randomly with a background $\mu_a^{bkg} = 0.05mm^{-1}$ and $\mu_s^{'bkg} = 4mm^{-1}$. We solve the forward problem by using the finite element method as described in [152]. We use a circular mesh of radius 20 mm, containing different inclusion sizes and shapes, with 30 504 nodes and 82 332 tetrahedral elements for the forward problem. To avoid inverse crime [144], we use different meshes for the inverse problem with 18 635 nodes and 41 723 tetrahedral elements. The number of anomalies is taken randomly in the range 1 to 3. The regularization parameter λ is set to be equal to 10^{-10}. We solve the inverse problem of DOT by using the NADA-p algorithm presented in the pseudo-code below. For comparison purposes, we use Nadam and AMSGrad algorithms. For more information and details, we refer the reader to [139].

The choice of the learning rate parameters of both NADAM and AMSGrad is based on the results from our work [149]. For Nadam, we choose a learning rate that gives the best possible reconstructed image quality without paying attention to the resulting low convergence speed. On the other hand, for AMSGrad we take a learning rate value that gives the best convergence rate despite the resulting low quality of reconstructed images. Based on the results in [149], α_{Nadam} and $\alpha_{AMSGrad}$ are set to be equal to 0.01 and 0.2, respectively. We fixed all the other hyperparameters for all optimizers to the recommended values from the corresponding literature with $\rho_1 = 0.9$ and $\rho_2 = 0.999$. The algorithm stops until the stopping criterion is satisfied. The complete data analysis was carried out in R [153].

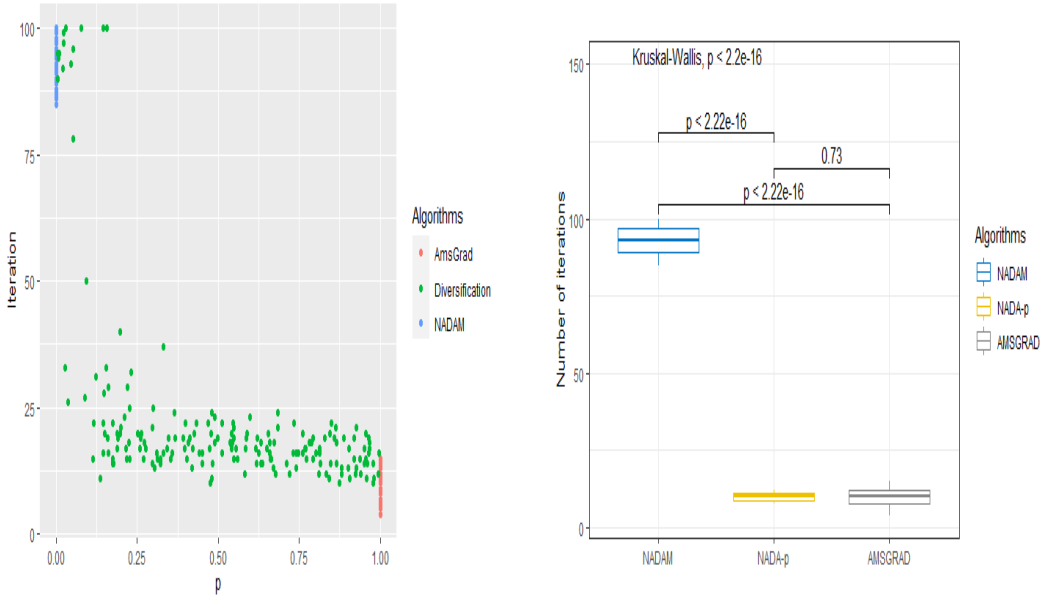
6.6 Results and Discussion

Our study is divided into two parts [154]. For the first section, we discuss the performance of the proposed algorithm in the case of free noise measurement data. In the second section, we will investigate our algorithm's performance and robustness when white Gaussian noise is added to measurement data. For both cases, we study the behavior of speed of convergence and the quality of the reconstructed images with respect to the output of the three algorithms, namely NADAM, AMSGrad, and NADA-p. To statistically compare different groups of resulted simulation samples, Kruskal-Wallis paired test [145] will be carried out. The resulting p-value of the test is at the top of each graph.

6.6.1 Simulation with free of noise Data measurements

We discuss in this first part the results from the simulation with free noise measurement data. We present the simulation results of 200 instances. For this simulation, the maximal number of iterations has been set to be equal to 120 iterations.

Figure 6.1(a) presents the impact of the ratio p on the number of iterations. We take $p = 0$ for the case of Nadam (which is equivalent to NADA-0) and $p = 1$ for the case of AMSGrad (which is equivalent to NADA-1) algorithm. From Figure 6.1(a), we note that our proposed algorithm converges faster than the NADAM algorithm when p is greater than 0.25. However, it takes more iteration number when p ranges in $[0, 0.25]$. To gain more credible evidence about this remark, we compare the mean of the number of iteration for each algorithm as shown in Figure 6.1(b), restricting our attention to the NADA-p where p is in the range $[0.2, 0.99]$. Box plot in Figure 6.1(b) show that there is a significant difference $(p - value < 0.05)$ for the speed of convergence between NADAM on the one hand and NADA-p and AMSGrad on the other hand. At the same time, there



(a) (b)

Figure 6.1: (a) Relationship between number of iterations and the choice of the ratio p for Nada-p, data from Nadam (p=0) and AMSGrad (p=1) is add for visual comparison. (b) Kruskal–Wallis test results for the number of iterations between different optimizers.

is no significant difference between NADA-p and AMSGrad in terms of speed of convergence.

Concerning the quality of the reconstructed images, the PSNR, and the SSIM values have been calculated as described in Section 6.4, for each simulated instance. The SSIM values range between 0 and 1. A result with a high value of SSIM is considered a good reconstructed image. Figure 6.2 (a) shows the relationship between the PSNR and the probability ratio p parameter choice for the NADA-p algorithm. From Figure 6.2 (a), we note that the PSNR of our proposed algorithm is the same as NADAM algorithm when p is in the range $[0.25, 1[$. The same observation can be concluded about the SSIM from Figure 6.2(b). Figure 6.2 (c) and Figure 6.2 (d) show the Kruskal-Wallis test results for PSNR and SSIM, respectively, between different pairs of algorithms. From Figure 6.2(c) and Figure 6.2(d), we can conclude that there is significant statistical differences ($p - value < 0.05$) between AMSGrad and the other two algorithms, and no significant statistical difference between NADAM and NADA-p ($p - value > 0.05$) in terms of PSNR and SSIM.

To further illustrate these observations, we present the relationship between PSNR/SSIM and the number of iterations to localize the NADA-p algorithm's

