# Solving Inverse Source Problems by the Orthogonal Solution and Kernel Correction Algorithm (OSKCA) with Applications in Fluorescence Tomography

Shui-Nee Chow \* Ke Yin<sup> $\dagger$ \*</sup> Hao-Min Zhou \* Ali Behrooz <sup>‡</sup>

#### Abstract

We present a new approach to solve the inverse source problem arising in Fluorescence Tomography (FT). In general, the solution is non-unique and the problem is severely illposed. It poses tremendous challenges in image reconstructions. In practice, the most widely used methods are based on Tikhonov-type regularizations, which minimize a cost function consisting of a regularization term and a data fitting term. We propose an alternative method which overcomes the major difficulties, namely the non-uniqueness of the solution and noisy data fitting, in two separate steps. 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 other regularity and physical requirements. The key ideas are that the correction function in the kernel has no impact on the data fitting, so that there is no parameter needed to balance the data fitting and additional constraints on the solution. Moreover, we use an efficient basis to represent the source function, and introduce a hybrid strategy combining spectral methods and finite element methods in the proposed algorithm. The resulting algorithm can dramatically increase the computation speed over the existing methods. Also the numerical evidence shows that it significantly improves the image resolution and robustness against noise.

# 1 Introduction

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, such as CT and MRI, FT uses harmless non-ionizing near-infrared (NIR) radiation (instead of X-ray or powerful magnetic field), and highly specific fluorescent probes to capture molecular specific information that cannot be obtained otherwise [25]. For this reason, it is considered to be a promising method in early cancer detection and drug monitoring [24, 39, 5].

In the experimental setup of FT (See Figure 1 for a cartoon demonstration),

NIR radiation (wavelength 650-900 nm) is pumped into the examined 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 NIR light at a longer wavelength, which propagates in the tissue. Then the intensity of the fluorescent emission

<sup>\*</sup>School of Mathematics, Georgia Institute of Technology, 686 Cherry Street, Atlanta, GA 30332-0160, USA. {chow,kyin,hmzhou}@math.gatech.edu. This work is partially supported by NSF Faculty Early Career Development (CAREER) Award DMS-0645266 and DMS-1042998.

<sup>&</sup>lt;sup>†</sup>Corresponding author.

<sup>&</sup>lt;sup>‡</sup>School of Electrical and Computer Engineering, Georgia Institute of Technology, 777 Atlantic Drive NW Atlanta GA 30332 USA. alibehrooz@gatech.edu.



Figure 1: An illustration of Fluorescence Tomography.<sup>1</sup>

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) [27]. 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 [1, 35, 14, 13]. The boundary condition associated with the DA model is Robin (mixed) type [17, 23, 36, 30], which accounts for the partial reflection and transmission of the light on the boundary.

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

Mathematically, finding the distribution of the fluorophores can be written as an inverse source problem,

$$g = Af + \eta, \tag{1}$$

where f represents the distribution of fluorophores which is the unknown. The noise  $\eta$  mainly comes from inaccurate modeling, which includes estimates of the parameters of the biological tissues and the approximation errors caused by using DA. The noise also contains the measurement error. It will be defined by the DA model that A is the operator mapping f to the boundary measurement g, which can be computed by solving the DA equation. We describe it in detail in the next section.

It is well known that the solution to (1) is non-unique and the problem is severely ill-posed [37]. This leads to many difficulties in its numerical reconstructions. The two major challenges are: 1) additional physical restrictions become necessary to uniquely determine the solution. Commonly used restrictions in FT include positivity constrains and certain regularity (stability) requirements. 2) Noise and inevitable rounding perturbations in numerical computations may cause huge artifacts in the reconstructions.

To cope with these challenges in FT applications, Tikhonov-type regularization is often applied, which can be written in the following form

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

<sup>&</sup>lt;sup>1</sup>This illustration was generated by and presented in P. Mohajerani, et al., Proc. of SPIE Vol. 7174 717413-2.

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, where  $\Psi_{\text{reg}}(f)$  is the smooth functional. Recently, some non-smooth regularization terms have become popular, such as  $L^1$  norm  $||f||_1$ , which promotes the sparsity [18], and total variation (TV) semi-norm  $||f||_{\text{TV}}$ , which tends to preserve edges in the image [14, 4]. In order to get the benefits of both, authors in [13, 19] 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) [14].

In all 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 [22] is a common strategy for the selection of the parameter  $\mu$ .

In this paper, 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 are many choices, but we choose the one that is robust against noise. 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 regularization methods.

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

$$f = f^* + f^0, (3)$$

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

$$f^* = A^* (AA^*)^{-1} g, (4)$$

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 [15].

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, e.g, the iterative Tikhonov method [31]. The result produced by the iterative regularization method is robust against noise, and we use it as a particular solution which has good data fitting. Obviously, that particular solution may not satisfy some regularity requirements from physics and prior information in the applications, such as positivity, visually smooth features, and 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

$$\hat{f}^{0} = \underset{f^{0} \in \mathcal{N}(A)}{\operatorname{argmin}} \Psi_{\operatorname{reg}}(f^{*} + f^{0}) \quad \text{such that } f^{*} + f^{0} \ge 0,$$
(5)

where the regularization functional  $\Psi_{\text{reg}}$  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.

In this paper, we call the proposed two-step approach the Orthogonal Solution and Kernel Correction Algorithm (OSKCA). Our contribution is that we select a stable solution  $f^*$  that achieves data fitting in the first step, and correct it in the kernel space by adding  $f^0$  in the second step, so that the final solution  $f = f^* + f^0$  satisfies other constraints, while preserving good data fitting of  $f^*$ .

The two steps (4) and (5) in our algorithm can be solved by existing methods. For example, several algorithms can be used to solve (4), such as the gradient based Landweber iterations [3]. However, the ill-conditioning in A can cause very slow convergence in FT application. To address the problem, we develop a method to compute  $(AA^*)^{-1}g$  by iterative Tikhonov regularization [31]. 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 [20].

