

Discretization Error Analysis and Adaptive Meshing Algorithms for Fluorescence Diffuse Optical Tomography in the Presence of Measurement Noise

Lu Zhou, and Birsen Yazıcı*, *Senior Member, IEEE*

Abstract—Quantitatively accurate Fluorescence Diffuse Optical Tomographic (FDOT) image reconstruction is a computationally demanding problem that requires repeated numerical solutions of two coupled partial differential equations and an associated inverse problem. Recently, adaptive finite element methods have been explored to reduce the computation requirements of the FDOT image reconstruction. However, existing approaches ignore the ubiquitous presence of noise in boundary measurements. In this paper, we analyze the effect of finite element discretization on the FDOT forward and inverse problems in the presence of measurement noise and develop novel adaptive meshing algorithms for FDOT that take into account noise statistics.

We formulate the FDOT inverse problem as an optimization problem in the maximum *a posteriori* framework to estimate the fluorophore concentration in a bounded domain. We use the Mean-Square-Error (MSE) between the exact solution and the discretized solution as a figure of merit to evaluate the image reconstruction accuracy, and derive an upper bound on the MSE which depends on the forward and inverse problem discretization parameters, noise statistics, *a priori* information of fluorophore concentration, source and detector geometry, as well as background optical properties. Next, we use this error bound to develop adaptive meshing algorithms for the FDOT forward and inverse problems to reduce the MSE due to discretization in the reconstructed images. Finally, we present a set of numerical simulations to illustrate the practical advantages of our adaptive meshing algorithms for FDOT image reconstruction.

Index Terms—Fluorescence diffuse optical tomography, adaptive meshing algorithms, error analysis.

I. INTRODUCTION

FLUORESCENCE Diffuse Optical Tomography (FDOT) is an emerging molecular imaging modality with applications in small animal and deep tissue imaging [1], [2]. FDOT uses visible or near infrared light to reconstruct the concentration, pharmacokinetics, as well as the life time of fluorophores injected into the tissue. Similar to its analogue Diffuse Optical

Tomography (DOT), FDOT poses a computationally intense imaging problem. This stems from the requirement of numerically solving both the forward problem, comprised of a set of diffusion equations, and the inverse problem, which is typically represented by a nonlinear integral equation. The numerical solutions of FDOT forward and inverse problems contain error due to discretization, and this discretization error together with the measurement noise deteriorate the final reconstruction accuracy. Thus, the discretization presents a tradeoff between the accuracy and the computational efficiency of the image reconstruction. To improve the reconstruction accuracy, one can reduce the mesh size and increase the number of discretization points. However, this also increases the size of the discretized forward and inverse problems, thereby decreasing the computational efficiency of the image reconstruction. Recently, a number of adaptive discretization techniques for the Partial Differential Equation (PDE) based inverse coefficient problems have been developed [3]–[16]. However, these approaches ignore the presence of noise in boundary measurements. In this paper, we analyze the effect of finite element discretization on the FDOT forward and inverse problems in the presence of measurement noise and develop novel adaptive meshing algorithms that take into account noise statistics.

There is extensive research on the analysis of discretization error in the numerical solutions of PDEs [17]–[22]. However, in the area of PDE-based inverse coefficient problems, where the objective is to estimate primarily the coefficients of PDEs, relatively little has been published (see [3]–[7]). As an application of the error analysis, Beilina *et al.* derived an *a posteriori* error estimate and developed an adaptive meshing method for the solution of an inverse acoustic scattering problem [8], [9]. In the area of FDOT, in [10], Bangerth *et al.* formulated the image reconstruction problem as a PDE-constrained optimization problem, and employed a mesh refinement method suggested in a dual weighted residual framework [3]. In [12], to achieve fast and robust parameter mapping between the adaptively refined/derefining meshes of forward and inverse problems, Lee *et al.* developed an algorithm to identify and resolve the intersections of tetrahedral finite elements. In [11], this algorithm was utilized in FDOT reconstruction within a dual adaptive meshing scheme in which the meshes for the forward and inverse problems are independently refined based on an *a posteriori* error estimate. In our previous work [13],

Copyright (c) 2009 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

Manuscript received May 28, 2010; revised August 7, 2010. This work was supported by US Army Medical Research-W81XWH-04-1-0559 and by the Center for Subsurface Sensing and Imaging Systems, under the Engineering Research Centers Program of the National Science Foundation (Award Number EEC-9986821). *Asterisk indicates corresponding author.*

L. Zhou is with Bloomberg L.P., New York, NY, USA.

*B. Yazıcı is with the Departments of Electrical, Computer and Systems Engineering and Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA (e-mail: yazici@ecse.rpi.edu).

[15], we presented a Finite Element Method (FEM) based approach to analyze the effect of discretization on the accuracy of DOT and FDOT reconstructions under the assumption that the measurements are noise-free. These studies further led to the development of new adaptive mesh generation algorithms for these two imaging modalities, that can effectively reduce the error due to discretization [14], [16]. Although most of these studies [8]–[10], [13]–[16] take into account the interdependence of forward and inverse problems and their proposed adaptive meshing methods can effectively reduce the discretization error in the reconstructed images, the effect of measurement noise was not considered in the error analysis and adaptive meshing schemes.

In this paper, we focus on analyzing the effect of measurement noise in the FDOT forward and inverse problem discretizations, and develop adaptive meshing algorithms that take into account noise statistics and can effectively reduce the discretization error. We assume that FDOT boundary measurements are collected using a continuous wave (CW) imaging system. We model the forward problem of FDOT by a pair of diffusion equations at the excitation and emission wavelengths, and use FEM to solve these equations. We formulate the FDOT inverse problem as an optimization problem in the Maximum *A Posteriori* (MAP) framework to estimate the fluorophore concentration, and use the Mean-Square-Error (MSE) between the exact solution and the discretized solution of the inverse problem as a figure of merit to assess the error due to discretization. We analyze the effect of discretization on the two components of the MSE, namely the bias and variance of the reconstructed image, and derive upper bounds for these quantities. These upper bounds depend on the forward and inverse problem discretization parameters, noise statistics, *a priori* information of fluorophore concentration, source and detector geometry, as well as background optical properties. We next utilize these upper bounds to design local error indicators to use in the adaptive discretization of the FDOT forward and inverse problems. Unlike the algorithms in [16], the new adaptive meshing algorithms take into account noise statistics and *a priori* information of fluorophore concentration. The numerical simulation results show that the new adaptive meshing algorithms can effectively improve the reconstruction accuracy and resolution when noise statistics are taken into account as compared to the uniform meshing and adaptive meshing algorithms presented in [16]. Similar results are also reported in [23] where we compare the accuracy of reconstruction for different meshing schemes using real measurements from a phantom experiment.

The outline of the paper is as follows: In Section II, III and IV, we introduce the FDOT forward and inverse problems, and their discretizations, respectively. In Section V, we derive the upper bounds for the bias, variance and MSE of the reconstructed image. In Section VI, we present adaptive meshing algorithms for FDOT forward and inverse problems based on the results in Section V. In Section VII, we present simulation results to demonstrate the performance of adaptive meshing algorithms. Finally, in Section VIII, we conclude our discussion.

TABLE I
DEFINITION OF FUNCTION SPACES AND NORMS.

Notation	Explanation
$C(\Omega)$	Space of continuous functions on Ω
$L^\infty(\Omega)$	$L^\infty(\Omega) = \{f \text{ess sup}_\Omega f(\mathbf{x}) < \infty\}$
$L^p(\Omega)$	$L^p(\Omega) = \{f (\int_\Omega f(\mathbf{x}) ^p dx)^{1/p} < \infty\}, p \in [1, \infty)$
$H^p(\Omega)$	$H^p(\Omega) = \{f (\sum_{ z \leq p} \ D_w^z f\ _0^2)^{1/2} < \infty\}, p \in [1, \infty)$
$\ f\ _0$	The $L^2(\Omega)$ norm of f
$\ f\ _p$	The $H^p(\Omega)$ norm of f
$\ f\ _\infty$	The $L^\infty(\Omega)$ norm of f
$\ f\ _{0,m}$	The L^2 norm of f over the m^{th} finite element Ω_m
$\ f\ _{p,m}$	The H^p norm of f over the m^{th} finite element Ω_m
$\ f\ _{\infty,m}$	The L^∞ norm of f over the m^{th} finite element Ω_m

II. FDOT FORWARD PROBLEM

A. Notational Conventions

Throughout the paper, we use capital cursive letters (\mathcal{A}) for operators and bold capital letters ($\mathbf{\Sigma}$) for matrices. We denote functions by lowercase letters (g and ϕ etc.) and their finite dimensional approximations by corresponding uppercase letters (G and Φ etc.). We use bold to denote vectorized quantities such as \mathbf{r} , $\mathbf{\Gamma}$. Table I provides a summary of key variables and function spaces used throughout the paper.

B. Diffusion Model for Light Propagation

We assume that the CW light sources are used to estimate the fluorophore concentration in a bounded domain $\Omega \subset \mathbb{R}^3$. Therefore, we use a pair of coupled frequency-domain diffusion equations, with modulation frequency $\omega = 0$, and the corresponding boundary conditions on $\partial\Omega$ to model light propagation [24]:

$$-\nabla \cdot D(\mathbf{r})\nabla\phi_x(\mathbf{r}, \mathbf{r}_i) + \mu_{ax}(\mathbf{r})\phi_x(\mathbf{r}, \mathbf{r}_i) = S_i(\mathbf{r}), \quad \mathbf{r} \in \Omega, \quad (1)$$

$$2D(\mathbf{r})\frac{\partial\phi_x(\mathbf{r}, \mathbf{r}_i)}{\partial n} + \rho\phi_x(\mathbf{r}, \mathbf{r}_i) = 0, \quad \mathbf{r} \in \partial\Omega, \quad (2)$$

$$-\nabla \cdot D(\mathbf{r})\nabla\phi_m(\mathbf{r}, \mathbf{r}_i) + \mu_{am}(\mathbf{r})\phi_m(\mathbf{r}, \mathbf{r}_i) = \phi_x(\mathbf{r}, \mathbf{r}_i)\eta\mu_{axf}(\mathbf{r}), \quad \mathbf{r} \in \Omega, \quad (3)$$

$$2D(\mathbf{r})\frac{\partial\phi_m(\mathbf{r}, \mathbf{r}_i)}{\partial n} + \rho\phi_m(\mathbf{r}, \mathbf{r}_i) = 0, \quad \mathbf{r} \in \partial\Omega, \quad (4)$$

where ϕ_x and ϕ_m are the photon densities at the excitation and emission wavelengths, respectively. D is the isotropic diffusion coefficient. μ_{ax} and μ_{am} are the absorption coefficients of the medium at the excitation and emission wavelengths, respectively. η and μ_{axf} are the quantum efficiency and absorption coefficient of the fluorophore. ρ is a parameter governing the internal reflection at the boundary $\partial\Omega$, and $\partial/\partial n$ denotes the directional derivative along the unit normal vector on the domain boundary. S_i is the i^{th} excitation source, modeled by a Gaussian function centered at the source position \mathbf{r}_i , for $i = 1, \dots, N_S$, where N_S is the number of sources. Note that since $\omega = 0$, we drop the frequency dependency of ϕ_x and ϕ_m to simplify our notation.

We make use of the adjoint problem associated with (3) and (4) to express the relationship between the fluorophore

concentration and the measurements [25]:

$$-\nabla \cdot D(\mathbf{r})\nabla g_m^*(\mathbf{r}, \mathbf{r}_j) + \mu_{am}(\mathbf{r})g_m^*(\mathbf{r}, \mathbf{r}_j) = 0, \quad \mathbf{r} \in \Omega, \quad (5)$$

$$2D(\mathbf{r})\frac{\partial g_m^*(\mathbf{r}, \mathbf{r}_j)}{\partial n} + \rho g_m^*(\mathbf{r}, \mathbf{r}_j) = S_j^*(\mathbf{r}), \quad \mathbf{r} \in \partial\Omega, \quad (6)$$

where $g_m^*(\mathbf{r}, \mathbf{r}_j)$ is the solution of the adjoint problem for the j^{th} adjoint source S_j^* located at the detector position $\mathbf{r}_j \in \partial\Omega$, for $j = 1, \dots, N_D$, where N_D is the number of detectors.

Given N_S sources and N_D detectors, we define $\Gamma_{i,j}$ to be the measurement obtained by the j^{th} detector, $j = 1, \dots, N_D$, due to the i^{th} source, $i = 1, \dots, N_S$. Using (1)-(2) and (5)-(6), we write $\Gamma_{i,j}$ as followings:

$$\begin{aligned} \Gamma_{i,j} &= \int_{\Omega} g_m^*(\mathbf{r}, \mathbf{r}_j)\phi_x(\mathbf{r}, \mathbf{r}_i)\eta\mu_{axf}(\mathbf{r})d\mathbf{r} \\ &= \int_{\Omega} g_j^*(\mathbf{r})\phi_i(\mathbf{r})\mu(\mathbf{r})d\mathbf{r}, \end{aligned} \quad (7)$$

where we define $g_j^*(\mathbf{r}) := g_m^*(\mathbf{r}, \mathbf{r}_j)$ and $\phi_i(\mathbf{r}) := \phi_x(\mathbf{r}, \mathbf{r}_i)$, suppressing the excitation (x) and emission (m) wavelengths dependency of these functions to simplify our notation.

We define $\mu(\mathbf{r}) := \eta\mu_{axf}(\mathbf{r})$, and refer to $\mu(\mathbf{r})$ as the fluorophore concentration, the quantity to be reconstructed. Then we group the individual measurements into the following vector:

$$\mathbf{\Gamma} := [\Gamma_{1,1}, \dots, \Gamma_{1,N_D}, \Gamma_{2,1}, \dots, \Gamma_{N_S,N_D}]^T, \quad (8)$$

and define the vector valued operator $\mathcal{A} : L^2(\Omega) \rightarrow \mathbb{R}^{N_S N_D}$ as

$$(\mathcal{A}\mu)_{ij} := \int_{\Omega} a_{ij}(\mathbf{r})\mu(\mathbf{r})d\mathbf{r}, \quad (9)$$

where $a_{ij}(\mathbf{r}) := g_j^*(\mathbf{r})\phi_i(\mathbf{r})$.

Combining (7), (8) and (9), we write $\mathbf{\Gamma} = \mathcal{A}\mu$, where the integration is understood elementwise.

C. Iterative Linearization

The integral equation in (9) is nonlinear in μ due to the dependency of ϕ_i and g_j^* to μ_{axf} . We use the Born approximation to linearize (9) around a known background fluorophore concentration μ_0 . Note that the Born approximation is valid when the perturbation of the fluorophore absorption coefficient is relatively small as compared to the known background absorption coefficient [24], [26].

Let ϕ_i^0 and $g_j^{*,0}$ be the solutions of (1)-(2) and (5)-(6) for $\mu = \mu_0$, and $a_{ij}^0(\mathbf{r}) = \phi_i^0(\mathbf{r})g_j^{*,0}(\mathbf{r})$. Then, the model (9) becomes

$$(\mathcal{A}_0\mu)_{ij} := \int_{\Omega} a_{ij}^0(\mathbf{r})\mu(\mathbf{r})d\mathbf{r},$$

or $\mathbf{\Gamma} = \mathcal{A}_0\mu$, where a_{ij}^0 , $i = 1, \dots, N_S$ and $j = 1, \dots, N_D$, is the kernel of \mathcal{A}_0 . Note that \mathcal{A}_0 is now linear in μ .

The solution at each linearization step can be iteratively refined based on the following model:

$$\mathbf{\Gamma} = \mathcal{A}_k\hat{\mu}_{k+1},$$

where $\hat{\mu}_{k+1}$ is the estimate of the fluorophore concentration at the $(k+1)^{\text{th}}$ iteration and

$$(\mathcal{A}_k\mu)_{ij} := \int_{\Omega} a_{ij}^k(\mathbf{r})\mu(\mathbf{r})d\mathbf{r},$$

with $a_{ij}^k = \phi_i^k(\mathbf{r})g_j^{*,k}(\mathbf{r})$, $\phi_i^k(\mathbf{r})$ and $g_j^{*,k}(\mathbf{r})$ are computed based on the fluorophore concentration $\hat{\mu}_k$ estimated at k^{th} iteration. Note that we drop 0 and k superscripts on ϕ_i^0 , $g_j^{*,0}$, ϕ_i^k , and $g_j^{*,k}$ for the rest of paper to simplify our notation.

