OPEN ACCESS
Journal of Innovative Optical Health Sciences
Vol. 14, No. 2 (2021) 2150002 (17 pages)

DOI: 10.1142/S1793545821500024

World Scientific
www.worldscientific.com

Inertial gradient method for fluorescence molecular tomography

Lei Wang*
College of Physics and Optoelectronic Engineering
Harbin Engineering University
145 Nantong Avenue, Harbin
150001, P. R. China
wangl@hrbeu.edu.cn

Hui Huang
School of Basic Medical Sciences
Harbin Medical University
157 Baojian Road, Harbin
150081, P. R. China
huangh81@sina.com

Received 2 September 2020
Accepted 16 November 2020
Published 9 December 2020

Image reconstruction in fluorescence molecular tomography involves seeking stable and meaningful solutions via the inversion of a highly under-determined and severely ill-posed linear mapping. An attractive scheme consists of minimizing a convex objective function that includes a quadratic error term added to a convex and nonsmooth sparsity-promoting regularizer. Choosing ℓ_1-norm as a particular case of a vast class of nonsmooth convex regularizers, our paper proposes a low per-iteration complexity gradient-based first-order optimization algorithm for the ℓ_1-regularized least squares inverse problem of image reconstruction. Our algorithm relies on a combination of two ideas applied to the nonsmooth convex objective function: Moreau–Yosida regularization and inertial dynamics-based acceleration. We also incorporate into our algorithm a gradient-based adaptive restart strategy to further enhance the practical performance. Extensive numerical experiments illustrate that in several representative test cases (covering different depths of small fluorescent inclusions, different noise levels and different separation distances between small fluorescent inclusions), our algorithm can significantly outperform three state-of-the-art algorithms in terms of CPU time taken by reconstruction, despite almost the same reconstructed images produced by each of the four algorithms.

Keywords: Biomedical imaging; image reconstruction; inverse problems; tomography.

*Corresponding author.

1. Introduction

Fluorescence molecular tomography (FMT) is a promising optical imaging modality which noninvasively probes cellular and molecular activity *in vivo* by using the fluorescent makers that interact specifically with the biomolecules of interest.[1] This technique is mainly used for small animal preclinical research in fields such as pharmacokinetics, oncology, inflammation, response to drug therapy, and infection progression.[2] Furthermore, it can also be applied to breast tumors diagnosis in humans.[3]

In FMT, a visible laser excitation light source rotating around the object illuminates the body surface at different boundary sites successively, and fluorescent makers absorb some of the excitation photons and emit fluorescence photons.[4,5] To collect fluorescence emission measurements, we perform detection by using a charged-couple device (CCD) camera placed opposite to the excitation source. We then infer the distribution of fluorescent makers from these noisy tomographic measurements.[6]

Due to the nature of the diffusion process of light when propagating through highly scattering media like tissue, the inverse problem of FMT is intrinsically ill-posed and computationally challenging, making the reconstructed images suffer from low spatial resolution.[7] To address this issue satisfactorily, some sort of regularization method can be used for tackling the ill-posedness of the inverse problem and producing meaningful and numerically stable solutions, but the resulting iterative inverse problem is computationally burdensome.[8–10] Motivated by the sparsity of the distribution of fluorescent makers,[11,12] the main stream of recent investigations of FMT inverse problem involves several types of convex nonsmooth regularizers, including total-variation (TV),[13] ℓ_1-norm,[14–16] $\ell_{2,1}$-mixed-norm,[17] and anisotropic diffusion,[18] to name a few. We are well aware that some contributions in FMT have employed nonconvex nonsmooth regularizers,[19] but in this paper we will focus mainly on convex nonsmooth regularizers such as those listed above. Unfortunately, the nonsmoothness of these regularizers makes the objective function of the inverse problem significantly more difficult to minimize.[20–22]

Among classical convex optimization methods designed to solve the nonsmooth regularized inverse problems of FMT, perhaps one of the popular candidates is the fast iterative shrinkage-thresholding algorithm (FISTA).[23] In recent years, theory and methods along this direction have been developed and have led to an explosion in the number of different algorithmic variants, including an accelerated first-order primal-dual method proposed by Chambolle and Pock (ACPM),[24] a proximal extension of optimized gradient method (POGM),[25] and an adaptive restart variant of FISTA (FISTA-R),[26] to name a few. The performance of the three state-of-art algorithms mentioned above is reported to all significantly exceed that of the plain version of FISTA.[24–26]

In stark contrast with the existing algorithms, our approach relies on a combination of three ideas applied to the convex nonsmooth objective function: Moreau–Yosida regularization,[20] inertial dynamics-based acceleration,[27] and adaptive restart scheme.[26] Our algorithm requires minimizing only one smooth function, that is, the Moreau–Yosida regularization of the original nonsmooth objective function.[20] Moreover, the first-order smooth optimization scheme is derived from the discrete time version of an inertial dynamic involving viscous damping, where the introduction of two additional terms, i.e., both Hessian-driven damping and time scale factor, provides dramatical acceleration.[27] Finally, we incorporate a heuristic adaptive momentum restart strategy that can further improve the performance of our algorithm.[26] In order to illustrate the performance of our algorithm in image reconstruction of FMT, we compare it versus the three state-of-art algorithms that are all mentioned in the preceding paragraph.[24–26] Our evaluation of reconstruction speeds for these four algorithms can help guide the selection of efficient reconstruction algorithms for researchers of FMT experimental systems when employing convex nonsmooth regularizers.

This paper is organized as follows. In Sec. 2.1, we describe the forward model of light propagation for FMT. In Sec. 2.2, we formulate the inverse problem using ℓ_1-regularized least-squares scheme, and present our algorithm. In Sec. 2.3, we give the settings of numerical experiments. Then, we present results in Sec. 3, comparing four algorithms in terms of reconstruction speed. Finally, we conclude this paper in Sec. 4.

2. Methods

2.1. *Forward problem*

We are concerned with the continuous-wave FMT, in which the forward model is described by the

steady-state coupled diffusion equations with the Robin-type boundary condition[7,28,29]:

$$-\nabla \cdot D_x(\mathbf{r})\nabla\Phi_x(\mathbf{r}) + \mu_{a_x}(\mathbf{r})\Phi_x(\mathbf{r}) = q(\mathbf{r}), \quad \mathbf{r} \in \Omega, \tag{1}$$

$$-\nabla \cdot D_m(\mathbf{r})\nabla\Phi_m(\mathbf{r}) + \mu_{a_m}(\mathbf{r})\Phi_m(\mathbf{r}) = \Phi_x(\mathbf{r})\eta\mu_{a_f}(\mathbf{r}), \quad \mathbf{r} \in \Omega, \tag{2}$$

$$\Phi_{x,m}(\mathbf{b}) + 2\zeta D_{x,m}(\mathbf{b})\hat{n} \cdot \nabla\Phi_{x,m}(\mathbf{b}) = 0, \quad \mathbf{b} \in \partial\Omega, \tag{3}$$

where the subscripts x and m denote excitation and emission wavelengths, respectively. $\hat{n}$ is the outer normal at boundary $\partial\Omega$ of domain Ω. Φ is the photon density. q is the excitation source distribution. μ_a is the absorption coefficient, and $D = (1/3)(\mu_a + \mu_s')^{-1}$ is the diffusion coefficient with μ_s' being the reduced scattering coefficient. ζ is a boundary term depending upon the refractive index mismatch at $\partial\Omega$. We let f denote the fluorescent yield $\eta\mu_{a_f}$, which incorporates the fluorophore's quantum efficiency η and its absorption coefficient μ_{a_f}.

From Eqs. (1) and (2), we write the boundary fluorescence photon density as

$$\Phi_m(\mathbf{b}) = \int_\Omega G_m(\mathbf{b};\mathbf{r}')\Phi_x(\mathbf{r}')f(\mathbf{r}')\mathrm{d}\mathbf{r}', \tag{4}$$

where $G_m(\mathbf{b};\mathbf{r}')$ is the Green function of (2) due to a point source at $\mathbf{r}'$, and can be efficiently computed based on the reciprocity theorem[7] $G_m(\mathbf{b};\mathbf{r}') = G_m(\mathbf{r}';\mathbf{b})$. The measurable existence on $\partial\Omega$ due to a source q can be defined as

$$\Gamma_{x,m}(\mathbf{b};q) = -D_{x,m}(\mathbf{b})\hat{n} \cdot \nabla\Phi_{x,m}(\mathbf{b}), \tag{5}$$

which combined with Eq. (3) gives the simpler form $\Gamma_{x,m}(\mathbf{b};q) = (1/2\zeta)\Phi_{x,m}(\mathbf{b})$.

We consider the s'th source q_s and the measurement projection of the d'th detector $\mathcal{M}_d$, leading to the following normalized Born ratio[30,31]:

$$y(s,d) = \frac{\mathcal{M}_d[\Gamma_m(\mathbf{b};q_s)]}{\mathcal{M}_d[\Gamma_x(\mathbf{b};q_s)]}, \tag{6}$$

where we divide the fluorescence measurements with corresponding excitation (*in situ*) measurements, in order to account for background heterogeneities and cancel out experimental factors such as system gains and coupling losses.

We apply the finite element method[32] (FEM) for numerically solving the forward problem Eq. (6). We use an unstructured mesh basis to discretize the diffusion equations (1) and (2), while in the inverse model the fluorescence parameter $\boldsymbol{f}$ is expanded into a solution basis on a regular grid, with N being the dimension of the basis expansion. We adopt a linear transformation for mapping between the two basis representations. Then, for a total of M source-detector pairs the forward model can be written in matrix form as

$$\boldsymbol{y} = \mathbf{A}\boldsymbol{f}, \tag{7}$$

where $\boldsymbol{f} \in \mathbb{R}^N$ is the set of representation coefficients f_i $(i = 1, \ldots, N)$ of $\boldsymbol{f}$ in the solution basis, $\boldsymbol{y} \in \mathbb{R}^M$ consists of measurements arising from all source-detector pairs, as shown in Eq. (6), and $\mathbf{A} \in \mathbb{R}^{M\times N}$ is the weight matrix.