In addition, to further speed up the computation, we introduce different bases to represent the fluorophore distribution f in our algorithms. Most of the existing methods for FT applications use finite element method (FEM) to solve the differential equations in the model. Naturally, the FEM basis, consisting of nodal functions (point-wise basis functions), are used to represent the unknown. We propose to use other more efficient bases, such as Harmonic functions or wavelets, 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.

The organization of the paper is as follows: in Section 2 we present the governing equation for the light propagation, as well as the forward and inverse problems. In Section 3, we describe the framework of our proposed algorithm. In Section 4, we discuss the details of the algorithm, including the computation of orthogonal solution  $f^*$  as well as the correction in the kernel  $f^0$ . Finally we present numerical examples including comparison with existing methods in Section 5. A brief conclusion is given in the last section.

# 2 Mathematical Models for FT

## 2.1 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. The DA equation is well accepted as an accurate model when the light propagation is highly scattering and low absorptive, such as the biological tissues [1]. The DA is given by

$$\frac{\partial\phi(r,t)}{c\partial t} + \mu_a\phi(r,t) - \nabla\cdot[\kappa\nabla\phi(r,t)] = q(r,t),\tag{6}$$

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.

In the frequency domain, we have the equation for the coefficients corresponding to the modulation frequency  $\omega$  [2]:

$$-\nabla \cdot [\kappa \nabla \Phi(r,\omega)] + (\mu_a + \frac{i\omega}{c})\Phi(r,\omega) = q(r,\omega) \quad \text{in } \Omega,$$
(7)

where  $\Omega$  is the region occupied by the examined tissue. We note  $\Phi(r, \omega)$  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

$$-\nabla \cdot [\kappa \nabla \Phi(r)] + \mu_a \Phi(r) = q(r) \quad \text{in } \Omega.$$
(8)

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 [36]:

$$\vec{n} \cdot [\kappa \nabla \Phi(r)] + R \Phi(r) = 0 \quad \text{on } \partial \Omega,$$
(9)

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

For the simplicity of notations, the DA equation (7) with Robin boundary condition (9) is written in short as

$$F(\kappa, \mu_a, R)\Phi = q, \text{ or } F\Phi = q.$$
(10)

From the classical PDE theory [29],  $\Phi$  in equation (10) has a unique solution in the Sobolev space  $H^1(\Omega)$  given  $q(\cdot) \in L^2(\Omega)$  and  $\Omega$  is a Lipschitz domain.

### 2.2 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

$$F_m \Phi_m(r) = \Phi_x(r) f(r), \tag{11}$$

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

$$F_x \Phi_x(r) = q(r), \tag{12}$$

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

$$g = T_r \Phi_m. \tag{13}$$

By (11) and (13), the forward model that maps unknown f to measurements g is formulated as

$$g = T_r F_m^{-1}(\Phi_x f), \tag{14}$$

where  $F_m^{-1}(\Phi_x f) \in H^1(\Omega)$  is the solution of (11).

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.

### 2.3 FT model

In order to improve the conditioning of the forward model in (14), multiple light sources are used for excitation [12]. 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 (14), we have

$$g^{(i)} = T_r F_m^{-1}(\Phi_x^{(i)} f), (15)$$

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

$$g = Af, \tag{16}$$

where

$$g = \begin{bmatrix} g^{(1)} \\ \vdots \\ g^{(s)} \end{bmatrix}, \quad Af = \begin{bmatrix} T_r F_m^{-1}(\Phi_x^{(1)}f) \\ \vdots \\ T_r F_m^{-1}(\Phi_x^{(s)}f) \end{bmatrix}.$$
(17)

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.

# 3 Proposed method: Orthogonal Solution with Kernel Correction Algorithm (OSKCA)

It is shown in [37] that the solution of (16) is non-unique. Also, it is severely ill-conditioned. Because 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, regularization techniques are often needed to handle the ill-posedness of the problem. 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 costly. 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 OSKCA to compute the solution in two steps. In this section, we describe the algorithm in detail.

We notice that any solution f to (16) can be decomposed as

$$f = f^* + f^0, (18)$$

where  $f^*$  is a particular solution, and  $f^0$  is a function in the kernel  $\mathcal{N}(A)$ . Then to solve the equation (16) we just need to determine  $f^*$  and  $f^0$ .

First, we choose  $f^*$  such that it satisfies

$$g = Af^* \text{ and } f^* \in \mathcal{N}(A)^{\perp},$$
(19)

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 3.1 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.,

$$\hat{f}^{0} = \underset{f^{0} \in \mathcal{N}(A)}{\operatorname{argmin}} \Psi_{\operatorname{reg}}(f^{*} + f^{0}) \quad \text{such that} \quad f^{*} + f^{0} \ge 0,$$
 (20)

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

In theory, OSKCA is equivalent to solving the optimization problem

$$\hat{f} = \underset{f \ge 0}{\operatorname{argmin}} \Psi_{\operatorname{reg}}(f) \quad \text{such that} \quad Af = g.$$
 (21)

Compared with Tikhonov regularization (2), the data fitting in (21) can be enforced without jeopardizing the regularity requirements. This is a desirable property, because A is severely underdetermined. Also, the equality constraint in (21) 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.

#### 3.1 The orthogonal solution

The orthogonal solution to the inverse problem g = Af is

$$f^* = A^* (AA^*)^{-1} g, (22)$$

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 [34].

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 (22) is strictly applicable. Though it is not needed in the computation, since the algorithm for (22) 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 4.2.

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

$$\bar{f} = A^{\dagger}g_{\dagger}$$

where  $A^{\dagger}$  is the Moore-Penrose (generalized) inverse of A (See [15]). The minimal norm solution and the orthogonal solution are closely related, and their connection is stated in the following:

**Proposition 3.1.** 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 (22), and the two solutions coincide when  $g \in \mathcal{R}(AA^*)$ .