III. FDOT INVERSE PROBLEM FORMULATION

A. Models for Measurement Noise and Fluorophore Concentration

We assume that the measurements are contaminated by additive noise and write:

$$\mathbf{\Gamma} = \mathcal{A}_0\mu + \boldsymbol{\varepsilon}, \quad (10)$$

where $\boldsymbol{\varepsilon} = [\varepsilon_{1,1}, \dots, \varepsilon_{1,N_D}, \varepsilon_{2,1}, \dots, \varepsilon_{N_S,N_D}]^T$ is the noise vector. Without loss of generality, we assume that the components of the noise vector are mutually statistically independent Gaussian random variables with zero-mean and known variance $\sigma_{\boldsymbol{\varepsilon},ij}^2$, for $i = 1, \dots, N_S$ and $j = 1, \dots, N_D$. We denote the covariance matrix of $\boldsymbol{\varepsilon}$ with

$$\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} = \text{diag}([\sigma_{\boldsymbol{\varepsilon},11}^2, \dots, \sigma_{\boldsymbol{\varepsilon},1N_D}^2, \sigma_{\boldsymbol{\varepsilon},21}^2, \dots, \sigma_{\boldsymbol{\varepsilon},N_S N_D}^2]^T).$$

We model the fluorophore concentration image as a Gaussian random field and assume that it is statistically independent of the noise. Furthermore, we assume that the fluorophore concentration μ has mean μ_0 equal to the known background fluorophore concentration. Without loss of generality, we assume that $\mu(\mathbf{r})$ and $\mu(\hat{\mathbf{r}})$, $\mathbf{r} \neq \hat{\mathbf{r}}$, are mutually statistically independent. Note that at the $(k+1)^{\text{th}}$ iteration of the iterative reconstruction, we assume $\hat{\mu}_{k+1}$ has mean $\hat{\mu}_k$, the estimate obtained at the k^{th} iteration. Thus, we define

$$\begin{aligned} \mathbb{E}[\mu(\mathbf{r})] &= \mu_0(\mathbf{r}), \\ \text{Cov}_{\mu\mu}(\mathbf{r}, \hat{\mathbf{r}}) &= \mathbb{E}[(\mu(\mathbf{r}) - \mu_0(\mathbf{r}))(\mu(\hat{\mathbf{r}}) - \mu_0(\hat{\mathbf{r}}))] \\ &=: \kappa(\mathbf{r})\delta(\mathbf{r} - \hat{\mathbf{r}}), \end{aligned}$$

where \mathbb{E} denotes expectation and $\kappa(\mathbf{r}) \geq 0$ for all $\mathbf{r} \in \Omega$.

B. The Maximum A Posteriori Estimators for Fluorophore Concentration

We consider the MAP estimator for μ which is given by the following constrained minimization problem:

$$\mu_{MAP} = \min_{\mu \in L^2(\Omega)} [J_{LH}(\mu) + J_{PR}(\mu)], \quad (11)$$

where

$$\begin{aligned} J_{LH}(\mu) &= \sum_{i,j}^{N_S, N_D} \frac{1}{\sigma_{\boldsymbol{\varepsilon},ij}^2} [\Gamma_{i,j} - (\mathcal{A}_0\mu)_{i,j}]^2, \\ J_{PR}(\mu) &= \int_{\Omega} \frac{1}{\kappa(\mathbf{r})} [\mu(\mathbf{r}) - \mu_0(\mathbf{r})]^2 d\mathbf{r}. \end{aligned}$$

Taking the Gâteaux derivative of (11) with respect to μ , and setting it equal to zero, we obtain an integral equation for which the MAP estimate, μ_{MAP} , of the fluorophore concentration satisfies:

$$\begin{aligned} (\mathcal{A}_0^* \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1} \mathcal{A}_0 \mu_{MAP})(\mathbf{r}) + \frac{\mu_{MAP}(\mathbf{r})}{\kappa(\mathbf{r})} &= (\mathcal{A}_0^* \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1} \mathbf{\Gamma})(\mathbf{r}) \\ &+ \frac{\mu_0(\mathbf{r})}{\kappa(\mathbf{r})}, \end{aligned} \quad (12)$$

where $\mathcal{A}_0^* : \mathbb{R}^{N_S N_D} \rightarrow L^2(\Omega)$ is the adjoint operator of \mathcal{A}_0 [15].

Note that when the *a priori* information on the fluorophore concentration is not available, one can formulate the inverse problem in the Maximum Likelihood (ML) framework with $J_{PR}(\mu) = 0$. In that case, the ML estimate, μ_{ML} , satisfies

$$(\mathcal{A}_0^* \Sigma_{\mathcal{E}}^{-1} \mathcal{A}_0 \mu_{ML}) (\mathbf{r}) = (\mathcal{A}_0^* \Sigma_{\mathcal{E}}^{-1} \Gamma) (\mathbf{r}).$$

Since \mathcal{A}_0 and \mathcal{A}_0^* are both compact [27], we consider the following regularized form for the solution of μ_{ML} :

$$\left(\left(\mathcal{A}_0^* \Sigma_{\mathcal{E}}^{-1} \mathcal{A}_0 + \frac{\lambda_0}{\sigma_{\mathcal{E},max}^2} \mathcal{I} \right) \mu_{ML} \right) (\mathbf{r}) = (\mathcal{A}_0^* \Sigma_{\mathcal{E}}^{-1} \Gamma) (\mathbf{r}),$$

where $\mathcal{I} : L^2(\Omega) \rightarrow L^2(\Omega)$ is the identity operator, $\sigma_{\mathcal{E},max}^2$ is the maximum value of $\sigma_{\mathcal{E},ij}^2$, for $i = 1, \dots, N_S$ and $j = 1, \dots, N_D$, and λ_0 is a small positive constant. Note that there are several methods for choosing appropriate regularization parameters λ_0 (see [28]–[32]). In this paper, we assume that λ_0 is appropriately chosen based on the spectral decomposition of the operator $\mathcal{A}_0^* \mathcal{A}_0$ [32].

To simplify our notation, we define the operator $\mathcal{B} : L^2(\Omega) \rightarrow L^2(\Omega)$ as

$$\mathcal{B} := \mathcal{A}_0^* \Sigma_{\mathcal{E}}^{-1} \mathcal{A}_0,$$

and express (12) as follows:

$$(\mathcal{B} \mu_{MAP}) (\mathbf{r}) + \frac{\mu_{MAP}(\mathbf{r})}{\kappa(\mathbf{r})} = (\mathcal{A}_0^* \Sigma_{\mathcal{E}}^{-1} \Gamma) (\mathbf{r}) + \frac{\mu_0(\mathbf{r})}{\kappa(\mathbf{r})}. \quad (13)$$

We use the Galerkin method [27] to solve the integral equation defined in (13). Thus, we first define the variational form of (13):

$$\mathcal{F}_{MAP}(\psi, \mu_{MAP}) = (\psi, \mathcal{A}_0^* \Sigma_{\mathcal{E}}^{-1} \Gamma) + (\psi, \frac{\mu_0}{\kappa}), \quad (14)$$

where

$$\mathcal{F}_{MAP}(\psi, \mu) := (\psi, \mathcal{B} \mu) + (\psi, \frac{\mu}{\kappa}), \quad (15)$$

(\cdot, \cdot) denotes inner product in $L^2(\Omega)$ and ψ is any test function in $L^2(\Omega)$. Then, the FDOT inverse problem involves recovering μ_{MAP} based on (14).

Similarly, the ML estimate satisfies the following variational form:

$$\mathcal{F}_{ML}(\psi, \mu_{ML}) = (\psi, \mathcal{A}_0^* \Sigma_{\mathcal{E}}^{-1} \Gamma), \quad (16)$$

where

$$\mathcal{F}_{ML}(\psi, \mu) := (\psi, \mathcal{B} \mu) + (\psi, \frac{\lambda_0 \mu}{\sigma_{\mathcal{E},max}^2}),$$

and ψ is any test function in $L^2(\Omega)$.

Finally we note that a unique solution for the inverse problem (14) or (16) exists when κ and λ_0 are appropriately chosen [15].

IV. DISCRETIZATION OF THE FORWARD AND INVERSE PROBLEMS

In the following subsections, we, first, discuss the variational formulation and finite element discretization of the forward problem to obtain a finite-dimensional approximation of the forward problem solution. Next, we use these finite element solutions of the forward problem in the inverse problem formulation and discuss the discretization of the resulting approximate inverse problem using the Galerkin method.

A. Forward Problem Discretization

We express the forward problem defined in (1)–(2) and (5)–(6) in variational forms and next apply the FEM to discretize and solve the resulting problems. To do so, we first multiply (1) and (5) by two test functions $\xi_1 \in H^1(\Omega)$ and $\xi_2 \in H^1(\Omega)$, respectively, and apply Green's theorem to the second-order derivative terms. Then, using the boundary conditions in (2) and (6), we obtain

$$\begin{aligned} \int_{\Omega} (\nabla \xi_1 \cdot D \nabla \phi_i + \mu_{ax} \xi_1 \phi_i) d\mathbf{r} + \frac{1}{2\rho} \int_{\partial\Omega} \xi_1 \phi_i dl \\ = \int_{\Omega} \xi_1 S_i d\mathbf{r}, \end{aligned} \quad (17)$$

$$\begin{aligned} \int_{\Omega} (\nabla \xi_2 \cdot D \nabla g_j^* + \mu_{am} \xi_2 g_j^*) d\mathbf{r} + \frac{1}{2\rho} \int_{\partial\Omega} \xi_2 g_j^* dl \\ = \frac{1}{2\rho} \int_{\partial\Omega} \xi_2 S_j^* dl. \end{aligned} \quad (18)$$

Let L_k denote the piecewise linear Lagrange basis functions used in discretizing the forward problem. We define $Y_i \subset H^1(\Omega)$, $i = 1, \dots, N_S$, as the finite dimensional subspace spanned by $\{L_k\}$, $k = 1, \dots, N_i$. Note that $\{L_k\}$ are associated with the set of points $\{\mathbf{r}_p\}$, $p = 1, \dots, N_i$, on Ω . We further let $\{\Omega_{ni}\}$ denote the corresponding set of elements used to discretize Ω , for $n = 1, \dots, N_{\Delta}^i$; such that $\bigcup_n \Omega_{ni} = \Omega$ and h_{ni} is the diameter of the smallest ball that contains the n^{th} element. Similarly, we define $Y_j^* \subset H^1(\Omega)$, $j = 1, \dots, N_D$, as the finite-dimensional subspace spanned by $\{L_k\}$, $k = 1, \dots, N_j$, which are associated with the set of points $\{\mathbf{r}_p\}$, $p = 1, \dots, N_j$. We further let $\{\Omega_{mj}\}$ denote the corresponding set of elements used to discretize Ω , for $m = 1, \dots, N_{\Delta}^{*j}$; such that $\bigcup_m \Omega_{mj} = \Omega$ and h_{mj} is the diameter of the smallest ball that contains the m^{th} element. Next, we replace ξ_1 , ϕ_i in (17) and ξ_2 , g_j^* in (18) by their finite-dimensional approximations defined as

$$\Xi_1(\mathbf{r}) := \sum_{k=1}^{N_i} p_k L_k(\mathbf{r}), \quad \Phi_i := \sum_{k=1}^{N_i} c_k L_k(\mathbf{r}), \quad (19)$$

$$\Xi_2(\mathbf{r}) := \sum_{k=1}^{N_j} p_k L_k(\mathbf{r}), \quad G_j^* := \sum_{k=1}^{N_j} d_k L_k(\mathbf{r}), \quad (20)$$

and obtain the matrix equations

$$\mathbf{M} \mathbf{c}_i = \mathbf{q}_i, \quad (21)$$

$$\mathbf{M}^* \mathbf{d}_j^* = \mathbf{q}_j^*, \quad (22)$$

for coefficients $\mathbf{c}_i = [c_1, c_2, \dots, c_{N_i}]^T$ and $\mathbf{d}_j^* = [d_1, d_2, \dots, d_{N_j}]^T$.

In (21) and (22), \mathbf{M} and \mathbf{M}^* are the finite element matrices and \mathbf{q}_i and \mathbf{q}_j^* are the load vectors resulting from the finite element discretization of the forward problem.

B. Inverse Problem Discretization

For the inverse problem discretization, we first substitute Φ_i and G_j^* into the operators \mathcal{A}_0 , \mathcal{A}_0^* and \mathcal{B} to obtain the approximate operators denoted by $\widetilde{\mathcal{A}}_0$, $\widetilde{\mathcal{A}}_0^*$ and $\widetilde{\mathcal{B}}$. We, next, substitute $\widetilde{\mathcal{A}}_0$, $\widetilde{\mathcal{A}}_0^*$ and $\widetilde{\mathcal{B}}$ into (14) and (15) to obtain an approximate inverse problem formulation:

$$\widetilde{\mathcal{F}}_{MAP}(\psi, \tilde{\mu}_{MAP}) = (\psi, \widetilde{\mathcal{A}}_0^* \Sigma_{\varepsilon}^{-1} \Gamma) + (\psi, \frac{\mu_0}{\kappa}), \quad (23)$$

for all $\psi \in L^2(\Omega)$, with

$$\widetilde{\mathcal{F}}_{MAP}(\psi, \mu) := (\psi, \widetilde{\mathcal{B}}\mu) + (\psi, \frac{\mu}{\kappa}),$$

where $\tilde{\mu}_{MAP} \in L^2(\Omega)$ is the solution of (23).

Next, we define the finite-dimensional subspace $V(\Omega) \subset L^2(\Omega)$ spanned by the first-order Lagrange basis functions $\{L_k\}$, $k = 1, \dots, N$, which are associated with the set of points $\{\mathbf{r}_p\}$, $p = 1, \dots, N$, on Ω . We use $\{\Omega_t\}$, $t = 1, \dots, N_{\Delta}$, to denote the corresponding set of elements used to discretize Ω with vertices at $\{\mathbf{r}_p\}$, $p = 1, \dots, N$, such that $\bigcup_t^{N_{\Delta}} \Omega_t = \Omega$, and h_t is the diameter of the smallest ball that contains the t^{th} element. We substitute $\tilde{\mu}_{MAP}$ and ψ in (23) with their finite-dimensional approximations $\mu_{MAP}^D \in V(\Omega)$ and $\Psi \in V(\Omega)$ defined, respectively, by

$$\mu_{MAP}^D(\mathbf{r}) := \sum_{k=1}^N m_k L_k(\mathbf{r}), \quad \Psi(\mathbf{r}) := \sum_{k=1}^N p_k L_k(\mathbf{r}), \quad (24)$$

and obtain the following fully discretized inverse problem formulation:

$$\widetilde{\mathcal{F}}_{MAP}(\Psi, \mu_{MAP}^D) = (\Psi, \widetilde{\mathcal{A}}_0^* \Sigma_{\varepsilon}^{-1} \Gamma) + (\Psi, \frac{\mu_0}{\kappa}). \quad (25)$$

The resulting inverse problem formulation can be expressed as the following matrix-vector equation:

$$\mathbf{F}_N \mathbf{m} = \mathbf{G}_N, \quad (26)$$

where $\mathbf{m} = [m_1, \dots, m_N]^T$ represents the unknown coefficients in the finite approximation of μ_{MAP}^D , and \mathbf{F}_N and \mathbf{G}_N are the finite element matrix and the load vector resulting from (25).

