# Multilevel bioluminescence tomography based on radiative transfer equation Part 2: total variation and 11 data fidelity

## Hao Gao and Hongkai Zhao

<sup>1</sup>Department of Mathematics, University of California, Irvine, CA 92697, USA haog@uci.edu, zhao@math.uci.edu

**Abstract:** In this paper we study the regularization with both 11 and totalvariation norm for bioluminescence tomography based on radiative transfer equation, compare 11 data fidelity with 12 data fidelity for different type of noise, and propose novel interior-point methods for solving related optimization problems. Simulations are performed to show that our approach is not only capable of preserving shapes, details and intensities of bioluminescent sources in the presence of sparse or non-sparse sources with angular-resolved or angular-averaged data, but also robust to noise, and thus is potential for efficient high-resolution imaging with only boundary data.

©2009 Optical Society of America

**OCIS codes:** (100.3190) Inverse problems; (110.6960) Tomography; (170.3010) Image reconstruction techniques; (170.6280) Spectroscopy, fluorescence and luminescence.

#### **References and links**

- C. H. Contag and B. D. Ross, "It's not just about anatomy: in vivo bioluminescence imaging as an eyepiece into biology," Journal of Magnetic Resonance Imaging 16, 378-387 (2002).
- 2. G. Wang, E. A. Hoffman, G. McLennan, L. V. Wang, M. Suter, and J. Meinel, "Development of the first bioluminescent CT scanner," Radiology 566 (2003).
- G. Wang, Y. Li, and M. Jiang, "Uniqueness theorems in bioluminescence tomography," Med. Phys. 31, 2289-2299 (2004).
- 4. A. R. Arridge, "Optical tomography in medical imaging," Inverse Problems 15, R41-93 (1999).
- W. Cong, G. Wang, D. Kumar, Y. Liu, M. Jiang, L. Wang, E. Hoffman, G. McLennan, P. McCray, J. Zabner, and A. Cong, "Practical reconstruction method for bioluminescence tomography," Opt. Express 13, 6756-6771 (2005).
- 6. H. Dehghani, S. C. Davis, S. Jiang, B. W. Pogue, K. D. Paulsen, and M. S. Patterson, "Spectrally resolved bioluminescence optical tomography," Opt. Lett. **31**, 365-367 (2006).
- A. J. Chaudhari, F. Darvas, J. R. Bading, R. A. Moats, P. S. Conti, D. J. Smith, S. R. Cherry, and R. M. Leahy, "Hyperspectral and multispectral bioluminescence optical tomography for small animal imaging," Phys. Med. Biol. 50, 5421-5441 (2005).
- X. Gu, Q. Zhang, L. Larcom, and H. Jiang, "Three-dimensional bioluminescence tomography with modelbased reconstruction," Opt. Express 12, 3996-4000 (2004).
- G. Alexandrakis, F. R. Rannou, and A. F. Chatziioannou, "Tomographic bioluminescence imaging by use of a combined optical-PET (OPET) system: a computer simulation feasibility study," Phys. Med. Biol. 50, 4225-4241 (2005).
- G. Wang, H. Shen, K. Durairaj, X. Qian, and W. Cong, "The first bioluminescence tomography system for simultaneous acquisition of multi-view and multi-spectral data," International Journal of Biomedical Imaging 1-8 (2006).
- 11. Y. Lv, J. Tian, W. Cong, and G. Wang, "Experimental study on bioluminescence tomography with multimodality fusion," Int. J. Biomed. Imaging 2007, 86741 (2007).
- C. Kuo, O. Coquoz, T. Troy, D. Zwarg, and B. Rice, "Bioluminescent tomography for in vivo localization and quantification of luminescent sources from a multiple-view imaging system," Molecular Imaging 4, 370 (2005).
- 13. S. Chandrasekhar, Radiative Transfer (Dover Publications, 1960).
- 14. K. M. Case and P. F. PF Zweifel, *Linear Transport Theory* (Addison-Wesley Educational Publishers Inc., 1967).
- 15. A. Ishimaru, Wave Propagation and Scattering in Random Media (Academic Press, 1978).
- 16. E. E. Lewis and W. F. Miller, Computational Methods of Neutron Transport (Wiley, 1984).