#### 3.2 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 coefficients for  $f^0$  under the basis W. The problem of finding the kernel correction can be written as a constrained optimization problem

$$\hat{\mu} = \underset{\mu}{\operatorname{argmin}} \Psi_{\operatorname{reg}}(f^* + W\mu) \quad \text{subject to} \quad f^* + W\mu \ge 0, \tag{23}$$

where  $\Psi_{\text{reg}}$  is a regularization functional. Here we note that (23) is equivalent to (21). 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 (23) becomes

$$\hat{\mu} = \underset{\mu}{\operatorname{argmin}} \|f^* + W\mu\| \quad \text{subject to} \quad f^* + W\mu \ge 0.$$
(24)

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

$$\hat{\mu} = \underset{\mu}{\operatorname{argmin}} \| f^* + W\mu \|_{TV} \quad \text{subject to} \quad f^* + W\mu \ge 0, \tag{25}$$

Likewise, other regularization 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.

#### 3.3 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

$$f = \sum_{i=1}^{\infty} c_i \xi_i,\tag{26}$$

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

$$B: (c_1, \dots, c_i, \dots) \mapsto \sum_{i=1}^{\infty} c_i \xi_i.$$
(27)

Then (26) implies the relation

$$f = Bc. (28)$$

By (28), the inverse problem in (16) is rewritten as

$$g = ABc, \text{ or } g = Mc, \tag{29}$$

where M denotes the composition of A, B.

The idea of OSKCA still applies to the new formulation (29). c is decomposed as

$$c = c^* + c^0, (30)$$

where  $c^*$  is the orthogonal solution to (29) 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

$$e^0 = K\lambda. \tag{31}$$

By (28)(30), the solution to (16) is written as

$$f = B(c^* + c^0). (32)$$

By (31), (32) is equivalent to

$$f = B(c^* + K\lambda). \tag{33}$$

B is determined by the chosen basis functions and K can be computed if M is given.  $c^*$  is computed by an analogy to (22), which is written as

$$c^* = M^* (MM^*)^{-1} g. aga{34}$$

We have the following lemma relating  $c^*$  to  $f^*$ .

**Lemma 3.2.** If the basis functions  $\{\xi_i\}_{i=1}^{\infty}$  are orthonormal, then

$$f^* = Bc^*. aga{35}$$

*Proof.* By the condition in the lemma, B as defined in (27) is an unitary transformation, and its adjoint operator  $B^*$  satisfies

$$BB^* = I(\text{identity}). \tag{36}$$

In this case,  $c^*$  as defined in (29) can be written as

$$c^* = (AB)^* (ABB^*A^*)^{-1}g = B^*A^* (AA^*)^{-1}g,$$
(37)

 $\mathbf{SO}$ 

$$Bc^* = BB^*A^*(AA^*)^{-1}g = A^*(AA^*)^{-1}g = f^*.$$
(38)

In practice, we choose orthonormal basis functions. In general, if the basis functions are not orthonormal, (35) does not hold. However, (32) is always satisfied.

The computation of the kernel correction term  $c^0$  follows the same way as (23). We note the solution f in the form (33), where  $\lambda$  is obtained by solving the optimization problem

$$\hat{\lambda} = \underset{\lambda}{\operatorname{argmin}} \Psi_{\operatorname{reg}}(B(c^* + K\lambda)) \quad \text{subject to} \quad B(c^* + K\lambda) \ge 0.$$
(39)

Here  $\Psi_{\text{reg}}(\cdot)$  also denotes the regularization functional as in (23).

In the following section, we address the implementation issues in detail.

# 4 Implementations

#### 4.1 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 (11) and (12), which are second order elliptic differential equations. They can be solved numerically by Finite Element Method (FEM) [6]. Let  $n_p$  be the number of nodes in the mesh for FEM. Suppose  $\{\delta_1, \ldots, \delta_{n_p}\}$  are the associated shape functions, which are actually the point-wise basis functions for the solution space in FEM. Under this basis, f is written as

$$f = \sum_{j=1}^{n_p} f_j \delta_j,\tag{40}$$

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

$$\mathbf{g} = \mathbf{A}\mathbf{f},\tag{41}$$

where  $\mathbf{f} = [f_1, \dots, f_{n_p}]^{\top}$ .  $\mathbf{g}$  is the discrete form of the boundary measurement g as defined in (17).  $\mathbf{A}$  is called the forward model matrix, which is the matrix form of the operator A in (16). By (17)(40), we have

$$\mathbf{A} = \begin{bmatrix} T_r F_m^{-1}(\Phi_x^{(1)}\delta_1) & \dots & T_r F_m^{-1}(\Phi_x^{(1)}\delta_{n_p}) \\ \vdots & \ddots & \vdots \\ T_r F_m^{-1}(\Phi_x^{(s)}\delta_1) & \dots & T_r F_m^{-1}(\Phi_x^{(s)}\delta_{n_p}) \end{bmatrix},$$
(42)

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 (12):

$$F_m u_{ij} = \Phi_x^{(i)} \delta_j, \tag{43}$$

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 d detector locations. Then  $T_r F_m^{-1}(\Phi_x^{(i)}\delta_j)$  is an  $d \times 1$  vector. Therefore, the forward model matrix **A** as defined in (42) has the size  $sd \times n_p$ .

As discussed in Section 3.3, 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 (26) and truncate it at the *n*-th term as

$$f = \sum_{i=1}^{n} c_i \xi_i. \tag{44}$$

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 (40).

By (29), the discrete form of the inverse problem is

$$\mathbf{g} = \mathbf{M}\mathbf{c},\tag{45}$$

where  $\mathbf{c} = [c_1, \ldots, c_n]^\top$ ,  $\mathbf{g}$  is the same as in (41), and  $\mathbf{M}$  is the forward model matrix corresponding to the new basis. Similar to (42), we write

$$\mathbf{M} = \begin{bmatrix} T_r F_m^{-1}(\Phi_x^{(1)}\xi_1) & \dots & T_r F_m^{-1}(\Phi_x^{(1)}\xi_n) \\ \vdots & \ddots & \vdots \\ T_r F_m^{-1}(\Phi_x^{(s)}\xi_1) & \dots & T_r F_m^{-1}(\Phi_x^{(s)}\xi_n) \end{bmatrix},$$
(46)

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 (46) is still obtained by solving the corresponding DA equations using the same FEM solver as (43).  $T_r$  has the same definition as in (42). **M** as defined above has the size  $sd \times n$ .