V. ANALYSIS OF THE DISCRETIZATION ERROR IN THE PRESENCE OF MEASUREMENT NOISE AND A PRIORI INFORMATION ON FLUOROPHORE CONCENTRATION

In this section, we analyze the effect of forward and inverse problem discretizations on the accuracy of FDOT reconstruction in the presence of measurement noise and *a priori* information of the fluorophore concentration. In this respect, we quantify the error in the mean square sense, and derive an upper bound for the MSE in FDOT reconstruction due to discretization. Next, we discuss the case of the ML estimate, as well as the case involving correlated noise and *a priori* fluorophore concentration models. Finally, we comment on the implications of the MSE bound for the discretizations of the FDOT forward and inverse problems.

A. Error Bound on the MSE due to Discretization

We are interested in quantifying the difference between the exact estimate μ_{MAP} and the estimate μ_{MAP}^D obtained after forward and inverse problem discretizations. Thus, we define

$$e_{MAP}(\mathbf{r}) := \mu_{MAP}(\mathbf{r}) - \mu_{MAP}^D(\mathbf{r}),$$

and quantify the difference between μ_{MAP} and μ_{MAP}^D in term of the MSE defined as follows:

$$\text{MSE}[\mu_{MAP}^D] := \int_{\Omega} \mathbb{E} [|e_{MAP}(\mathbf{r})|^2] d\mathbf{r}. \quad (27)$$

We further express (27) as

$$\text{MSE}[\mu_{MAP}^D] = \text{Bias}^2[\mu_{MAP}^D] + \text{Var}[\mu_{MAP}^D],$$

where

$$\text{Bias}^2[\mu_{MAP}^D] := \int_{\Omega} |\mathbb{E} [e_{MAP}(\mathbf{r})]|^2 d\mathbf{r}, \quad (28)$$

$$\text{Var}[\mu_{MAP}^D] := \int_{\Omega} \mathbb{E} [|e_{MAP}(\mathbf{r}) - \mathbb{E}[e_{MAP}(\mathbf{r})]|^2] d\mathbf{r}. \quad (29)$$

We refer to $\text{Bias}[\mu_{MAP}^D]$ as the bias of μ_{MAP}^D with respect to the exact MAP estimate μ_{MAP} and $\text{Var}[\mu_{MAP}^D]$ as the variance of μ_{MAP}^D .

In Theorem 1 and 2, we present upper bounds for $\text{Bias}[\mu_{MAP}^D]$ and $\text{Var}[\mu_{MAP}^D]$; and next use these bounds to develop new adaptive meshing algorithms for FDOT in the following sections.

Theorem 1:

Consider the Galerkin projection of the variational problems (17), (18) and (23) described in Section IV.

Let $\bar{\mu}_{MAP}(\mathbf{r}) := \mathbb{E}[\mu_{MAP}(\mathbf{r})]$. Then,

1) $\bar{\mu}_{MAP}$ satisfies the following variational problem:

$$\mathcal{F}_{MAP}(\psi, \bar{\mu}_{MAP}) = (\psi, \mathcal{A}_0^* \Sigma_{\varepsilon}^{-1} \bar{\Gamma}) + (\psi, \frac{\mu_0}{\kappa}),$$

for all $\psi \in L^2(\Omega)$, where $\bar{\Gamma} = \mathbb{E}[\Gamma] = \mathcal{A}_0 \mu_0$ and

2) Assume that $\bar{\mu}_{MAP} \in H^1(\Omega)$, then

$$\text{Bias}^2[\mu_{MAP}^D] \leq C_B [B_1 + B_2 + B_3]^2, \quad (30)$$

where

$$\begin{aligned} B_1 &= \sum_{i=1}^{N_S} \sum_{n,j}^{N_{\Delta}^i, N_D} (F_{ij}^1 \|g_j^* \bar{\mu}_{MAP}\|_{0,ni} \\ &\quad + F_{ij}^2 \|g_j^*\|_{\infty,ni}) \|\phi_i\|_{1,ni} h_{ni}, \\ B_2 &= \sum_{j=1}^{N_D} \sum_{m,i}^{N_{\Delta}^j, N_S} (F_{ij}^1 \|\phi_i \bar{\mu}_{MAP}\|_{0,mj} \\ &\quad + F_{ij}^2 \|\phi_i\|_{\infty,mj}) \|g_j^*\|_{1,mj} h_{mj}, \\ B_3 &= \sum_{t=1}^{N_{\Delta}} \left(\sum_{i,j}^{N_S, N_D} I_{ij}^1 \|G_j^* \Phi_i\|_{0,t} + I_t^2 \right) \\ &\quad \cdot \|\bar{\mu}_{MAP}\|_{1,t} h_t, \end{aligned}$$

with

$$F_{ij}^1 = \frac{2\|\kappa\|_\infty \|g_j^* \phi_i\|_0}{\sigma_{\boldsymbol{\varepsilon},ij}^2}, \quad F_{ij}^2 = \frac{\|\kappa\|_\infty |\bar{\Gamma}_{i,j}|}{\sigma_{\boldsymbol{\varepsilon},ij}^2},$$

$$I_{ij}^1 = \frac{\|\kappa\|_\infty \|G_j^* \Phi_i\|_0}{\sigma_{\boldsymbol{\varepsilon},ij}^2}, \quad I_t^2 = \|\kappa\|_\infty \left\| \frac{1}{\kappa} \right\|_{\infty,t},$$

and C_B is a constant independent of the discretization parameters h_{ni} , h_{mj} and h_t .

Proof: See Appendix A. \square

Theorem 2:

Consider the Galerkin projection of the variational problems (17), (18) and (23) described in Section IV. Let $\pi_{ij} \in L^2(\Omega)$ be the solution of the following variational problem:

$$\mathcal{F}_{MAP}(\psi, \pi_{ij}) = (\psi, \mathcal{A}_0^* \boldsymbol{\Sigma} \boldsymbol{\varepsilon}^{-1} \mathbf{e}_{ij}),$$

for all $\psi \in L^2(\Omega)$, and $\mathbf{e}_{ij} \in \mathbb{R}^{N_S N_D}$ is the $[N_D(i-1) + j]^{\text{th}}$ column vector of the $N_S N_D \times N_S N_D$ identity matrix. Assume $\pi_{ij} \in H^1(\Omega)$, then $\text{Var}[\mu_{MAP}^D]$ satisfies the following inequality:

$$\text{Var}[\mu_{MAP}^D] \leq C_V [V_1 + V_2 + V_3]^2, \quad (31)$$

where

$$V_1 = \sum_{i=1}^{N_S} \sum_{n,j}^{N_\Delta^i, N_D} \left(F_{ij}^1 \sum_{i',j'}^{N_S, N_D} \|g_j^* D_{i'j'} \pi_{i'j'}\|_{0,ni} \right. \\ \left. + F_{ij}^3 D_{ij} \|g_j^*\|_{\infty,ni} \right) \|\phi_i\|_{1,ni} h_{ni},$$

$$V_2 = \sum_{j=1}^{N_D} \sum_{m,i}^{N_\Delta^j, N_S} \left(F_{ij}^1 \sum_{i',j'}^{N_S, N_D} \|\phi_i D_{i'j'} \pi_{i'j'}\|_{0,mj} \right. \\ \left. + F_{ij}^3 D_{ij} \|\phi_i\|_{\infty,mj} \right) \|g_j^*\|_{1,mj} h_{mj},$$

$$V_3 = \sum_{t=1}^{N_\Delta} \left(\sum_{i,j}^{N_S, N_D} I_{ij}^1 \|G_j^* \Phi_i\|_{0,t} + I_t^2 \right) \\ \cdot \left(\sum_{i,j}^{N_S, N_D} \|D_{ij} \pi_{ij}\|_{1,t} \right) h_t,$$

with

$$F_{ij}^1 = \frac{2\|\kappa\|_\infty \|g_j^* \phi_i\|_0}{\sigma_{\boldsymbol{\varepsilon},ij}^2}, \quad F_{ij}^3 = \frac{\|\kappa\|_\infty}{\sigma_{\boldsymbol{\varepsilon},ij}^2},$$

$$I_{ij}^1 = \frac{\|\kappa\|_\infty \|G_j^* \Phi_i\|_0}{\sigma_{\boldsymbol{\varepsilon},ij}^2}, \quad I_t^2 = \|\kappa\|_\infty \left\| \frac{1}{\kappa} \right\|_{\infty,t},$$

$$D_{ij} = \left[\sigma_{\boldsymbol{\varepsilon},ij}^2 + \int_{\Omega} \kappa(\mathbf{r}) a_{ij}^{*,0}(\mathbf{r}) a_{ij}^0(\mathbf{r}) d\mathbf{r} \right]^{1/2},$$

and C_V is a constant independent of the discretization parameters h_{ni} , h_{mj} and h_t .

Proof: See Appendix B. \square

We note that, for the ML estimate, μ_{ML} , of the fluorophore concentration, the upper bounds for $\text{Bias}^2[\mu_{ML}^D]$ and $\text{Var}[\mu_{ML}^D]$

are given as in (30) and (31) with $D_{ij} = \sigma_{\boldsymbol{\varepsilon},ij}$ and $\kappa(\mathbf{r}) = \sigma_{\boldsymbol{\varepsilon},max}^2/\lambda_0$. We also note that, when the measurements are noise-free, the inverse problem formulation in this work can be reduced to the one in [15], by setting $\sigma_{\boldsymbol{\varepsilon},ij}^2 = 1$ for $i = 1, \dots, N_S$ and $j = 1, \dots, N_D$, and $\kappa(\mathbf{r}) = \lambda_1$. Then, the error bound for $\text{Bias}^2[\mu_{MAP}^D]$ can be reduced to the combination of the two error bounds in [15], and $\text{Var}[\mu_{MAP}^D]$ vanishes as shown in the proof of Theorem 2.

Finally, we can combine Theorem 1 and 2 to obtain an upper bound for $\text{MSE}[\mu_{MAP}^D]$ using the fact that B_i and V_i , for $i = 1, 2, 3$, are all positive terms:

$$\text{MSE}[\mu_{MAP}^D] = \text{Bias}^2[\mu_{MAP}^D] + \text{Var}[\mu_{MAP}^D] \\ \leq C_M \left[\sum_{i=1}^3 B_i + \sum_{i=1}^3 V_i \right]^2, \quad (32)$$

where $C_M = \max\{C_B, C_V\}$, and B_i and V_i , $i = 1, 2, 3$, are defined in Theorem 1 and 2.

B. General Models for Noise and Fluorophore Concentration

For a more general *a priori* second order statistical model of the fluorophore concentration, we consider $\text{Cov}_{\mu\mu}(\mathbf{r}, \hat{\mathbf{r}}) = \kappa(\mathbf{r}, \hat{\mathbf{r}})$ as the kernel of a positive definite operator $\mathcal{K} : L^2(\Omega) \rightarrow L^2(\Omega)$. Similarly, for a general noise model, we assume that the covariance matrix, $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$, of the measurement noise is a positive definite matrix that is not necessarily diagonal. Then, the variational form of the inverse problem formulation becomes

$$\mathcal{F}_{MAP}(\psi, \mu_{MAP}) = (\psi, \mathcal{A}_0^* \boldsymbol{\Sigma} \boldsymbol{\varepsilon}^{-1} \boldsymbol{\Gamma}) + (\psi, \mathcal{K}^{-1} \mu_0),$$

where

$$\mathcal{F}_{MAP}(\psi, \mu) := (\psi, \mathcal{B}\mu) + (\psi, \mathcal{K}^{-1} \mu).$$

In this case, it can be shown that the upper bounds for $\text{Bias}^2[\mu_{MAP}^D]$ and $\text{Var}[\mu_{MAP}^D]$ are also of the same forms as in (30) and (31) with new coefficients given by

$$F_{ij}^1 = \frac{2}{\|\kappa^{-1}\|_\infty} \sum_{p=1}^{N_S N_D} (\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1})_{p,(i-1)N_D+j} \|g_j^* \phi_i\|_0,$$

$$F_{ij}^2 = \frac{1}{\|\kappa^{-1}\|_\infty} \sum_{p=1}^{N_S N_D} (\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1})_{p,(i-1)N_D+j} |\bar{\Gamma}_{i,j}|,$$

$$F_{ij}^3 = \frac{1}{\|\kappa^{-1}\|_\infty} \sum_{p=1}^{N_S N_D} (\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1})_{p,(i-1)N_D+j},$$

$$I_{ij}^1 = \frac{1}{\|\kappa^{-1}\|_\infty} \sum_{p=1}^{N_S N_D} (\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1})_{p,(i-1)N_D+j} \|G_j^* \Phi_i\|_0,$$

$$I_t^2 = \frac{1}{\|\kappa^{-1}\|_\infty} \|\kappa_0\|_0 \|\kappa_0\|_{\infty,t},$$

$$D_{ij} = \left[(\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})_{(i-1)N_D+j,(i-1)N_D+j} \right. \\ \left. + \int_{\Omega} \int_{\Omega} \kappa(\mathbf{r}, \hat{\mathbf{r}}) a_{ij}^{*,0}(\mathbf{r}) a_{ij}^0(\hat{\mathbf{r}}) d\mathbf{r} d\hat{\mathbf{r}} \right]^{1/2},$$

where κ^{-1} denotes the kernel of the operator $\mathcal{K}^{-1} : L^2(\Omega) \rightarrow L^2(\Omega)$, i.e.,

$$(\mathcal{K}^{-1}\mu)(\mathbf{r}) = \int_{\Omega} \kappa^{-1}(\mathbf{r}, \hat{\mathbf{r}})\mu(\hat{\mathbf{r}})d\hat{\mathbf{r}},$$

and $\kappa_0(\mathbf{r})$ is the Kolmogorov decomposition of $\kappa^{-1}(\mathbf{r}, \hat{\mathbf{r}})$ [33]; $(\Sigma_{\mathcal{E}})_{p,q}$ and $(\Sigma_{\mathcal{E}^{-1}})_{p,q}$, for $p, q = 1, \dots, N_S N_D$, denote the entries on the p^{th} row and the q^{th} column of $\Sigma_{\mathcal{E}}$ and $\Sigma_{\mathcal{E}^{-1}}$, respectively. Clearly, these coefficients reduce to those in Theorem 1 and 2 when the independent noise and fluorophore concentration models are considered in FDOT reconstruction.

C. Implications of Theorem 1 and 2 for Discretizations of Forward and Inverse Problems

In this subsection, we discuss the implications of the error bounds given in Theorem 1 and 2 for the discretization of the forward and inverse problems of FDOT.

Equation (30) in Theorem 1 presents an upper bound for $\text{Bias}^2[\mu_{MAP}^D]$, which takes into account the noise statistics and the *a priori* information of fluorophore concentration, in addition to the factors such as the interdependence between the forward and inverse problem solutions, the source-detector configuration, and their positions with respect to the fluorophore heterogeneity. In this error bound, B_1 and B_2 represent the contribution from the forward problem discretization. To keep these quantities small, the mesh parameters h_{ni} and h_{mj} of the n^{th} and m^{th} elements in the meshes used in solving Φ_i and G_j^* , respectively, have to be chosen small when their corresponding scaling factors

$$\sum_{j=1}^{N_D} (F_{ij}^1 \|g_j^* \bar{\mu}_{MAP}\|_{0,ni} + F_{ij}^2 \|g_j^*\|_{\infty,ni}) \|\phi_i\|_{1,ni},$$

and

$$\sum_{i=1}^{N_S} (F_{ij}^1 \|\phi_i \bar{\mu}_{MAP}\|_{0,mj} + F_{ij}^2 \|\phi_i\|_{\infty,mj}) \|g_j^*\|_{1,mj},$$

are large on those elements. Further examination of these factors suggests an adaptive refinement scheme within each mesh, because $\|g_j^* \bar{\mu}_{MAP}\|_{0,ni}$, $\|g_j^*\|_{\infty,ni}$, $\|\phi_i\|_{1,ni}$, $\|\phi_i \bar{\mu}_{MAP}\|_{0,mj}$, $\|\phi_i\|_{\infty,mj}$, and $\|g_j^*\|_{1,mj}$ all vary within the mesh. This mesh refinement scheme is similar to the one suggested by Theorem 1 in our previous work [15]: For the i^{th} source or the j^{th} detector, it refines the mesh close to that source or detector, as well as around the fluorophore heterogeneity and other detectors or sources. At the same time, the coefficients $F_{ij}^1 = 2\|g_j^* \phi_i\|_0 \|\kappa\|_{\infty} / \sigma_{\mathcal{E},ij}^2$ and $F_{ij}^2 = |\bar{\Gamma}_{i,j}| \|\kappa\|_{\infty} / \sigma_{\mathcal{E},ij}^2$ in B_1 and B_2 may vary for different source-detector pairs. To keep B_1 and B_2 low, one has to generate finer meshes for the source-detector pairs with smaller noise variances (higher F_{ij}^1 and F_{ij}^2), as compared to those pairs with larger noise variances. In this respect, the error bound in Theorem 1 suggests a new adaptive mesh refinement scheme across different meshes in solving Φ_i and G_j^* based on the measurements and the noise statistics. We note that this is a major difference between the implications of the error bounds in this paper and those presented in our previous work [15].