2.2. *Inverse problem*

We approach image reconstruction in FMT as the following model-based convex optimization problem[7]:

$$\hat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \mathcal{E}(\boldsymbol{f}) := \frac{1}{2}\|\mathbf{A}\boldsymbol{f} - \boldsymbol{y}\|_2^2 + \alpha\|\boldsymbol{f}\|_1, \tag{8}$$

where $\|\cdot\|_2$ denotes the ℓ_2 norm, $\|\boldsymbol{f}\|_1 = \sum_i |f_i|$ is the ℓ_1 norm of $\boldsymbol{f}$, and $\alpha > 0$ is the regularization parameter. Letting $\mathcal{H}(\boldsymbol{f})$ and $\mathcal{R}(\boldsymbol{f})$ denote the least-squares data-fitting term $(1/2)\|\mathbf{A}\boldsymbol{f} - \boldsymbol{y}\|_2^2$ and the ℓ_1 regularizer term $\alpha\|\boldsymbol{f}\|_1$, respectively, we can rewrite Eq. (8) as

$$\hat{\boldsymbol{f}} = \arg\min_{\boldsymbol{f}} \mathcal{E}(\boldsymbol{f}) := \mathcal{H}(\boldsymbol{f}) + \mathcal{R}(\boldsymbol{f}), \tag{9}$$

where $\mathcal{H}$ is differentiable with a L-Lipschitz continuous gradient $\nabla\mathcal{H}(\cdot) = \mathbf{A}^T(\mathbf{A}\cdot - \boldsymbol{y})$. The Lipschitz constant L of $\nabla\mathcal{H}$ satisfies $L = \lambda_{\max}(\mathbf{A}^T\mathbf{A})$, with $\lambda_{\max}(\cdot)$ denoting the maximum eigenvalue.[23]

Due to its lack of smoothness, $\mathcal{R}$ is employed by the minimization algorithms via its proximal operator $\mathbf{prox}_{\delta\mathcal{R}} : \mathbb{R}^N \to \mathbb{R}^N$, where $\delta > 0$, defined by[33]

$$\mathbf{prox}_{\delta\mathcal{R}}(\boldsymbol{f}) := \arg\min_{\boldsymbol{g}} \mathcal{R}(\boldsymbol{g}) + \frac{1}{2\delta}\|\boldsymbol{f} - \boldsymbol{g}\|_2^2. \tag{10}$$

In the case $\mathcal{R}(\cdot) = \alpha\|\cdot\|_1$, we easily check from Eq. (10) that $\mathbf{prox}_{\delta\mathcal{R}}$ reduces to the soft-threshold operator $\mathcal{S}_{\delta\alpha}$, that is $[\mathcal{S}_{\delta\alpha}(\boldsymbol{f})]_i = \mathrm{sgn}(f_i)\max(|f_i| - \delta\alpha, 0)$.[34]

The presence of ℓ_1 regularizer in Eq. (8) typically encourages the solution to have relatively few nonzero components, thus promoting sparsity in reconstructed images. It is important to point out

that though this paper focuses mainly on algorithms for minimizing Eq. (9) in a particular case Eq. (8) where the ℓ_1 regularizer is used, our algorithm presented below is versatile enough to handle other more general cases of Eq. (9), in which the ℓ_1 regularizer is replaced with other interesting types of convex nonsmooth regularizers.[13,17,18]

2.2.1. *Moreau–Yosida regularization*

Recalling that the regularizer $\mathcal{R}$ is nonsmooth, we are now in position to tackle the resulting nonsmoothness of the objective function $\mathcal{E}$. Denoting the identity operator on $\mathbb{R}^N$ by $\mathbf{I}$, we let $\mathbf{K} = \delta^{-1}\mathbf{I} - \mathbf{A}^T\mathbf{A}$, which is symmetric positive definite for $0 < \delta < 1/L$. We denote $\langle \mathbf{K}\cdot|\cdot\rangle$ the scalar product of $\mathbb{R}^N$ in the metric $\mathbf{K}$ and $\|\cdot\|_{\mathbf{K}}$ the associated norm. The Moreau–Yosida regularization $\mathcal{E}_{\mathbf{K}}$ and proximal operator $\mathbf{prox}_{\mathcal{E}}^{\mathbf{K}}$ of $\mathcal{E}$ in the metric $\mathbf{K}$ are given by[20]:

$$\mathcal{E}_{\mathbf{K}}(\boldsymbol{f}) := \min_{\boldsymbol{g}} \mathcal{E}(\boldsymbol{g}) + \frac{1}{2}\|\boldsymbol{f}-\boldsymbol{g}\|_{\mathbf{K}}^2 \tag{11}$$

and

$$\mathbf{prox}_{\mathcal{E}}^{\mathbf{K}}(\boldsymbol{f}) := \arg\min_{\boldsymbol{g}} \mathcal{E}(\boldsymbol{g}) + \frac{1}{2}\|\boldsymbol{f}-\boldsymbol{g}\|_{\mathbf{K}}^2, \tag{12}$$

respectively. Obviously, $\mathbf{prox}_{\mathcal{E}}^{\mathbf{K}}$ is the solution of Eq. (11). We see that $\mathcal{E}_{\mathbf{K}}$ is a convex differentiable function, and its gradient computed in the metric $\mathbf{K}$ is written as follows[20]:

$$\nabla\mathcal{E}_{\mathbf{K}}(\boldsymbol{f}) = \boldsymbol{f} - \mathbf{prox}_{\mathcal{E}}^{\mathbf{K}}(\boldsymbol{f}), \tag{13}$$

which is 1-Lipschitz continuous in the metric $\mathbf{K}$.

An important observation is that both $\mathcal{E}$ and $\mathcal{E}_{\mathbf{K}}$ have the same set of minimizers. Thus, the problems of minimizing $\mathcal{E}$ and $\mathcal{E}_{\mathbf{K}}$ are equivalent, with the latter being a smooth optimization problem.[34] To take advantage of this attractive property, in Eq. (9) we replace $\mathcal{E}$ with its regularized and smoothed form $\mathcal{E}_{\mathbf{K}}$.

Moreover, it can be proved that $\mathbf{prox}_{\mathcal{E}}^{\mathbf{K}}$ is precisely the same as the forward–backward operator,[35] i.e.,

$$\mathbf{prox}_{\mathcal{E}}^{\mathbf{K}}(\boldsymbol{f}) = \mathbf{prox}_{\delta\mathcal{R}}[\boldsymbol{f} - \delta\nabla\mathcal{H}(\boldsymbol{f})], \tag{14}$$

whose evaluation is straightforward. Then it follows from Eq. (13) that $\nabla\mathcal{E}_{\mathbf{K}}$ can be easily evaluated.

2.2.2. *Inertial gradient algorithm*

We are now ready to minimize $\mathcal{E}_{\mathbf{K}}$. Our algorithm relies on the asymptotic behavior, as time $t \to +\infty$, of the trajectories of the following second-order continuous inertial dynamic[27]:

$$\ddot{\boldsymbol{f}}(t) + \frac{\sigma}{t}\dot{\boldsymbol{f}}(t) + \tau\nabla^2\mathcal{E}_{\mathbf{K}}[\boldsymbol{f}(t)]\dot{\boldsymbol{f}}(t) + \left(1+\frac{\tau}{t}\right)\nabla\mathcal{E}_{\mathbf{K}}[\boldsymbol{f}(t)] = 0, \tag{15}$$

with $\sigma \geq 3, \tau > 0$. This inertial system shares three characteristic features. One is the viscous damping coefficient σ/t which vanishes asymptotically as $t \to +\infty$. Another is the fact that the Hessian-driven damping is involved in the form $\nabla^2\mathcal{E}_{\mathbf{K}}[\boldsymbol{f}(t)]\dot{\boldsymbol{f}}(t)$, which is the time derivative of $\nabla\mathcal{E}_{\mathbf{K}}[\boldsymbol{f}(t)]$. A final feature is that the vanishing time scaling parameter τ/t is introduced in front of the gradient term. We say that any trajectory of $\mathcal{E}_{\mathbf{K}}[\boldsymbol{f}(t)]$ generated by Eq. (15) can converge to the minimum of $\mathcal{E}_{\mathbf{K}}$ with the asymptotic rate $\mathcal{O}(1/t^2)$. We will see that all of the three features described above are useful for acceleration of the algorithm derived below.

Now we can obtain a first-order optimization algorithm for minimizing $\mathcal{E}_{\mathbf{K}}$ from a discrete time version of the inertial dynamic given in Eq. (15). We take a fixed time step size $\gamma > 0$, and denote the Lipschitz constant of $\nabla\mathcal{E}_{\mathbf{K}}$ by L'. An admissible choice which ensures convergence is to take $\gamma \leq 1/\sqrt{L'}$, and $\tau < 2\gamma$.[27] Recall that $\nabla\mathcal{E}_{\mathbf{K}}$ is 1-Lipschitz continuous, as mentioned in Sec. 2.2.1, we therefore choose $\gamma = 1$, i.e., the largest possible time step size, and $\tau < 2$.

Denoting the iteration counter by the subscript k, an explicit centered finite-difference time discretization[35] for Eq. (15) gives

$$\begin{aligned}\boldsymbol{f}_k - 2\boldsymbol{f}_{k-1} + \boldsymbol{f}_{k-2} + \frac{\sigma}{k-1}(\boldsymbol{f}_{k-1} - \boldsymbol{f}_{k-2}) \\ + \tau[\nabla\mathcal{E}_{\mathbf{K}}(\boldsymbol{f}_{k-1}) - \nabla\mathcal{E}_{\mathbf{K}}(\boldsymbol{f}_{k-2})] \\ + \frac{\tau}{k-1}\nabla\mathcal{E}_{\mathbf{K}}(\boldsymbol{f}_{k-2}) + \nabla\mathcal{E}_{\mathbf{K}}(\boldsymbol{p}_{k-1}) = 0.\end{aligned} \tag{16}$$

Note that for the last term in the left-hand side of Eq. (16), inspired by Nesterov's extrapolation strategy,[23,36] we replace $\boldsymbol{f}_{k-2}$ with some auxiliary variable $\boldsymbol{p}_{k-1}$. Then, we choose $\boldsymbol{p}_{k-1}$ as

$$\begin{aligned}\boldsymbol{p}_{k-1} = \boldsymbol{f}_{k-1} + \left(1 - \frac{\sigma}{k-1}\right)(\boldsymbol{f}_{k-1} - \boldsymbol{f}_{k-2}) \\ - \tau[\nabla\mathcal{E}_{\mathbf{K}}(\boldsymbol{f}_{k-1}) - \nabla\mathcal{E}_{\mathbf{K}}(\boldsymbol{f}_{k-2})] \\ - \frac{\tau}{k-1}\nabla\mathcal{E}_{\mathbf{K}}(\boldsymbol{f}_{k-2}).\end{aligned} \tag{17}$$

Hence, combining Eqs. (16) and (17) leads to the following iteration steps:

$$\boldsymbol{f}_k = \boldsymbol{p}_{k-1} - \nabla\mathcal{E}_{\mathbf{K}}(\boldsymbol{p}_{k-1}), \tag{18}$$

$$\begin{aligned}\boldsymbol{p}_k = \boldsymbol{f}_k &+ \left(1-\frac{\sigma}{k}\right)(\boldsymbol{f}_k - \boldsymbol{f}_{k-1}) \\ &- \tau[\nabla\mathcal{E}_{\mathbf{K}}(\boldsymbol{f}_k) - \nabla\mathcal{E}_{\mathbf{K}}(\boldsymbol{f}_{k-1})] \\ &- \frac{\tau}{k}\nabla\mathcal{E}_{\mathbf{K}}(\boldsymbol{f}_{k-1}).\end{aligned} \tag{19}$$

We call the above iterations the inertial gradient algorithm (IGA), since they consist in a gradient descent step in Eq. (18) and an inertial momentum extrapolation step in Eq. (19). We claim that the sequence of $\mathcal{E}_{\mathbf{K}}(\boldsymbol{f}_k)$ generated by IGA can converge to the minimum of $\mathcal{E}_{\mathbf{K}}$ with a worst-case optimal rate $\mathcal{O}(1/k^2)$. Observe that when combined with Eqs. (13) and (14), the gradient descent step Eq. (18) with respect to $\mathcal{E}_{\mathbf{K}}$ becomes the following equivalent formulation:

$$\boldsymbol{f}_k = \mathbf{prox}_{\delta\mathcal{R}}[\boldsymbol{p}_{k-1} - \delta\nabla\mathcal{H}(\boldsymbol{p}_{k-1})], \tag{20}$$

which is in fact the forward–backward step[34] with respect to the original objective function $\mathcal{E}$, and is straightforward to implement.

Let us emphasize that the last three terms in the right-hand side of the inertial momentum extrapolation step Eq. (19) play a central role in acceleration of IGA. To provide a deep insight to this algorithm, we list these three terms below in detail: a momentum term $\boldsymbol{f}_k - \boldsymbol{f}_{k-1}$ of consecutive iterations with asymptotically growing momentum coefficient $1-\sigma/k$, a momentum term $\nabla\mathcal{E}_{\mathbf{K}}(\boldsymbol{f}_k) - \nabla\mathcal{E}_{\mathbf{K}}(\boldsymbol{f}_{k-1})$ of gradients at consecutive iterations with constant momentum coefficient τ and a gradient term $\nabla\mathcal{E}_{\mathbf{K}}(\boldsymbol{f}_{k-1})$ with asymptotically vanishing time scaling coefficient τ/k. We remark that these three extrapolation terms arise from the three characteristic features, mentioned above, of the inertial system Eq. (15), respectively.