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 **M**, compared to  $sn_p$  for **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 (27) has the discrete form as

$$\mathbf{B} = [\xi_1, \dots, \xi_n],\tag{47}$$

which is an  $n_p \times n$  matrix. Then by (28), we can recover the point-wise representation of f by

$$\mathbf{f} = \mathbf{B}\mathbf{c},\tag{48}$$

where **c** is the solution to (45). In the implementation, we use the formulation (45) for the discrete form of the inverse problem (16) and the final solution f is expressed by (48).

### 4.2 Computation of the orthogonal solution

The orthogonal solution to (45) is given by (34) after discretization. As mentioned earlier, regularization is needed to compute the orthogonal solution. Various regularization methods are proposed to compute the minimal norm solution [15], which is the orthogonal solution in our problem. We present first the Landweber iteration that is used in inverse scattering problems [3]. 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

$$\mathbf{c}^{(k)} = \mathbf{c}^{(k-1)} + \omega \mathbf{M}^* (\mathbf{g} - \mathbf{M} \mathbf{c}^{(k-1)}), \tag{49}$$

where  $0 < \omega < (\|\mathbf{M}^*\|_2 \|\mathbf{M}\|_2)^{-1}$  is a relaxation parameter which ensures that (49) defines a contraction mapping. By the Banach fixed-point theorem, if  $g \in \mathcal{R}(\mathbf{M})$ , then (49) produces a sequence  $\{\mathbf{c}^{(k)}\}$  converging to a fixed-point, which turns out to be the orthogonal norm solution to (45).

Landweber iteration converges to the true solution as the number of iterations go to infinity, but when it is applied to the problem (45) in the FT applications, the convergence is slow due to the ill-conditioning of  $\mathbf{M}$ . Also, if noise presents in  $\mathbf{g}$ , the iterations do not converge correctly, or even diverge [15].

Iterated Tikhonov regularization is another well-known approach [31] to compute the orthogonal solution. It iteratively refines the current solution by applying Tikhonov regularization to the residual equation. The iteration scheme is as follows:

$$\mathbf{c}^{(0)} = 0, \quad \mathbf{r}^{(0)} = \mathbf{g},$$
 (50)

and for k = 1, 2, ...

$$\mathbf{c}^{(k)} = \mathbf{c}^{(k-1)} + \mathbf{M}^* (\mathbf{M}\mathbf{M}^* + h^2 \mathbf{I})^{-1} \mathbf{r}^{(k-1)}, \quad \mathbf{r}^{(k)} = \mathbf{g} - \mathbf{M}\mathbf{c}^{(k)}, \tag{51}$$

where h is the regularization parameter, and **I** is the identity matrix. There are two layers of iterations in (51). 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, ( $\mathbf{MM}^* + h^2\mathbf{I}$ )<sup>-1</sup> is actually implemented by iterative methods such as GMRES or CG [21] that form the inner iterations, which converge linearly. Unlike Landweber iteration which converges to the exact solution, iterative Tikhonov regularization computes a smooth approximation to it. 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 (22).

#### 4.3 Computation of the kernel space

Let **K** be a matrix with columns an orthonormal basis for  $\mathcal{N}(\mathbf{M})$ . Suppose the size of **K** is  $n \times m$ , where *n* is equal to the number of columns of **M**. *n* is always larger than *m*. We use **K** to represent the computed kernel space. However, it is unstable to compute **K** by solving  $\mathcal{N}(\mathbf{M})$  directly. We note that each row of **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 **M** are nearly identical, which makes **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 **M** is much smaller than the number of rows in **M**. It is very unstable to compute  $\mathcal{N}(\mathbf{M})$  directly if **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  $(sd \times n)$  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 [7]. It is formulated as the convex optimization problem

$$\tilde{\mathbf{M}} = \underset{\mathbf{X} \in \mathbb{R}^{sd \times n}}{\operatorname{argmin}} \tau \|\mathbf{X}\|_* + \frac{1}{2} \|\mathbf{M} - \mathbf{X}\|_F^2,$$
(52)

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

$$\tilde{\mathbf{M}} = \underset{\mathbf{X} \in \mathbb{R}^{sd \times n}}{\operatorname{argmin}} \operatorname{rank}(\mathbf{X}) + \frac{\rho}{2} \|\mathbf{M} - \mathbf{X}\|_{F}^{2},$$
(53)

which is intractable. In contrast, the problem (52) can be solved very efficiently. Suppose

$$\mathbf{M} = \mathbf{U} \Sigma \mathbf{V}^{\top} \tag{54}$$

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

$$\mathcal{D}_{\tau}(\Sigma) = \operatorname{diag}(\{(\sigma_i - \tau)_+\}),\tag{55}$$

where

$$(\sigma_i - \tau)_+ = \max(\sigma_i - \tau, 0). \tag{56}$$

Then the problem (52) has the explicit solution

$$\tilde{\mathbf{M}} = \mathbf{U}\mathcal{D}_{\tau}(\Sigma)\mathbf{V}^{\top} = \tilde{\mathbf{U}}\tilde{\Sigma}\tilde{\mathbf{V}}^{\top},\tag{57}$$

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 4.1. Suppose M has r nonzero singular values, then K satisfies

$$\frac{\|\mathbf{M}\mathbf{K}\|_F}{\|\mathbf{M}\|_F} \le \frac{\sqrt{\sum_{i=1}^r \min(\sigma_i, \tau)^2}}{\sqrt{\sum_{i=1}^r \sigma_i^2}}.$$
(58)

Proof.

$$\begin{aligned} \|\mathbf{M}\mathbf{K}\|_{F}^{2} &\leq \|(\mathbf{M} - \tilde{\mathbf{M}})\mathbf{K}\|_{F}^{2} + \|\tilde{\mathbf{M}}\mathbf{K}\|_{F}^{2} \\ &= \|(\mathbf{M} - \tilde{\mathbf{M}})\mathbf{K}\|_{F}^{2} \quad (\tilde{\mathbf{M}}\mathbf{K} = 0) \\ &\leq \|(\mathbf{M} - \tilde{\mathbf{M}})\|_{F}^{2} \quad (\mathbf{K} \text{ has orthonormal columns}) \\ &\leq \sum_{i=1}^{r} \min(\sigma_{i}, \tau)^{2} \text{ (by definition of } \tilde{\mathbf{M}} \text{ in } ((57)). \end{aligned}$$

Note that

$$\|\mathbf{M}\|_{F}^{2} = \sum_{i=1}^{r} \sigma_{i}^{2},\tag{59}$$

the inequality in this theorem follows.