In B_3 , which corresponds to the contribution from the inverse problem discretization, the discretization parameter h_t of the inverse mesh is not only scaled by the inverse problem solution $\|\bar{\mu}_{MAP}\|_{1,t}$, but also scaled by the finite element solutions of the forward problem, the noise variance, and the *a priori* information of the fluorophore concentration:

$$\sum_{i,j}^{N_S, N_D} I_{ij}^1 \|G_j^* \Phi_i\|_{0,t} + I_t^2,$$

where $I_{ij}^1 = \|\kappa\|_{\infty} \|G_j^* \Phi_i\|_0 / \sigma_{\mathcal{E},ij}^2$, and $I_t^2 = \|\kappa\|_{\infty} \|1/\kappa\|_{\infty, t}$. This result also suggests a new adaptive meshing criteria for the inverse problem, not only based on the forward and inverse problem solutions, but also based on the noise statistics and *a priori* information of the fluorophore concentration. More specifically, to keep B_3 low, one has to refine the mesh around the heterogeneity of fluorophore concentration, the source-detector pairs with low noise variances, as well as the region where the fluorophore concentration has low variance.

Equation (31) in Theorem 2 shows the effect of the forward and inverse problem discretizations, the *a priori* information of the fluorophore concentration, as well as the noise on $\text{Var}[\mu_{MAP}^D]$.

In this error bound, V_1 and V_2 correspond to the contribution from the forward problem discretization, and V_3 corresponds to the contribution from the inverse problem discretization. This error bound has a similar form as the one in (30), but $\bar{\Gamma}_{ij}$ is replaced with the standard deviation, D_{ij} , of the $(i, j)^{\text{th}}$ measurement, and $\bar{\mu}_{MAP}$ is replaced with $D_{ij}\pi_{ij}$, where π_{ij} is the image reconstructed by the imaging system using the basis vector \mathbf{e}_{ij} in the measurement space $\mathbb{R}^{N_S N_D}$. This result indicates that $\text{Var}[\mu_{MAP}^D]$ is independent of the fluorophore concentration, but depends explicitly on the noise statistics, as well as the factors related to the imaging geometry and the background optical properties, which are incorporated into the error bound through the functions π_{ij} . More specifically, $D_{ij}\pi_{ij}$ indicates where μ_{MAP} may have high variance due to the $(i, j)^{\text{th}}$ measurement. Therefore, to keep the error bound in (31) low, one has to refine the mesh in the region where $D_{ij}\pi_{ij}$, has high value, in addition to the region close to the sources and detectors.

VI. ADAPTIVE MESHING ALGORITHMS IN THE PRESENCE OF MEASUREMENT NOISE AND *A Priori* INFORMATION ON FLUOROPHORE CONCENTRATION

In this section, based on Theorem 1 and 2 given in Section V, we present two new adaptive meshing algorithms for FDOT forward and inverse problems. Taking the noise statistics and *a priori* information on fluorophore concentration into account and using $\text{MSE}[\mu_{MAP}^D]$ as a figure of merit, these algorithms adaptively discretize the FDOT problem to minimize the error due to discretization in the mean square sense. In the following, we first give the error indicator based on the error bound in (32) for each element in the mesh used in solving the forward or inverse problem, then we describe the steps of the algorithms in detail.

For the forward problem discretization, we aim to minimize the summation $\sum_{i=1}^2 [B_i + V_i]$ in (32). Rearranging the terms

in this summation, we obtain

$$\sum_{i=1}^2 [B_i + V_i] = \sum_{i=1}^{N_S} \sum_{n=1}^{N_{\Delta}^i} \varepsilon_f^i(n) + \sum_{j=1}^{N_D} \sum_{m=1}^{N_{\Delta}^{*j}} \varepsilon_f^j(m), \quad (33)$$

where

$$\begin{aligned} \varepsilon_f^i(n) := & \\ & \sum_{j=1}^{N_D} \left(F_{ij}^1 \left(\|g_j^* \bar{\mu}_{MAP}\|_{0,ni} + \sum_{i',j'}^{N_S, N_D} \|g_j^* D_{i'j'} \pi_{i'j'}\|_{0,ni} \right) \right. \\ & \left. + (F_{ij}^2 + F_{ij}^3 D_{ij}) \|g_j^*\|_{\infty,ni} \right) \|\phi_i\|_{1,ni} h_{ni}, \quad (34) \end{aligned}$$

$$\begin{aligned} \varepsilon_f^j(m) := & \\ & \sum_{i=1}^{N_S} \left(F_{ij}^1 \left(\|\phi_i \bar{\mu}_{MAP}\|_{0,mj} + \sum_{i',j'}^{N_S, N_D} \|\phi_i D_{i'j'} \pi_{i'j'}\|_{0,mj} \right) \right. \\ & \left. + (F_{ij}^2 + F_{ij}^3 D_{ij}) \|\phi_i\|_{\infty,mj} \right) \|g_j^*\|_{1,mj} h_{mj}. \quad (35) \end{aligned}$$

Each $\varepsilon_f^i(n)$ (or $\varepsilon_f^j(m)$) entails the contribution of the n^{th} (or m^{th}) element of the mesh used in solving Φ_i (or G_j^*), to the MSE. Equation (33) shows that, the contribution of the forward problem discretization to MSE can be expressed as a summation of the contribution of each element in all meshes used in solving Φ_i and G_j^* . Thus, we use $\varepsilon_f^i(n)$ and $\varepsilon_f^j(m)$ as the error indicators in adaptive mesh refinement for the forward problem solution.

As discussed in Section V, since both theorems suggest an adaptive refinement scheme across all meshes used in solving Φ_i and G_j^* , the new adaptive meshing algorithm limits the total number of nodes in all forward problem meshes, instead of separately limiting the number of nodes in each mesh used for solving Φ_i or G_j^* . For the adaptive refinement process, the algorithm is initiated with a set of coarse uniform meshes. With each sweep of refinement and for each source or detector, it computes the error indicator $\varepsilon_f^i(n)$ or $\varepsilon_f^j(m)$ on every element and the average value $\bar{\varepsilon}_f$ of the error indicators on all elements in all meshes. Every element with $\varepsilon_f^i(n) > \bar{\varepsilon}_f$ or $\varepsilon_f^j(m) > \bar{\varepsilon}_f$ is refined thereafter. By doing so, the resulting meshes provide spatially varying resolution not only within each mesh, but also among all forward problem meshes. The algorithm is stopped when the total number of nodes in all forward problem meshes reaches a predetermined allowable limit. Algorithm 1 describes the detailed steps of this refinement process in the form of a pseudocode.

For the inverse problem discretization, we aim to minimize the summation $[B_3 + V_3]$ in (32). Rearranging the terms in this summation, we obtain

$$B_3 + V_3 = \sum_{t=1}^{N_{\Delta}} \varepsilon_i(t),$$

Algorithm 1 The pseudocode of the adaptive meshing algorithm for the forward problem.

◇ Generate the initial uniform meshes for all forward problems:

(Δ^i, N_{Δ}^i) , $\Delta^i = \bigcup_{n=1}^{N_{\Delta}^i} \{\Delta_n\}$, $i = 1, \dots, N_S$, and

$(\Delta^{*j}, N_{\Delta}^{*j})$, $\Delta^{*j} = \bigcup_{m=1}^{N_{\Delta}^{*j}} \{\Delta_m\}$, $j = 1, \dots, N_D$

◇ Set the maximum number of nodes N_{max}^f in all meshes

while number of nodes in all meshes less than N_{max}^f

for $i = 1, \dots, N_S$ and $j = 1, \dots, N_D$

for each element $\Delta_n \in \Delta^i$ with mesh parameter h_{ni} or $\Delta_m \in \Delta^{*j}$ with mesh parameter h_{mj}

if first linearization

 • Use analytical solutions for ϕ_i and g_j^* and *a priori* information about $\bar{\mu}_{MAP}$ and π_{ij} to compute $\varepsilon_f^i(n)$ in (34) or $\varepsilon_f^j(m)$ in (35)

else

 • Use current solution updates Φ_i , G_j^* , $\bar{\mu}_{MAP}^D$ and Π_{ij} to compute $\varepsilon_f^i(n)$ in (34) or $\varepsilon_f^j(m)$ in (35)

end

end

 • Compute $\bar{\varepsilon}_f$

 • Refine the elements with $\varepsilon_f^i(n) > \bar{\varepsilon}_f$ or $\varepsilon_f^j(m) > \bar{\varepsilon}_f$

 • Update the mesh Δ^i , $i = 1, \dots, N_S$, and Δ^{*j} , $j = 1, \dots, N_D$.

end

◇ Solve for Φ_i , $i = 1, \dots, N_S$, and G_j^* , $j = 1, \dots, N_D$

Algorithm 2 The pseudocode of the adaptive meshing algorithm for the inverse problem.

◇ Generate an initial uniform mesh:

(Δ, N_{Δ}) , $\Delta = \bigcup_{t=1}^{N_{\Delta}} \{\Delta_t\}$

◇ Set the maximum number of nodes N_{max}^i

while number of nodes N less than N_{max}^i

for each element $\Delta_t \in \Delta$ with mesh size parameter h_t

if first linearization

 • Use current solution updates Φ_i , G_j^* and *a priori* information about $\bar{\mu}_{MAP}$, π_{ij} to compute $\varepsilon_i(t)$ in (36)

else

 • Use current solution updates Φ_i , G_j^* , $\bar{\mu}_{MAP}^D$ and Π_{ij} to compute $\varepsilon_i(t)$ in (36)

end

 • Compute $\bar{\varepsilon}_i$

 • Refine the elements with $\varepsilon_i(t) > \bar{\varepsilon}_i$

 • Update the mesh Δ

end

◇ Solve for μ_{MAP}^D and Π_{ij}

where

$$\begin{aligned} \varepsilon_i(t) := & \left(\sum_{i,j}^{N_S, N_D} I_{ij}^1 \|G_j^* \Phi_i\|_{0,t} + I_t^2 \right) \\ & \cdot \left(\|\bar{\mu}_{MAP}\|_{1,t} + \sum_{i,j}^{N_S, N_D} \|D_{ij} \pi_{ij}\|_{1,t} \right) h_t. \quad (36) \end{aligned}$$

Clearly, $\varepsilon_i(t)$ is the contribution of the t^{th} element of the mesh, used in solving the inverse problem, to the MSE. Therefore, we use $\varepsilon_i(t)$ as the error indicators in adaptive mesh refinement for the inverse problem solution.

Our new algorithm for the inverse problem starts from a coarse uniform mesh. In each sweep of the refinement, it computes ε_i for each element and the average value $\bar{\varepsilon}_i$ for all elements, and refines those elements with $\varepsilon_i(t) > \bar{\varepsilon}_i$. The algorithm stops when the total number of nodes exceeds a predetermined allowable limit. Algorithm 2 describes the detailed steps of this refinement process in the form of a pseudocode.

The practical implementations of both algorithms require several adjustments: Since ϕ_i , g_j^* , $\bar{\mu}_{MAP}$, and π_{ij} in (34), (35) and (36) can not be computed exactly, we use the analytical solution of the diffusion equation on an unbounded domain to approximate g_j^* and ϕ_i [16] and *a priori* information about $\bar{\mu}_{MAP}$ and π_{ij} in the first iteration and the updated finite-dimensional solutions thereafter. Also note that in both Algorithm 1 and 2, one needs to solve for π_{ij} , $i = 1, 2, \dots, N_S$, and $j = 1, 2, \dots, N_D$. π_{ij} s can be computed once numerically given the source-detector geometry and background optical properties, and stored for the computation of the error indicators in adaptive mesh refinement.

Finally, for the forward and inverse problem discretizations, the computational complexity of our adaptive meshing algorithms can be reduced from $\mathcal{O}(N_D N_\Delta^i)$, $\mathcal{O}(N_S N_\Delta^{*j})$ or $\mathcal{O}(N_S N_D N_\Delta)$ to $\mathcal{O}(N_\Delta^i)$, $\mathcal{O}(N_\Delta^{*j})$ or $\mathcal{O}(N_\Delta)$, respectively, by using approximations similar to those given in our previous works [14], [16]. With these modifications, our new adaptive meshing algorithms have the same computational complexity as that of the conventional method.

VII. NUMERICAL SIMULATION

In this section, we demonstrate the implications of our error analysis and the performance of our new adaptive meshing algorithms in a set of numerical simulations. We primarily focus on showing the effect of measurement noise on the discretization as well as the FDOT image reconstruction in the simulation study. For the effect of *a priori* information, see [34] for a detailed simulation study.

In the following sections, we first describe the setup of our simulation study, then we present the results of adaptive mesh generation and FDOT image reconstruction.

A. Simulation Setup

In the numerical simulation study, we considered a 6 cm \times 6 cm \times 3 cm cubic domain Ω shown in Fig. 1. We set the homogeneous background absorption coefficient $\mu_{axe} = \mu_{ame} = 0.05 \text{ cm}^{-1}$ and diffusion coefficient $D = 0.0410 \text{ cm}$ for both excitation and emission wavelengths, and set the refractive index mismatch parameter $\rho = 3$ for the boundary $\partial\Omega$. At the center of the domain, we placed a fluorophore heterogeneity with 3 mm radius and constant absorption coefficient $\mu_{axf} = 0.015 \text{ cm}^{-1}$. In the rest of the domain, we assumed $\mu_{axf} = 0$. To reconstruct the fluorophore concentration image, we placed 36 sources and 36 detectors evenly on two

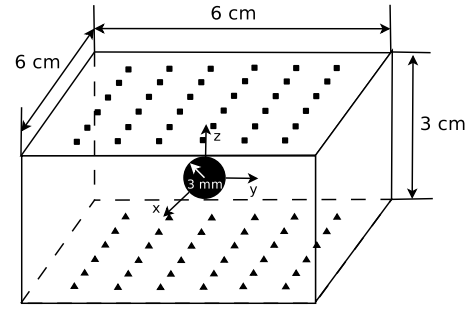


Fig. 1. The optical domain and source-detector configuration for the simulation study. The squares and triangles denote the detectors and sources, respectively.

6 \times 6 grids at the bottom and top surfaces of the domain, as shown in Fig. 1. We simulated both the excitation and emission light fields by solving the coupled diffusion equations (1) and (3) with their corresponding boundary conditions (2) and (4), using the parameters above on a fine uniform grid with $81 \times 81 \times 41$ nodes.