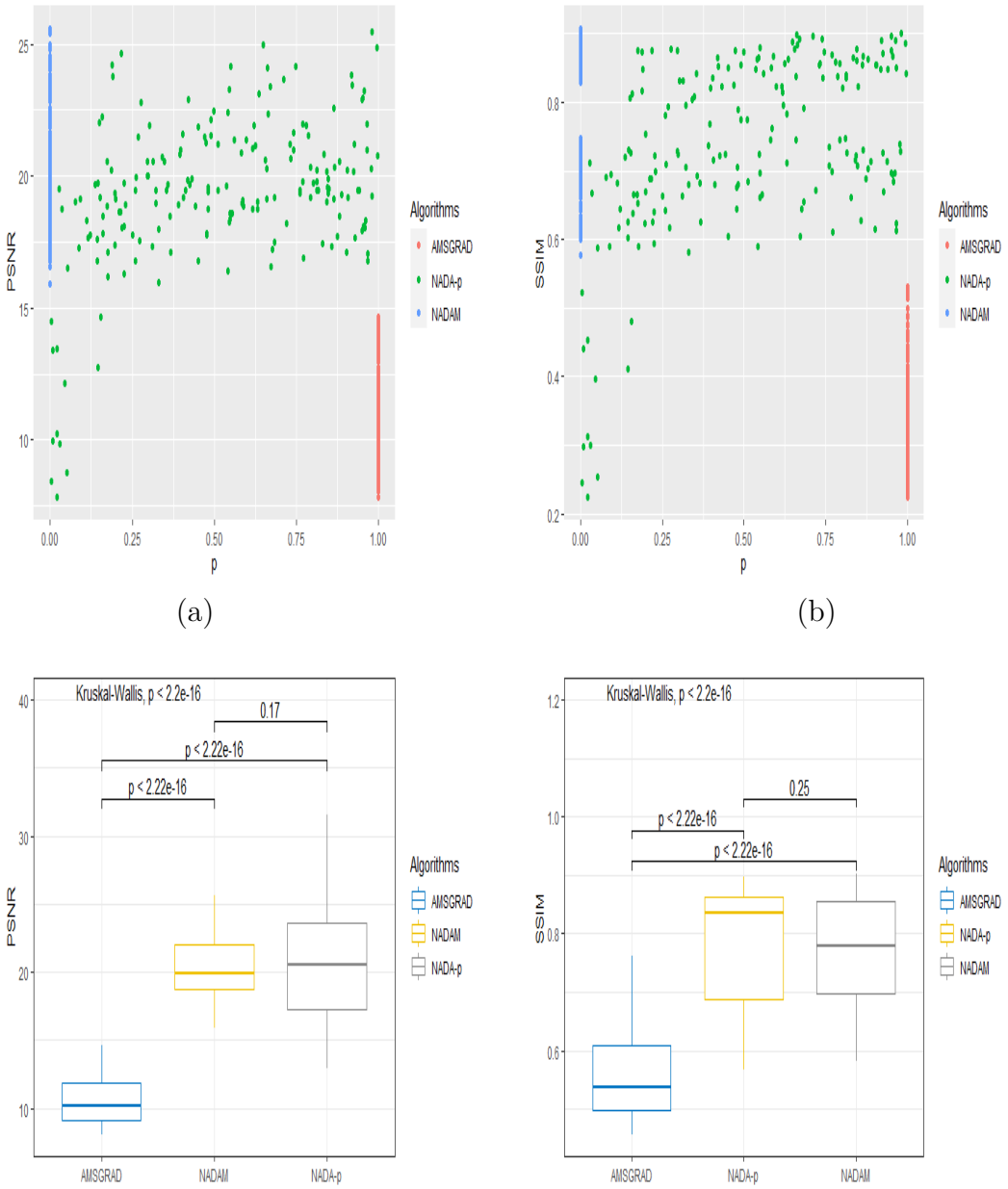
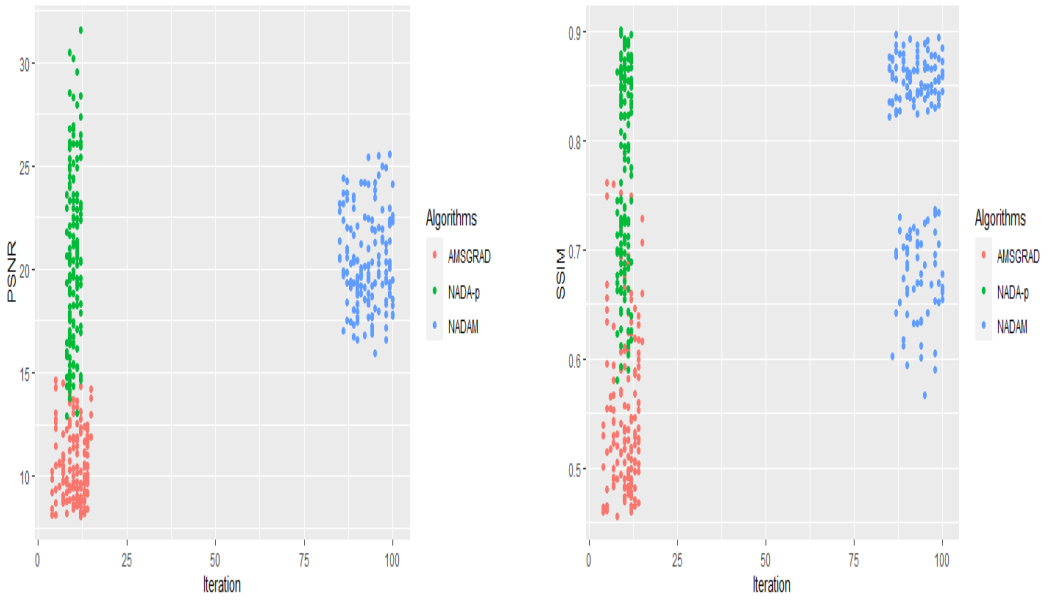


Figure 6.2: (a) Impact of ratio p on PSNR. (b) Impact of ratio p on SSIM. (c) Kruskal–Wallis test results for the PSNR difference between optimizers. (d) Kruskal–Wallis test for the SSIM difference between different optimizers.

performance relative to both others. From Figure 6.3(a) and Figure 6.3(b), we



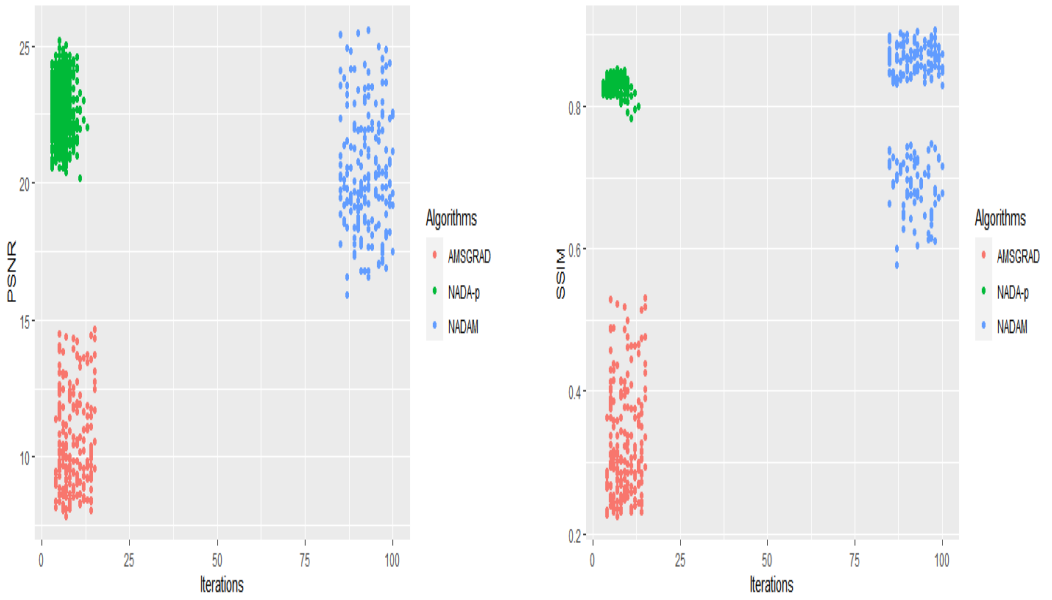
(a) (b)

Figure 6.3: (a) Relationship between PSNR and the number of iterations. (b) Relationship between SSIM and the number of iterations.

outline that NADA-p algorithm can achieve high values of PSNR and SSIM with a few numbers of iterations contrarily to NADAM that performs the same in terms of quality reconstruction but with a much higher number of iterations.

The most remarkable feature of the NADA-p algorithm that we note from Figure 6.3 is that, even though NADA-p is in a way an "average" algorithm from NADAM and AMSGrad, it does not yield an "average" response with respect to the rate of convergence and quality of reconstructed images. Still, it inherits the good features from both NADAM and AMSGrad, which is a very interesting behavior that allows us to construct a better optimizer from combination of two lesser good ones. This fact is significantly exhibited when we simulate with a specific value of ratio p around the middle (like $p = 0.6$) as shown in Figure 6.4.

Figure 6.5 shows the reconstruction results of the absorption coefficient with three inclusions by different reconstruction algorithms. The first row presents the target image and the initial guess image from left to right. The second row shows the reconstruction images using NADAM after 100 iterations (left) and AMSGrad after 15 iterations (right). The third row presents the reconstruction images using NADA-0.3 , NADA-0.5 , and NADA-0.6 after 40, 25 and 10 iterations. Results demonstrated in Figure 6.5 show that our algorithm can localize the shape and the size of inclusions in a fewer number of iteration compared to NADAM, with



(a) (b)

Figure 6.4: (a) Relationship between the PSNR and the number of iterations for NADAM, AMSGrad and NADA-0.6. (b) Relationship between the SSIM and the number of iterations for NADAM, AMSGrad and NADA-0.6.

high contrast compared to AMSGrad.

6.6.2 Simulation with white noise contaminated Data measurements

Since the measurement data is often contaminated with noise, and to further assess the performance and effectiveness of the proposed algorithm against noisy data, we add 5% Gaussian noise to the boundary measured data. We increase the maximum number limit of iterations to 200 due to the slow convergence rate of the algorithms in the presence of noise. We run the experiment for 200 simulation instances.