In practice, we select  $\tau$  to be

$$\tau = \epsilon \sqrt{\frac{1}{r} \sum_{i=1}^{r} \sigma_i^2} = \epsilon \sqrt{\frac{1}{r} \|\mathbf{M}\|_F^2},\tag{60}$$

where  $\epsilon$  is a small number. By Theorem 4.1, **K** satisfies

$$\frac{\|\mathbf{M}\mathbf{K}\|_F}{\|\mathbf{M}\|_F} \le \epsilon.$$
(61)

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

#### 4.4 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.

#### 4.4.1 The positivity constraint for the kernel correction

One important regularity requirement of the solution is the positivity constraint. By (33), after discretization, the point-wise representation of f is

$$\mathbf{f} = \mathbf{B}(\mathbf{c}^* + \mathbf{K}\boldsymbol{\lambda}). \tag{62}$$

Then the positivity constraint can be written as discrete form

$$\mathbf{B}(\mathbf{c}^* + \mathbf{K}\boldsymbol{\lambda}) \ge \mathbf{0},\tag{63}$$

where  $\mathbf{B} \in \mathbb{R}^{n_p \times n}$  is computed in (47),  $\mathbf{c}^* \in \mathbb{R}^n$  is computed by (51),  $\mathbf{K} \in \mathbb{R}^{n \times m}$  is given by the algorithm described in Section 4.3.  $\boldsymbol{\lambda} \in \mathbb{R}^m$  is the unknown.

The feasible points of (63) may not be easy to find. The Algebraic Reconstruction Technique (ART) [26] 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 (63). In the following, we apply ART to (63).

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  $\lambda$  is updated via

$$\boldsymbol{\lambda}_{j+1} = \begin{cases} \boldsymbol{\lambda}_j & \text{if } b_i \ge (\mathbf{h}_i, \boldsymbol{\lambda}_j) \\ \boldsymbol{\lambda}_j + \alpha_j \frac{b_i - (\mathbf{h}_i, \boldsymbol{\lambda}_j)}{\|\mathbf{h}_i\|^2} \mathbf{h}_i & \text{if } b_i < (\mathbf{h}_i, \boldsymbol{\lambda}_j) \end{cases}$$
(64)

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  $\boldsymbol{\lambda}$  is smaller than a prescribed value.

Due to the noise in the measurements, (63) may be infeasible. In that case, the cyclic convergence of ART will happen [8].

#### 4.4.2 The TV minimization for the kernel correction

TV minimization has been demonstrated to have edge-preserving property in image recovery [33]. This approach can be incorporated in our framework, which is addressed in (25). After discretization and change of basis (39), it can be proposed as the optimization problem

$$\underset{\boldsymbol{\lambda}}{\operatorname{argmin}} \| \mathbf{B}(\mathbf{c}^* + \mathbf{K}\boldsymbol{\lambda}) \|_{TV} \quad \text{subject to} \quad \mathbf{B}(\mathbf{c}^* + \mathbf{K}\boldsymbol{\lambda}) \ge 0, \tag{65}$$

where the computation of  $\mathbf{c}^*$  and  $\mathbf{K}$  are described before. Inspired by the operator splitting technique [38], we introduce two auxiliary variables  $\mathbf{f}$  and  $\mathbf{w}$ .  $\mathbf{f}$  is defined in (62) 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 (65) is

$$\underset{\boldsymbol{\lambda}, \mathbf{f} \ge 0, \mathbf{w}}{\operatorname{argmin}} \| \mathbf{w} \|_{1} \quad \text{subject to} \begin{cases} \mathbf{f} = \mathbf{B}(\mathbf{c}^{*} + \mathbf{K}\boldsymbol{\lambda}) \\ \mathbf{w} = \mathbf{D}\mathbf{f} \end{cases}$$
(66)

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 (66), which is written as [40]

$$\mathscr{L}(\boldsymbol{\lambda}, \mathbf{f}, \mathbf{w}, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2) = \alpha \|\mathbf{w}\|_1 + (\boldsymbol{\mu}_1, \mathbf{D}\mathbf{f} - \mathbf{w}) + \frac{\rho_1}{2} \|\mathbf{D}\mathbf{f} - \mathbf{w}\|^2 + (\boldsymbol{\mu}_2, \mathbf{f} - \mathbf{B}(\mathbf{c}^* + \mathbf{K}\boldsymbol{\lambda})) + \frac{\rho_2}{2} \|\mathbf{f} - \mathbf{B}(\mathbf{c}^* + \mathbf{K}\boldsymbol{\lambda})\|^2,$$
(67)

which is called the augmented Lagrangian functional. The equivalent saddle point problem for (66) is written as

$$\min_{\boldsymbol{\lambda}, \mathbf{f} \ge 0, \mathbf{w}} \max_{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2} \mathscr{L}(\boldsymbol{\lambda}, \mathbf{f}, \mathbf{w}, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$$
(68)

Here  $\alpha, \rho_1, \rho_2$  are the regularization parameters that are selected by the user, and  $\mu_1, \mu_2$  are the Lagrange multipliers. The saddle point problem (68) can be solved iteratively [16], which is described in Algorithm 4.1. 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 4.1 OSKCA with TV minimization for kernel correction (OSCKA-TV)