- G. Bal and A. Tamasan, "Inverse source problems in transport equations," SIAM J. Math. Anal. 39, 57-76 (2007).
- 18. P. Stefanov and G. Uhlmann, "An inverse source problem in optical molecular imaging," Analysis and PDE 1, 115-126 (2008).
- R. Tibshirani, "Regression shrinkage and selection via the Lasso," Journal of the Royal Statistical Society. Series B 58, 267-288 (1996).
- 20. D. Donoho, "Compresse sensing," IEEE Transactions on Information Theory 52, 1289-1306 (2006).
- E. J. Candes and T. Tao, "Near optimal signal recovery from random projections: universal encoding strategies," IEEE Transactions on Information Theory 52, 5406-5425 (2006).
- Y. Zhang, "Theory of compressive sensing via l1-minimization: a non-RIP analysis and extensions," Rice University CAAM Technical Report TR08-11, (2008).
- H. Gao and H. Zhao, "Multilevel bioluminescence tomography based on radiative transfer equation Part 1: 11 regularization," Preprint, (2009).
- L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," J. Phys. D 60, 259-268 (1992).
- D. Strong and T. Chan, "Edge-preserving and scale-dependent properties of total variation regularization," Inverse Problems 19, 165-187 (2003).
- K. Ren, G. Bal, and A. H. Hielscher, "Frequency domain optical tomography based on the equation of radiative transfer," SIAM J. Sci. Comput. 28, 1463-1489 (2006).
- M. Nikolova, "Minimizers of cost-functions involving non-smooth data fidelity terms. Application to the processing of outliers," SIAM J. Nume. Ana. 40, 965-994 (2002).
- M. Nikolova, "A variational approach to remove outliers and impulse noise," Journal of Mathematical Imaging and Vision 20, 99-120 (2004).
- 29. S. Alliney, "Digital filters as absolute norm regularizers," IEEE Trans. Signal Process. 40, 1548-1562 (1992).
- 30. W. Yin, T. Chen, X. S. Zhou, and A. Chakraborty, "Background correction for cDNA microarray image using the TV + L1 model," Bioinformatics **21**, 2416 (2005).
- T. Chen, T. Huang, W. Yin, and X. S. Zhou, "A new coarse-to-fine framework for 3D brain MR image registration," Computer Vision for Biomedical Image 3765, 114-124 (2005).
- 32. T. Chen, W. Yin, X. S. Zhou, D. Comaniciu, and T. Huang, "Total variation models for variable lighting face recognition," IEEE Trans. Pattern Analysis and Machine Intelligence **28**, 1519-1524 (2006).
- T. F. Chan and S. Esedoglu, "Aspects of total variation regularized L1 function approximation," SIAM Journal on Applied Mathematics 65, 1817-1837 (2005).
- 34. H. Gao and H. K. Zhao, "A fast forward solver of radiative transfer equation," Transport Theory and Statistical Physics **38**, 149-192 (2009).
- 35. E. Giusti, Minimal surfaces and functions of bounded variation (Birkhäuser, 1984).
- 36. S. Boyd and L. Vandenberghe, Convex Optimization (Cambridge university press, 2004).
- 37. J. Demmel, Applied Numerical Linear Algebra (Cambidge Univ. Press, 1997).
- S. J. Kim, K. Koh, M. Lustig, and S. Boyd, "An efficient method for compressed sensing," IEEE International Conference on Image Processing 3, 117-120 (2007).
- S. J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method for large-scale l1regularized least squares," IEEE Journal of Selected Topics in Signal Processing 1, 606-617 (2007).
- 40. J. Nocedal and S. J. Wright, Numerical Optimization (Springer, 1999).
- K. Levenberg, "A method for the solution of certain nonlinear problems in least squares," Quart. Appl. Math 2, 164-168 (1944).
- D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," J. Soc. Indust. Appl. Math 11, 431-441 (1963).
- 43. 42. K. Madsen, H. B. Nielsen, and O. Tingleff, *Methods for non-linear least squares problems* (Technical University of Denmark, 1999).
- 44. J. Yang, Y. Zhang, and W. Yin, "An Efficient TVL1 Algorithm for Deblurring Multichannel Images Corrupted by Impulsive Noise," SIAM J. Sci. Comput. **31**, 2842-2865 (2009).
- T. Goldstein and S. Osher, "The split bregman method for 11 regularized problems," SIAM J. Imaging Sci. 2, 323-343 (2009).
- E. Esser, X. Zhang, and T. Chan, "A general framework for a class of first order primal-dual algorithms for TV minimization," UCLA CAM Report 09-67, (2009).

