# NEW ALGORITHMS FOR SOLVING INVERSE SOURCE PROBLEMS IN IMAGING TECHNIQUES WITH APPLICATIONS IN FLUORESCENCE TOMOGRAPHY

A Thesis
Presented to
The Academic Faculty

by

Ke Yin

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Mathematics

Georgia Institute of Technology
August 2013

# NEW ALGORITHMS FOR SOLVING INVERSE SOURCE PROBLEMS IN IMAGING TECHNIQUES WITH APPLICATIONS IN FLUORESCENCE TOMOGRAPHY

Approved by:

Professor Hao-Min Zhou,
Committee Chair
School of Mathematics
*Georgia Institute of Technology*

Professor Hao-Min Zhou, Advisor
School of Mathematics
*Georgia Institute of Technology*

Professor Ali Adibi
School of Electrical and Computer Engineering
*Georgia Institute of Technology*

Professor Shui-Nee Chow
School of Mathematics
*Georgia Institute of Technology*

Professor Luca Dieci
School of Mathematics
*Georgia Institute of Technology*

Professor Sung Ha Kang
School of Mathematics
*Georgia Institute of Technology*

Date Approved: 25 April 2013

*To my parents.*

# ACKNOWLEDGEMENTS

I owe a lot of thanks to many people, who have guided and supported me during my five-year journey of graduate studies in Georgia Tech.

First, I want to thank my advisor, Dr. Hao-Min Zhou, who has guided me on the path to an academia life, which I could not imagine five years ago. Not only did he teach me mathematics, he also helped me strengthen my faith and confidence in enduring the hardships in doing science. I am deeply impressed by his optimistic attitude towards difficulties, and often inspired by his thoughtful comments in the discussions. I would also like to express my sincere gratitude to him for his constant care and advisement in my daily life which help me going through the hard times.

I would like to thank Professors Ali Adibi, Shui-Nee Chow, Luca Dieci, and Sung Ha Kang for serving as members of my thesis committee. I would like to convey my sincere thanks to Professor Kang for her insightful instructions and lectures, which broadened my mathematical knowledge and brought the field of image processing to my attention. I also owe thanks to Professor Chow for his constant mentorship, and many inspiring discussions. I am grateful to Professor Dieci for recruiting me to Georgia Tech, helping me make a smooth transition to the new environment, and providing continuing support during his tenure as our graduate coordinator.

It is my pleasure to thank Ms. Cathy Jacobson, not only for improving my English skills, but also for teaching me life tips for living in America. I would also like to thank Professor Klara Grodzinsky who leads the teaching assistant training program in the School of Math. I benefit greatly from the training she provided.

I want to thank Ms. Karen Hinds, Ms. Genola Turner, Ms. Sharon McDowell and the School of Math IT group for their unfailing support.

I am indebted to many of my colleagues, including fellow graduate students and visiting scholars. I'm especially grateful to Ruoting Gong, Xun Huang, Huy Huynh, Jinyong Ma, Jie Ma, and Lei Wang for their precious help to me in my first year. I am thankful to Weizhe Zhang, Yi Huang, Tianjun Ye, Maria Reguera Rodriguez, Linwei Xin, Thao Vuong, and Hyunshik Shin who came to Georgia Tech in the same year as me, and went through the hard time together. I am also grateful to many other friends I met in Georgia Tech, especially Yunlong He, Ruidong Wang, Jingfang Liu, Jun Lu, Xiaoling Wang, Yao Li, Lili Hu, Ruodu Wang, Wuchen Li, Lei Zhang, Romeo Awi, Giang Do, Kai Ni and Yan Shu. I owe thanks to Dr. Xiaojing Ye for sharing his valuable experience in academia life with me. My work benefited from collaboration with Ali Behrooz, and I thank him for sharing his expertise in optics. I thank Rui Tuo, Xiaopeng Shao and Guojing He for their friendship during their visit to Georgia Tech.

I would like also thank some of my friends I knew since my college days: Gongjun Xu, Xiang Zhang, Dewei Wang, and Ke Han, all of whom are now working or studying in USA. Our last reunion in March 2013 in Atlanta is memorable.

Last, yet most importantly, I would like to express my deep thanks to my parents, Zuohan Yin and Xiaoling Zhang, without whose constant and unconditional love and support, I would not have survived this long journey. To them I delicate this thesis.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# SUMMARY

This thesis is devoted to solving the inverse source problems arising in image reconstruction techniques. In general, the solution is non-unique and the problem is severely ill-posed. Therefore. It poses tremedous challenges in the numerical computation.

In Chapter I, we discuss several existing methods to solve this problem, especially the most widely used Tikhonov-type regularization method. Tikhonov regularization minimizes a cost function which is a linear combination of a regularization term and a data fitting term. The coefficients for the linear combination are called the regularization parameters, which balance between regularization and data fitting. The parameter tuning can be achieved by L-curve or cross validation [64, 59]. Besides parameter tuning, solving Tikhonov regularization has been an active area of research. Some proposed approaches are the Interior Point (IP) method [56, 20], the Algebraic Reconstruction Technique (ART) [68], and the newly revived Augmented Lagrangian (AL) method [104, 58].

In Chapter II, we propose a method which overcomes the major difficulties, namely the non-uniqueness of the solution and noisy data fitting, separately. First we find a particular solution called the orthogonal solution that satisfies the data fitting term. Then we add to it a correction function in the kernel space so that the final solution fulfills the regularization and other physical requirements. The key idea is that the correction function in the kernel has no impact to the data fitting, and the regularization is imposed in a smaller space. Moreover, there is no parameter needed to balance the data fitting and regularization terms. In addition, we use an efficient basis to represent the source function to form a hybrid strategy using spectral

methods and finite element methods in the algorithm. The resulting algorithm can dramatically increase the computation speed over the existing methods.

In Chapter III, we apply the proposed method to Fluorescence Tomography (FT). Fluorescence Tomography (FT) is an emerging, *in vivo* non-invasive 3-D imaging technique which reconstructs images that characterize the distribution of molecules that are tagged by fluorophores. Compared to other medical imaging modalities, FT is less harmful. It uses near-infrared (NIR) light (650-900 nm wavelength) which is within the spectrum of the sunlight, compared to the X-ray that is used in CT and powerful magnetic fields deployed in MRI, where the dosage must be limited for safety concerns. Another advantage of FT is that it can capture molecular specific information that cannot be obtained otherwise [67]. For this reason, it is considered to be a promising method in early cancer detection and drug monitoring [66, 103, 22]. However, this imaging modality has not yet become popular in clinical practice, partially because the severe ill-posedness of the problem, low resolution in the reconstructed image by existing state of the art algorithm, and huge computation cost. We demonstrate by theory, algorithms, and examples with synthetic as well as real data, that some of these challenges can be addressed by developing efficient mathematical tools.

# CHAPTER I

# INTRODUCTION

Many imaging techniques, such as X-Ray Computed Tomography (CT), Magnetic Resonance Imaging (MRI), and an emerging modality called Fluorescence Tomography (FT), can be modeled by a linear model, which is written as

$$g = Af. \tag{1}$$

It relates the unknown $f$ to the collected data $g$ through a linear operator $A$. For example, in CT, $f$ denotes the distribution of the attenuation coefficient in the examined tissue. It is modeled by a 3D image, where each voxel (volumetric pixel) contains the value of the attenuation at that point. $A$ is the discrete form of the line integral of the attenuation coefficient along each X-ray path in use, which simply sums up the values in the voxels that the X-ray passes through. $g$ is the measurement of the total attenuation in the X-ray along each path.

The above model (1) is often called the forward model, where given $f$, we can compute $g$. In practice, the story goes in the reverse way. $g$ is given, or partially given, and $f$ is to be computed. It is called an inverse problem. If we consider $f$ to be the source for the incidence signal $g$, then this is an inverse source problem.

Even if the forward model is linear, the method for solving the unknown source term $f$ is often nonlinear [85]. There are several reasons for this. First, the forward model is often under-determined, hence the solution is non-unique. For example, in CT and MRI, due to the limited amount of dose in use, the data is not fully collected, and the more data are missing, the more the problem is under-determined. Some other models are under-determined in theory, such as FT, where $A$ is a diffusion operator, and the data $g$ is obtained only on the boundary. Non-uniqueness also present in

some other inverse source problems, such as in acoustics and electromagnetics, where there exist physically realizable sources for which the measurements are exactly zero [18]. In order to overcome the non-uniqueness of the solution, some prior information is needed as a constraint for the solution, and the problem is often proposed as a constraint optimization program, which is usually solved by iterative algorithms.

Another reason that the solution method being nonlinear is because the forward model is usually ill-posed. After discretization, the linear operator $A$ in the model (1) often turns out to be an ill-conditioned matrix. Therefore, small perturbations, such as noise in the data $g$, and modeling imperfection in $A$ itself, can cause huge errors in the computation of $f$. Regularization techniques are need in order to make the solution robust against certain perturbations.

Both under-determinacy and ill-posedness pose tremendous challenges in solving inverse source problems. In history, people have developed various approaches specific to difference imaging modalities. For instance, the filtered back projection has been the commercial standard for CT, and the Inverse Fourier Transform is the staple algorithm built in to the MRI machines. See [47] for a comprehensive introduction. Recently, Tikhonov-type regularization has revived as a promising approach alternative to previous methods [33, 53, 44]. It can be written in the following form

$$\hat{f} = \underset{f \geq 0}{\operatorname{argmin}} \|Af - g\|^2 + \mu \Psi_{\text{reg}}(f). \tag{2}$$

Here $\|Af - g\|$ is the data fitting term to match the boundary measurements, where $\|\cdot\|$ denotes the $L^2$ norm by default. $\Psi_{\text{reg}}(f)$ is the regularization term to impose the regularity of the solution, which also alleviates the ill-posedness of the numerical computations.

Both smooth and non-smooth functionals are used for the regularization term $\Psi_{\text{reg}}(f)$. The original Tikhonov regularization seeks a smooth solution [97], where $\Psi_{\text{reg}}(f)$ is the smooth functional. It is demonstrated that the popular $L^2$ regularization is robust against the noise and can overcome ill-posedness of the the problem.

However, it is recognized that the $L^2$ regularization can make the solution overly smoothed, and features that are important in clinical and engineering applications, such as the resolution of small objects and sharp changes like edges, are often missing. Recently, some non-smooth regularization terms have become popular, such as $L^1$ norm $\|f\|_1$, which promotes the sparsity [56], and total variation (TV) semi-norm $\|f\|_{\mathrm{TV}}$, which tends to preserve edges in the image [45, 15]. In order to get the benefits of both, authors in [44, 57] use linear combinations of $L^1$ norm and TV semi-norm as the regularization term. For each choice of those regularization terms, Tikhonov regularization (2) defines a unique solution, which can be regarded as a stable approximation to the original problem (1) [45].

In Tikhonov regularization methods, the regularization parameter $\mu$ plays an important role. It balances the data fidelity and the regularity of the solution. If $\mu$ is relatively small so the data fidelity is good but the regularization is not strongly enforced. The story is the opposite, when $\mu$ is large: data fidelity is poor but the regularization is well enforced. There are many studies focusing on how to choose $\mu$. Among different approaches, the L-curve method [64] and the cross validation method [59] are common strategies for the selection of the parameter $\mu$.

In this thesis, we propose a new approach to tackle the challenges. Our main idea is to handle the two major challenges separately so that each one can be addressed more efficiently. Briefly speaking, we first find a particular solution that matches the boundary measurements. In this step, there is no regularity requirement for this particular solution, and there are many choices. After finding the particular solution, we then add to it a function in the kernel space of $A$ so that the final solution meets the regularization requirements. Since the correction is only in the kernel space, it does not alter the data matching property obtained by the particular solution. Moreover, the regularity is achieved only in the kernel space, which is smaller than the entire space used in Tikhonov regularization methods.

More precisely, the true solution to the inverse source problem (1) can be decomposed as

$$f = f_* + f_0, \tag{3}$$

where $f_*$ is a particular solution used to match the boundary measurements, and $f_0$ is a function in the kernel space $\mathcal{N}(A)$ to fulfill regularity constrains. In the first step, the particular solution $f_*$ is chosen to be

$$f_* = A^*(AA^*)^{-1}g, \tag{4}$$

where $A^*$ is the adjoint operator of $A$. In theory, $f_*$ is the solution that is orthogonal to the kernel $\mathcal{N}(A)$. For this reason, we call it the orthogonal solution, which is also known as minimal norm solution or Moore-Penrose pseudo inverse in the literature [46, 61].

If there is no noise in the measurements and no errors in the modeling, $f_*$ has the perfect data fitting property. Otherwise, we cannot solve $f_*$ exactly. Instead, we compute an approximation to it by an iterative regularization procedure. Obviously, the particular solution may not satisfy regularity constrains that are required in the applications, such as positivity, or visually smooth features, or sparsity. These constrains are addressed in the second step of the method. We choose a correction $f_0$ in the kernel space $\mathcal{N}(A)$ such that the combined solution $f$ given in (3) satisfies the desirable regularity requirements. We realize this step by solving a constrained optimization problem

$$\hat{f}_0 = \underset{f_0 \in \mathcal{N}(A)}{\operatorname{argmin}} \Psi_{\text{reg}}(f_* + f_0) \quad \text{such that } f_* + f_0 \geq 0, \tag{5}$$

where the regularization $\Psi_{\text{reg}}$ is chosen as $L^2$, or $L^1$ norms, or TV semi-norm. Since $f_0$ is in the kernel and does not affect the data fitting achieved by $f_*$, the final solution $f$ preserves the correct data fitting of $f_*$ while having the desired regularity. In addition, there is no regularization parameter selection needed to balance the data fidelity and regularity requirements.

In this paper, we call the proposed two-step approach the Orthogonal Solution and Kernel Correction Algorithm (OSKCA).

The two steps (4) and (5) in our algorithm can be solved by some existing methods. For example, several algorithms can be used to solve (4), such as the gradient based Landweber iterations [13]. However, the ill-conditioning in $A$ can cause very slow convergence in FT application. To address the problem, people develop a method to compute $(AA^*)^{-1}g$ by iterative Tikhonov regularization [82]. For the second step (5), if $\Psi_{\text{reg}}$ is taken as $L^2$-norm, $L^1$-norm, or TV semi-norm, we can take advantage of existing fast algorithms such as the Augmented Lagrangian, also known as the split Bregman iteration in the literature [58] .

In addition, to further speed up the computation, we introduce different bases to represent the solution $f$ in our algorithms. Most of the existing methods use point-wise basis to represent the solution. In this setting, the number of unknowns in the solution is equal to the number of pixels (or voxels for 3D images). We notice that the solution $f$, represented by an image either in 2D or 3D, usually have certain regularities, such as being $L^2$ integrable and locally smooth with possibly sharp transitions. Therefore, some bases other than the point-wise basis, such as harmonic functions or wavelets, are much more efficient to represent the unknowns. This dramatically reduces the dimension of the unknowns. It is shown in our numerical experiments that the new method gains significant speedup over the existing methods.

# CHAPTER II

# ORTHOGONAL SOLUTION WITH KERNEL CORRECTION ALGORITHM (OSKCA)

As we have discussed, regularization techniques are often needed to handle the ill-posedness of the problem (1). The Tikhonov regularization as in (2) is widely used in the existing methods. However, there are limitations associated with this approach. For example, the regularization parameter tuning can be difficult and expensive. Also the reconstruction has low resolution, because it may be overly smooth or too noisy if the regularization parameter is not properly chosen. In order to overcome such difficulties, we propose the OSKCA method to compute the solution. In this section, we describe the algorithm in detail.

We notice that any solution $f$ to (1) can be decomposed as

$$f = f_* + f_0, \tag{6}$$

where $f_*$ is a particular solution, and $f_0$ is a function in the kernel $\mathcal{N}(A)$. Then to solve the equation (1) we just need to determine $f_*$ and $f_0$.

First, we choose $f_*$ such that it satisfies

$$g = Af_* \text{ and } f_* \in \mathcal{N}(A)^\perp, \tag{7}$$

where $\mathcal{N}(A)^\perp$ refers to the orthogonal complement of $\mathcal{N}(A)$. Later we show that such defined $f_*$ exists, and it is uniquely determined by $g$ and $A$. We call $f_*$ the *orthogonal solution* because it is perpendicular to the kernel $\mathcal{N}(A)$. Once $f_*$ is determined, we choose $f_0 \in \mathcal{N}(A)$ such that $f_* + f_0$ satisfies the regularity requirements. We call $f_0$ the *kernel correction*. In summary, OSKCA is given as follows:

**Algorithm 2.0.1** Orthogonal Solution and Kernel Correction Algorithm (OSKCA)

1. Formulate $A$ and $g$ for the inverse problem $Af = g$.
2. Compute the orthogonal solution $f_*$.
3. Compute the correction in the kernel $f_0 \in \mathcal{N}(A)$ such that $f = f_* + f_0$ satisfies the regularity requirements, i.e.,

$$\hat{f}_0 = \underset{f_0 \in \mathcal{N}(A)}{\operatorname{argmin}} \Psi_{\text{reg}}(f_* + f_0) \quad \text{such that} \quad f_* + f_0 \geq 0, \tag{8}$$

   where $\Psi_{\text{reg}}$ is a regularization functional.

In theory, OSKCA is equivalent to solving the optimization problem

$$\hat{f} = \underset{f \geq 0}{\operatorname{argmin}} \Psi_{\text{reg}}(f) \quad \text{such that} \quad Af = g. \tag{9}$$

Compared with Tikhonov regularization (2), the data fitting in (9) can be enforced without jeopardizing the regularity requirements. This is a desirable property, because $A$ is severely under-determined. Also, the equality constraint in (9) can be strongly enforced if the noise level is low, and loosely enforced if the noise level is high.

In the remaining part of this section we describe the orthogonal solution and kernel correction in more details. We also demonstrate that common regularization techniques (like $L^2$, $L^1$, and TV minimization) can be incorporated into the proposed framework.

## *2.1 The orthogonal solution*

The orthogonal solution to the inverse problem $g = Af$ is

$$f_* = A^*(AA^*)^{-1}g, \tag{10}$$

if $g \in \mathcal{R}(AA^*)$, where $\mathcal{R}(\cdot)$ denotes the range. We note that the closure of $\mathcal{R}(A)$, $\overline{\mathcal{R}(A)}$, is the same as $\overline{\mathcal{R}(AA^*)}$ by standard results in functional analysis [88].

If $g$ contains noise, it may not be in $\mathcal{R}(A)$ or $\mathcal{R}(AA^*)$. We can project $g$ onto

$\mathcal{R}(AA^*)$ so that (10) is strictly applicable. Though it is not needed in the computation, since we can make the algorithm for (10) is robust against the noise.

Since $A$ is a compact operator, $A^*(AA^*)^{-1}$ is unbounded. Therefore, the orthogonal solution may not depend continuously on the right hand side $g$. So regularization techniques become necessary to compute it. We propose a numerical method for the orthogonal solution in Section 3.3.3.

If $g \in \mathcal{R}(A)$, then the well-known minimal norm solution is given by

$$\bar{f} = A^\dagger g,$$

where $A^\dagger$ is the Moore-Penrose (generalized) inverse of $A$ (See [46]). The minimal norm solution and the orthogonal solution are closely related, and their connection is stated in the following:

**Proposition 1** *1. For every $g \in \mathcal{R}(AA^*)$, $A^\dagger g = A^*(AA^*)^{-1}g$.*

*2. For every $g \in \mathcal{R}(A) \setminus \mathcal{R}(AA^*)$, $A^\dagger g \neq A^*(AA^*)^{-1}g$.*

In short, the minimal norm solution is defined in a larger space than the orthogonal solution in (10), and the two solutions coincide when $g \in \mathcal{R}(AA^*)$.

## *2.2 The kernel correction*

The kernel correction $f_0$ is chosen such that $f = f_* + f_0$ satisfies the regularity requirements. Suppose $W = \{w_i\}_{i=1}^{\infty}$ is an orthonormal basis for $\mathcal{N}(A)$, then $f_0 = W\mu$ where $\mu$ is the auxiliary variable which denotes the coefficient for $f_0$ under the basis $W$. The problem of finding the kernel correction can be written as a constrained optimization problem

$$\hat{\mu} = \underset{\mu}{\operatorname{argmin}}\ \Psi_{\text{reg}}(f_* + W\mu) \quad \text{subject to} \quad f_* + W\mu \geq 0, \tag{11}$$

where $\Psi_{\text{reg}}$ is a regularization functional. The final solution is $f_* + W\mu$, which is required to be non-negative since it is a distribution. Here we note that (11) is

equivalent to (9). However, the auxiliary variable $\mu$ has a smaller dimension than $f$, so the size of the problem is reduced.

The regularization functional $\Psi_{\text{reg}}$ can be chosen differently. For instance, if $L^2$ minimization is used, the problem (11) becomes

$$\hat{\mu} = \underset{\mu}{\operatorname{argmin}} \|f_* + W\mu\| \quad \text{subject to} \quad f_* + W\mu \geq 0. \tag{12}$$

We can also use some other regularization requirements. For example, the famous total variation minimization can be applied, which is helpful if sharp transitions such as edges are expected in the reconstructed image. In that case, we solve the constrained optimization problem

$$\hat{\mu} = \underset{\mu}{\operatorname{argmin}} \|f_* + W\mu\|_{TV} \quad \text{subject to} \quad f_* + W\mu \geq 0, \tag{13}$$

Likewise, other requlariza tion techniques can be formulated similarly.

The kernel correction is used for enforcing regularity of the solution. Since it is solved in the kernel space, it does not affect the data fitting of the solution.

### *2.3 Representation of the solution under a chosen basis*

Inspired by the spectral method, other than representing the solution $f$ by its point values, we can also choose a basis $\{\xi_i\}_{i=1}^{\infty}$ for the solution space and write $f$ as

$$f = \sum_{i=1}^{\infty} c_i \xi_i, \tag{14}$$

Let $c = (c_1, \ldots, c_i, \ldots)$ be the coefficient for $f$ under the basis. Let $B$ denote the linear transformation from the spectral domain to the physical domain, which is defined by

$$B : (c_1, \ldots, c_i, \ldots) \mapsto \sum_{i=1}^{\infty} c_i \xi_i. \tag{15}$$

Then (14) implies the relation

$$f = Bc. \tag{16}$$

By (16), the inverse problem in (66) is rewritten as

$$g = ABc, \text{ or } g = Mc, \tag{17}$$

where $M$ denotes the composition of $A, B$.

The idea of OSKCA still applies to the new formulation (17). $c$ is decomposed as

$$c = c_* + c_0, \tag{18}$$

where $c_*$ is the orthogonal solution to (17) and $c_0 \in \mathcal{N}(M)$ is the kernel correction term. Let $K$ be an orthonormal basis for $\mathcal{N}(M)$. $\lambda$ is the auxiliary variable for $c_0$, and

$$c_0 = K\lambda. \tag{19}$$

By (16)(18), the solution to (66) is written as

$$f = B(c_* + c_0). \tag{20}$$

By (19), (20) is equivalent to

$$f = B(c_* + K\lambda). \tag{21}$$

$B$ is determined by the chosen basis functions and $K$ can be computed if $M$ is given. $c_*$ is computed by an analogy to (10), which is written as

$$c_* = M^*(MM^*)^{-1}g. \tag{22}$$

We have the following lemma relating $c_*$ to $f_*$.

**Lemma 2.3.1** *If the basis functions $\{\xi_i\}_{i=1}^{\infty}$ are orthonormal, then*

$$f_* = Bc_*. \tag{23}$$

**Proof** By the condition in the lemma, $B$ as defined in (15) is an unitary transformation, and its adjoint operator $B^*$ satisfies

$$BB^* = I(\text{identity}). \tag{24}$$

In this case, $c_*$ as defined in (17) can be written as

$$c_* = (AB)^*(ABB^*A^*)^{-1}g = B^*A^*(AA^*)^{-1}g, \tag{25}$$

so

$$Bc_* = BB^*A^*(AA^*)^{-1}g = A^*(AA^*)^{-1}g = f_*. \tag{26}$$

In the theory the orthonormal basis functions have some nice analytical properties as described above. In practice, the basis functions do not have to be orthonormal. More generally, we can choose a frame, instead of a basis, to represent the solution. The approach that analyzing the solution using a frame, for example the wavelet frame, has become popular in the field of image and signal processing [29, 26]. One obvious reason is that some analytical properties of vectors in the underlying function space, such as orthogonality (24), are not preserved in their discrete form. The frame approach provides some useful tools to study the regularity of the solution, such as the Total Variation semi-norm and the $L^1$ norm. For non-orthogonal bases or the frames, (23) does not hold. However, (20) is always satisfied.

The computation of the kernel correction term $c_0$ follows the same way as (11). We note the solution $f$ in the form (21), where $\lambda$ is obtained by solving the optimization problem

$$\hat{\lambda} = \underset{\lambda}{\operatorname{argmin}}\ \Psi_{\text{reg}}(B(c_* + K\lambda)) \quad \text{subject to} \quad B(c_* + K\lambda) \geq 0. \tag{27}$$

Here $\Psi_{\text{reg}}(\cdot)$ also denotes the regularization functional as in (11).

# CHAPTER III

# FLUORESCENCE TOMOGRAPHY (FT)

We now apply the OSKCA method described in the previous chapter to the image reconstruction problems arising in an emerging imaging technique call the Fluorescence Tomography (FT). In the experimental setup of FT (See Figure 1 for a cartoon demonstration), NIR radiation (wavelength 650-900 nm) is pumped into the examined

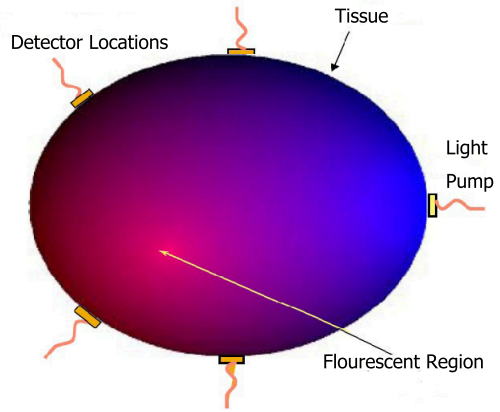



Figure 1: An illustration of Fluorescence Tomography

biological tissue through fibers placed on the tissue surface. The light is scattered and absorbed partially in the tissue and the fluorophores are excited by the diffuse excitation. The excited fluorophores then emit near-infrared (NIR) light at a longer wavelength, which propagates in the tissue. Then the intensity of the fluorescent emission is measured by the detectors placed on the tissue surface. The goal of FT is to reconstruct the distribution of fluorophores from boundary measurements, knowing the scattering and absorption parameters of the light.

The NIR light is strongly scattered in the biological tissue, and this can be modeled

by diffusive photons governed by the Radiative Transfer Equation (RTE) [70]. Although the RTE has been intensively studied in many other problems, and a number of schemes have been proposed to solve it numerically, it is still considered expensive to solve due to the high dimensionality in FT applications. To ease the challenge, the Diffusion Approximation (DA) to the RTE is introduced. It is a second order diffusion equation generally accepted as an accurate model in the regime of highly scattering and low absorptive media, such as the biological tissues [5, 90, 45, 44]. The boundary condition associated with the DA model is Robin (mixed) type [50, 65, 91, 80], which accounts for the partial reflection and transmission of the light on the boundary.