Input: B, K, c<sup>\*</sup>,  $\alpha$ ,  $\rho_1$ ,  $\rho_2$ ,  $\epsilon_0$ Output: f Initialization:  $\mathbf{f}^{(0)} = \mathbf{0}$ ,  $\lambda^{(0)} = \mathbf{0}$ ,  $\mu_1^{(0)} = \mathbf{0}$ ,  $\mu_2^{(0)} = \mathbf{0}$ while  $\frac{\|\mathbf{f}^{(k)} - \mathbf{f}^{(k-1)}\|}{\|\mathbf{f}^{(k-1)}\|} \leq \epsilon_0$  do 1.  $\mathbf{w}^{(k+1)} = \operatorname*{argmin}_{\mathbf{w}} \alpha \|\mathbf{w}\|_1 + \frac{\rho_1}{2} \|\mathbf{D}\mathbf{f}^{(k)} - \mathbf{w} + \frac{\mu_1^{(k)}}{\rho_1}\|^2$ . 2.  $\mathbf{f}^{(k+1)} = \operatorname*{argmin}_{\mathbf{f} \geq 0} \frac{\rho_1}{2} \|\mathbf{D}\mathbf{f} - \mathbf{w}^{(k+1)} + \frac{\mu_1^{(k)}}{\rho_1}\|^2 + \frac{\rho_2}{2} \|\mathbf{f} - \mathbf{B}(\mathbf{c}^* + \mathbf{K}\lambda) + \frac{\mu_2^{(k)}}{\rho_2}\|^2$ . 3.  $\lambda^{(k+1)} = \operatorname*{argmin}_{\lambda} \|\mathbf{f}^{(k+1)} - \mathbf{B}(\mathbf{c}^* + \mathbf{K}\lambda) + \frac{\mu_2^{(k)}}{\rho_2}\|^2$ . 4.  $\mu_1^{(k+1)} = \mu_1^{(k)} + \rho_1(\mathbf{D}\mathbf{f}^{(k+1)} - \mathbf{w}^{(k+1)})$ . 5.  $\mu_2^{(k+1)} = \mu_2^{(k)} + \rho_2(\mathbf{f}^{(k+1)} - \mathbf{B}(\mathbf{c}^* + \mathbf{K}\lambda^{(k+1)}))$ . end while

We note that in each iteration of Algorithm 4.1, Step 1 is solved by soft thresholding [20]. Step 2 is a constraint quadratic program, which can be solved by Projected Barzilai-Borwein (PBB) method [10, 9]. It is an iterative method based on the gradient of the object function. The computation cost for the gradient is dominated by the matrix-vector multiplication with the  $n_p \times n_p$  matrix

$$\frac{\rho_1}{\rho_2} \mathbf{D}^\top \mathbf{D} + \mathbf{I}.$$
 (69)

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 [11]. Therefore the overall cost for Step 2 is  $O(n_p)$ . Step 3 is a Least-Squares (LS) problems, where the variable  $\lambda$  is an *m*-vector, and the computation cost  $O(mn_p)$ . Moreover, if an orthonormal basis is chosen to represent the solution, then by (47), **B** is orthogonal. We also note that **K** has orthonormal columns, so **BK** is an orthogonal matrix as well. Then this step has the explicit solution

$$\boldsymbol{\lambda}^{(k+1)} = (\mathbf{B}\mathbf{K})^{\top} (\mathbf{f}^{(k+1)} - \mathbf{B}\mathbf{c}^* + \frac{\boldsymbol{\mu}_2^{(k)}}{\rho_2}), \tag{70}$$

which has complexity  $O(mn_p)$ . Step 4 and 5 are simple matrix-vector computations, and the computation cost are  $n_p$  and  $O(mn_p)$  respectively. So the overall complexity is  $O(mn_p)$ .

#### 4.4.3 Comparison with direct Augmented Lagrangian method

As a comparison, we may also apply Augmented Lagrangian method directly to the problem

$$\hat{f} = \underset{f \ge 0}{\operatorname{argmin}} \|f\|_{TV} \quad \text{such that} \quad Af = g, \tag{71}$$

which is a special case of (21). After discretization and change of basis, it is proposed as the optimization problem

$$\underset{\mathbf{c}}{\operatorname{argmin}} \|\mathbf{B}\mathbf{c}\|_{TV} \quad \text{subject to } \mathbf{B}\mathbf{c} \ge \mathbf{0} \text{ and } \mathbf{M}\mathbf{c} = \mathbf{g}, \tag{72}$$

where the final solution is given by  $\mathbf{f} = \mathbf{Bc}$ . By introducing another auxiliary variable  $\mathbf{w} = \mathbf{Df}$ , (72) has the un-constraint formulation as

$$\min_{\mathbf{w},\mathbf{f}\geq\mathbf{0},\mathbf{c}} \max_{\boldsymbol{\mu}_{1},\boldsymbol{\mu}_{2},\boldsymbol{\mu}_{3}} \gamma \|\mathbf{w}\|_{1} + (\boldsymbol{\mu}_{1},\mathbf{D}\mathbf{f}-\mathbf{w}) + \frac{\beta_{1}}{2} \|\mathbf{D}\mathbf{f}-\mathbf{w}\|^{2} + (\boldsymbol{\mu}_{2},\mathbf{B}\mathbf{c}-\mathbf{f}) + \frac{\beta_{2}}{2} \|\mathbf{B}\mathbf{c}-\mathbf{f}\|^{2} + (\boldsymbol{\mu}_{3},\mathbf{M}\mathbf{c}-\mathbf{g}) + \frac{\beta_{3}}{2} \|\mathbf{M}\mathbf{c}-\mathbf{g}\|^{2},$$
(73)

which can be solved in the same way as Algorithm 4.1.  $\mathbf{w}, \mathbf{f}$  are updated by the same formulas as in 4.1. The major difference is that in (73),  $\mathbf{c} \in \mathbb{R}^n$  instead of  $\lambda \in \mathbb{R}^m$  is updated in each iteration, which has the formulation

$$\mathbf{c}^{(k+1)} = \underset{\mathbf{c}}{\operatorname{argmin}} \ \frac{\beta_2}{2} \|\mathbf{B}\mathbf{c} - \mathbf{f}^{(k+1)} + \frac{\boldsymbol{\mu}_2^{(k)}}{\beta_2}\|^2 + \frac{\beta_3}{2} \|\mathbf{M}\mathbf{c} - \mathbf{g} + \frac{\boldsymbol{\mu}_3^{(k)}}{\beta_3}\|^2.$$
(74)

It is another LS problem similar to Step 3 in Algorithm 4.1, but the unknown is larger in size. The system matrix in the normal equation for (74) is

$$\frac{\beta_2}{\beta_3} \mathbf{B}^\top \mathbf{B} + \mathbf{M}^\top \mathbf{M}.$$
 (75)

It is an  $n \times n$  dense matrix, and cannot be inverted by fast transforms to our knowledge. Compared to formula (70) in Algorithm 4.1, solving this LS problem is more computationally involved. In fact, formula (70) has only one matrix vector multiplication with time complexity  $O(mn_p)$ . For (74), 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  $O(nn_p) + O(n^2)$ , compared to  $O(mn_p)$  for OSKCA. Here we note  $m < n < n_p$ , so OSKCA has smaller complexity. 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 the next section.