As shown in Figure 6.6(a), regarding the speed that the NADA-p algorithm takes to converge on average, it is much faster than the NADAM algorithm when the ratio p ranges in $[0.125, 1[$. Moreover, the speed of convergence of the proposed algorithm can be considered similar to the speed of convergence of AMSGrad algorithm when p is greater than 0.125. To statistically confirm this observation, we run the Kruskal-Wallis test restricting our attention only to the instances with values of p in the range $[0.125, 0.99]$. Kruskal-Wallis test results are shown in Figure 6.6(b). From Figure 6.6(b), we note that there is a significant statistical difference

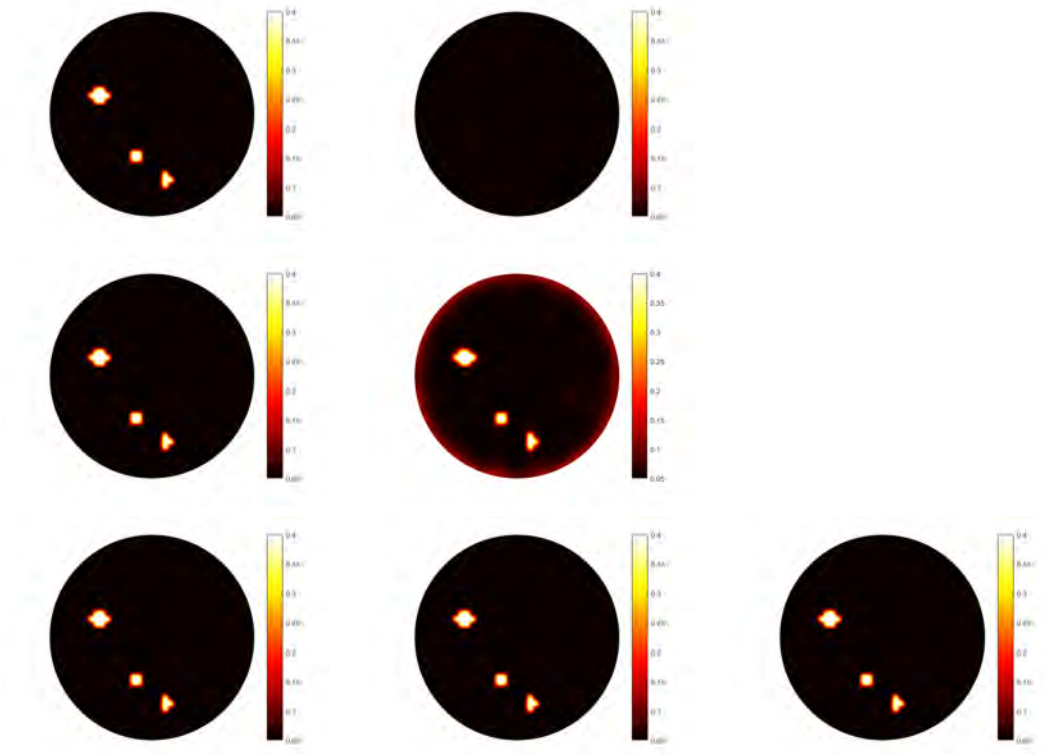
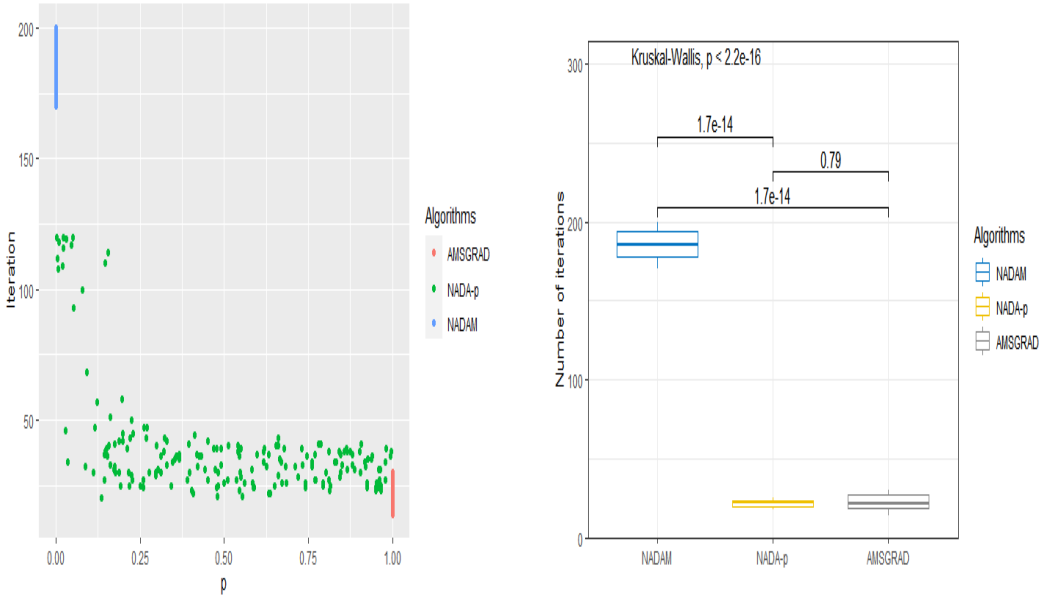


Figure 6.5: Reconstruction results of the absorption coefficient μ_a with three inclusions with different algorithms in the case of free noise measurement data. The first row presents the target image (left) and initial guess image (right). The second row, presents the reconstruction images using NADA-0 (left), and NADA-1 (right). The third row, presents the reconstruction images using NADA-0.3, NADA-0.5, and NADA-0.6 from left to right,respectively.

between NADAM and the other two algorithms ($p-value < 0.05$) in terms of rate of convergence. However, we note that there is no significant difference between AMSGrad and NADA-p algorithms ($p - value > 0.05$).

Finishing our evaluating of the impact of noise on the performance of our algorithm, this time in terms of the quality of reconstructed images. We notice first from Figure 6.7(a) and Figure 6.7(b) that the NADA-p algorithm behaves as NADAM algorithm when p ranges in $[0.125, 1[$ while outperforming AMSGrad in terms of PSNR and SSIM in that same range. We run the Kruskal-Wallis test to confirm this observation as shown in Figure 6.7(c) and Figure 6.7(d). From Figure 6.7(a) and Figure 6.7(b), we conclude that there is no significant statistical differences ($p - value > 0.05$) between NADAM and NADA-p in term of PSNR and SSIM. However, we notice a significant statistical difference between AMSGrad



(a) (b)

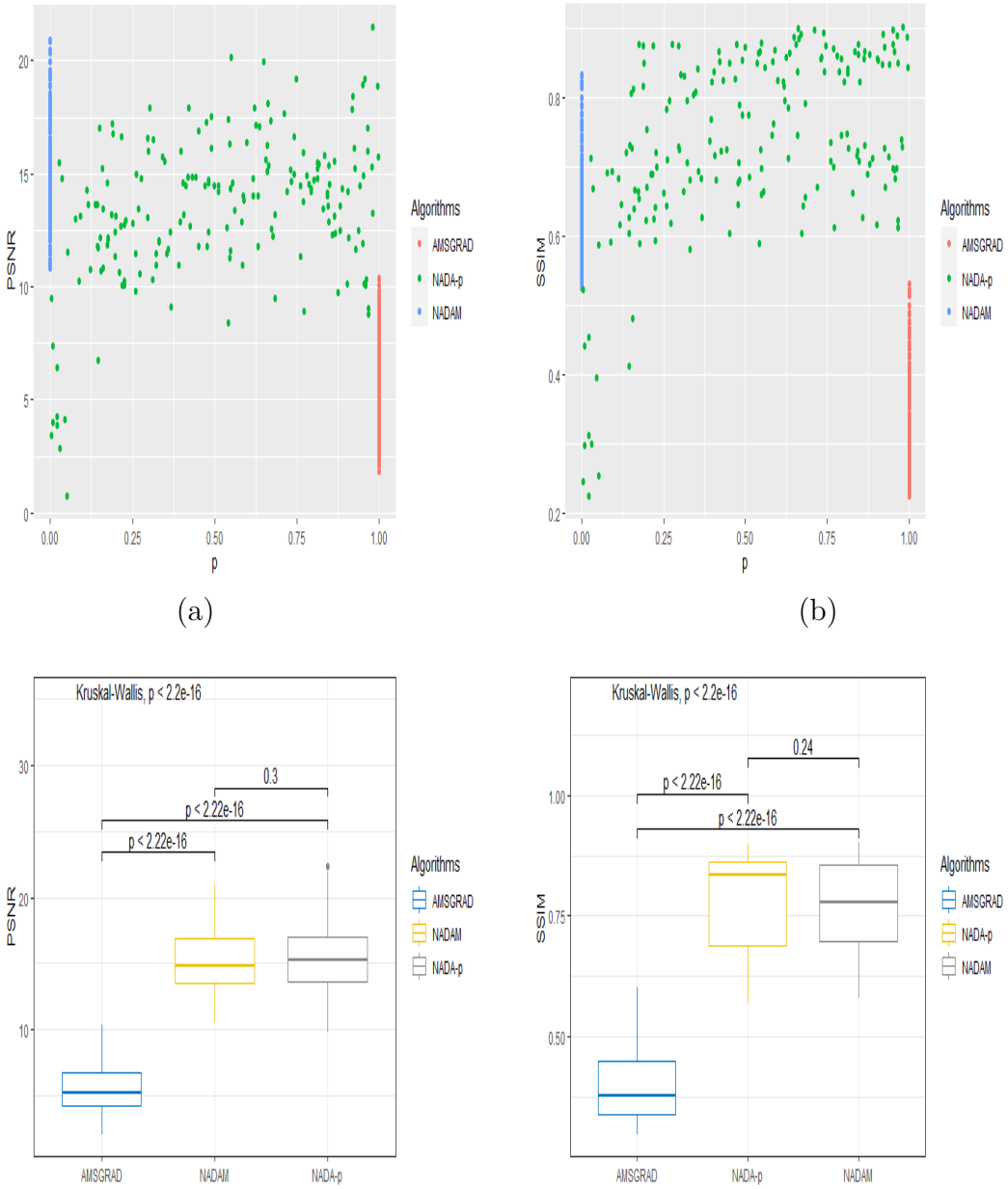
Figure 6.6: (a) Relationship between number of iterations and the choice of the ratio p for NADA-p. (b) Kruskal–Wallis test results for the number of iterations between different optimizers.

and both other two algorithms regarding the quality of reconstructed images.