In this thesis, our goal is to design an efficient numerical method to reconstruct the image of the fluorophore distribution from the boundary measurements.

## *3.1 Mathematical models*

### 3.1.1 The governing equations

As described in the introduction, there are two radiative fields at different wavelengths: the excitation and the emission photons propagating in the examined tissue. Both of them can be modeled by the same radiative transfer equation (RTE) with different parameters:

$$\begin{aligned} \frac{\partial L(r,\hat{s},t)/c}{\partial t} &= -\hat{s}\cdot\nabla L(r,\hat{s},t) - (\mu_a+\mu_s)L(r,\hat{s},t) \\ &+ \mu_s \textstyle\int_{S^{d-1}} L(r,\hat{s}',t)P(\hat{s}',\hat{s})\,\mathrm{d}\hat{s}' + q(r,\hat{s},t). \end{aligned} \tag{28}$$

where $L$ denotes the radiance of the photon; $\mu_a$ and $\mu_s$ are absorption and scattering coefficients respectively; $d$ is the dimension of the space; $S^{d-1}$ is the collection of unit vectors; $P$ is the scattering kernel and $q$ is the light source. In the RTE, there are three spatial variables in $r$ and two angular variables in $\hat{s}$, in addition to the time variable $t$ if the problem is in 3-D. As this is a high dimensional PDE, it is usually computationally expensive to solve directly. There are some recent effort to compute it such as in [55]. To reduce the dimension, the diffusion approximation (DA) model

is introduced in practice, which is well accepted as an accurate approximation to the RTE when the light propagates in high scattering and low absorptive medium, such as the biological tissues [5]. More precisely, let us denote the light intensity, also called fluence rate, by

$$\Phi(r,t) = \int_{S^{d-1}} L(r,\hat{s},t)\,\mathrm{d}\hat{s}, \tag{29}$$

which is the sum of radiance in all directions. Then the DA is given by

$$\frac{\partial \Phi(r,t)}{c\partial t} + \mu_a \Phi(r,t) - \nabla\cdot[\kappa\nabla\Phi(r,t)] = q(r,t), \tag{30}$$

where $c$ is the light speed in the tissue, $\kappa = \frac{1}{3(\mu_a+\mu_s')}$ with $\mu_s'$ being the effective scattering coefficient, and $q(r,t)$ models the light source.

The derivation of DA from RTE follows that in [101]. It is important to assume that the light we consider should be highly scattering, and lowly absorptive in the examined medium. More exactly, the scattering events happen much more often than the absorption events. After numerous scattering events, few absorption events will occur and the radiance ($L(r,\hat{s},t)$) becomes nearly isotropic. Also, the change in the fluence rate ($\Phi(r,t)$) over one transport mean free path $1/\mu_s$ (which is the average distance travelled by a photon before being absorbed) is negligible. As a result, it is required that the distance from the interested region to the detector should be much larger than the mean free path (usually in the 0.1mm scale). These assumptions are often satisfied in the biological tissues, where Fluorescence Tomography is applied. To separate the intensity from the angular components, the radiance $L(r,\hat{s},t)$ is expanded into spherical harmonics, and truncate at the first order terms in $\hat{s}$ to get an approximation. That is,

$$L(r,\hat{s},t) = \frac{1}{4\pi}\Phi(r,t) + \frac{3}{4\pi}\vec{J}(r,t)\cdot\hat{s}, \tag{31}$$

where $\frac{1}{4\pi}\Phi(r,t)$ and $\frac{3}{4\pi}\vec{J}(r,t)$ are the coefficients for zeroth and first order spherical harmonics respectively. $\Phi(r,t)$ is given in (29) and

$$\vec{J}(r,t) = \int_{S^{N-1}} \hat{s}L(r,\hat{s},t)\,\mathrm{d}\hat{s}. \tag{32}$$

$\vec{J}(r,t)$ is called the current density. We note the assumption that after several mean free path, the radiance is nearly isotropic, and we write

$$\int_{S^{d-1}} \hat{s}' P(\hat{s},\hat{s}')\,\mathrm{d}\hat{s}' = g\hat{s}, \tag{33}$$

which means the scattering kernel $P$ is symmetric. Here

$$g = \int_{S^{d-1}} \hat{s}\cdot\hat{s}' P(\hat{s},\hat{s}')\,\mathrm{d}\hat{s}' \tag{34}$$

for all $\hat{s} \in S^{d-1}$. Also $P$ should satisfy the identity

$$\int_{S^{d-1}} P(\hat{s},\hat{s}')\,\mathrm{d}\hat{s}' = 1. \tag{35}$$

By plugging (31) into (28), and matching up the coefficients for the zeroth and first order spherical harmonics, we have the following two equations

$$\frac{\partial\Phi(r,t)}{c\partial t} = -\nabla\cdot\vec{J}(r,t) - \mu_a\Phi(r,t) + q(r,t) \tag{36}$$

and

$$\frac{\partial\vec{J}(r,t)}{c\partial t} = -\frac{1}{3}\nabla\Phi(r,t) - (\mu_a+\mu_s')\vec{J}(r,t), \tag{37}$$

where $q(r,t) = \int_{S^{N-1}} q(r,\hat{s},t)\,\mathrm{d}\hat{s}$, and $\mu_s' = (1-g)\mu_s$ is the reduced scattering coefficient.

Also we note the assumption that the change in fluence rate $\Phi(r,t)$ over one transport mean free path $(1/\mu_s)$ is negligible, i.e.,

$$\mu_a \ll \mu_s. \tag{38}$$

Hence the change in current density $\vec{J}(r,t)$ over one transport mean free path is negligible as well. Then in (37), we can assume

$$\frac{\partial\vec{J}(r,t)}{c\partial t} = 0, \tag{39}$$

then (37) becomes Fick's Law

$$\vec{J}(r,t) = -\frac{\nabla\Phi(r,t)}{3(\mu_a+\mu_s')}. \tag{40}$$

By plugging (40) into (36), we obtain the DA for RTE as

$$\frac{\partial \Phi(r,t)}{c\partial t} = \nabla \cdot \kappa \nabla \Phi(r,t) - \mu_a \Phi(r,t) + q(r,t), \tag{41}$$

where $\kappa = \frac{1}{3(\mu_a + \mu_s')}$ is the diffusion coefficient.

In the frequency domain, we have the equation for the coefficients corresponding to the modulation frequency $\omega$ [8]:

$$-\nabla \cdot [\kappa \nabla \Phi(r,\omega)] + (\mu_a + \frac{i\omega}{c})\Phi(r,\omega) = q(r,\omega) \quad \text{in } \Omega, \tag{42}$$

where $\Omega$ is the region occupied by the examined tissue. $\Phi(r,\omega)$ is called the fluence of radiation at frequency $\omega$ in optics. In intensity-based fluorescence tomography, the modulation frequency ($\omega$) is zero, resulting in a CW (continuous wave) DA which is written as

$$-\nabla \cdot (\kappa \nabla \Phi(r)) + \mu_a \Phi(r) = q(r) \quad \text{in } \Omega. \tag{43}$$

The above equation is usually accompanied by the partially reflecting and partially absorbing boundary condition, which is given by the following Robin (mixed) boundary condition [91, 41]:

$$\vec{n} \cdot [\kappa \nabla \Phi(r)] + R\Phi(r) = 0 \quad \text{on } \partial\Omega, \tag{44}$$

where $\partial\Omega$ is the boundary of $\Omega$, and the parameters $\kappa, R$ are given.

For the simplicity of notations, the DA equation (42) with Robin boundary condition (44) is written in short as

$$F(\kappa, \mu_a, R)\Phi = q, \text{ or } F\Phi = q. \tag{45}$$

By the classical PDE theory [75], $\Phi$ in equation (1) has a unique solution in the Sobolev space $H^1(\Omega)$ given $q(\cdot) \in L^2(\Omega)$ and $\Omega$ is a Lipschitz domain.

### 3.1.2 The forward and inverse problems

Let $\Phi_m$ be the fluorophore emission fluence. The mathematical problem in FT is to compute $f$ from the boundary measurements of $\Phi_m$ in the DA equation

$$F_m \Phi_m(r) = \Phi_x(r) f(r), \tag{46}$$

where $\Phi_x$ is the excitation fluence and $F_m = F(\kappa_m, \mu_{am}, R)$. The subscript $m$ represents the emission model. The point-wise product $\Phi_x(r)f(r)$ models the source of fluorophore emission. $\Phi_x$ is excitation light field induced by the boundary light sources $q$, which can be modeled by another DA equation

$$F_x \Phi_x(r) = q(r), \tag{47}$$

where $F_x = F(\kappa_x, \mu_{ax}, R)$. The subscript $x$ is for the excitation model.

In FT applications, only the boundary values of $\Phi_m$, denoted by $g$, can be obtained from detectors. Let $T_r$ be the Sobolev trace operator which takes the boundary value of a function in $H^1(\Omega)$, then

$$g = T_r \Phi_m. \tag{48}$$

By (46) and (48), the forward model that maps unknown $f$ to measurements $g$ is formulated as

$$g = T_r F_m^{-1}(\Phi_x f), \tag{49}$$

where $F_m^{-1}(\Phi_x f) \in H^1(\Omega)$ is the solution of (46).

The inverse problem is to find $f$ from given measurement $g$, the parameters in $T_r$ and $F_m$, and precomputed excitation field $\Phi_x$. Since $f$ is a distribution, it is naturally a non-negative function.

### 3.1.3 A fourth order PDE approach to find a particular solution for the inverse problem

The inverse source problem (49) can be reformulated as a PDE problem

$$\begin{cases} -\nabla \cdot (\kappa_m \nabla \Phi_m(r)) + \mu_{am} \Phi_m(r) = \Phi_x(r) f(r) \quad \text{in } \Omega, \\ \vec{n} \cdot [\kappa_m \nabla \Phi_m(r)] + R \Phi_m(r) = 0 \quad \text{on } \partial\Omega, \\ \Phi_m(r) = g \quad \text{on } \partial\Omega, \end{cases} \tag{50}$$

where the parameters $\kappa_m, \mu_{am}, R$ and the boundary data $g$ are given, $\Phi_x$ is computed by solving the excitation model, and $f$ is the unknown. Using previous notations, we denote the elliptic operator

$$F_m \equiv -\nabla \cdot (\kappa_m \nabla) + \mu_{am}. \tag{51}$$

Obviously, this problem does not have a unique solution. However, the following fourth order PDE is related to (50):

$$\begin{cases} F_m^2 \widetilde{\Phi}_m(r) = 0 \quad \text{in } \Omega, \\ \vec{n} \cdot [\kappa_m \nabla \widetilde{\Phi}_m(r)] + R \widetilde{\Phi}_m(r) = 0 \quad \text{on } \partial\Omega, \\ \widetilde{\Phi}_m(r) = g \quad \text{on } \partial\Omega, \end{cases} \tag{52}$$

where $F_m^2 = F_m \circ F_m$. In order to study the solvability of the above fourth order equation, we note the following result on $F_m^2$.

**Definition 3.1.1** *[75] In a bounded domain $\Omega \subset \mathbb{R}^d$, we consider the system of equations*

$$F\mathbf{u} \equiv \sum_{|i|,|j| \le m} (-1)^{|i|} D^i (a_{i,j}(x) D^j \mathbf{u}) = \mathbf{f}, \tag{53}$$

*where $D^i = \partial^{|i|} / \partial x_1^{i_1} \dots \partial x_n^{i_n}$, $i = (i_1, \dots, i_n)$, $|i| = i_1 + \dots + i_n$. $\mathbf{u}$ and $\mathbf{f}$ are $N$-dimensional vector-functions, and $a_{i,j}(x)$ is the $(i,j)$-th entry of the $N \times N$ symmetric matrix $\mathbf{A}(x)$. The system (53) is called strongly elliptic if for any real vector $\boldsymbol{\zeta} = (\zeta_1, \dots, \zeta_N) \neq 0$ and for arbitrary real numbers $\xi_1, \dots, \xi_d$ not all zeros, the quadratic*

*form*

$$\mathbf{A}(x;\xi)\boldsymbol{\zeta}\cdot\boldsymbol{\zeta} \equiv \sum_{|i|,|j|=m} a_{i,j}(x)\xi_1^{i_1+j_1}\dots\xi_d^{i_d+j_d}\boldsymbol{\zeta}\cdot\boldsymbol{\zeta} \tag{54}$$

*is positive.*

**Lemma 3.1.2** (52) *is a strongly elliptic equation.*

**Proof** $F_m$ as defined in (51) is a strongly elliptic operator for $\kappa_m > 0$, whose principal part (second order derivatives) corresponds to the quadratic form

$$\kappa_m(\xi_1^2+\xi_2^2)\zeta^2, \tag{55}$$

which is positive for any real numbers $\xi_1, \xi_2$ and scalar $\zeta$. Then $F_m^2$ has the quadratic form

$$(\kappa_m(\xi_1^2+\xi_2^2))^2\zeta^2, \tag{56}$$

which is also positive. By Definition 3.1.1, (52) is a strongly elliptic equation.

**Theorem 3.1.3** *The fourth order PDE* (52) *is uniquely solvable in* $H^2(\Omega)$ *if it is a strongly elliptic equation, and the parameters* $\kappa_m, \mu_{am}$ *are* $C^2(\Omega)$ *and bounded.*

**Proof** By Lemma 3.1.2, (52) is a strongly elliptic equation. The complete proof can be found (with slight adaptation to our problem) in [75].

We have the following theorem describing the relation between (50) and (52).

**Theorem 3.1.4** *We consider the equation*

$$F_m\widetilde{\Phi}_m(r) = \Phi_x(r)\widetilde{f}(r), \tag{57}$$

*where* $\widetilde{\Phi}_m(r)$ *is the solution to* (52)*. Then*

1. $\widetilde{f}(r)$ *in* (57) *is a particular solution to the problem* (50)*, and it is a solution to the inverse problem* (49) *as well.*

*2. For any $f_0 \in \mathcal{N}(T_r F_m^{-1} \Phi_x)$, that is,*

$$T_r F_m^{-1}(\Phi_x f_0) = 0, \tag{58}$$

*we have*

$$\langle \Phi_x \widetilde{f}, \Phi_x f_0 \rangle_{L^2(\Omega)} = 0. \tag{59}$$

**Proof** 1. It can be verified directly by plugging $\tilde{f}$ for $f$, and $\widetilde{\Phi}_m$ for $\Phi_m$ in (50).

2. We denote $\Phi_0 = F_m^{-1}(\Phi_x f_0)$. Then we have $\Phi_x f_0 = F_m \Phi_0$. Since $\Phi_0$ is the solution to the DA equation, it satisfies the Robin boundary condition

$$\vec{n} \cdot [\kappa_m \nabla \Phi_0] + R\Phi_0 = 0 \quad \text{on } \partial\Omega, \tag{60}$$

and by (58),

$$\Phi_0 = 0 \quad \text{on } \partial\Omega, \tag{61}$$

and also

$$\vec{n} \cdot [\kappa_m \nabla \Phi_0] = 0 \quad \text{on } \partial\Omega. \tag{62}$$

We use integration by parts and the above boundary conditions of $\Phi_0$ to calculate the left-hand side of (59),

$$\begin{aligned}
\langle \Phi_x \widetilde{f}, \Phi_x f_0 \rangle_{L^2(\Omega)} &= \int_\Omega F_m \widetilde{\Phi}_m \cdot F_m \Phi_0 \, \mathrm{d}x \\
&= \int_\Omega F_m \widetilde{\Phi}_m \cdot (-\nabla \cdot (\kappa_m \nabla \Phi_0) + \mu_{am} \Phi_0) \, \mathrm{d}x \\
&= \int_\Omega (\kappa_m \nabla \Phi_0 \cdot \nabla (F_m \widetilde{\Phi}_m) + \mu_{am} \Phi_0 F_m \widetilde{\Phi}_m) \, \mathrm{d}x \\
&\quad - \int_{\partial\Omega} \vec{n} \cdot (\kappa_m \nabla \Phi_0) F_m \widetilde{\Phi}_m \, \mathrm{d}\Gamma \\
&= \int_\Omega \Phi_0 \big( - \nabla \cdot (\kappa_m \nabla (F_m \widetilde{\Phi}_m)) + \mu_{am} F_m \widetilde{\Phi}_m \big) \, \mathrm{d}x \\
&\quad + \int_{\partial\Omega} \big( \vec{n} \cdot (\kappa_m \nabla (F_m \widetilde{\Phi}_m)) \Phi_0 - \vec{n} \cdot (\kappa_m \nabla \Phi_0) F_m \widetilde{\Phi}_m \big) \, \mathrm{d}\Gamma.
\end{aligned} \tag{63}$$

By the boundary conditions (61)(62), the boundary integral in (63) is equal to zero. The remaining integral over $\Omega$ is equal to

$$\int_{\Omega} \Phi_0 \cdot F_m^2 \widetilde{\Phi}_m \, \mathrm{d}x. \tag{64}$$

By (52), $F_m^2 \widetilde{\Phi}_m = 0$. Then the left-hand side of (63) is equal to zero. This completes the proof of part 2.

Although this theorem provides a well-defined particular solution, it is not efficient to solve in practice. It is well known that fourth order PDEs are difficult to solve [81, 62]. Moreover, this approach is valid for the case where only on light source is used. It is unknown whether it can be generalized to the case where multiple light sources are in use. We will discuss alternative approaches to solve the inverse problem (49) in the remaining part of this thesis.

### 3.1.4 FT model

In order to improve the conditioning of the forward model, multiple light sources are used for excitation [43]. Suppose that $s$ is the number of different light sources used in the experiments. For the $i$-th light source $q^{(i)}$, the excitation field is $\Phi_x^{(i)}$ and the emission field is $\Phi_m^{(i)}$. The boundary measurement of the emission field is $g^{(i)}$. By (49), we have

$$g^{(i)} = T_r F_m^{-1}(\Phi_x^{(i)} f), \tag{65}$$

where $i = 1, \ldots, s$. By vertically concatenating the above $s$ equations, we write the inverse problem as finding $f$ in

$$g = Af, \tag{66}$$

where

$$g = \begin{bmatrix} g^{(1)} \\ \vdots \\ g^{(s)} \end{bmatrix}, \quad Af = \begin{bmatrix} T_r F_m^{-1}(\Phi_x^{(1)} f) \\ \vdots \\ T_r F_m^{-1}(\Phi_x^{(s)} f) \end{bmatrix}. \tag{67}$$

For a vector space $\mathcal{V}$, we denote multiple tensor product by

$$\otimes^p \mathcal{V} = \underbrace{\mathcal{V} \otimes \ldots \otimes \mathcal{V}}_{p \text{ copies}}. \tag{68}$$

Then $A$ is a bounded linear operator mapping from $L^2(\Omega)$ to $\otimes^s L^2(\partial\Omega)$. Then the adjoint of $A$, denoted by $A^*$, is a linear mapping from $\otimes^s L^2(\partial\Omega)$ to $L^2(\Omega)$. In the rest of the paper, $A$ denotes the forward model operator for multiple light sources, and $g$ represents the concatenation of multiple boundary measurements corresponding to the light sources. It is shown in [100] that the solution of (66) is non-unique. $A$ is the composition of the Sobolev trace operator and the solution operator for elliptic equations (which are DA equations in this case), it is a compact operator which has very large condition number after discretization. Therefore, (66) is severely ill-conditioned. In the following, we apply OSKCA method described in Chapter 2 to solve (66).

## *3.2 An analysis of the orthogonal solution*

In order to compute the orthogonal solution defined in (10), we need to clarify the notion of the adjoint operator $A^*$ of $A$, and verify the existence of inverse of $AA^*$.

In what follows, we denote the inner product in Hilbert space $H$ by $\langle \cdot, \cdot \rangle_H$, and for a linear functional $F$ on $H$. For topological vector spaces $\mathcal{X}$ and $\mathcal{Y}$, $\mathcal{B}(\mathcal{X}, \mathcal{Y})$ will denote the collection of all bounded linear mappings of $\mathcal{X}$ into $\mathcal{Y}$.

In the definition of $A$ in (67), $F_m^{-1} \in \mathcal{B}(L^2(\Omega), H^1(\Omega))$[49], $T_r \in \mathcal{B}(H^1(\Omega), L^2(\partial\Omega))$ [42], and the composed operator $T_r F_m^{-1} \in \mathcal{B}(L^2(\Omega), L^2(\partial\Omega))$. If $F_m^{-1*}$ and $T_r^*$ are adjoint operators of $F_m^{-1}$ and $T_r$ respectively, then the adjoint operator of $T_r F_m^{-1}$, denoted by $F_m^{-1*} T_r^*$, is in $\mathcal{B}(L^2(\partial\Omega), L^2(\Omega))$, which satisfies

$$\langle T_r F_m^{-1} f, y \rangle_{L^2(\partial\Omega)} = \langle f, F_m^{-1*} T_r^* y \rangle_{L^2(\Omega)} \tag{69}$$

for all $f \in L^2(\Omega)$ and $y \in L^2(\partial\Omega)$.

We have the following characterization of $T_r^*$ and $F_m^{-1*}$.

**Lemma 3.2.1** *For the Sobolev trace operator $T_r : H^1(\Omega) \to L^2(\partial\Omega)$, its adjoint $T_r^* \in \mathcal{B}(L^2(\partial\Omega), H^1(\Omega))$, which satisfies*

$$\langle z, T_r^* y\rangle_{H^1(\Omega)} = \langle T_r z, y\rangle_{L^2(\Omega)} \tag{70}$$

*for any pair $(z, y) \in H^1(\Omega) \times L^2(\partial\Omega)$. Specifically, for any $y \in H^{1/2}(\partial\Omega)$, $v = T^* y$ solves the PDE:*

$$\begin{cases} v - \Delta v = 0 & \text{in } \Omega \\ \frac{\partial v}{\partial n} = y & \text{on } \partial\Omega \end{cases} \tag{71}$$

**Proof** We show that $T_r^* y$ solves (71) for any given $y \in H^{1/2}(\partial\Omega)$. Multiply both sides of (71) by a test function $z \in H^1(\Omega)$ and integrate over $\Omega$, we have

$$\begin{aligned} 0 &= \int_\Omega (vz + \nabla v \cdot \nabla z)\,\mathrm{d}x - \int_{\partial\Omega} \tfrac{\partial v}{\partial n} z\,\mathrm{d}\Gamma \\ &\qquad \text{(using boundary condition)} \\ &= \int_\Omega (vz + \nabla v \cdot \nabla z)\,\mathrm{d}x - \int_{\partial\Omega} yz\,\mathrm{d}\Gamma, \end{aligned} \tag{72}$$

which is nothing but

$$\langle z, v\rangle_{H^1(\Omega)} = \langle T_r z, y\rangle_{L^2(\partial\Omega)}. \tag{73}$$

The above equality holds for all $z \in H^1(\Omega)$, and by the uniqueness of $T_r^* y$ we have $v = T_r^* y$. So $T_r \in \mathcal{B}(H^{1/2}(\partial\Omega), H^1(\Omega))$. Since $H^{1/2}(\partial\Omega)$ is dense in $L^2(\partial\Omega)$, $T_r$ can be uniquely extended to a bounded linear operator from $L^2(\partial\Omega)$ to $H^1(\Omega)$, which is still denoted by $T_r$.

**Lemma 3.2.2** *For the solution operator $F_m^{-1} : L^2(\Omega) \to H^1(\Omega)$ for (46), its adjoint operator $F_m^{-1^*} \in \mathcal{B}(H^1(\Omega), L^2(\Omega))$, and it satisfies*

$$\langle F_m^{-1} f, v\rangle = \langle f, F_m^{-1^*} v\rangle \tag{74}$$

*for any pair $(f, v) \in L^2(\Omega) \times H^1(\Omega)$. In particular, if $v \in H^2(\Omega)$, then*

$$w = F_m^{-1^*} v \tag{75}$$

*is the solution to the PDE:*

$$\begin{cases} -\nabla\cdot(\kappa_m\nabla w)+\mu_{am}w = v-\Delta v & \text{in } \Omega, \\ \kappa_m\frac{\partial w}{\partial \vec{n}}+Rw=\frac{\partial v}{\partial n} & \text{on } \partial\Omega. \end{cases} \tag{76}$$

**Proof** Multiply both sides of (43) by $w$ and integrate over $\Omega$,

$$\begin{aligned} \textstyle\int_\Omega f w\,dx &= \textstyle\int_\Omega \kappa_m\nabla u\cdot\nabla w\,\mathrm{d}x - \int_{\partial\Omega}\kappa\frac{\partial u}{\partial n}w\,\mathrm{d}\Gamma + \int_\Omega \mu_{am}uw\,\mathrm{d}x \\ &\quad \text{(integration by parts)} \\ &= \textstyle\int_\Omega\big(-u\nabla\cdot(\kappa_m\nabla w)+\mu_{am}uw\big)\,\mathrm{d}x + \int_{\partial\Omega}\big(u\kappa_m\frac{\partial w}{\partial n}-\kappa_m\frac{\partial u}{\partial n}w\big)\,\mathrm{d}\Gamma \\ &\quad \text{(boundary condition (44))} \\ &= \textstyle\int_\Omega\big(-u\nabla\cdot(\kappa_m\nabla w)+\mu_{am}uw\big)\,\mathrm{d}x + \int_{\partial\Omega}\big(u\kappa_m\frac{\partial w}{\partial n}+Ruw\big)\,\mathrm{d}\Gamma. \end{aligned} \tag{77}$$

The we use (76) and its boundary condition,

$$\begin{aligned} \textstyle\int_\Omega f w\,dx &= \textstyle\int_\Omega u(v-\Delta v)\,\mathrm{d}x + \int_{\partial\Omega}u\frac{\partial v}{\partial n}\mathrm{d}\Gamma \\ &= \textstyle\int_\Omega(uv+\nabla u\cdot\nabla v)\,\mathrm{d}x \end{aligned} \tag{78}$$

which is exactly $\langle f,w\rangle_{L^2(\Omega)}=\langle u,v\rangle_{H^1(\Omega)}$, i.e., $\langle F_m^{-1}f,v\rangle=\langle f,F_m^{-1*}v\rangle$. By definition $u=F_m^{-1}f$, and by the uniqueness of $F_m^{-1*}v$ we conclude $w=F_m^{-1*}v$. Since $H^2(\Omega)$ is dense in $H^1(\Omega)$, we can continuously extend $F_m^{-1*}$ to be a bounded linear operator from $H^1(\Omega)$ to $L^2(\Omega)$.

With the understanding of above two lemmas, we have $F_m^{-1*}T_r^*$ characterized in the following theorem:

**Theorem 3.2.3** *$F_m^{-1*}T_r^*$ is a bounded linear operator from $L^2(\partial\Omega)$ to $L^2(\Omega)$. For any $y\in H^{1/2}(\partial\Omega)$, $w=F_m^{-1*}T_r^*y$ is the solution to the following PDE:*

$$\begin{cases} -\nabla\cdot(\kappa_m\nabla w)+\mu_{am}w = 0 & \text{in } \Omega \\ \kappa_m\frac{\partial w}{\partial \vec{n}}+Rw=y & \text{on } \partial\Omega \end{cases} \tag{79}$$

**Proof** It is proved by combining two PDE's in Lemmas 3.2.1 and 3.2.2.

With $F_m^{-1*}T_r^*$ characterized in Theorem 3.2.3, we have the following theorem for the computation of $A^*$.