We also note that the proposed new method OSKCA is closely related to the Augmented Lagrangian method for ROF model as described in [40]. OSKCA uses augmented Lagrangian functional to transform the constrained minimization problem (66) to an unconstrained saddle point problem (68). However, in OSKCA, the data fitting constraint ( $\mathbf{Mc} = \mathbf{g}$ ) is treated separately, which does not appear in the augmented Lagrangian functional. We show that OSCKA is more efficient if the data fitting constraint is severely under-determined ( $m \ll n$ ).

# 5 Numerical Examples

### 5.1 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 2a. 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 2b. Totally we have  $40 \times 60 = 2400$ measurements. Different levels of noise are added to the simulated measurements for comparison.



Figure 2: (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.

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 [28]. The linear equation for FEM is solved by an implementation of the Algebraic Multi-grid Method [32].

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

For Tikhonov regularization, nodal basis functions are used to represent the reconstruction result. We use ART for  $L^2$  regularization, and Bregman Operator Splitting (BOS) for TV regularization.

For OSKCA, the basis functions are chosen as

$$\{\cos(2\pi(p\frac{x}{w}+q\frac{y}{h})), \sin(2\pi(p\frac{x}{w}+q\frac{y}{h}))\},\tag{76}$$



Figure 3: Reconstructed fluorescent distributions for 2D simulated data.

where  $|p|, |q| \leq 10$ . After removing those duplicates, we actually have n = 441 basis functions. We generate matrix **M** in the inverse problem (45). And then apply iterative Tikhonov regularization (51) with parameter  $h = 10^{-2}$  for the computation of the orthogonal solution. Algorithm described in Section 4.3 provides a basis for the kernel of **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 4.1 is used if TV minimization requirement is used.



Figure 4: 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.

We compare the results obtained by OSKCA and by Tikhonov regularization methods. Figure 3 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 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 4. 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 4 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.

Tikhonov $+$ L2	Tikhonov + TV	OSKCA + positivity	OSKCA + TV
2812	5919	320	230

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

## 5.2 The effect of using more detectors

In the following examples, we consider the experiment setup with different number of detectors. The domain has the size 50mm×50mm, and the parameters are  $\mu_s = 1$ mm<sup>-1</sup> throughout the domain and  $\mu_a = 0.01$ mm<sup>-1</sup>.

In the first example illustrated in Figure 5, three circular inclusions of different sizes and intensities are implanted in the homogeneous medium. We compare the reconstruction when different numbers of detectors around the boundary are used to collect measurements. We can see that by



Figure 5: Three circular inclusions of different sizes and intensities. More detectors are helpful to achieve better resolutions.

increasing the number of detectors, hence collecting more boundary measurements, the resolutions in the reconstructions are improved.

# 5.3 Comparison between OSKCA and direct Augmented Lagrangian method

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 (76). The mesh for the forward model has 7839 nodes and the mesh for the inverse problem has 4096 nodes. We apply Augmented Lagrangian method directly to the problem (71) by solving the formulation (73). We also use OSKCA-TV (Algorithm 4.1) 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 6(b)(c). We can see that OSKCA has better resolution than the direct Augmented Lagrangian method.

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 7. We can see that OSKCA has significantly better data fidelity in 200 iterations. The main



Figure 6: (a) Ground Truth. (b) Augmented Lagrangian with TV minimization regularization. (c) OSKCA with TV minimization regularization.



Figure 7: The plot of  $\|\mathbf{Mc} - \mathbf{g}\| / \|\mathbf{g}\|$  against the number of iterations.

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

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

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 (58).  $\tau$  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.

### 5.4 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  $\mu_s = 1$ mm<sup>-1</sup> throughout the domain and  $\mu_a = 0.01$ mm<sup>-1</sup>. By varying the depth of the inclusions, we compare the reconstruction results, which are shown in Figure 8. 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.



Figure 8: The first row are the images of the ground truth, and the second row are the reconstructions on a mesh with 4096 nodes.

We can see that as the depth of the source increases, the resolution of the reconstruction gets worse. As shown in Figure 8, 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 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 8. The resolution of the reconstructed image is significantly better, which is shown in Figure 9b.



Figure 9: The fluorescent inclusions are 20mm deep. Reconstruction is done on a mesh with 262144 nodes.

# 6 Conclusions

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

# Acknowledgment

The authors would like to thank Prof. Xiaojing Ye for his suggestions on the implementation of Augmented Lagrangian method, Prof. Xuecheng Tai and two other anonymous referees for their helpful comments and suggestions for improving this paper.

# References

[1] S R Arridge. Optical tomography in medical imaging. *Inverse Problems*, 15(2):R41–R93, April 1999.