Finally, to conclude this performance comparison as we did in the previous section about free noise data, Figure 6.8 illustrate clearly the relative performance of the three algorithms with respect to the rate of convergence and quality of reconstruction. From Figure 6.8, we notice that NADA-p and NADAM algorithm outperform AMSGrad algorithm in terms of quality of reconstructed images. Furthermore, the NADA-p algorithm can achieve the same quality of reconstructed images as the NADAM algorithm in fewer iterations. Thus, we can conclude that the NADA-p algorithm is more efficient and robust even in noisy measurement data and can inherit the advantages of NADAM and AMSGrad.

Here too, we outline the same remarkable feature of the NADA-p algorithm that we noted before in the case of the free noise data. From Figure 6.8 is that, even though NADA-p is in a way an "average" algorithm from NADAM and AMSGrad, it does not yield an "average" response with respect to the rate of convergence and quality of reconstructed images. Still, it inherits the good features from both NADAM and AMSGrad, which is a very interesting behavior that allows us to construct a better optimizer from combination of two lesser good ones.

Figure 6.9 shows the reconstruction results of the absorption coefficient μ_a

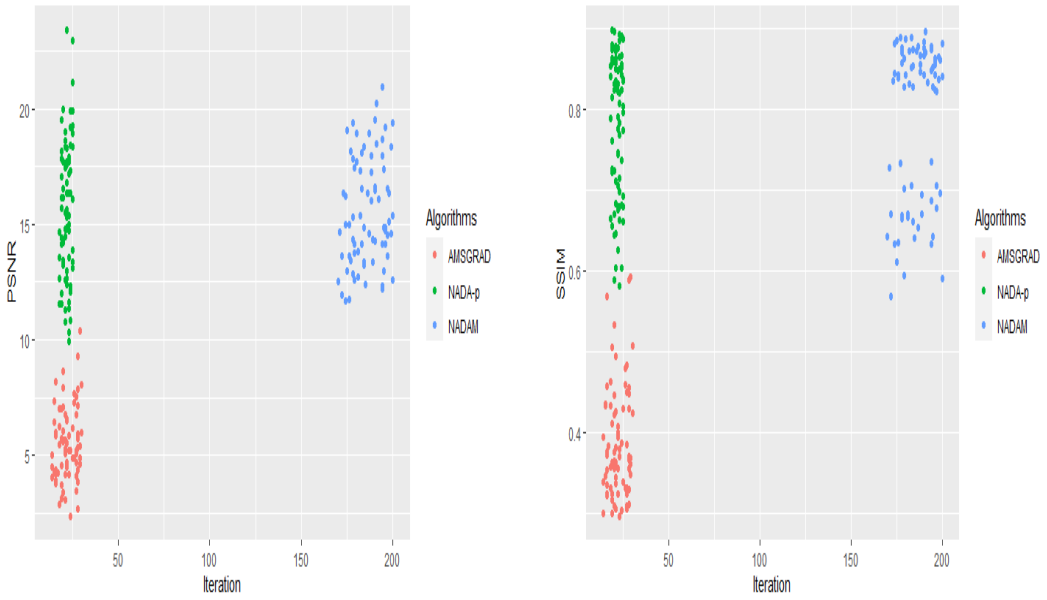


(a) (b)

(c) (d)

Figure 6.7: (a) Impact of ratio p on PSNR. (b) Impact of ratio p on SSIM. (c) Kruskal–Wallis test results for the PSNR difference between optimizers. (d) Kruskal–Wallis test for the SSIM difference between different optimizers.

with two inclusions by different algorithms in the case when 5% of white Gaussian



(a) (b)

Figure 6.8: (a) Relationship between the PSNR and the number of iterations for each algorithm. (b) Relationship between the SSIM and the number of iterations for each algorithm.

noise is added to boundary measurement data. The first row presents the actual image (left) and the initial guess image (right). The second row shows the reconstruction images using NADAM with 200 iterations (left) and AMSGrad with 30 iterations (right). The third row presents the reconstruction images using NADA-0.1, NADA-0.4, and NADA-0.6 from left to right, with 100, 55, and 40 iterations. Reconstruction results show that AMSGrad fails to localize the inclusions in the case of noisy data. For the NADAM algorithm, we notice that the shape of inclusions matches those of the true image but with some artifacts. For the case of the reconstructed images obtained by NADA-p algorithm, we notice that, the shape and the size of inclusions match with those figuring in the actual image. Moreover,the image reconstructed by NADA-p algorithm has better contrast than those obtained by NADAM and AMSGrad.

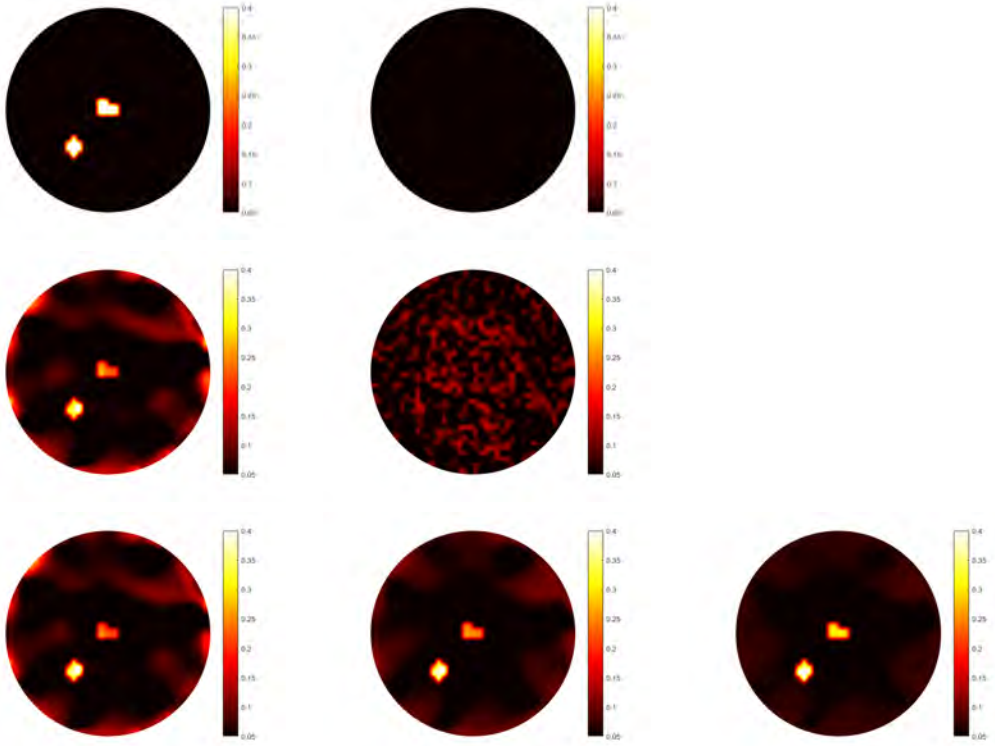


Figure 6.9: Reconstruction results of the absorption coefficient μ_a with two inclusions by different algorithms in the case of noisy measurement data. The first row presents the target image (left) and initial guess image (right). The second row, presents the reconstruction images using Nadam (left), and AMSGrad (right). The third row, presents the reconstruction images using NADA-0.1, NADA-0.4, and NADA-0.6 from left to right,respectively.

6.7 Conclusion

In the presented work, a new optimizer algorithm to solve the inverse problem of DOT has been presented and discussed. A large amount of random simulation has been carried out to evaluate the convergence behavior and quality of reconstructed images for both cases of free of noise and contaminated measurement data. Results and their analysis show that the proposed algorithm can achieve convergence significantly faster compared to NADAM algorithm. Also according to the quality of reconstructed images, results show that the proposed algorithm can achieve much better quality of reconstruction compared to AMSGrad algorithm.

By diversifying the two algorithms, NADAM and AMSGrad, in the way we used to design our NADA-p algorithm, the resulting optimizer exhibits a very interesting

feature, by inheriting the speed of convergence from AMSGrad, alongside with the better quality of image reconstruction from NADAM. And even though being an algorithm that follows an "average" behavior from NADAM and AMSGrad, the resulting performance of NADA-p, as a hybrid mixed algorithm, is significantly better than the average performance of the mix components.

Chapter 7

Conclusion and outlook

In this chapter we provide some conclusions and outline some possible directions for future work.

In this thesis, some novel non-standard image reconstruction techniques for Diffuse Optical Tomography were introduced and discussed. The main novelty in this thesis lies on the application of some trendeous optimizers to handle diffusion equation for the application to DOT.

The purpose of our thesis is a proof-of-concept (rather than dealing with real data), hence we intended to demonstrate that these sophisticated approaches can be coupled. We limited ourselves to a two-dimensional physical setup in order to deal with this combination and evaluate a large number of potential methods. However, we think that all of the methodologies and conclusions provided in this thesis apply easily to the more realistic three-dimensional configuration, with the exception that calculation time may still be long with current computer equipment.

