Ultrafast and Ultrahigh-Resolution Diffuse Optical Tomography for Brain Imaging with Sensitivity Equation based Noniterative Sparse Optical Reconstruction (SENSOR)

Hyun Keol Kim [a,b,*], Yongyi Zhao [c], Ankit Raghuram [c], Ashok Veeraraghavan [c], Jacob Robinson [c], Andreas H. Hielscher [b]

[a] *Department of Radiology, Columbia University Irvine Medical Center, New York, NY 10032*
[b] *Department of Biomedical Engineering, New York University – Tandon School of Engineering, New York, NY 10010*
[c] *Department of Electrical and Computer Engineering, Rice University, Houston, TX 77005*

ARTICLE INFO

ABSTRACT

We introduce a novel image reconstruction method for time-resolved diffuse optical tomography (DOT) that yields submillimeter resolution in less than a second. This opens the door to high-resolution real-time DOT in imaging of the brain activity. We call this approach the sensitivity equation based noniterative sparse optical reconstruction (SENSOR) method. The high spatial resolution is achieved by implementing an asymptotic l_0-norm operator that guarantees to obtain sparsest representation of reconstructed targets. The high computational speed is achieved by employing the nontruncated sensitivity equation based noniterative inverse formulation combined with reduced sensing matrix and parallel computing. We tested the new method with numerical and experimental data. The results demonstrate that the SENSOR algorithm can achieve 1 mm^3 spatial-resolution optical tomographic imaging at depth of $\sim$60 mean free paths (MFPs) in 20$\sim$30 milliseconds on an Intel Core i9 processor.

1. Introduction

Over the last decades, functional brain imaging has become a viable tool for exploring and understanding of brain function [1,2], diagnosing of neurological diseases such as Alzheimer [3] or stroke [4], and assessing effects of rehabilitation therapeutics [5]. Traditional brain imaging modalities include electro/magneto encephalography (EEG/MEG), positron emission tomography (PET), and functional magnetic resonance imaging (fMRI). In recent years, optical neuroimaging methods have gained increasing prominence in this area due to their unique advantages over traditional methods. Optical techniques provide safe (non-ionizing), portable, low-cost, fast and high-sensitive quantitative imaging of physiological changes induced by brain activities and diseases.

Functional near infrared spectroscopy (fNIRS) [6-8] retrieves changes in the concentrations of oxyhemoglobin (HbO$_2$) and deoxyhemoglobin (Hb), and total hemoglobin (tHb = HbO$_2$ + Hb) at various depths, with a multitude of different source-detector distances. Due to its simple setup and high portability, fNIRS has been widely used for bedside imaging. This technology is particularly useful in clinical settings, where traditional imaging

modalities such as fMRI are deemed unsafe or impractical, for example, imaging of vulnerable patients (e.g., infants). In addition to fNIRS, diffuse optical tomographic (DOT) techniques [9-11] have also been employed to improve spatial resolution. Using measurements available from all source-detector pairs, DOT provides three dimensional tomographic images of oxygenation and blood volume changes in the brain. One of the latest developments of DOT brain imaging is high-density diffuse optical tomography (HD-DOT) [12-15], which employs a high-density array of sources and detectors. With a maximum distance of 15 mm between nearest neighboring sources and detectors, HD-DOT systems have provided rich data sets of measurements that can further improve spatial resolution. The HD-DOT systems have been applied for the 3D mapping of various brain functions in response to tasks or at resting state [7,12], and extensively validated on their performance against fMRI [16-17].

All optical methods rely upon appropriate models of light propagation and image reconstruction. Light propagation in the human head can be accurately modeled by the equation of radiative transfer (ERT) [18-22] or Monte-Carlo (MC) method [23]. These methods, while highly accurate, require considerable computational resources, e.g., CPU times and memory. The diffusion equation (DE) [24-26] is much faster and allows for an excellent approximation to the RTE in scattering dominant region. But it becomes inac-

* Corresponding author.
E-mail addresses: hkk2107@cumc.columbia.edu, hk3363@nyu.edu (H.K. Kim).

curate in void-like regions, i.e., cerebrospinal fluid (CSF) between the head and the brain [27-31]. The SP_N method is a high-order approximation to RTE, and produces higher accuracy than the DE while maintaining computational efficiency comparable to DE [32-34]. Image reconstruction in brain DOT has been focused on absorption changes in response to neural activities, assuming that variations in blood volume and oxygenation are small between two different states (baseline vs target). Based on this assumption, the linear perturbation approach [23-26] given by $\mathbf{y}_{\text{meas}} = \mathbf{J}\Delta\boldsymbol{\mu}$ has been used to find absorption variation $\Delta\boldsymbol{\mu}$, in which $\mathbf{y}_{\text{meas}}$ is the vector of difference measurements between two states and $\mathbf{J}$ is the Jacobian (sensitivity) matrix.

While DOT techniques have demonstrated promising results, there still remain challenges in achieving higher temporal and spatial resolution. Especially, in recent years, cellular resolution ($\sim 1 \text{ mm}^3$) deep-tissue imaging through non-invasive optical methods has attracted increasing attention. This approach allows direct imaging of neuronal electric activities via absorption activation within a volume of $\sim 1 \text{ mm}^3$. However, ultrahigh spatial and temporal resolution is currently out of reach for existing DOT techniques. Spatial resolution is negatively affected mainly by superficial, extracerebral tissues such as the skull and scalp. Temporal resolution is mostly affected by the long computation times for solving the complex inverse problem. Yet, subsecond temporal resolution is needed to monitor various neural activities that occur in milliseconds [11,35-36].

Motivated by this problem, we present here a novel image reconstruction algorithm called Sensitivity Equation based Noniterative Sparse Optical Reconstruction (SENSOR). The new method makes use of a nontruncated sensitivity function combined with a novel asymptotic l_0-norm operator and a dimensional reduction scheme. In this paper we demonstrate that spatial resolution of $\sim 1 \text{ mm}^3$ can be achieved by enforcing sparsest representation of absorption activation to conform sparse expression of actual neural activities. Furthermore, we show that temporal resolution of $20 \sim 30$ milliseconds can be reached by the software and hardware advances. The software acceleration is made possible by employing the nontruncated sensitivity-based noniterative inverse formulation combined with the dimensional reduction scheme. Further acceleration is accomplished by using a multicore processor with OpenMP for parallel computing of the solutions.

Time-domain data is chosen here because it allows to exploit useful temporal features that are sensitive to absorption changes occurring within a $\sim 1 \text{ mm}^3$ voxel. Furthermore, time-domain data can effectively filter out signal interference by other tissues surrounding the brain. In particular we use the parameterized temporal data – zero moment E, 1st moment M^1 and exponential feature $L(s)$ – available from Mellin and Laplace transforms of temporal point spread function (TPSF), as demonstrated in [37-42] for their computational efficiency, accuracy and robustness to noise. We evaluate the performance of the proposed algorithm through numerical simulations and experimental data, focusing on spatial resolution and computational speed.

The remainder of the paper is organized as follows. We first describe a general framework of the noniterative inverse formulation based on nontruncated sensitivity equation and asymptotic l_0-norm function in Section 2. Numerical and experimental results are given in Section 3, addressing the performance evaluation of the SENSOR algorithm. Finally, we draw conclusions and summarize our work in Section 4.

2. Methods

In general, an optical tomographic problem is to find a set of optical properties that leads to a close match between measurements and predictions of light intensities on the tissue surface [26-27]. The search process starts with making an initial prediction P_0 of measured light intensities z from an initial guess x_0 of unknown optical properties. Due to strong nonlinearity between optical properties x and light intensities u, this search process is carried out with an iterative updating of optical properties, i.e., $x = x_0 + \Delta x$ and $P = P_0 + \Delta P$, until the difference between predictions P and measurements z is minimized. i.e., $z - P \approx 0$. This requires repeated solving of the forward model for a new prediction P with updated optical properties x, which leads to large computation time. To mitigate this iterative process with nonlinearity, we reformulate in this paper the original nonlinear inverse problem into the linear inverse problem based on the nontruncated sensitivity equation and the adjoint theorem.

2.1. Nontruncated sensitivity equation and noniterative inverse formulation

In this section, we derive a generalized sensitivity formulation for estimating the exact variation in the measurement that results from an arbitrary change in the properties of the system, by following a generalized perturbation and adjoint theory [20,43–46]. To evaluate the exact variational formulation, the derivation takes into accounts not only the direct (first-order) effect by the change in the optical properties of the medium but also the indirect (second-order) effect by the change in the light intensity, both due to the system variation.

Consider the time dependent equation of radiative transfer (TD-ERT) that describes transient light propagation in tissue:

$$\frac{1}{c}\frac{\partial \psi(t, \mathbf{r}, \boldsymbol{\Omega})}{\partial t} = -(\nabla \cdot \Omega)\psi(t, \mathbf{r}, \boldsymbol{\Omega}) - (\mu_a + \mu_s)\psi(t, \mathbf{r}, \boldsymbol{\Omega})$$
$$+ \mu_s \int_{4\pi} p(\boldsymbol{\Omega'}, \boldsymbol{\Omega})\psi'(t, \mathbf{r}, \boldsymbol{\Omega'})d\Omega' + S(t, \mathbf{r}, \boldsymbol{\Omega}) \quad \text{(1a)}$$

with the reflective boundary condition:

$$\psi(t, \mathbf{r_b}, \boldsymbol{\Omega})|_{\vec{n}_b \cdot \boldsymbol{\Omega} < 0} = R(\boldsymbol{\Omega^i} \rightarrow \boldsymbol{\Omega})\psi(t, \mathbf{r}, \boldsymbol{\Omega^i})|_{\vec{n}_b \cdot \boldsymbol{\Omega^i} > 0}. \quad \text{(1b)}$$

where $\psi(t, \mathbf{r}, \boldsymbol{\Omega})$ is the radiance in units of $\text{Wcm}^{-2}\text{sr}^{-1}$; μ_a and μ_s are the absorption coefficient and the scattering coefficient, respectively, both in units of cm^{-1}; $p(\boldsymbol{\Omega'}, \boldsymbol{\Omega})$ is the scattering phase function that describes scattering from direction $\boldsymbol{\Omega'}$ into $\boldsymbol{\Omega}$; $S(t, \mathbf{r}, \boldsymbol{\Omega})$ is the source term inside the medium, and $c = c_0/n$ is the speed of light in the medium where c_0 is the speed of light in vacuum and n is the refractive index of the medium ($n = 1.4$ is used in this work). We employ here the Henyey-Greenstein phase function [47] commonly used in tissue optics, given by $p(\boldsymbol{\Omega'}, \boldsymbol{\Omega}) = 1/4\pi (1 - g^2)/(1 - g^2 - 2g\cos(\theta))^{3/2}$ with an anisotropic factor g and scattering cosine $\cos(\theta)$. Also, we use here the reflective boundary condition given by Eq. (1b) where $R(\boldsymbol{\Omega^i}, \boldsymbol{\Omega})$ is the reflectivity at Fresnel interface (i.e., air-tissue) from direction $\boldsymbol{\Omega^i}$ into direction $\boldsymbol{\Omega}$, and the subscript b denotes the boundary surface of the medium and $\vec{n}_b$ is the unit normal vector pointing outward the boundary surface.

Together with the boundary conditions Eq. (1b), the Mellin or Laplace transform of Eq. (1a) [37,38], can be obtained in the following discretized form as:

$$A\boldsymbol{\psi} = \mathbf{b} \quad \text{(2)}$$

where $\mathbf{A}$ is the matrix of coefficients that depend on optical properties $\mu = (\mu_a, \mu_s)$ and the geometry of the medium, and $\mathbf{b}$ is the right-hand side vector that incorporates boundary conditions and internal light sources (e.g., bioluminescent or fluorescent emission) inside the medium. For discretization, the node-centered finite-volume method is used for the spatial domain and the discrete ordinates method for the angular domain, respectively [21,22,32],. With proper boundary conditions and optical properties of the medium, the forward operator $\mathbf{A}$ of light propagation

is well-defined (i.e., square and non-singular), and thus invertible. Therefore, Eq. (2) always has a unique and exact solution given by $\boldsymbol{\psi} = \mathbf{A}^{-1}\mathbf{b}$, which provides a spatial distribution of light intensities at all points inside the medium and on the medium surface. In this work, Eq. (2) is solved with a Krylov subspace based block biconjugate gradient stabilized (BiCGSTAB) method [19]. With the intensity vector $\boldsymbol{\psi}$ available, a physical quantity of interest such as surface measurement z at a certain location can then be expressed as a linear functional of the intensity vector $\boldsymbol{\psi}$:

$$z = \mathbf{Q}\boldsymbol{\psi} \tag{3}$$

where $\mathbf{Q}$ is the measurement operator that maps $\boldsymbol{\psi}$ onto z and depends on imaging systems, i.e., contact and non-contact systems take different forms of $\mathbf{Q}$.

Noting that $\boldsymbol{\psi}$ in Eq (3) is subject to the forward solution of Eq. (2), we can write Eqs. (2) and (3) together into the constrained form by using a functional called Lagrangian as:

$$L(\boldsymbol{\psi}; \boldsymbol{\lambda}) = \mathbf{Q}\boldsymbol{\psi} + \boldsymbol{\lambda}^{\mathrm{T}}(\mathbf{A}\boldsymbol{\psi} - \mathbf{b}) \tag{4}$$

where the vector $\boldsymbol{\lambda}$ is an adjoint variable called a Lagrange multiplier that constrains the measurement $\mathbf{Q}\boldsymbol{\psi}$ to the forward model $\mathbf{A}\boldsymbol{\psi} - \mathbf{b} = 0$. Following a variational principle [44], this functional $L(\boldsymbol{\psi}; \boldsymbol{\lambda})$ is stationary about functions $\boldsymbol{\psi}$ and $\boldsymbol{\lambda}$ for which the gradient $\nabla \mathbf{L}$ is therefore required to be zero with respect to $\boldsymbol{\psi}$ and $\boldsymbol{\lambda}$ as:

$$\nabla_{\boldsymbol{\lambda}} \mathbf{L}(\boldsymbol{\psi}; \boldsymbol{\lambda}) = \mathbf{A}\boldsymbol{\psi} - \mathbf{b} = 0 \rightarrow \mathbf{A}\boldsymbol{\psi} = \mathbf{b} \tag{5a}$$

$$\nabla_{\boldsymbol{\psi}} \mathbf{L}(\boldsymbol{\psi}; \boldsymbol{\lambda}) = \mathbf{Q}^{\mathrm{T}} + \mathbf{A}^{\mathrm{T}}\boldsymbol{\lambda} = 0 \rightarrow \mathbf{A}^{\mathrm{T}}\boldsymbol{\lambda} = -\mathbf{Q}^{\mathrm{T}} \tag{5b}$$

where Eq. (5a) is equal to the forward model and Eq. (5b) is the adjoint model that will be used later for the inverse formulation. Substituting Eqs. (5a) and (5b) into Eq. (4) also gives the following expression for measurement z as

$$L(\boldsymbol{\psi}; \boldsymbol{\lambda}) = \mathbf{Q}\boldsymbol{\psi} = -\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{b} = z. \tag{6}$$

Eq. (6) indicates that measurement z can be described in terms of two equivalent mathematical forms: one in terms of forward solution $\boldsymbol{\psi}$ and another in terms of adjoint solution $\boldsymbol{\lambda}$. This, therefore, can suggest that two different approaches are possible for evaluation of the variation in the measurement with respect to the variation in the system parameters, i.e., one approach based on $z = \mathbf{Q}\boldsymbol{\psi}$ and another approach on $z = -\boldsymbol{\lambda}^{\mathrm{T}}\mathbf{b}$. However, it should be noted that the two approaches lead to the same sensitivity formulation, although the different procedures are followed, as shown in Appendix A. Therefore, we present here the procedure based on $z = \mathbf{Q}\boldsymbol{\psi}$.