# 1. Introduction

In bioluminescence tomography (BLT) [1-3] with proper discretization, boundary measurements  $X^m := \{x_j^m, 1 \le j \le M\}$  are linearly dependent on the bioluminescent

sources  $Q := \{q_i, 1 \le i \le N_s\}$ . For conventional reconstruction methods for BLT, people use diffusion approximation (DA) [4] as forward model, and then formulate it as a standard least-square problem, i.e.,

$$\min_{Q} \|JQ - X^{m}\|_{2}^{2} + \frac{\lambda_{2}}{2} \|Q\|_{2}^{2}, \qquad (1)$$

in which J is the system matrix usually called Jacobian or sensitivity matrix corresponding to the forward model and  $\lambda_2$  is the l2-regularzation parameter.

However the reconstructed images based on DA approximation from (1) are not satisfactory mainly for two reasons. First physically DA approximation may not be accurate, especially in the non-diffusive regime, e.g., close to the source, near the boundary, forward-peaking scattering, discontinuous absorption or scattering, or when absorption is smaller or comparable with the scattering. Second, mathematically BLT with DA as forward model is severely ill-posed. Since the boundary measurement provides only angular-averaged data whose dimension does not match with that of the interior region, i.e., number of measurements M is much smaller than the total number of unknowns  $N_s$  numerically, it was shown mathematically [3] that there is no unique solution for this inverse source problem based on DA. Moreover, J comes from a smoothing diffusion operator which destroys image singularities and features. In practice all the above means that one has to solve a very ill-conditioned and underdetermined system for DA-based BLT. In general, one does not expect a good image quality with clean background from l2-regularized BLT with DA forward model alone without either more data or reduction of unknowns [5-12].

As a remedy for poor reconstruction due to forward modeling by DA, we use radiative transfer equation (RTE) [13-16] as the forward model. Physically, RTE is an accurate model for photon migration in tissues. Moreover, with RTE one can utilize much richer data, such as boundary angular-resolved measurements, whose dimension can match that in the interior region. Mathematically, it was shown [17, 18] that the RTE-based inverse source problem with angular-resolved boundary measurements almost always has a unique solution with certain stability. Numerically it means that a better conditioned and less underdetermined system resulted from RTE-based BLT than from DA-based BLT. So in those applications in non-diffusive regime, it is more appropriate to use RTE model and better image reconstruction is expected.

A powerful and general framework in dealing with inverse problem is through variational formulation. It usually contains two key ingredients in the energy functional, data fidelity, which is to match the solution with measurements based on an appropriate forward model, and regularization, which is to regularize an ill-posed problem. With probability interpretation, the form of data fidelity should be determined by noise model and the form of regularization should be determined by *a priori* knowledge. The balance between this two is usually determined by signal to noise ratio of the data and the smallest scale one wants to resolve in the reconstruction.

In general 12 regularization tends to penalize large elements and create spurious small elements which smears the reconstructed image and produces noisy background when the inverse problem is severely ill-posed. On the other hand, 11 regularization weights all elements equally and tends to produce a sparse solution with clean background which would be a suitable choice for sparse sources [19-22]. As a result, we studied 11-regularization strategy in a multilevel framework and showed its superiority over the conventional 12 regularization for sparse sources in our previous work [23] in which we minimize

$$\min_{Q} \|JQ - X^{m}\|_{2}^{2} + \lambda_{1} \|Q\|_{1}.$$
 (2)

However, 11 regularization tends to find sparse solutions, which on one hand, makes it successful for sparse source reconstruction, but on the other hand, may shrink the support for

non-sparse sources. Therefore, it is desirable to design more appropriate formulations for nonsparse sources that also inherit the merits of 11 regularization. In this paper we propose and compare a few different choices of data fidelity and regularization combinations for RTEbased BLT and justify their performances in different scenarios.