In this thesis, we investigated the behavior of three optimizers when applied to the inverse problem of DOT in terms of convergence speed and reconstruction quality. The three optimizers under consideration, Nadam, Adam, and AmsGrad, are improved variants of the simple gradient descent method that have proven to be very successful at handling optimization issues in various areas of application, including Deep Learning model search. Our work is based on a well structured randomized numerical simulation, with the goal of obtaining convincing statistical data on the real performance of these optimizers when used to solving the DOT inverse issue. We focused on the effect of the number of anomalies, learning rate selection, and initial guess error on the speed of convergence. We also investigated the effect of these same factors on the quality of reconstructed images. The results obtained using mainly non-parametric statistical approaches provide a scientifically credible quantification of the actual performance of these optimizers, with respect to the choice of learning rate, and under the constraint of the true number of anomalies and the arbitrariness of the optimization's initial starting

point. The study provided valuable guidelines in terms of statistical evidence of the importance of selecting a good learning rate for the three algorithms, as well as statistically proving the robustness of Nadam and Adam to the initial guess and the number of anomalies. These findings can help improve and promote the use of DOT in practical medical applications.

Then we moved on to provide such an innovative reconstruction approach based on combining two types of adaptive moments optimizers. We demonstrated that such a technique may dependably get the position and an estimate of the contrast value of the absorption coefficient. Furthermore, we found that in many cases, the batch type converges quicker and provides greater resolution than Nadam and AmsGrad reconstructions. Furthermore, Nada-p reconstructions are frequently more accurate than Adam-type reconstructions in resolving multiple inclusions. On a more technical level, the Nadam-AmsGrad step size criterion performs pretty well as a step size criterion. In this case, our results show that the Nada-p optimizer is more efficient than the Nadam and AmsGrad optimizers in the hyperparameter ranges indicated above.

Finally, we would like to provide some bullet-point ideas for prospective future study. The use of modern optimizers to inverse problems governed by the diffusion approximation is yet relatively unexplored, leaving this study topic wide open for further investigation. The following list is merely meant to provide some general guidelines and is by no means exhaustive.

- Investigate the accuracy of the proposed algorithms in the case of Frequency domain and Time domain.
- Reconstruct optical parameters directly by using a developed Neural Network.
- Apply Deep learning algorithms in real life experiments by choosing a suitable datasets, accurate data acquisition, and optimal loss function.

Bibliography

[1] Baritaux, J.-C. (2012). Sparsity-Inducing Reconstruction Methods for Fluorescence Diffuse Optical Tomography. PhD thesis, Ecole polytechnique fédérale de Lausanne.

[2] Bone, B., Pentek, Z., Perbeck, L., and Veress, B. (1997). Diagnostic accuracy of mammography and contrast enhanced MR imaging in 238 histologically verified breast lesions. Acta Radiol., 38:489–496.

[3] Nass, S., I, H., and Lashof, J. (2001). Mammography and beyond: developing technologies for the early detection of breast cancer. National Academy Press: Washington DC.

[4] Ernst, M. and Roukema, J. (2002). Diagnosis of non-palpable breast cancer: a review. The Breast, 11:13–21.

[5] Prakash, J. (2014). Development of next generation image reconstruction algorithms for Diffuse Optical and Photoacoustic Tomography. PhD thesis, Supercomputer Education and Research Centre (SERC), Indian Institute of Science.

[6] Irishina, N. (2009). Microwave medical imaging using level set techniques. PhD thesis, Universidad III Carlos de Madrid.

[7] Kerlikowske, K., Grady, D., Rubin, S., Sandrock, C., and Ernster, V. (1995). Efficacy of screening mammography. A meta-analysis. JAMA-Journal of the American Medical Association, pages 149–154.

[8] Pisano, E., Gatsonis, C., Hendrick, E., Yaffe, M., Baum, J., Acharyya, S., and et. al. (2005). Diagnostic performance of digital versus film mammography for breast-cancer screening. Annals Academy of Medicine Singapore, 353:1773–1783.

[9] Fear, E., Meaney, P., and Stuchly, M. (2003). Microwaves for breast detection? IEEE Potencials, pages 12–18.

[10] Muttarak, M., Pojchamarnwiputh, S., and Chaiwun, B. (2003). Preoperative detection by mammography. Annals Academy of Medicine Singapore, 32:433–437.

[11] Bushberg, J., Seibert, J., Leidholdt, E., and Boone, J. (2002). The essential physics of medical imaging. Lippincott Williams and Wiklkins, second edition edition.

[12] Schweiger, M., Gibson, A., and Arridge, S. (2003). Computational aspects of diffuse optical tomography. IEEE Signal Process Mag., 5(6):33–41.

[13] Hata, T., Takahashi, H., Watanabe, K., Takahashi, M., Taguchi, K., Itoh, T., and Todo, S. (2004). Magnetic resonance imaging for preoperative evaluation of breast cancer: a comparative study with mamography and ultrasonography. J. Am Coll. Surg., 198:190–197.

[14] Hwang, E., Kinkel, K., Esserman, L., Lu, Y., Weidner, N., and Hylton, N. (2003). Magnetic resonance imaging in patients diagnosed with ductal carcinoma-in-situ value in the diagnosis of residual desease, occult invasion, and multicentricity. Ann. Surg. Oncol., 10:381–388.

[15] Ogiela, M. and Tadeusiewicz, R. (2008). Modern Computational Intelligence Methods for the Interpretation of Medical Images, volume 84 of Studies in Computational Intelligence, chapter Sources of medical images and their general characteristics, pages 7–46. Springer.

[16] Keith, L., Oleszczuk, J., and Laguens, M. (2002). Are Mammography and Palpation Sufficient for Breast Cancer Screening? A Dissenting Opinion. Journal of Wmen's Health and Gender-based Medicine, 11:17–25.

[17] T Tarvainen, M Vauhkonen, V Kolehmainen, S R Arridge, and J P Kaipio. Coupled radiative transfer equation and diffusion approximation model for photon migration in turbid medium with low-scattering and non-scattering regions. Physics in Medicine and Biology, 50(20):4913–4930, 2005.

[18] K Ren, G S Abdoulaev, G Bal, and A H Hielscher. Algorithm for solving the equation of radiative transfer in the frequency domain. Optics Letters, 29(6):578–580, 2004.

[19] H Dehghani, M E Eames, P K Yalavarthy, S C Davis, S Srinivasan, C M Carpenter, B W Pogue, and K D Paulsen. Near infrared optical tomography using NIRFAST: Algorithm for numerical model and image reconstruction. Communications in Numerical Methods in Engineering, 25(6):711–732, 2008.

[20] M Schweiger and S Arridge. The Toast++ software suite for forward and inverse modeling in optical tomography. Journal of Biomedical Optics, 19(4), 2014.

[21] V Ntziachristos, A G Yodh, M Schnall, and B Chance. Concurrent MRI and diffuse optical tomography of breast after indocyanine green enhancement. In Proceedings of the National Academy of Sciences of the United States of America, volume 97, pages 2767–2772, Department of Bioengineering, University of Pennsylvania, Philadelphia, PA 19104-6089, United States, 2000.

[22] Samir Kumar Biswas, Rajan Kanhirodan, Ram Mohan Vasu, and Debasish Roy. Practical fully three-dimensional reconstruction algorithms for diffuse optical tomography. Journal of the Optical Society of America A, 29(6):1017–1026, 2012.

[23] Arridge, S., J., K., Kolehmainen, V., and Tarvainen, T. (2011). Handbook of Mathematical Methods in Medical Imaging, chapter Optical Imaging, pages 735–780. Springer (New York).

[24] Tarvainen, T., Vauhkonen, M., Kolehmainen, V., Arridge, S., and Kaipio, J. (2005). Coupled radiative transfer equation and diffusion approximation model for photon migration in turbid medium with low-scattering and non-scattering regions. Physics in medicine and bioloby, 50(20):4913–4930.

[25] Boas, D., Brooks, D., Miller, E., DiMarzio, C., Kilmer, M., Gaudette, R., and Zhang, Q. (2001). Imaging the body with diffuse optical tomography. IEEE Signal Process Mag., 18(6):57–75.

[26] Dorn, O. and Prieto, K. (2011). From data to images: A shape based approach for fluorescence tomography. In Science: Image in Action: Proceedings of the 7th International Workshop on Data Analysis in Astronomy., pages 255–266.

[27] Gao, H. and Zhao, H. (2010b). Multilevel bioluminescence tomography based on radiative transfer equation Part 2: total variation and l1 data fidelity. Optics Express, 18(3):2894–2912.

[28] A Gibson and H Dehghani. Diffuse optical imaging. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 367(1900):3055–3072, 2009.

[29] M. Cutler. Transillumination as an aid in the diagnosis of breast lesions. Surg. Gynecol. Obstet, 48:721–728, 1929.

[30] C Gros, Y Quenneville, and Y Hummel. Breast diaphanology . Journal de radiologie, d'electrologie, et de medecine nucleaire, 53(4):297–306, 1972.

[31] Frans F. Jobsis. Noninvasive, infrared monitoring of cerebral and myocardial oxygen sufficiency and circulatory parameters. Science, 198(4323):1264–1267, 1977.