To simulate measurement noise, we considered a shot-noise model described in [35]. When a sufficiently large number of photons are detected, the Poisson distribution can be approximated by a Gaussian distribution with the variance proportional to the magnitude of the measurements. In this case, the variance, $\sigma_{\varepsilon,ij}^2$, of each noise component is given by $\alpha |\Gamma_{0,ij}|$, where $\Gamma_{0,ij}$ is the noise-free measurement obtained at the j^{th} detector due to the i^{th} source. We define the Signal-to-Noise-Ratio (SNR) of the measurements as

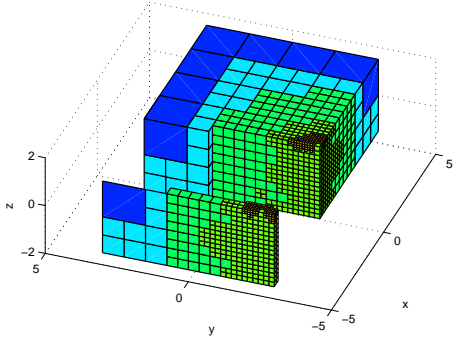
$$\text{SNR}_{ij} = 10 \log_{10} \frac{|\Gamma_{0,ij}|^2}{\sigma_{\varepsilon,ij}^2} = 10 \log_{10} \frac{|\Gamma_{0,ij}|}{\alpha}.$$

Note that, each measurement, Γ_{ij} , has a different SNR proportional to $\log_{10} |\Gamma_{0,ij}|$. We simulated the noise ε for 3 different values of α : 5×10^{-11} , 1×10^{-9} and 5×10^{-9} , corresponding to approximately 40, 26 and 20 dB average SNR over all measurements Γ_{ij} , $i = 1, \dots, N_S$ and $j = 1, \dots, N_D$. For each value of α , we generated 100 different realizations of noise and obtain three sets of noise contaminated measurements with approximately 1%, 5% and 10% noise.

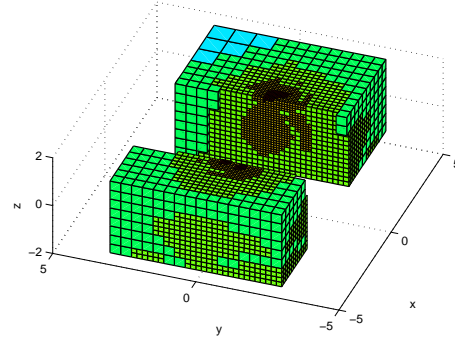
In the FDOT reconstruction, we considered a simplified *a priori* model for the fluorophore concentration, and set $\kappa(\mathbf{r}) = \kappa_0$, where $\kappa_0 = 5 \times 10^{-6}$ is a constant chosen empirically. Finally, we note that we performed our simulation study using deal.II finite element C++ library [36] and used hexahedral finite elements with trilinear Lagrange basis functions to discretize both the forward and inverse problems. We used the Gaussian quadrature method to evaluate the integrals in the variational problems (17), (18) and (25), as well as in calculating the function norms of the finite dimensional solutions on an element.

B. Simulation Results - Mesh Generation

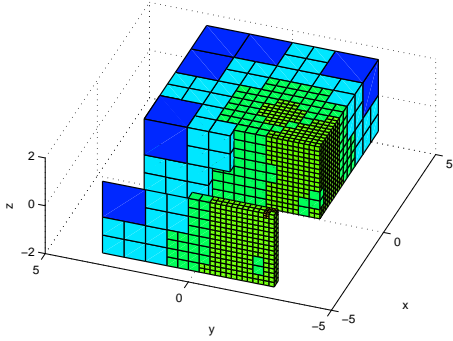
We used three different types of coarse meshes: uniform meshes, the adaptive meshes generated by our previous algorithms in [16], and the adaptive meshes generated by our new algorithms described in Section VI, to reconstruct the



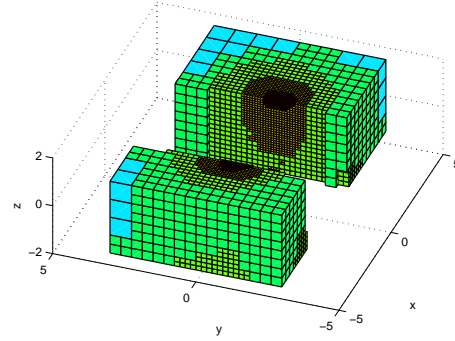
(a) The adaptive mesh with 3289 nodes generated by our new algorithm for the detector located at $(-2.5, -2.5, 1.5)$ for the 1% noise case.



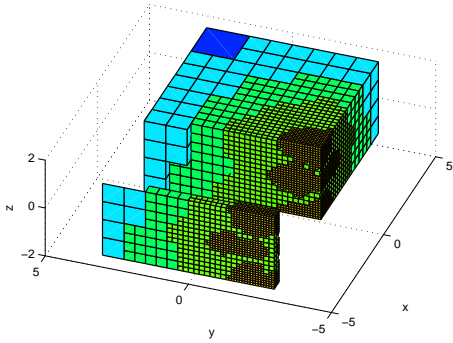
(b) The adaptive mesh with 18876 nodes generated by our new algorithm for the detector located at $(-0.5, -0.5, 1.5)$ for the 1% noise case.



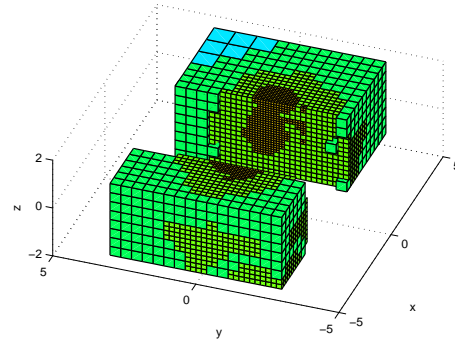
(c) The adaptive mesh with 3588 nodes generated by our new algorithm for the detector located at $(-2.5, -2.5, 1.5)$ for the 10% noise case.



(d) The adaptive mesh with 18054 nodes generated by our new algorithm for the detector located at $(-0.5, -0.5, 1.5)$ for the 10% noise case.



(e) The adaptive mesh with 8304 nodes generated by our previous algorithm for the detector located at $(-2.5, -2.5, 1.5)$.



(f) The adaptive mesh with 7973 nodes generated by our previous algorithm for the detector located at $(-0.5, -0.5, 1.5)$.

Fig. 2. Examples of the adaptive meshes for the forward problem used in the simulation study. The mesh is cut through to show the mesh structure inside.

fluorophore concentration image. For the forward problem, the total number of nodes in the meshes used to solve all Φ_i and G_j^* , $i = 1, \dots, N_S$ and $j = 1, \dots, N_D$, ranges from 500,000 to 650,000 (roughly 7000 to 9000 for each mesh); and for the inverse problem, it ranges from 2000 to 3000. Note that the uniform meshes used in solving the forward and inverse problems have $25 \times 25 \times 13$ nodes and $17 \times 17 \times 9$ nodes, respectively.

For the forward problem, the examples of the adaptive meshes generated for the detectors located at $(-2.5, -2.5, 1.5)$ and $(-0.5, -0.5, 1.5)$ in the 1% and 10% noise cases are

shown in Fig. 2. Figs. 2(a) - 2(d) show the meshes generated by our new algorithm. We observe there are more nodes in the meshes for the detector located at $(-0.5, -0.5, 1.5)$ than in the meshes for the detector located at $(-2.5, -2.5, 1.5)$. Figs. 2(e) and 2(f) show the corresponding meshes generated by our previous algorithm, and these two meshes have approximately same number of nodes. We plotted the relationship between the number of nodes in the meshes generated by our new and previous algorithms for a certain source or detector and the distance of that source or detector to the center of the fluorophore heterogeneity in Fig. 3. Note that for the sources and

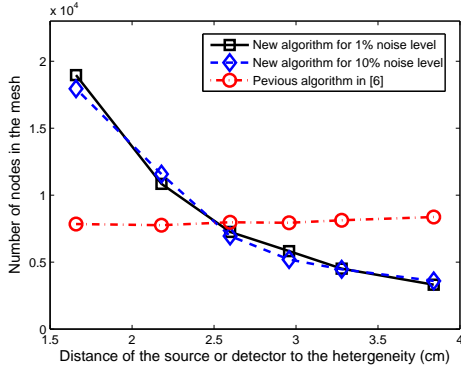


Fig. 3. The relationship between the number of nodes in the forward adaptive mesh for a certain source (or detector) and the distance of the source (or detector) to the center of the fluorophore heterogeneity in 1% and 10% noise cases.

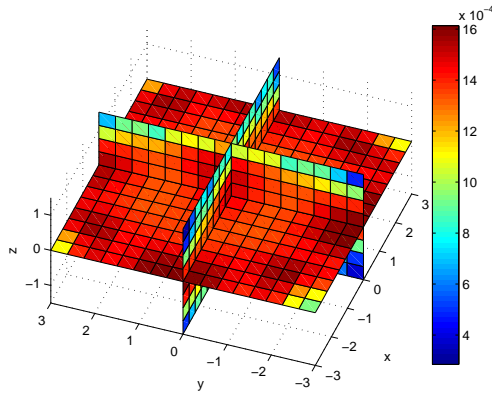
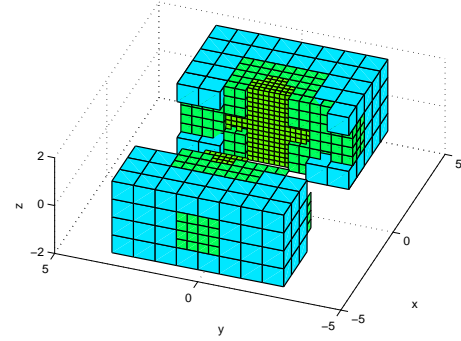


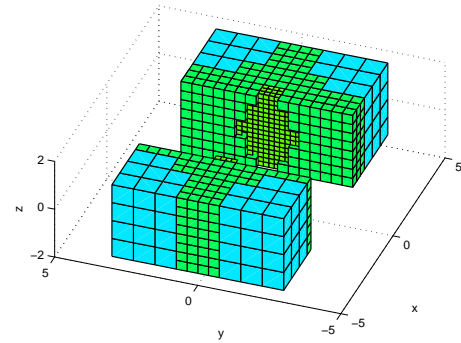
Fig. 4. The cross-section of the image $\sum_{i,j}^{N_S, N_D} D_{ij} \pi_{ij}$, reconstructed using coarse uniform meshes using data with 10% noise.

detectors which have the same distance to the heterogeneity, we plotted the average number of nodes in the corresponding meshes. For our new algorithm, we observe that the closer the sources or detectors to the heterogeneity, the larger the number of nodes is in the associated meshes. This can be explained with the fact that for those source-detector pairs closer to the heterogeneity, the measurements have higher SNR. As a result, our new algorithm generates finer meshes for these source-detector pairs, so that the accuracy of the corresponding forward problem solutions can match the accuracy of the measurements. This result in our forward problem meshes with varying resolution for different source-detector pairs unlike our previous algorithm.

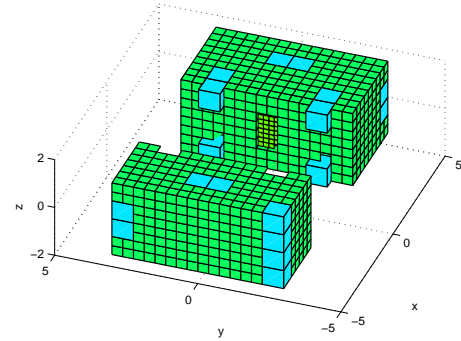
We note the difference between the adaptive meshes generated for the cases of 1% and 10% measurement noise in Figs. 2(a) - 2(d), which illustrates the impact of the noise level on the forward adaptive meshes generated by our new algorithm. Since the value of coefficient D_{ij} increases as $\sigma_{\epsilon,ij}^2$ increases, the images $D_{ij} \pi_{ij}$, $i = 1, \dots, N_S$, $j = 1, \dots, N_D$, have more contribution to the mesh refinement when the noise level is high. Fig. 4 shows the cross-section of $\sum_{i,j} D_{ij} \pi_{ij}$ for the 10% noise case. These results indicate that our new algorithm can refine the meshes adaptively according to the measurement noise level, while our previous algorithm gener-



(a) The adaptive mesh with 2721 nodes generated by our new algorithm for the inverse problem for the 1% noise case.



(b) The adaptive mesh with 2785 nodes generated by our new algorithm for the inverse problem for the 10% noise case.



(c) The adaptive mesh with 2652 nodes generated by our previous algorithm for the inverse problem.

Fig. 5. Examples of the adaptive meshes for the inverse problem used in the simulation study. The mesh is cut through to show the mesh structure inside.

ates the same mesh for different noise levels.

Fig. 5 shows sample adaptive meshes for the inverse problem. Figs. 5(a) and 5(b) show two different meshes generated by our new algorithm for 1% and 10% noise levels, respectively. These meshes are refined around the fluorophore heterogeneity as well as around the nearby sources and detectors. This shows that our new algorithm takes into account the noise statistics, and adaptively refines the mesh according to the noise level. On the other hand, our previous algorithm neglects the impact of noise on the discretization, and generates the same mesh for different noise levels as shown in Fig. 5(c).

TABLE II
MEAN-SQUARE-ERROR, BIAS AND VARIANCE OF THE IMAGES RECONSTRUCTED BY USING DIFFERENT MESHES IN THE SIMULATION STUDY.

Noise Level	Images	Bias ² ($\times 10^{-4}$)		Var ($\times 10^{-6}$)		MSE ($\times 10^{-4}$)	
1%	$\mu_{MAP,U}^D$	3.295	100%	0.037	100%	3.295	100%
	$\mu_{MAP,NA}^D$	0.857	26.00%	0.009	23.49%	0.857	26.00%
	$\mu_{MAP,A}^D$	2.040	61.93%	0.038	101.45%	2.041	61.93%
5%	$\mu_{MAP,U}^D$	3.251	100%	0.893	100%	3.260	100%
	$\mu_{MAP,NA}^D$	0.753	23.16%	0.199	22.27%	0.755	23.16%
	$\mu_{MAP,A}^D$	2.028	62.38%	0.990	110.92%	2.038	62.52%
10%	$\mu_{MAP,U}^D$	3.217	100%	3.670	100%	3.253	100%
	$\mu_{MAP,NA}^D$	0.768	23.89%	0.872	23.77%	0.777	23.89%
	$\mu_{MAP,A}^D$	2.045	63.57%	4.301	117.19%	2.088	64.18%

C. Simulation Results - Image Reconstruction

In this part of the simulation study, we considered 3 sets of reconstructions using 3 sets of measurements at different noise levels. To obtain the exact solutions of the forward and inverse problems, we solved the forward and inverse problems on a fine mesh with $61 \times 61 \times 31$ nodes. We assumed that the error due to discretization in the resulting image, denoted by μ_{MAP} , is negligible with respect to the images reconstructed using the three types of coarse meshes; and used this image as a baseline to compute the MSE. In each reconstruction set, we used $\mu_{MAP,U}^D$, $\mu_{MAP,A}^D$, and $\mu_{MAP,NA}^D$ to denote the images reconstructed using the coarse uniform meshes, the adaptive meshes generated by our previous algorithm in [16] and the new algorithms described in Section VI, respectively.

We calculated the bias, variance and the MSE of the reconstructed images for each set of reconstructions by averaging all reconstructed image samples for 100 realizations of noise. The results are tabulated in Table II. Additionally, we tabulated the percentage of each quantity as compared to the one of the images, $\mu_{MAP,U}^D$, reconstructed by using the coarse uniform meshes: The left column is the absolute value, and the right column is the corresponding percentage. The results in Table II show that the bias squares of the images, reconstructed using different types of meshes, remain at a fixed level when the noise level changes, while the variances of the images increase as the noise level increases. The bias square, the variance as well as the MSE of $\mu_{MAP,NA}^D$ are approximately reduced by 75% as compared to $\mu_{MAP,U}^D$, when our new algorithm is used. On the other hand, our previous algorithm in [16] provides about 40% reduction in the bias square, but no reduction in the variance of $\mu_{MAP,A}^D$ with respect to $\mu_{MAP,U}^D$.