For nearly piecewise-constant source distribution in BLT, a natural choice is to use total variation (TV) regularization, which was first proposed in [24], and has wide applications in various aspects in image processing. Now we consider the following minimization problem

$$\min_{Q} \|JQ - X^{m}\|_{2}^{2} + \lambda_{TV} \|Q\|_{TV}, \qquad (3)$$

in which  $\lambda_{TV}$  is the TV-regularization parameter and  $\|\cdot\|_{TV}$  denotes the TV norm that we shall specify in the next section.

It is well known that TV regularization [25] tends to be effective for piecewise-constant reconstruction and can preserve the boundary of large object well while removing small features and thin or small objects such as small inclusions or sparse sources. Therefore we propose a combination of 11 and TV regularization which has an interesting geometric balance between them which will be explained in the next section. We shall call it 11+TV regularization, so that it combines the merits from both 11 and TV regularization in the sense that it successfully reconstructs piecewise-constant image that preserve both geometry and details with little background noise for both sparse and non-sparse sources. The optimization problem becomes

$$\min_{Q} \|JQ - X^{m}\|_{2}^{2} + \lambda_{1} \|Q\|_{1} + \lambda_{TV} \|Q\|_{TV}.$$
(4)

However, if the source distribution is smooth, it is well known that H1 seminorm, which is l2 norm of the gradient, is a good regularization. For example, H1 regularization with l2 data fidelity was used for RTE-based optical tomography in [26]. For comparison purpose, we present mathematical formulation and simulation results with H1 regularization in Section 3.

Most often in practice the data may be contaminated by various types of noises. For Gaussian type of noise, the l2 data fidelity term is a good and natural choice. However, l2 data fidelity term does not handle *outliers* well since it is biased towards large residuals. On the other hand, nonsmooth data fidelity [27, 28] can reconstruct correct images under some assumptions. As a successful example of nonsmooth data fidelity, l1 data fidelity was first introduced in [29], and later found applications in medical imaging [30, 31], computer vision [32] and image processing [33]. Therefore, we consider the following optimization problem with l1 data fidelity to deal with practical noise involving *outliers* 

$$\min_{Q} \|JQ - X^{m}\|_{1} + \lambda_{1} \|Q\|_{1} + \lambda_{TV} \|Q\|_{TV}.$$
<sup>(5)</sup>

Besides, from our simulations, we find that 11 data fidelity with 11+TV regularization tends to reconstruct as good results as 12 data fidelity with H1 regularization for the smooth non-piecewise-constant sources.

Finally, our multilevel reconstruction developed in [23] can be implemented for all the above formulations. If the source is sparse, coarse mesh reconstruction, which costs much less computation effort and is better conditioned, not only provides a good initial guess but also a good localization for the fine mesh. Hence, the reconstruction on fine mesh can be restricted to the localized region, which again reduces the computation cost. As another consequence, the amount of measurements can be reduced. On the other hand, if the source is not sparse, coarse mesh solution can still provide a good initial guess for the fine mesh and improves overall computational efficiency and stability. Moreover, coarse mesh reconstruction can also help to choose better r to balance between 11 and TV norm for the fine mesh reconstruction.

Here is the outline of this paper. In section 2, we shall first briefly introduce RTE-based BLT, define TV norm on triangular mesh and describe our interior-point methods for solving (4) and (5); in section 3, we shall validate our algorithms with simulations; in section 4, we shall conclude and comment on our future work.

#### 2. Methods

## 2.1. RTE-based BLT

In this paper, we use RTE as forward model for reasons described in the introduction. The details of an efficient solver of RTE can be found in [34]. Next we shall introduce notations for presentation purpose.

For forward problem, *in vivo* light propagation is modeled by the following RTE with the vacuum boundary condition

$$\hat{s} \cdot \nabla \Psi(\vec{r}, \hat{s}) + (\mu_a + \mu_s) \Psi(\vec{r}, \hat{s}) = \mu_s \int_{\hat{s}^{n-1}} f(\hat{s}, \hat{s}') \Psi(\vec{r}, \hat{s}') d\hat{s}' + q(\vec{r}, \hat{s})$$
(6)  
$$\Psi(\vec{r}, \hat{s}) = 0, \text{ if } \hat{s} \cdot \hat{n} < 0$$

where the quantities in the equation are the photon flux  $\Psi$  which depends on both space  $\vec{r}$  and angle  $\hat{s}$ , the light source q, the Henyey-Greenstein scattering function f with anisotropic factor g, the absorption coefficient  $\mu_a$ , the scattering coefficient  $\mu_s$ , the outer normal  $\hat{n}$  at boundary and the angular space  $\hat{S}^{n-1}$ , i.e., the unit circle in 2D or the unit sphere in 3D. After Discontinuous Galerkin discretization of (6), we numerically denote piecewise-