**Theorem 3.2.4** *$A^*$ is a bounded linear operator mapping from $\otimes^s L^2(\partial\Omega)$ to $L^2(\Omega)$. In particular, for $y^{(i)} \in L^2(\partial\Omega) \quad (i = 1, \ldots, s)$,*

$$A^* \begin{bmatrix} y^{(1)} \\ \vdots \\ y^{(s)} \end{bmatrix} = \sum_{i=1}^{s} \Phi_x^{(i)} F_m^{-1*} T_r^* y^{(i)}, \tag{80}$$

*where $F_m^{-1*}T_r^*$ is characterized in Theorem 3.2.3.*

$AA^*$ has following properties:

**Proposition 2** *1. $AA^* \in \mathcal{B}(\otimes^s L^2(\partial\Omega))$ is a compact operator.*

*2. If $AA^*$ is invertible, then its inverse is unbounded.*

**Proof** 1. We note that the solution operator to for DA equation (43), which is strongly elliptic, is a compact operator, and the trace operator is also compact. Therefore both $A$ and $A*$ are compact, and $AA^*$ is compact.

2. Since $\dim(\otimes^s L^2(\partial\Omega)) = \infty$ and $AA^*$ is compact, $0 \in \sigma(AA^*)$. Therefore, the inverse of $AA^*$ is unbounded if it exists.

**Remark** This proposition tells us that we cannot use (10) to compute the orthogonal solution, since $(AA^*)^{-1}$ is unbounded if it exists. Instead, we iterative regularization methods, which is built on $A$ and $A^*$. Early termination of the iterations exhibits regularization effects [46]. It aims to approximate the orthogonal solution such that the solution depends continuously on the data $g$. It is discussed in more details in Section 3.3.3.

## *3.3 Implementation of OSKCA for FT*

### 3.3.1 Discretization of the forward and inverse problems

In what follows, the matrix form of linear operators are denoted by bold capital letters.

The forward problem involves solving DA equations (46) and (47), which are second order elliptic differential equations. They can be solved numerically by Finite Element Method (FEM) [23]. Let $n_p$ be the number of nodes in the mesh for FEM. Suppose $\{\delta_1, \ldots, \delta_{n_p}\}$ are the associated FEM basis functions, which are actually the point-wise basis functions for the solution space in FEM. Under this basis, $f$ is written as

$$f = \sum_{j=1}^{n_p} f_j \delta_j, \tag{81}$$

where $f_1, \ldots, f_{n_p}$ are values of $f$ at the mesh nodes.

Under the point-wise basis, the discrete form of the inverse problem is

$$\mathbf{g} = \mathbf{A}\mathbf{f}, \tag{82}$$

where $\mathbf{f} = [f_1, \ldots, f_{n_p}]^\top$. $\mathbf{g}$ is the discrete form of the boundary measurement $g$ as defined in (67). $\mathbf{A}$ is called the forward model matrix, which is the matrix form of the operator $A$ in (66). By (67)(81), we have

$$\mathbf{A} = \left[A\delta_1 \ldots A\delta_{n_p}\right] = \begin{bmatrix} T_r F_m^{-1}(\Phi_x^{(1)}\delta_1) & \ldots & T_r F_m^{-1}(\Phi_x^{(1)}\delta_{n_p}) \\ \vdots & \ddots & \vdots \\ T_r F_m^{-1}(\Phi_x^{(s)}\delta_1) & \ldots & T_r F_m^{-1}(\Phi_x^{(s)}\delta_{n_p}) \end{bmatrix}, \tag{83}$$

where $\delta_j$ and $\Phi_x^{(i)}$ are both $n_p$-vectors representing their point values at the mesh nodes, and $\Phi_x^{(i)}\delta_j$ is the point-wise product of the two vectors. $F_m^{-1}(\Phi_x^{(i)}\delta_j)$ is represented by an $n_p$-vector defined at the mesh nodes, which is the FEM solution to the DA equation defined in (47):

$$F_m u_{ij} = \Phi_x^{(i)}\delta_j, \tag{84}$$

where $u_{ij}$ is the unknown. $T_r$ is the discrete form of the Sobolev trace operator, which is an interpolation of an $n_p$-vector defined on the mesh nodes at $n_d$ detector locations. Then $T_r F_m^{-1}(\Phi_x^{(i)}\delta_j)$ is an $n_d \times 1$ vector. Therefore, the forward model matrix $\mathbf{A}$ as defined in (83)has the size $sd \times n_p$.

As discussed in Section 2.3, we can also choose some $L^2$-basis other than the point-wise functions to represent $f$. More precisely, we denote the new basis by $\{\xi_i\}$. We recall (14), and with a slight abuse of notation denote its truncation at the $n$-th term as

$$f = \sum_{i=1}^{n} c_i \xi_i. \tag{85}$$

Assuming $f$ to have certain regularity, we can choose $\{\xi_i\}$ to be an efficient basis such as the harmonic functions or wavelets, so that $n$ can be much smaller than $n_p$, while the accuracy of the representation is the same as (81).

By (17), the discrete form of the inverse problem is

$$\mathbf{g} = \mathbf{M}\mathbf{c}, \tag{86}$$

where $\mathbf{c} = [c_1, \ldots, c_n]^\top$, $\mathbf{g}$ is the same as in (82), and $\mathbf{M}$ is the forward model matrix corresponding to the new basis. Similar to (83), we write

$$\mathbf{M} = [A\xi_1 \ldots A\xi_n] = \begin{bmatrix} T_r F_m^{-1}(\Phi_x^{(1)}\xi_1) & \ldots & T_r F_m^{-1}(\Phi_x^{(1)}\xi_n) \\ \vdots & \ddots & \vdots \\ T_r F_m^{-1}(\Phi_x^{(s)}\xi_1) & \ldots & T_r F_m^{-1}(\Phi_x^{(s)}\xi_n) \end{bmatrix}, \tag{87}$$

where $\xi_j$ is represented by a vector of its point values at the mesh nodes, and each $\Phi_x^{(i)}\xi_j$ $(i = 1, \ldots, s; j = 1, \ldots, n)$ is the point-wise product of the two $n_p$-vectors. Each $F_m^{-1}(\Phi_x^{(i)}\xi_j)$ in (87) is still obtained by solving the corresponding DA equations using the same FEM solver as (84). $T_r$ has the same definition as in (83). $\mathbf{M}$ as defined above has the size $sn_d \times n$.

Here we use the spectral method to represent the solution $f$ to the inverse problem, while using FEM to solve the PDE's in the forward model. Although $f$ is represented

in the spectral domain, it is not involved in FEM, so there are no convolutions in the computation. This hybrid approach takes the advantages of the efficient representation of the solution by spectral method, and the flexibility of handling complicated domains by FEM.

The computation cost is also saved in this approach. We note that there are totally $sn$ PDE's to be solved to form $\mathbf{M}$, compared to $sn_p$ for $\mathbf{A}$. Therefore, by using efficient basis instead of point-wise basis, we may achieve a speedup of $\frac{n_p}{n}$ by solving proportionally fewer PDE's when forming the forward model matrix. Moreover, the dimension of the unknown in $f$ decreases from the number of mesh nodes $n_p$ to the number of chosen basis functions $n$, so the computational complexity for solving the inverse problem is reduced accordingly.

The linear transformation $B$ as defined in (15) has the discrete form as

$$\mathbf{B} = [\xi_1, \ldots, \xi_n], \tag{88}$$

which is an $n_p \times n$ matrix. Then by (16), we can recover the point-wise representation of $f$ by

$$\mathbf{f} = \mathbf{B}\mathbf{c}, \tag{89}$$

where $\mathbf{c}$ is the solution to (86). In the implementation, we use the formulation (86) for the discrete form of the inverse problem (66) and the final solution $f$ is expressed by (89).

### 3.3.2 The choice of bases

**Tensor product construction**. As it is described in (14), the solution $f$ is expanded under a basis $\{\xi_i\}_{i=1}^{\infty}$ in the underlying Hilbert space, which is usually taken to be $L^2(\Omega)$, where $\Omega$ is the domain. In the computation, it is truncated at the $n$-th term as in (85).

There are several candidates for $\{\xi_i\}$, such as the sinusoidal functions, orthogonal polynomials, and wavelets, to name a few. They are used extensively in spectral

methods for solving PDEs [98]. Recently, they also arise in solving stochastic differential equations [105, 69], and modeling stochastic phenomena [12, 9]. In the classical setup of spectral method for solving multi-dimensional PDE's, the solution is based on a tensor product construction in the spectral domain. More exactly, assuming $n_x, n_y, n_z$ grid points are used in each dimension, the tensor product spectral domain has $n = n_x n_y n_z$ variables.

**Fourier basis**. Based on the tensor product construction, we can choose Fourier basis of the form

$$\{\exp\{2\pi i(p\omega_x x + q\omega_y y + r\omega_z z)\}\}, \tag{90}$$

where $p, q, r \in \mathbb{N}$, and $\omega_x, \omega_y, \omega_z$ are constants which determine the period in each dimension. For the purpose of representing real valued functions, the Fourier basis can be equivalently formed as

$$\{\cos(2\pi(p\omega_x x + q\omega_y y + r\omega_z z)), \sin(2\pi(p\omega_x x + q\omega_y y + r\omega_z z))\}, \tag{91}$$

where $p, q, r \in \mathbb{N}$. For finite computation, the Fourier basis is truncated, with $|p| \leq n_x, |q| \leq n_y, |r| \leq n_z$. Also the duplicate functions need to be removed. e.g., only one of $\sin(2\pi(p\omega_x x + q\omega_y y + r\omega_z z))$ and $\sin(2\pi(-p\omega_x x - q\omega_y y - r\omega_z z)))$ is kept in the basis. After removing those duplicates, we have the Fourier basis (91) with the index set

$$\{p, q, r : 0 \leq p \leq n_x, |q| \leq n_y, |r| \leq n_z\}. \tag{92}$$

In the case of 2D, the truncated Fourier basis is in the form of

$$\{\exp\{2\pi i(p\omega_x x + q\omega_y y)\} : 0 \leq p \leq n_x, |q| \leq n_y\}. \tag{93}$$

**Wavelet basis**. Tensor product wavelet basis consists of tensor products of 1D scaling and wavelet functions for reconstruction in each dimension. More exactly, we construct tensor product wavelet basis on a 3D Cartesian grid. In each dimension, we treat each point-wise basis function as a low frequency wavelet basis function at level 0

(the finest level), and construct coarser level wavelet basis functions by the the wavelet reconstruction [17]. In each dimension, the space spanned by the low frequency wavelet basis functions at level $l$ are denoted by $\mathcal{V}^x_{-l}, \mathcal{V}^y_{-l}, \mathcal{V}^z_{-l}$ respectively, and the spaces spanned by the high frequency wavelet basis functions are $\mathcal{W}^x_{-l}, \mathcal{W}^y_{-l}, \mathcal{W}^z_{-l}$. Then the tensor product wavelet space at level $l$ is

$$\mathcal{V}_{-l} = \mathcal{V}^x_{-l} \otimes \mathcal{V}^y_{-l} \otimes \mathcal{V}^z_{-l}. \tag{94}$$

Similarly, the tensor product wavelet space at level $l-1$ is written as

$$\mathcal{V}_{-l+1} = \mathcal{V}^x_{-l+1} \otimes \mathcal{V}^y_{-l+1} \otimes \mathcal{V}^z_{-l+1}. \tag{95}$$

By wavelet decomposition,

$$\mathcal{V}^x_{-l+1} = \mathcal{V}^x_{-l} \oplus \mathcal{W}^x_{-l}, \quad \mathcal{V}^y_{-l+1} = \mathcal{V}^y_{-l} \oplus \mathcal{W}^y_{-l}, \quad \mathcal{V}^z_{-l+1} = \mathcal{V}^z_{-l} \oplus \mathcal{W}^z_{-l}, \tag{96}$$

so

$$\mathcal{V}_{-l} \subset \mathcal{V}_{-l+1}. \tag{97}$$

In general, we have a multi-resolution analysis as

$$\ldots \subset \mathcal{V}_{-l} \subset \mathcal{V}_{-l+1} \subset \ldots \subset \mathcal{V}_0, \tag{98}$$

where $\mathcal{V}_0$ is the space spanned by the point-wise basis functions. This observation is useful when we design a multi-level algorithm, which is discussed in Section 3.5. The construction of the basis for $\mathcal{V}_{-l+1}$ follows the decomposition

$$\mathcal{V}_{-l+1} = (\mathcal{V}^x_{-l} \oplus \mathcal{W}^x_{-l}) \otimes (\mathcal{V}^y_{-l} \oplus \mathcal{W}^y_{-l}) \otimes (\mathcal{V}^z_{-l} \oplus \mathcal{W}^z_{-l}). \tag{99}$$

That is, in $x$ dimension, we choose the set of wavelet basis functions for reconstruction at level $l$ as the basis. Other two dimensions are similar. Then the tensor products of the basis functions in all three dimensions form the basis for $\mathcal{V}_{-l+1}$.

**A discussion on the number of basis functions**. From FEM point of view, each basis function $\xi_i (i = 1, \ldots, n)$ is represented by its nodal values in the triangulation $\mathcal{T}^h$ of the domain $\Omega$, or more exactly, its linear interpolation in $\mathcal{T}^h$. Here $h$

characterizes the fineness of the mesh, which is defined as the smallest number such that [23]

$$\max\{\mathrm{diam}(T) : T \in \mathcal{T}^h\} \leq h\,\mathrm{diam}(\Omega). \tag{100}$$

The interpolation operator is denoted by $\mathcal{I}^h$. A natural questions arises as, how do we determine the place of truncation $n$, based on the mesh size and the accuracy requirements?

We consider DA equation (43) with source term $q = f$, which is written as

$$-\nabla \cdot (\kappa \nabla \Phi) + \mu_a \Phi = f \quad \text{in } \Omega. \tag{101}$$

The accompanying boundary condition is (44). Suppose we solve it using FEM with piece-wise linear, conformal finite elements defined over triangulation $\mathcal{T}^h$. The nodal variables are denoted by $\{x_1, \ldots, x_{n_p}\}$, and the FEM basis functions are piece-wise linear tent functions denoted by $\{\psi_1, \ldots, \psi_{n_p}\}$, where $n_p$ is the number of nodes in the triangulation. Let $V^h$ be the space spanned by the FEM basis functions of $\mathcal{T}^h$, then $V^h$ is the solution space for FEM in this setting, and $V^h \subset H^1(\Omega)$. By Galerkin method, the variational formulation of DA equation is written as

$$\int_{\partial\Omega} R\Phi\psi_i \,\mathrm{d}s + \int_{\Omega} (\kappa \nabla\Phi \cdot \nabla\psi_i + \mu_a \Phi\psi_i)\,\mathrm{d}x = \int_{\Omega} f\psi_i \,\mathrm{d}x, \quad i = 1, \ldots, n_p. \tag{102}$$

We denote

$$B[u, v] = \int_{\partial\Omega} Ruv \,\mathrm{d}s + \int_{\Omega} (\kappa \nabla u \cdot \nabla v + \mu_a uv)\,\mathrm{d}x, \tag{103}$$

which is a bi-linear functional on $H^1(\Omega) \times H^1(\Omega)$. Then (102) is denoted by the weak formulation

$$B[\Phi^h, \psi_i] = \langle f, \psi_i \rangle, \tag{104}$$

where $\langle \cdot, \cdot \rangle$ is the inner product in $L^2(\Omega)$, and $\Phi^h \in V^h$ is the solution to this weak formulation. Under mild assumptions on the parameters $\kappa, \mu_a, R$, we can prove that $B[\cdot, \cdot]$ satisfies the continuity and coercivity conditions as follows

**Theorem 3.3.1** *Assume $\kappa, \mu_a > \gamma > 0$ for some constant $\alpha$, and $\Omega$ is a polyhedral domain. Then for $u, v \in H^1(\Omega)$,*

$$|B[u,v]| \leq C(R,\kappa,\mu_a,\Omega)\|u\|_{H^1(\Omega)}\|v\|_{H^1(\Omega)}, \tag{105}$$

*where*

$$C(R,\kappa,\mu_a,\Omega) = \max(\sqrt{2}\|R\|_{L^\infty(\partial\Omega)} + \|\mu_a\|_{L^\infty(\Omega)}, \|\kappa\|_{L^\infty(\Omega)}), \tag{106}$$

*and*

$$|B[u,u]| \geq \alpha(\kappa,\mu_a,\Omega)\|u\|^2_{H^1(\Omega)}, \tag{107}$$

*where*

$$\alpha(\kappa,\mu_a,\Omega) = \min_{\Omega} \min(\kappa,\mu_a). \tag{108}$$

**Remark** The proof of this theorem is standard [49]. The assumption that $\Omega$ is polyhedral is natural in FEM: we approximate $\Omega$ by a polyhedron in the triangulation. Under this assumption, we obtain specific continuity constant $C$ and coercivity constant $\alpha$ for the bi-linear form $B[\cdot,\cdot]$.

Going back to the variational formulation (104), we note $\Phi^h$ is the solution to it in $V^h$, and $\Phi$ is the solution to the original equation (101). Then by Céa's Theorem [23], we have the error estimate of FEM solution for $\Phi$ as

$$\|\Phi - \Phi^h\|_{H^1(\Omega)} \leq \frac{C(R,\kappa,\mu_a,\Omega)}{\alpha(\kappa,\mu_a,\Omega)} \min_{v\in V^h} \|v - \Phi\|_{H^1(\Omega)}, \tag{109}$$

where $C, \alpha$ are defined in (106)(108) respectively. We note that

$$\operatorname*{argmin}_{v\in V^h} \|v - \Phi\|_{H^1(\Omega)} = \mathcal{I}^h\Phi, \tag{110}$$

where $\mathcal{I}^h\Phi$ is the interpolation of $\Phi$ in $V^h$.

Similar to (101), we consider (1) with $q = f_n$, which is written as

$$-\nabla\cdot(\kappa\nabla\Phi) + \mu_a\Phi = f_n \quad \text{in } \Omega. \tag{111}$$

Its boundary condition is also (44). Its variational formulation is

$$B[\Phi_n^h, \Psi_i] = \langle f_n, \Psi_i \rangle, \quad i = 1 \ldots, n_p, \tag{112}$$

where $\Phi_n^h$ is its solution in $V^h$. By comparing (104) and (112), we have

$$B[\Phi^h - \Phi_n^h, \Psi_i] = \langle f - f_n, \Psi_i \rangle. \tag{113}$$

Then by coercivity of $B$ and Cauchy-Schwarz inequality,

$$\|\Phi^h - \Phi_n^h\|_{H^1(\Omega)} \leq \frac{\|f - f_n\|_{L^2(\Omega)}}{\alpha(\kappa, \mu_a, \Omega)}, \tag{114}$$

which gives an estimate of the error for $\Psi$ caused by the truncation error of the source term $f$. We require that

$$\|\Phi^h - \Phi_n^h\|_{H^1(\Omega)} \leq \|\Phi - \Phi^h\|_{H^1(\Omega)} \tag{115}$$

which means the error for $\Phi$ caused by the truncation error of source term does not exceed that caused by FEM. By (109), a necessary condition for this is

$$C(R, \kappa, \mu_a, \Omega)\|\mathcal{I}^h\Phi - \Phi\|_{H^1(\Omega)} \geq \|f - f_n\|_{L^2(\Omega)}. \tag{116}$$

By the interpolation theory ([23], Theorem 4.4.20),

$$\|\mathcal{I}^h\Phi - \Phi\|_{H^1(\Omega)} \leq C_1 h^{m-1}\|\Phi\|_{H^m(\Omega)}, \tag{117}$$

where constant $C_1$ depends on $\mathcal{T}^h$ and $m$, if $\Phi \in H^m(\Omega)$. Here $m$ characterizes the regularity of the solution to DA equation (43). Under the condition that $\kappa \in C^1(\Omega), R, \mu_a \in L^\infty(\Omega)$, we have $\Phi \in H^m(\Omega)$ for $m \geq 2$ by the theory of interior regularity [49]. In particular for $m = 2$,

$$\|\Phi\|_{H^2(\Omega)} \leq C_2\|f\|_{L^2(\Omega)}, \tag{118}$$

where $C_2$ is a constant depending on $\Omega$ and the coefficients in the DA equation. Then by (116)(117)(118) we have

$$\frac{\|f - f_n\|_{L^2(\Omega)}}{\|f\|_{L^2(\Omega)}} \leq CC_1C_2h, \tag{119}$$

which gives a requirement for the truncation error.

Let $V_n$ be the solution space spanned by $\{\xi_i\}_{i=1}^n$. Based on the tensor product construction of $\{\xi_i\}$ in the spectral domain, a typical estimate of the approximation error of $f$ in $V_n$ is [93]

$$\min_{f_n \in V_n} \|f - f_n\|_{L^2(\Omega)} \leq C_4 n^{-r/d} \|f\|_{H^r(\Omega)}, \tag{120}$$

where $d$ is the dimension of the domain, $C_4$ is independent of $f$ and $n$. Here we assume $f \in H^r(\Omega), r > 1$. A sufficient condition for (119) is that

$$n \geq \left( \frac{\|f\|_{H^r(\Omega)}}{\|f\|_{L^2(\Omega)}} \frac{C_4}{C C_1 C_2 h} \right)^{d/r}. \tag{121}$$

In general, $\frac{\|f\|_{H^r(\Omega)}}{\|f\|_{L^2(\Omega)}}$ is not bounded. However, as $f$ characterizes the physical quantity (the distribution of fluorophores) which changes smoothly in the domain, we can often assume that $\frac{\|f\|_{H^r(\Omega)}}{\|f\|_{L^2(\Omega)}}$ is actually bounded for $r = 2$ in applications. Under this assumption, we have

$$n = \mathcal{O}(h^{-d/2}). \tag{122}$$

Then for each dimension in the spectral domain, the number of variables is of the order $\mathcal{O}(h^{-1/2})$.

### 3.3.3 Computation of the orthogonal solution

After previous necessary steps, we have set up the inverse problem (86). The orthogonal solution to (86) is given by (22) after discretization. As mentioned earlier, regularization is needed to compute the orthogonal solution. Various regularization methods are proposed to compute the minimal norm solution [46], which is the orthogonal solution in our problem. We present first the Landweber iteration that is used in inverse scattering problems [13]. Then we describe a more robust approach that we use to solve this problem.

**Landweber iteration** is a popular choice for linear as well as nonlinear inverse problems, which has the form

$$\mathbf{c}^{(k)} = \mathbf{c}^{(k-1)} + \omega \mathbf{M}^*(\mathbf{g} - \mathbf{M}\mathbf{c}^{(k-1)}), \tag{123}$$

where $0 < \omega < (\|\mathbf{M}^*\|_2 \|\mathbf{M}\|_2)^{-1}$ is a relaxation parameter which ensures that (123) defines a contraction mapping. By the Banach fixed-point theorem, if $g \in \mathcal{R}(\mathbf{M})$, then (123) produces a sequence $\{\mathbf{c}^{(k)}\}$ converging to a fixed-point, which turns out to be the orthogonal solution to (86).

Landweber iteration is not an effective method, which often exhibits slow convergence. When it is applied to the problem (86) in the FT applications, the convergence is slow due to the ill-conditioning of $\mathbf{M}$.

**Iterated Tikhonov regularization** is another well-known approach [82] to compute the orthogonal solution. It iteratively regularizes the current solution by applying Tikhonov regularization to the residual equation. The iteration scheme is as follows:

$$\mathbf{c}^{(0)} = 0, \quad \mathbf{r}^{(0)} = \mathbf{g}, \tag{124}$$

and for $k = 1, 2, \ldots$

$$\mathbf{c}^{(k)} = \mathbf{c}^{(k-1)} + \mathbf{M}^*(\mathbf{M}\mathbf{M}^* + h^2\mathbf{I})^{-1}\mathbf{r}^{(k-1)}, \quad \mathbf{r}^{(k)} = \mathbf{g} - \mathbf{M}\mathbf{c}^{(k)}, \tag{125}$$

where $h$ is the regularization parameter, and $\mathbf{I}$ is the identity matrix. There are two layers of iterations in (125). The outer iterations update $\mathbf{c}^{(k)}$ directly, where the number of iterations is usually small in practice. The first outer iteration is equivalent to the standard Tikhonov regularization, and a few more iterations can improve the accuracy of the solution. In each outer iteration, $(\mathbf{M}\mathbf{M}^* + h^2\mathbf{I})^{-1}$ is actually implemented by iterative methods such as GMRES or CG [60] that form the inner iterations, which converge linearly. Iterative Tikhonov regularization is considered as a preconditioned Landweber iteration, and in practice it needs less

number of iterations than Landweber iteration to reduce the error of approximation to the same amount [82]. It has the benefit of being robust against noise, which is inherited from Tikhonov regularization. We adopt it to compute the orthogonal solution defined in (10).

**Remark** The orthogonal solution serves as a particular solution to the inverse problem, which can be computed by a stabilized algorithm such as Lanweber iteration and Iterative Tihkonov regularization described above. Other particular solutions can also be used in place of the orthogonal solution. A very important requirement of the particular solution is that it must depend continuously on the data, so that it is robust against noise.

### 3.3.4 Computation of the kernel space

Let $\mathbf{K}$ be a matrix with columns an orthonormal basis for $\mathcal{N}(\mathbf{M})$. Suppose the size of $\mathbf{K}$ is $n \times m$, where $n$ is equal to the number of columns of $\mathbf{M}$. $n$ is always larger than $m$. We use $\mathbf{K}$ to represent the computed kernel space. However, it is unstable to compute $\mathbf{K}$ by solving $\mathcal{N}(\mathbf{M})$ directly. We note that each row of $\mathbf{M}$ represents the measurements generated by all basis functions at the location of one detector. Nearby detectors have almost the same measurements, so their corresponding rows of $\mathbf{M}$ are nearly identical, which makes $\mathbf{M}$ not (numerically) full rank in rows. The numerical rank is defined to be the number of singular values that are larger than machine epsilon, or more generally a prescribed small threshold. If large number of detectors are in use, the numerical rank of $\mathbf{M}$ is much smaller than the number of rows in $\mathbf{M}$. It is very unstable to compute $\mathcal{N}(\mathbf{M})$ directly if $\mathbf{M}$ is numerically rank deficient.

In order to handle the numerical stability issue, we find a low rank approximation of $\mathbf{M}$, denoted by $\tilde{\mathbf{M}}$, which has the same size ($sn_d \times n$) as $\mathbf{M}$. Then we take $\mathcal{N}(\tilde{\mathbf{M}})$ as the approximation to $\mathcal{N}(\mathbf{M})$.