2.2.3. *Adaptive restart*

Though the asymptotically growing momentum coefficient $1-\sigma/k$ in Eq. (19) results in an amount of momentum growing from one iteration to the next, which is useful for providing acceleration of IGA, an unpleasant aspect of $1-\sigma/k$ is that if the momentum is too high so that it exceeds the optimal momentum which we should apply, it can cause ripples in the trace of the objective function $\mathcal{E}$, having a severe detrimental effect on acceleration of IGA.[26]

To address this issue, we apply an adaptive restart technique to IGA in the spirit of O'Donoghue.[26] Specifically, whenever we observe that the rippling behavior occurs, we adaptively restart IGA, with the current iteration chosen as the new initial guess, and reset the momentum coefficient $1-\sigma/k$ back to its initial value $1-\sigma$. Since the accurate value of optimal momentum derived by Nesterov is generally difficult to estimate, we cannot use it for deciding when to restart. Here, following the rule introduced by O'Donoghue,[26] we employ a heuristic gradient-based restart criterion given by

$$\langle -\nabla\mathcal{E}_{\mathbf{K}}(\boldsymbol{p}_{k-1}), \boldsymbol{f}_k - \boldsymbol{f}_{k-1}\rangle < 0. \tag{21}$$

That is, we restart IGA whenever the current negative gradient $-\nabla\mathcal{E}_{\mathbf{K}}(\boldsymbol{p}_{k-1})$ and the momentum term $\boldsymbol{f}_k - \boldsymbol{f}_{k-1}$ have an obtuse angle. In other words, this criterion encourages IGA to restart whenever the momentum takes an undesirable direction. Observe that the criterion in Eq. (21) does not introduce any extra computational cost, as $\nabla\mathcal{E}_{\mathbf{K}}(\boldsymbol{p}_{k-1})$ has been already evaluated before by the gradient descent step in Eq. (18). Note that for convergence guarantee, IGA requires us to also reset the time scaling coefficient τ/k back to its initial value τ whenever activating restart.

Finally, let us combine all the ingredients presented previously in Secs. 2.2.1–2.2.3 to obtain a Regularized Inertial Gradient Algorithm with Restart (RIGA-R) for solving Eq. (9), as summarized in Algorithm 13.

Algorithm 1 RIGA-R

Set $k=1$, $j=1$, $0<\delta<1/L$, $\sigma\geq 3$, $0<\tau<2$.
Choose $\boldsymbol{f}_0\in\mathbb{R}^N$, $\boldsymbol{p}_0\in\mathbb{R}^N$, $\boldsymbol{u}_0 = \boldsymbol{f}_0 - \mathcal{S}_{\delta\alpha}[\boldsymbol{f}_0 - \delta\mathbf{A}^T(\mathbf{A}\boldsymbol{f}_0 - \boldsymbol{y})]$.
repeat
$\boldsymbol{v}_{k-1} = \boldsymbol{p}_{k-1} - \mathcal{S}_{\delta\alpha}[\boldsymbol{p}_{k-1} - \delta\mathbf{A}^T(\mathbf{A}\boldsymbol{p}_{k-1} - \boldsymbol{y})]$
$\boldsymbol{f}_k = \boldsymbol{p}_{k-1} - \boldsymbol{v}_{k-1}$
$\boldsymbol{u}_k = \boldsymbol{f}_k - \mathcal{S}_{\delta\alpha}[\boldsymbol{f}_k - \delta\mathbf{A}^T(\mathbf{A}\boldsymbol{f}_k - \boldsymbol{y})]$
if $\langle -\boldsymbol{v}_{k-1}, \boldsymbol{f}_k - \boldsymbol{f}_{k-1}\rangle < 0$ **then**
$j \leftarrow 1$
end if
$\boldsymbol{p}_k = \boldsymbol{f}_k + (1-\sigma/j)(\boldsymbol{f}_k - \boldsymbol{f}_{k-1}) - \tau(\boldsymbol{u}_k - \boldsymbol{u}_{k-1}) - (\tau/j)\boldsymbol{u}_{k-1}$
$k \leftarrow k+1$
$j \leftarrow j+1$
until stopping criterion is satisfied.

2.2.4. *Stopping criterion*

RIGA-R involves the following robust stopping criterion. Given a tolerance $\epsilon > 0$, we terminate RIGA-R when the relative variation of the objective function $\mathcal{E}$ at two consecutive iterations is small, i.e.,

$$\frac{|\mathcal{E}(\boldsymbol{f}_k) - \mathcal{E}(\boldsymbol{f}_{k-1})|}{\mathcal{E}(\boldsymbol{f}_{k-1})} \leq \epsilon. \tag{22}$$

To fairly compare the reconstruction speeds of RIGA-R and other algorithms in a way independent of the stopping rule for each algorithm, we first run RIGA-R using the stopping criterion described in Eq. (22) to set a benchmark value of the objective function $\mathcal{E}$, and then ask the other algorithms to run until each of them reaches the same benchmark.

2.3. *Numerical experiments*

We present numerical experiments comparing Algorithm 1 with the following three optimization methods: ACPM,[24] POGM[25] and FISTA-R.[26] Our choice of these three very recent state-of-the-art algorithms for comparison was not accidental. Indeed, they all significantly outperform the plain version of FISTA.[24–26] Given the already existing relative merits of ACPM, POGM and FISTA-R with respect to FISTA, here we illustrate the behavior of Algorithm 1 versus the three chosen state-of-the-art algorithms. Our experiments were executed on a desktop with an Intel Core i3-2120 CPU 3.3 GHz and 16 GB of memory and running Linux 64 bit Ubuntu 16.04 operating system. All algorithms were implemented in MATLAB R2012a.

2.3.1. *Two-dimensional numerical phantom*

We considered a circular object of diameter 25 mm. For simplicity, by choosing $\mu_{a_{x,m}} = 0.025\,\mathrm{mm}^{-1}$ and $\mu'_{s_{x,m}} = 1\,\mathrm{mm}^{-1}$, we ignored the heterogeneities and the dependence on wavelengths of the optical parameters.[37,38] Two circular fluorescent inclusions of diameter 2 mm and contrast $1\,\mathrm{mm}^{-1}$ ($0\,\mathrm{mm}^{-1}$ outside) were placed symmetrically about the x-axis. A total of 18 source sites were located on the boundary at equispaced intervals. For each source, measurements of Born ratio were detected at 37 boundary sites evenly spaced by 5° over the 180° range on the opposite side to the source. The resulting number of measurements $M = 666$. The synthetic measurements were then contaminated with the additive Gaussian noise.

We set up the inhomogeneous fluorophore distribution $\boldsymbol{f}^*$ on a 130×130 fine regular grid. $\boldsymbol{f}^* \in \mathbb{R}^{13040}$, after excluding grid points not overlapping with Ω. To generate the synthetic measurements $\boldsymbol{y}$ with the FEM forward model, we mapped $\boldsymbol{f}^*$ into a fine mesh basis consisting of 13825 nodes and 26860 linear triangles. To reduce the dimension of the solution and thus improve the ill-posedness of the inverse problem, we performed the reconstruction on a 65×65 coarse regular grid. The solution $\hat{\boldsymbol{f}} \in \mathbb{R}^{3445}$, after excluding grid points not overlapping with Ω. In the context of the inverse solver, the FEM forward model used a coarse mesh basis consisting of 3879 nodes and 7360 linear triangles.

2.3.2. *Three-dimensional numerical phantom*

We considered a cylindrical object of height 30 mm (between the planes $z = 0$ mm and $z = 30$ mm) and diameter 25 mm. We chose the same values of the background optical parameters as in the circular phantom.[37,38] Two cylindrical fluorescent inclusions of height 4 mm and diameter 2 mm and contrast $1\,\mathrm{mm}^{-1}$ ($0\,\mathrm{mm}^{-1}$ outside) were placed symmetrically about the x-axis. The centers of these two fluorescent inclusions were in the plane $z = 15$ mm. Sources and detectors were arranged in five rings ($z = -7.5, -3.75, 0, 3.75$ and 7.5 mm, respectively) around the mantle of the cylinder. In each ring, the x- and y-coordinates of the source and detector positions were the same as in the circular phantom. The resulting number of measurements $M = 16650$. The synthetic measurements were then contaminated with the additive Gaussian noise.

We set up the inhomogeneous fluorophore distribution $\boldsymbol{f}^*$ on a $50 \times 50 \times 60$ fine regular grid. $\boldsymbol{f}^* \in \mathbb{R}^{112560}$, after excluding grid points not overlapping with Ω. To generate the synthetic measurements $\boldsymbol{y}$ with the FEM forward model, we mapped $\boldsymbol{f}^*$ into a fine mesh basis consisting of 116771 nodes and 590804 linear tetrahedrons. To reduce the dimension of the solution and thus improve the ill-posedness of the inverse problem, we performed the reconstruction on a $25 \times 25 \times 30$ coarse regular grid. The solution $\hat{\boldsymbol{f}} \in \mathbb{R}^{15990}$, after excluding grid points not overlapping with Ω. In the context of the inverse solver, the FEM forward model used a coarse mesh basis consisting of 28193 nodes and 133967 linear tetrahedrons. In both of the circular and cylindrical phantoms, to assess the

quality of reconstruction, we used a vector $\tilde{\boldsymbol{f}}$ obtained by mapping $\hat{\boldsymbol{f}}$ back into the fine regular grid on which $\boldsymbol{f}^*$ was set up.

We constructed the forward modeling of light propagation with the TOAST FEM code.[32] We mention here that the construction of the weight matrix $\mathbf{A}$ with the reciprocity-based adjoint method[7] introduced in Sec. 2.1 is computationally cheap, taking around 0.5 s. Hence, we claim that even in the case of nonlinear inverse problem of diffuse optical tomography[39,40] (DOT), where the Jacobian matrix dynamically changes and thus needs to be computed repeatedly in the outer loop, this efficient adjoint method is still feasible and leads to no computational challenge.

2.3.3. *Test cases*

Let β, ξ and ρ denote the ratio of the x-coordinates of the centers of the two fluorescent inclusions to the radius of the object, the edge-to-edge distance (EED) of the two fluorescent inclusions and the noise level, respectively. To simulate different scenes of realistic FMT experiments, we designed six typical two-dimensional test cases 1–6 on the two-dimensional circular phantom, as shown in Table 1. Specifically, we chose case 1 as a baseline case, let cases 2 and 3 represent the particular cases of deep locations of fluorescent inclusions and short distance by which fluorescent inclusions were separated, respectively, and let cases 4–6 represent the particular cases of three high noise levels 5%, 15% and 25%, respectively. For all of the six 2D test cases considered, we adaptively selected the values of regularization parameter α via the L-curve method,[41] as also shown in Table 1.

Furthermore, we designed six typical 3D test cases 7–12 on the 3D cylindrical phantom, in each of which we chose the same values of β, ξ and ρ as those of each of the 2D test cases 1–6, respectively.

Table 1. Six test cases and the corresponding chosen regularization parameters α.