[32] Y Yamashita, A Maki, and H Koizumi. Measurement system for noninvasive dynamic optical topography, 1999.

[33] D Grosenick, H Wabnitz, H H Rinneberg, K T Moesta, and P M Schlag. Development of a time-domain optical mammograph and first in vivo applications. Applied Optics, 38(13):2927–2943, 1999.

[34] J C Hebden, S R Arridge, and D T Delpy. Optical imaging in medicine: I. Experimental techniques. Physics in Medicine and Biology, 42(5):825–840, 1997.

[35] S B Colak, M B Van Der Mark, G W 'T Hooft, J H Hoogenraad, E S Van Der Linden, and F A Kuijpers. Clinical optical tomography and NIR spectroscopy for breast cancer detection. IEEE Journal on Selected Topics in Quantum Electronics, 5(4):1143–1158, 1999.

[36] A D Klose and A H Hielscher. Optical tomography with the equation of radiative transfer. International Journal of Numerical Methods for Heat and Fluid Flow, 18(3-4):443–464, 2008.

[37] B W Pogue, S C Davis, X Song, B A Brooksby, H Dehghani, and K D Paulsen. Image analysis methods for diffuse optical tomography. Journal of Biomedical Optics, 11(3), 2006.

[38] H Egger, M Freiberger, and M Schlottbom. On forward and inverse models in fluorescence diffuse optical tomography. Inverse Problems and Imaging, 4(3):411–427, 2010.

[39] D A Boas, C E Elwell, M Ferrari, and G Taga. Twenty years of functional near-infrared spectroscopy: Introduction for the special issue. NeuroImage, 85:1–5, 2014.

[40] A P Gibson, J C Hebden, and S R Arridge. Recent advances in diffuse optical imaging. Physics in Medicine and Biology, 50(4):R1–R43, 2005.

[41] R Roy and E M Sevick-Muraca. Truncated Newton's optimization scheme for absorption and fluorescence optical tomography: Part II reconstruction from synthetic measurements. Optics Express, 4(10):372–382, 1999.

[42] Scholkmann, F., Kleiser, S., Metz, A. J., Zimmermann, R., Pavia, J. M., Wolf, U., Wolf, M. (2014). A review on continuous wave functional near-infrared spectroscopy and imaging instrumentation and methodology. Neuroimage, 85, 6-27.

[43] Xiaofeng Zhang. Instrumentation in Diffuse Optical Imaging. Photonics, 1(1):9–32, mar 2014.

[44] C H Schmitz, M Löcker, J M Lasker, A H Hielscher, and R L Barbour. Instrumentation for fast functional optical tomography. Review of Scientific Instruments, 73(2 I):429, 2002.

[45] A. Hielscher, G. Abdoulaev, A. Klose, J. Lasker, M. Stewart, and A. Bluestone. Near-Infrared diffuse optical tomography. Disease Markers Functional Imaging of Early Markers of Disease, 18:313–337, 2002.

[46] K L Perdue, Q Fang, and S G Diamond. Quantitative assessment of diffuse optical tomography sensitivity to the cerebral cortex using a whole-head probe. Physics in Medicine and Biology, 57(10):2857–2872, 2012.

[47] V Ntziachristos and R Weissleder. Experimental three-dimensional fluorescence reconstruction of diffuse media by use of a normalized Born approximation. Optics Letters, 26(12):893–895, 2001.

[48] S R Arridge and W R B Lionheart. Nonuniqueness in diffusion-based optical tomography. Optics Letters, 23(11):882–884, 1998.

[49] Y Pei, H L Graber, and R L Barbour. Normalized-constraint algorithm for minimizing inter-parameter crosstalk in DC optical tomography. Optics Express, 9(2):97–109, 2001.

[50] V Ntziachristos, J Ripoll, L V Wang, and R Weissleder. Looking and listening to light: The evolution of whole-body photonic imaging. Nature Biotechnology, 23(3):313–320, 2005.

[51] A Corlu, R Choe, T Durduran, M A Rosen, M Schweiger, S R Arridge, M D Schnall, and A G Yodh. Three-dimensional in vivo fluorescence diffuse optical tomography of breast cancer in humans. Optics Express, 15(11):6696–6716, 2007.

[52] Y Pei, H L Graber, and R L Barbour. Influence of systematic errors in reference states on image quality and on stability of derived information for dc optical imaging. Applied Optics, 40(31):5755–5769, 2001.

[53] Farina, A., Tagliabue, S., Di Sieno, L., Martinenghi, E., Durduran, T., Arridge, S., ... Dalla Mora, A. (2017). Time-domain functional diffuse optical tomography system based on fiber-free silicon photomultipliers. Applied Sciences, 7(12), 1235.

[54] Andreas H Hielscher. Optical tomographic imaging of small animals. Current opinion in biothecnology, 16(1):79–88, 2005.

[55] Puszka, A., Di Sieno, L., Dalla Mora, A., Pifferi, A., Contini, D., Planat-Chrétien, A., ... Dinten, J. M. (2015). Spatial resolution in depth for time-resolved diffuse optical tomography using short source-detector separations. Biomedical optics express, 6(1), 1-10.

[56] Nissilä, I. T., Hebden, J. C., Jennions, D., Heino, J., Schweiger, M., Kotilahti, K. M., ... Katila, T. E. (2006). Comparison between a time-domain and a frequency-domain system for optical tomography. Journal of biomedical optics, 11(6), 064015.

[57] Vivek Venugopal, Jin Chen, and Xavier Intes. Development of an optical imaging platform for functional imaging of small animals using wide- field excitation. Biomedical Optics Express, 1(1):143–156, 2010.

[58] F Leblond, S C Davis, P A Valdés, and B W Pogue. Pre-clinical whole-body fluorescence imaging: Review of instruments, methods and applications. Journal of Photochemistry and Photobiology B: Biology, 98(1):77–94, 2010.

[59] L. Hundley, T. Coburn, E. Garwin, and L. Stryer. Nanosecond fluorimeter. Review of Scientific Instruments, 38:488–492, 1967.

[60] D.T. Delpy, M. Cope, P. Vanderzee, Simon Arridge, S. Wray, and J. Wyatt. Estimation of optical pathlength through tissue from direct time of flight measurement. Physics in Medicine and Biology, 33:1433–1442, 1988.

[61] B. Chance, Shoko Nioka, J. Kent, K. McCully, M. Fountain, R. Greenfeld, and G. Holtom. Time-resolved spectroscopy of hemoglobin and myoglobin in resting and ischemic muscle. Analytical Biochemistry, 174:698–707, 1988.

[62] M S Patterson, B Chance, and B C Wilson. Time resolved reflectance and transmittance for the non-invasive measurement of tissue optical properties. Appl. Opt., 28(12):2331–2336, 1989.

[63] V Ntziachristos, X H Ma, and B Chance. Time-correlated single photon counting imager for simultaneous magnetic resonance and near-infrared mammography. Review of Scientific Instruments, 69(12):4221–4233, 1998.

[64] David K. Jennions. Time-Resolved Optical Tomography Instrumentation for Fast 3D Functional Imaging. PhD thesis, University College London, 2008.

[65] E M C Hillman, J C Hebden, M Schweiger, H Dehghani, F E W Schmidt, D T Delpy, and S R Arridge. Time resolved optical tomography of the human forearm. Physics in Medicine and Biology, 46(4):1117–1130, 2001.

[66] Vasilis Ntziachristos, A G Yodh, Mitchel Schnall, and Britton Chance. Comparison between intrinsic and extrinsic contrast for malignancy detection using NIR mammography. In Proceedings of SPIE - The International Society for Optical Engineering, volume 3597, pages 565–570, 1999.

[67] R M Danen, Yong Wang, X D Li, C Dean Kurth, W S Thayer, and A G Yodh. Near-infrared imaging of brains. In Conference on Lasers and Electro-Optics Europe - Technical Digest, pages 86–87, 1998.

[68] Bérubé-Lauzière, Y., Crotti, M., Boucher, S., Ettehadi, S., Pichette, J., Rech, I. (2016). Prospects on time-domain diffuse optical tomography based on time-correlated single photon counting for small animal imaging. Journal of Spectroscopy, 2016.

[69] Pifferi, A., Contini, D., Dalla Mora, A., Farina, A., Spinelli, L., Torricelli, A. (2016). New frontiers in time-domain diffuse optics, a review. Journal of biomedical optics, 21(9), 091310.

[70] N Ducros, A Da Silva, J.-M. Dinten, C S Seelamantula, M Unser, and F Peyrin. Time resolved fluorescence diffuse optical tomography using multi-resolution exponential B-splines. In Proceedings - 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2009, pages 157–160, CEA, LETI, MINATEC, 17 rue des Martyrs, F-38054 Grenoble, France, 2009.

[71] Elizabeth M. C. Hillman. Experimental and theoretical investigations of near infrared tomographic imaging methods and clinical applications. PhD thesis, University of London, 2002.

[72] D A Boas. Diffuse photon probes of structural and dynamical properties of turbid media: theory and biomedical applications. Optics express, 8(5):263–270, 1996.