To compute the low rank approximation of $\mathbf{M}$, we adopt a fast algorithm based on singular value thresholding [25]. It is formulated as the convex optimization problem

$$\tilde{\mathbf{M}} = \underset{\mathbf{X}\in\mathbb{R}^{sd\times n}}{\operatorname{argmin}} \tau\|\mathbf{X}\|_* + \frac{1}{2}\|\mathbf{M}-\mathbf{X}\|_F^2, \tag{126}$$

where $\|\cdot\|_*$ denotes the nuclear norm, which is the sum of singular values. $\|\cdot\|_F$ is the Frobenius norm. It is a convex relaxation of the combinatorial problem

$$\tilde{\mathbf{M}} = \underset{\mathbf{X}\in\mathbb{R}^{sd\times n}}{\operatorname{argmin}} \operatorname{rank}(\mathbf{X}) + \frac{\rho}{2}\|\mathbf{M}-\mathbf{X}\|_F^2, \tag{127}$$

which is intractable. In contrast, the problem (126) can be solved very efficiently. Suppose

$$\mathbf{M} = \mathbf{U}\Sigma\mathbf{V}^\top \tag{128}$$

is the singular value decomposition (SVD) of $\mathbf{M}$, where $\Sigma = \operatorname{diag}(\{\sigma_i\})$, and $\{\sigma_i\}$ are the singular values of $\mathbf{M}$. Define the soft-thresholding of singular values by

$$\mathcal{D}_\tau(\Sigma) = \operatorname{diag}(\{(\sigma_i-\tau)_+\}), \tag{129}$$

where

$$(\sigma_i-\tau)_+ = \max(\sigma_i-\tau, 0). \tag{130}$$

Then the problem (126) has the explicit solution

$$\tilde{\mathbf{M}} = \mathbf{U}\mathcal{D}_\tau(\Sigma)\mathbf{V}^\top = \tilde{\mathbf{U}}\tilde{\Sigma}\tilde{\mathbf{V}}^\top, \tag{131}$$

where $\tilde{\mathbf{V}}$ is the sub-matrix of $V$ whose columns correspond to the nonzero singular values of $\mathcal{D}_\tau(\Sigma)$. Let $\mathbf{K}$ be the $n\times m$ matrix whose columns are complement to $\tilde{\mathbf{V}}$ in $\mathbf{V}$, so $\tilde{\mathbf{M}}\mathbf{K} = 0$. Let $\operatorname{col}(\mathbf{K})$ be the column space of $\mathbf{K}$, which has dimension $m$. $\operatorname{col}(\mathbf{K})$ is considered to be a good approximation to $\mathcal{N}(\mathbf{M})$ if $\|\mathbf{M}\mathbf{K}\|_F$ is sufficiently small. The following theorem gives an estimate of $\|\mathbf{M}\mathbf{K}\|_F$.

**Theorem 3.3.2** *Suppose* $\mathbf{M}$ *has* $r$ *nonzero singular values, then* $\mathbf{K}$ *satisfies*

$$\frac{\|\mathbf{M}\mathbf{K}\|_F}{\|\mathbf{M}\|_F} \le \frac{\sqrt{\sum_{i=1}^r \min(\sigma_i,\tau)^2}}{\sqrt{\sum_{i=1}^r \sigma_i^2}}. \tag{132}$$

**Proof**

$$\begin{aligned}
\|\mathbf{MK}\|_F^2 &\leq \|(\mathbf{M}-\tilde{\mathbf{M}})\mathbf{K}\|_F^2 + \|\tilde{\mathbf{M}}\mathbf{K}\|_F^2 \\
&= \|(\mathbf{M}-\tilde{\mathbf{M}})\mathbf{K}\|_F^2 \quad (\tilde{\mathbf{M}}\mathbf{K}=0) \\
&\leq \|(\mathbf{M}-\tilde{\mathbf{M}})\|_F^2 \quad (\mathbf{K} \text{ has orthonormal columns}) \\
&\leq \sum_{i=1}^{r} \min(\sigma_i,\tau)^2 \text{ (by definition of } \tilde{\mathbf{M}} \text{ in ((131)).}
\end{aligned}$$

Note that

$$\|\mathbf{M}\|_F^2 = \sum_{i=1}^{r} \sigma_i^2, \tag{133}$$

the inequality in this theorem follows.

**Remark** In practice, we select $\tau$ to be

$$\tau = \epsilon\sqrt{\frac{1}{r}\sum_{i=1}^{r}\sigma_i^2} = \epsilon\sqrt{\frac{1}{r}\|\mathbf{M}\|_F^2}, \tag{134}$$

where $\epsilon$ is a small number. By Theorem 3.3.2, $\mathbf{K}$ satisfies

$$\frac{\|\mathbf{MK}\|_F}{\|\mathbf{M}\|_F} \leq \epsilon. \tag{135}$$

Then col($\mathbf{K}$) is a good approximation to $\mathcal{N}(\mathbf{M})$.

### 3.3.5 Kernel correction

We note that the algorithm for kernel correction depends on the specific regularization requirement. For some popular regularization techniques like $L^2$, $L^1$ and $TV$ minimization, people have developed fast algorithms, which can be used in OSKCA. Here we present two examples to illustrate this idea.

**The positivity constraint for the kernel correction.** One important regularity requirement of the solution is the positivity constraint. By (21), after discretization, the point-wise representation of $f$ is

$$\mathbf{f} = \mathbf{B}(\mathbf{c}_* + \mathbf{K}\boldsymbol{\lambda}). \tag{136}$$

Then the positivity constraint, which is a very important physical requirement for the solution [24], can be written as discrete form

$$\mathbf{B}(\mathbf{c}_* + \mathbf{K}\boldsymbol{\lambda}) \geq \mathbf{0}, \tag{137}$$

where $\mathbf{B} \in \mathbb{R}^{n_p \times n}$ is computed in (88), $\mathbf{c}_* \in \mathbb{R}^n$ is computed by (125), $\mathbf{K} \in \mathbb{R}^{n \times m}$ is given by the algorithm described in Section 3.3.4. $\boldsymbol{\lambda} \in \mathbb{R}^m$ is the unknown.

The feasible points of (137) may not be easy to find. The Algebraic Reconstruction Technique (ART) [68, 21] is a common algorithm to find one feasible point from any given initial point, by successively projecting the point onto the half-spaces defined by each line of inequality in (137). In the following, we apply ART to (137).

Denote $\mathbf{H} = -\mathbf{B}\mathbf{K}$ and $\mathbf{b} = \mathbf{B}\mathbf{c}_*$. For $\mathbf{H}\lambda \leq \mathbf{b}$, in $j$-th iteration $\boldsymbol{\lambda}$ is updated via

$$\boldsymbol{\lambda}_{j+1} = \begin{cases} \boldsymbol{\lambda}_j & \text{if } b_i \geq (\mathbf{h}_i, \boldsymbol{\lambda}_j) \\ \boldsymbol{\lambda}_j + \alpha_j \frac{b_i - (\mathbf{h}_i, \boldsymbol{\lambda}_j)}{\|\mathbf{h}_i\|^2} \mathbf{h}_i & \text{if } b_i < (\mathbf{h}_i, \boldsymbol{\lambda}_j) \end{cases} \tag{138}$$

where $(\cdot, \cdot)$ denotes the inner product. $\mathbf{h}_i$ is the $i$'th row of $\mathbf{H}$, $b_i$ is the $i$'th entry of $\mathbf{b}$, and $\alpha_j \in (0, 1)$ is preselected. In practice, the initial value for $\boldsymbol{\lambda}$ is the zero vector. The iterations are terminated when the change of $\lambda$ is smaller than a prescribed value.

Due to the noise in the measurements, (137) may be infeasible. In that case, the cyclic convergence of ART will happen [27].

**The TV minimization for the kernel correction**. TV minimization has been demonstrated to have edge-preserving property in image recovery [87]. This approach can be incorporated in our framework, which is addressed in (13). After discretization and change of basis (27), it can be proposed as the optimization problem

$$\underset{\boldsymbol{\lambda}}{\operatorname{argmin}} \ \|\mathbf{B}(\mathbf{c}_* + \mathbf{K}\boldsymbol{\lambda})\|_{TV} \quad \text{subject to} \quad \mathbf{B}(\mathbf{c}_* + \mathbf{K}\boldsymbol{\lambda}) \geq 0, \tag{139}$$

where the computation of $\mathbf{c}_*$ and $\mathbf{K}$ are described before. Inspired by the operator splitting technique [102], we introduce two auxiliary variables $\mathbf{f}$ and $\mathbf{w}$. $\mathbf{f}$ is defined in (136) and $\mathbf{w}$ is given by $\mathbf{w} = \mathbf{D}\mathbf{f}$ ($\mathbf{D}$ is the finite difference operator used to approximate the gradient). By introducing these two auxiliary variables, an equivalent

formulation of (139) is

$$\underset{\boldsymbol{\lambda},\mathbf{f}\geq 0,\mathbf{w}}{\operatorname{argmin}} \|\mathbf{w}\|_1 \quad \text{subject to} \begin{cases} \mathbf{f} = \mathbf{B}(\mathbf{c}_* + \mathbf{K}\boldsymbol{\lambda}) \\ \mathbf{w} = \mathbf{D}\mathbf{f} \end{cases} \tag{140}$$

Let $\mathbf{w}_i$ be the value of $\mathbf{w}$ at $i$-th node, then $\|\mathbf{w}\|_1 = \sum_i \|\mathbf{w}_i\|$.

The Augmented Lagrangian method uses the unconstrained objective for (140), which is written as [104]

$$\begin{aligned} \underset{\boldsymbol{\lambda},\mathbf{f}\geq 0,\mathbf{w},\boldsymbol{\mu}_1,\boldsymbol{\mu}_2}{\operatorname{argmin}} \quad & \alpha\|\mathbf{w}\|_1 + (\boldsymbol{\mu}_1, \mathbf{D}\mathbf{f} - \mathbf{w}) + \frac{\rho_1}{2}\|\mathbf{D}\mathbf{f} - \mathbf{w}\|^2 + \\ & (\boldsymbol{\mu}_2, \mathbf{f} - \mathbf{B}(\mathbf{c}_* + \mathbf{K}\boldsymbol{\lambda})) + \frac{\rho_2}{2}\|\mathbf{f} - \mathbf{B}(\mathbf{c}_* + \mathbf{K}\boldsymbol{\lambda})\|^2. \end{aligned} \tag{141}$$

Here $\alpha, \rho_1, \rho_2$ are the regularization parameters that are selected by the user, and $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$ are the Lagrange multipliers that are unknown. The unknowns are solved iteratively [48], which is described in Algorithm 3.3.1. If the iteration is terminated in $l$ steps by some criteria such as $\frac{\|\mathbf{f}^{(l)} - \mathbf{f}^{(l-1)}\|}{\|\mathbf{f}^{(l-1)}\|} \leq \epsilon_0$, where $\epsilon_0$ can be a small number, then $\mathbf{f} = \mathbf{f}^{(l)}$ is the final solution.

**Algorithm 3.3.1** OSKCA with TV minimization for kernel correction (OSCKA-TV)

**Input**: $\mathbf{B}, \mathbf{K}, \mathbf{c}_*, \alpha, \rho_1, \rho_2, \epsilon_0$
**Output**: $\mathbf{f}$
**Initialization**: $\mathbf{f}^{(0)} = \mathbf{0}, \boldsymbol{\lambda}^{(0)} = \mathbf{0}, \boldsymbol{\mu}_1^{(0)} = \mathbf{0}, \boldsymbol{\mu}_2^{(0)} = \mathbf{0}$
**while** $\frac{\|\mathbf{f}^{(k)} - \mathbf{f}^{(k-1)}\|}{\|\mathbf{f}^{(k-1)}\|} \leq \epsilon_0$ **do**

1. $\mathbf{w}^{(k+1)} = \underset{\mathbf{w}}{\operatorname{argmin}}\ \alpha\|\mathbf{w}\|_1 + \frac{\rho_1}{2}\|\mathbf{D}\mathbf{f}^{(k)} - \mathbf{w} + \frac{\boldsymbol{\mu}_1^{(k)}}{\rho_1}\|^2$.
2. $\mathbf{f}^{(k+1)} = \underset{\mathbf{f}\geq 0}{\operatorname{argmin}}\ \frac{\rho_1}{2}\|\mathbf{D}\mathbf{f} - \mathbf{w}^{(k+1)} + \frac{\boldsymbol{\mu}_1^{(k)}}{\rho_1}\|^2 + \frac{\rho_2}{2}\|\mathbf{f} - \mathbf{B}(\mathbf{c}_* + \mathbf{K}\boldsymbol{\lambda}) + \frac{\boldsymbol{\mu}_2^{(k)}}{\rho_2}\|^2$.
3. $\boldsymbol{\lambda}^{(k+1)} = \underset{\boldsymbol{\lambda}}{\operatorname{argmin}}\ \|\mathbf{f}^{(k+1)} - \mathbf{B}(\mathbf{c}_* + \mathbf{K}\boldsymbol{\lambda}) + \frac{\boldsymbol{\mu}_2^{(k)}}{\rho_2}\|^2$.
4. $\boldsymbol{\mu}_1^{(k+1)} = \boldsymbol{\mu}_1^{(k)} + \rho_1(\mathbf{D}\mathbf{f}^{(k+1)} - \mathbf{w}^{(k+1)})$.
5. $\boldsymbol{\mu}_2^{(k+1)} = \boldsymbol{\mu}_2^{(k)} + \rho_2(\mathbf{f}^{(k+1)} - \mathbf{B}(\mathbf{c}_* + \mathbf{K}\boldsymbol{\lambda}^{(k+1)}))$.

**end while**

We note that in each iteration of Algorithm 3.3.1, Step 1 is solved by soft-thresholding [58]. Step 2 is a constraint quadratic program, which can be solved

by Projected Barzilai-Borwein (PBB) method [38, 32]. It is an iterative method based on gradient projection. The computation cost for the gradient is dominated by the matrix-vector multiplication with the $n_p \times n_p$ matrix

$$\frac{\rho_1}{\rho_2}\mathbf{D}^\top\mathbf{D} + \mathbf{I}. \tag{142}$$

It can be computed in $O(n_p)$ time, since $\mathbf{D}^\top\mathbf{D}$ is a discrete Laplacian matrix with only $O(n_p)$ nonzero entries. PBB is shown to have $R$-linear convergence [37]. Therefore the overall cost for Step 2 is $O(n_p)$. Step 3 is a Least-Squares (LS) problems, where the variable $\boldsymbol{\lambda}$ is an $m$-vector. $m < n \ll n_p$, so the computation cost is small. Moreover, if an orthonormal basis is chosen to represent the solution, then by (88), $\mathbf{B}$ is orthogonal. We also note that $\mathbf{K}$ has orthonormal columns, so $\mathbf{BK}$ is an orthogonal matrix as well. Then this step has the explicit solution

$$\boldsymbol{\lambda}^{(k+1)} = (\mathbf{BK})^\top(\mathbf{f}^{(k+1)} - \mathbf{B}\mathbf{c}_* + \frac{\boldsymbol{\mu}_2^{(k)}}{\rho_2}), \tag{143}$$

which is very cheap to solve. Step 4 and 5 are simple matrix-vector computations, and the computation cost is small.

**Comparison with direct application of Augmented Lagrangian method.** As a comparison, we also apply Augmented Lagrangian method directly to the problem

$$\hat{f} = \underset{f\geq 0}{\operatorname{argmin}} \|f\|_{TV} \quad \text{such that} \quad Af = g, \tag{144}$$

which is a special case of (9). After discretization and change of basis, it is proposed as the optimization problem

$$\underset{\mathbf{c}}{\operatorname{argmin}} \|\mathbf{Bc}\|_{TV} \quad \text{subject to } \mathbf{Bc} \geq \mathbf{0} \text{ and } \mathbf{Mc} = \mathbf{g}, \tag{145}$$

where the final solution is given by $\mathbf{f} = \mathbf{Bc}$. By introducing another auxiliary variable $\mathbf{w} = \mathbf{Df}$, (145) has the un-constraint formulation as

$$\begin{aligned}\underset{\mathbf{w},\mathbf{f}\geq\mathbf{0},\mathbf{c},\boldsymbol{\mu_1},\boldsymbol{\mu_2},\boldsymbol{\mu_3}}{\operatorname{argmin}} \quad & \gamma\|\mathbf{w}\|_1 + (\boldsymbol{\mu}_1, \mathbf{Df} - \mathbf{w}) + \frac{\beta_1}{2}\|\mathbf{Df} - \mathbf{w}\|^2 + (\boldsymbol{\mu}_2, \mathbf{Bc} - \mathbf{f}) + \frac{\beta_2}{2}\|\mathbf{Bc} - \mathbf{f}\|^2 \\ & + (\boldsymbol{\mu}_3, \mathbf{Mc} - \mathbf{g}) + \frac{\beta_3}{2}\|\mathbf{Mc} - \mathbf{g}\|^2,\end{aligned} \tag{146}$$

which can be solved in the same way as Algorithm 3.3.1. $\mathbf{w}, \mathbf{f}$ are updated by the same formulas as in 3.3.1. The major difference is that in (146), $\mathbf{c} \in \mathbb{R}^n$ instead of $\boldsymbol{\lambda} \in \mathbb{R}^m$ is updated in each iteration, which has the formulation

$$\mathbf{c}^{(k+1)} = \underset{\mathbf{c}}{\operatorname{argmin}} \frac{\beta_2}{2}\|\mathbf{B}\mathbf{c} - \mathbf{f}^{(k+1)} + \frac{\boldsymbol{\mu}_2^{(k)}}{\beta_2}\|^2 + \frac{\beta_3}{2}\|\mathbf{M}\mathbf{c} - \mathbf{g} + \frac{\boldsymbol{\mu}_3^{(k)}}{\beta_3}\|^2. \tag{147}$$

It is another LS problem similar to Step 3 in Algorithm 3.3.1, but the unknown is larger in size. The system matrix in the normal equation for (147) is

$$\frac{\beta_2}{\beta_3}\mathbf{B}^\top\mathbf{B} + \mathbf{M}^\top\mathbf{M}. \tag{148}$$

It is an $n \times n$ dense matrix, and cannot be inverted by fast transforms to our knowledge. Compared to formula (143) in Algorithm 3.3.1, solving this LS problem is more computationally involved. In fact, formula (143) has only one matrix vector multiplication with time complexity $O(mn_p)$. For (147), a matrix vector multiplication is needed to form the normal equation. In addition, another cost of $O(n^2)$ is needed for solving the normal equation if iterative method is used, so the total cost is $O(nn_p + n^2)$. Here we note the comparison $m < n \ll n_p$. Besides having more computation cost for each iteration, the direct Augmented Lagrangian method converges slower, which is demonstrated in one of the numerical studies in Section 3.6.

## *3.4 Some practical issues*

### 3.4.1 The handling of large amount of boundary measurements

In the discrete form of the inverse problem for FT (86), the system matrix $\mathbf{M}$ has the size $sn_d \times n$, where $n$ is the number of basis function to represent the solution $f$, $s$ is the number of light sources in use, and $n_d$ is the number of collected boundary measurements. Because the original inverse problem is severely under-determined, $\mathbf{M}$ would represent an under-determined linear system. However, in practice, it may turn out to be that $sn_d > n$, when multiple light sources and large number of data points are used, so that the total number of measurements exceeds that of the unknowns.

In this case, the system matrix $\mathbf{M}$ and measurement data $\mathbf{g}$ must be pre-processed before putting into the inverse solver.

One simple strategy is to discard some small measurement data below the chosen threshold, and also eliminate the corresponding entries in the columns of the system matrix. This method is valid, as we can prove below, the intensity of the fluorophore radiation follows exponential decay pattern, and the largest and smallest boundary values can differ several orders of magnitude. Because the maximum range of the measurements can be no more than 4 orders of magnitude, so some small measurements are more likely to be contaminated by the measurement noise. However, blind thresholding can cause lost of information. As we will see later, small measurements near the light source implies important information on the distribution of the fluorophore.

In order to analyze the thresholding procedure in a more quantitative way, we consider a semi-infinite domain, which can model the slab shaped phantom that is often used in transillumination FT applications [15]. More exactly, we assume $\Omega = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : 0 \leq x_3 \leq l\}$, where $l$ is the thickness of the slab. Let $\Phi^{(i)}$ be the incidence field induced by the $i$-th light source, which is the solution to the DA equation (47). The source term $q = q_i$ is modeled as a point source

$$q_i(x) = \delta(x - \xi_i), \tag{149}$$

where $\xi_i$ is located at a depth of one transport scattering distance $(1/\mu_s')$ [6] below the surface $\pi_l = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_3 = l\}$. Let $G(x, y)$ be the Green's function [63] for the region $\Omega$, then

$$\Phi_x^{(i)}(x) = G(x, \xi_i). \tag{150}$$

**Theorem 3.4.1** *For semi-infinite domain $\Omega$ with slab geometry,*

$$G(x, \xi_i) \leq \Phi(x - \xi_i) + \Phi(x - \tilde{\xi}_i) \tag{151}$$

*where $\tilde{\xi}_i$ is the reflection of $\xi_i$ in the plane $\pi_0 = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_3 = 0\}$, and*

$$\Phi(r) = \frac{1}{4\pi\kappa} \frac{e^{-\mu_e \|r\|}}{\|r\|}$$

*is the Green's function in infinite medium,* $\mu_e = \sqrt{\frac{\mu_a}{\kappa}} = \sqrt{3\mu_a(\mu_a + \mu_s')}$ *is called the effective attenuation coefficient.*

**Proof** To prove this, let $\phi_i(x) = \Phi(x - \xi_i)$. By standard computation, on $\partial\Omega$,

$$\kappa\frac{\partial\Phi(x-\xi_i)}{\partial\vec{n}} + R\Phi(x-\xi_i) = \frac{e^{-\mu_e\|x-\xi_i\|}}{4\pi\|x-\xi_i\|}\Big(\frac{R}{\kappa} - \Big(\mu_e + \frac{1}{\|x-\xi_i\|}\Big)\Big(\frac{x-\xi_i}{\|x-x_i\|}\cdot\vec{n}\Big)\Big). \quad (152)$$

By the facts that $R > 1$ [7],$\|x - \xi_i\| \geq \mu_s'$ for $x \in \partial\Omega$, and $\frac{x-\xi_i}{\|x-x_i\|} \cdot \vec{n} \leq 1$, we have

$$\kappa\frac{\partial\Phi(x-\xi_i)}{\partial\vec{n}} + R\Phi(x-\xi_i) \geq 0 \quad \text{on } \partial\Omega, \quad (153)$$

That is,

$$\kappa\frac{\partial\phi_i}{\partial\vec{n}} + R\phi_i \geq 0 \quad \text{on } \partial\Omega. \quad (154)$$

Therefore we have

$$\begin{cases} -\nabla\cdot\kappa\nabla\phi_i + \mu_a\phi_i = q_i & \text{in } \Omega \\ \kappa\frac{\partial\phi_i}{\partial\vec{n}} + R\phi_i \geq 0 & \text{on } \partial\Omega. \end{cases} \quad (155)$$

Denote $\tilde{\phi}_i(x) = \phi_i(x) - \Phi^{(i)}(x)$. $\tilde{\phi}_i$ solves

$$\begin{cases} -\nabla\cdot\kappa\nabla\tilde{\phi}_i + \mu_a\tilde{\phi}_i = 0 & \text{in } \Omega \\ \kappa\frac{\partial\tilde{\phi}_i}{\partial\vec{n}} + R\tilde{\phi}_i \geq 0 & \text{on } \partial\Omega \end{cases} \quad (156)$$

So $\tilde{\phi}_i \geq 0$ in $\Omega$. Therefore we have

$$0 < \Phi_x^{(i)}(x) \leq \phi_i(x) = \Phi(x - \xi_i). \quad (157)$$

Furthermore, let

$$\Omega_D = \{y \in \Omega : \text{dist}(y, \partial\Omega) \geq \mu_s'\} \quad (158)$$

be the region where DA equation is valid, then by (150),

$$G(x, y) \leq \Phi(x - y), \quad (159)$$

for $x \in \Omega, y \in \Omega_D$.

We assume that $f$ has its support on a bounded region $B_D$, and

$$\mathrm{supp}(f) \subset B_D \subset \Omega_D. \tag{160}$$

In the FT applications, $B_D$ is the space that the examined tissue occupies.

The fluorophore emission in the $i$-th excitation is $\Phi_m^{(i)}$, which is the solution to DA equation (46) with source term $\Phi_x^{(i)} f$. By (157), for $x \in \Omega_D$,

$$\begin{aligned}
\Phi_m^{(i)}(x) &= \int_\Omega G(x,y)\Phi_x^{(i)}(y) f(y)\,\mathrm{d}y \\
&< \int_\Omega \Phi(x-y)\Phi(y-\xi_i) f(y)\,\mathrm{d}y \\
&= \tfrac{1}{(4\pi\kappa)^2} \left| \int_{B_D} \tfrac{1}{\|x-y\|} \exp\left(-\mu_e\|x-y\|\right) \tfrac{1}{\|y-\xi_i\|} \exp\left(-\mu_e\|y-\xi_i\|\right) f(y)\,\mathrm{d}y \right| \\
&\le \tfrac{1}{(4\pi\kappa)^2} \exp\left(-\mu_e\|x-\xi_i\|\right) \left| \int_{B_D} \tfrac{1}{\|x-y\|\|y-\xi_i\|} f(y)\,\mathrm{d}y \right| \\
&\le \tfrac{1}{(4\pi\kappa)^2} \exp\left(-\mu_e\|x-\xi_i\|\right) \left( \int_{B_D} \tfrac{1}{\|x-y\|^2} |f(y)|\,\mathrm{d}y \cdot \int_{B_D} \tfrac{1}{\|\xi_i-y\|^2} |f(y)|\,\mathrm{d}y \right)^{1/2} \\
&\le C_7 \tfrac{1}{(4\pi\kappa)^2} \exp\left(-\mu_e\|x-\xi_i\|\right) \|f\|_\infty,
\end{aligned} \tag{161}$$

where

$$C_7 = \sup_{z\in\mathbb{R}^d} \int_{B_D} \frac{1}{\|z-y\|^2}\,\mathrm{d}y \tag{162}$$

is a constant depending only on $B_D$. (161) tells us that the fluorophore emission roughly decays exponentially as the distance to the light source $\|x-\xi_i\|$ increases. We denote the intensity profile function for the point light source located at $\xi_i$ by

$$p_i(x) = \exp\big(-\mu_e\|x-\xi_i\|\big), \tag{163}$$

which holds for $x \in \Omega_D$. For $x$ on the measurement plane, $p_i(x)$ attains its maximum value

$$p_i(\xi_i') = \exp\big(-\mu_e\|\xi_i'-\xi_i\|\big) \tag{164}$$

at $\xi'$, which is the projection of $\xi_i$ onto the measurement plane. For any $x$ on the

measurement plane,

$$\frac{p_i(x)}{p_i(\xi_i')} = \exp\big(-\mu_e(\|x-\xi_i\| - \|\xi_i'-\xi_i\|)\big), \tag{165}$$

which characterizes the exponential decay property of the boundary measurement. Our criteria of discarding small measurements is based on the intensity profile function $p_i(x)$. We only keep the measurement at $x$ if it belongs to

$$\{x : \frac{p_i(x)}{p_i(\xi_i')} \geq \tau\} \tag{166}$$

for the selected threshold $\tau$, which in practice can be between 0.001 and 0.01. (166) is equivalent to

$$B_i = \left\{x : \|x-\xi_i\| \leq \|\xi_i'-\xi_i\| + \frac{1}{\mu_e}\log\frac{1}{\tau}\right\}. \tag{167}$$

The measurements outside of the ball $B_i$ are discarded.