After Discontinuous Galerkin discretization of (6), we numerically denote piecewiseconstant isotropic bioluminescent source by  $Q = [q_i]$  and piecewise-linear photon flux by  $\Phi = [\phi_{ij,m}]$  with  $1 \le i \le N_s, 1 \le j \le n_d + 1, 1 \le m \le N_a$ , where  $N_a$ ,  $N_s$  and  $n_d$  are the number of angles, the number of spatial mesh elements and problem dimension respectively. On the other hand, we use  $\{P_j, 1 \le j \le M\}$  to denote measure operators for reconstruction. In this paper, we shall use two types of measurements, i.e., boundary angular-resolved measurement  $\Psi$  and boundary angular-averaged measurement  $\int_{\hat{s}\cdot\hat{n}>0} \hat{s}\cdot\hat{n}\Psi d\hat{s}$ . For RTE based BLT, which is a linear inverse source problem, the Jacobian matrix J is composed of Green's function of RTE. We refer readers to [23] for the computation of Jacobian J.

#### 2.2. Regularization by TV

Although 11 regularization performs well for sparse sources and is capable of recovering high-resolution details as demonstrated in [23], it tends to shrink the support of large or non-sparse sources, especially in the isotropic scattering regime or diffusive regime. To overcome this, we introduce the regularization by TV norm  $||Q||_{TV}$  for nearly piecewise-constant source reconstruction.

Since our algorithm is discretized on triangular mesh instead of Cartesian grid, we shall use a discretized version of TV coarea formula [35] for piecewise-constant Q. Let  $\{\Gamma_k, 1 \le k \le N_e\}$  be edges, then we have

$$\|Q\|_{TV} = \sum_{k=1}^{N_e} L_k |q^{k,l} - q^{k,r}|, \qquad (7)$$

in which  $L_k$  denotes the length of the k th edge, and  $q^{k,l}$  and  $q^{k,r}$  represent the left element and the right element with respect to the k th edge with the convention that the spatial index of the left element  $q^{k,l}$  is always smaller than that of right element  $q^{k,r}$  to avoid the repetitive counting of edges. For TV on the domain boundary, we let  $q^{k,l}$  be the corresponding boundary element and  $q^{k,r}$  be empty, i.e.,  $L_k | q^{k,l} |$ .

## 2.3. Optimization Algorithm for 11 and TV regularization

In this section, we shall describe an interior-point method [36] for solving (4), which can be easily modified for minimization problem (2) and (3).

First we recast (4) as a constrained convex optimization problem

$$\min_{Q} \|JQ - X^{m}\|_{2}^{2} + \lambda_{1} (\sum_{i=1}^{N_{s}} A_{i}u_{i} + r \sum_{k=1}^{N_{e}} L_{k}v_{k}), \qquad (8)$$

with constraints  $-u_i \leq q_i \leq u_i, 1 \leq i \leq N_s$  and  $-v_k \leq q^{k,l} - q^{k,r} \leq v_k, 1 \leq k \leq N_e$ . Here  $A_i$  is area of the *i* th element and  $r := \lambda_{TV} / \lambda_1$ . Here we incorporate area weights  $A_i$  and edge weights  $L_k$  into the scheme to account for non-uniformity of the mesh, which is not necessary on a uniform mesh.

Next we transform (8) into an sequence of t -dependent unconstrained convex optimization problems

$$\Phi_t(Q, u, v) = ||JQ - X^m||_2^2 + [\lambda_1 \sum_{i=1}^{N_s} A_i u_i + \frac{1}{t} \phi_1(Q, u)] + [\lambda_{TV} \sum_{k=1}^{N_e} L_k v_k + \frac{1}{t} \phi_{TV}(Q, v)]$$
(9)

by penalizing the element and edge constraints corresponding with logarithmic barrier functions  $\phi_1(Q, u) = -\sum_{i=1}^{N_s} (\log u_i^+ + \log u_i^-)$  and  $\phi_{TV}(Q, v) = -\sum_{k=1}^{N_e} (\log v_k^+ + \log v_k^1)$  with  $u_i^+ = u_i + q_i$ ,  $u_i^- = u_i - q_i$ ,  $v_k^+ = v_k + q^{k,l} - q^{k,r}$  and  $v_k^- = v_k - q^{k,l} + q^{k,r}$ .