Figs 6 and 7 show the cross-section of the sample images at $z = 0$ plane reconstructed using different types of meshes when the noise level is 1% and 10%. The cross-section of the baseline images are shown in Fig. 6(a) and Fig. 7(a). We observe that the variability of images in Fig. 7 is more visible as compared to that of the images in Fig. 6 due to increased noise level in the measurements. The shape of the small fluorophore heterogeneity is better resolved in $\mu_{MAP,A}^D$ and $\mu_{MAP,NA}^D$ as compared to the one in $\mu_{MAP,U}^D$, due to the spatially varying resolution provided by the adaptive meshes. Additionally, we observe a higher variability in $\mu_{MAP,A}^D$ than that of $\mu_{MAP,NA}^D$ in Fig. 7(d) and Fig. 7(c), while the difference between the images in Fig. 6(d) and Fig. 6(c) is

not as noticeable due to lower noise level. These observations can be seen more clearly in Fig. 8, where the reconstructed images along the y -axis on $z = 0$ plane are shown for 1% and 10% noise cases. The solid lines in Fig. 8(a) and Fig. 8(b) represent the baseline image μ_{MAP} which is assumed to have negligible error. We observe that the image, $\mu_{MAP,NA}^D$, is the best approximation to μ_{MAP} in all three reconstructed images, which has higher response at the center of the fluorophore heterogeneity and lower background variation, as compared to those of $\mu_{MAP,U}^D$ and $\mu_{MAP,A}^D$.

In summary, the simulation study shows that

- 1) The new adaptive meshing algorithms can adaptively discretize the FDOT forward and inverse problems according to the noise level, and, unlike the algorithms in [16], generates the forward problem mesh with varying resolution for different source-detector pairs.
- 2) The new adaptive meshing algorithms can effectively reduce the bias, variance, as well as the MSE of the reconstructed images with respect to the uniform meshing scheme while keeping the sizes of the discretized forward and inverse problems under predetermined allowable limits.
- 3) As compared to our previous adaptive meshing algorithms [16], the new algorithms is more effective in reducing the variance and MSE of reconstructed images, particularly for high levels of measurement noise.

VIII. CONCLUSION

In this work, we analyzed the effect of discretization on the accuracy of FDOT reconstruction in the presence of measurement noise. We formulated the FDOT inverse problem as an optimization problem in the MAP framework to estimate the fluorophore concentration in a bounded domain. To quantitatively assess the accuracy of FDOT reconstruction, we first defined the MSE between the exact solution and the discretized solution of the inverse problem. We, then, identified two components of the MSE: the bias and the variance of the reconstructed image, and derived an upper bound for each component. These upper bounds identify the key factors that determine the extent to which the forward and inverse problem discretizations can affect the accuracy of FDOT reconstruction. These factors include the noise statistics and the *a priori* information of fluorophore concentration in addition to the interdependence between the forward and inverse problem

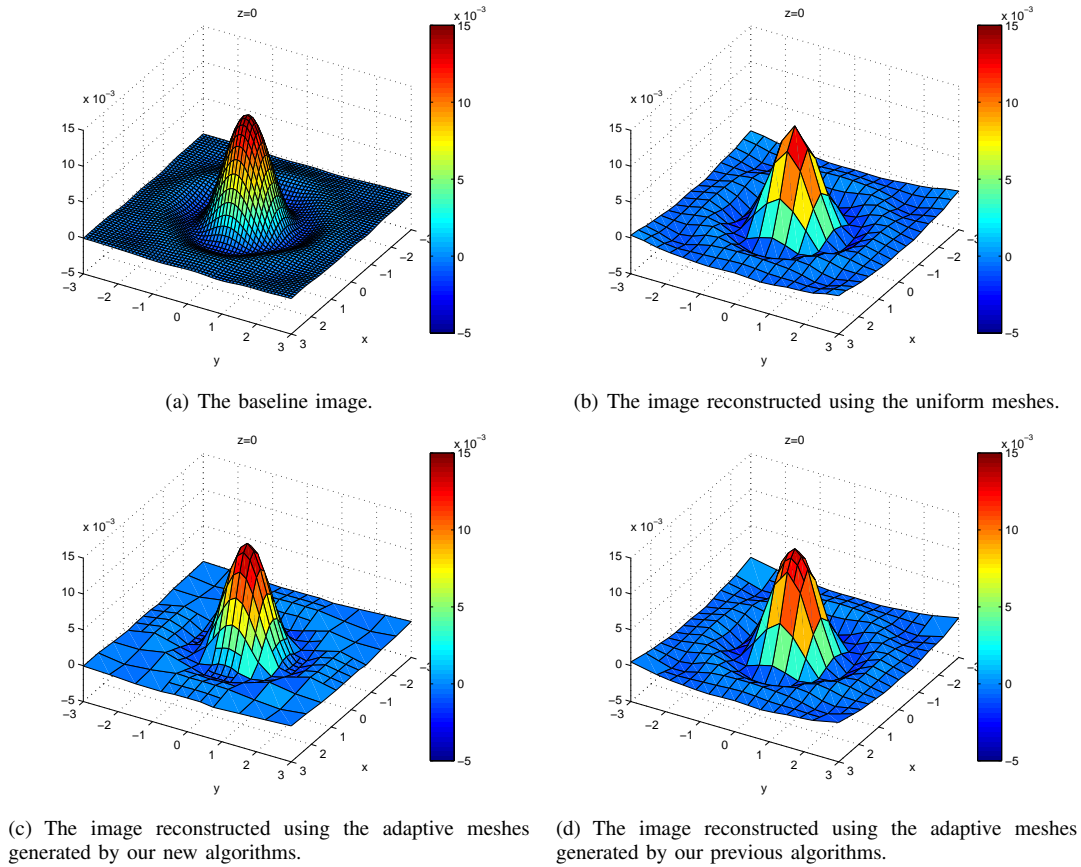


Fig. 6. The cross-section of the reconstructed sample images on $z = 0$ plane in 1% noise case in the simulation study.

solutions, the source-detector configuration, and their positions with respect to the fluorophore heterogeneity.

Based on these bounds, we developed new adaptive meshing algorithms for the FDOT forward and inverse problems to reduce the MSE in reconstructed images. Unlike the algorithms in [16], these algorithms take into account noise statistics, as well as *a priori* information on fluorophore concentration in the adaptive mesh refinement process. We demonstrated the performance of these algorithms in a set of numerical simulations. The simulation results showed that the new algorithms generate adaptive forward meshes with varying resolution not only within each mesh for a certain source-detector pair, but also across the meshes for all source-detector pairs. Additionally, we showed that the meshes generated by the new algorithms can effectively reduce both the bias and the variance of the reconstructed images, thereby effectively reducing the total MSE as compared to the algorithms in [16], as well as the uniform meshing scheme for a fixed number of nodes.

In our inverse problem formulation, the regularization as well as the *a priori* information on fluorophore concentration introduce bias into the reconstructed images with respect to the true fluorophore concentration. This type of error may sometimes overwhelm other types of errors in the reconstructed images. The regularization parameter and the variance of fluorophore concentration provide a way to balance the bias and the error due to the measurement noise [37], [38].

In the development of our adaptive meshing algorithms, we assumed that the optimal regularization parameter as well as the variance of the fluorophore concentration are known *a priori*, and therefore we kept them fixed for different meshing schemes. Taking into account the severely ill-posed nature of the FDOT inverse problem, we compared the performance of the reconstruction for different meshing schemes with the optimally regularized solution obtained with uniform fine meshing to isolate the error due to discretization only. However, since the error bounds given in Theorem 1 and 2 take into account both the regularization and *a priori* information, it is possible to study the interplay between these parameters and the problem discretization, and to adaptively refine the mesh while adjusting parameters to reduce the overall error in the reconstructed images.

Finally, we note that the error analysis approach introduced in this work is not limited to FDOT imaging, and can be extended to analyze the error due to discretization in other PDE-based inverse coefficient estimation problems, such as DOT, bioluminescence tomography, electrical impedance tomography and microwave tomography.

REFERENCES

- [1] R. Weissleder and V. Ntziachristos, "Shedding light onto live molecular targets," *Nature Medicine*, vol. 9, pp. 123–128, 2003.
- [2] V. Ntziachristos, J. Ripoll, L. V. Wang, and R. Weissleder, "Looking and listening to light: the evolution of whole-body photonic imaging," *Nature Biotechnology*, vol. 23, pp. 313–320, 2005.

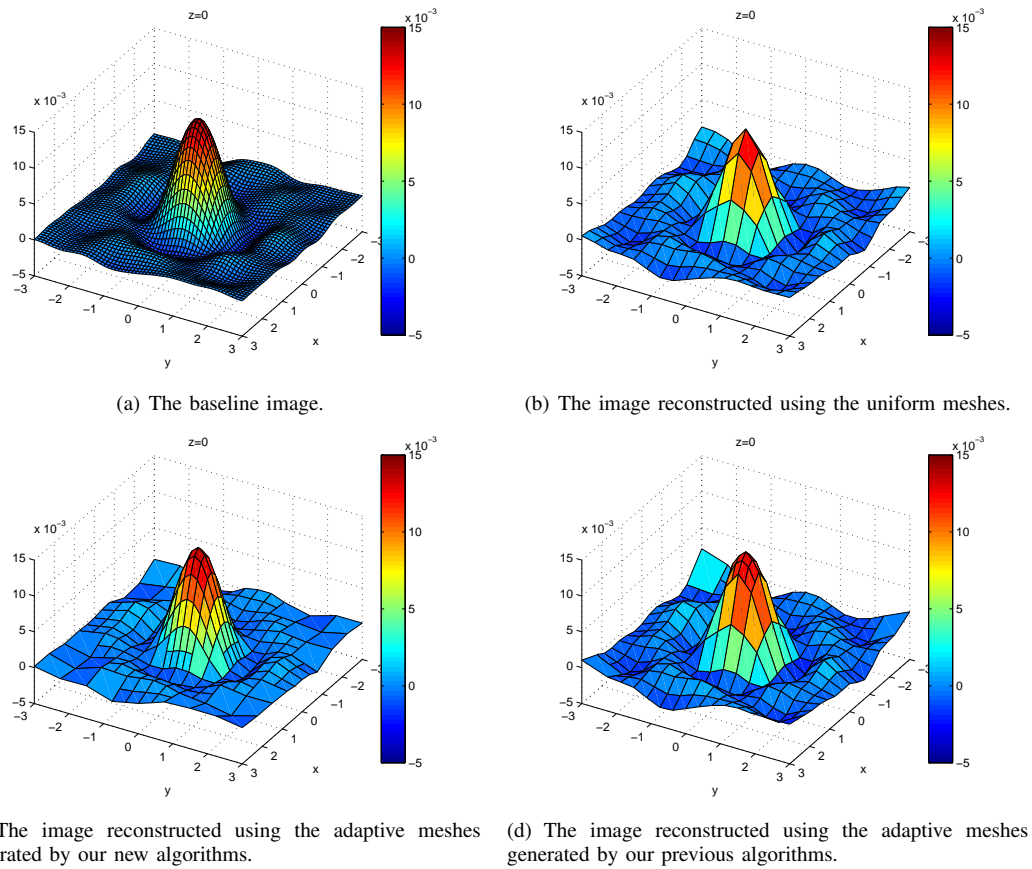


Fig. 7. The cross-section of the reconstructed sample images on $z = 0$ plane in 10% noise case in the simulation study.

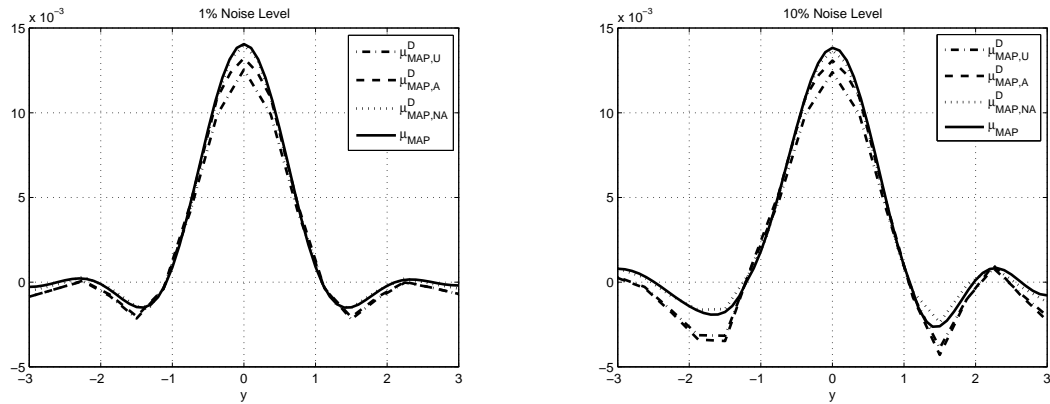


Fig. 8. The sample images along the y -axis in 1% and 10% noise cases.

[3] W. Bangerth, “Adaptive finite element methods for the identification of distributed parameters in partial differential equations,” Ph.D. dissertation, University of Heidelberg, 2002.

[4] R. Li, W. Liu, H. Ma, and T. Tang, “Adaptive finite element approximation for distributed elliptic optimal control problems,” *SIAM J. Control Optim.*, vol. 41, pp. 1321–49, 2002.

[5] R. Becker and B. Vexler, “A posteriori error estimation for finite element discretization of parameter identification problems,” *Numer. Math.*, vol. 96, pp. 435–59, 2003.

[6] D. Meidner and B. Vexler, “Adaptive space-time finite element methods for parabolic optimization problems,” *SIAM J. Control Optim.*, vol. 46, pp. 116–42, 2007.

[7] B. Vexler and W. Wollner, “Adaptive finite elements for elliptic opti-

mization problems with control constraints,” *SIAM J. Control Optim.*, vol. 47, pp. 509–34, 2008.

[8] L. Beilina and C. Johnson, “A posteriori error estimation in computational inverse scattering,” *Mathematical Models and Methods in Applied Sciences*, vol. 15, pp. 23–37, 2005.

[9] —, “Adaptive finite element/difference method for inverse elastic scattering waves,” *Applied and Computational Mathematics*, vol. 2, pp. 158–174, 2003.

[10] W. Bangerth and A. Joshi, “Adaptive finite element methods for the solution of inverse problems in optical tomography,” *Inverse Problems*, vol. 24, p. 034011, 2008.

[11] J. Lee, A. Joshi, and E. Sevick-Muraca, “Fully adaptive finite element based tomography using tetrahedral dual-meshing for fluorescence en-