To find the variational relationship between measurement z and system properties $\mathbf{A}$, $\boldsymbol{\psi}$ and $\mathbf{b}$, we consider two systems at different states in terms of system properties: one with the reference (or base) state and another with the target (or altered) state. The reference state can be defined as a state that is already known or can be reasonably assumed based on *a priori* knowledge or data. In practice, the reference state can refer to an initial guess of unknown optical properties in static imaging or baseline assumption in dynamic imaging. The target state can be defined as a state that deviates from the reference state by an unknown amount. Therefore, the target state can imply an unknown distribution of optical properties for static imaging, or an unknown next time point for dynamic imaging. Hence, the deviation between the reference state and the target state can be formally expressed as the differences in the properties of the system $\mathbf{A}\boldsymbol{\psi} = \mathbf{b}$ as:

$$\boldsymbol{\Delta A} = \mathbf{A} - \overline{\mathbf{A}}, \; \boldsymbol{\Delta \psi} = \boldsymbol{\psi} - \overline{\boldsymbol{\psi}}, \; \text{and} \; \boldsymbol{\Delta b} = \mathbf{b} - \overline{\mathbf{b}} \tag{7}$$

and, for the measurement z,

$$\Delta z = z - \bar{z} = \mathbf{Q}\boldsymbol{\psi} - \mathbf{Q}\overline{\boldsymbol{\psi}} = \mathbf{Q}\boldsymbol{\Delta\psi} \tag{8}$$

where the properties $\overline{\mathbf{A}}$, $\overline{\boldsymbol{\psi}}$, $\overline{\mathbf{b}}$ and $\bar{z}$ represent the reference state and the properties $\mathbf{A}$, $\boldsymbol{\psi}$, $\mathbf{b}$ and z indicate the unknown target state.

Therefore, with Eq. (7) into $\mathbf{A}\boldsymbol{\psi} = \mathbf{b}$, the target state system can be rewritten in terms of the reference state and the deviation amount as:

$$(\overline{\mathbf{A}} + \boldsymbol{\Delta A})(\overline{\boldsymbol{\psi}} + \boldsymbol{\Delta \psi}) = \overline{\mathbf{b}} + \boldsymbol{\Delta b} \tag{9}$$

which, using the reference state $\overline{\mathbf{A}}\overline{\boldsymbol{\psi}} = \overline{\mathbf{b}}$, also reduces to:

$$\overline{\mathbf{A}}\boldsymbol{\Delta \psi} = \boldsymbol{\Delta b} - \boldsymbol{\Delta A}\overline{\boldsymbol{\psi}} - \boldsymbol{\Delta A}\boldsymbol{\Delta \psi}. \tag{10}$$

Eq. (10) is often referred to as the exact perturbed system, which represents the exact variational relationship between the reference state and the target state when optical properties are varied from $\overline{\boldsymbol{\mu}}$ to $\overline{\boldsymbol{\mu}} + \boldsymbol{\Delta \mu}$, the intensity from $\overline{\boldsymbol{\psi}}$ to $\overline{\boldsymbol{\psi}} + \boldsymbol{\Delta \psi}$ and the right-hand side from $\overline{\mathbf{b}}$ to $\overline{\mathbf{b}} + \boldsymbol{\Delta b}$. Here the variation $\boldsymbol{\Delta \mu}$ can be a spectral absorption change $\boldsymbol{\Delta \mu_a}$, scattering change $\boldsymbol{\Delta \mu_s}$ or wavelength-independent chromophore concentration change $\boldsymbol{\Delta c}$ based on the relation $\boldsymbol{\Delta \mu_a} = \boldsymbol{\epsilon}_{\mathrm{M}}^{\mathrm{T}}\boldsymbol{\Delta c}$ where $\boldsymbol{\varepsilon}_{\mathrm{M}}$ denotes spectral molecular extinction coefficients.

The intensity variation $\boldsymbol{\Delta \psi}$ given by Eq. (10) can now be related to the measurement variation Δz given by Eq. (8) through the adjoint model (Eq. (5b)). Using the reference state adjoint model $\overline{\mathbf{A}}^{\mathrm{T}}\overline{\boldsymbol{\lambda}} = -\mathbf{Q}^{\mathrm{T}}$, Eq. (8) can be rewritten as:

$$\Delta z = \mathbf{Q}\boldsymbol{\Delta \psi} = -\overline{\boldsymbol{\lambda}}^{\mathrm{T}}\overline{\mathbf{A}}\boldsymbol{\Delta \psi}. \tag{11}$$

Taking the inner product of Eq. (10) with $\overline{\boldsymbol{\lambda}}$ and subtracting Eq. (11) gives the generalized exact sensitivity relationship between the measurement difference and the variations in the system properties as:

$$\Delta z = -\overline{\boldsymbol{\lambda}}^{\mathrm{T}}(\boldsymbol{\Delta b} - \boldsymbol{\Delta A}\overline{\boldsymbol{\psi}}) + \overline{\boldsymbol{\lambda}}^{\mathrm{T}}\boldsymbol{\Delta A}\boldsymbol{\Delta \psi}. \tag{12}$$

Depending on the inverse problems to be solved, Eq. (12) can be recast into two different sensitivity formulations as:

$$\Delta z = \overline{\boldsymbol{\lambda}}^{\mathrm{T}}(\boldsymbol{\Delta A}\overline{\boldsymbol{\psi}} + \boldsymbol{\Delta A}\boldsymbol{\Delta \psi}) = \overline{\boldsymbol{\lambda}}^{\mathrm{T}}\boldsymbol{\Delta A}(\overline{\boldsymbol{\psi}} + \boldsymbol{\Delta \psi}) \tag{13a}$$

and

$$\Delta z = -\overline{\boldsymbol{\lambda}}^{\mathrm{T}}(\boldsymbol{\Delta b}). \tag{13b}$$

The first formulation given by Eq. (13a) is obtained, assuming that the right-hand side $\mathbf{b}$ is not varied between the reference state and the new state, i.e., $\boldsymbol{\Delta b} = 0$. Therefore Eq. (13a) describes the sensitivity of intensity variation $\boldsymbol{\Delta \psi}$ onto optical parameter variation $\boldsymbol{\Delta \mu}$ alone, and thus can be used for standard DOT problems of reconstructing $\boldsymbol{\Delta \mu}$. On the other hand, the second formulation given by Eq. (13b) is obtained under the assumption that $\overline{\mathbf{A}} + \boldsymbol{\Delta A} \approx \overline{\mathbf{A}}$ with $\boldsymbol{\Delta A} \approx 0$ between the two states. This assumption is often widely made in inverse source problems such as bioluminescence and fluorescence tomographic problems, i.e., the background optical properties can be assumed to be the same both for excitation and emission states. As a result, Eq. (13b) reveals the linear relationship between source-term variation $\boldsymbol{\Delta b}$ and intensity variation $\boldsymbol{\Delta \psi}$. Thus, because Eq. (13b) is already formulated into the linear inverse system, we focus here on Eq. (13a) to derive the linear inverse formulation of standard DOT problems. It should also be noted that the whole expression Eq. (12) should be used when source term also changes with absorption and scattering coefficients in the medium.

Since the coefficient matrix $\mathbf{A}$ is linear with respect to optical properties $\boldsymbol{\mu}$, the difference matrix $\boldsymbol{\Delta A}$ can be written as $\boldsymbol{\Delta A} = (\partial \mathbf{A}/\partial \boldsymbol{\mu})\boldsymbol{\Delta \mu}$. For the diffusion equation (DE), this linearity is not valid in general, but can be used for the cases where the

diffusion coefficient $D = 1/3(\mu_a + \mu_s')$ can be reasonably approximated as $D \approx 1/3\mu_s'$. Then, Eq. (13a) can be expressed as

$$\Delta z = \left(\overline{\boldsymbol{\lambda}}^{\mathrm{T}} \frac{\partial \mathbf{A}}{\partial \boldsymbol{\mu}} \overline{\boldsymbol{\psi}}\right) \Delta \boldsymbol{\mu} + \left(\overline{\boldsymbol{\lambda}}^{\mathrm{T}} \frac{\partial \mathbf{A}}{\partial \boldsymbol{\mu}} \Delta \boldsymbol{\psi}\right) \Delta \boldsymbol{\mu}$$

$$= \overline{\boldsymbol{\lambda}}^{\mathrm{T}} \frac{\partial \mathbf{A}}{\partial \boldsymbol{\mu}} \left(\overline{\boldsymbol{\psi}} + \Delta \boldsymbol{\psi}\right) \Delta \boldsymbol{\mu} = \mathbf{S_{nt}} \Delta \boldsymbol{\mu} \tag{14a}$$

or, equivalently,

$$\ln z - \ln \overline{z} = \frac{1}{\overline{z}}\left(\overline{\boldsymbol{\lambda}}^{\mathrm{T}} \frac{\partial \mathbf{A}}{\partial \boldsymbol{\mu}} \overline{\boldsymbol{\psi}}\right) \Delta \boldsymbol{\mu} + \frac{1}{\overline{z}}\left(\overline{\boldsymbol{\lambda}}^{\mathrm{T}} \frac{\partial \mathbf{A}}{\partial \boldsymbol{\mu}} \Delta \boldsymbol{\psi}\right) \Delta \boldsymbol{\mu}$$

$$= \frac{1}{\overline{z}} \overline{\boldsymbol{\lambda}}^{\mathrm{T}} \frac{\partial \mathbf{A}}{\partial \boldsymbol{\mu}} \left(\overline{\boldsymbol{\psi}} + \Delta \boldsymbol{\psi}\right) \Delta \boldsymbol{\mu} = \frac{1}{\overline{z}} \mathbf{S_{nt}} \Delta \boldsymbol{\mu} \tag{14b}$$

when the measurement variation is expressed in the exponential form as $z = \overline{z}e^{\Delta z/\overline{z}}$ instead of $z = \overline{z} + \Delta z$. Here $\mathbf{S_{nt}}$ is the nontruncated sensing matrix that comprises both the first-order sensitivity matrix (first term) and the correction matrix (second term). When the variation $\Delta \boldsymbol{\mu}$ of optical properties is sufficiently small, i.e., a differential change, then the intensity variation $\Delta \boldsymbol{\psi}$ can also be assumed to be sufficiently small. In this case, by neglecting the second order term $\Delta \mathbf{A} \Delta \boldsymbol{\psi}$ ($\ll \Delta \mathbf{A}$ and $\Delta \boldsymbol{\psi}$) in Eq. (10) and also $\Delta \boldsymbol{\mu} \Delta \boldsymbol{\psi}$ ($\ll \Delta \boldsymbol{\mu}$ and $\Delta \boldsymbol{\psi}$) in Eq. (14), we can obtain the following first-order sensitivity (i.e., linearized perturbation) form of Eq. (10) as:

$$\overline{\mathbf{A}} \Delta \boldsymbol{\psi} + \Delta \mathbf{A} \overline{\boldsymbol{\psi}} = 0 \tag{15}$$

and Eq. (14) also reduces to

$$\Delta z = \left(\overline{\boldsymbol{\lambda}}^{\mathrm{T}} \frac{\partial \mathbf{A}}{\partial \boldsymbol{\mu}} \overline{\boldsymbol{\psi}}\right) \Delta \boldsymbol{\mu} = \left[\frac{\partial z}{\partial \boldsymbol{\mu}}\right] \Delta \boldsymbol{\mu} \tag{16}$$

which forms a basis for the linearized perturbation inverse model based on the first-order sensitivity (Jacobian) coefficient given by $\partial z/\partial \boldsymbol{\mu}$ [20,24–26],. However, in this work we retain the cross-product correction term $\Delta \mathbf{A} \Delta \boldsymbol{\psi}$ in Eqs. (10) and (14) to not compromise accuracy. It should be noted from Eq. (14) that the nontruncated sensing matrix $\mathbf{S_{nt}}$ can be obtained once the flux difference $\Delta \boldsymbol{\psi}$ is determined. In the following, we present the method to estimate this flux difference $\Delta \boldsymbol{\psi}$ by deriving the projection matrix that maps the measurement change Δz onto the flux change $\Delta \boldsymbol{\psi}$.