We note that our thresholding strategy is not based on the value of the specific measurements, but depending on the locations of them. Thresholding based purely on the value can loss useful spacial information of the distribution $f$, which is demonstrated below.

A slightly different estimate of $\Phi_m^{(i)}(x)$ is as follows:

$$\begin{aligned}
\Phi_m^{(i)}(x) &= \tfrac{1}{(4\pi\kappa)^2}\int\limits_{B_D} \tfrac{1}{\|x-y\|}\exp\left(-\mu_e\|x-y\|\right)\tfrac{1}{\|y-\xi_i\|}\exp\left(-\mu_e\|y-\xi_i\|\right)f(y)\,\mathrm{d}y \\
&\leq \tfrac{\exp\left(-\frac{\mu_e}{2}\|x-\xi_i\|\right)}{(4\pi\kappa)^2}\int\limits_{B_D}\tfrac{1}{\|x-y\|\|y-\xi_i\|}\exp\left(-\tfrac{\mu_e}{2}\|x-y\| - \tfrac{\mu_e}{2}\|y-\xi_i\|\right)f(y)\,\mathrm{d}y \\
&\leq \tfrac{\exp\left(-\frac{\mu_e}{2}\|x-\xi_i\|\right)}{(4\pi\kappa)^2}\left(\int\limits_{B_D}\tfrac{\exp(-\mu_e\|x-y\|)}{\|x-y\|^2}|f(y)|\,\mathrm{d}y\cdot\int\limits_{B_D}\tfrac{\exp(-\mu_e\|\xi_i-y\|)}{\|\xi_i-y\|^2}|f(y)|\,\mathrm{d}y\right)^{1/2}.
\end{aligned} \tag{168}$$

By (65), the boundary measurement in the $i$-th light source is

$$g^{(i)} = T\Phi_m^{(i)}. \tag{169}$$

Therefore, for $x \in \partial\Omega$

$$\begin{aligned} & g^{(i)}(x) \\ \leq \ & \tfrac{1}{(4\pi D)^2} \exp\Big(-\tfrac{\mu_e\|x-\xi_i\|}{2}\Big) \left( \int\limits_{\Omega_D} \tfrac{\exp(-\mu_e\|x-y\|)}{\|x-y\|^2} |f(y)| \,\mathrm{d}y \cdot \int\limits_{\Omega_D} \tfrac{\exp(-\mu_e\|\xi_i-y\|)}{\|\xi_i-y\|^2} |f(y)| \,\mathrm{d}y \right)^{1/2}. \end{aligned} \tag{170}$$

In the above inequality, $\xi_i$ is fixed, so

$$\begin{aligned} & \int\limits_{\Omega_D} \tfrac{\exp(-\mu_e\|\xi_i-y\|)}{\|\xi_i-y\|^2} |f(y)| \,\mathrm{d}y \\ \leq \ & \|f\|_{L^\infty(\Omega)} \exp\Big( -\mu_e \mathrm{dist}\big(\xi_i, \mathrm{supp}(f)\big)\Big) \int\limits_{B_D} \tfrac{1}{\|\xi_i-y\|^2} \,\mathrm{d}y \\ \leq \ & C_7 \|f\|_{L^\infty(\Omega)} \exp\Big( -\mu_e \mathrm{dist}\big(\xi_i, \mathrm{supp}(f)\big)\Big). \end{aligned} \tag{171}$$

Similarly,

$$\begin{aligned} & \int\limits_{\Omega_D} \tfrac{\exp(-\mu_e\|x-y\|)}{\|x-y\|^2} |f(y)| \,\mathrm{d}y \\ \leq \ & C_7 \|f\|_{L^\infty(\Omega)} \exp\Big( -\mu_e \mathrm{dist}\big(x, \mathrm{supp}(f)\big)\Big). \end{aligned} \tag{172}$$

Summarizing above inequalities, we have an estimate for $g^{(i)}$ in the following theorem.

**Theorem 3.4.2** *Suppose the fluorophore distribution is bounded and supported on* $\mathrm{supp}(f)$*, then the boundary measurement of the i-th fluorophore radiation satisfies*

$$g^{(i)}(x) \leq C_7 \|f\|_{L^\infty(\Omega)} \exp\left( -\frac{\mu_e}{2}\Big(\|x-\xi_i\| + \mathrm{dist}\big(x, \mathrm{supp}(f)\big) + \mathrm{dist}\big(\xi_i, \mathrm{supp}(f)\big)\Big)\right) \tag{173}$$

*for some constant $C_7$ depending only on $B_D$.*

**Remark** In the FT applications, the fluorophore inclusions usually concentrate in small regions compared to the whole domain. It is justified to assume that the distribution of fluorophore $f$ is compactly supported on a set of small volume.

From (173) we can see that, when $\mathrm{dist}\big(x, \mathrm{supp}(f)\big) + \mathrm{dist}\big(\xi_i, \mathrm{supp}(f)\big)$ is large enough, even if $\|x-\xi_i\|$ is small, the measurement $g^{(i)}(x)$ can still be smaller than

any given threshold. In this case, important spacial information on $f$ is lost if the data $g^{(i)}(x)$ is discarded.

Based on the criteria (167), we do the "truncation" on $\mathbf{g}$, that is, data points that are outside of $B_i$ is cut off from $g^{(i)}$, and denote the remaining data by $g_t^{(i)}$. The truncated data $\mathbf{g_t}$ is the concatenation of $g_t^{(i)}$ for $i = 1, \ldots, s$. Correspondingly, $\mathbf{M}$ is truncated to $\mathbf{M_t}$. The resulting formulation for the inverse problem is

$$\mathbf{M_t c} = \mathbf{g_t}. \tag{174}$$

We may further compress the boundary data and the system matrix in the wavelet domain. This is because the measurements are boundary values of the the solutions to diffusion equations, which are very smooth, without sharp transitions. Therefore, compression in the wavelet domain is very efficient, while preserving most spacial and intensity information of the signal. However, our compression method is slightly different from conventional one. We keep all approximation coefficients, which corresponds to the low frequency components of the boundary measurements. The reason of keeping small approximation coefficients is the same as keeping small measurements, which is explained previously. We do the compression only on details coefficients that are high frequency components of the measurements. Compression on the approximation coefficient by thresholding is equivalent to throwing away small measurements in $\mathbf{g_t}$, which is shown to lose useful information.

We assume that the boundary measurements are taken from part of a plane, which is called the measurement plane. We take a 2D wavelet decomposition of the measurements $\mathbf{g_t}$ in 2D up to a certain level, then do a hard thresholding on the detail coefficients, while preserving all the approximation coefficients. After the thresholding, the index set of the remaining wavelet coefficients is denoted by $I$, and the associated sampling operator in the wavelet domain is denoted by $S_I$. Then the

boundary measurement in the wavelet domain after compression is denoted by

$$\mathbf{g}' = \begin{bmatrix} S_I W g_t^{(1)} \\ \vdots \\ S_I W g_t^{(s)} \end{bmatrix}, \tag{175}$$

where $W$ is the wavelet filter for decomposition. Correspondingly, we perform the wavelet decomposition on each column of $\mathbf{M_0}$, and apply the sampling operator $S_I$ to get a new system matrix $\mathbf{M}'$, and a new inverse problem

$$\mathbf{g}' = \mathbf{M}'\mathbf{c}. \tag{176}$$

Then (176) is solved by OSKCA as described previously.

**Remark** We note (167) and consider the set

$$B = \bigcup_{i=1}^{s} B_i. \tag{177}$$

We claim that our reconstruction of $f$ must be restricted to this set, which is called the region of interest. Outside of this region, the intensities of both the excitation radiation and fluorophore emissions are very weak under all light sources in use. Therefore, fluorophore emission that takes place outside of the region of interest can hardly be detected on the boundary. This is the physical limitation of FT, which cannot be overcome in our mathematical framework.

### 3.4.2 The correction on the optical parameters

In our modeling, we assume that in the DA equation (43) with boundary condition (44), all parameters are known. $R$ depends on the reflective index on the interface of tissue and air, which can be measured quite accurately. But diffusion and scattering coefficients are not easily obtained, because they are non-uniform in the medium. In practice, we assume these parameters are homogeneous inside specific regions of the tissue, such as lungs, heart, and bones. They are measured separately, and used

for similar scenarios without making changes. However, in the regions where the fluorophores are concentrated, the absorption coefficient is larger than normal. If the difference becomes significant, the modeling errors caused by it could lead to unreliable results. In the following we describe a strategy that can be used to correct perturbations in the absorption coefficient.

Recall that in the excitation model, the light source $q$ is given, so that with the prescribed parameters we can compute the excitation light field. Also, the true excitation light field can be measured on the surface, which is compared against the synthetic data. Let the prescribed absorption coefficient be $\mu_a$, and the DA equation is written as

$$\begin{cases} -\nabla \cdot (\kappa \nabla \tilde{u}) + \mu_a \tilde{u} = q \quad \text{in } \Omega, \\ \kappa \frac{\partial \tilde{u}}{\partial n} + R\tilde{u} = 0 \quad \text{on } \partial\Omega. \end{cases} \tag{178}$$

Let the true value be $\mu_a + \delta\mu_a$, where $\delta\mu_a$ models the small perturbation. Then we have the DA equation with true parameters

$$\begin{cases} -\nabla \cdot (\kappa \nabla u) + (\mu_a + \delta\mu_a) u = q \quad \text{in } \Omega, \\ \kappa \frac{\partial u}{\partial n} + Ru = 0 \quad \text{on } \partial\Omega. \end{cases} \tag{179}$$

The solution $\tilde{u}$ to (178) is denoted by

$$\tilde{u} = S(\mu_a, q), \tag{180}$$

and for fixed $q$, $S(\mu_a, q)$ is denoted in short as $S_q(\mu_a)$. Let $\delta u = u - \tilde{u}$, then $\delta u = \mathcal{O}(\delta\mu_a)$. By taking the difference of (179) and (178) we have

$$\begin{cases} -\nabla \cdot (\kappa \nabla \delta u) + \mu_a \delta u = -\delta\mu_a u \quad \text{in } \Omega, \\ \kappa \frac{\partial \delta u}{\partial n} + R\delta u = 0 \quad \text{on } \partial\Omega. \end{cases} \tag{181}$$

Recall that $T$ denotes the trace operator, we define the map $\Psi$ by

$$\Psi(\mu_a, q) = T\tilde{u}. \tag{182}$$

$\Psi$ is linear in $q$, and nonlinear in $\mu_a$. For fixed $q$, we denote $\Psi(\mu_a, q) = \Psi_q(\mu_a)$, and write (182) as

$$\Psi_q(\mu_a) = T\tilde{u} = TS_q(\mu_a). \tag{183}$$

Similarly by (179) we have

$$\Psi_q(\mu_a + \delta\mu_a) = Tu = TS_q(\mu_a + \delta\mu_a). \tag{184}$$

In (184), $Tu$ is the boundary measurement, $\mu_a$ is prescribed, and $\delta\mu_a$, which is the correction to $\mu_a$, is the unknown. In order to linearize (184), we need Fréchet derivative of $\Psi_q$, denoted by $D\Psi_q$, and its adjoint $D\Psi_q^*$. Following [4, 11] and noting (181), we have the following lemma.

**Lemma 3.4.3** *Let $\delta\mu_a \in L^2(\Omega)$, then*

$$\|S_q(\mu_a + \delta\mu_a) - S_q(\mu_a) - S(\mu_a, \tilde{u}\delta\mu_a)\|_{H^1(\Omega)} \leq C_5\|\delta\mu_a\|^2_{L^2(\Omega)}\|q\|_{L^2(\Omega)}, \tag{185}$$

*where $\tilde{u} = S_q(\mu_a)$, $C_5$ is a constant depending on $\Omega, \kappa, R, \mu_a$.*

By Lemma 3.4.3 and the definition of Fréchet derivative, $D\Psi_q(\mu_a) : L^2(\Omega) \to L^2(\partial\Omega)$ is defined by

$$D\Psi_q(\mu_a)\delta\mu_a = TS(\mu_a, \tilde{u}\delta\mu_a). \tag{186}$$

Here we use the fact that $T$ is bounded and linear, which is guaranteed by the trace theorem [1]:

**Lemma 3.4.4** *Assume $\Omega$ is a $C^{0,1}$ (Lipschitz) domain, then the trace operator*

$$T : H^1(\Omega) \to H^{1/2}(\partial\Omega) \tag{187}$$

*satisfies*

$$\|Tu\|_{L^2(\partial\Omega)} \leq C_0\|u\|_{L^2(\Omega)} \tag{188}$$

*for each $u \in H^1(\Omega)$, where $C_0$ depends only on $\Omega$.*

Note that $S(\cdot,\cdot)$ is linear in the second variable, so $S(\mu_a, \tilde{u}\delta\mu_a) = S(\mu_a, \tilde{u})\delta\mu_a$. Formally we write

$$D\Psi_q(\mu_a) = TS(\mu_a, \tilde{u}). \tag{189}$$

We need another lemma for the definition of $D\Psi^*$:

**Lemma 3.4.5** *For any $y \in H^{1/2}(\Omega)$, let $v \in H^1(\Omega)$ be the solution to*

$$\begin{cases} -\nabla \cdot (\kappa \nabla v) + \mu_a v = 0 \quad \text{in } \Omega, \\ \kappa \frac{\partial v}{\partial n} + Rv = y \quad \text{on } \partial\Omega. \end{cases} \tag{190}$$

*Then for any $\delta\mu_a \in L^2(\Omega)$,*

$$\langle D\Psi_q(\mu_a)\delta\mu_a, y\rangle_{L^2(\partial\Omega)} = \langle \delta\mu_a, \tilde{u}v\rangle_{L^2(\Omega)}. \tag{191}$$

*Moreover,*

$$\|\tilde{u}v\|_{H^1(\Omega)} \le C_6 \|q\|_{L^2(\Omega)} \|y\|_{L^2(\partial\Omega)}, \tag{192}$$

*where $C_6$ is a constant depending on $\Omega, \kappa, \mu_a, R$.*

By the above lemma, $D\Psi_q^*(\mu_a) : L^2(\partial\Omega) \to L^2(\Omega)$ is a bounded linear operator defined as

$$D\Psi_q^*(\mu_a)y = \tilde{u}v, \tag{193}$$

for any $y \in H^{1/2}(\partial\Omega)$, and

$$\|D\Psi_q^*(\mu_a)\| \le C_6 \|q\|_{L^2(\Omega)}. \tag{194}$$

Since $H^{1/2}(\Omega)$ is dense in $L^2(\partial\Omega)$, the definition of $D\Psi_q^*$ can be continuously extended to the domain $L^2(\partial\Omega)$, while preserving the boundedness as in (194) [88].

With the knowledge of $D\Psi_q$ and its adjoint $D\Psi_q^*$, we use Landweber iteration or Iterative Tikhonov regularization to solve (184) iteratively. For example, the $k$-th step Landweber iteration can be described as

$$\begin{cases} \delta\mu_a^{(k)} = \beta_k D\Psi_q^*(\mu_a^{(k)})(Tu - \Psi_q(\mu_a^{(k)})), \\ \mu_a^{(k+1)} = \mu_a^{(k)} + \delta\mu_a^{(k)}. \end{cases} \tag{195}$$

Here $\beta_k$ is the step size for $k$-th iteration, which can be a constant for each iteration. In the computation, $\Psi_q$ involves solving (179) once, and $D\Psi_q^*$ needs solving (190) once. So in each iteration, the computation cost is equivalent to solving DA equations twice. We note that this approach of tuning the parameter works only if the full boundary measurement can be obtained. However, it is not practical in FT applications, where only partial boundary is visible to the detectors. It is a challenging problem to our knowledge so far [72].

## *3.5 A multi-level framework based on adaptive wavelet basis*

The above algorithm (OSKCA) is applicable to any choice of basis (or frames) for the solution. Using tensor product wavelets for the basis allows a multi-level strategy, which is a useful way to reduce the computation cost. The framework is inherited from the multi-resolution analysis [94] in the wavelet theory. Based on the multi-resolution analysis as defined in (98), we propose a multi-level framework for the computation of the orthogonal solution and the kernel correction.

Unlike the usual "bottom-up" approach that is used in wavelet compression [30], our approach is mainly "top-down". In wavelet compression, the original signal is transformed to the wavelet domain, decomposed from fine level to coarse level, and then compressed by thresholding. In our approach, we reconstruct the solution from coarse level to fine level. First, a smaller size problem is solved in coarse level wavelet domain, then the solution is refined in fine level wavelet domain.

The idea can be illustrated by a two-level example. For simplicity of the notations, we describe the 2D version of it, while 3D version follows the same way. The coarse level and the fine level problems correspond to the wavelet decomposition at level 2 and level 1 respectively. The coarse level equation for the inverse problem (86) is denoted by

$$\mathbf{M}_2\mathbf{c}_2 = \mathbf{g}, \tag{196}$$

and the fine level equation is denoted by

$$\mathbf{M}_1\mathbf{c}_1 = \mathbf{g}. \tag{197}$$

We describe the construction of both equations in the following.

We first describe the construction of the coarse level equation. The key step is to select the wavelet basis functions. Not all wavelet basis functions are selected. Because the solution is restricted to the region of interest $B$ as defined in (177), we select level-2 wavelet basis functions that have support overlapping with $B$. In practice, we choose the smallest rectangular cuboid $Q$ that covers $B$, which is depicted in Figure 2(a)), and perform two levels of wavelet decomposition of an all-one function supported on $Q$. The resulting nonzero low frequency coefficients at level 2 are denoted by $\mathbf{c}_2^A$, and the coefficients having the same relative locations in the high frequency sub-bands are $\mathbf{c}_2^V, \mathbf{c}_2^H, \mathbf{c}_2^D$ (see Figure 2(b)). The concatenation of the selected low and high frequency coefficients form $\mathbf{c}_2 = [\mathbf{c}_2^A\ \mathbf{c}_2^V\ \mathbf{c}_2^H\ \mathbf{c}_2^D]$, which is the variable to solve in coarse level equation (196).

All level-2 wavelet basis functions corresponding to the wavelet coefficients $\mathbf{c}_2$ form the basis $\{\xi_1, \ldots, \xi_n\}$, which are used to form the system matrix $\mathbf{M}_2$, which is computed column-wise by the formula

$$\mathbf{M}_2 = [A\xi_1 \ldots A\xi_n], \tag{198}$$

where $A$ is the linear operator defined in (67).

The fine level equation is constructed adaptively based on the orthogonal solution to the coarse level equation (196), which is described in the two-level algorithm for the orthogonal solution next.

**Two-level approach for the orthogonal solution.** Let $\mathbf{c}_2^*$ be the orthogonal solution to (196), which is computed by the method in Section 3.3.3.

The construction of the system matrix $\mathbf{M}_1$ for the fine level equation (197) is by adaptively selecting basis functions based on $\mathbf{c}_2^*$. First, we apply one level of wavelet

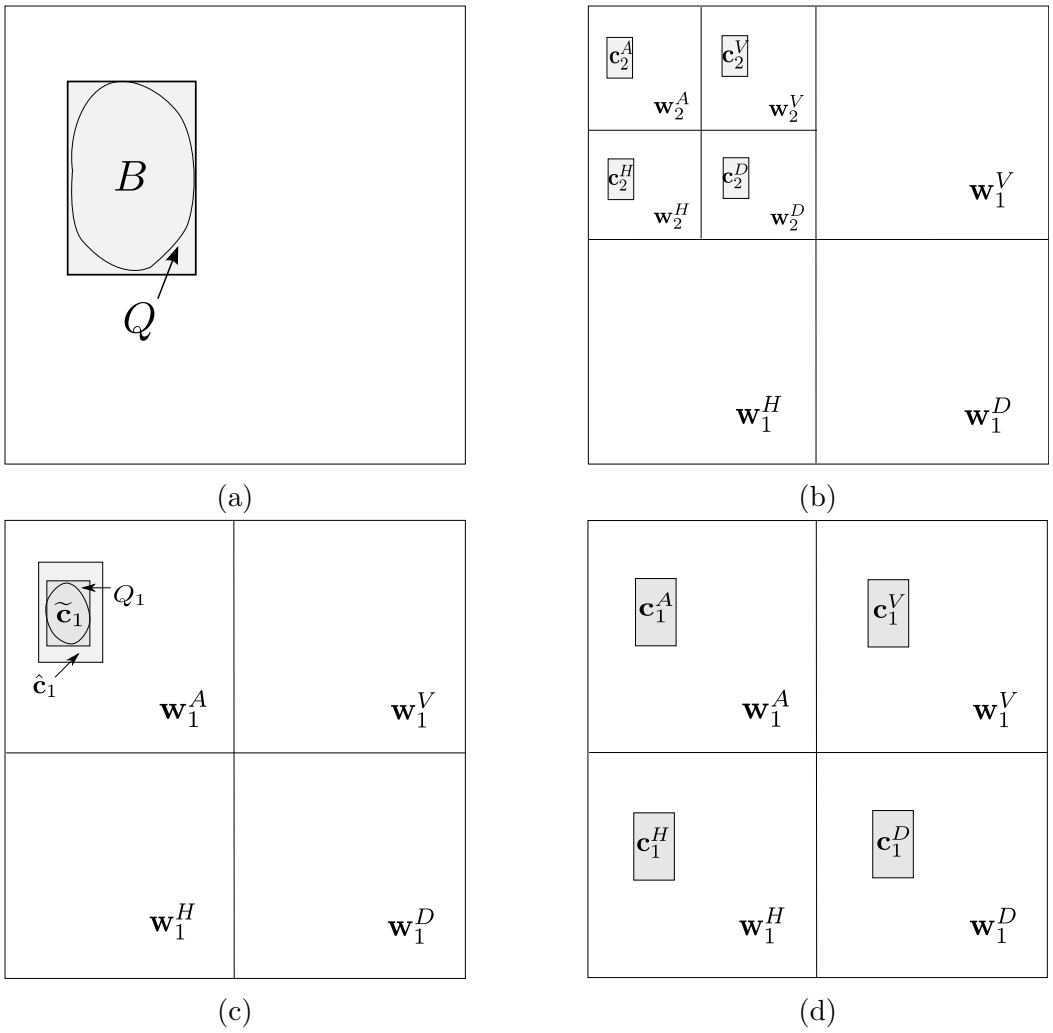



Figure 2: (a) Q is the smallest rectangular cuboid that covers the region of interest $B$. (b) The shaded area denotes the level-2 wavelet coefficients selected for coarse level equation. (c) $\hat{\mathbf{c}}_1$ is the result of one level of wavelet reconstruction of $\mathbf{c}_2$. After thresholding, the nonzeros are $\widetilde{\mathbf{c}}_1$. $Q_1$ is the smallest rectangular cuboid covering $\widetilde{\mathbf{c}}_1$. (d) The shaded area depicts the level-1 wavelet coefficients chosen for fine level equation. $\mathbf{w}_2^A$ is the low frequency sub-band at level 2, and $\mathbf{w}_2^V$, $\mathbf{w}_2^H$, $\mathbf{w}_2^D$ are the high frequency sub-bands at level 2. $\mathbf{w}_1^A$, $\mathbf{w}_1^V$, $\mathbf{w}_1^H$, $\mathbf{w}_1^D$ have similar meanings.

reconstruction $W_S$ to $\mathbf{c}_2^*$ and obtain level-1 wavelet coefficients

$$\hat{\mathbf{c}}_1^* = W_S \mathbf{c}_2^*. \tag{199}$$

We note that only low frequency coefficients in $\hat{\mathbf{c}}_1^*$ are nonzero. Then we do a hard thresholding on $\hat{\mathbf{c}}_1^*$, by keeping only the coefficients that are larger than $\epsilon \|\hat{\mathbf{c}}_1^*\|_\infty$, where $\epsilon$ can be between 0.001 and 0.01 in practice. The remaining coefficients are denoted by:

$$\widetilde{\mathbf{c}}_1^* = T_\epsilon \hat{\mathbf{c}}_1^*, \tag{200}$$

where $T_\epsilon$ is the thresholding operator. The purpose of the thresholding is to reduce the number of unknowns in the wavelet domain, and shrink the support of the solution.

We choose the smallest rectangular cuboid $Q_1$ that covers $\widetilde{\mathbf{c}}_1^*$ in the level-1 wavelet domain, which is depicted in Figure 2(c). We select coefficients $\mathbf{c}_1^A$ having the same support as $Q_1$ in the low frequency sub-band, and coefficients $\mathbf{c}_1^V,\ \mathbf{c}_1^H,\ \mathbf{c}_1^D$ having the same relative locations as $\mathbf{c}_1^A$ in the high frequency sub-bands. Then the variable $\mathbf{c}_1$, given by $\mathbf{c}_1 = [\mathbf{c}_1^A\ \mathbf{c}_1^V\ \mathbf{c}_1^H\ \mathbf{c}_1^D]$ is the variable to solve in the fine level equation. The wavelet basis functions corresponding to the wavelet coefficients $\mathbf{c}_1$ are used to compute $\mathbf{M}_1$, in the same way as (198).

$\widetilde{\mathbf{c}}_1^*$ as defined in (199)(200) is used as an approximation to the orthogonal solution for (197), which is refined by solving the residual equation

$$\mathbf{M}_1 \Delta \mathbf{c}_1 = \Delta \mathbf{g}_1, \tag{201}$$

where

$$\Delta \mathbf{g}_1 = \mathbf{g} - \mathbf{M}_2 \mathbf{c}_2^*. \tag{202}$$

The orthogonal solution to the residual equation (201) is denoted by $\Delta \mathbf{c}_1^*$. Then the orthogonal solution to fine level equation (197) is

$$\mathbf{c}_1^* = \widetilde{\mathbf{c}}_1^* + \Delta \mathbf{c}_1^*. \tag{203}$$

**Two-level approach for the kernel correction.** The two-level approach for the kernel correction is slightly different. Given the orthogonal solution $\mathbf{c}_1^*$ to (197), we first transform the solution to level 2 wavelet domain, and do the kernel correction at that level, then transform the result back to level 1, and do the kernel correction again.