- hanced optical imaging in tissue,” *Optics Express*, vol. 15, no. 11, p. 6955C6975, 2007.
- [12] —, “Fast intersections on nested tetrahedrons (fint): An algorithm for adaptive finite element based distributed parameter estimation,” *Journal of Computational Physics*, vol. 227, pp. 5778–5798, 2008.
- [13] M. Guven, B. Yazici, K. Kwon, E. Giladi, and X. Intes, “Effect of discretization error and adaptive mesh generation in diffuse optical absorption imaging: I,” *Inverse Problems*, vol. 23, pp. 1115–1133, 2007.
- [14] —, “Effect of discretization error and adaptive mesh generation in diffuse optical absorption imaging: II,” *Inverse Problems*, vol. 23, pp. 1135–1160, 2007.
- [15] M. Guven, L. Reilly-Raska, L. Zhou, and B. Yazici, “Discretization error analysis and adaptive meshing algorithms for fluorescence diffuse optical tomography: Part I,” *IEEE Transactions on Medical Imaging*, vol. 29, no. 2, pp. 217–229, 2010.
- [16] M. Guven, L. Zhou, L. Reilly-Raska, and B. Yazici, “Discretization error analysis and adaptive meshing algorithms for fluorescence diffuse optical tomography: Part II,” *IEEE Transactions on Medical Imaging*, vol. 29, no. 2, pp. 230–245, 2010.
- [17] M. Ainsworth and J. T. Oden, “A unified approach to a posteriori error estimation using elemental residual methods,” *Numerische Mathematik*, vol. 65, pp. 23–50, 1993.
- [18] I. Babuška and W. C. Rheinboldt, “Error estimates for adaptive finite element computations,” *SIAM Journal on Numerical Analysis*, vol. 15, pp. 736–754, 1978.
- [19] I. Babuška, O. C. Zienkiewicz, J. Gago, and E. R. de A. Oliveira, *Accuracy Estimates and Adaptive Refinements in Finite Element Computations*. John Wiley and Sons, 1986.
- [20] R. E. Bank and A. Weiser, “Some a posteriori error estimators for elliptic partial differential equations,” *Mathematics of Computation*, vol. 44, pp. 283–301, 1985.
- [21] T. Strouboulis and K. A. Hague, “Recent experiences with error estimation and adaptivity, part I: Review of error estimators for scalar elliptic problems,” *Computer Methods in Applied Mechanics and Engineering*, vol. 97, pp. 399–436, 1992.
- [22] R. Verfurth, *A Review of A Posteriori Error Estimation and Adaptive Mesh Refinement Techniques*. Teubner-Wiley, 1996.
- [23] L. Zhou, B. Yazici, A. B. F. Ale, and V. Ntziachristos, “Performance evaluation of adaptive meshing algorithms for fluorescence diffuse optical tomography using experimental data,” *submitted to Optics Letters*, 2010.
- [24] E. M. Sevick-Muraca, G. Lopez, J. S. Reynolds, T. L. Troy, and C. L. Hutchinson, “Fluorescence and absorption contrast mechanisms for biomedical optical imaging using frequency-domain techniques,” *Photochem. Photobiol.*, vol. 66, pp. 56–64, 1997.
- [25] S. R. Arridge, “Optical tomography in medical imaging,” *Inverse Problems*, vol. 15, pp. R41–93, 1999.
- [26] V. Ntziachristos and R. Weissleder, “Experimental three-dimensional fluorescence reconstruction of diffuse media by use of a normalized born approximation,” *Optics Letters*, vol. 26, no. 12, pp. 893–895, 2001.
- [27] R. Kress, *Linear integral equations*, 2nd ed., ser. Applied Mathematical Sciences. Springer-Verlag, 1999, vol. 82.
- [28] N. P. Galatsanos and A. K. Katsaggelos, “Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation,” *IEEE Transactions on Image Processing*, vol. 1, pp. 322–336, 1992.
- [29] M. Hanke and T. Raus, “A general heuristic for choosing the regularization parameter in ill-posed problems,” *SIAM Journal on Scientific Computing*, vol. 17, pp. 956–972, 1996.
- [30] R. Molina, A. K. Katsaggelos, and J. Mateos, “Bayesian and regularization methods for hyperparameter estimation in image restoration,” *IEEE Transactions on Image Processing*, vol. 8, pp. 231–246, 1999.
- [31] M. Heath, G. Golub, and G. Wahba, “Generalized cross-validation as a method for choosing a good ridge parameter,” *Technometrics*, vol. 21, p. 215C223, 1979.
- [32] P. Hansen and D. O’Learly, “The use of the l-curve in the regularization of discrete ill-posed problems,” *SIAM J. Comput.*, vol. 14, p. 1487C1503, 1993.
- [33] I. I. Gikhman and A. V. Skorokhod, *The Theory of Stochastic Processes I*. Berlin Heidelberg New York: Springer-Verlag, 1974.
- [34] L. Zhou, “Adaptive finite element methods for fluorescence diffuse optical tomography,” Ph.D. dissertation, Rensselaer Polytechnic Institute, Troy NY, 2010.
- [35] J. Ye, K. J. Webb, and C. A. Bouman, “Optical diffusion tomography by iterative-coordinate-descent optimization in a bayesian framework,” *J. Opt. Soc. Am. A*, vol. 16, no. 10, pp. 2400–12, 1999.
- [36] W. Bangerth, R. Hartmann, and G. Kanschat, “deal.II — a general-purpose object-oriented finite element library,” *ACM Trans. Math. Softw.*, vol. 33, no. 4, 2007.
- [37] N. P. Galatsanos and A. K. Katsaggelos, “Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation,” *IEEE Trans. on Image Processing*, vol. 1, no. 3, pp. 322–36, 1992.
- [38] R. Molina, A. K. Katsaggelos, and J. Mateos, “Bayesian and regularization methods for hyperparameter estimation in image restoration,” *IEEE Trans. on Image Processing*, vol. 8, no. 2, pp. 231–46, 1999.
- [39] G. Fubini, “Sugli integrali multipli,” *Opere scelte*, vol. 2, pp. 243–49, 1958.
- [40] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods*. Springer Verlag, 2002.
- [41] R. A. Horn, *Matrix Analysis*. Cambridge: Cambridge University Press, 1985.

APPENDIX A

PROOF OF THEOREM 1: UPPER BOUND FOR $\text{BIAS}^2[\mu_{MAP}^D]$

Let $\bar{\mu}_{MAP}(\mathbf{r}) := \mathbb{E}[\mu_{MAP}(\mathbf{r})]$ and $\bar{\mu}_{MAP}^D(\mathbf{r}) := \mathbb{E}[\mu_{MAP}^D(\mathbf{r})]$, then $\mathbb{E}[e_{MAP}(\mathbf{r})] = \bar{\mu}_{MAP}(\mathbf{r}) - \bar{\mu}_{MAP}^D(\mathbf{r})$. We can further express $\text{Bias}^2[\mu_{MAP}^D]$ in (28) as

$$\begin{aligned} \text{Bias}^2[\mu_{MAP}^D] &= \int_{\Omega} |\bar{\mu}_{MAP}(\mathbf{r}) - \bar{\mu}_{MAP}^D(\mathbf{r})|^2 d\mathbf{r} \\ &= \|\bar{\mu}_{MAP}(\mathbf{r}) - \bar{\mu}_{MAP}^D(\mathbf{r})\|_0^2. \end{aligned}$$

It is clear that $\text{Bias}^2[\mu_{MAP}^D]$ is the square of $L^2(\Omega)$ norm of the difference between $\bar{\mu}_{MAP}$ and $\bar{\mu}_{MAP}^D$.

Taking the expectation on both sides of (14) and (25), and applying Fubini’s theorem [39], we can show that $\bar{\mu}_{MAP}$ and $\bar{\mu}_{MAP}^D$ satisfy the following variational forms:

$$\mathcal{F}_{MAP}(\psi, \bar{\mu}_{MAP}) = (\psi, \mathcal{A}_0^* \Sigma_{\epsilon}^{-1} \bar{\Gamma}) + (\psi, \frac{\mu_0}{\kappa}), \quad (37)$$

$$\tilde{\mathcal{F}}_{MAP}(\Psi, \bar{\mu}_{MAP}^D) = (\Psi, \tilde{\mathcal{A}}_0^* \Sigma_{\epsilon}^{-1} \bar{\Gamma}) + (\Psi, \frac{\mu_0}{\kappa}), \quad (38)$$

where $\bar{\Gamma} = \mathbb{E}[\Gamma] = \mathcal{A}_0 \mu_0$, according to the noise model and the *a priori* model for fluorophore concentration. Further, we let $\tilde{\mu}_{MAP} \in L^2(\Omega)$ be the solution of the following approximate inverse problem:

$$\tilde{\mathcal{F}}_{MAP}(\psi, \tilde{\mu}_{MAP}) = (\psi, \tilde{\mathcal{A}}_0^* \Sigma_{\epsilon}^{-1} \bar{\Gamma}) + (\psi, \frac{\mu_0}{\kappa}),$$

for all $\psi \in L^2(\Omega)$. Then we have

$$\begin{aligned} \text{Bias}^2[\mu_{MAP}^D] &\leq [\|\bar{\mu}_{MAP} - \tilde{\mu}_{MAP}\|_0 + \|\tilde{\mu}_{MAP} - \bar{\mu}_{MAP}^D\|_0]^2, \quad (39) \end{aligned}$$

by the triangular inequality.

For the first term $\|\bar{\mu}_{MAP} - \tilde{\mu}_{MAP}\|_0$ in (39), we follow the similar procedures given in [15] and obtain

$$\begin{aligned} \|\bar{\mu}_{MAP} - \tilde{\mu}_{MAP}\|_0 &\leq \|\kappa\|_{\infty} \left[\|(\mathcal{B} - \tilde{\mathcal{B}})\bar{\mu}_{MAP}\|_0 \right. \\ &\quad \left. + \left\| (\tilde{\mathcal{A}}_0^* - \mathcal{A}_0^*) \Sigma_{\epsilon}^{-1} \bar{\Gamma} \right\|_0 \right]. \quad (40) \end{aligned}$$

For term $\|(\mathcal{B} - \tilde{\mathcal{B}})\bar{\mu}_{MAP}\|_0$, we have [13]

$$\begin{aligned} & \|(\mathcal{B} - \tilde{\mathcal{B}})\bar{\mu}_{MAP}\|_0 \\ & \approx 2 \left\| \sum_{i,j}^{N_S, N_D} \frac{g_j^* \phi_i}{\sigma_{\tilde{\mathcal{E}}, ij}^2} \int_{\Omega} (g_j^* e_i + \phi_i e_j^*) \bar{\mu}_{MAP} d\mathbf{r} \right\|_0 \\ & \leq 2 \sum_{i,j}^{N_S, N_D} \frac{\|g_j^* \phi_i\|_0}{\sigma_{\tilde{\mathcal{E}}, ij}^2} \int_{\Omega} |(g_j^* e_i + \phi_i e_j^*) \bar{\mu}_{MAP}| d\mathbf{r}, \end{aligned}$$

where $e_i = \phi_i - \Phi_i$ and $e_j^* = g_j^* - G_j^*$. Decomposing the integral on Ω into a summation of the integrals on the finite elements Ω_{mj} , $m = 1, \dots, N_{\Delta}^j$, and Ω_{ni} , $n = 1, \dots, N_{\Delta}^i$, which are used to discretize the forward problem, we arrive at

$$\begin{aligned} & \|(\mathcal{B} - \tilde{\mathcal{B}})\bar{\mu}_{MAP}\|_0 \\ & \leq 2 \left(\sum_{i=1}^{N_S} \sum_{n,j}^{N_{\Delta}^i, N_D} \frac{\|g_j^* \phi_i\|_0}{\sigma_{\tilde{\mathcal{E}}, ij}^2} \|g_j^* \bar{\mu}_{MAP}\|_{0, ni} \|e_i\|_{0, ni} \right. \\ & \quad \left. + \sum_{j=1}^{N_D} \sum_{m,i}^{N_{\Delta}^j, N_S} \frac{\|g_j^* \phi_i\|_0}{\sigma_{\tilde{\mathcal{E}}, ij}^2} \|\phi_i \bar{\mu}_{MAP}\|_{0, mj} \|e_j^*\|_{0, mj} \right). \end{aligned} \quad (41)$$

Similarly, for term $\|(\tilde{\mathcal{A}}_0^* - \mathcal{A}_0^*)\Sigma_{\tilde{\mathcal{E}}}^{-1}\bar{\Gamma}\|_0$, we have its upper bound as

$$\begin{aligned} & \|(\tilde{\mathcal{A}}_0^* - \mathcal{A}_0^*)\Sigma_{\tilde{\mathcal{E}}}^{-1}\bar{\Gamma}\|_0 \\ & \approx \left\| \sum_{i,j}^{N_S, N_D} \frac{g_j^* e_i + \phi_i e_j^*}{\sigma_{\tilde{\mathcal{E}}, ij}^2} \bar{\Gamma}_{i,j} \right\|_0 \\ & \leq \left(\sum_{i=1}^{N_S} \sum_{n,j}^{N_{\Delta}^i, N_D} \frac{|\bar{\Gamma}_{i,j}|}{\sigma_{\tilde{\mathcal{E}}, ij}^2} \|g_j^*\|_{\infty, ni} \|e_i\|_{0, ni} \right. \\ & \quad \left. + \sum_{j=1}^{N_D} \sum_{m,i}^{N_{\Delta}^j, N_S} \frac{|\bar{\Gamma}_{i,j}|}{\sigma_{\tilde{\mathcal{E}}, ij}^2} \|\phi_i\|_{\infty, mj} \|e_j^*\|_{0, mj} \right). \end{aligned} \quad (42)$$

In the end, substituting (41) and (42) into (40) and using the discretization error bounds given by [40]

$$\begin{aligned} \|e_i\|_{0, ni} & \leq C \|\phi_i\|_{1, ni} h_{ni}, \\ \|e_j^*\|_{0, mj} & \leq C \|g_j^*\|_{1, mj} h_{mj}, \end{aligned}$$

lead to B_1 and B_2 in Theorem 1.

For the second term $\|\bar{\mu}_{MAP} - \bar{\mu}_{MAP}^D\|_0$ in (39), we also follow the procedures in [15] and obtain

$$\begin{aligned} \|\bar{\mu}_{MAP} - \bar{\mu}_{MAP}^D\|_0 & \leq \|\kappa\|_{\infty} \left\| \tilde{\mathcal{B}}(\bar{\mu}_{MAP} - \varphi) \right\|_0 \\ & \quad + \left\| \frac{\bar{\mu}_{MAP} - \varphi}{\kappa} \right\|_0. \end{aligned} \quad (43)$$

Let $\varphi \in V(\Omega)$ be the interpolant of $\bar{\mu}_{MAP}$ and $e_{\mu} := \bar{\mu}_{MAP} -$

φ be the interpolation error, we have

$$\begin{aligned} & \left\| \tilde{\mathcal{B}}(\bar{\mu}_{MAP} - \varphi) \right\|_0 \\ & = \left\| \sum_{i,j}^{N_S, N_D} \frac{G_j^*(\cdot)\Phi_i(\cdot)}{\sigma_{\tilde{\mathcal{E}}, ij}^2} \int_{\Omega} G_j^*(\mathbf{r}')\Phi_i(\mathbf{r}')e_{\mu}(\mathbf{r}') d\mathbf{r}' \right\|_0 \\ & \leq \sum_{i,j}^{N_S, N_D} \frac{\|G_j^*\Phi_i\|_0}{\sigma_{\tilde{\mathcal{E}}, ij}^2} \int_{\Omega} |G_j^*(\mathbf{r}')\Phi_i(\mathbf{r}')e_{\mu}(\mathbf{r}')| d\mathbf{r}' \\ & \leq \sum_{i,j}^{N_S, N_D} \frac{\|G_j^*\Phi_i\|_0}{\sigma_{\tilde{\mathcal{E}}, ij}^2} \sum_{t=1}^{N_{\Delta}} \|G_j^*\Phi_i\|_{0,t} \|e_{\mu}\|_{0,t}, \end{aligned} \quad (44)$$

and

$$\left\| \frac{\bar{\mu}_{MAP} - \varphi}{\kappa} \right\|_0 \leq \sum_{t=1}^{N_{\Delta}} \left\| \frac{e_{\mu}}{\kappa} \right\|_{0,t} \leq \sum_{t=1}^{N_{\Delta}} \left\| \frac{1}{\kappa} \right\|_{\infty,t} \|e_{\mu}\|_{0,t}. \quad (45)$$

Assume that $\bar{\mu}_{MAP} \in H^1(\Omega)$. Then an upper bound for the discretization error can be given by

$$\|e_{\mu}\|_{0,t} \leq C \|\bar{\mu}_{MAP}\|_{1,t} h_t.$$

Approximating $\bar{\mu}_{MAP}$ by $\bar{\mu}_{MAP}$, and substituting (44), (45) and the discretization error bound into (43), we obtain B_3 in Theorem 1.