Test case	β	ξ (mm)	ρ	α
1	0.65	2.5	1%	2.29×10^{-7}
2	0.15	2.5	1%	6.31×10^{-7}
3	0.65	1	1%	5.01×10^{-8}
4	0.65	2.5	5%	9.15×10^{-7}
5	0.65	2.5	15%	2.66×10^{-6}
6	0.65	2.5	25%	3.94×10^{-6}

For these six 3D test cases considered, the values of regularization parameter α adaptively selected via the L-curve method[41] are 3.06×10^{-9}, 1.26×10^{-9}, 1.04×10^{-9}, 1.71×10^{-8}, 5.42×10^{-8} and 9.23×10^{-8}, respectively.

2.3.4. *Figures of merit*

We used the following two quantitative metrics for objectively evaluating the quality of the reconstructed images. One is the root mean square error (RMSE) defined as[19]

$$\mathrm{RMSE} = \frac{\|\tilde{\boldsymbol{f}} - \boldsymbol{f}^*\|_2}{\|\boldsymbol{f}^*\|_2}. \tag{23}$$

RMSE is a measure of the difference between the solution and the target. The other is the contrast-to-noise ratio (CNR) given by[28]:

$$\mathrm{CNR} = \frac{\Upsilon_{\mathrm{ROI}} - \Upsilon_{\mathrm{B}}}{(\Psi_{\mathrm{ROI}}\Theta_{\mathrm{ROI}} + \Psi_{\mathrm{B}}\Theta_{\mathrm{B}})^{1/2}}, \tag{24}$$

Table 2. Performance metrics for the comparison of the four algorithms in the 2D test cases 1–6. RIGA-R terminates when satisfying the stopping criterion in Eq. (22), while each of the other three terminates after the same iterations as required by RIGA-R to satisfy the stopping criterion in Eq. (22).

Test case	Algorithm	Time (s)	RMSE	CNR
1	RIGA-R	7.99	0.40	20.38
	ACPM	4.00	0.65	10.87
	POGM	4.11	0.70	9.62
	FISTA-R	4.15	0.71	9.31
2	RIGA-R	12.42	0.42	18.97
	ACPM	6.22	0.69	9.52
	POGM	6.39	0.72	8.75
	FISTA-R	6.45	0.72	8.53
3	RIGA-R	20.78	0.43	18.83
	ACPM	10.40	0.60	11.80
	POGM	10.69	0.62	11.35
	FISTA-R	10.79	0.63	11.04
4	RIGA-R	5.40	0.42	19.04
	ACPM	2.70	0.64	11.14
	POGM	2.78	0.68	10.04
	FISTA-R	2.80	0.70	9.35
5	RIGA-R	3.35	0.50	15.38
	ACPM	1.68	0.67	10.23
	POGM	1.72	0.72	8.82
	FISTA-R	1.74	0.74	8.34
6	RIGA-R	2.40	0.59	12.05
	ACPM	1.20	0.73	8.59
	POGM	1.23	0.77	7.52
	FISTA-R	1.25	0.78	7.28

with the region of interest (ROI) Referring to the actual fluorescent inclusions. Υ_{ROI} and Υ_{B} are the mean values, Θ_{ROI} and Θ_{B} are the variances, and Ψ_{ROI} and Ψ_{B} are the relative areas of the ROI and the background (B), respectively. A higher CNR indicates a better quality of the reconstructed image.

2.3.5. *Parameters settings*

We chose the stopping tolerance $\epsilon = 10^{-3}$, and set the initial estimates of both the solution $\boldsymbol{f}$ and the auxiliary variable $\boldsymbol{p}$ as $\boldsymbol{f} = \boldsymbol{p} = \boldsymbol{0}$, with the symbol $\boldsymbol{0}$ denoting the zero vector in $\mathbb{R}^N$. Note that the CPU time taken by each algorithm was averaged over 10 independent runs.

Recalling that the choice of the step size δ is under the constraint[27] $0 < \delta < 1/L$, we set $\delta = 0.9/L$, as it is preferable to try to use a large step size. We experimentally found that the performance of RIGA-R is not sensitive to the initial values σ and τ of the viscous damping coefficient and the time scaling parameter, respectively. With the constraint $3 \leq \sigma < \infty$ and $0 < \tau < 2$ in mind,[27] we chose 3.5 and 1.5 as their default values, respectively. We remark that to make the comparisons as fair as possible, when running ACPM, POGM and FISTA-R, we manually tuned their parameters to achieve their optimal performance.

3. Results

In all test cases, we found that to satisfy the stopping criterion in Eq. (22), RIGA-R requires the fewest iterations among the four algorithms. Hence, for comparison, ACPM, POGM and FISTA-R were run twice: they were first run for the same number of iterations as required by RIGA-R to satisfy the stopping criterion in Eq. (22); see Tables 2 and 3, respectively, for results. Second, they were run using the same stopping criterion as RIGA-R does; see Tables 4 and 5, respectively, for results.

Table 3. Performance metrics for the comparison of the four algorithms in the 3D test cases 7–12. RIGA-R terminates when satisfying the stopping criterion in Eq. (22), while each of the other three terminates after the same iterations as required by RIGA-R to satisfy the stopping criterion in Eq. (22).

Test case	Algorithm	Time (s)	RMSE	CNR
7	RIGA-R	34.96	0.31	70.26
	ACPM	17.50	0.68	26.23
	POGM	17.98	0.73	22.36
	FISTA-R	18.16	0.75	21.53
8	RIGA-R	43.32	0.38	64.09
	ACPM	21.69	0.83	17.23
	POGM	22.28	0.86	15.46
	FISTA-R	22.50	0.86	15.02
9	RIGA-R	58.79	0.35	64.36
	ACPM	29.43	0.66	27.97
	POGM	30.23	0.71	24.45
	FISTA-R	30.57	0.72	23.57
10	RIGA-R	21.08	0.33	65.80
	ACPM	10.55	0.62	30.58
	POGM	10.84	0.68	26.02
	FISTA-R	10.95	0.72	23.57
11	RIGA-R	12.96	0.38	55.96
	ACPM	6.49	0.65	27.95
	POGM	6.67	0.72	23.11
	FISTA-R	6.73	0.74	22.02
12	RIGA-R	9.77	0.41	51.53
	ACPM	4.89	0.66	27.54
	POGM	5.02	0.70	24.78
	FISTA-R	5.07	0.72	22.93

Table 4. Performance metrics for the comparison of the four algorithms in the 2D test cases 1–6. Each algorithm terminates when satisfying the stopping criterion in Eq. (22).

Test case	Algorithm	Time (s)	RMSE	CNR
1	RIGA-R	7.99	0.40	20.38
	ACPM	23.46	0.40	20.28
	POGM	29.48	0.40	20.28
	FISTA-R	31.49	0.40	20.27
2	RIGA-R	12.42	0.42	18.97
	ACPM	49.67	0.42	18.95
	POGM	63.80	0.42	18.96
	FISTA-R	69.17	0.42	18.95
3	RIGA-R	20.78	0.43	18.83
	ACPM	82.55	0.44	18.15
	POGM	96.03	0.44	18.16
	FISTA-R	107.06	0.44	18.11
4	RIGA-R	5.40	0.42	19.04
	ACPM	13.05	0.42	19.02
	POGM	15.74	0.42	19.02
	FISTA-R	17.83	0.42	19.01
5	RIGA-R	3.35	0.50	15.38
	ACPM	7.71	0.50	15.36
	POGM	10.09	0.50	15.35
	FISTA-R	11.33	0.50	15.36
6	RIGA-R	2.40	0.59	12.05
	ACPM	5.58	0.59	12.04
	POGM	7.43	0.59	12.04
	FISTA-R	8.04	0.59	12.03

Table 5. Performance metrics for the comparison of the four algorithms in the 3D test cases 7–12. Each algorithm terminates when satisfying the stopping criterion in Eq. (22).

Test case	Algorithm	Time (s)	RMSE	CNR
7	RIGA-R	34.96	0.31	70.26
	ACPM	117.29	0.31	70.25
	POGM	148.35	0.31	70.24
	FISTA-R	159.13	0.31	70.25
8	RIGA-R	43.32	0.38	64.09
	ACPM	156.28	0.38	64.03
	POGM	195.61	0.38	64.07
	FISTA-R	209.54	0.38	64.06
9	RIGA-R	58.79	0.35	64.36
	ACPM	207.50	0.35	64.34
	POGM	260.15	0.35	64.34
	FISTA-R	279.01	0.35	64.35
10	RIGA-R	21.08	0.33	65.80
	ACPM	59.46	0.33	65.81
	POGM	71.73	0.33	65.80
	FISTA-R	81.37	0.33	65.82
11	RIGA-R	12.96	0.38	55.96
	ACPM	31.68	0.38	55.94
	POGM	40.71	0.38	55.95
	FISTA-R	44.13	0.38	55.94
12	RIGA-R	9.77	0.41	51.53
	ACPM	23.15	0.41	51.54
	POGM	27.30	0.41	51.52
	FISTA-R	30.53	0.41	51.52

For 2D test cases 1−6, we show the reconstruction results in Figs. 1–6, respectively, which correspond to the quantitative analysis results listed in Tables 2 and 4. For 3D test cases 7–12, we show the reconstruction results in Figs. 7–12, respectively, which correspond to the quantitative analysis results listed in Tables 3 and 5. Based on these results, we chose the 2D test case 3 as an example, and illustrate the relative performance of the four algorithms in terms of CPU time in two aspects as follows.

First, we observe that after the same iterations as required by RIGA-R to satisfy the stopping criterion in Eq. (22), only RIGA-R can successfully resolve the two fluorescent targets with an EED of 2 mm [shown in Fig. 3(e)], whereas ACPM, POGM and FISTA-R fail entirely to do that [shown in Figs. 3(b)–3(d)], as well as lead to much higher RMSE and lower CNR. Recall the fact that for each of ACPM, POGM and FISTA-R the main per-iteration cost is roughly one matrix-vector multiplication involving $\mathbf{A}$, and one involving $\mathbf{A}^T$, while for RIGA-R the main per-iteration cost is roughly two matrix-vector multiplications involving $\mathbf{A}$, and two involving $\mathbf{A}^T$. Hence, for the same iterations, RIGA-R is almost twice slower than each of ACPM, POGM and FISTA-R.

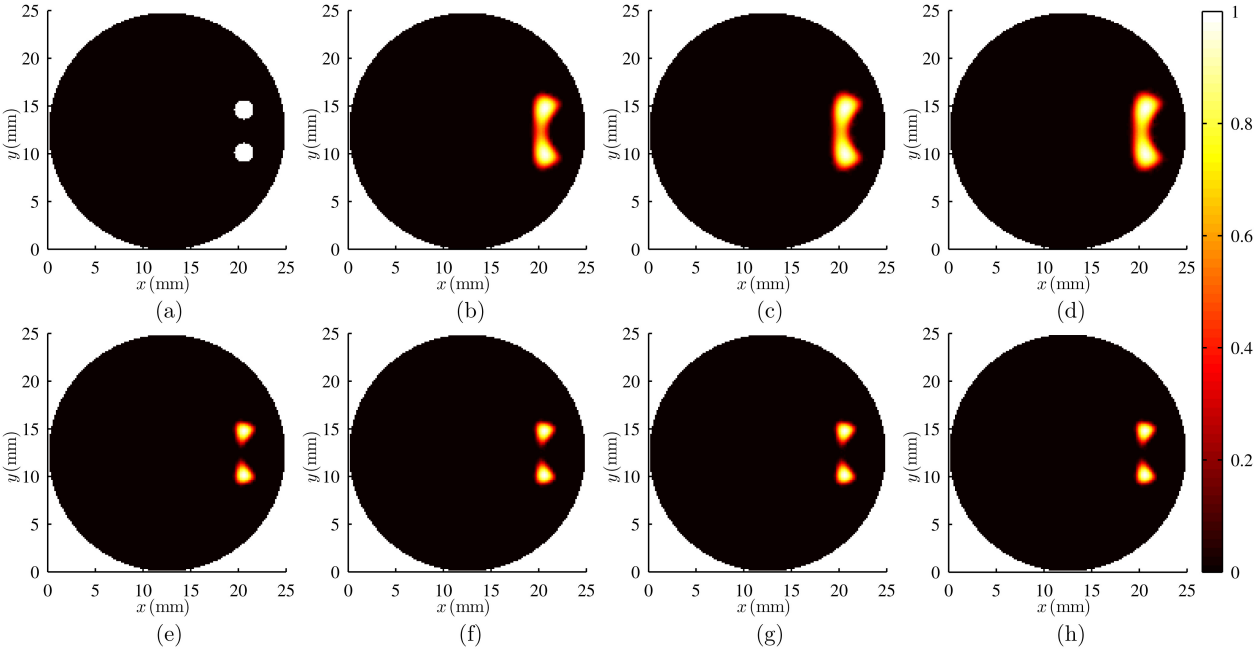

Fig. 1. Reconstructed images in 2D test case 1 using (e) RIGA-R, (b) and (f) ACPM, (c) and (g) POGM and (d) and (h) FISTA-R. (e)–(h): Each algorithm terminates when satisfying the stopping criterion in Eq. (22). (b)–(d): Each algorithm terminates after the same iterations as required by RIGA-R to satisfy the stopping criterion in Eq. (22). (a): The fluorophore inclusions. Each reconstructed image is normalized by its largest pixel-value.