The above procedure is described more precisely as follows. In the first step, we apply one level of wavelet decomposition to $\mathbf{c}_1^*$, and keep only the wavelet coefficients at level 2, which is written as

$$\bar{\mathbf{c}}_2^* = P_2 W_A \mathbf{c}_1^*, \tag{204}$$

where $W_A$ denotes wavelet decomposition, $P_2$ is the restriction of wavelet coefficients to level 2. We perform the kernel correction for $\bar{\mathbf{c}}_2^*$ in the null space of $\mathbf{M}_2$, as described in Section 3.3.5. The result is denoted by $\mathbf{c}_2^0$. By applying one level of wavelet reconstruction to $\mathbf{c}_2^0$, we obtain wavelet coefficients at level 1, which is denoted by

$$\bar{\mathbf{c}}_1^0 = W_S \mathbf{c}_2^0, \tag{205}$$

which is called the coarse level correction for $\mathbf{c}_1^*$. The solution to (197) is updated to be

$$\hat{\mathbf{c}}_1 = \mathbf{c}_1^* + \bar{\mathbf{c}}_1^0. \tag{206}$$

In the next step, which we call the fine level correction, we use $\hat{\mathbf{c}}_1$ as a particular solution to (197), and compute kernel correction for it in the null space of $\mathbf{M}_1$. The result is denoted by $\hat{\mathbf{c}}_1^0$. Combining the coarse and fine level corrections, the kernel correction for $\mathbf{c}_1^*$ is written as

$$\mathbf{c}_1^0 = \bar{\mathbf{c}}_1^0 + \hat{\mathbf{c}}_1^0. \tag{207}$$

The two level algorithm can be easily generalized to a multi-level version, which is summarized in Algorithms 3.5.1 and 3.5.2.

**Remark** For the computation of the system matrix $\mathbf{M}_{l-i}$ for $i = 1, \ldots, l-1$, we can make use of the already computed matrix $\mathbf{M}_{l-i+1}$ to form some columns of $\mathbf{M}_{l-i}$.

**Algorithm 3.5.1** Multi-level algorithm for the orthogonal solution (MLOS)

**function** $\mathbf{c}^*$=MLOS$(l, \mathbf{g})$ ▷ $l$ denotes the coarsest level.
  COMPUTE $\mathbf{M}_l$
  $\mathbf{c}_l^* = \text{OS}(\mathbf{M}_l, \mathbf{g})$ ▷ solved by the method in Section 3.3.3
  $\mathbf{g}_l = \mathbf{g}$
  **for** $i = 1$ to $l - 1$ **do**
    $\Delta\mathbf{g}_{l-i} = \mathbf{g}_{l-i+1} - \mathbf{M}_{l-i+1}\mathbf{c}_{l-i+1}^*$
    $\hat{\mathbf{c}}_{l-i}^* = W_S\mathbf{c}_{l-i+1}^*$ ▷ wavelet synthesis
    $\widetilde{\mathbf{c}}_{l-i}^* = \text{THRESHOLD}(\hat{\mathbf{c}}_{l-i}^*)$ ▷ hard thresholding on the solution
    COMPUTE $\mathbf{M}_{l-i}$
    $\Delta\mathbf{c}_{l-i}^* = \text{OS}(\mathbf{M}_{l-i}, \Delta\mathbf{g}_{l-i})$
    $\mathbf{c}_{l-i}^* = \bar{\mathbf{c}}_{l-i}^* + \Delta\mathbf{c}_{l-i}^*$
  **end for**
  $\mathbf{c}^* = \mathbf{c}_1^*$
  **return** $\mathbf{c}^*$
**end function**

**Algorithm 3.5.2** Multi-level algorithm for the kernel correction (MLKC)

**function** $\mathbf{c}^0$=MLKC$(l, \mathbf{c}_1^*)$ ▷ $l$ denotes the coarsest level.
  **if** $l = 1$ **then**
    $\mathbf{c}^0 = \text{KC}(\mathbf{M}_l, \bar{\mathbf{c}}_l^*)$ ▷ the coarsest level problem.
    **return** $\mathbf{c}^0$
  **else**
    $\bar{\mathbf{c}}_2^* = P_2 W_A^L \mathbf{c}_1^*$ ▷ wavelet analysis as defined in (204)
    $\mathbf{c}_2^0 = \text{MLKC}(l - 1, \bar{\mathbf{c}}_2^*)$ ▷ recursively call itself until the coarsest level
  **end if**
  $\hat{\mathbf{c}}_1^0 = W_S\mathbf{c}_2^0$
  $\hat{\mathbf{c}}_1 = \mathbf{c}_1^* + \hat{\mathbf{c}}_1^0$
  $\bar{\mathbf{c}}_1^0 = \text{KC}(\mathbf{M}_1, \hat{\mathbf{c}}_1)$ ▷ $\bar{\mathbf{c}}_1^0$ is the finest level correction.
  $\mathbf{c}^0 = \bar{\mathbf{c}}_1^0 + \hat{\mathbf{c}}_1^0$ ▷ total correction
  **return** $\mathbf{c}^0$
**end function**

Columns of $\mathbf{M}_{l-i}$ are generated by selected wavelet basis functions at level $l-i$. Recall that each low frequency wavelet basis function at level $l-i$ is a linear combination of wavelet basis functions at level $l-i+1$, and the coefficients are exactly the entries of the wavelet filter for reconstruction. Therefore, columns of $\mathbf{M}_{l-i}$ that are generated by low frequency wavelet functions at level $l-i$ can be obtained by applying one level of wavelet reconstruction to the rows of $\mathbf{M}_{l-i+1}$. Besides, the columns generated by selected high frequency wavelet functions at level $l-i$ have to be computed separately by solving DA equations and assembling the solutions in the form of (87).

## *3.6 Numerical examples with synthetic data*

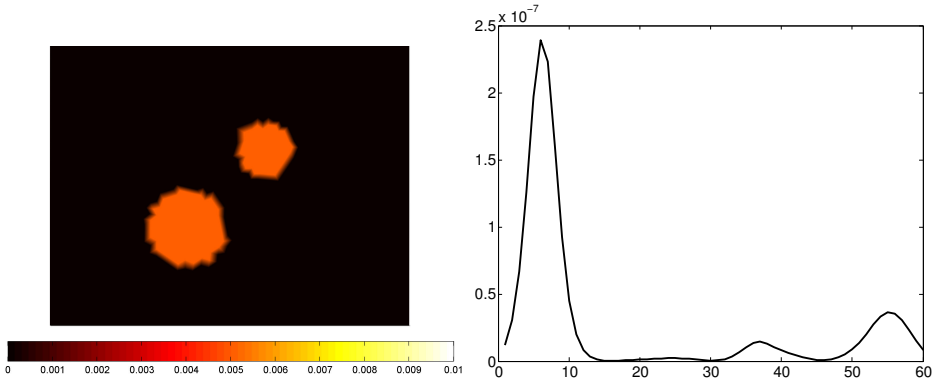
### 3.6.1 Comparison between OSKCA and Tikhonov regularization

In our first simulation, we consider a square domain with two fluorescent inclusions in it, which is shown in Figure 3a. The domain has the size $w \times h$, where the width $w$ is 91.6mm and the height $h$ is 71.5mm. The parameters are $\mu_s = 1\text{mm}^{-1}$, $\mu_a = 0.01\text{mm}^{-1}$, and $R = 1.4$. 40 light sources and 60 detectors are put on the boundary. Each time we turn on one source with others off and get the measurements from all detectors, which is illustrated in Figure 3b. Totally we have $40 \times 60 = 2400$ measurements. Different levels of noise are added to the simulated measurements for comparison.

In this example, a FEM mesh with 7938 triangles and 4096 nodes is generated for the formulation of the forward problem, and a mesh with 4352 triangles and $n_p = 2253$ nodes is generated for solving the inverse problem. The two meshes are different to avoid the "inverse crime" known in the literature [73]. The linear equation for FEM is solved by an implementation of the Algebraic Multi-grid Method [83].

We apply Tikhonov regularization and OSKCA to solve this problem respectively.

For Tikhonov regularization, point sources are used as the basis functions. We




(a) (b)

Figure 3: (a) An illustration of the fluorophore distribution. (b) The boundary measurement of the emission field for one light source. All detectors are arranged counter-clockwise, and their measurements form a 1D signal. We can see that the 1D signal is very smooth, with many places nearly zero, and decays exponentially away from the peak value points.

use ART for $L^2$ regularization, and Bregman Operator Splitting (BOS) for TV regularization.

For OSKCA, the basis functions are chosen as

$$\{\cos(2\pi(p\frac{x}{w}+q\frac{y}{h})),\sin(2\pi(p\frac{x}{w}+q\frac{y}{h}))\}, \tag{208}$$

where $|p|,|q|\leq 10$. After removing those duplicates, we actually have $n=441$ basis functions. We generate matrix $\mathbf{M}$ in the inverse problem (86). And then apply iterative Tikhonov regularization (125) with parameter $h=10^{-2}$ for the computation of the orthogonal solution. Algorithm described in Section 3.3.4 provides a basis for the kernel of $\mathbf{M}$, where parameter $\tau=10^{-4}\times\|\mathbf{M}\|_F$. The ART is used if the positivity constraint is applied for the kernel correction. Algorithm 3.3.1 is used if TV minimization requirement is used.

We compare the results obtained by OSKCA and by Tikhonov regularization methods. Figure 4 shows the reconstructed images. As one can see that OSKCA has an advantage of achieving cleaner images and being more robust against noise. As the noise level increase, Tikhonov regularization method needs to penalize the regularization term more, which results in a blurry reconstruction. In OSKCA, the

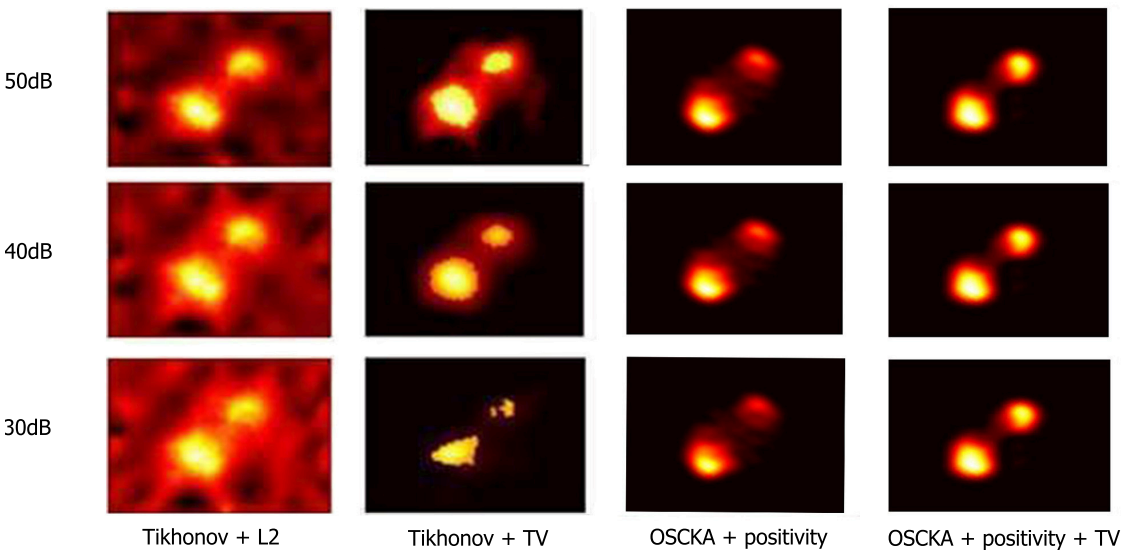



Figure 4: Reconstructed fluorescent distributions for 2D simulated data.

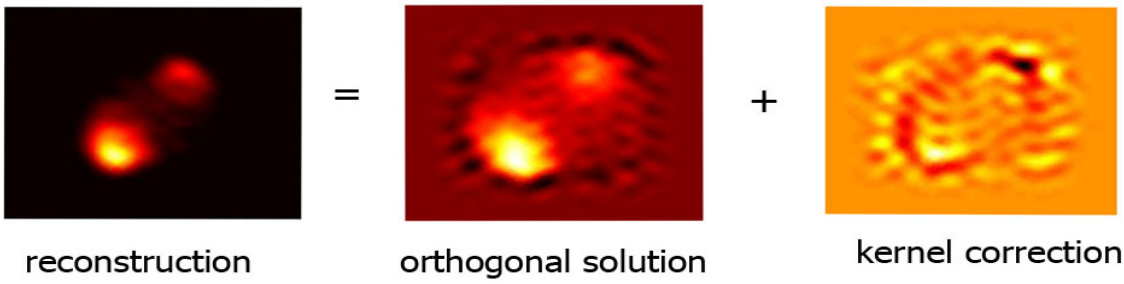



Figure 5: The final reconstruction by OSKCA is decomposed into the orthogonal solution and the kernel correction. Noise of 30dB is added in the synthetic boundary measurement.

orthogonal solution is computed so that it fits the data and is robust against the noise, though it is quite blurry, as shown in Figure 5. The kernel correction, which is done in the kernel space, regularizes the solution without affecting the data fidelity, so that the regularization requirement for the solution can be better satisfied. Figure 5 illustrates the effect of kernel correction.

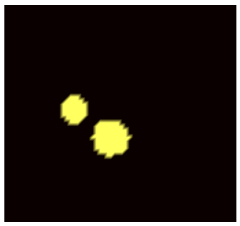
We also compare the computation time in Table 1. All the computations are performed on a laptop with 2.53GHz Intel Core2 Duo CPU. The programming interface is MATLAB with C++ subroutines. It shows OSKCA gains a dramatic speedup. This is partly due to the much smaller number of basis functions in use for OSKCA, and reduced size of the system matrix by the compression of measurement data.

Table 1: CPU time of different methods (in seconds)

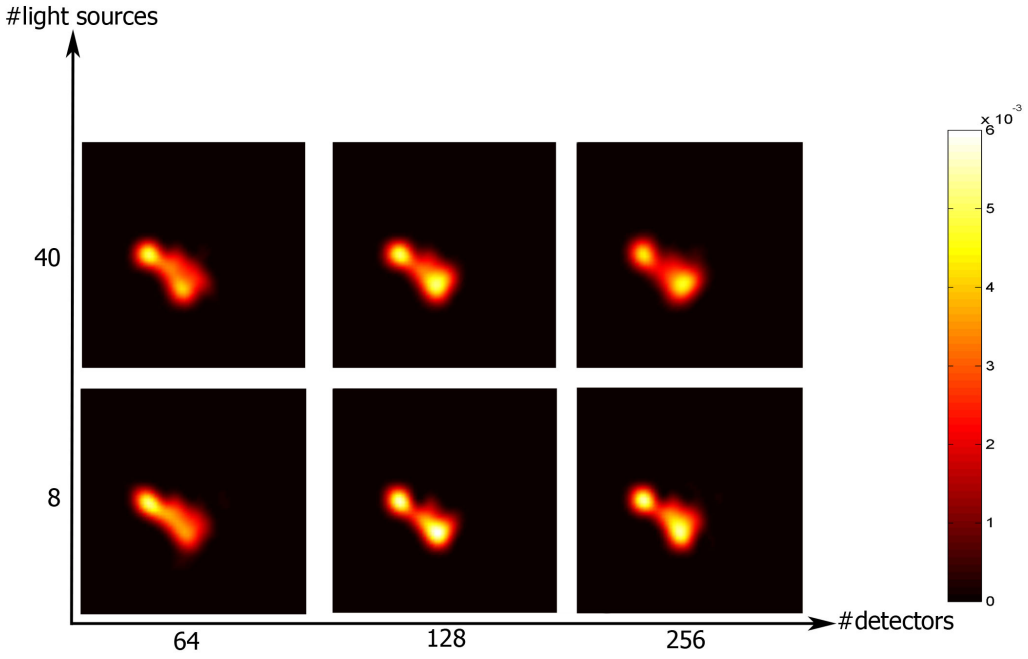
| Tikhonov + L2 | Tikhonov + TV | OSKCA + positivity | OSKCA + TV |
|---|---|---|---|
| 2812 | 5919 | 320 | 230 |

### 3.6.2 The effect of using more detectors and light sources

In the next example, a 50mm×50mm medium is implanted with two circular inclusion. Their radius are 3mm and 4mm respectively, and their center-to-center distance is 10mm (see Figure 6a). The optical properties are the same as the previous example. The basis for the solution space is the same as (208). The mesh for the forward model has 7839 nodes and the mesh for the inverse problem has 4096 nodes. In each simulation, we use 64, 128 or 256 detectors, and 8 or 40 light sources. Totally we have 6 configurations of detectors and light sources. For each configuration, we compute the synthetic boundary measurements, and use OSKCA with positivity constraint regularization method to reconstruct the solution. The reconstructions are displayed in Figure 6b. We can see that increasing the number of detectors from 64 to 128, and adding the number of light sources from 8 to 40 will improve the resolution of the reconstruction results. However, further increasing the number of detectors and light



(a) The ground truth.



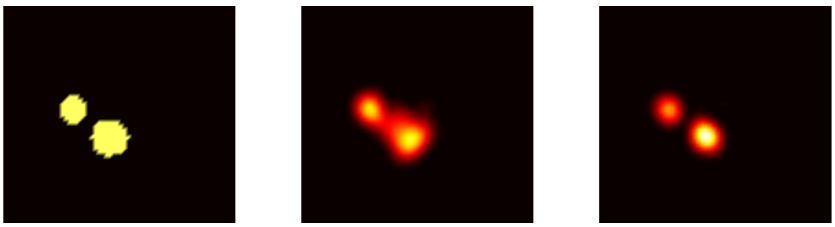

(b) The reconstruction results for each configuration of detectors and light sources.

Figure 6: A comparison of the reconstruction results by using different number of detectors and light sources.

source doesn't help much. This can be overcome by taking advantage of some other regularization techniques that will preserve the visual features in the reconstructions, which is demonstrated in the following simulation.

### 3.6.3 Comparison between different regularization techniques for the kernel correction

In this example, we still use the same phantom as shown in Figure 6a. The number of the light sources in use is 40, and the total number of detectors is 256. Besides using only the positivity constraint for the kernel correction as in Figure 6b, we also use OSKCA-TV (Algorithm 3.3.1), which uses both TV regularization and positivity constraint. The comparison of the results is shown in Figure 7. We can see that the celebrated TV regularization techniques can resolve the edges in the reconstructed images.



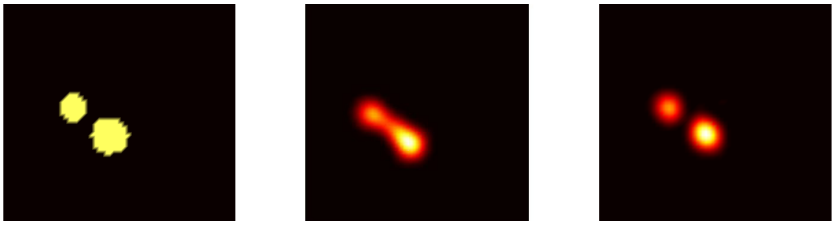
(a) The ground truth. (b) positivity (c) positivity + TV

Figure 7: Comparison between positivity constraint regularization and positivity plus TV regularization.

### 3.6.4 Comparison between OSKCA and direct Augmented Lagrangian method

We continue to use the same phantom as shown in Figure 6a. We apply Augmented Lagrangian method directly to the problem (144) by solving the formulation (146). We also use OSKCA-TV (Algorithm 3.3.1) to solve the same problem. For OSKCA-TV, we use the set of parameters as $\alpha = 10^{-5}, \rho_1 = \rho_2 = 1$. And for direct Augmented Lagrangian method, the parameters are $\gamma = 10^{-5}, \beta_1 = \beta_2 = 1, \beta_3 = 10^4$. The maximum numbers of iterations are both 1000. The reconstructed distribution $f$

from these two approaches are shown in Figures 8(b)(c). We can see that OSKCA has better resolution than the direct Augmented Lagrangian method.



(a) (b) (c)

Figure 8: (a) Ground Truth. (b) Augmented Lagrangian with TV minimization regularization. (c) OSKCA with TV minimization regularization.

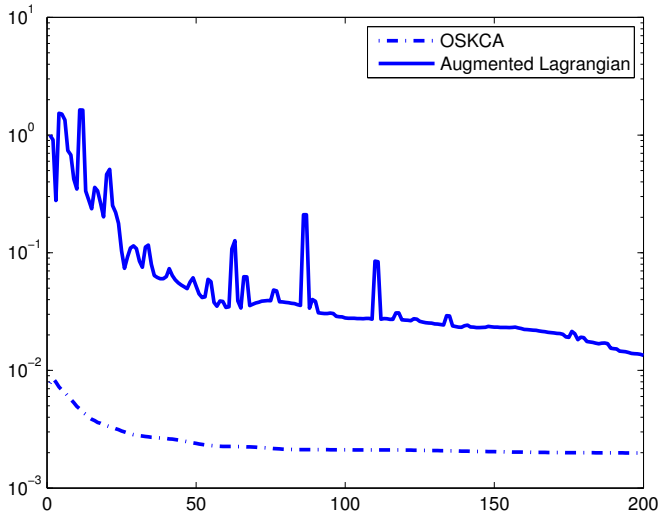



Figure 9: The plot of $\|\mathbf{Mc} - \mathbf{g}\|/\|\mathbf{g}\|$ against the number of iterations.

We also compare the relative error of the data fitting term $\frac{\|\mathbf{Mc}-\mathbf{g}\|}{\|g\|}$ in these two methods, as shown in Figure 9. We can see that OSKCA has significantly better data fidelity in 200 iterations. The main reason is that the initial value for $f$ in OSKCA is the orthogonal solution, which is intended to satisfy the data fitting requirement. Later changes in the data fitting are caused by the numerical error in the computed kernel space, which is controlled by $\tau$ in (132). $\tau$ can be chosen small enough, so that the changes are negligible. In this example, the kernel correction

actually improves the data fidelity, and the relative error decreases monotonically. In comparison with OSKCA, the direct Augmented Lagrangian method starts with relative error $\frac{\|\mathbf{Mc}-\mathbf{g}\|}{\|g\|} = 0$ for $\mathbf{c} = \mathbf{0}$. The relative error is oscillatory in the iterations, and the overall convergence rate is low. We also have a time comparison of these two methods in Table 2. The cost for the formulation of the forward model is not included, which is the same for both these methods. We can see that OSKCA is much faster than Augmented Lagrangian.

Table 2: CPU time (in seconds) of OSKCA and Augmented Lagrangian

| | Augmented Lagrangian | OSKCA |
|---|---|---|
| kernel space | N/A | 0.58 |
| orthogonal solution | | 0.39 |
| kernel correction | | 2.66 |
| total | 27.24 | 3.63 |

### 3.6.5 The resolution of the reconstruction with respect to the depth of the source

We consider a 50mm×50mm homogeneous medium implanted with two circular fluorescent inclusions with radius 3mm, and their center-to-center distance is 7mm. The optical parameters are $\mu_s = 1\text{mm}^{-1}$ throughout the domain and $\mu_a = 0.01\text{mm}^{-1}$. By varying the depth of the inclusions, we compare the reconstruction results, which are shown in Figure 10. In each case, the forward model is computed on a fine mesh with 1789305 nodes so that it can be considered as the physical truth, and the inverse problem is solved on a coarse mesh with only 4096 nodes, which models the situation when the modeling error is not negligible.

We can see that as the depth of the source increases, the resolution of the reconstruction gets worse. As shown in Figure 10, when the centers of the fluorescent inclusions are 20mm deep, the two inclusions become indistinguishable. This is partly due to the diffusive nature of light propagation. Also the mesh for the inverse problem is not fine enough, so that the PDE solver is not accurate and the modeling error

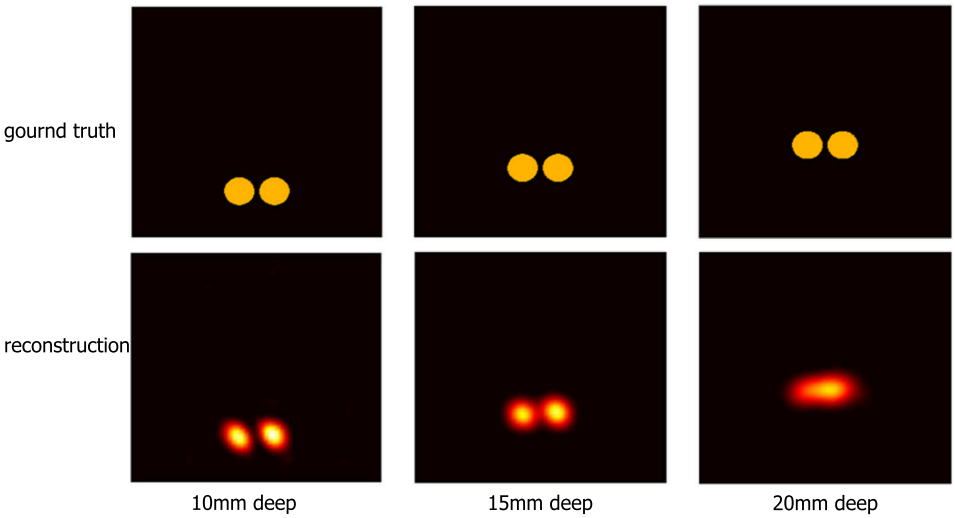



Figure 10: The first row are the images of the ground truth, and the second row are the reconstructions on a mesh with 4096 nodes.

is large. It can be improved if a finer mesh and a better PDE solver are used for the reconstruction of the solution. To demonstrate this, we use a mesh with 262144 nodes for the same inverse problem as illustrated in the third column of Figure 10. The resolution of the reconstructed image is significantly better, which is shown in Figure 11b.

(a) Ground truth (b) Reconstruction

Figure 11: The fluorescent inclusions are 20mm deep. Reconstruction is done on a mesh with 262144 nodes.

## *3.7 An experimetal 3D phantom study with real data*

### 3.7.1 Experiment setup

We use a non contact continuous-wave (CW) transillumination phantom-based FT system for the validation of our proposed approach. Figure 12 depicts the setup of the

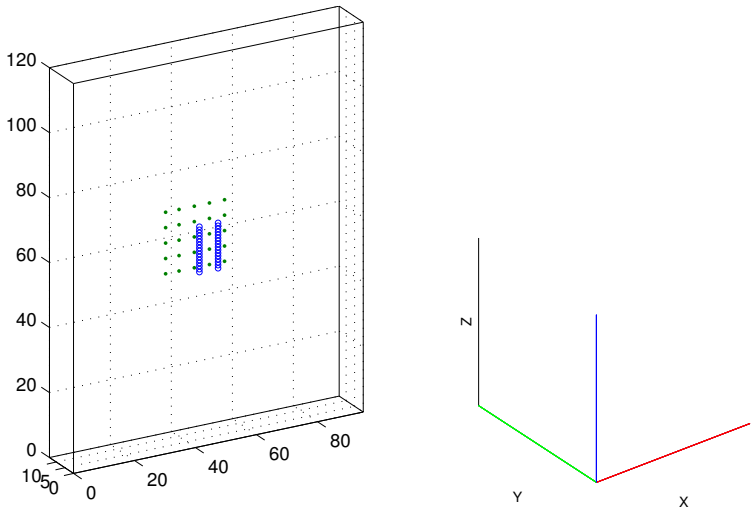



Figure 12: The phantom used in the experiment is depicted as a slab shaped domain. Two tube inclusions are shown in blue and the locations of light tips are displayed as a 2D array of green dots.

experiment. The phantom has the dimension of 95mm $\times$ 14mm $\times$ 120mm, filled with an intra-lipid liquid, which mimics the optical property of biological tissues. Indian ink is added to it, in order to match the tissue absorption. Two thin glass tubes, encapsulating a solution of fluorescent dye, is implanted to the phantom. The two tubes, both 2mm in diameter and 15mm in length, are parallel to each other with center-to-center distance of 6mm. Their depth with respect to the front surface (the plane $y = 0$mm) is 6mm. They are small inclusions, and their induced changes to optical parameters is negligible. (Though the study of optical inhomogeneity at the anatomical level, called optical tomography, has been an active area of research [80, 79, 5, 13, 77, 31], and Section 3.4.2 of this thesis also sheds light on this topic.) Thus the phantom is considered to be homogeneous, with a scattering coefficient 3.6mm$^{-1}$, and absorption coefficient 0.03mm$^{-1}$, at both excitation and emission wavelength.