APPENDIX B

PROOF OF THEOREM 2: UPPER BOUND FOR $\text{VAR}[\mu_{MAP}^D]$

We express $\text{Var}[\mu_{MAP}^D]$ in (29) using μ_{MAP} , μ_{MAP}^D , $\bar{\mu}_{MAP}$ and $\bar{\mu}_{MAP}$ as

$$\begin{aligned} \text{Var}[\mu_{MAP}^D] & := \int_{\Omega} \mathbf{E} \left[|\mu_{MAP}(\mathbf{r}) - \mu_{MAP}^D(\mathbf{r}) \right. \\ & \quad \left. - [\bar{\mu}_{MAP}(\mathbf{r}) - \bar{\mu}_{MAP}^D(\mathbf{r})] \right|^2 d\mathbf{r} \\ & = \int_{\Omega} \mathbf{E} \left[|\mu_{MAP}(\mathbf{r}) - \bar{\mu}_{MAP}(\mathbf{r}) \right. \\ & \quad \left. - [\mu_{MAP}^D(\mathbf{r}) - \bar{\mu}_{MAP}^D(\mathbf{r})] \right|^2 d\mathbf{r}. \end{aligned}$$

Subtracting (37) and (38) from (14) and (25), respectively, we obtain

$$\mathcal{F}_{MAP}(\psi, \mu_{MAP}(\mathbf{r}) - \bar{\mu}_{MAP}) = (\psi, \mathcal{A}_0^* \Sigma_{\tilde{\mathcal{E}}}^{-1} (\Gamma - \bar{\Gamma})), \quad (46)$$

$$\tilde{\mathcal{F}}_{MAP}(\Psi, \mu_{MAP}^D(\mathbf{r}) - \bar{\mu}_{MAP}^D) = (\Psi, \tilde{\mathcal{A}}_0^* \Sigma_{\tilde{\mathcal{E}}}^{-1} (\Gamma - \bar{\Gamma})). \quad (47)$$

Let $\mathbf{e}_{ij} \in \mathbb{R}^{N_S N_D}$, $i = 1, \dots, N_S$ and $j = 1, \dots, N_D$, be the unit vector given by

$$\mathbf{e}_{ij} = [0, \dots, 1, \dots, 0]^T,$$

with only non-zero entry at $[(i-1)N_D + j]^{\text{th}}$ position. Then, we define $\pi_{ij}(\mathbf{r}) \in L^2(\Omega)$ and $\Pi_{ij}(\mathbf{r}) \in V(\Omega)$ as the solutions of the following variational problems:

$$\begin{aligned} \mathcal{F}_{MAP}(\psi, \pi_{ij}) & = (\psi, \mathcal{A}_0^* \Sigma_{\tilde{\mathcal{E}}}^{-1} \mathbf{e}_{ij}), \\ \tilde{\mathcal{F}}_{MAP}(\Psi, \Pi_{ij}) & = (\Psi, \tilde{\mathcal{A}}_0^* \Sigma_{\tilde{\mathcal{E}}}^{-1} \mathbf{e}_{ij}), \end{aligned}$$

for all $\psi \in L^2(\Omega)$ and $\Psi \in V(\Omega)$. Clearly, π_{ij} and Π_{ij} are the solutions of (46) and (47) given the unit base vector \mathbf{e}_{ij} in the measurement space $\mathbb{R}^{N_S N_D}$. Due to the linearity of the

bilinear form, any solutions of (46) and (47) can be given as the linear combinations of π_{ij} s and Π_{ij} s, respectively.

We further define $\boldsymbol{\pi}(\mathbf{r})$ and $\boldsymbol{\Pi}(\mathbf{r})$ as

$$\begin{aligned}\boldsymbol{\pi}(\mathbf{r}) &= [\pi(\mathbf{r})_{11}, \dots, \pi(\mathbf{r})_{1N_S}, \pi(\mathbf{r})_{21}, \dots, \pi(\mathbf{r})_{N_S N_D}]^T, \\ \boldsymbol{\Pi}(\mathbf{r}) &= [\Pi(\mathbf{r})_{11}, \dots, \Pi(\mathbf{r})_{1N_S}, \Pi(\mathbf{r})_{21}, \dots, \Pi(\mathbf{r})_{N_S N_D}]^T.\end{aligned}$$

Then $\mu_{MAP}(\mathbf{r}) - \bar{\mu}_{MAP}(\mathbf{r})$ and $\mu_{MAP}^D(\mathbf{r}) - \bar{\mu}_{MAP}^D(\mathbf{r})$ can be given as

$$\mu_{MAP}(\mathbf{r}) - \bar{\mu}_{MAP}(\mathbf{r}) = \boldsymbol{\pi}(\mathbf{r})^T (\boldsymbol{\Gamma} - \bar{\boldsymbol{\Gamma}}), \quad (48)$$

$$\mu_{MAP}^D(\mathbf{r}) - \bar{\mu}_{MAP}^D(\mathbf{r}) = \boldsymbol{\Pi}(\mathbf{r})^T (\boldsymbol{\Gamma} - \bar{\boldsymbol{\Gamma}}). \quad (49)$$

Substituting (48) and (49) into (46) and (47), we have

$$\begin{aligned}& \text{Var}[\mu_{MAP}^D] \\ &= \int_{\Omega} \mathbb{E} \left[\left| [\boldsymbol{\pi}(\mathbf{r}) - \boldsymbol{\Pi}(\mathbf{r})]^T (\boldsymbol{\Gamma} - \bar{\boldsymbol{\Gamma}}) \right|^2 \right] d\mathbf{r} \\ &= \int_{\Omega} [\boldsymbol{\pi}(\mathbf{r}) - \boldsymbol{\Pi}(\mathbf{r})]^T \mathbb{E} [(\boldsymbol{\Gamma} - \bar{\boldsymbol{\Gamma}})(\boldsymbol{\Gamma} - \bar{\boldsymbol{\Gamma}})^T] \\ & \quad \cdot [\boldsymbol{\pi}(\mathbf{r}) - \boldsymbol{\Pi}(\mathbf{r})] d\mathbf{r} \\ &= \int_{\Omega} [\boldsymbol{\pi}(\mathbf{r}) - \boldsymbol{\Pi}(\mathbf{r})]^T \boldsymbol{\Sigma}_{\boldsymbol{\Gamma}} [\boldsymbol{\pi}(\mathbf{r}) - \boldsymbol{\Pi}(\mathbf{r})] d\mathbf{r}, \quad (50)\end{aligned}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\Gamma}}$ is the covariance matrix of $\boldsymbol{\Gamma}$. From the property of Gaussian random field, it can be shown that $\mathcal{A}_0\mu$ in our measurement model (10) is a multivariate Gaussian random variable statistically independent to the noise $\boldsymbol{\varepsilon}$. More specifically, using the Fubini's theorem, we can derive the mean of $\mathcal{A}_0\mu$ and the covariance between each pair of its entries:

$$\mathbb{E}[\mathcal{A}_0\mu] = \mathcal{A}_0\mu_0,$$

$$\text{Cov}[(\mathcal{A}_0\mu)_{ij}, (\mathcal{A}_0\mu)_{kl}] = \int_{\Omega} \kappa(\mathbf{r}) a_{ij}^{*,0}(\mathbf{r}) a_{kl}^0(\mathbf{r}) d\mathbf{r}.$$

Then, due to the independence between the noise and fluorophore concentration, $\boldsymbol{\Sigma}_{\boldsymbol{\Gamma}}$ can be obtained as the sum of the covariance matrices of $\mathcal{A}_0\mu$ and noise $\boldsymbol{\varepsilon}$. We use $(\boldsymbol{\Sigma}_{\boldsymbol{\Gamma}})_{p,q}$, for $p = 1, \dots, N_S N_D$, and $q = 1, \dots, N_S N_D$, to denote the entry at the p^{th} row and the q^{th} column of $\boldsymbol{\Sigma}_{\boldsymbol{\Gamma}}$, then we have

$$(\boldsymbol{\Sigma}_{\boldsymbol{\Gamma}})_{p,q} = \delta_{pq} \sigma_{\boldsymbol{\varepsilon},ij}^2 + \int_{\Omega} \kappa(\mathbf{r}) a_{ij}^{*,0}(\mathbf{r}) a_{kl}^0(\mathbf{r}) d\mathbf{r},$$

where δ_{pq} is the Kronecker delta function and the indices i, j, k, l, p and q have the following relationship:

$$p = (i-1)N_D + j, \quad q = (k-1)N_D + l,$$

for $i, k = 1, \dots, N_S$, $j, l = 1, \dots, N_D$, and $p, q = 1, \dots, N_S N_D$.

In this respect, we can express (50) as

$$\begin{aligned}& \text{Var}[\mu_{MAP}^D] \\ &= \int_{\Omega} \sum_{i,j}^{N_S, N_D} (\boldsymbol{\Sigma}_{\boldsymbol{\Gamma}})_{(i-1)N_D+j, (i-1)N_D+j} \\ & \quad \cdot |\pi_{ij}(\mathbf{r}) - \Pi_{ij}(\mathbf{r})|^2 d\mathbf{r} \\ & \quad + \int_{\Omega} \sum_{i,j}^{N_S, N_D} \sum_{k,l, kl \neq ij}^{N_S, N_D} (\boldsymbol{\Sigma}_{\boldsymbol{\Gamma}})_{(i-1)N_D+j, (k-1)N_D+l} \\ & \quad \cdot [\pi_{ij}(\mathbf{r}) - \Pi_{ij}(\mathbf{r})] [\pi_{kl}(\mathbf{r}) - \Pi_{kl}(\mathbf{r})] d\mathbf{r}.\end{aligned}$$

Applying Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}& \text{Var}[\mu_{MAP}^D] \\ & \leq \sum_{i,j}^{N_S, N_D} (\boldsymbol{\Sigma}_{\boldsymbol{\Gamma}})_{(i-1)N_D+j, (i-1)N_D+j} \|\pi_{ij}(\mathbf{r}) - \Pi_{ij}(\mathbf{r})\|_0^2 \\ & \quad + \sum_{i,j}^{N_S, N_D} \sum_{k,l, kl \neq ij}^{N_S, N_D} (\boldsymbol{\Sigma}_{\boldsymbol{\Gamma}})_{(i-1)N_D+j, (k-1)N_D+l} \\ & \quad \cdot \|\pi_{ij}(\mathbf{r}) - \Pi_{ij}(\mathbf{r})\|_0 \|\pi_{kl}(\mathbf{r}) - \Pi_{kl}(\mathbf{r})\|_0.\end{aligned}$$

Since $\boldsymbol{\Sigma}_{\boldsymbol{\Gamma}}$ is positive semi-definite, we have [41]

$$\left| (\boldsymbol{\Sigma}_{\boldsymbol{\Gamma}})_{p,q} \right| \leq \left| (\boldsymbol{\Sigma}_{\boldsymbol{\Gamma}})_{p,p} \right|^{1/2} \left| (\boldsymbol{\Sigma}_{\boldsymbol{\Gamma}})_{q,q} \right|^{1/2},$$

and

$$\begin{aligned}& \text{Var}[\mu_{MAP}^D] \\ & \leq \sum_{i,j}^{N_S, N_D} (\boldsymbol{\Sigma}_{\boldsymbol{\Gamma}})_{(i-1)N_D+j, (i-1)N_D+j} \|\pi_{ij}(\mathbf{r}) - \Pi_{ij}(\mathbf{r})\|_0^2 \\ & \quad + \sum_{i,j,k,l, ij \neq kl}^{N_S, N_D, N_S, N_D} \left| (\boldsymbol{\Sigma}_{\boldsymbol{\Gamma}})_{(i-1)N_D+j, (i-1)N_D+j} \right|^{1/2} \\ & \quad \cdot \left| (\boldsymbol{\Sigma}_{\boldsymbol{\Gamma}})_{(k-1)N_D+l, (k-1)N_D+l} \right|^{1/2} \\ & \quad \cdot \|\pi_{ij}(\mathbf{r}) - \Pi_{ij}(\mathbf{r})\|_0 \|\pi_{kl}(\mathbf{r}) - \Pi_{kl}(\mathbf{r})\|_0 \\ & \leq \left[\sum_{i,j}^{N_S, N_D} \left| (\boldsymbol{\Sigma}_{\boldsymbol{\Gamma}})_{(i-1)N_D+j, (i-1)N_D+j} \right|^{1/2} \right. \\ & \quad \left. \cdot \|\pi_{ij}(\mathbf{r}) - \Pi_{ij}(\mathbf{r})\|_0 \right]^2. \quad (51)\end{aligned}$$

Note that when the measurements are noise-free and the fluorophore concentration is also deterministic, we have $\boldsymbol{\Sigma}_{\boldsymbol{\Gamma}} = \mathbf{E}[(\boldsymbol{\Gamma} - \bar{\boldsymbol{\Gamma}})(\boldsymbol{\Gamma} - \bar{\boldsymbol{\Gamma}})^T] = \mathbf{0}$, thereby $\text{Var}[\mu_{MAP}^D] = 0$.

For each $\|\pi_{ij}(\mathbf{r}) - \Pi_{ij}(\mathbf{r})\|_0$ in (51), we let $\bar{\boldsymbol{\Gamma}} = \mathbf{e}_{ij}$ and further assume $\pi_{ij} \in H^1(\Omega)$. Following a similar approach as in Appendix A, we obtain an upper bound for $\|\pi_{ij}(\mathbf{r}) - \Pi_{ij}(\mathbf{r})\|_0$ given by

$$\begin{aligned}\|\pi_{ij}(\mathbf{r}) - \Pi_{ij}(\mathbf{r})\|_0 & \leq C \\ & \cdot \left[\sum_{i'=1}^{N_S} \sum_{n,j'}^{N_{\Delta}^i, N_D} \frac{2\|\kappa\|_{\infty} \|g_{j'}^* \phi_{i'}\|_0}{\sigma_{\boldsymbol{\varepsilon},i'j'}^2} \|g_{j'}^* \pi_{ij}\|_{0,ni'} \right. \\ & \quad \cdot \|\phi_{i'}\|_{1,ni'} h_{ni'} + \frac{\|\kappa\|_{\infty}}{\sigma_{\boldsymbol{\varepsilon},ij}^2} \sum_{n=1}^{N_{\Delta}^i} \|g_j^*\|_{\infty,ni} \|\phi_i\|_{1,ni} h_{ni} \\ & \quad + \sum_{j'=1}^{N_D} \sum_{m,i'}^{N_{\Delta}^{j'}, N_S} \frac{2\|\kappa\|_{\infty} \|g_{j'}^* \phi_{i'}\|_0}{\sigma_{\boldsymbol{\varepsilon},i'j'}^2} \|\phi_{i'} \pi_{ij}\|_{0,mj'} \\ & \quad \cdot \|g_{j'}^*\|_{1,mj'} h_{mj'} + \frac{\|\kappa\|_{\infty}}{\sigma_{\boldsymbol{\varepsilon},ij}^2} \sum_{m=1}^{N_{\Delta}^j} \|\phi_i\|_{\infty,mj} \\ & \quad \left. + \sum_{t=1}^{N_{\Delta}} \left(\sum_{i',j'}^{N_S, N_D} \frac{\|\kappa\|_{\infty} \|G_{j'}^* \Phi_{i'}\|_0 \|G_{j'}^* \Phi_{i'}\|_{0,t}}{\sigma_{\boldsymbol{\varepsilon},i'j'}^2} \right. \right. \\ & \quad \left. \left. + \|\kappa\|_{\infty} \left\| \frac{1}{\kappa} \right\|_{\infty,t} \right) \|\pi_{ij}\|_{1,t} h_t \right]. \quad (52)\end{aligned}$$

Finally, letting $D_{ij} = \left| (\Sigma_{\Gamma})_{(i-1)N_D+j, (i-1)N_D+j} \right|^{1/2}$ and substituting (52) into (51), we arrive at (31) in Theorem 2.



Lu Zhou received the M.S. degree in biomedical engineering from Tsinghua University, China, and the Ph.D degree in electrical engineering from Rensselaer Polytechnic Institute, Troy, NY, in 2006 and 2010, respectively. He currently is a software engineer at Bloomberg L.P., New York, NY.



Birsen Yazici received BS degrees in Electrical Engineering and Mathematics in 1988 from Bogazici University, Istanbul Turkey and MS and Ph.D. degrees in Mathematics and Electrical Engineering both from Purdue University, W. Lafayette IN, in 1990 and 1994, respectively. From September 1994 until 2000, she was a research engineer at the General Electric Company Global Research Center, Schenectady NY. During her tenure in industry, she worked on radar, transportation, industrial and medical imaging systems. Her work on industrial systems received the 2nd best paper award in 1997 given by IEEE Transactions in Industrial Applications. From 2001 to June 2003, she was an assistant professor at Drexel University, Electrical and Computer Engineering Department. In Fall 2003, she joined Rensselaer Polytechnic Institute where she is currently an associate professor in the Department of Electrical, Computer and Systems Engineering and in the Department of Biomedical Engineering.

Prof. Yazici's research interests span the areas of statistical signal processing, inverse problems in imaging, image reconstruction, biomedical optics, radar and X-ray imaging. She currently serves as an associate editor for the IEEE Transactions on Image Processing and SIAM Journal on Imaging Science. She is the recipient of the Rensselaer Polytechnic Institute 2007 School of Engineering Research Excellence Award. She holds 11 US patents.