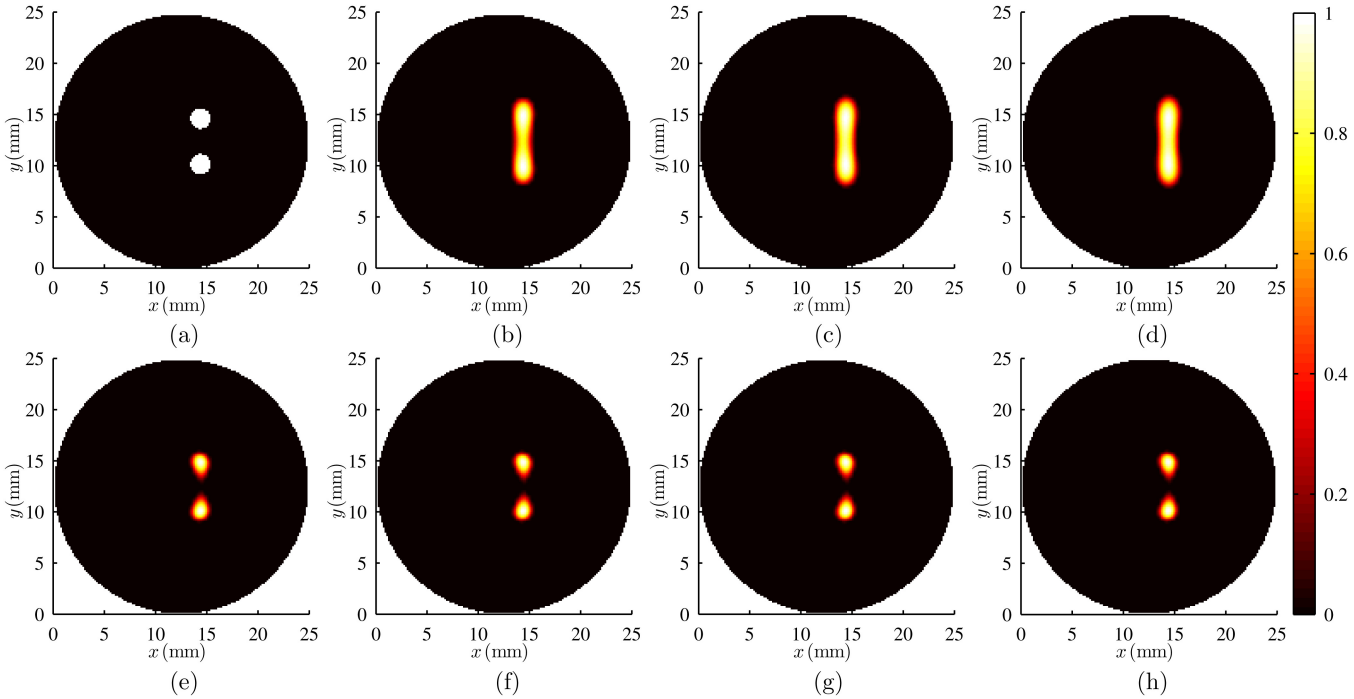

Fig. 2. Reconstructed images in 2D test case 2 using (e) RIGA-R, (b) and (f) ACPM, (c) and (g) POGM and (d) and (h) FISTA-R. (e)–(h): Each algorithm terminates when satisfying the stopping criterion in Eq. (22). (b)–(d): Each algorithm terminates after the same iterations as required by RIGA-R to satisfy the stopping criterion in Eq. (22). (a): The fluorophore inclusions. Each reconstructed image is normalized by its largest pixel-value.

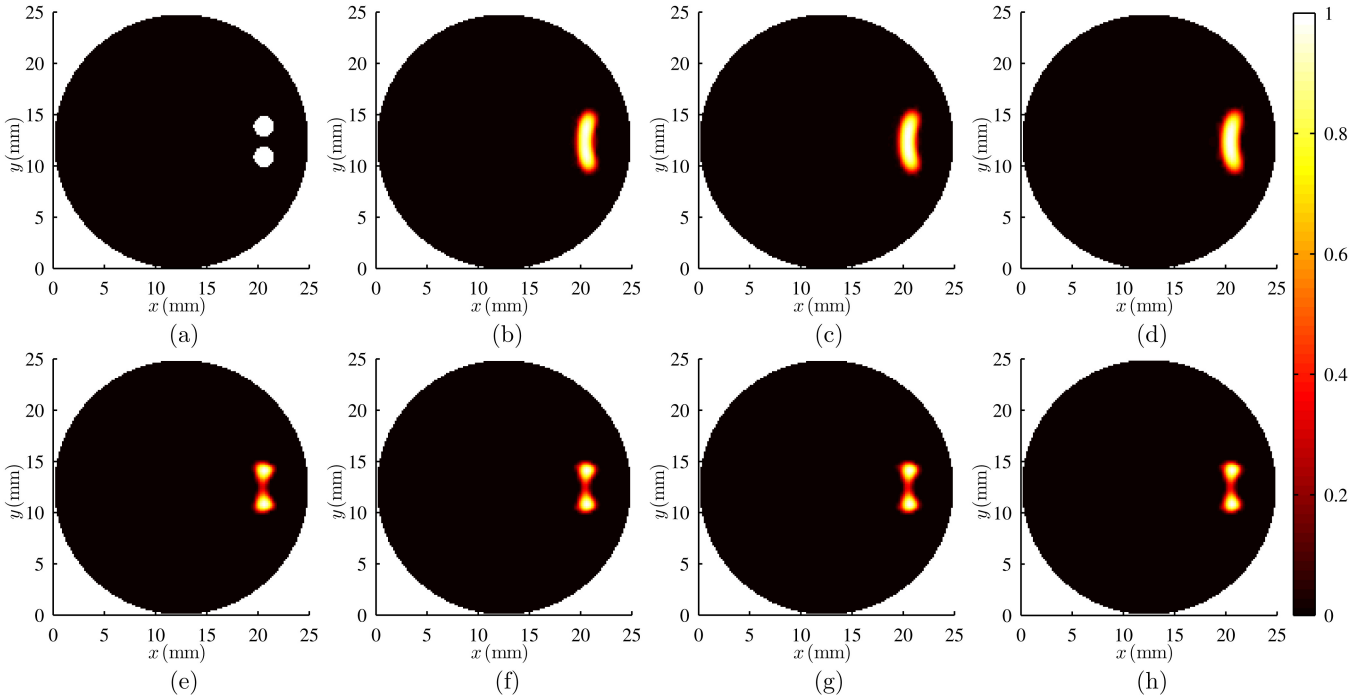

Fig. 3. Reconstructed images in 2D test case 3 using (e) RIGA-R, (b) and (f) ACPM, (c) and (g) POGM and (d) and (h) FISTA-R. (e)–(h): Each algorithm terminates when satisfying the stopping criterion in Eq. (22). (b)–(d): Each algorithm terminates after the same iterations as required by RIGA-R to satisfy the stopping criterion in Eq. (22). (a): The fluorophore inclusions. Each reconstructed image is normalized by its largest pixel-value.

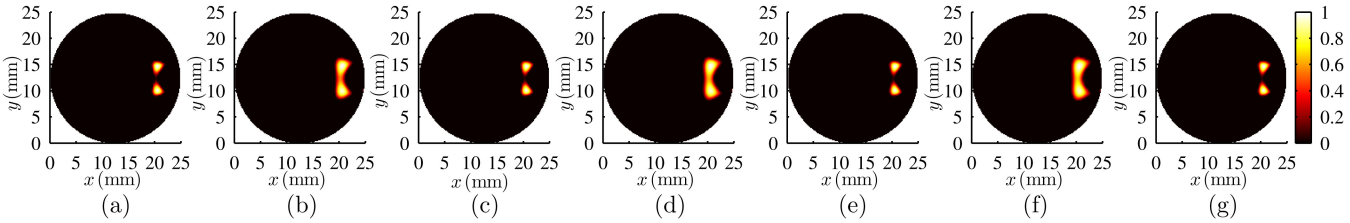

Fig. 4. Reconstructed images in 2D test case 4 using (a) RIGA-R, (b) and (c) ACPM, (d) and (e) POGM and (f) and (g) FISTA-R. (a),(c),(e) and (g): Each algorithm terminates when satisfying the stopping criterion in Eq. (22). (b),(d) and (f): Each algorithm terminates after the same iterations as required by RIGA-R to satisfy the stopping criterion in Eq. (22). The fluorophore inclusions are shown in Fig. 1(a). Each reconstructed image is normalized by its largest pixel-value.

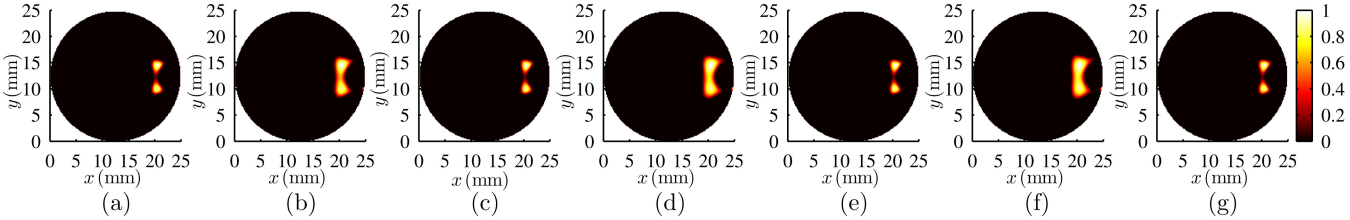

Fig. 5. Reconstructed images in 2D test case 5 using (a) RIGA-R, (b) and (c) ACPM, (d) and (e) POGM and (f) and (g) FISTA-R. (a),(c),(e) and (g): Each algorithm terminates when satisfying the stopping criterion in Eq. (22). (b),(d) and (f): Each algorithm terminates after the same iterations as required by RIGA-R to satisfy the stopping criterion in Eq. (22). The fluorophore inclusions are shown in Fig. 1(a). Each reconstructed image is normalized by its largest pixel-value.