Using the relation $\overline{\boldsymbol{\psi}} + \Delta \boldsymbol{\psi} = \boldsymbol{\psi}$, Eq. (13a) can be written as the following general form:

$$\Delta z = \overline{\boldsymbol{\lambda}}^{\mathrm{T}} \Delta \mathbf{A} \boldsymbol{\psi}, \tag{17a}$$

and, given M measurements at locations $r_i (i = 1, ..., M)$, Eq. (17a) gives the following linear system as:

$$\begin{bmatrix} \Delta z_1 \\ \vdots \\ \Delta z_M \end{bmatrix} = \begin{bmatrix} \overline{\lambda}_{1,1} & \cdots & \overline{\lambda}_{1,N} \\ \vdots & \ddots & \vdots \\ \overline{\lambda}_{M,1} & \cdots & \overline{\lambda}_{M,N} \end{bmatrix} \begin{bmatrix} (\Delta A \psi)_1 \\ \vdots \\ (\Delta A \psi)_N \end{bmatrix}. \tag{17b}$$

For N_s source illuminations at locations $r_j (j = 1, ..., N_s)$ particularly with the fixed number of detectors, Eq. (17b) can be expanded into the matrix-matrix linear system with multiple right-hand sides as:

$$\boldsymbol{\Phi} \mathbf{X} = \Delta \mathbf{Z} \tag{18a}$$

with

$$\boldsymbol{\Phi} = \begin{bmatrix} \overline{\lambda}_{1,1} & \cdots & \overline{\lambda}_{1,N} \\ \vdots & \ddots & \vdots \\ \overline{\lambda}_{M,1} & \cdots & \overline{\lambda}_{M,N} \end{bmatrix},$$
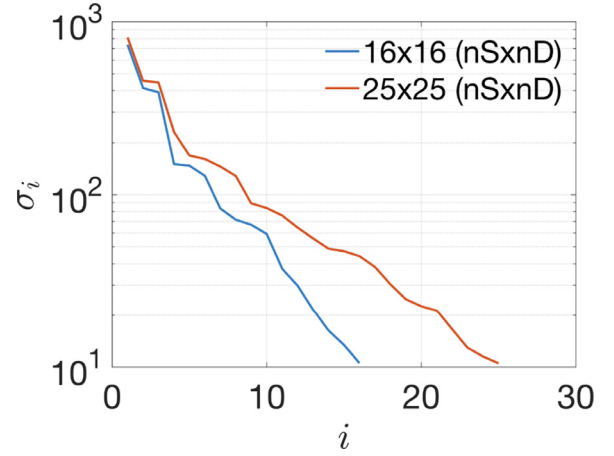

Fig. 1. Singular value (σ_i) spectrum of the adjoint matrix $\boldsymbol{\Phi}$ truncated at a small positive value ($\sim 10^{-6}$) for measurement geometries with 16×16 and 25×25 source-detector pairs.

$$\mathbf{X} = \begin{bmatrix} (\Delta A \psi)_1^1 & \cdots & (\Delta A \psi)_1^{N_s} \\ \vdots & \ddots & \vdots \\ (\Delta A \psi)_N^1 & \cdots & (\Delta A \psi)_N^{N_s} \end{bmatrix},$$

$$\Delta \mathbf{Z} = \begin{bmatrix} \Delta z_1^1 & \cdots & \Delta z_1^{N_s} \\ \vdots & \ddots & \vdots \\ \Delta z_M^1 & \cdots & \Delta z_M^{N_s} \end{bmatrix}. \tag{18b}$$

where $\boldsymbol{\Phi} \in R^{M \times N}$ is the adjoint sensing matrix, $\mathbf{X} \in R^{N \times N_s}$ is the system variation matrix, and $\Delta \mathbf{Z} \in R^{M \times N_s}$ is the measurement variation matrix. Here the measurement $\mathbf{Z}$ is known for the target state and the reference prediction $\overline{\mathbf{Z}}$, and the adjoint matrix $\boldsymbol{\Phi}$ can also be precalculated for the reference state, so Eq. (18a) can be solved for $\mathbf{X}$. But note that the adjoint matrix $\boldsymbol{\Phi}$ typically has full row rank ($Rank(\boldsymbol{\Phi}) < N$, i.e., rank deficient) as shown in Fig. 1 for its non-zero singular value spectrum. Therefore, we solve the linear system given by Eq. (17b) or Eq. (18) by using the rank deficient Moore-Penrose least squares method, which leads to the solution $\mathbf{X} = \boldsymbol{\Phi}^{\mathrm{T}}(\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{T}})^{-1}\Delta \mathbf{Z}$ with sparsity constraints.

Once the system variation matrix $\mathbf{X} = [\Delta \mathbf{A}\boldsymbol{\psi}_1, \Delta \mathbf{A}\boldsymbol{\psi}_2 ..., \Delta \mathbf{A}\boldsymbol{\psi}_{N_s}]$ is obtained for all N_s sources, the intensity variation vector $\Delta \boldsymbol{\psi}$ can be easily calculated for all N_s sources by using Eq. (10) as:

$$\Delta \boldsymbol{\Psi} = -\left(\overline{\mathbf{A}}\right)^{-1} \mathbf{X} = -\left(\overline{\mathbf{A}}\right)^{-1} \boldsymbol{\Phi}^{\mathrm{T}}\left(\boldsymbol{\Phi}\boldsymbol{\Phi}^{\mathrm{T}}\right)^{-1} \Delta \mathbf{Z} = \mathbf{H} \Delta \mathbf{Z} \tag{19}$$

where the matrix $\Delta \boldsymbol{\Psi} = [\Delta \boldsymbol{\psi}_1, \Delta \boldsymbol{\psi}_2, ..., \Delta \boldsymbol{\psi}_{N_s}] \in R^{N \times N_s}$ and $\mathbf{H} \in R^{N \times M}$ is the projection operator that transforms measurement variation $\Delta \mathbf{Z}$ back into flux variation $\Delta \boldsymbol{\Psi}$. Eq. (19) is now used into Eq. (14) to evaluate the nontruncated sensing matrix $\mathbf{S_{nt}}$. Since $\mathbf{A}$ and $\boldsymbol{\Phi}$ are defined at the reference state, $\mathbf{S_{nt}}$ can also be precalculated before reconstruction is started. Therefore, the unknown optical property change $\Delta \boldsymbol{\mu}$ can be found by solving the following linear least squares problem of Eq. (14) as

$$\Delta \hat{\boldsymbol{\mu}} = \underset{\Delta \boldsymbol{\mu}}{\operatorname{argmin}} \| \Delta \mathbf{Z} - \mathbf{S_{nt}} \Delta \boldsymbol{\mu} \|_2^2 \tag{20}$$

where $\Delta \mathbf{Z} \in R^{(N_s M \times 1)}$, $\mathbf{S_{nt}} \in R^{(N_s M \times N)}$ and $\Delta \boldsymbol{\mu} \in R^{(N \times 1)}$.

Thus, given the target measurement z, the reference prediction $\overline{z}$ and the reference system matrix $\overline{\mathbf{A}}(\overline{\mu})$, the unknown optical property variation $\Delta \boldsymbol{\mu}$ can be found in a noniterative manner through Eqs. (15)-(20). The term 'noniterative' is used here to emphasize that the SENSOR algorithm does not require repeated solving of the forward model while iteratively updating $\Delta \boldsymbol{\mu}$. This feature of the SENSOR method is distinct from traditional nonlinear
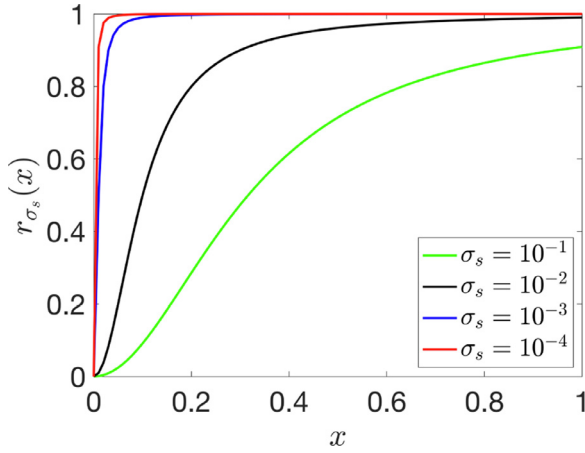
Fig. 2. The sparse function values $r_{\sigma_s}(x)$ with different sparsity parameter σ_s.

minimization algorithms that solve the forward problem repeatedly with an iterative update of $\Delta\mu$, which often leads to large CPU time in the total reconstruction time. Also, it should be noted that the approach has no restriction, in principle, placed on the variation amount ΔA for its validity, while the linear perturbation approximation requires the variation $\Delta\mu$ to be sufficiently small. Also it should be noted that the formulations presented here can be directly applied to a multispectral tomographic model [22] as given in Appendix B.

2.2. Asymptotic l_0 norm minimization and sparsity enhancement

The sparsest representation of the solution $\Delta\mu$ to Eq. (20) can be achieved with the following l_0-norm minimization as:

$$\underset{\Delta\mu}{argmin}\ \|\Delta\mu\|_0 \ \text{subject to}\ \|\Delta Z - S_{\mathrm{nt}}\Delta\mu\|_2^2 \qquad (21)$$

where $\|\Delta\mu\|_0$ is the l_0 norm that represents the number of non-zero elements in $\Delta\mu$. Mathematically, the optimization problem (21) subject to the l_0 norm is a non-convex problem, which means it is very difficult to find a l_0 constrained global minimum and thus the solution can easily fall into one of local minima. Another difficulty with l_0 norm comes from its non-differentiability. Current methods for solving such non-convex problems as Eq. (21) have been focused mainly on so-called greedy algorithms of matching pursuit (MP) type [47-51] and convex relaxation algorithms [52-56] based on a l_1 norm and linear programming. These two kinds are useful for finding sparse solution but have disadvantages of requiring *a priori* information about the sparsity, relatively high sensitivity to measurement noise, or limited performance in the asymmetric measurement systems with non-fixed source-detector configuration (i.e., varying numbers of sources and detectors) [57].

To address this problem, we propose here a new approach that achieves a l_0 norm asymptotically from a smooth l_2 norm, similarly to the work [58-59] as done for a sparse signal recovery in the field of signal processing. Therefore, the approach is started with a globally convergent convex method (l_2) and gradually converted to a non-convex method that achieves the sparsest solution. To this end, the asymptotic sparse function is designed here that has the following form:

$$r_{\sigma_s}(x_i) = \frac{x_i^2}{x_i^2 + \sigma_s} \qquad (22)$$

where x_i is the i-th element of the sparse solution vector $\Delta\mu$ and σ is the tuning parameter that adjusts a level of sparsity in the solution. Fig. 2 shows the values of the sparse function $r_\sigma(x_i)$ as a function of x_i. We observe that as the value of σ goes close to zero,

$r_{\sigma_s}(x_i) \approx 1$ when $x_i^2 >> \sigma$ and $r_{\sigma_s}(x_i) \approx 0$ when $x_i^2 \ll \sigma$, whereas as σ increases $r_{\sigma_s}(x_i)$ resembles a smooth l_2-norm function. Therefore, with the asymptotic function $r_{\sigma_s}(x_i)$, the l_0-norm of $\Delta\mu$ can be approximated as

$$\|\Delta\mu\|_0 \approx R = \sum_i r_\sigma(x_i) = \sum_i \left(\frac{x_i^2}{x_i^2 + \sigma_s}\right). \qquad (23)$$

2.3. Minimization with SENSOR algorithm

With the asymptotic l_0-norm function given by Eq. (22), the original optimization problem Eq. (21) is now rewritten into the following constrained form:

$$\Delta\hat{\mu} = \underset{\Delta\mu}{argmin}\left\{ f(\Delta\mu) = \|\Delta Z - S_{\mathrm{nt}}\Delta\mu\|_2^2 + \beta\sum_i r_\sigma(\Delta\mu_i)\right\}$$
$$(24)$$

where the first term is the data fit error and the second term is the l_0 norm constrained to enforce the sparsest solution and β is the regularization parameter that balances the efforts of minimizing the data fit error and the l_0 term at once. Therefore, the sparsest solution to Eq. (24) can be found at the point that minimizes the mismatch $\|\Delta Z - S\Delta\mu\|_2^2$ and the l_0 norm simultaneously.

The minimization of Eq. (24) starts with a large value of σ_s, and gradually decreases it to a smaller value to approximate the l_0 norm asymptotically. The large value of σ_s is required in the beginning of this minimization process because it ensures that an initial search $\Delta\mu^0$ is already in the vicinity (i.e., local convex) of global minimum and thus the risk of falling into a local minimum can be avoided. The inverse solution to Eq. (24) with a very large value of $\sigma_s \to \infty$ is given by $\Delta\mu^0 = S^T(SS^T)^{-1}\Delta Z$, which is often referred to as the pseudo inverse solution and can be efficiently evaluated using the truncated singular value decomposition (SVD) method. From this initial smooth solution $\Delta\mu^0$, we then find a nonzero maximum value $|\Delta\mu^0|$ and set it to the initial value σ_s^0 for the sparsity tuning parameter. This initial value σ^0 is then adjusted to decrease monotonically by a predefined factor as $\sigma_s^k = t\sigma_s$ with $0.5 \le t < 1$ at k-th iteration.

Starting with $\Delta\mu^0$, we solve the minimization problem Eq. (24) with the projected gradient descent method [60]. The gradient calculation of $f(\Delta\mu)$ is straightforward and given as

$$g = \nabla f(\Delta\mu) = -S^T(\Delta Z - S\Delta\mu) + \beta\ \nabla R(\Delta\mu) \qquad (25)$$

where $\nabla R(X)$ is the gradient vector of the l_0 norm Eq. (23) given as

$$\nabla R(\Delta\mu) = \left[\frac{2\sigma_s x_1 x_1^2}{(x_1^2 + \sigma_s)^2} \quad \cdots \quad \frac{2\sigma_s x_m x_m^2}{(x_m^2 + \sigma_s)^2}\right]^T. \qquad (26)$$

With the gradient g available, the iterate is generated as

$$\Delta\mu^k \leftarrow \Delta\mu^k - \rho^k g^k, \qquad (27)$$

and $\Delta\mu$ in Eq. (27) is further updated into the feasible solution set through orthogonal projection as

$$\Delta\mu^{k+1} \leftarrow \Delta\mu^k + \hat{S}\left(\Delta Z - S\Delta\mu^k\right) \qquad (28)$$

where $\hat{S}$ is the orthogonal projection matrix defined as

$$\hat{S} = S^T\left(SS^T\right)^{-1}\ \text{or}\ \left(S^T S\right)^{-1}S^T \qquad (29)$$

depending on the rank of the matrix S. The regularization parameter β is started with $\beta_0\|S^T\Delta Z\|$ and adjusted iteratively according to the data fit error $\|S^T(\Delta Z - S\Delta\mu)\|$ as $\beta = \beta_0\|S^T(\Delta Z - S\Delta\mu)\|$. Based on all the descriptions so far, the detailed steps of the SENSOR algorithm are given in the following table (see Table 1).

Table 1
Sensitivity Equation based Noniterative Sparse Optical Reconstruction (SENSOR) algorithm.

Initialize $\Delta \mathbf{Z}, \mathbf{H}, \mathbf{S}, \sigma_s^T, \alpha, \beta_0, t, KMAX, IMAX, err, \epsilon, tol$
Find initial support region $supp = \{j \mid Cov = \Delta \mathbf{Z}^t S_{i,j} > tol\}$
Obtain reduced $\mathbf{S} \equiv \mathbf{S}_{supp} = \{S^j \mid j \in supp(j)\}$ where S^j is the j-th column vector of $\mathbf{S}$
Set $\Delta \boldsymbol{\mu}^0 = \mathbf{S}^\mathrm{T}(\mathbf{SS}^\mathrm{T})^{-1} \Delta \mathbf{Z}, \sigma_s^1 = \max(|\Delta \boldsymbol{\mu}^0|), \beta = \beta_0 \|\mathbf{S}^\mathrm{T} \Delta \mathbf{Z}\|_\infty$
Compute sparsity-level sequence $\{\sigma_s^1, t\sigma_s^1, \ldots, t^{KMAX-1}\sigma_s^1\}$ with $\sigma_s^k = \max(t^{k-1}\sigma_s^1, \sigma_s^T)$
FOR $k = 1, KMAX$
    Set $\sigma_s = \sigma_s^k$ and $i = 1$
    WHILE ($err > \varepsilon$ AND $i < IMAX$)
        Update $\Delta \boldsymbol{\mu} \leftarrow \Delta \boldsymbol{\mu} + \rho^k \mathbf{d}^k$ with $\mathbf{d} = -\mathbf{g}$ in Eqs. (23)-(27)
        Update $\Delta \boldsymbol{\mu} \leftarrow \Delta \boldsymbol{\mu} + \hat{\mathbf{S}}(\Delta \mathbf{Z} - \mathbf{S}\Delta \boldsymbol{\mu}^k)$ with $\hat{\mathbf{S}}$ in Eqs. (28)-(29)
        Compute residual $err = \|\rho^k \mathbf{d}^k\|$
        Set $i = i + 1$
    END
    Update support region $supp = \{j \mid Cov = (\Delta \mathbf{Z} - \mathbf{S}\Delta \boldsymbol{\mu})^\mathrm{T} S_{i,j} > tol\}$
    Update reduced $S_{supp}^k = \{S^j \mid j \in supp(j)\}$ where S^j is the j-th column vector of S
    Set $\Delta \boldsymbol{\mu}^k = \Delta \boldsymbol{\mu}^i$ and $k = k + 1$
END
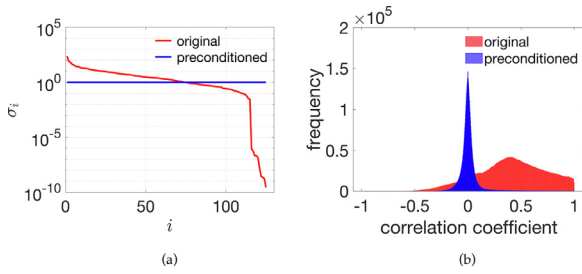Output $\Delta \boldsymbol{\mu}^{optimal} = \Delta \boldsymbol{\mu}^k$

Fig. 3. Properties of the sensing matrix **S** with and without preconditioning with 25×25 source-detector measurements: (a) singular value spectrum; (b) correlation distribution between column vectors.