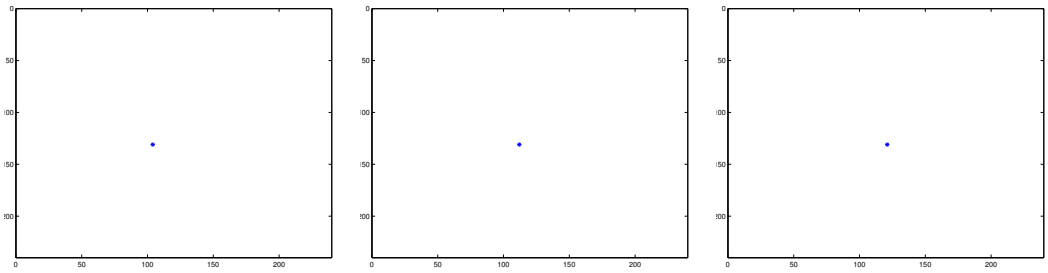
These optical parameters are measured in a separate experiment.

A He-Ne laser fiber produces CW light at 632nm wavelength, whose tip lies on the rear surface ($y = 14$mm) to direct the light into the phantom. The tip of the fiber is translated on a 2D $5 \times 5$ uniform grid of 25 points (see Figure 12), each one serves as a different light source location. The grid spacing is 5mm. A cooled CCD camera captures the light signals transmitting through the front surface (called the measurement plane), and form images representing the intensity of the signals. We note that the signals are the mixture transillumination excitation ($\Phi_x^{(i)}$) and the fluorescent emission ($\Phi_m^{(i)}$). They are separated by specific filters before being received by the camera. So for each light source, two images are obtained. The images of transillumination excitation are used to calibrate the intensities of the light source, and to normalize the images of fluorescent emission to the same scale. Then the normalized fluorescent emission images are ready for the reconstruction algorithm. In this experiment, the resolution of the images is $240 \times 240$ pixels. Figure 13 shows some data images obtained in this experiment.

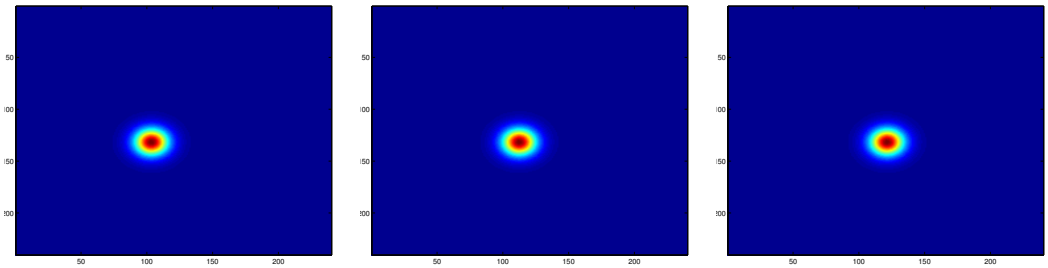
### 3.7.2 Solution process

**FEM discretization.** A tetrahedral mesh with 228800 uniformly spaced nodes are generated for the discretization of the forward and inverse problems. The light sources are modeled by point-wise functions, which is a shape function in the finite element framework. The excitation radiation in the whole domain given a light source can be computed by solving the DA equation in the forward model. The computed excitation radiation is calibrated by the corresponding transillumination image, as described previously.

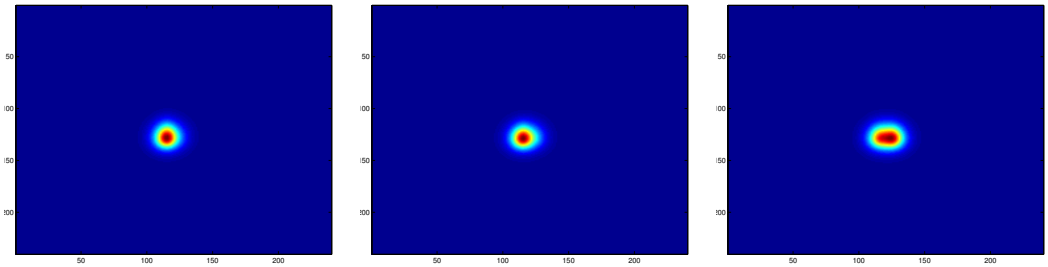
**The system matrix for the inverse problem.** Totally 25 data images of the size $170 \times 180$ are captured by the camera. So the amount of data is huge, with $n_s = 25$ and $n_d = 30600$ in (87). Using all of them to form the system matrix



(a)



(b)



(c)

Figure 13: (a) Three light sources, which are 5mm apart, viewed from the measurement plane. (b) The images of the corresponding transillumination excitation. (c) The images of the corresponding fluorophore emission.

$\mathbf{M} \in \mathbb{R}^{n_s n_d \times n}$ in (87) will cause excessive storage and computation cost. Even worse, the resulting system matrix $\mathbf{M}$ will have more rows than columns, which contradicts the under-determined nature of the inverse problem. Therefore, the strategy described in Section 3.4.1 is useful here. We first discard some small measurements based on the intensity profile defined in (163) and the criteria (167). More exactly, In this experiment, $\mu_e = 0.57\text{mm}^{-1}$, $\|\xi_i' - \xi_i\| = 14\text{mm}$. We choose $\tau = 0.01$, so

$$B_i = \{x : \|x - \xi_i\| \leq 22.1\text{mm}\}. \tag{209}$$

The fluorophore emission image for the $i$-th light source is truncated by cutting off data points outside of $B_i$. Then the truncated data images are further compressed in the wavelet domain as described in (175), where the wavelet filters are taken to be bi-orthogonal wavelet 3/3. The threshold for the wavelet compression on the details coefficients is taken to be the 95% of total energy. Based on the criteria in (177), we choose the region of interest for reconstruction of $f$, which has the size $50\text{mm} \times 14\text{mm} \times 50\text{mm}$. The mesh nodes inside the region of interest form a tensor grid of $58 \times 16 \times 60$. For the representation of the solution $f$, the basis consists of 3D bi-orthogonal 2/2 wavelets for reconstruction. We choose wavelet basis functions at level 1, which is the collection of tensor product of 33 wavelets in $x$ dimension, 13 wavelets in $y$ dimension, and 35 wavelets in $z$ dimension. We have a new system for the inverse problem (17) with $\mathbf{g}'$ a $6233 \times 1$ vector, $\mathbf{M}'$ a $6233 \times 15015$ matrix. The unknown $\mathbf{c}$, which is the coefficient vector for $f$ under the tensor product wavelet basis, has the size $15015 \times 1$.

**Choice of algorithms for orthogonal solution and kernel correction**. The computation of the orthogonal solution can still use Iterative Tihkonov regularization as described in Section 3.3.3. However, in the setup of transillumination FT in this case, the boundary data are take from only one surface of the phantom, and large amount of data, especially the boundary measurement on the opposite side of the measurement plane, are missing. If we still apply the old formulation (125), we will

get a solution that matches the boundary data on the measurement surface, while the boundary data on the un-measured surfaces are essentially treated as arbitrary, making this inverse problem even more ill-posed. To alleviate the ill-posedness caused by missing data, we incorporate more physical constraints into the modeling. For example, we require that the synthetic boundary data of the computed orthogonal solution should be non-negative on both measured and unmeasured surfaces of the phantom. If we consider the boundary data in the wavelet domain of bi-orthogonal 3/3, this requirement is translated to the approximation coefficients being non-negative, as the reconstruction scaling function is positive on its support. Let $\bar{\mathbf{M}}'$ be the system matrix that is generated by the same set of basis as $\mathbf{M}'$, but taking boundary measurements on the opposite side of the measurement surface. We therefore consider the following formulation for the orthogonal solution

$$\mathbf{c}^* = \underset{\mathbf{c}}{\operatorname{argmin}} \|\mathbf{M}'\mathbf{c} - \mathbf{g}'\|_2^2 + \alpha\|\mathbf{c}\|_2^2 \tag{210}$$

such that

$$\begin{cases} \mathbf{M}'\mathbf{c} & \geq & 0 \\ \bar{\mathbf{M}}'\mathbf{c} & \geq & 0 \end{cases} \tag{211}$$

This can be solved by Augmented Lagrangian algorithm that has been addressed previously. For the choice of the kernel correction, we impose the $l^1$ minimization on the wavelet coefficients $\mathbf{c}$. This is because the sparsity in the wavelet domain is an important feature for images [78], and $l^1$ minimization, which is a convexification of $l^0$ minimization, has been known to enforce sparsity. In light of this, we solve the following optimization problem for the kernel correction, which minimizes the $l^1$-norm in addition to TV semi-norm of the solution

$$\min_{\boldsymbol{\lambda}} \|\mathbf{c}_* + \mathbf{K}\boldsymbol{\lambda}\|_1 + \beta\|\mathbf{c}_* + \mathbf{K}\boldsymbol{\lambda}\|_{TV} \tag{212}$$

such that

$$W_S^0(\mathbf{c}_* + \mathbf{K}\boldsymbol{\lambda}) \geq 0, \tag{213}$$

where $\mathbf{K}$ is a set of orthonormal basis for $\mathcal{N}(\mathbf{M}')$, $W_S^0$ is the wavelet reconstruction operator that transforms wavelet coefficients to level 0. Since level 0 wavelets are the point-wise basis function in our construction of wavelets, $W_S^0(\mathbf{c}_* + \mathbf{K}\boldsymbol{\lambda})$ represents a function in the physical domain, which is the final reconstruction of the solution $\mathbf{f}$.

**Multi-level reconstruction.** Based on the multi-level algorithms described in Algorithm 3.5.1 and 3.5.2 in Section 3.5, we use a two-level strategy for the implementation of OSKCA. More exactly, we use wavelet basis functions at level 1 for the fine level computation of the orthogonal solution and kernel correction, and wavelet basis functions at level 2 for the coarse level computation. The system matrix for the coarse level basis is $\hat{\mathbf{M}}'$, which has the size $6233 \times 4851$.

**An alternative approach for computing the orthogonal solution**. Besides the formulation (210)(211) for computing the orthogonal solution, we can also consider other approaches. For instance, Tihkonov regularization with $l^1$ minimization in the wavelet domain is proposed as

$$\mathbf{c} = \underset{\mathbf{c}}{\operatorname{argmin}} \|\mathbf{M}'\mathbf{c} - \mathbf{g}'\|_2^2 + \alpha\|\mathbf{c}\|_1, \tag{214}$$

which can be solved by ISTA [28, 35, 40, 52] and FISTA , algorithms [14]. We note that our multi-level framework also applies to Tihkonov regularization as (214), in a way similar to Algorithm 3.5.2. To elaborate this idea, we implement a two-level version. Let $\mathbf{c}_2, \mathbf{c}_1$ be the solutions to (214) at the coarse and fine levels respectively. $\mathbf{c}_2$ solves

$$\mathbf{c}_2 = \underset{\hat{\mathbf{c}}}{\operatorname{argmin}} \|\hat{\mathbf{M}}'\hat{\mathbf{c}} - \mathbf{g}'\|_2^2 + \alpha\|\hat{\mathbf{c}}\|_1. \tag{215}$$

The fine level correction $\Delta\mathbf{c}_1$ solves

$$\Delta\mathbf{c}_1 = \underset{\mathbf{c}}{\operatorname{argmin}} \|\mathbf{M}'\mathbf{c} - \Delta\mathbf{g}'\|_2^2 + \alpha\|\mathbf{c}\|_1, \tag{216}$$

where

$$\Delta\mathbf{g}' = \mathbf{g}' - \hat{\mathbf{M}}'\mathbf{c}_2. \tag{217}$$

Then the fine level solution is written as

$$\mathbf{c}_1 = W_S\mathbf{c}_2 + \Delta\mathbf{c}_1, \tag{218}$$

where we recall that $W_S$ denoted one level of wavelet reconstruction. Here we perform a simple comparison between the solution obtained by two-level algorithm, and the true solution to (214). Obviously,

$$\|\mathbf{M}'\mathbf{c} - \mathbf{g}'\|_2^2 + \alpha\|\mathbf{c}\|_1 \leq \|\mathbf{M}'\mathbf{c}_1 - \mathbf{g}'\|_2^2 + \alpha\|\mathbf{c}_1\|_1. \tag{219}$$

Suppose $\mathbf{c}$ solves (214), and it is decomposed as

$$\mathbf{c} = W_S\mathbf{c}_2 + \Delta\bar{\mathbf{c}}_1, \tag{220}$$

where $W_S\mathbf{c}_2$ is the same as in (218). A simple calculation shows

$$\|\mathbf{M}'\mathbf{c}_1 - \mathbf{g}'\|_2^2 + \alpha\|\mathbf{c}_1\|_1 \leq \|\mathbf{M}'\mathbf{c} - \mathbf{g}'\|_2^2 + \alpha\|\mathbf{c}\|_1 + \alpha(\|\Delta\bar{\mathbf{c}}_1\|_1 + \|\Delta\mathbf{c}_1\|_1). \tag{221}$$

So the difference between the optimal objective and the computed objective by two level method for (214) is no larger than $\alpha(\|\Delta\bar{\mathbf{c}}_1\|_1 + \|\Delta\mathbf{c}_1\|_1)$. Under the assumption that

$$\|\Delta\mathbf{c}_1\|_1, \|\Delta\bar{\mathbf{c}}_1\|_1\| \ll \|W_S\mathbf{c}_2\|_1, \tag{222}$$

meaning that the $l^1$ norm of fine level correction is much smaller than that of the coarse level solution, the two-level algorithm produces a near optimal solution. The kernel correction corresponding to this choice of the orthogonal solution can be proposed the same as (212).

### 3.7.3 Results

**Comparison between OSKCA and Tikhonov regularizations with $L^2$ penalty term.** Using the same data, we perform the following reconstruction algorithms:

(a) $L^2$ **OSKCA**: OSKCA with tensor product wavelet basis. The orthogonal solution is computed by $L^2$ regularization (210)(211). The kernel correction is solved by (212) ($L^1$+$TV$+ positivity constraint).

(b) $L^2$ **Tikhonov**: Tikhonov regularization with point-wise basis. The classical $L^2$ penalty is imposed on the solution. The regularization parameter is tuned so that the solution is positive.

(c) $L^1$ **OSKCA**: OSCKA with tensor product wavelet basis. The orthogonal solution is computed by $L^1$ regularization (214). The kernel correction is solved by (212) ($L^1$+$TV$+ positivity constraint).

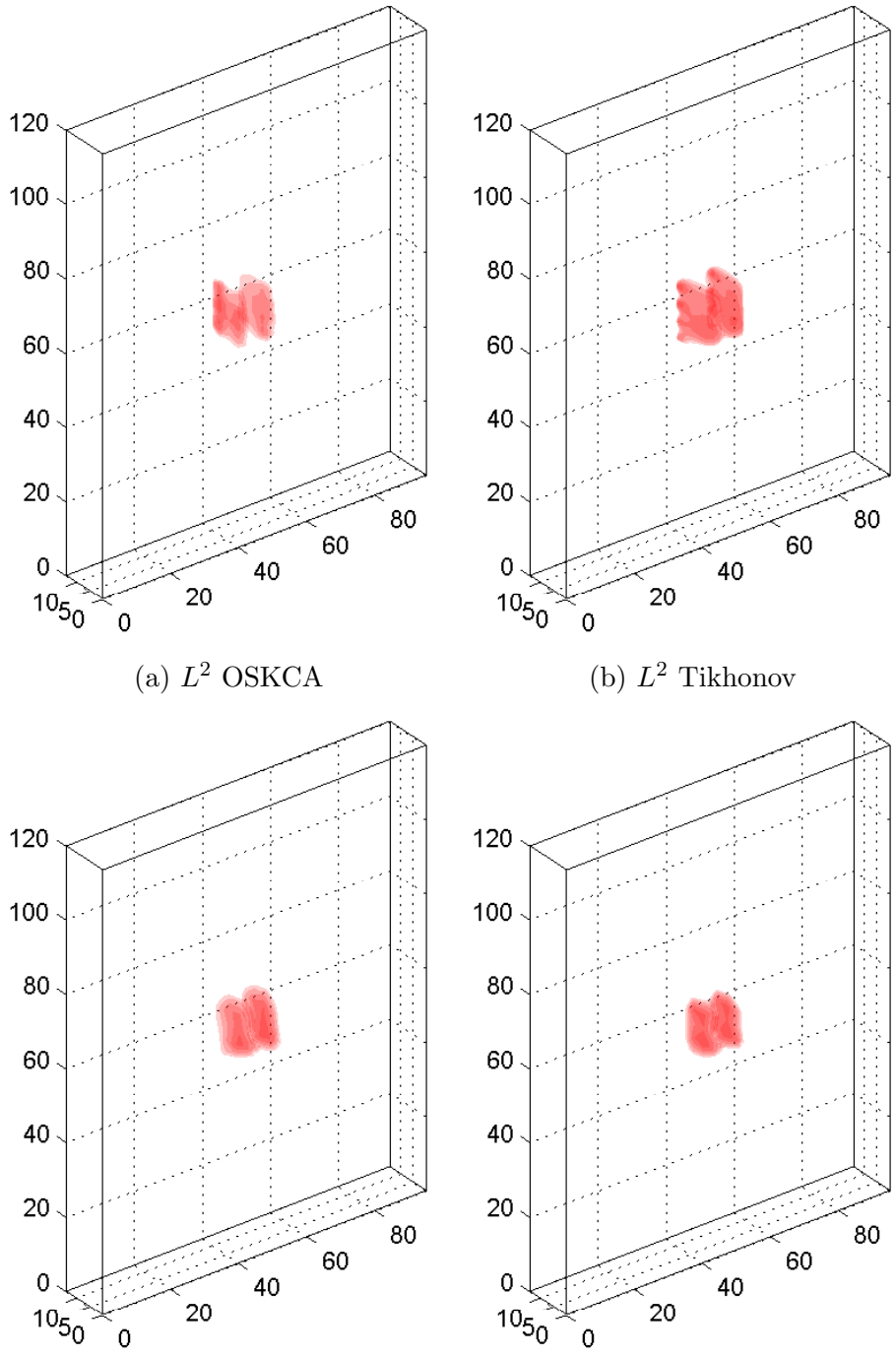
(d) $L^1$ **Tikhonov**: Tikhonov regularization with tensor product wavelet basis. $L^1$ penalty in imposed on the wavelet coefficient of the solution. The regularization parameter is tuned so that the solution is positive.

The reconstructions of fluorophore distribution $f$ by the above three approaches are shown in Figure 14.

The viewpoint of Figure 14 is not informative to tell the resolution of the reconstruction results. Especially, we want to examine the resolution in the $y$-dimension, which is perpendicular to the measurement plane. For this purpose, we plot the projections of the 3D images onto three coordinate planes: $xy$, $yz$ and $xz$, which are displayed in Figures 15 and 16. The tubes are marked with solid curves.

### 3.7.4 An analysis on the resolution of the reconstruction results

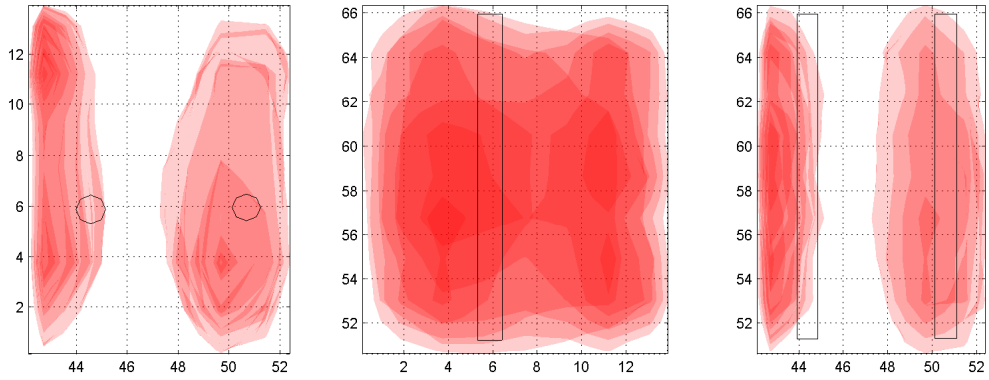
Figure 15 shows that the solution obtained by $L^2$ OSKCA has much better resolutions viewed from $xy$ and $xz$ planes. Figure 16 shows that the resolution of the solutions by $L^1$ OSKCA and $L^1$ Tikhonov are comparable, though $L^1$ OSKCA shows slightly better resolution from the view of $xz$ plane. For every solution, the $yz$ view is very blurry. More exactly, the blurring of the solution is much more severe in $y$ dimension than other two dimensions. This can be partially explained through the following analysis. We note that the size of the slab shaped phantom is much larger in $x$ and $z$



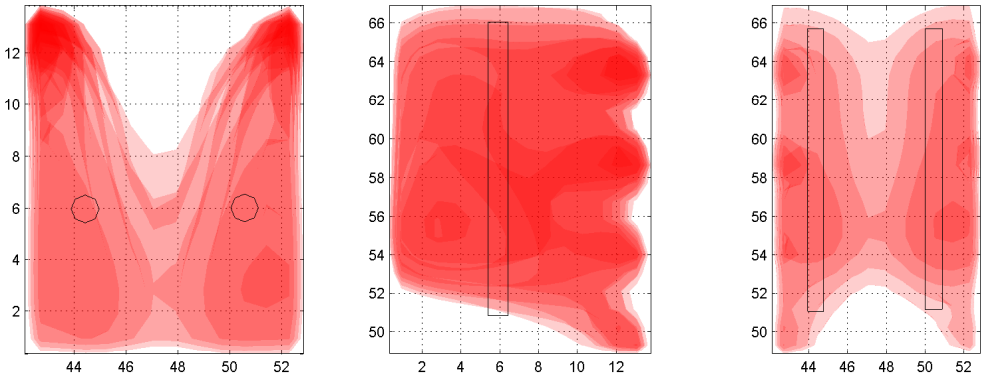

(a) $L^2$ OSKCA (b) $L^2$ Tikhonov

(c) $L^1$ OSKCA (d) $L^1$ Tikhonov

Figure 14: The comparison of the reconstruction results between (a)(c) OSKCA and (b)(d) Tihkonov regularization.

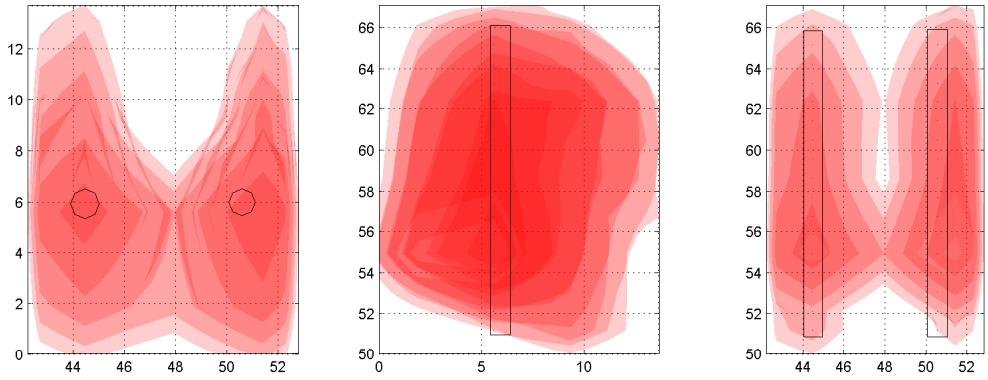



(a) $L^2$ OSKCA. relative error of data fitting = 0.14

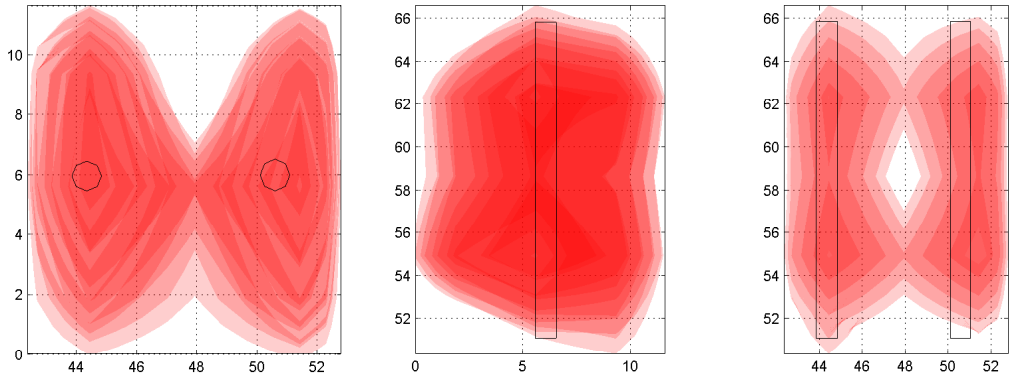



(b) $L^2$ Tikhonov. relative error of data fitting = 0.17

Figure 15: A comparison between $L^2$ OSKCA and $L^2$ Tikhonov methods. From left to right are the projections of 3D images onto $xy$, $yz$ and $xz$ planes. On the first column, the two circles depict the cross sections of the tubes. On the second and the third columns, the rectangles denote the placement of the tubes.




(a) $L^1$ OSKCA. relative error of data fitting = 0.11




(b) $L^1$ Tikhonov. relative error of data fitting = 0.14

Figure 16: A comparison between $L^1$ OSKCA and $L^1$ Tikhonov methods. From left to right are the projections of 3D images onto $xy$, $yz$ and $xz$ planes. On the first column, the two circles depict the cross sections of the tubes. On the second and the third columns, the rectangles denote the placement of the tubes.

dimensions than in $y$ dimension, so it can be approximated by a semi-infinite domain

$$\Omega_d = \{(x, y, z) : -\infty < x, z < \infty, 0 < y < d\}, \tag{223}$$

where $d$ is the thickness of the slab. The light source is placed on the plane $y = d$, and the measurement plane is $y = 0$.