The standard interior point method for solving constrained convex problem (8) is through minimizing a sequence of unconstrained convex problems (9), for which the solution converges as  $t \to \infty$ . Actually the solution of (9) is no more than  $2(N_s + N_e)/t$ -suboptimal. This algorithm is what so called barrier method (algorithm 11.1) [36].

For subproblems (9), a simple gradient descent method (algorithm 9.3) [36] is implemented with the efficient computation of the search direction by a preconditioned conjugate gradient method (PCG) [37-39] and backtracking line search algorithm [36, 40].

Next we shall describe our PCG method for computing the search direction by solving

$$\nabla^2 \Phi_t \begin{bmatrix} \Delta Q \\ \Delta u \\ \Delta v \end{bmatrix} = -\nabla \Phi_t . \tag{10}$$

Before that, we first compute  $\nabla \Phi_t$  and the Hessian  $\nabla^2 \Phi_t$ , i.e.,

$$\nabla \Phi_{t} = 2J^{T}(JQ - X^{m}) + \lambda_{1}[A]_{u} + \lambda_{TV}[L]_{v} + \frac{1}{t}([\nabla \phi_{1}] + [\nabla \phi_{TV}]), \qquad (11)$$

$$\nabla^{2} \Phi_{t} = 2J^{T}J + \frac{1}{t} (\nabla^{2} \phi_{1} + \nabla^{2} \phi_{TV}).$$
(12)

Note the equality above should be understood in the sense the left hand side is assembled from the matrices on the right hand side with correct ordering of unknowns. Here  $[A]_u$  denotes the element area vector for unknown u and  $[L]_v$  for the edge length vector for unknown v. (11) and (12) can be easily implemented numerically.

For the convenience of presentation, we define  $f^+(a,b) := 1/a^2 + 1/b^2$ ,  $f^-(a,b) := 1/a^2 - 1/b^2$ ,  $g^+(a,b) := 1/a + 1/b$  and  $g^-(a,b) := 1/a - 1/b$ . And then we assemble the following local vectors and matrices corresponding to each pair of constraints for either some element or edge into both sides of (10).

First, with the first column or row for  $q_i$  and the second for  $u_i$ , the local vector and matrix for the pair of constraints corresponding to the *i* th element are  $[\nabla \phi_1]_i = -\begin{bmatrix} \alpha_i^1 \\ \alpha_i^2 \end{bmatrix}$  and  $\begin{bmatrix} -1 & -2 \end{bmatrix}$ 

$$\begin{bmatrix} \nabla^2 \phi_1 \end{bmatrix}_i = \begin{bmatrix} \sigma_i^i & \sigma_i^2 \\ \sigma_i^2 & \sigma_i^1 \end{bmatrix}, \ 1 \le i \le N_s \text{ , with } \alpha_i^1 = g^-(u_i^+, u_i^-) \text{ , } \alpha_i^2 = g^+(u_i^+, u_i^-) \text{ , } \sigma_i^1 = f^+(u_i^+, u_i^-) \text{ and } \sigma_i^2 = f^-(u_i^+, u_i^-).$$

Second, with the first column or row for  $q^{k,l}$ , the second for  $q^{k,r}$  and the third for  $v_k$ , the local vector and matrix for the pair of constraints corresponding to the k th edge

are 
$$[\nabla \phi_{TV}]_k = -\begin{bmatrix} \alpha_k^3 \\ -\alpha_k^3 \\ \alpha_k^4 \end{bmatrix}$$
 and  $[\nabla^2 \phi_{TV}]_k = \begin{bmatrix} \sigma_k^3 & -\sigma_k^3 & \sigma_k^4 \\ -\sigma_k^3 & \sigma_k^3 & -\sigma_k^4 \\ \sigma_k^4 & -\sigma_k^4 & \sigma_k^3 \end{bmatrix}$ ,  $1 \le k \le N_e$ , with