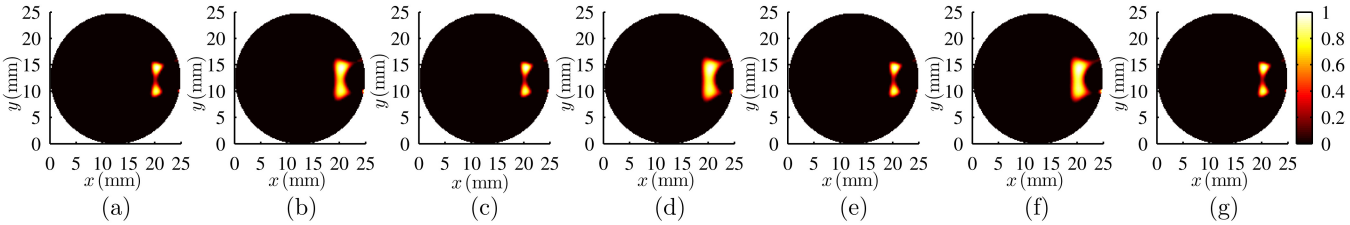

Fig. 6. Reconstructed images in 2D test case 6 using (a) RIGA-R, (b) and (c) ACPM, (d) and (e) POGM and (f) and (g) FISTA-R. (a),(c),(e) and (g): Each algorithm terminates when satisfying the stopping criterion in Eq. (22). (b),(d) and (f): Each algorithm terminates after the same iterations as required by RIGA-R to satisfy the stopping criterion in Eq. (22). The fluorophore inclusions are shown in Fig. 1(a). Each reconstructed image is normalized by its largest pixel-value.

Next, we see that after more iterations until the same stopping criterion in Eq. (22) as used by RIGA-R is also satisfied by ACPM, POGM and FISTA-R, they can produce Figs. 3(f)–3(h), respectively, all of which are visually undistinguishable from Fig. 3(e) produced by RIGA-R, as well as have almost the same values of RMSE and CNR as those of Fig. 3(e). However, RIGA-R is around 3.97, 4.62, and 5.15 times faster than ACPM, POGM and FISTA-R, respectively.

In addition, for the other five 2D test cases and all of the six 3D test cases, let us observe that the relative merits of RIGA-R with respect to ACPM, POGM and FISTA-R, similar to test case 3 illustrated above, also exist. Specifically, in all of the 12 test cases considered, on the one hand, after the same iterations as required by RIGA-R to satisfy the stopping criterion in Eq. (22), RIGA-R can reconstruct images with significantly better quality than ACPM, POGM and FISTA-R do, and on the other hand, when reconstructing images with almost the same visual effect and quantitative metrics, in 2D test cases RIGA-R is faster than ACPM, POGM and FISTA-R by 2.30 to 4.00 times, 2.91 to 5.14 times and 3.30 to 5.57 times, respectively, while in 3D test cases RIGA-R is faster than ACPM, POGM and FISTA-R by 2.37 to 3.61 times, 2.79 to 4.52 times and 3.12 to 4.84 times, respectively.

Finally, it may be interesting to point out that besides the case of FMT concerned here, the

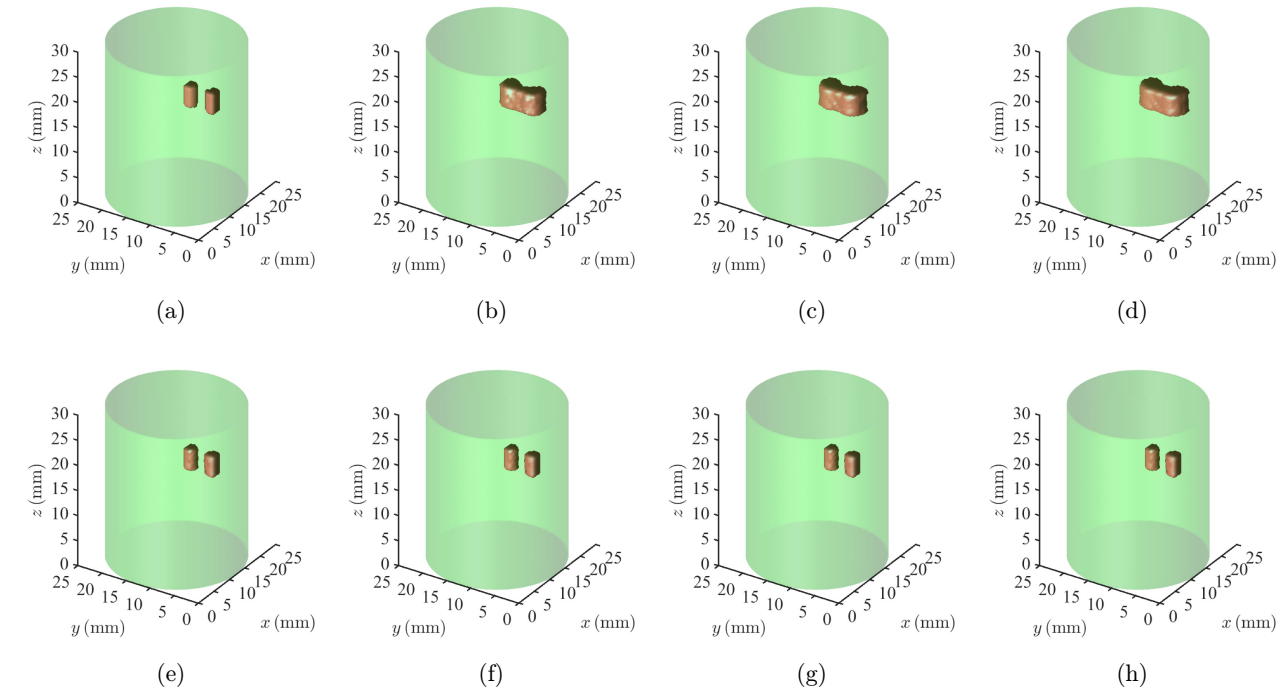

Fig. 7. Reconstructed images in 3D test case 7 using (e) RIGA-R, (b) and (f) ACPM, (c) and (g) POGM and (d) and (h) FISTA-R. (e),(f),(g) and (h): Each algorithm terminates when satisfying the stopping criterion in Eq. (22). (b), (c) and (d): Each algorithm terminates after the same iterations as required by RIGA-R to satisfy the stopping criterion in Eq. (22). (a): The fluorophore inclusions. Each reconstructed image is normalized by its largest pixel-value.

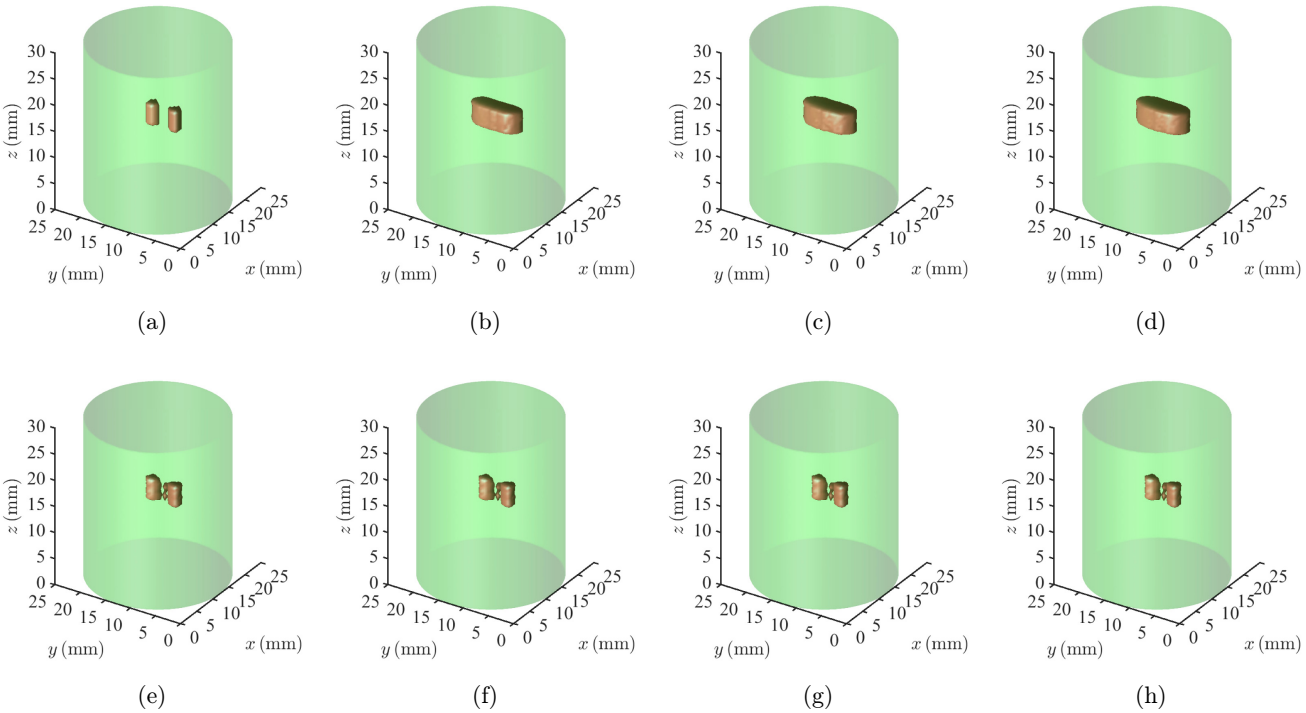

Fig. 8. Reconstructed images in three-dimensional test case 8 using (e) RIGA-R, (b) and (f) ACPM, (c) and (g) POGM and (d) and (h) FISTA-R. (e)–(h): Each algorithm terminates when satisfying the stopping criterion in Eq. (22). (b),(c) and (d): Each algorithm terminates after the same iterations as required by RIGA-R to satisfy the stopping criterion in Eq. (22). (a): The fluorophore inclusions. Each reconstructed image is normalized by its largest pixel-value.

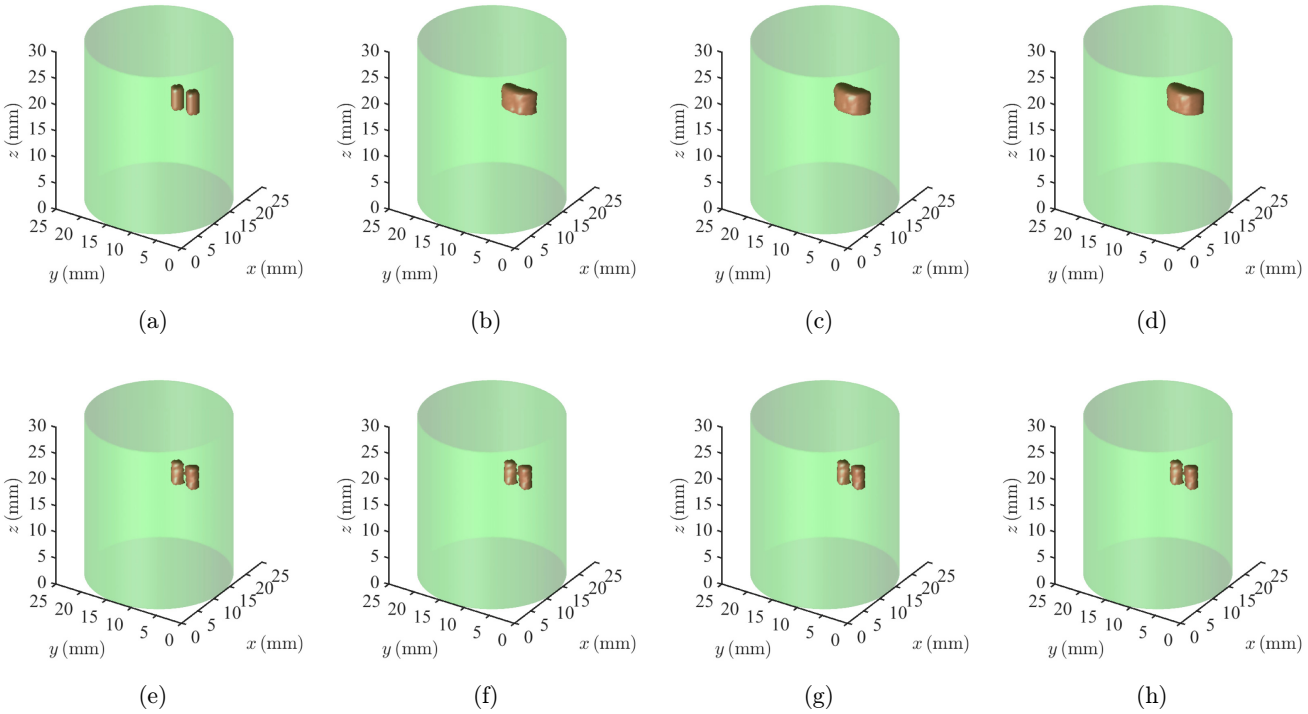

Fig. 9. Reconstructed images in three-dimensional test case 9 using (e) RIGA-R, (b) and (f) ACPM, (c) and (g) POGM and (d) and (h) FISTA-R. (e)–(h): Each algorithm terminates when satisfying the stopping criterion in Eq. (22). (b), (c) and (d): Each algorithm terminates after the same iterations as required by RIGA-R to satisfy the stopping criterion in Eq. (22). (a): The fluorophore inclusions. Each reconstructed image is normalized by its largest pixel-value.