2.4. Preconditioning and correlation-based dimensional reduction

In the linear inverse formulation $\Delta \mathbf{Z} = \mathbf{S}\Delta \boldsymbol{\mu}$ as described earlier, the sensing matrix **S** is ill-conditioned because, in a typical setting of optical imaging, the system is under-determined ($M \ll N$) and the singular values of **S** cluster around zero, i.e., the columns of the sensing matrix are highly correlated. Therefore, we precondition the sensing matrix **S** towards making the correlations between the columns smaller (i.e., making the column vectors linearly more independent). To this end, we compute in this work the preconditioner by minimizing the Frobenius norm as $\mathbf{P} = \arg\min\|\mathbf{PS} - \mathbf{I}\|_F$, which leads to the solution that makes the **PS** close to an identity matrix. For full row rank (e.g., under-determined) system, the preconditioner **P** is given as $\mathbf{P} = \mathbf{S}^\mathrm{T}(\mathbf{SS}^\mathrm{T})^{-1}$, also known as the Moore-Penrose least squares solution matrix, while for full column rank (e.g., over-determined) system, **P** is given as $\mathbf{P} = (\mathbf{S}^\mathrm{T}\mathbf{S})^{-1}\mathbf{S}^\mathrm{T}$. In practice, this pseudo-inverse preconditioner matrix **P** can be easily evaluated by using the single value decomposition (SVD) method. By multiplying the preconditioner $\mathbf{P} = \mathbf{S}^\mathrm{T}(\mathbf{SS}^\mathrm{T})^{-1}$ on both sides of $\Delta \mathbf{Z} = \mathbf{S}\Delta \boldsymbol{\mu}$, we then obtain the preconditioned sensing matrix $\mathbf{S}^*$ and the preconditioned measurement matrix $\mathbf{Z}^*$ as

$$\mathbf{S}^* = \mathbf{PS}, \text{ and } \Delta \mathbf{Z}^* = \mathbf{P}\Delta \mathbf{Z}. \tag{30}$$

Fig. 3 shows the singular value spectrum and the distribution of correlation values between column vectors of the original sensing matrix **S** and the preconditioned sensing matrix $\mathbf{S}^*$ with the 25×25 source-detector measurements. The results clearly demonstrate that the preconditioned system has uniform singular value distribution at 1 and clustered correlations around zero (i.e., highly independent column vectors in $\mathbf{S}^*$, while for the original non-preconditioned matrix **S** the correlation values are higher

and spread over the larger range and also many singular values are close to zero (i.e. highly correlated column vectors in **S**). Thus, the preconditioner $\mathbf{P} = \mathbf{S}^\mathrm{T}(\mathbf{SS}^\mathrm{T})^{-1}$ can stabilize the inversion process by making the sensing matrix better conditioned.

Furthermore, we employ here an efficient dimensional reduction scheme that reduces the sensing matrix iteratively by making use of the correlation information between measurements and columns of the sensing matrix. The j-th column S^j of the sensing matrix can be viewed as a vector of weights that projects the optical property change $\Delta \mu_j$ of the j-th element (or voxel) onto the measurement space $\Delta \mathbf{Z}$, as

$$\Delta \mathbf{Z} = \sum_j \mathbf{S^j}\Delta \mu_j \tag{31}$$

$$\Delta \mathbf{Z} = \mathbf{S}^1\Delta \mu_1 + \mathbf{S}^2\Delta \mu_2 + \ldots \mathbf{S^j}\Delta \mu_j + \cdots + \mathbf{S^N}\Delta \mu_N. \tag{32}$$

Therefore, the correlation $Cor(\Delta \mathbf{Z}, \mathbf{S})$ between $\Delta \mathbf{Z}$ and columns **S**, which is given by $Cor(\Delta \mathbf{Z}, \mathbf{S}) = \Delta \mathbf{Z}^\mathrm{T}\mathbf{S}$, can be used to determine which elements of $\Delta \boldsymbol{\mu}$ have high correlation with $\Delta \mathbf{Z}$ and thus should be included into reconstruction and which elements of $\Delta \boldsymbol{\mu}$ have low correlation with $\Delta \mathbf{Z}$ and can be excluded from reconstruction. To maximize computational efficiency, this reduction can be applied recursively throughout the optimization process, at each step of iterations, in order to keep adjusting the sensing matrix into the reduced one that consists of high-correlation columns alone. As a result, this can lead to a great reduction in the sensing matrix and make sparse reconstruction computationally very efficient.

2.5. Parallel implementation

Solving large linear systems is a performance bottleneck both in the forward and inverse computations in terms of computation time because it comprises a large number of matrix-matrix and matrix-vector products and operations. Therefore, parallel solving of these linear systems is the most effective way to improve computational efficiency. In this work, we focus on parallel solving of the linear systems resulting from multiple source illuminations given by $\Delta \mathbf{Z} = \mathbf{S}\Delta \boldsymbol{\mu}$ (Eqs. (17)-(18)). Parallel implementation of solving $\Delta \mathbf{Z} = \mathbf{S}\Delta \boldsymbol{\mu}$ is done on a framework of OpenMP [61] with multicore CPUs. To reduce communication overhead between K processors taking part in parallel computing, we use static scheduling so that K tasks are optimally distributed over K available processors. In addition, synchronization is further optimized so that threads do not wait until other threads are finished, leaving none of K processors waiting or idling during computation.
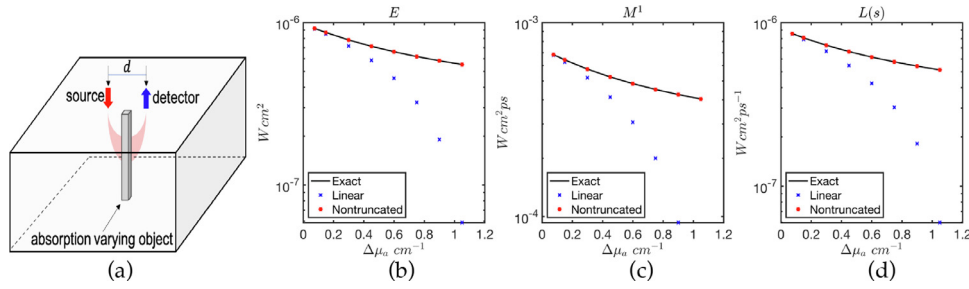
Fig. 4. Comparison in accuracy between the linearized approximation and the nontruncated sensitivity formulation: (a) problem setup; (b)-(d) calculated E, M^1, and $L(s)$, respectively. Note that $s = 0.0001$ is used in $L(s)$.

3. Results and Discussions

In this section, we first show the validity and accuracy of the nontruncated sensitivity equation Eq. (10), which is crucial to the noniterative inverse formulation of our SENSOR algorithm. Then we discuss in details the performance of the SENSOR algorithm using both numerical and experimental data. In both cases, we focus on the reconstruction of absorption coefficients in the medium. However, it should be noted that the method presented here can be used without loss of generality across all domains of DOT applications, including the simultaneous reconstruction of both absorption and scattering. The code performance is evaluated in terms of CPU time and accuracy with an emphasis on spatial and temporal resolution. To illustrate the performance metrics, we compare the new algorithm with our existing PDE-constrained reduced-Hessian sequential quadratic programming (rSQP) method [21-22]. This approach has been known to be the fastest iterative method of its kind to date since it solves the forward and inverse problems simultaneously within a framework of PDE-constrained optimization [21-22].

3.1. Validity and accuracy of nontruncated sensitivity formulation

To validate the accuracy of nontruncated sensitivity formulation, we provide here a comparison between the linearized sensitivity formulation (Eq. (15)) and the nontruncated sensitivity equation (Eq. (10)) on their computed values of three parameterized time data E, M^1, $L(s)$, with different variations $\Delta\mu_a$ in absorption coefficient. For this comparison, the 4 cm $\times$ 4 cm $\times$ 2 cm block phantom is used with the background absorption and reduced scattering coefficients given by of $\mu_a^b = 0.15$ cm^{-1} and $\mu_s' = 9$ cm^{-1}, and with a 0.4 cm $\times$ 0.4 cm $\times$ 2 cm cylindric inclusion of absorption inhomogeneity $\mu_a = \mu_a^b + \Delta\mu_a$ embedded within the medium (Fig. 4a). The amount $\Delta\mu_a$ is varied from $\Delta\mu_a = 0.075$ cm^{-1} to 1.05 cm^{-1}, corresponding to 50% to 700% of μ_a^b respectively.

The computations of three parameterized time data are done with varying single-detector distance Fig. 4a). The accuracy of the two formulations Eqs. (10) and (15) is compared against the exact solution obtained with solving the original equation (Eq. (2)) with given optical properties. Fig. 4 shows the results with source-detector separation $d = 3$ cm, in which the parameterized data are computed using the two formulations as a function of absorption variation $\Delta\mu_a$. As shown in Fig. 4b-4d, the nontruncated formulation gives a perfect agreement with the exact solution Eq. (2) for all variations, while the linearized formulation deteriorates rapidly with increasing variation, with error up to 100%. Based on extensive comparisons on accuracy for other cases, although not presented here, we found that the nontruncated sensitivity formulation provides accurate predictions regardless of absorption variations and thus can be used for quantitative imaging in all
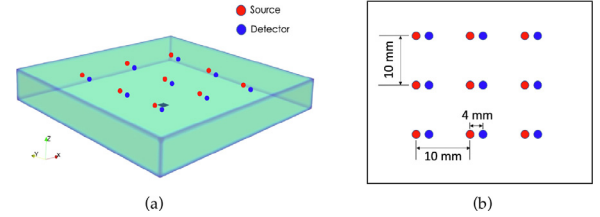

Fig. 5. 3D numerical model and source-detector configuration.

range of absorption variations, whereas the linearized formulation can only be used within the limit of linearization.

3.2. Numerical experiments

To simulate the situation of neural activity in the cortex below the skull, we consider a 5 cm $\times$ 5 cm $\times$ 0.7 cm numerical model as shown in Fig. 5. The background optical properties of the numerical phantom are given as absorption coefficient of $\mu_a^b = 0.15$ cm^{-1}, scattering coefficient of $\mu_s^b = 90$ cm^{-1}, anisotropy factor of $g = 0.9$ and refractive index of $n = 1.4$. The 7 mm thickness and $\mu_s^b = 90$ cm^{-1} scattering coefficient of the domain, which corresponds to ~ 60 mean free paths (MFPs), are set to mimic the human skull within acceptable range [62,63]. All targets are located on the bottom (below the "skull") at depth of 7 mm with absorption inhomogeneity of $\mu_a^t = 0.65$ cm^{-1} that is $\sim$4 times larger than the background absorption: absorption contrast is set high in order to show that the SENSNOR algorithm performs well beyond typical absorption contrast (20~40%) in brain activities. The medium is also discretized with a mesh of $N = 51 \times 51 \times 8 = 20808$ voxels. As shown in Fig. 5, 9 sources and 9 detectors are placed at regular intervals of 1 cm on the top surface of the medium. This yields a total of 81 source-detector pairs that can be used for simulated measurement and reconstruction.

For numerical experiments, we generate noise-free data $\bar{z}(t)$ by solving the time-dependent equation of radiative transfer (TD-ERT) with a known distribution of optical properties for test cases, with the same source-detector configuration as shown in Fig. 5. Synthetic measurements containing noise are simulated by adding a noise term to $\bar{z}$ in the form $z(t)^{obs} = \bar{z}(t) + \omega\sigma_{noise}$, where σ_{noise} is the standard deviation of measurement noises and ω is the random variable with normal distribution. This simulated noisy time-of-flight (ToF) measurement $z(t)^{obs}$ are then processed with Mellin and Laplace transforms to give new parameterized data sets given by E, M^1, $L(s)$. For these three parameterized data E, M^1 and L, the corresponding sensitivity and adjoint equations are expressed in the explicit forms as follows:

$$\Delta z_E = \bar{\boldsymbol{\lambda}}_E^T \boldsymbol{\Delta A} \left(\overline{\boldsymbol{\psi}}_E + \boldsymbol{\Delta \psi}_E \right), \text{ and } \overline{\mathbf{A}}^T \bar{\boldsymbol{\lambda}}_E = -\mathbf{Q}^T \tag{33a}$$
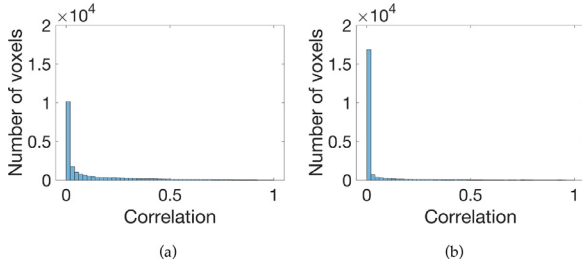
Fig. 6. Normalized correlation distribution between measurement and columns of the sensing matrix: (a) $\Delta\mathbf{Z}$ and $\mathbf{S}$ without preconditioning; (b) with preconditioned $\mathbf{Z}^*$ and $\mathbf{S}^*$.