- [2] Simon R Arridge and John C Schotland. Optical tomography: forward and inverse problems. *Inverse Problems*, 25(12):123010, December 2009.
- [3] G. Bao, S. Hou, and P. Li. Inverse scattering by a continuation method with initial guesses from a direct imaging algorithm. *Journal of Computational Physics*, 227(1):755–762, 2007.
- [4] A. Behrooz, H.M. Zhou, A.A. Eftekhar, and A. Adibi. Total variation regularization for 3d reconstruction in fluorescence tomography: experimental phantom studies. *Applied Optics*, 51(34):8216–8227, 2012.
- [5] C. Bremer, V. Ntziachristos, and R. Weissleder. Optical-based molecular imaging: contrast agents and potential medical applications. *European radiology*, 13(2):231–243, 2003.
- [6] S.C. Brenner and R. Scott. The mathematical theory of finite element methods, volume 15. Springer, 2007.
- [7] J.F. Cai, E.J. Candès, and Z. Shen. A singular value thresholding algorithm for matrix completion. SIAM Journal on Optimization, 20(4):1956–1982, 2010.
- [8] Y. Censor and E. Tom. Convergence of string-averaging projection schemes for inconsistent convex feasibility problems. *Optimization Methods and Software*, 18(5):543–554, 2003.
- [9] Y. Chen, W.W. Hager, M. Yashtini, X. Ye, and H. Zhang. Bregman operator splitting with variable stepsize for total variation image reconstruction. *Comput. Optim. Appl*, 2012.
- [10] Y.H. Dai and R. Fletcher. Projected barzilai-borwein methods for large-scale boxconstrained quadratic programming. *Numerische Mathematik*, 100(1):21–47, 2005.
- [11] Y.H. Dai and L.Z. Liao. R-linear convergence of the barzilai and borwein gradient method. IMA Journal of Numerical Analysis, 22(1):1–10, 2002.
- [12] J. Dutta, S. Ahn, A.A. Joshi, and R.M. Leahy. Optimal illumination patterns for fluorescence tomography. In *Biomedical Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium on*, pages 1275–1278. IEEE, 2009.
- [13] Joyita Dutta, Sangtae Ahn, Changqing Li, Simon R Cherry, and Richard M Leahy. Joint 11 and total variation regularization for fluorescence molecular tomography. *Physics in medicine and biology*, 57(6):1459–1476, March 2012. PMID: 22390906.
- [14] H. Egger, M. Freiberger, and M. Schlottbom. On forward and inverse models in fluorescence diffuse optical tomography. *Inverse Problems and Imaging*, 4(3):411–427, 2010.
- [15] H.W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375. Springer, 1996.
- [16] E. Esser. Applications of lagrangian-based alternating direction methods and connections to split bregman. CAM report, 9:31, 2009.
- [17] T J Farrell, M S Patterson, and B Wilson. A diffusion theory model of spatially resolved, steady-state diffuse reflectance for the noninvasive determination of tissue optical properties in vivo. *Medical physics*, 19(4):879–888, August 1992. PMID: 1518476.
- [18] H. Gao and H. Zhao. Multilevel bioluminescence tomography based on radiative transfer equation part 1: l1 regularization. Optics express, 18(3):1854–1871, 2010.

- [19] H. Gao and H. Zhao. Multilevel bioluminescence tomography based on radiative transfer equation part 2: total variation and l1 data fidelity. *Optics express*, 18(3):2894–2912, 2010.
- [20] T. Goldstein and S. Osher. The split bregman method for l1-regularized problems. SIAM Journal on Imaging Sciences, 2(2):323–343, 2009.
- [21] G.H. Golub and C.F. Van Loan. *Matrix computations*, volume 3. Johns Hopkins University Press, 1996.
- [22] P.C. Hansen and D.P. O'Leary. The use of the L-curve in the regularization of discrete ill-posed problems. *SIAM Journal on Scientific Computing*, 14(6):1487–1503, 1993.
- [23] R C Haskell, L O Svaasand, T T Tsay, T C Feng, M S McAdams, and B J Tromberg. Boundary conditions for the diffusion equation in radiative transfer. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, 11(10):2727–2741, October 1994. PMID: 7931757.
- [24] D.J. Hawrysz and E.M. Sevick-Muraca. Developments toward diagnostic breast cancer imaging using near-infrared optical measurements and fluorescent contrast agents. *Neoplasia (New York, NY)*, 2(5):388, 2000.
- [25] J.C. Hebden, S.R. Arridge, and D.T. Delpy. Optical imaging in medicine: I. experimental techniques. *Physics in medicine and biology*, 42(5):825, 1999.
- [26] G.T. Herman, A. Lent, and S.W. Rowland. 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*, 42(1):1–32, 1973.
- [27] A. Ishimaru. Wave propagation and scattering in random media, volume 2. IEEE press Piscataway, NJ, 1997.
- [28] J. Kaipio and E. Somersalo. Statistical and computational inverse problems, volume 160. Springer, 2004.
- [29] Olga Aleksandrovna Ladyzhenskaia and Nina Nikolaevna Uraltseva. Linear and Quasilinear Elliptic Equations. Academic Press, 1968.
- [30] Vadim A Markel and John C Schotland. Inverse problem in optical diffusion tomography.
   II. role of boundary conditions. Journal of the Optical Society of America. A, Optics, Image Science, and Vision, 19(3):558–566, March 2002. PMID: 11876321.
- [31] A. Neumaier. Solving ill-conditioned and singular linear systems: A tutorial on regularization. SIAM Review, 40(3):636–666, 1998.
- [32] Yvan Notay. An aggregation-based algebraic multigrid method. Electronic Transactions on Numerical Analysis, 37(6):123–146, 2010.
- [33] L.I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.
- [34] W. Rudin. Functional Analysis. McGraw-Hill, New York, 1991.
- [35] J C Schotland and V A Markel. Inverse scattering with diffusing waves. Journal of the Optical Society of America. A, Optics, Image Science, and Vision, 18(11):2767–2777, November 2001. PMID: 11688867.

- [36] M. Schweiger, S R Arridge, M Hiraoka, and D.T. Delpy. The finite element method for the propagation of light in scattering media: Boundary and source conditions. *Medical Physics*, 22(11):1779, 1995.
- [37] Ge Wang, Yi Li, and Ming Jiang. Uniqueness theorems in bioluminescence tomography. Medical Physics, 31(8):2289, 2004.
- [38] Yilun Wang, Junfeng Yang, Wotao Yin, and Yin Zhang. A new alternating minimization algorithm for total variation image reconstruction. SIAM Journal on Imaging Sciences, 1(3):248, 2008.
- [39] R. Weissleder, C.H. Tung, U. Mahmood, and A. Bogdanov. In vivo imaging of tumors with protease-activated near-infrared fluorescent probes. *Nature biotechnology*, 17(4):375–378, 1999.
- [40] C. Wu and X.C. Tai. Augmented lagrangian method, dual methods, and split bregman iteration for rof, vectorial tv, and high order models. SIAM J. Imaging Sci, 3(3):300–339, 2010.