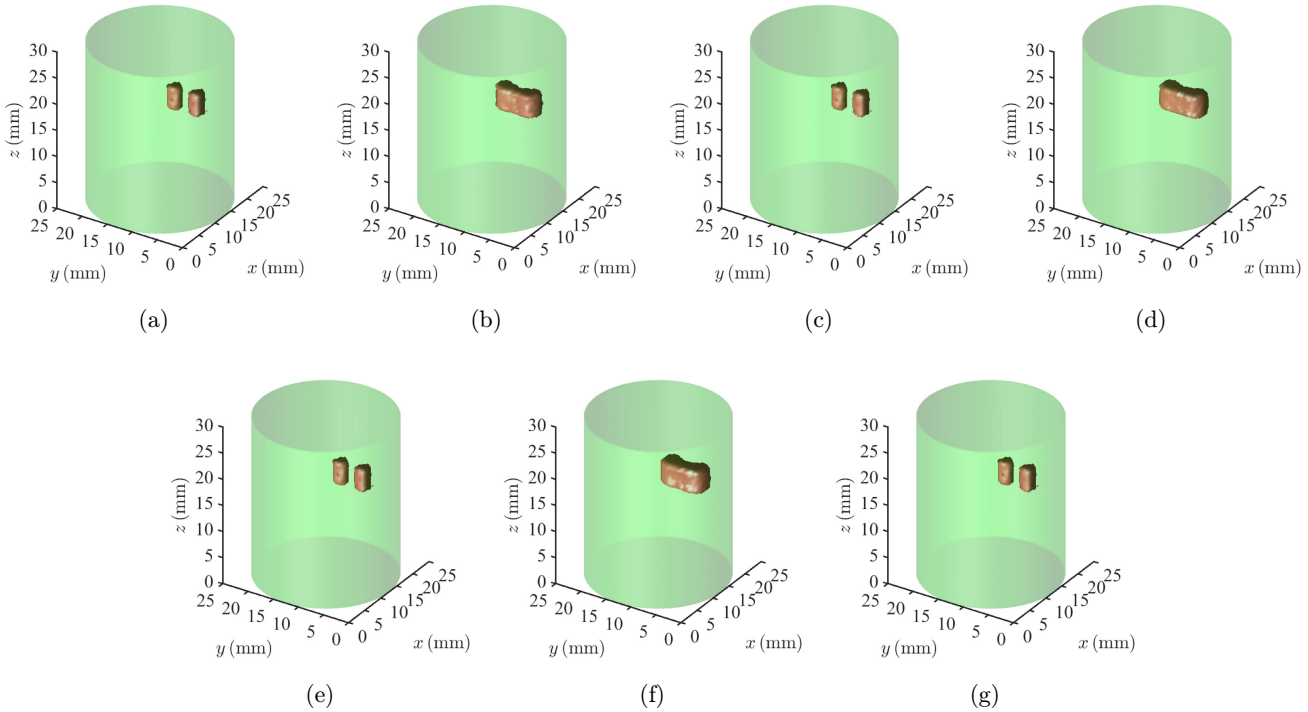

Fig. 10. Reconstructed images in three-dimensional test case 10 using (a) RIGA-R, (b) and (c) ACPM, (d) and (e) POGM and (f) and (g) FISTA-R. (a),(c),(e) and (g): Each algorithm terminates when satisfying the stopping criterion in Eq. (22). (b),(d) and (f): Each algorithm terminates after the same iterations as required by RIGA-R to satisfy the stopping criterion in Eq. (22). The fluorophore inclusions are shown in Fig. 7(a). Each reconstructed image is normalized by its largest pixel-value.

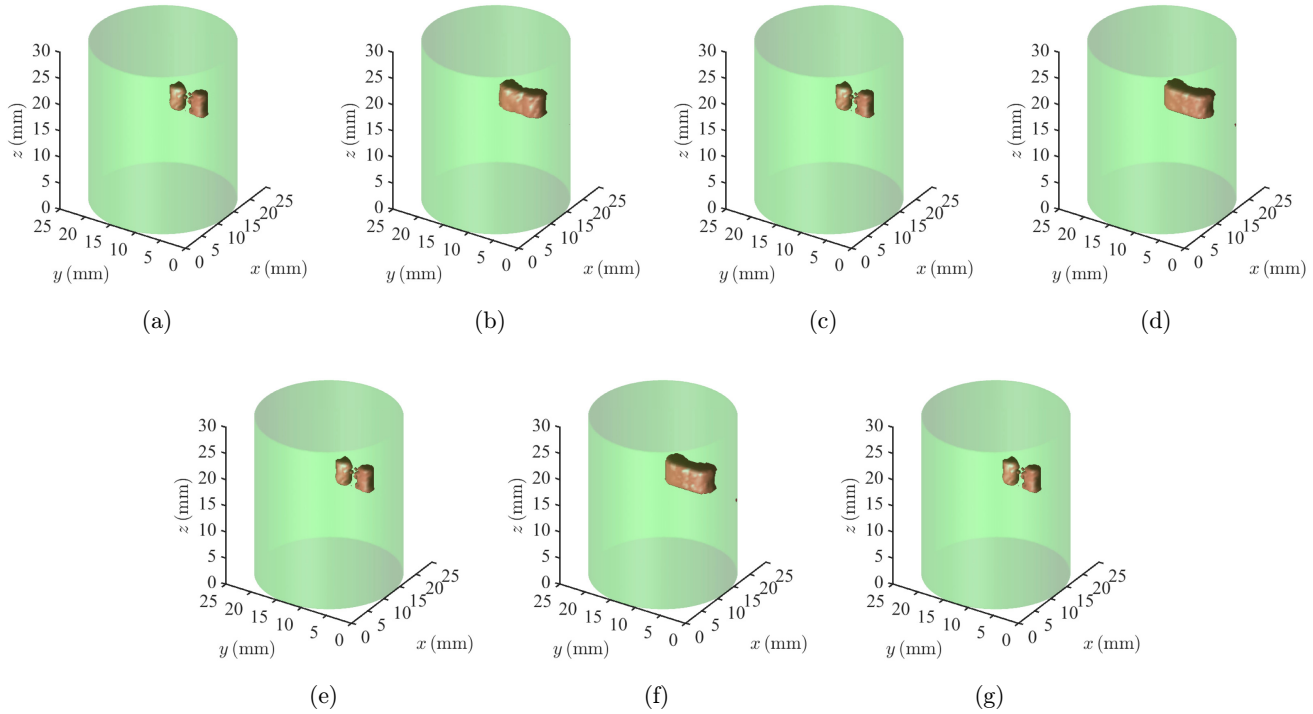

Fig. 11. Reconstructed images in three-dimensional test case 11 using (a) RIGA-R, (b) and (c) ACPM, (d) and (e) POGM and (f) and (g) FISTA-R. (a),(c),(e) and (g): Each algorithm terminates when satisfying the stopping criterion in Eq. (22). (b), (d) and (f): Each algorithm terminates after the same iterations as required by RIGA-R to satisfy the stopping criterion in Eq. (22). The fluorophore inclusions are shown in Fig. 7(a). Each reconstructed image is normalized by its largest pixel-value.

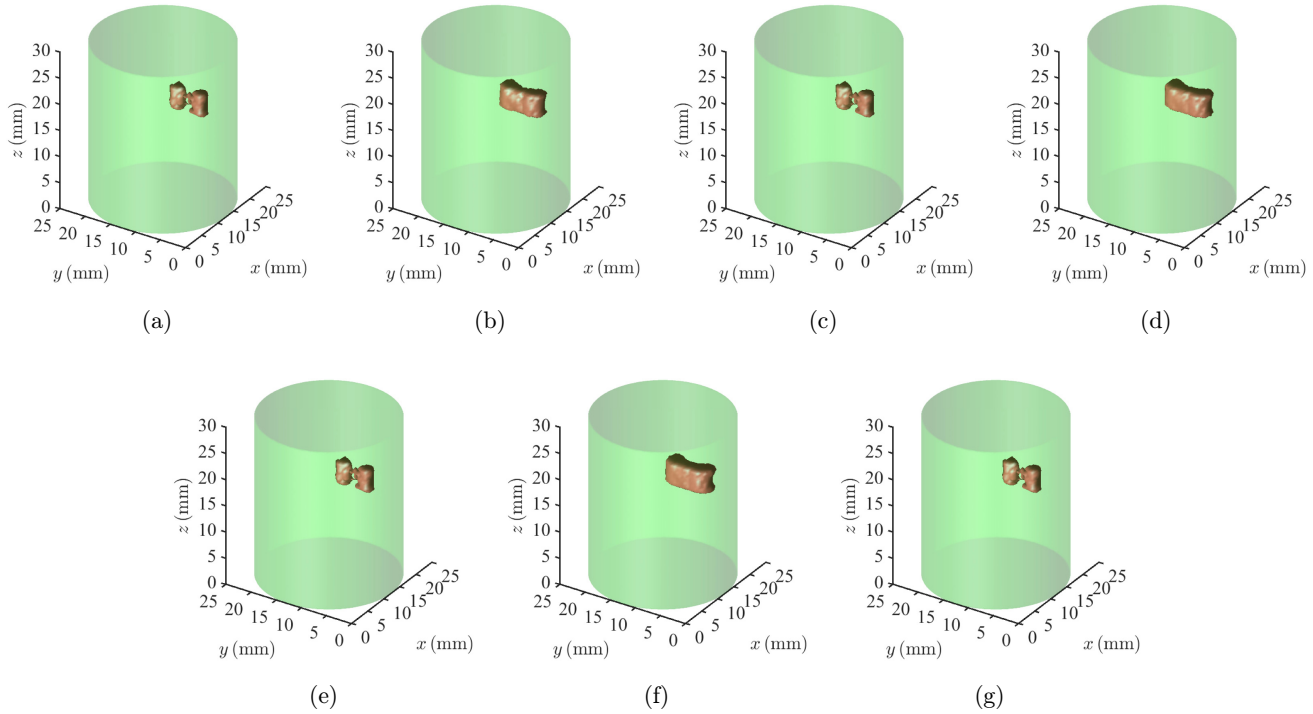

Fig. 12. Reconstructed images in three-dimensional test case 12 using (a) RIGA-R, (b) and (c) ACPM, (d) and (e) POGM and (f) and (g) FISTA-R. (a),(c),(e) and (g): Each algorithm terminates when satisfying the stopping criterion in Eq. (22). (b),(d) and (f): Each algorithm terminates after the same iterations as required by RIGA-R to satisfy the stopping criterion in Eq. (22). The fluorophore inclusions are shown in Fig. 7(a). Each reconstructed image is normalized by its largest pixel-value.