$$\Delta z_M = \overline{\boldsymbol{\lambda}}_{M^1}^T \left\{ \Delta \boldsymbol{A}\left(\overline{\boldsymbol{\psi}}_{M^1} + \Delta\boldsymbol{\psi}_{M^1}\right) - \boldsymbol{B}\Delta\boldsymbol{\psi}_E \right\}, \text{ and } \overline{\boldsymbol{A}}^T\overline{\boldsymbol{\lambda}}_{M^1} = -\mathbf{Q}^T \quad (33\text{b})$$

$$\Delta z_L = \overline{\boldsymbol{\lambda}}_L^T \Delta\boldsymbol{A}\left(\overline{\boldsymbol{\psi}}_L + \Delta\boldsymbol{\psi}_L\right), \text{ and } \overline{\boldsymbol{A}}^T\overline{\boldsymbol{\lambda}}_L = -\mathbf{Q}^T \quad (33\text{c})$$

which gives $\Delta\boldsymbol{\psi}_E$, $\Delta\boldsymbol{M}^1$ and $\Delta\boldsymbol{\psi}_L$ with Eqs. (17)-(19), and leads to the inverse formulation for the normalized data $\langle t \rangle = M^1/E$ and $\hat{L}(s) = L(s)/E$ as

$$\Delta \langle \mathbf{t} \rangle = \frac{1}{\overline{\mathbf{z}}_E} \overline{\boldsymbol{\lambda}}_{M^1}^T \left\{ \Delta\boldsymbol{A}\left(\overline{\boldsymbol{\psi}}_{M^1} + \Delta\boldsymbol{\psi}_{M^1}\right) - \frac{\delta V}{c}\Delta\boldsymbol{\psi}_E \right\}$$
$$- \frac{\overline{\langle \mathbf{t} \rangle}}{\overline{\mathbf{z}}_M} \overline{\boldsymbol{\lambda}}_E^T \Delta\boldsymbol{A}\left(\overline{\boldsymbol{\psi}}_E + \Delta\boldsymbol{\psi}_E\right) = \mathbf{S}_{nt}^{\langle \mathbf{t} \rangle} \Delta\boldsymbol{\mu} \quad (34\text{a})$$

$$\Delta\hat{\mathbf{L}}(s) = \frac{1}{\overline{\mathbf{z}}_E} \overline{\boldsymbol{\lambda}}_L^T \Delta\boldsymbol{A}\left(\overline{\boldsymbol{\psi}}_L + \Delta\boldsymbol{\psi}_L\right) - \frac{\overline{\hat{L}(s)}}{\overline{\mathbf{z}}_L} \overline{\boldsymbol{\lambda}}_E^T \Delta\boldsymbol{A}\left(\overline{\psi}_E + \Delta\boldsymbol{\psi}_E\right) = \mathbf{S}_{nt}^{\hat{L}} \Delta\boldsymbol{\mu} \quad (34\text{b})$$

The above two normalized quantities $\langle t \rangle$ and $\hat{L}(s)$ are used as input to the reconstruction code as described in Table 1. Also, by definition $L(s) = \int e^{-st}\psi \, dt$, the Laplace parameter s in $L(s)$ describes an exponential change rate in the temporal point spread function (TPSF) curve. Therefore, the small s involves late arriving photons (i.e., photons that travelled deeper and longer inside the medium), whereas the larger s corresponds to early arriving photons (i.e., photons that travelled near the line of sight between the source and the detector). Thus, the parameter s brings region specific data to reconstruction and affects spatial resolution. For our reconstructions, we use $L(s = 0.0001)$ since it brings high sensitivity to absorption changes in the target region ($\sim$at depth of 7 mm).

As we described earlier, we also present here how the correlation between measurement $\Delta\mathbf{Z}$ and sensing matrix $\boldsymbol{\Phi}$ can be used to reduce the sensing matrix for computational efficiency. The correlation distribution $\Delta\mathbf{Z}^T\mathbf{S}$ is computed both with and without the preconditioner given by Eq. (31) for the problem setup with 9 sources and 9 detectors (Fig. 5). Fig. 6 shows a histogram of normalized correlation coefficients between measurements and columns of the sensing matrix (corresponding to voxel contribution). It can be clearly seen that most of voxels have zeros or close-to-zero correlation and only a fraction ($\sim$1600 voxels) has non-zero correlation with measurement and thus needs to be used for reconstruction. Fig. 6 also demonstrates that the preconditioning makes the correlation distribution more favorable towards the sensing matrix reduction. That is, almost 18500 voxels have near-zero correlation for the preconditioned system while, without the preconditioning, fewer elements ($\sim$10000 elements) exhibit near-zero correlation. As a result, with the preconditioning the original sensing matrix can be reduced to 10 % of its original size in the first reduction step alone, and further reduced close to the target sparsity level ($\ll N$) during optimization, which can lead to a great saving in the reconstruction time. Based on this observation, we apply our correlation-based reduction scheme to all reconstructions presented in this work.

3.2.1. Impact of noise

We examine here the effects of noise on the code performance using various absorption inclusions: one to three targets at the same depth and five targets at different depths. The noise impact is tested with noise free and 15 dB noise level based on $SNR = 10\log(z(t)/\sigma)$ where $z(t)$ is the temporal point spread function (TPSF) of light. As mentioned earlier, we use parameterized data (e.g., E, M, L), not TPSF data, for reconstruction, and these parameterized data are analytically obtained with Mellin or Laplace transform of the original TPSF data. So to correctly simulate parameterized data containing noise, we first generate noise-free TPSF ($\bar{z}(t)$) by solving the TD-RTE, and then add a random noise $\varepsilon(t)$ of 15 dB to $\bar{z}(t)$ in order to obtain noise-added TPSF as $z(t) = \bar{z}(t) + \epsilon(t)$. Then we perform Mellin and Laplace transforms of $z(t)$ to compute noise-added parameterized data for E, M and L, respectively, which are used as input to the reconstruction code. As a result, the noise level of parameterized data differs from the noise level of the original TPSF data. For the TPSF data with 15 dB noise, the resulting noise level of parameterized data, after Mellin or Laplace transformation, is reduced to approximately 50 dB. So it should be noted that our reconstruction is done on these $\sim$50 dB parameterized data that is extracted from the 15 dB TPSF data $z(t)$. All targets are of the same size of 1 mm^3. The two-target cases are further divided into two different cases having the targets separated by 1 mm and 2 mm, respectively. Fig. 7 shows the target orientations for all cases considered here. With $\langle t \rangle$ and $\hat{L}(s = 0.0001)$ collected for all 81 source-detector pairs, the reconstructions are performed using the SENSOR method and the rSQP method. The parameters used for the SENSOR reconstruction are as follows: the regularization adjusting factor $\beta_0 = 0.1 \sim 0.01$, the target sparsity level parameter $\sigma_s^T = 1.0 \times 10^{-6}$, the sparsity decreasing factor $t = 0.5$ and the number of maximum iterations $KMAX = 20$. The rSQP reconstruction is done with l_2 norm regularization and terminated when either the objective function value or its relative change becomes as small as 1.0×10^{-6}. Note that numerical results reported here are with the $\langle t \rangle$ data as we found that $\langle t \rangle$ and $\hat{L}(s)$ give almost identical results both in terms of CPU time and accuracy.

Fig. 8 shows the cross-sectional maps at depth of 7 mm from the 3D reconstruction results obtained with the SENSOR and the PDE-constrained method using noise free data (∞ dB). Note that the targets are not seen elsewhere except at depth of 7 mm. It can clearly be seen that the SENSOR algorithm accurately retrieves the targets in all cases, achieving maximum sparsity through l_0-norm minimization, whereas the rSQP method only leads to underestimated and blurry results as commonly observed for l_2-norm based image reconstruction. The line profiles passing across the targets are presented in Fig. 9 to provide a close look at the target area. As shown in the figures, the targets are resolved very well in all cases, even in the 1mm-separation two-target case.

Fig. 10 shows that the results obtained with 50 dB parameterized data ($\sim$15 dB TPSF data) are almost identical to those with noise-free data (also see Table 2), which is true both for the SENSOR method and the rSQP method. This observation demonstrates the robustness of parameterized data to noise. This robustness is attributed mainly to an integral operator of Mellin or Laplace transform used to obtain E, M, and L. In other words, fundamentally, the integral operator serves as low-pass filters, because time varying random noises distributed around true values are cancelled out when integrated over time (Appendix B). As a result, this makes parameterized data E, M, and L essentially far less affected by random noises than the original temporal point spread function (TPSF) that is easily compromised by such noises.

In addition, we also examine the depth resolution with the SENSOR algorithm. To this end, we consider the case with five absorption targets located at different depths: one target at depth of
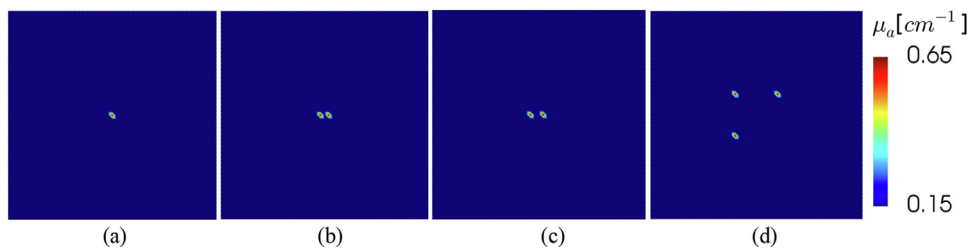
Fig. 7. Exact target orientations and their absorption coefficients for all test cases: (a) single target; (b) two targets 2 mm apart from center to center (1 mm from side to side); (c) two targets 3 mm apart from center to center(2 mm from side to side); (d) three targets. The target size is 1 mm³, the same for all cases.
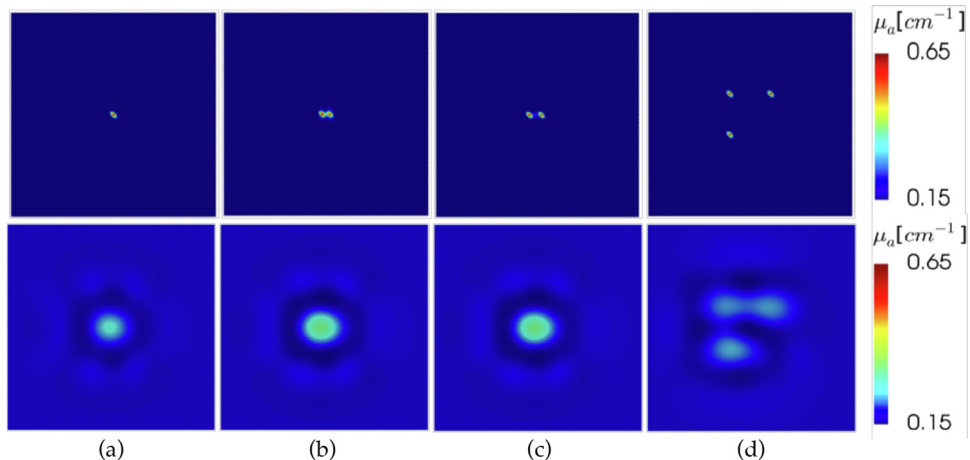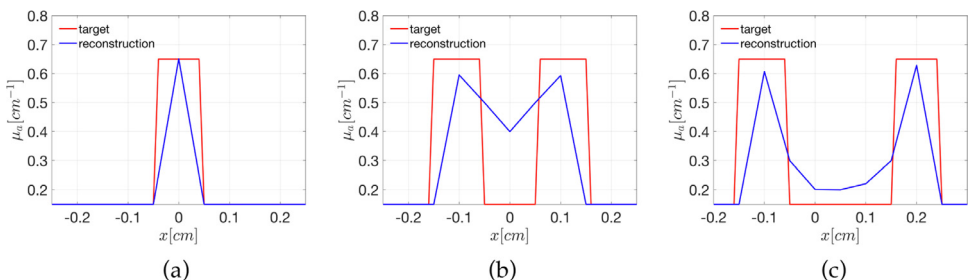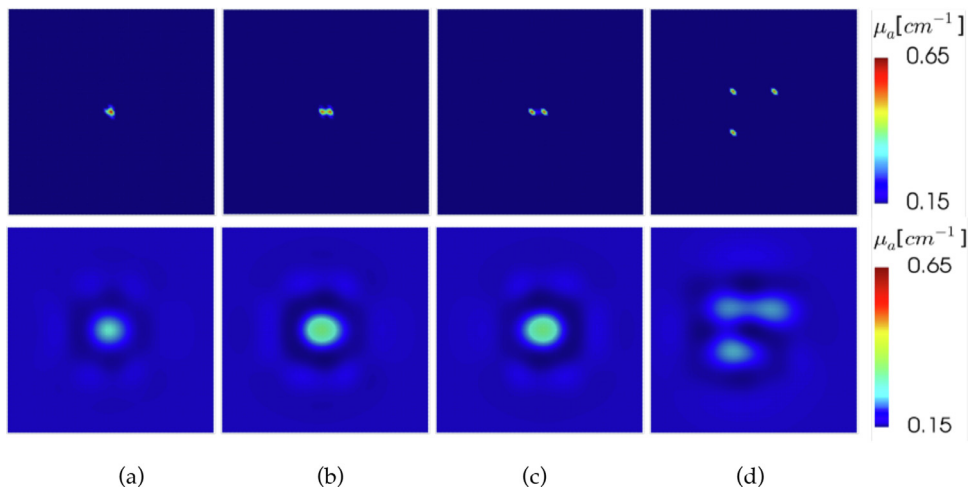

Fig. 8. Reconstructions with noise free data obtained using the SENSOR method (top row) and the rSQP method (bottom row): (a) single target; (b) two targets 2 mm apart from center to center (1 mm from side to side); (c) two targets 3 mm apart from center to center (2 mm from side to side); (d) three targets. The target size is 1 mm³, the same for all cases.

Fig. 9. Line profiles extracted from the images presented in Fig. 8, obtained with the SENSOR method; (a) single target; (b) two targets 2 mm apart from center to center (1 mm from side to side); (c) two targets 3 mm apart from center to center (2 mm from side to side). The target size is 1 mm³, the same for all cases.

Fig. 10. Cross-sectional maps from the 3D reconstruction with parameterized temporal data with 50 dB noise level (corresponding to the TPSF data $z(t)$ with approximately 15dB noise level) by using the SENSOR method (top row) and the rSQP method (bottom row): (a) single target; (b) two targets 2 mm apart from center to center (1 mm from side to side); (c) two targets 3 mm apart from center to center (2 mm from side to side); (d) three targets. The target size is 1 mm³, the same for all cases.

Table 2

The reconstruction CPU times and accuracy obtained with the SENSOR and rSQP methods using noise-free data and noise-added data, respectively. The numbers in parenthesis indicate the speedup factors by SENSOR over rSQP. Note that the 50 dB noise in E, M, L corresponds to approximately 15 dB noise in TPSF $z(t)$.