For the light source $i$, the fluorophore emission is written as

$$\Phi_m^{(i)}(x, y, z) = \int_{\Omega_d} G_d(x, y, z, x', y', z')\Phi_x^{(i)}(x', y', z')f(x', y', z')\,\mathrm{d}x'\,\mathrm{d}y'\,\mathrm{d}z', \tag{224}$$

where $G_d$ is the Green's function for the DA equation in domain $\Omega_d$, $\Phi_x^{(i)}$ is the transillumination excitation for light source $i$, $f$ is the fluorophore distribution. For $\Omega_d$ as defined in (223), its Green's function is translation invariant in $x$ and $z$. More exactly, we can rewrite $G_d$ as

$$G_d(x, y, z, x', y', z') = G_d(x - x', z - z'; y, y'). \tag{225}$$

The boundary measurements for light source $i$ is formally written as

$$\Phi_m^{(i)}(x, 0, z) = \int_{\Omega_d} G_d(x, 0, z, x', y', z')\Phi_x^{(i)}(x', y', z')f(x', y', z')\,\mathrm{d}x'\,\mathrm{d}y'\,\mathrm{d}z', \tag{226}$$

and by (225),

$$\Phi_m^{(i)}(x, 0, z) = \int_0^d \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G_d(x - x', z - z'; 0, y')\Phi_x^{(i)}(x', y', z')f(x', y', z')\,\mathrm{d}x'\,\mathrm{d}z'\,\mathrm{d}y'. \tag{227}$$

For simplicity of notations, we denote

$$\Psi^{(i)}(x', y', z') = \Phi_x^{(i)}(x', y', z')f(x', y', z'). \tag{228}$$

Then for the light source $i$, the boundary measurement is

$$\Phi_m^{(i)}(x, 0, z) = \int_0^d \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} G_d(x - x', z - z'; 0, y')\Psi^{(i)}(x', y', z')\,\mathrm{d}x'\,\mathrm{d}z'\,\mathrm{d}y'. \tag{229}$$

The righthand side of the above equation can be viewed as the convolution in variables $x$ and $z$. Ideally, we assume that all boundary measurements on the measurement plane

$$\{(x, y, z) : -\infty < x, z < \infty, y = 0\} \tag{230}$$

can be obtained. By taking the Fourier transform of (229) with respect to the variables $x$ and $z$ we have

$$\widehat{\Phi_m^{(i)}}(\omega_x, 0, \omega_z) = \int_0^d \widehat{G_d}(\omega_x, \omega_z; 0, y')\widehat{\Psi^{(i)}}(\omega_x, y', \omega_z)\,\mathrm{d}y', \tag{231}$$

where

$$\widehat{\Psi^{(i)}}(\omega_x, y', \omega_z) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x', y', z')\Phi_x^{(i)}(x', y', z')e^{-j(\omega_x x' + \omega_y z')}\,\mathrm{d}x'\,\mathrm{d}z'. \tag{232}$$

In our assumption, given the Fourier transform of the boundary measurements $\widehat{\Phi_m^{(i)}}(\omega_x, 0, \omega_z)$, and the Fourier transform of the Green's function $\widehat{G_d}(\omega_x, \omega_z; 0, y')$ can be computed for any given $\omega_x, \omega_z, y'$. The inverse problem is to first solve the integral equation (231) for $\widehat{\Psi^{(i)}}(\omega_x, y', \omega_z)$ with any given $\omega_x, \omega_z, y'$, then solve for $f$ in (232) by Inverse Fourier Transform for any given $x', y', z'$. The first step is a severely ill-posed problem. The reason is that in (231), given the weighted integral of a function $\widehat{\Psi^{(i)}}(\omega_x, y', \omega_z)$ with respect to $y'$, that function is not uniquely determined. This partially explains the reason that our numerical results is severely blurred in $y$ dimension, as the solution is strongly regularized in that dimension in order to alleviate the indeterminacy. In order to further alleviate the ill-posedness of solving (231), more information on $\widehat{\Psi^{(i)}}(\omega_x, y', \omega_z)$ in variable $y'$ is needed, though it is difficult to obtain in this experimental setting.

The second step is also ill-posed, because in (150), $\Phi_x^{(i)}$ decays exponentially, which numerically behaves like a function with compact support. In light of this, the righthand side of (232) can be considered as a Windowed Fourier Transform of $f$. Therefore solving (232) alone will encounter stability issues. This is alleviated

by using multiple light sources, that is, we have (232) for various $\Phi_x^{(i)}$'s, where $i = 1, \ldots, s$. Therefore, solving $f$ in $s$ simultaneous equations (232) is approximately solving the Inverse Windowed Fourier Transform of $f$.

We perform the above analysis to illustrate challenges that come from the ill-posedness of the formulation of the problem. In practice, the problem is handled by numerical inversion techniques, rather than the analytical approach as shown in (231)(232). With prior information of the solution, such as positivity, sparsity, and smoothness, combined with proper choices of regularization techniques, the regularized solution can be computed, as it is demonstrated in the previous examples. However, the artifacts that caused by regularization is evident. In our opinion, this can be overcome by incorporating other imaing modalities in the experimental setting, so that additional information on the $y$ dimension of the unknown can be obtained.

# CHAPTER IV

# CONCLUSION

As demonstrated in the numerical examples, the proposed OSKCA has advantages over the Tikhonov type regularization methods in two ways. First, in OSKCA, the regularization can be enforced better than that in the Tikhonov regularization methods. OSKCA solves regularization without the constraint of data fitting, while in Tikhonov regularization, the regularization term is minimized together with the data fitting term. Two terms compete with each other in the minimization process and a compromise has to be taken between them. Therefore, the reconstruction results of OSKCA have more regularity and less artifacts than that of Tikhonov regularization. Second, in Tikhonov regularization, the point source basis is used, and it is not necessarily an efficient basis to represent the reconstructed source distribution. In OSKCA, the reconstructed source term is expressed under a more efficient basis. In this way, the dimension of the unknowns is greatly reduced. As a result, a considerable speedup is gained in both the formulation of the forward model matrix and the reconstruction process. Also, we can increase the resolution of the reconstruction by adding more basis functions or changing the basis locally.

There are several interesting features of OSKCA. It is known that error and artifacts are considered as the bottleneck for the existing methods for FT applications. But for OSKCA, besides its computational efficiency, we demonstrate through our numerical studies that it is robust against noise and perturbations, while having the potential to improve the resolution in image reconstructions dramatically. Since the kernel correction step regularizes the solution in the kernel space, the proposed approach is particularly useful for severely under-determined system.

Besides the above strategies to improve the efficiency of the algorithm, we incorporate a multi-level framework, which has two interesting features: one is to remove artifacts in the solution and improve the resolution, another is to speed up the computation.

Since OSKCA does not depend on the imaging modality, it may be applied to solve other inverse source problems in imaging.

# REFERENCES


[1] Adams, R. A. and Fournier, J. J., *Sobolev spaces*, vol. 140. Academic press, 2003.

[2] Adcock, B., "Multivariate modified fourier series and application to boundary value problems," *Numerische Mathematik*, vol. 115, no. 4, pp. 511–552, 2010.

[3] Adcock, B., "Convergence acceleration of modified fourier series in one or more dimensions," *Mathematics of Computation*, vol. 80, no. 273, pp. 225–261, 2011.

[4] Ammari, H. and Bao, G., "Analysis of the scattering map of a linearized inverse medium problem for electromagnetic waves," *Inverse Problems*, vol. 17, no. 2, p. 219, 2001.

[5] Arridge, S. R., "Optical tomography in medical imaging," *Inverse Problems*, vol. 15, pp. R41–R93, Apr. 1999.

[6] Arridge, S. R., Cope, M., and Delpy, D. T., "The theoretical basis for the determination of optical pathlengths in tissue: temporal and frequency analysis," *Physics in Medicine and Biology*, vol. 37, pp. 1531–1560, July 1992.

[7] Arridge, S. R. and Hebden, J. C., "Optical imaging in medicine: II. modelling and reconstruction," *Physics in Medicine and Biology*, vol. 42, pp. 841–853, May 1997.

[8] Arridge, S. R. and Schotland, J. C., "Optical tomography: forward and inverse problems," *Inverse Problems*, vol. 25, p. 123010, Dec. 2009.

[9] Badieirostami, M., Adibi, A., Zhou, H.-M., and Chow, S.-N., "Wiener chaos expansion and simulation of electromagnetic wave propagation excited by a spatially incoherent source," *Multiscale Modeling & Simulation*, vol. 8, no. 2, pp. 591–604, 2010.

[10] Bal, G., "Radiative transfer equations with varying refractive index: a mathematical perspective," *Journal of the Optical Society of America A*, vol. 23, no. 7, p. 1639, 2006.

[11] Bao, G., Chen, Y., and Ma, F., "Regularity and stability for the scattering map of a linearized inverse medium problem," *Journal of mathematical analysis and applications*, vol. 247, no. 1, pp. 255–271, 2000.

[12] Bao, G., Chow, S.-N., Li, P., and Zhou, H., "Numerical solution of an inverse medium scattering problem with a stochastic source," *Inverse Problems*, vol. 26, no. 7, p. 074014, 2010.

[13] Bao, G. and Li, P., "Inverse medium scattering problems for electromagnetic waves," *SIAM Journal on Applied Mathematics*, vol. 65, pp. 2049–2066, Jan. 2005.

[14] Beck, A. and Teboulle, M., "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, p. 183, 2009.

[15] Behrooz, A., Zhou, H.-M., Eftekhar, A. A., and Adibi, A., "Total variation regularization for 3D reconstruction in fluorescence tomography: experimental phantom studies," *Applied Optics*, vol. 51, pp. 8216–8227, Dec. 2012.

[16] Bertero, M., Mol, C. D., and Pike, E. R., "Linear inverse problems with discrete data: II. stability and regularisation," *Inverse Problems*, vol. 4, pp. 573–594, Aug. 1988.

[17] Beylkin, G., Coifman, R., and Rokhlin, V., "Fast wavelet transforms and numerical algorithms i," *Communications on Pure and Applied Mathematics*, vol. 44, pp. 141–183, Oct. 2006.

[18] Bleistein, N. and Cohen, J. K., "Nonuniqueness in the inverse source problem in acoustics and electromagnetics," *Journal of Mathematical Physics*, vol. 18, p. 194, 1977.

[19] Bornemann, F. A. and Deuflhard, P., "The cascadic multigrid method for elliptic problems," *Numerische Mathematik*, vol. 75, no. 2, p. 135152, 1996.

[20] Boyd, S. and Vandenberghe, L., *Convex optimization*. Cambridge university press, 2004.

[21] Bregman, L. M., Censor, Y., Reich, S., and Zepkowitz-Malachi, Y., "Finding the projection of a point onto the intersection of convex sets via projections onto half-spaces," *Journal of Approximation Theory*, vol. 124, pp. 194–218, Oct. 2003.

[22] Bremer, C., Ntziachristos, V., and Weissleder, R., "Optical-based molecular imaging: contrast agents and potential medical applications," *European radiology*, vol. 13, no. 2, pp. 231–243, 2003.

[23] Brenner, S. C. and Scott, L. R., *The mathematical theory of finite element methods*, vol. 15. Springer, 2008.

[24] Bruckstein, A. M., Elad, M., and Zibulevsky, M., "On the uniqueness of nonnegative sparse solutions to underdetermined systems of equations," *IEEE Transactions on Information Theory*, vol. 54, pp. 4813–4820, Nov. 2008.

[25] Cai, J.-F., Cands, E. J., and Shen, Z., "A singular value thresholding algorithm for matrix completion," *SIAM J. on Optimization*, vol. 20, p. 19561982, Mar. 2010.

[26] Cai, J.-F., Dong, B., Osher, S., and Shen, Z., "Image restoration: Total variation, wavelet frames, and beyond," *Journal of the American Mathematical Society*, vol. 25, no. 4, pp. 1033–1089, 2012.

[27] Censor, Y. and Tom, E., "Convergence of string-averaging projection schemes for inconsistent convex feasibility problems," *Optimization Methods and Software*, vol. 18, no. 5, pp. 543–554, 2003.

[28] Chambolle, A., De Vore, R. A., Lee, N.-Y., and Lucier, B. J., "Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage," *Image Processing, IEEE Transactions on*, vol. 7, no. 3, pp. 319–335, 1998.

[29] Chan, R. H., Riemenschneider, S. D., Shen, L., and Shen, Z., "Tight frame: an efficient way for high-resolution image reconstruction," *Applied and Computational Harmonic Analysis*, vol. 17, no. 1, pp. 91–115, 2004.

[30] Chang, S. G., Yu, B., and Vetterli, M., "Adaptive wavelet thresholding for image denoising and compression," *Image Processing, IEEE Transactions on*, vol. 9, no. 9, pp. 1532–1546, 2000.

[31] Chaudhari, A. J., Ahn, S., Levenson, R., Badawi, R. D., Cherry, S. R., and Leahy, R. M., "Excitation spectroscopy in multispectral optical fluorescence tomography: methodology, feasibility and computer simulation studies," *Physics in Medicine and Biology*, vol. 54, pp. 4687–4704, Aug. 2009.

[32] Chen, Y., Hager, W. W., Yashtini, M., Ye, X., and Zhang, H., "Bregman operator splitting with variable stepsize for total variation image reconstruction," *Comput. Optim. Appl*, 2012.

[33] Chen, Y., Hager, W., Huang, F., Phan, D., Ye, X., and Yin, W., "Fast algorithms for image reconstruction with application to partially parallel mr imaging," *SIAM Journal on Imaging Sciences*, vol. 5, no. 1, pp. 90–118, 2012.

[34] Chen, Z. and Liu, X., "An adaptive perfectly matched layer technique for time-harmonic scattering problems," *SIAM journal on numerical analysis*, p. 645671, 2006.

[35] Combettes, P. L. and Wajs, V. R., "Signal recovery by proximal forward-backward splitting," *Multiscale Modeling & Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.

[36] Cong, A. X. and Wang, G., "A finite-element-based reconstruction method for 3D fluorescence tomography," *Optics Express*, vol. 13, no. 24, p. 9847, 2005.

[37] Dai, Y.-H., "R-linear convergence of the barzilai and borwein gradient method," *IMA Journal of Numerical Analysis*, vol. 22, pp. 1–10, Jan. 2002.

[38] Dai, Y.-H. and Fletcher, R., "Projected barzilai-borwein methods for large-scale box-constrained quadratic programming," *Numerische Mathematik*, vol. 100, pp. 21–47, Mar. 2005.

[39] Daubechies, I., Defrise, M., and De Mol, C., "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on Pure and Applied Mathematics*, vol. 57, no. 11, p. 14131457, 2004.

[40] Daubechies, I., Defrise, M., and De Mol, C., "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Communications on pure and applied mathematics*, vol. 57, no. 11, pp. 1413–1457, 2004.

[41] Dehghani, H., Eames, M. E., Yalavarthy, P. K., Davis, S. C., Srinivasan, S., Carpenter, C. M., Pogue, B. W., and Paulsen, K. D., "Near infrared optical tomography using NIRFAST: algorithm for numerical model and image reconstruction," *Communications in Numerical Methods in Engineering*, vol. 25, pp. 711–732, June 2009.

[42] Ding, Z., "A proof of the trace theorem of sobolev spaces on lipschitz domains," *Proceedings of the American Mathematical Society*, vol. 124, no. 2, p. 591600, 1996.

[43] Dutta, J., Ahn, S., Joshi, A., and Leahy, R., "Optimal illumination patterns for fluorescence tomography," in *Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on*, pp. 1275–1278, IEEE, 2009.

[44] Dutta, J., Ahn, S., Li, C., Cherry, S. R., and Leahy, R. M., "Joint l1 and total variation regularization for fluorescence molecular tomography," *Physics in medicine and biology*, vol. 57, pp. 1459–1476, Mar. 2012. PMID: 22390906.

[45] Egger, H., Freiberger, M., and Schlottbom, M., "On forward and inverse models in fluorescence diffuse optical tomography," *Inverse Problems and Imaging*, vol. 4, no. 3, p. 411427, 2010.

[46] Engl, H., Hanke, M., and Neubauer, A., *Regularization of inverse problems*, vol. 375. Springer, 1996.

[47] Epstein, C. L., *Introduction to the mathematics of medical imaging*. Society for Industrial and Applied Mathematics, 2007.

[48] Esser, E., "Applications of lagrangian-based alternating direction methods and connections to split bregman," *CAM report*, vol. 9, p. 31, 2009.

[49] Evans, L. C., *Partial Differential Equations*, vol. 19 of *Graduate Studies in Mathematics*. Providence, Rhode Island: American Mathematical Society, 1998.

[50] Farrell, T. J., Patterson, M. S., and Wilson, B., "A diffusion theory model of spatially resolved, steady-state diffuse reflectance for the noninvasive determination of tissue optical properties in vivo," *Medical physics*, vol. 19, pp. 879–888, Aug. 1992. PMID: 1518476.

[51] Fefferman, C., "Pointwise convergence of fourier series," *Annals of Mathematics*, vol. 98, pp. 551–571, Nov. 1973.

[52] Figueiredo, M. A. and Nowak, R. D., "An em algorithm for wavelet-based image restoration," *Image Processing, IEEE Transactions on*, vol. 12, no. 8, pp. 906–916, 2003.

[53] Gao, H., Yu, H., Osher, S., and Wang, G., "Multi-energy CT based on a prior rank, intensity and sparsity model (PRISM)," *Inverse problems*, vol. 27, Nov. 2011. PMID: 22223929 PMCID: PMC3249839.

[54] Gao, H., Yu, H., Osher, S., and Wang, G., "Multi-energy CT based on a prior rank, intensity and sparsity model (PRISM)," *Inverse problems*, vol. 27, Nov. 2011. PMID: 22223929 PMCID: PMC3249839.

[55] Gao, H. and Zhao, H., "A fast-forward solver of radiative transfer equation," *Transport Theory and Statistical Physics*, vol. 38, no. 3, pp. 149–192, 2009.

[56] Gao, H. and Zhao, H., "Multilevel bioluminescence tomography based on radiative transfer equation part 1: l1 regularization," *Optics Express*, vol. 18, no. 3, pp. 1854–1871, 2010.

[57] Gao, H. and Zhao, H., "Multilevel bioluminescence tomography based on radiative transfer equation part 2: total variation and l1 data fidelity," *Optics Express*, vol. 18, pp. 2894–2912, Feb. 2010.

[58] Goldstein, T. and Osher, S., "The split bregman method for l1-regularized problems," *SIAM Journal on Imaging Sciences*, vol. 2, pp. 323–343, Jan. 2009.

[59] Golub, G. H., Heath, M., and Wahba, G., "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, no. 2, pp. 215–223, 1979.

[60] Golub, G. and Van Loan, C., *Matrix computations*, vol. 3. Johns Hopkins University Press, 1996.

[61] Gorodnitsky, I. and Rao, B., "Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm," *Signal Processing, IEEE Transactions on*, vol. 45, pp. 600 –616, Mar. 1997.

[62] Greer, J. B., Bertozzi, A. L., and Sapiro, G., "Fourth order partial differential equations on general geometries," *Journal of Computational Physics*, vol. 216, no. 1, pp. 216–246, 2006.

[63] Grter, M. and Widman, K.-O., "The green function for uniformly elliptic equations," *Manuscripta Mathematica*, vol. 37, no. 3, pp. 303–342, 1982.

[64] Hansen, P. C. and OLeary, D. P., "The use of the l-curve in the regularization of discrete ill-posed problems," *SIAM Journal on Scientific Computing*, vol. 14, pp. 1487–1503, Nov. 1993.

[65] Haskell, R. C., Svaasand, L. O., Tsay, T. T., Feng, T. C., McAdams, M. S., and Tromberg, B. J., "Boundary conditions for the diffusion equation in radiative transfer," *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, vol. 11, pp. 2727–2741, Oct. 1994. PMID: 7931757.

[66] Hawrysz, D. J. and Sevick-Muraca, E. M., "Developments toward diagnostic breast cancer imaging using near-infrared optical measurements and fluorescent contrast agents," *Neoplasia (New York, N.Y.)*, vol. 2, pp. 388–417, Sept. 2000. PMID: 11191107 PMCID: PMC1507982.

[67] Hebden, J. C., Arridge, S. R., and Delpy, D. T., "Optical imaging in medicine: I. experimental techniques," *Physics in Medicine and Biology*, vol. 42, pp. 825–840, May 1997.

[68] Herman, G. T., Lent, A., and Rowland, S. W., "ART: mathematics and applications: A report on the mathematical foundations and on the applicability to real data of the algebraic reconstruction techniques," *Journal of Theoretical Biology*, vol. 42, pp. 1–32, Nov. 1973.

[69] Hou, T. Y., Luo, W., Rozovskii, B., and Zhou, H.-M., "Wiener chaos expansions and numerical solutions of randomly forced equations of fluid mechanics," *Journal of Computational Physics*, vol. 216, no. 2, pp. 687–706, 2006.

[70] Ishimaru, A., *Wave propagation and scattering in random media*, vol. 2. IEEE press Piscataway, NJ, 1997.

[71] Jain, S., Tsiotras, P., and Zhou, H.-M., "A hierarchical multiresolution adaptive mesh refinement for the solution of evolution pdes," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 1221–1248, 2008.

[72] Jiang, M., Zhou, T., Cheng, J., Cong, W., and Wang, G., "Image reconstruction for bioluminescence tomography from partial measurement," *Optics Express*, vol. 15, pp. 11095–11116, Sept. 2007. PMID: 19547465.

[73] Kaipio, J. and Somersalo, E., "Statistical inverse problems: Discretization, model reduction and inverse crimes," *Journal of Computational and Applied Mathematics*, vol. 198, pp. 493–504, Jan. 2007.

[74] Kelly, S. E., "Gibbs phenomenon for wavelets," *Applied and Computational Harmonic Analysis*, vol. 3, no. 1, p. 7281, 1996.

[75] Ladyzhenskaia, O. A. and Uraltseva, N. N., *Linear and Quasilinear Elliptic Equations*. Academic Press, 1968.

[76] Lanzani, L. and Shen, Z., "On the robin boundary condition for laplace's equation in lipschitz domains," *Communications in Partial Differential Equations*, vol. 29, no. 1-2, pp. 91–109, 2005.

[77] Lukic, V., Markel, V. A., and Schotland, J. C., "Optical tomography with structured illumination," *Optics letters*, vol. 34, no. 7, p. 983985, 2009.

[78] Mallat, S., *A wavelet tour of signal processing: the sparse way*. Academic press, 2008.

[79] Markel, V. A., Mital, V., and Schotland, J. C., "Inverse problem in optical diffusion tomography. III. inversion formulas and singular-value decomposition," *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, vol. 20, pp. 890–902, May 2003. PMID: 12747436.

[80] Markel, V. A. and Schotland, J. C., "Inverse problem in optical diffusion tomography. II. role of boundary conditions," *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, vol. 19, pp. 558–566, Mar. 2002. PMID: 11876321.

[81] Ming, W. and Xu, J., "The Morley element for fourth order elliptic equations in any dimensions," *Numerische Mathematik*, vol. 103, no. 1, pp. 155–169, 2006.

[82] Neumaier, A., "Solving ill-conditioned and singular linear systems: A tutorial on regularization," *SIAM Review*, vol. 40, pp. 636–666, Jan. 1998.

[83] Notay, Y., "An aggregation-based algebraic multigrid method," *Electronic Transactions on Numerical Analysis*, vol. 37, no. 6, pp. 123–146, 2010.

[84] Ren, K., Gao, H., and Zhao, H., "A hybrid reconstruction method for quantitative PAT," *SIAM Journal on Imaging Sciences*, vol. 6, pp. 32–55, Jan. 2013.

[85] Ribes, A. and Schmitt, F., "Linear inverse problems in imaging," *IEEE Signal Processing Magazine*, vol. 25, pp. 84–99, July 2008.

[86] Ripoll, J., Ntziachristos, V., Carminati, R., and Nieto-Vesperinas, M., "Kirchhoff approximation for diffusive waves," *Physical Review. E, Statistical, Nonlinear, and Soft Matter Physics*, vol. 64, p. 051917, Nov. 2001. PMID: 11735978.

[87] Rudin, L. I., Osher, S., and Fatemi, E., "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, pp. 259–268, Nov. 1992.

[88] Rudin, W., *Functional analysis*. McGraw-Hill, Jan. 1991.

[89] Saad, Y., *Iterative Methods for Sparse Linear Systems*. SIAM, Apr. 2003.

[90] Schotland, J. C. and Markel, V. A., "Inverse scattering with diffusing waves," *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, vol. 18, pp. 2767–2777, Nov. 2001. PMID: 11688867.

[91] Schweiger, M., Arridge, S. R., Hiraoka, M., and Delpy, D., "The finite element method for the propagation of light in scattering media: Boundary and source conditions," *Medical Physics*, vol. 22, no. 11, p. 1779, 1995.

[92] Sezan, M. I. and Stark, H., "Image restoration by the method of convex projections: part 2 applications and numerical results," *IEEE Transactions on Medical Imaging*, vol. 1, no. 2, pp. 95–101, 1982. PMID: 18238262.

[93] Shen, J. and Wang, L.-L., "Sparse spectral approximations of high-dimensional problems based on hyperbolic cross," *SIAM Journal on Numerical Analysis*, vol. 48, pp. 1087–1109, Jan. 2010.

[94] Strang, G. and Nguyen, T., *Wavelets and filter banks*. Wellesley Cambridge Press, 1996.

[95] Tarvainen, T., Kolehmainen, V., Kaipio, J. P., and Arridge, S. R., "Corrections to linear methods for diffuse optical tomography using approximation error modelling," *Biomedical Optics Express*, vol. 1, pp. 209–222, July 2010. PMID: 21258459 PMCID: PMC3005182.

[96] Temlyakov, V., "Approximation of functions of several variables by trigonometric polynomials with harmonics from hyperbolic crosses," *Ukrainian Mathematical Journal*, vol. 41, no. 4, pp. 451–456, 1989.

[97] Tikhonov, A., "Solution of incorrectly formulated problems and the regularization method," in *Soviet Math. Dokl.*, vol. 5, p. 1035, 1963.

[98] Trefethen, L. N., *Spectral methods in MATLAB*, vol. 10. Society for Industrial and Applied Mathematics, 2000.

[99] Vogel, C. R., "A multigrid method for total variation-based image denoising," *Progress in Systems and Control Theory*, vol. 20, p. 323323, 1995.

[100] Wang, G., Li, Y., and Jiang, M., "Uniqueness theorems in bioluminescence tomography," *Medical Physics*, vol. 31, no. 8, p. 2289, 2004.

[101] Wang, L. V. and Wu, H.-i., *Biomedical optics: principles and imaging*. Wiley-Interscience, 2007.

[102] Wang, Y., Yang, J., Yin, W., and Zhang, Y., "A new alternating minimization algorithm for total variation image reconstruction," *SIAM Journal on Imaging Sciences*, vol. 1, no. 3, p. 248, 2008.

[103] Weissleder, R., Tung, C., Mahmood, U., and Bogdanov, A., "In vivo imaging of tumors with protease-activated near-infrared fluorescent probes," *Nature Biotechnology*, vol. 17, no. 4, pp. 375–378, 1999.

[104] Wu, C. and Tai, X.-C., "Augmented Lagrangian method, dual methods, and split Bregman iteration for ROF, vectorial TV, and high order models," *SIAM Journal on Imaging Sciences*, vol. 3, pp. 300–339, Jan. 2010.

[105] Xiu, D. and Karniadakis, G. E., "The Wiener–Askey polynomial chaos for stochastic differential equations," *SIAM Journal on Scientific Computing*, vol. 24, no. 2, pp. 619–644, 2002.