application of RIGA-R developed in this paper can be straightforwardly extended to the cases of other types of optical tomography modalities, such as bioluminescence tomography[42–44] (BLT), photoacoustic tomography[45] (PAT) and linearized DOT.[39,40]

4. Conclusion

We have proposed RIGA-R, i.e., a regularized inertial gradient algorithm equipped with an adaptive restart strategy, for solving a class of convex nonsmooth minimization problems arising in image reconstruction of FMT. In extensive numerical experiments, we compared reconstruction speed of RIGA-R versus those of three state-of-the-art optimization methods, including ACPM, POGM and FISTA-R. RIGA-R is easy to implement and works well across multiple representative test cases in which deep locations of fluorescent inclusions, short distance by which fluorescent inclusions were separated, and high noise level were taken into account. We found that though generating almost the same reconstructed images, RIGA-R is significantly faster than ACPM, POGM and FISTA-R, especially in the cases of very deep locations of small fluorescent inclusions (test case 2) and very short separation distance between small fluorescent inclusions (test case 3). Our numerical experiments for applications in FMT evidence the outperformance of the proposed Moreau–Yosida regularization-based restart inertial gradient algorithm over the three chosen state-of-the-art methods. Besides the ℓ_1-norm concerned here, RIGA-R can be straightforwardly extended to incorporate other nonsmooth convex regularizers usually used by reconstruction in FMT, and is applicable to other optical tomography modalities such as BLT, PAT and DOT.

Conflicts of Interest

The authors have no relevant conflicts of interest to disclose.

Acknowledgment

This work has been supported by funding from the Fundamental Research Funds for the Central Universities under Grant Agreement No. GK2250260033.

References

1. C. Darne, Y. Lu, E. M. Sevick-Muraca, "Small animal fluorescence and bioluminescence tomography: A review of approaches, algorithms and technology update," *Phys. Med. Biol.* **59**, R1 (2014).
2. V. Ntziachristos, "Fluorescence molecular imaging," *Annu. Rev. Biomed. Eng.* **8**, 1–33 (2006).
3. S. C. Davis, H. Dehghani, J. Wang, S. Jiang, B. W. Pogue, K. D. Paulsen, "Image-guided diffuse optical fluorescence tomography implemented with laplacian-type regularization," *Opt. Express* **15**, 4066–4082 (2007).
4. K. Liu, X. Jiang, Y. Deng, "First-order approximation of fluorescence excitation-transmission to accelerate fluorescence molecular tomography image reconstruction," *Opt. Lett.* **44**(13), 3222–3225 (2019).
5. Y. He, X. Cao, F. Liu, J. Luo, J. Bai, "Influence of limited-projection on fluorescence molecular tomography," *J. Innov. Opt. Health Sci.* **5**, 1250020 (2012).
6. A. B. Milstein, S. Oh, K. J. Webb, C. A. Bouman, Q. Zhang, D. A. Boas, R. Millane, "Fluorescence optical diffusion tomography," *Appl. Opt.* **42**, 3081–3094 (2003).
7. S. R. Arridge, J. C. Schotland, "Optical tomography: forward and inverse problems," *Inverse Probl.* **25**, 123010 (2009).
8. D. Zhu, C. Li, "Accelerated image reconstruction in fluorescence molecular tomography using a nonuniform updating scheme with momentum and ordered subsets methods," *J. Biomed. Opt.* **21**, 016004 (2016).
9. H. Guo, J. Yu, X. He, Y. Hou, F. Dong, S. Zhang, "Improved sparse reconstruction for fluorescence molecular tomography with $\ell_{1/2}$ regularization," *Biomed. Opt. Express* **6**, 1648–1664 (2015).
10. W. Zou, J. Wang, D. D. Feng, "Image reconstruction of fluorescent molecular tomography based on the simplified matrix system," *J. Opt. Soc. Am. A* **30**(8), 1464–1475 (2013).
11. J. Shi, X. Cao, F. Liu, B. Zhang, J. Luo, J. Bai, "Greedy reconstruction algorithm for fluorescence molecular tomography by means of truncated singular value decomposition conversion," *J. Opt. Soc. Am. A* **30**(3), 437–447 (2013).
12. J. Yu, J. Cheng, Y. Hou, X. He, "Sparse reconstruction for fluorescence molecular tomography via a fast iterative algorithm," *J. Innov. Opt. Health Sci.* **7**, 1450008 (2014).

13. A. Behrooz, H.-M. Zhou, A. A. Eftekhar, A. Adibi, "Total variation regularization for 3d reconstruction in fluorescence tomography: experimental phantom studies," *Appl. Opt.* **51**, 8216–8227 (2012).
14. W. Xie, Y. Deng, D. Yan, X. Yang, Q. Luo, "Sparsity-promoting Bayesian approximation error method for compensating for the mismodeling of optical properties in fluorescence molecular tomography," *Opt. Lett.* **42**, 3024–3027 (2017).
15. Z. Wu, X. Wang, J. Yu, H. Yi, X. He, "Synchronization-based clustering algorithm for reconstruction of multiple reconstructed targets in fluorescence molecular tomography," *J. Opt. Soc. Am. A* **35**(2), 328–335 (2018).
16. H. Yi, P. Jiao, X. Li, J. Peng, X. He, "Three-way decision based reconstruction frame for fluorescence molecular tomography," *J. Opt. Soc. Am. A* **35**(11), 1814–1822 (2018).
17. S. Jiang, J. Liu, Y. An, G. Zhang, J. Ye, Y. Mao, K. He, C. Chi, J. Tian, "Novel $\ell_{2,1}$-norm optimization method for fluorescence molecular tomography reconstruction," *Biomed. Opt. Express* **7**, 2342–2359 (2016).
18. T. Correia, M. Koch, A. Ale, V. Ntziachristos, S. Arridge, "Patch-based anisotropic diffusion scheme for fluorescence diffuse optical tomography-part 2: Image reconstruction," *Phys. Med. Biol.* **61**, 1452 (2016).
19. L. Zhao, H. Yang, W. Cong, G. Wang, X. Intes, "ℓ_{p} regularization for early gate fluorescence molecular tomography," *Opt. Lett.* **39**, 4156–4159 (2014).
20. A. Chambolle, T. Pock, "An introduction to continuous optimization for imaging," *Acta Numer.* **25**, 161–319 (2016).
21. J. Ye, Y. Du, Y. An, Y. Mao, S. Jiang, W. Shang, K. He, X. Yang, K. Wang, C. Chi *et al.*, "Sparse reconstruction of fluorescence molecular tomography using variable splitting and alternating direction scheme," *Mol. Imaging. Biol.* **20**, 37–46 (2018).
22. X. He, H. Guo, J. Yu, X. Zhang, Y. Hou, "Effective and robust approach for fluorescence molecular tomography based on cosamp and SP_3 model," *J. Innov. Opt. Health Sci.* **9**, 1650024 (2016).
23. A. Beck, M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.* **2**, 183–202 (2009).
24. A. Chambolle, T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imaging Vis.* **40**, 120–145 (2011).
25. A. B. Taylor, J. M. Hendrickx, F. Glineur, "Exact worst-case performance of first-order methods for composite convex optimization," *SIAM J. Optim.* **27**, 1283–1313 (2017).
26. B. O'Donoghue, E. Candes, "Adaptive restart for accelerated gradient schemes," *Found. Comput. Math.* **15**, 715–732 (2015).
27. H. Attouch, Z. Chbani, J. Fadili, H. Riahi, "First-order optimization algorithms via inertial systems with hessian driven damping," arXiv:1907.10536.
28. J.-C. Baritaux, K. Hassler, M. Bucher, S. Sanyal, M. Unser, "Sparsity-driven reconstruction for fdot with anatomical priors," *IEEE Trans. Med. Imaging* **30**, 1143–1153 (2011).
29. D. Wang, X. Liu, Y. Chen, J. Bai, "A novel finite-element-based algorithm for fluorescence molecular tomography of heterogeneous media," *IEEE Trans. Inf. Technol. Biomed.* **13**(5), 766–773 (2009).
30. V. Ntziachristos, R. Weissleder, "Experimental three-dimensional fluorescence reconstruction of diffuse media by use of a normalized born approximation," *Opt. Lett.* **26**, 893–895 (2001).
31. L. Lian, Y. Deng, W. Xie, G. Xu, X. Yang, Z. Zhang, Q. Luo, "Enhancement of the localization and quantitative performance of fluorescence molecular tomography by using linear nborn method," *Opt. Express* **25**(3), 2063–2079 (2017).
32. M. Schweiger, S. R. Arridge, "The toast++ software suite for forward and inverse modeling in optical tomography," *J. Biomed. Opt.* **19**, 040801 (2014).
33. P. L. Combettes, J.-C. Pesquet, Proximal splitting methods in signal processing, in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, H. Wolkowicz, Eds., pp. 185–212, Springer, New York (2011).
34. P. L. Combettes, V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Model. Simul.* **4**, 1168–1200 (2005).
35. H. Attouch, A. Cabot, "Convergence of a relaxed inertial forward–backward algorithm for structured monotone inclusions," *Appl. Math. Optim.* **80**, 547–598 (2019).
36. H. Attouch, J. Peypouquet, "Convergence of inertial dynamics and proximal algorithms governed by maximally monotone operators," *Math. Program.* **174**, 391–432 (2019).
37. S. R. Arridge, M. Schweiger, "A gradient-based optimisation scheme for optical tomography," *Opt. Express* **2**, 213–226 (1998).
38. M. Schweiger, S. R. Arridge, I. Nissila, "Gauss-newton method for image reconstruction in diffuse optical tomography," *Phys. Med. Biol.* **50**, 2365–2386 (2005).
39. R. P. K. Jagannath, P. K. Yalavarthy, "Nonquadratic penalization improves near-infrared diffuse optical tomography," *J. Opt. Soc. Am. A* **30**(8), 1516–1523 (2013).
40. C. B. Shaw, P. K. Yalavarthy, "Performance evaluation of typical approximation algorithms for nonconvex ℓ_{p}-minimization in diffuse optical tomography," *J. Opt. Soc. Am. A* **31**(4), 852–862 (2014).

41. P. C. Hansen, D. P. O'Leary, "The use of the L-curve in the regularization of discrete ill-posed problems," *SIAM J. Sci. Comput.* **14**, 1487–1503 (1993).
42. A. Bentley, J. E. Rowe, H. Dehghani, "Single pixel hyperspectral bioluminescence tomography based on compressive sensing," *Biomed. Opt. Express* **10** (11), 5549–5564 (2019).
43. K. Liu, X. Yang, D. Liu, C. Qin, J. Liu, Z. Chang, M. Xu, J. Tian, "Spectrally resolved three-dimensional bioluminescence tomography with a level-set strategy," *J. Opt. Soc. Am. A* **27**(6), 1413–1423 (2010).
44. J. Yu, Q. Li, H. Wang, "Source reconstruction for bioluminescence tomography via $\ell_{1/2}$ regularization," *J. Innov. Opt. Health Sci.* **11**, 1750014 (2018).
45. M. Bhatt, S. Gutta, P. K. Yalavarthy, "Exponential filtering of singular values improves photoacoustic image reconstruction," *J. Opt. Soc. Am. A* **33**(9), 1785–1792 (2016).