	Schemes	Cases	CPU time (speedup factor)	Correlation ρ	Deviation δ
Noise free	SENSOR	Single target	**0.021 s (17619)**	**1.0**	**0.01**
		Two-target	**0.021 s (17667)**	**0.95/0.99**	**0.32/0.11**
		Three-target	**0.023 s (16043)**	**0.99**	**0.02**
	rSQP	Single target	370 s	0.22	0.98
		Two-target	371 s	0.27	0.96
		Three-target	369 s	0.18	0.98
50 dB ($\sim$15 dB in $z(t)$)	SENSOR	Single target	**0.021 s (17619)**	**0.99**	**0.01**
		Two-target	**0.023 s (16130)**	**0.91/0.99**	**0.43/0.14**
		Three-target	**0.022 s (16727)**	**0.99**	**0.02**
	rSQP	Single target	370 s	0.22	0.97
		Two-target	371 s	0.25	0.98
		Three-target	368 s	0.18	0.98
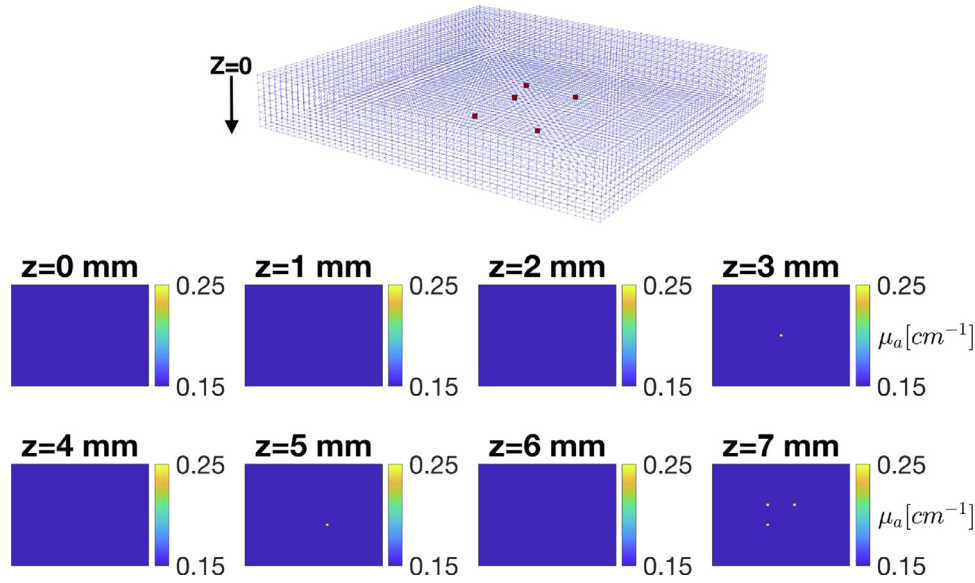
Fig. 11. True and reconstructed distributions of the 5 targets located at different depths: 3D true distribution of the 5 targets (top row); 2D cross-sectional maps at different depths extracted from the 3D reconstruction (second and third rows).

3 mm, one target at depth of 5 mm, and three targets at depth of 6 mm, respectively (Fig. 11). The background optical properties are kept the same as before with $\mu_a^b = 0.15$ cm^{-1} and $\mu_s^{b,t} = 90$ cm^{-1}, but this time the targets are set to have much lower absorption at $\mu_a^t = 0.25$ cm^{-1} to mimic absorption change (20$\sim$40% increase) due to neural activities and metabolism in the human brain. To examine depth resolution, the cross-sectional images are obtained at different depths from the 3D reconstruction. Fig. 11 shows the cross-sectional images at all depths from $z = 0$ cm to 0.7 cm below the top surface for the case with the five targets. The results clearly show that the SENSOR method performs excellent in depth resolution as well as spatial resolution between the targets – all the targets are reconstructed exactly where they are located, not visible elsewhere.

3.2.2. CPU times and accuracy quantification

The accuracy of each reconstruction is quantified using the correlation coefficient $\rho \in [-1, 1]$ and the deviation factor $\delta \in [0, \infty)$ defined as

$$\rho = \frac{\sum_{i=1}^n \left(\mu_i^r - \overline{\mu^r}\right)\left(\mu_i^t - \overline{\mu^t}\right)}{(n-1)\sigma(\mu^r) \cdot \sigma(\mu^t)}, \quad \delta = \frac{\sqrt{\sum_{i=1}^n \left(\mu_i^t - \mu_i^r\right)^2/n}}{\sigma(\mu^t)} \quad (35)$$

where μ^t is the true image, μ^r is the reconstruction image, $\overline{\mu^t}$ and $\overline{\mu^r}$ are the mean values and $\sigma(\mu^t)$ and $\sigma(\mu^r)$ are the standard de-

viations of μ^t and μ^r respectively, and n refers to the total number of distinct mesh nodes. By definition, the correlation coefficient ρ can take any value between -1 and 1, with $\rho = 1.0$ corresponding to an exact match between the true and reconstruction images, but it typically takes the positive value between 0 and 1 unless input data to reconstruction is extremely noisy. The deviation factor $\delta \in [0, \infty)$ denotes normalized root mean square error (RMSE). Accordingly, the closer ρ gets to 1 and δ to 0, the better is quality of reconstruction.

Table 2 provides the CPU times and accuracy in terms of correlation ρ and deviation δ factors as given in Eq. (35). As shown in Table 2, the SENSOR method gives a significant speedup in the reconstruction process in all cases considered here. For the case of the noise-free data, the SENSOR method achieves sparsest solution in $\sim$20 milliseconds, while the rSQP method takes about $\sim$370 seconds to convergence. Therefore, the SENSOR method is about 17000 times faster than the rSQP method. Similar CPU times are observed for the noise case: 20-23 milliseconds with the SENSOR that corresponds to a speedup factor of $\sim$17000 (see Table 2). Note that both of the two methods are implemented with OpenMP for parallel computing and executed on the same machine (Intel Core i9 process). Therefore, it is clear that the speedup of the SENSOR method is achieved by the algorithm advances made with the non-iterative nature of the inverse formulation and the sensing matrix reduction, which is not exploited by the iterative rSQP method.
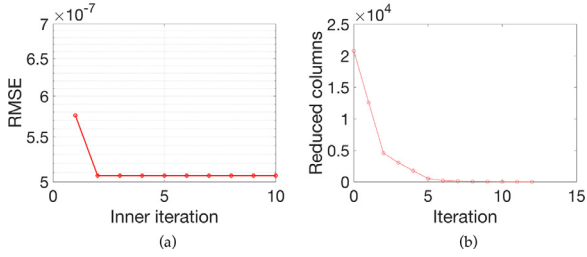
Fig. 12. Impacts of inner iterations and sensing matrix reduction: (a) RMSE with increasing inner iterations; (b) reduced sensing matrix with the number of outer iterations.

CPU times can also be affected by inner iterations that are performed for each σ_s in the sequence $\{\sigma_s^1, \sigma_s^2, \ldots, \sigma_s^{IMAX}\}$ as described in the algorithm (Table 1). Fig. 12a shows how the number of inner iterations affects convergence for the case with three targets. In this case, inner iterations improve the data fit only for the first two iterations, which suggests that the number of inner iterations can be reduced to a small number to avoid unnecessary inner iterations. Based on this observation, the number of inner iterations is fixed at 2 in all simulation cases presented here. The sensing matrix reduction is one of the major algorithmic features leading to fast convergence. Fig. 12b demonstrates how efficiently the sensing matrix can be truncated at each step of the outer iterations based on the correlation threshold: starting from the 20808 columns (corresponding to the 20808 voxels of the medium), the matrix is reduced to only 7 columns at convergence, which leads to a significant time saving in the total reconstruction time.

In addition to the CPU time, the accuracy of reconstruction is measured and reported here to address spatial resolution. The correlation factor $\rho(\mu_e, \mu_r) \in [-1, 1]$ and the deviation factor $\delta(\mu_e, \mu_r) \in [0, \infty)$ as defined in Eq. (35) are computed for the bottom area where the absorption targets are located. The values of $\rho(\mu_e, \mu_r)$ and $\delta(\mu_e, \mu_r)$ are given in Table 2. In all cases, the SENSOR method provides very high correlation factors of $\rho \approx 1$, which means highest correlation (i.e., highest spatial resolution) is achieved. The deviation factors δ, which indicates RMSE, is much lower with the SENSOR method in all cases than the rSQP method. In both methods, as expected, the 1mm-separation case gives larger deviation from true image than the 2mm-separation case, i.e., $\delta^{1mm} = 0.32$ and $\delta^{2mm} = 0.11$ with the SENSOR method, while $\delta^{1mm} = 0.43$ and $\delta^{2mm} = 0.14$. Also, there is no noticeable difference in both $\rho(\mu_e, \mu_r)$ and $\delta(\mu_e, \mu_r)$ between noise-free and 15 dB cases for both methods. Again, as discussed earlier, this is due to the robustness of parameterized data used here on noise.

3.3. Experimental results

We present here absorption reconstruction results using experimental data. Fig. 13 shows the time-domain (TD) system for measurements of time-of-flight (ToF) light intensities on the medium surface. The system consists of three main parts: illumination, detection, and target medium. For illumination, we use an ultrashort pulsed laser (NKT Photonics SuperK EXTREME) that emits 100 femtoseconds pulses at wavelength of 680 nm as a light source, with duration of 100 femtoseconds. The detection of temporal light intensities traveling back to the medium surface is performed by a time-correlated single photon counting (TCSPC) system from PicoQuant that employs a fast-gated single-photon avalanche diode (SPAD) from MPD with a temporal resolution of 50 picoseconds. As shown in Fig. 13, the galvo mirrors are used to allow precise light delivery onto and from the skull tissue phantom.

Similar to the numerical phantom, the 5 cm × 5 cm × 0.7 cm scattering medium is employed as the background medium.

The optical properties of this background medium are given by $\mu_a = 0.0$ cm^{-1}, $\mu_s = 90$ cm^{-1}, $g = 0.9$ and $n = 1.4$. Locations and absorption strengths of targets are generated by using an E-Ink display that is placed behind the scattering phantom block (Fig. 13). The display panel has an array of pixels that can be programmed to generate various patterns and structures at different gray levels (i.e., different absorption levels) on a white background. Note that the white background of the display panel is considered as perfect scatterer. With this experimental setup, we perform spatial resolution tests: one single target, two targets at different separations and more complex pattern target.

3.3.1. Single-target and two-target cases

We looked at the single and two-target cases. As in numerical experiments, the targets are of the same size of 1 mm^3 and all located on the bottom at depth of 7 mm. The two-target case has the targets separated by 1mm distance from side to side (2mm center to center). For computational efficiency, we use a dense source-detector configuration, in which the 5 × 5 sources and 5 × 5 detectors are located at intervals of 2mm approximately on the 0.5 cm × 0.5 cm top surface area over the targets located in the middle of the bottom surface. As the sources and detectors are clustered on smaller region than in numerical simulations, the computational domain for reconstruction is also reduced to the dimension of 2 cm × 2 cm × 0.7 cm accordingly (Fig. 14). This source-detector configuration leads to a total of 25 × 25 time-of-flight (ToF) measurements. Two separate measurements are made for calibration purpose: one without a target (so-called "reference state") and one with a target (so-called "target state").

Fig. 15 shows an example of ToF light intensity measurement for the reference state (blue) and the target state (red) obtained with source number 7. ToF data is treated with a 50-point Gaussian fit for noise removal and transformed with parameterized data to give integrated signal E, Mellin 1st moment M_1, Laplace transform $L(s)$. The new data E, M_1 and $L(s)$ are then normalized to the reference predictions to give the calibrated mean time $<t>_{calib}$ and normalized Laplace transform $\hat{L}_{calib}$ (Fig. 12b) as:

$$E_{calib} = \frac{E_{m,tar}}{E_{m,ref}} \cdot E_{p,ref}, \quad M_{calib} = \frac{M_{m,tar}}{M_{m,ref}} \cdot M_{pred,ref},$$

$$L(s)_{calib} = \frac{L_{m,tar}}{L_{m,ref}} \cdot L_{p,ref} \tag{35a}$$

$$<t>_{calib} = \frac{M_{calib}}{E_{calib}}, \quad \hat{L}(s)_{calib} = \frac{L_{calib}}{E_{calib}}. \tag{35b}$$

The differences $<t>_{calib,tar} - <t>_{calib,ref}$ and $\hat{L}(s)_{calib,tar} - \hat{L}(s)_{calib,ref}$ are computed for each source and detector pair and are used as input to the image reconstruction algorithm in the form of matrix Z as described in Sec. 2.1 and 2.2.

The data $\langle t \rangle$ and $\hat{L}(s = 0.0001)$, as mentioned earlier, are used for the reconstructions. Fig. 16 shows the cross-sectional maps of the reconstructed targets for the single and two-target cases. The line profiles across the targets are also extracted from the same images to provide clear visualization of spatial resolution achieved for the cases. As shown in Fig. 16, both single and two targets are accurately recovered with the targets. The code took about 6 milliseconds for the single target case, and 7 milliseconds for the two targets case, both on an Intel Core i9 processor.

3.3.2. Three-target and arbitrary pattern cases

We also test additional experimental data using 3 randomly distributed targets and arbitrary pattern cases. For the three-target case, the targets have the same size of 4 mm in side length and are separated approximately by 1 cm distance between the targets. The three randomly distributed targets case has 4mm-square size,
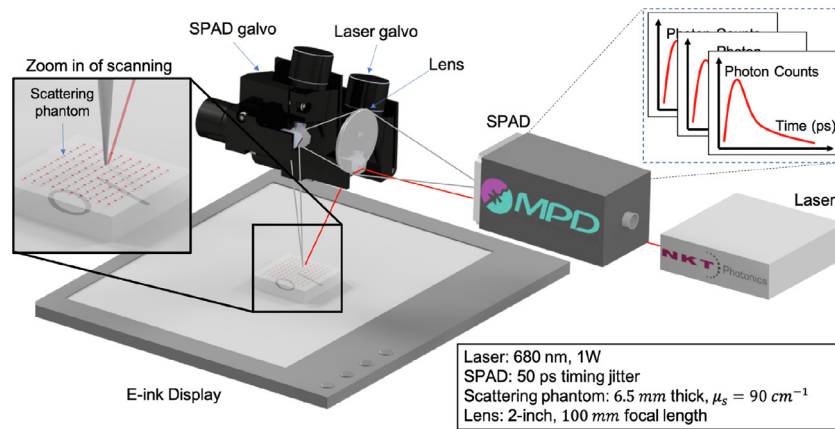
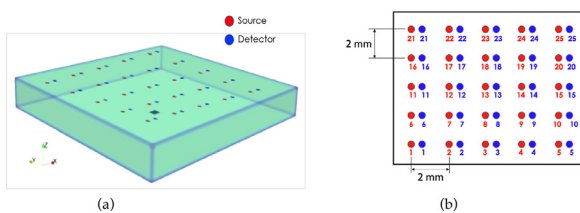**Fig. 13.** Experimental setup for time-of-flight light measurement.

Fig. 14. A 0.2 cm × 0.2 cm × 0.1cm voxel target case: (a) a 2cm × 2cm × 0.7cm experimental phantom with a single square target at the bottom; (b) source-detector configuration.