[73] C Darne, Y Lu, and E M Sevick-Muraca. Small animal fluorescence and bioluminescence tomography: A review of approaches, algorithms and technology update. Physics in Medicine and Biology, 59(1):R1–R64, 2014.

[74] R. D. Spencer and G. Weber. Measurements of subnanosecond fluorescence lifetimes with a cross-correlation phase fluorometer. Annals of the New York Academy of Sciences, 158:361–376, 1969.

[75] B. A. Feddersen, D. W. Piston, and E. Gratton. Digital parallel acquisition in frequency-domain fluorimetry. Review of Scientific Instruments, 60:2929–2936, 1989.

[76] M. A. Oleary, David A. Boas, Britton Chance, and Arjun G Yodh. Refraction of diffuse photon density waves. Physical Review Letters, 69:2658–2661, 1992.

[77] S. J. Madsen, E. R. Anderson, Richard C. Haskell, and Bruce J. Tromberg. Portable, high-bandwidth frequency-domain photon migration instrument for tissue spectroscopy. Optics Letters, 19:1934–1936, 1994.

[78] Sergio Fantini, Maria Angela Franceschini, B. Fishkin Joshua, Beniamino Barbieri, and Enrico Gratton. Quantitative determination of the absorption spectra of chromophores in strongly scattering media: A light emitting diode based technique. Applied Optics, 33(22):5204–5214, 1994.

[79] Q Q Zhang, X J Wu, C Wang, S W Zhu, Y L Wang, B Z Gao, and X.-C. Yuan. Scattering coefficients of mice organs categorized pathologically by spectral domain optical coherence tomography. BioMed Research International, 2014, 2014.

[80] Grosenick, D., Rinneberg, H., Cubeddu, R., Taroni, P. (2016). Review of optical breast imaging and spectroscopy. Journal of biomedical optics, 21(9), 091311.

[81] S Fantini, S. A. Walker, M A. Franceschini, M. Kaschke, P.M. Schlag, and K.T. Moesta. Assessnent of the size, position and optical properties of breast tumours in vivo by non-invasive optical methods. Appl. Opt., 37:1982–1989, 2005.

[82] E M C Hillman. Optical brain imaging in vivo: Techniques and applications from animal to man. Journal of Biomedical Optics, 12(5), 2007.

[83] K Ren, B Moa-Anderson, G Bal, X Gu, and A H Hielscher. Frequency domain tomography in small animals with the equation of radiative transfer. In Progress in Biomedical Optics and Imaging - Proceedings of SPIE, volume 5693, pages 111–120, Dept. of Appl. Physics and Appl. Mathematics, Columbia University, New York, NY 10027, United States, 2005.

[84] A Joshi, W Bangerth, and E M Sevick-Muraca. Adaptive finite element based tomography for fluorescence optical imaging in tissue. Optics Express, 12(22):5402–5417, 2004.

[85] Sevick-Muraca, E. M., Reynolds, J. S., Troy, T. L., Lopez, G., Paithankar, D. Y. (1998). Fluorescence lifetime spectroscopic imaging with measurements of photon migration. Annals of the New York Academy of Sciences, 838, 46-57.

[86] Q Zhu, T Durduran, V Ntziachristos, M Holboke, and A G Yodh. Imager that combines near-infrared diffusive light and ultrasound. Optics Letters, 24(15):1050–1052, 1999.

[87] B Li, F Maafi, R Berti, P Pouliot, E Rhéaume, J C Tardif, and F Lesage. Hybrid FMT-MRI applied to in vivo atherosclerosis imaging. Biomedical Optics Express, 5(5):1664–1676, 2014.

[88] S C Davis, S L Gibbs-Strauss, H Dehghani, B W Pogue, and K D Paulsen. MRI-coupled fluorescence tomography of murine glioma metabolic activity. In Biomedical Optics, BIOMED 2008, Thayer School of Engineering, Dartmouth College, Hanover NH 03755, United States, 2008.

[89] A Ale, V Ermolayev, E Herzog, C Cohrs, M H De Angelis, and V Ntziachristos. FMT-XCT: In vivo animal studies with hybrid fluorescence molecular tomography-X-ray computed tomography. Nature Methods, 9(6):615–620, 2012.

[90] J P Culver, R Choe, M J Holboke, L Zubkov, T Durduran, A Slemp, V Ntziachristos, B Chance, and A G Yodh. Three-dimensional diffuse optical tomography in the parallel plane transmission geometry: Evaluation of a hybrid frequency domain/continuous wave clinical system for breast imaging. Medical Physics, 30(2):235–247, 2003.

[91] Vasilis Ntziachristos, A G Yodh, and Britton Chance. Optical Tomography using multi-frequency intensity information. In Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings, volume 2, page 1100, 1999.

[92] S. J. Madsen, P. Wyss, Lars O. Svaasand, Richard C. Haskell, Y. Tadir, and Bruce J. Tromberg. Determination of the optical-properties of the human uterus using frequency-domain photon migration and steady-state techniques. Physics in Medicine and Biology, 39:1191–1202, 1994.

[93] M C W Van Rossum and Th.M. Nieuwenhuizen. Multiple scattering of classical waves: Microscopy, mesoscopy, and diffusion. Reviews of Modern Physics, 71(1):313–371, 1999.

[94] Brian C. Wilson and Steven L. Jacques. Optical reflectance and transmission of tissues: Principles and Applications. IEEEJ Quantum Electron, 26(12):2186–2199, 1990.

[95] C Dean Kurth and W S Thayer. A multiwavelength frequency-domain near infrared cerebral oximeter. Physics in Medicine and Biology, 44(3):727–740, 1999.

[96] S R Arridge. Optical tomography in medical imaging. Inverse Problems, 15(2):R41–R49, 1999.

[97] D A Boas, M A O'Leary, B Chance, and A G Yodh. Scattering of diffuse photon density waves by spherical inhomogeneities within turbid media: Analytic solution and applications. In Proceedings of the National Academy of Sciences of the United States of America, volume 91, pages 4887–4891, 1994.

[98] M A Franceschini, V Toronov, M E Filiaci, E Gratton, and S Fantini. Online optical imaging of the human brain with 160-ms temporal resolution, 2000.

[99] F Martelli, A Sassaroli, Y Yamada, and G Zaccanti. Analytical approximate solutions of the time-domain diffusion equation in layered slabs. Journal of the Optical Society of America A: Optics and Image Science, and Vision, 19(1):71–80, 2002.

[100] J C J Paasschens. Solution of the time-dependent Boltzmann equation, 1997.

[101] Fabrizio Martelli, Angelo Sassaroli, Antonio Pifferi, Alessandro Torricelli, Lorenzo Spinelli, and Giovanni Zaccanti. Heuristic Green's function of the time-dependent radiative transfer equation for a semi-infinite medium. Optics Express, 15(26):18168–18175, 2007.

[102] S.A. Prahl, M. Keijzer, S.L. Jacques, and A.J. Welch. A Monte Carlo model of light propagation in tissue. SPIE Institute series, IS(5):102–111, 1989.

[103] N Cao, M Ortner, and A Nehorai. Solutions for diffuse optical tomography using the Feynman-Kac formula and interacting particle method. In Progress in Biomedical Optics and Imaging - Proceedings of SPIE, volume 6434, 2007.

[104] A H Gandjbakhche, G H Weiss, R F Bonner, and R Nossal. Photon pathlength distributions for transmission through optically turbid slabs. Physical Review E, 48(2):810–818, 1993.

[105] Rinat Ankri, Haim Taitelbaum, and Dror Fixler. On phantom experiments of photon migration model in tissues. The open optics journal, 5:28–32, 2011.

[106] F A Grunbaum. Diffuse tomography: The isotropic case. Inverse Problems, 8(3):409–419, 1992.

[107] Phaneendra Yalavarthy. A generalized least-squares minimization method for near infrared diffuse optical tomography. PhD thesis, Dartmouth College, 2007.

[108] W. F. Ames. Numerical Methods for Partial Differential Equations. Computer science and scientific computing, Academic Press, New York, 3rd edition, 1992.

[109] B W Pogue, M S Patterson, H Jiang, and K D Paulsen. Initial assessment of a simple system for frequency domain diffuse optical tomography, 1995.

[110] A H Hielscher, R E Alcouffe, and R L Barbour. Comparison of finite difference transport and diffusion calculations for photon migration in homogeneous and heterogeneous tissues. Physics in Medicine and Biology, 43(5):1285–1302, 1998.

[111] M Schweiger, S R Arridge, and D T Delpy. Application of the finite-element method for the forward and inverse models in optical tomography. Journal of Mathematical Imaging and Vision, 3(3):263–283, 1993.

[112] K D Paulsen and H Jiang. Spatially varying optical property reconstruction using a finite element diffusion equation approximation. Medical Physics, 22(6):691–701, 1995.

[113] H Gao and H Zhao. A fast-forward solver of radiative transfer equation. Transport Theory and Statistical Physics, 38(3):149–192, 2009.

[114] H Dehghani, S Sri Nivasan, B W Pogue, and A Gibson. Numerical modelling and image reconstruction in diffuse optical tomography. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 367(1900):3073–3093, 2009.