Fig. 16. Image reconstruction results and line profiles for the 1 mm-square target (first column) and the two 1 mm-square targets in 1mm separation from side to side (second column).

and the arbitrary line pattern absorption case is also tested with line thickness of 0.2 cm. The absorption coefficients of the targets are set to be $\mu_a^t = 1.3$ cm^{-1}. To achieve desired spatial resolution in these cases, all reconstructions presented here are carried out using the 10 × 10 sources and 10 × 10 detectors distributed at regular intervals over the 4cm x 4cm x 0.7cm computational domain. The cross-sections from reconstruction are presented in Fig. 17 for each case, showing that the targets are accurately resolved from side to side in all test cases. The reconstruction time is around 30 milliseconds for all cases considered here.

3.3.3. Different target-background contrast cases

The reconstruction accuracy is also affected by the background absorption: for example, the reconstruction of a small absorption object embedded in the high absorbing background can often pose a challenge to achieving quality reconstruction of high resolution. To address the effects of absorption varying background on accuracy, we have further tested the algorithm on non-white background cases, in which the background medium is varied to different gray levels leading to different absorption contrasts to the
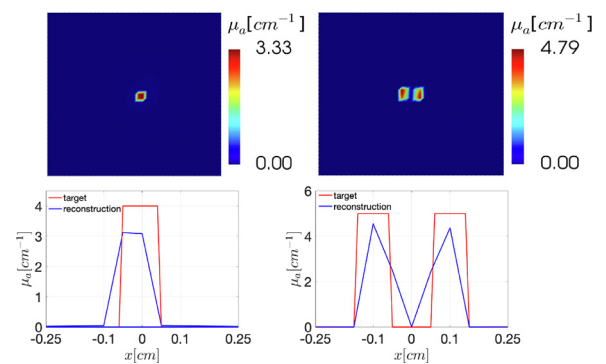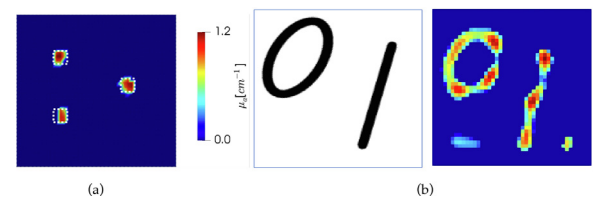

Fig. 17. Image reconstruction results of (a) three random targets; (b) arbitrary pattern. The white dot lines in (a) indicate the target boundaries.

target. To this end, we consider here the two-target case with the targets in size of 1 mm^3 separated by 1 mm from side to side, located at the depth of 7 mm as before. Four different contrasts are tested with varying background absorption as $\mu_a^b = 0.26$ cm^{-1},
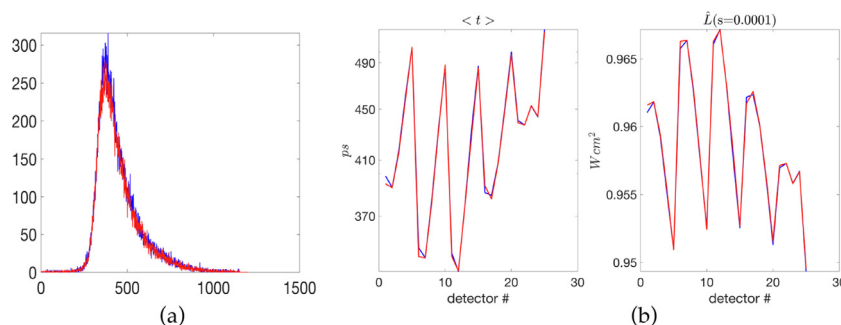

Fig. 15. Examples of ToF measurement and parameterized temporal responses: (a) ToF measurement with source 7 and detector 19; (b) parameterized data, E, $\langle t \rangle$, $\hat{L}(s)$ with $s = 0.0001$.
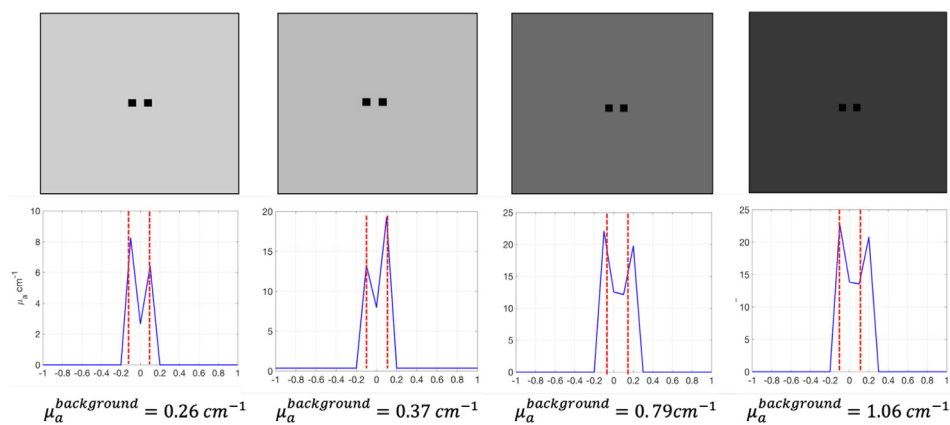
Fig. 18. Image reconstruction results with different target-background contrasts. The top row indicates the cross-sectional map of each case, while the bottom row shows the line profiles extracted from the cross-section maps of the 3D reconstructions.

0.37 cm^{-1}, 0.79 cm^{-1} and 1.06 cm^{-1}. Absorption targets (black dots) in this experiment are almost perfect absorbers of light. Measurements are taken with 5×5 sources and 5×5 detectors equidistantly spaced within the 2 cm $\times$ 2 cm surface area. Fig. 18 shows the line profiles extracted from the reconstructions (bottom row), along with the true distribution of the two targets (top row). It can be seen that in all cases the separation between the two targets are clearly visible, even in the case of the highly absorbing background given by $\mu_a^b = 1.06$ cm^{-1}. These results demonstrate that the SENSOR algorithm, together with the use of parameterized data, can effectively differentiate between the background and the targets to a great variety of contrast.

4. Summary and Conclusions

In this work, we have developed a novel image reconstruction algorithm called sensitivity-equation-based noniterative sparse optical reconstruction (SENSOR). This algorithm can be used to provide ultrahigh spatial and temporal resolution optical tomographic neuroimaging. The new method makes use of a nontruncated sensitivity equation, which allows to reformulate a nonlinear iterative inverse problem into a noniterative inverse problem for fast one-shot image reconstruction. Further acceleration has been achieved by implementing the sensing matrix reduction scheme that exploits the correlation information between measurements and columns of the sensing matrix. In addition, the new method also adopts an asymptotic l_0-norm function into the inverse model, which promotes sparsest representation of absorption changes scarcely distributed in the brain. The method has achieved additional speedup with parallel computing through OpenMP on a multicore processor.

The method has been tested using both numerical and experimental data. The numerical simulations are used to evaluate the performance of the SENSOR method with respect to CPU times and accuracy, with comparison to the fastest iterative image reconstruction method (rSQP). The spatial resolution and speed of the method is further tested on various experimental data sets: single target cases with different target sizes, multiple targets cases in different separations, and also arbitrary line pattern case. The reconstruction results demonstrate that the SENSOR method is capable of producing high-resolution image of absorption changes in approximately 1 mm^3 volumes at depth of 7 mm ($\sim$60 MFPs) within the medium in 20~30 milliseconds. This opens the door, for example, to non-invasive cellular-resolution real-time monitoring of neuronal electrical activities, which we will pursue in the future.

While the algorithm presented in this paper is based on time-domain data, the approach presented here can be easily adapted, without loss of generality, to continuous wave (CW) and frequency domain (FD) systems. It is therefore expected that the SENSOR method can greatly aid to provide further applications beyond neural imaging.

Author Statement

Hyun Keol Kim: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Data Curation, Visualization, Writing-Original Draft Preparation, Writing-Reviewing & Editing; **Yongyi Zhao:** Validation, Formal Analysis, Investigation, Resources, Data Curation, Visualization; **Ankit Raghuram:** Validation, Formal Analysis, Investigation, Resources, Data Curation, Visualization; **Ashok Veeraraghavan**; Resources, Supervision, Project administration, Funding Acquisition; **Jacob Robinson**; Resources, Supervision, Project Administration, Funding Acquisition; **Andreas Hielscher**; Resources, Supervision, Project Administration, Funding Acquisition; Writing – Reviewing & Editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix

Appendix A

The exact sensitivity equation can also be derived from using the relationship $z = -\boldsymbol{\lambda}^\mathsf{T}\mathbf{b}$. To this end, we write the target state

system in terms of the reference state and the variation (or perturbation) with a fixed source **b** as:

$$A\psi = \mathbf{b} \rightarrow \left(\overline{A} + \Delta A\right)\left(\overline{\psi} + \Delta\psi\right) = \mathbf{b}. \tag{A.1}$$

Taking the inner product of Eq. (A1) with $\overline{\lambda}$ gives

$$\overline{\lambda}^{\mathrm{T}}\overline{A}\left(\overline{\psi} + \Delta\psi\right) + \overline{\lambda}^{\mathrm{T}}\Delta A\left(\overline{\psi} + \Delta\psi\right) = \overline{\lambda}^{\mathrm{T}}\mathbf{b} \tag{A.2}$$

With the adjoint model $\overline{\lambda}^{\mathrm{T}}\overline{A} = -\mathbf{Q}$ and the identity $\mathbf{P} = \mathbf{Q}\psi = -\lambda^{\mathrm{T}}\mathbf{b}$, Eq. (A2) can be rewritten as

$$\mathbf{Q}\left(\overline{\psi} + \Delta\psi\right) + \overline{\lambda}^{\mathrm{T}}\Delta A\left(\overline{\psi} + \Delta\psi\right) = \overline{\lambda}^{\mathrm{T}}\mathbf{b} \tag{A.3}$$

$$z + \overline{\lambda}^{\mathrm{T}}\Delta A\left(\overline{\psi} + \Delta\psi\right) = \overline{\mathbf{P}} \tag{A.4}$$

$$\Delta z = z - \overline{z} = -\overline{\lambda}^{\mathrm{T}}\Delta A\left(\overline{\psi} + \Delta\psi\right) \tag{A.5}$$

Thus, the above formulation given by Eq. (A5) is equivalent to the inverse formulation given by Eq. (13a) based on $z = \mathbf{Q}\psi$.

Appendix B

In a multispectral tomographic model that exploits the linear relationship given by $\Delta\mu_{\mathbf{a}} = \epsilon_{\mathrm{M}}^{\mathrm{T}}\Delta\mathbf{c}$ where $\epsilon_{\mathbf{M}} = (\epsilon_{HbO2}^{w}, \epsilon_{Hb}^{w}, \ldots, \epsilon_{other}^{w})$ denotes molecular extinction coefficients between tissue chromophores and absorption coefficient, the SENSOR method can be applied with the following similar formulations for a tomographic image of chromophore concentration changes $\Delta\mathbf{c}$ such as $\Delta\mathbf{HbO}_2$ and $\Delta\mathbf{Hb}$. Given the multispectral data $\Delta z^{w^1}, \Delta z^{w^2}, \ldots, \Delta z^{w^m}$ from multiple wavelengths (w^1, w^2, $\ldots$, w^m), Eq. (14) can be reformulated to allow for a multispectral approach as:

$$\Delta z^{w^1} = \overline{\lambda}^{\mathrm{T},w^1}\sum_{i}\left(\frac{\partial \mathbf{A}^{w^1}}{\partial\mu_{\mathbf{a}}^{\mathbf{w^1}}} \times \frac{\partial\mu_{\mathbf{a}}^{w^1}}{\partial\mathbf{c}^{i}} \times \Delta\mathbf{c}^{i}\right)\left(\overline{\psi}^{w^1} + \Delta\psi^{w^1}\right) \tag{B.1}$$

$$\Delta z^{w^2} = \overline{\lambda}^{\mathrm{T},w^2}\sum_{i}\left(\frac{\partial \mathbf{A}^{w^2}}{\partial\mu_{\mathbf{a}}^{\mathbf{w^2}}} \times \frac{\partial\mu_{\mathbf{a}}^{w^2}}{\partial\mathbf{c}^{i}} \times \Delta\mathbf{c}^{i}\right)\left(\overline{\psi}^{w^2} + \Delta\psi^{w^2}\right) \tag{B.2}$$

$$\vdots$$

$$\Delta z^{w^m} = \overline{\lambda}^{\mathrm{T},w^m}\sum_{i}\left(\frac{\partial \mathbf{A}^{w^m}}{\partial\mu_{\mathbf{a}}^{\mathbf{w^m}}} \times \frac{\partial\mu_{\mathbf{a}}^{w^m}}{\partial\mathbf{c}^{i}} \times \Delta\mathbf{c}^{i}\right)\left(\overline{\psi}^{w^m} + \Delta\psi^{w^m}\right) \tag{B.3}$$

where the superscripts $w^1, w^2, \ldots, w^m$ denote the wavelengths, the variables $\overline{\lambda}$, A, $\overline{\psi}$ and $\Delta\psi$ are evaluated as a function of wavelength and $\Delta\mathbf{c} = (\Delta HbO_2, \Delta Hb, \Delta H_2O, \ldots)^{\mathrm{T}}$. As a result, Eqs. (B.1)-(B.3) can be put together to form a following matrix-vector system as:

$$\begin{bmatrix} \Delta z^{w^1} \\ \Delta z^{w^2} \\ \vdots \\ \Delta z^{w^m} \end{bmatrix} = \begin{bmatrix} \overline{\lambda}^{\mathrm{T},w^1}\left(\frac{\partial\mathbf{A}^{w^1}}{\partial\mu_{\mathbf{a}}^{\mathbf{w^1}}} \times \frac{\partial\mu_{\mathbf{a}}^{w^1}}{\partial\mathbf{c}}\right)\left(\overline{\psi}^{w^1} + \Delta\psi^{w^1}\right) \\ \overline{\lambda}^{\mathrm{T},w^2}\left(\frac{\partial\mathbf{A}^{w^2}}{\partial\mu_{\mathbf{a}}^{\mathbf{w^2}}} \times \frac{\partial\mu_{\mathbf{a}}^{w^2}}{\partial\mathbf{c}}\right)\left(\overline{\psi}^{w^2} + \Delta\psi^{w^2}\right) \\ \overline{\lambda}^{\mathrm{T},w^m}\left(\frac{\partial\mathbf{A}^{w^m}}{\partial\mu_{\mathbf{a}}^{\mathbf{w^m}}} \times \frac{\partial\mu_{\mathbf{a}}^{w^m}}{\partial\mathbf{c}}\right)\left(\overline{\psi}^{w^m} + \Delta\psi^{w^m}\right) \end{bmatrix}[\Delta c] \Leftrightarrow \Delta\mathbf{Z}^{\mathbf{w}}$$

$$= \mathbf{S}_{\mathbf{nt}}^{\mathbf{w}}\Delta\mathbf{c} \tag{B.4}$$

which can be solved following the same procedure given in Sections 2.2-2.4 for the chromophore change $\Delta\mathbf{c}$.
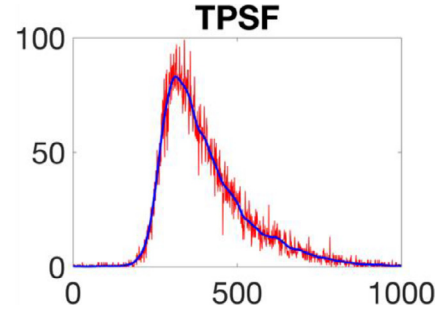

Fig. C.1. Example of noise free data (blue) and noise added data (red) of temporal point spread function (TPSF) of light.

Appendix C

As shown in Fig. B, temporal data $z(t)$ containing random noise can be written as a sum of true (noise-free) signal $\overline{z}(t)$ and noise term $\varepsilon(t)$ as

$$z(t) = \overline{z}(t) + \epsilon(t) \tag{C.1}$$

Since Mellin transform is defined as $M^n = \int_0^\infty t^n\Gamma(t)dt$, plugging (B1) into M^n gives the following:

$$M^n = \int_0^\infty t^n\left(\overline{\Gamma}(t) + \epsilon(t)\right)dt \tag{C.2a}$$

$$= \int_0^\infty t^n\overline{\Gamma}(t)dt + \int_0^\infty t^n\epsilon(t)dt \tag{C.2b}$$

$$\approx \int_0^\infty t^n\overline{\Gamma}(t)dt + 0 \tag{C.2c}$$

when $\varepsilon(t)$ is assumed to follow normal distribution. The same can be true for Laplace transformed data $L(s)$. Thus an integral operator of parameterized data serves as a low pass filter that effectively removes noise by cancelling out random noise distributed around true values when integrated over time (see Fig. C.1). As a result, this makes parameterized data *E, M,* and *L* essentially far less affected by random noises than the original temporal point spread function (TPSF) easily compromised by such noises.

References

[1] Power JD, Cohen AL, Nelson SM, Wig GS, Barnes KA, Church JA, Vogel AC, Laumann TO, Miezin FM, Schlaggar BL, Petersen SE. Functional Network Organization of the Human Brain. Neuron 2011;72:665.

[2] Yeo BT, Krienen FM, Sepulcre J, Sabuncu MR, Lashkari D, Hollinshead M, Roffman JL, Smoller JW, Zollei L, Polimeni JR, Fischl B, Liu H, Buckner RL. The organization of the human cerebellum estimated by intrinsic functional connectivity. J. Neurophysiol. 2011;106:1125.

[3] Buckner RL, Sepulcre J, Talukdar T, Krienen FM, Liu H, Hedden T, Andrews-Hanna JR, Sperling RA, Johnson KA. Cortical hubs revealed by intrinsic functional connectivity: mapping, assessment of stability, and relation to Alzheimer's disease. J. Neurosci. 2009;29:1860.

[4] Baldassarre A, Ramsey LE, Siegel JS, Shulman GL, Corbetta M. Brain connectivity and neurological disorders after stroke. Curr. Opin. Neurol. 2016;29:706.

[5] Crosson B, Ford A, McGregor KM, Meinzer M, Cheshkov S, Li X, Walker-Batson D, Briggs RW. Functional imaging and related techniques: an introduction for rehabilitation researchers. Journal of rehabilitation research and development 2010;47(2) vii–xxxiv.

[6] Boas DA, Dale AM, Franceschini MA. NeuroImage 2004;23(Suppl 1):S275–88.

[7] Ferradal SL, Liao SM, Eggebrecht AT, Shimony JS, Inder TE, Culver JP, Smyser CD. Functional Imaging of the Developing Brain at the Bedside Using Diffuse Optical Tomography. Cereb. Cortex 2016;26:1558–68.

[8] Chalia M, Lee CW, Dempsey LA, Edwards AD, Singh H, Michell AW, Everdell NL, Hill RW, Hebden JC, Austin T, Cooper RJ. Neurophotonics 2016;3:031408.

[9] Dai X, Zhang T, Yang H, Tang J, Carney PR, Jiang H. Fast noninvasive functional diffuse optical tomography for brain imaging. J Biophotonics 2018;11(3) p.e201600267.

[10] Xu H, Springett R, Dehghani H, Pogue BW, Paulsen KD, Dunn JF. Magnetic- res- onance-imagingcoupled broadband near-infrared tomography system for small animal brain studies. Appl. Opt. 2005;44:2177–88.

[11] Bluestone A, Abdoulaev G, Schmitz C, Barbour R, Hielscher A. Three-dimen- sional optical tomography of hemodynamics in the human head. Opt. Express 2001;9(6):272–86.

[12] Zeff BW, White BR, Dehghani H, Schlaggar BL, Culver JP. Retinotopic mapping of adult human visual cortex with high-density diffuse optical tomography. Proc. Natl. Acad. Sci. U.S.A. 2007;104(29):12169–74.

[13] Eggebrecht AT, Ferradal SL, Robichaux-Viehoever A, Hassanpour MS, De- hghani H, Snyder AZ, Hershey T, Culver JP. Mapping distributed brain function and networks with diffuse optical tomography. Nature Photonics 2014;8:448–54.

[14] Fishell AK, Burns-Yocum TM, Bergonzi KM, et al. Mapping brain function dur- ing naturalistic viewing using high-density diffuse optical tomography. Sci Rep 2019;9:11115.

[15] Chitnis D, Cooper RJ, Dempsey L, et al. Functional imaging of the human brain using a modular, fibre-less, high-density diffuse optical tomography system. Biomed. Opt. Express 2016;7:4275–88.

[16] Ferradal SL, Eggebrecht AT, Hassanpour M, Snyder AZ, Culver JP. Atlas-based head modeling and spatial normalization for high-density diffuse optical to- mography: in vivo validation against fMRI. Neuroimage 2014;85(Pt 1):117–26.

[17] Eggebrecht AT, White BR, Ferradal SL, Chen C, Zhan Y, Snyder AZ, Dehghani H, Culver JP. A quantitative spatial comparison of high-density diffuse optical to- mography and fMRI cortical mapping. Neuroimage 2012;61(4):1120–8.

[18] Asllanaj F, Contassot-Vivier S, Hohmann A, Kienle A. Light propagation in biological tissue. Journal of Quantitative Spectroscopy & Radiative Transfer 2019;224:78–90.

[19] Jia J, Kim HK, Hielscher AH. Fast linear solver for radiative transport equation with multiple right hand sides in diffuse optical tomography. J Quant Spectrosc Radiat Transf 2015;167:10–22.

[20] Kim H K, Charette A. A sensitivity function-based conjugate gradient method for optical tomography with the frequency-domain equation of radiative trans- fer. J. Quant. Spec. Rad. Trans. 2007;104:24–39.

[21] Kim HK, Hielscher AH. A PDE-constrained SQP algorithm for optical tomog- raphy based on the frequency-domain equation of radiative transfer. Inverse Problems 2009;25:015010 online-doi:. doi:10.1088/0266-5611/25/1/015010.

[22] Kim HK, Flexman M, Yamashiro DJ, Kandel J, Hielscher AH. PDE-constrained Multispectral Imaging of Tissue Chromophores Concentrations with the Equa- tion of Radiative Transfer. Biomedical Optics Express 2010;1(3):812–24.

[23] Wang L, Jacques SL, Zheng L. MCML–Monte Carlo modeling of light transport in multi- layered tissues. Computer Methods and Programs in Biomedicine 1995;47(2):131–46.

[24] Arridge SR. Optical tomography in medical imaging. Inverse Problems 2009;15:R41–93.

[25] Arridge SR, Schotland JC. Optical tomography: forward and inverse problems. Inverse Problems 2009;25(12):123010.

[26] O'Leary MA, Boas DA, Chance B, Yodh AG. Experimental images of heteroge- neous turbid media by frequency-domain diffusing-photon tomography. Opt. Lett. 1995;20:426–8.

[27] Hielscher AH, Alcouffe AE, Barbour RL. Comparison of finite-difference trans- port and diffusion calculations for photon migration in homogeneous and het- erogeneous tissues. Phys. Med. Biol. 1998;43:1285–302.

[28] Shen H, Cong W, Qian X, Durairaj K, Wang G. Numerical study on the validity of the diffusion approximation for computational optical biopsy. J Opt Soc Am A Opt Image Sci Vis 2007;24(2):423–9.

[29] Riley J, Dehghani H, Schweiger M, Arridge S R, Ripoll J, Nieto-Vesperi- nas M. 3D optical tomography in the presence of void regions. Opt. Express 2000;7:462–7.

[30] Dehghani H, Arridge SR, Schweiger M, Delpy DT. Optical tomography in the presence of void regions. J. Opt. Soc. Am. A 2000;17:1659–70.

[31] Elisee J, Bonnet M, Arridge S. Accelerated boundary element method for diffuse optical imaging. Opt. Lett. 2011;36:4101–3.

[32] Kim HK, Montejo L, Jia J, Hielscher AH. Frequency-domain optical tomo- graphic image reconstruction algorithm with the simplified spherical harmon- ics (SP3) light propagation model. International Journal of Thermal Science 2017;116:265–77.

[33] Dominguez JB, Berube-Lauziere Y. Diffuse light propagation in biological me- dia by a time- domain parabolic simplified spherical harmonics approximation with ray-divergence effects. Appl. Opt. Mar 2010;49:1414–29.

[34] Klose AD, Larsen EW. Light transport in biological tissue based on the simplified spherical harmonics equations. Journal of Computational Physics 2006;220(1):441–70.

[35] Ovsepian SV, Olefir I, Westmeyer G, Razansky D, Ntziachristos V. Pushing the Boundaries of Neuroimaging with Optoacoustics. Neuron 2017;96:966–88.

[36] Hillman EMC. Optical brain imaging in vivo: techniques and applications from animal to man. J Biomed Opt 2008;12:051402.

[37] Arridge SR, Schweiger M. Direct calculation of the moments of the distribution of photon time of flight in tissue with a finite-element method. Applied Optics 1995;34:2683–7.

[38] Schweiger M, Arridge SR. Direct calculation with a finite-element method of the Laplace transform of the distribution of photon time of flight in tissue. Applied Optics 1997;36:9042–9.

[39] Hillman EMC, Hebden JC, Schweiger M, Dehghani H, Schmidt FEW, Delpy DT, Arridge SR. Time resolved optical tomography of the human forearm. Phys. Med. Biol. 2002;46:1117–30.

[40] Gao F, Tanikawa Y, Zhao HJ, Yamada Y. Semi-three-dimensional algorithm for time-resolved diffuse optical tomography by use of the generalized pulse spec- trum technique. Appl. Opt. 2002;41:7346–58.

[41] Gao F, Zhao H, Tanikawa Y, Yamada Y. Optical tomographic mapping of cere- bral haemodynamics by time-domain detection: methodology and phantom validation. Phys. Med. Biol. 2004;49:1055–78.

[42] Zhao Huijuan, Gao Feng, Tanikawa Yukari, Homma Kazuhiro, Yamada Yukio. Time-resolved optical tomographic imaging for the provision of both anatomical and functional information about biological tissue. Appl. Opt. 2005;43:1905–16.

[43] Box MA, Loughlin PE, Samaras M, Trautmann T. Applications of radia- tive perturbation theory to changes in absorbing gas. J. Geophys. Res. 1997;102:4333–42.

[44] Lewins J. Time-Dependent Variational Theory. Nucl. Sei. Eng. 1968;31:160.

[45] Marchuk GI. Adjoint Equations and Analysis of Complex Systems. Amsterdam: Kluwer; 1995.

[46] Gerstl SAW. Application of the adjoint method in atmospheric radiative trans- fer calculations. In: Deepak A, editor. Atmospheric Aerosols: Their Forma- tion, Optical Properties, and Effects. Hampton, VA: Spec- trum Press; 1982. p. 241–54.

[47] Henyey LG, Greenstein LJ. Diffuse radiation in the galaxy Astrophys 1941;90:70.

[48] Tropp JA, Gilbert AC. Signal recovery from random measurements via or- thogonal matching pursuit. IEEE Trans- actions on Information Theory 2007;53(12):4655–66.

[49] Lee O, Kim J, Bresler Y, Ye J. Compressive diffuse optical tomography: non- iterative exact reconstruction using joint sparsity. IEEE Trans. Med. Imag. 2011;30:1129–42.

[50] Lee O, Kim J, Bresler Y, Ye JC. Diffuse optical tomography using generalized music algorithm. In: 2011 IEEE International Symposium on Biomedical Imag- ing: From Nano to Macro; 2011. p. 1142–5. doi:10.1109/ISBI.2011.5872603.

[51] Donoho DL, Tsaig Y, Drori I, Starck JL. Sparse solution of underdetermined lin- ear equations by stagewise orthogonal matching pursuit. IEEE Transactions on Information Theory 2012;2:1094–121.

[52] Zhu XX, Bamler R. Tomographic SAR inversion by L1- norm regularization-the compressive sensing approach. IEEE Transactions on Geoscience and Remote Sensing 2010;48(10):3839–46.

[53] Long T, Jiao W, He G. RPC estimation via l1-norm- regularized least squares (L1LS). IEEE Transactions on Geo- science and Remote Sensing 2015;8:4554–67.

[54] Bhowmik T, Liu H, Ye Z, Oraintara S. Dimensionality Reduction Based Opti- mization Algorithm for Sparse 3-D Image Reconstruction in Diffuse Optical To- mography. Scientific Reports 2016;6:22242.

[55] Kavuri VC, Lin ZJ, Tian F, Liu H. Sparsity enhanced spatial resolution and depth localization in diffuse optical tomography. Biomed. Opt. Express 2012;3:943–57.

[56] Shaw CB, Yalavarthy PK. Prior image-constrained l1-norm-based reconstruction method for effective usage of structural information in diffuse optical tomog- raphy. Opt. Lett. 2012;37:4353–5.

[57] Applegate MB, Istfan RE, Spink S, Tank A, Roblyer D. Recent advances in high speed diffuse optical imaging in biomedicine. APL Photonics 2020;5:040802.

[58] Bu HX, Tao R, Bai X, Zhao J. Regularized smoothed l0 norm algorithm and its application to CS-based radar imaging. Signal Process 2016;122:115–22.

[59] J. Liu, W. D. Zhou, F. H. Juwono, D. F. Huang, "Reweighted smoothed l0-norm based DOA estimation for MIMO radar", Signal Process. 137, 44–51 (2017).

[60] Eicke B. Iteration methods for convexly constrained ill-posed problems in Hilbert space. Numer. Funct. Anal. Optim. 1992;13:413–29.

[61] OpenMP, The OpenMP API specification for parallel programming, URL: https: //www.openmp.org.

[62] Bevilacqua F, Piguet D, Marquet P, Gross JD, Tromberg BJ, Depeursinge C. In vivo local determination of tissue optical prop- erties: Applications to human brain. Appl. Opt. 1999;38(22):4939–50.

[63] Li H, Ruan J, Xie Z, Wang H, Liu W. Investigation of the critical geometric char- acteristics of living human skulls utilising medical image analysis techniques. Int. J. Veh. Saf. 2007;2:345–67.