[115] V Y Soloviev, C D'Andrea, P Surya Mohan, G Valentini, R Cubeddu, and S R Arridge. Fluorescence lifetime optical tomography with discontinuous Galerkin discretisation scheme. Biomedical Optics Express, 1(3):998–1013, 2010.

[116] W Xie, Y Deng, L Lian, K Wang, Z Luo, and H Gong. Boundary element method for diffuse optical tomography. In Proceedings - 2013 7th International Conference on Image and Graphics, ICIG 2013, pages 5–8, 2013.

[117] F Fedele, M J Eppstein, J P Laible, A Godavarty, and E M Sevick-Muraca. Fluorescence photon migration by the boundary element method. Journal of Computational Physics, 210(1):109–132, 2005.

[118] S Srinivasan and H Ghadyani. 3-D Image-guided diffuse optical tomography using boundary element method and MPI implementation. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, pages 8452–8454, Thayer School of Engineering, Dartmouth College, 8000 Cummings Hall, Hanover, NH 03755, United States, 2011.

[119] Kolehmainen, V. (2001). Novel approaches to image reconstruction in diffusion tomography. PhD thesis, University of Kuopio.

[120] Bal, G. (2009). Inverse transport theory and applications. Inverse Problems, 25(5):48.

[121] Ahmad, S. U. (2019). Analytical and Iterative Regularization Methods for Nonlinear Ill-posed Inverse Problems: Applications to Diffuse Optical and Electrical Impedance Tomography (Doctoral dissertation, Clemson University).

[122] Dai, X.; Zhang, T.; Yang, H.; Tang, J.; Carney, P.R.; Jiang, H. Fast noninvasive functional diffuse optical tomography for brain imaging. Journal of biophotonics 2018, 11, e201600267. doi:10.1002/jbio.201600267.

[123] Zimmermann, B.B.; Deng, B.; Singh, B.; Martino, M.; Selb, J.J.; Fang, Q.; Sajjadi, A.Y.; Cormier, J.A.; Moore, R.H.; Kopans, D.B.; others. Multimodal breast cancer imaging using coregistered dynamic diffuse optical tomography and digital breast tomosynthesis. Journal of Biomedical Optics 2017, 22, 046008. doi:10.1117/1.JBO.22.4.046008.

[124] Yoo, J.; Sabir, S.; Heo, D.; Kim, K.H.; Wahab, A.; Choi, Y.; Lee, S.I.; Chae, E.Y.; Kim, H.H.; Bae, Y.M.; others. Deep learning diffuse optical tomography. IEEE transactions on medical imaging 2019, 39, 877–887. doi:10.1109/TMI.2019.2936522.

[125] Cochran, J.M.; Busch, D.R.; Lin, L.; Minkoff, D.L.; Schweiger, M.; Arridge, S.; Yodh, A.G. Hybrid time-domain and continuous-wave diffuse optical tomography instrument with concurrent, clinical magnetic resonance imaging

for breast cancer imaging. Journal of biomedical optics 2019, 24, 051409. doi:10.1117/1.JBO.24.5.051409.

[126] Vavadi, H.; Mostafa, A.; Zhou, F.; Uddin, K.S.; Althobaiti, M.; Xu, C.; Bansal, R.; Ademuyiwa, F.; Poplack, S.; Zhu, Q. Compact ultrasound-guided diffuse optical tomography system for breast cancer imaging. Journal of biomedical optics 2018, 24, 021203. doi:10.1117/1.JBO.24.2.021203.

[127] Taroni, P.; Pifferi, A.; Quarto, G.; Spinelli, L.; Torricelli, A.; Abbate, F.; Villa, A.M.; Balestreri, N.; Menna, S.; Cassano, E.; others. Noninvasive assessment of breast cancer risk using time-resolved diffuse optical spectroscopy. Journal of biomedical optics 2010, 15, 060501. doi:10.1117/1.3506043.

[128] Zhu, Q.; Hegde, P.U.; Ricci Jr, A.; Kane, M.; Cronin, E.B.; Ardeshirpour, Y.; Xu, C.; Aguirre, A.; Kurtzman, S.H.; Deckers, P.J.; others. Early-stage invasive breast cancers: potential role of optical tomography with US localization in assisting diagnosis. Radiology 2010, 256, 367–378. doi:10.1148/radiol.10091237.

[129] Ferradal, S.L.; Liao, S.M.; Eggebrecht, A.T.; Shimony, J.S.; Inder, T.E.; Culver, J.P.; Smyser, C.D. Functional imaging of the developing brain at the bedside using diffuse optical tomography. Cerebral cortex 2016, 26, 1558–1568. doi:10.1093/cercor/bhu320.

[130] Hernandez-Martin, E.; Gonzalez-Mora, J.L. Diffuse Optical Tomography Using Bayesian Filtering in the Human Brain. Applied Sciences 2020, 10, 3399. doi:10.3390/app10103399.

[131] Lee, C.W.; Cooper, R.J.; Austin, T. Diffuse optical tomography to investigate the newborn brain. Pediatric research 2017, 82, 376–386. doi:10.1038/pr.2017.107.

[132] Arridge, S.R.; Schotland, J.C. Optical tomography: forward and inverse problems. Inverse problems 2009, 25, 123010. doi:10.1088/0266-5611/25/12/123010.

[133] Klose, A.D.; Hielscher, A.H. Iterative reconstruction scheme for optical tomography based on the equation of radiative transfer. Medical physics 1999, 26, 1698–1707. doi:10.1118/1.598661.

[134] Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 2014.

[135] Duchi, J.; Hazan, E.; Singer, Y. Adaptive subgradient methods for online learning and stochastic optimization. Journal of machine learning research 2011, 12. doi:10.1118/1.598661.

[136] Tieleman, T.; Hinton, G. Lecture 6.5-rmsprop, coursera: Neural networks for machine learning. University of Toronto, Technical Report 2012.

[137] Reddi, S.J.; Hefny, A.; Sra, S.; Poczos, B.; Smola, A. Stochastic variance reduction for nonconvex optimization. International conference on machine learning, 2016, pp. 314–323.

[138] Tran, P.T.; others. On the convergence proof of amsgrad and a new version. IEEE Access 2019, 7, 61706–61716. doi:10.1109/ACCESS.2019.2916341.

[139] Ruder, S. An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747 2016.

[140] Zhou, D.; Tang, Y.; Yang, Z.; Cao, Y.; Gu, Q. On the convergence of adaptive gradient methods for nonconvex optimization. arXiv preprint arXiv:1808.05671 2018.

[141] Arridge, S.; Schweiger, M.; Hiraoka, M.; Delpy, D. A finite element approach for modeling photon transport in tissue. Medical physics 1993, 20, 299–309. doi:10.1118/1.597069.

[142] Rodrıguez, J.A.O. Regularization methods for inverse problems. PhD thesis, Citeseer, 2011.

[143] Colton, D.; Kress, R. Inverse acoustic and electromagnetic scattering theory; Vol. 93, Springer Nature, 2019.

[144] Kruskal, W.H.; Wallis, W.A. Use of ranks in one-criterion variance analysis. Journal of the American statistical Association 1952, 47, 583–621. doi:10.2307/2280779.

[145] Gehan, E.A. A generalized two-sample Wilcoxon test for doubly censored data. Biometrika 1965, 52, 650–653. 405 doi:10.2307/2333721.

[146] Y.Zhao, A.Raghuram, H.Kim, A.Hielscher, J. T.Robinson, A. N.Veeraraghavan. High Resolution, Deep Imaging Using Confocal Time-of-flight Diffuse Optical Tomography. IEEE Transactions on Pattern Analysis and Machine Intelligence. (2021)

[147] M. B.Applegate, R. E.Istfan, S.Spink, A.Tank, D.Roblyer, Recent advances in high speed diffuse optical imaging in biomedicine. APL Photonics, 5(4), 040802. (2020).

[148] N. Chakhim, M. Louzar, A. Lamnii, M. Alaoui, Image Reconstruction in Diffuse Optical Tomography Using Adaptive Moment Gradient Based Optimizers: A Statistical Study. Applied Sciences, 10(24), (2020), 9117.

[149] T. Dozat, Incorporating nesterov momentum into adam. (2016).

[150] Z. Wang, E. P. Simoncelli, A. C. Bovik, Multiscale structural similarity for image quality assessment. In The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003 (Vol. 2, pp. 1398-1402). Ieee.

[151] A. Hore, D. Ziou, Image quality metrics: PSNR vs. SSIM. In 2010 20th international conference on pattern recognition (pp. 2366-2369). (2010, August). IEEE.

[152] S. R. Arridge, M. Schweiger, M. Hiraoka, D. T.Delpy, A finite element approach for modeling photon transport in tissue. Medical physics, 20(2), (1993), 299-309.

[153] J. Verzani, Getting started with RStudio. O'Reilly Media, Inc. (2011).

[154] Nada Chakhim, Mohamed Louzar, Mohammed Alaoui, A new approach to improve optimizer performance through algorithms diversification for image reconstruction in diffuse optical tomography, Commun. Math. Biol. Neurosci., 2021 (2021) .