 $\alpha_k^3 = g^-(v_k^+, v_k^-), \alpha_k^4 = g^+(v_k^+, v_k^-), \sigma_k^3 = f^+(v_k^+, v_k^-) \text{ and } \sigma_k^4 = f^-(v_k^+, v_k^-).$ 

It is easy to see that the Hessian (12) is positive symmetric definite. In order to make PCG method work, we need to find a precondition matrix P which is not only fast to invert, but also a good approximation of the Hessian. Here we choose the sparse matrix  $2\text{diag}(J^TJ) + \frac{1}{t}(\nabla^2\phi_1 + \nabla^2\phi_{TV})$  as P, which is also symmetric positive definite, and thus  $M^{-1}g$  can be efficiently solved by Gauss-Seidel method for any vector g.

We summarize our minimization algorithm below. With  $\varepsilon$  between 0.00001 and 0.001 as stopping criterion,  $\lambda_1$  between 0.01 and 1, and *r* between 0.1 and 10 as the ratio of TV norm over 11 norm, our barrier method gives good results from our numerical tests. Also one needs to increase values of  $\varepsilon$  and  $\lambda_1$  when noise level is high. Note in backtracking line search, we check the constraint condition for (8) for feasibility.

Algorithm: barrier method for optimization problem (8) given tolerance  $\varepsilon > 0$ ,  $\lambda_1$  and r. Initialize  $\mu = 2$ ,  $t = 1/\lambda_1$ ,  $\lambda_{TV} = r\lambda_1$ , q = 0, u = 1, v = 1. Repeat optimization of subproblems (9)

- 1. compute the search direction  $(\Delta Q, \Delta u, \Delta v)$  by (12) through PCG with preconditioner P, the relative tolerance  $\varepsilon_{pcg} = \min(0.1, 0.01(N_s + N_e)/t)$  and the relative tolerance  $\varepsilon_{gs} = 0.0001$  for Gauss-Seidel iterations in PCG.
- 2. compute step size *s* by backtracking line search with  $\alpha = 0.01$ ,  $\beta = 0.5$ ,  $s_0 = 1$ .
- 3. update:  $(Q, u, v) = (Q, u, v) + s(\Delta Q, \Delta u, \Delta v)$
- 4. quit if  $2(N_s + N_e)/t < \varepsilon$
- 5.  $t = \mu t$

Next, we consider the balance between 11 and TV regularization. For source distribution that is piecewise constant with uniform intensity I, TV norm is approximately IC when C is the perimeter of boundary of the support and 11 norm is approximately AI, where A is the area of the support. So in this case a good choice of ratio r, should be roughly proportional to A/C. Note A/C can be independent of the length unit since one can always nondimensionalize RTE (6) so that A/C is dimensionless, to which r should be equal. This implies that ratio r gives a geometric balance of 11 norm and TV norm. For example, 11 regularization should weight more than TV regularization for sources with smaller ratio A/C, e.g., sources with small or thin support, which corresponds to sparse sources. As a result, with sparse sources as a priori, one should put more weights on 11 regularization by choosing small r in the algorithm to get better results. We use numerical experiments to demonstrate our rule of optimal choices for different configurations in Section 3. On the other hand, our tests seem to suggest that the results are not very sensitive to the choice of r. From our numerical experiments, the value of r is typically between 0.1 and 10.

Last but not the least, we normalize J by columns to get  $J^*$ , and then actually solve the following instead of (8)

$$\min_{Q^*} \|J^*Q^* - X^m\|_2^2 + \lambda_1 \sum_{i=1}^{N_s} A_i u_i + \lambda_{TV} \sum_{k=1}^{N_e} L_k v_k$$
(13)

This is crucial for the successful reconstruction. Physically this means the total energy of measurements with respect to each unknown is on the same level, while mathematically  $diag(J^{*T}J^*)$  is equal to identity matrix, which makes problem isotropic to all unknowns. After solving (13), we need to change correspondingly from  $Q^*$  to Q, which reflects the sensitivity of each unknown on measurements. This is also equivalent to optimization in weighted norm.

#### 2.4. Optimization Algorithm for 11 Data Fidelity

To simplify the discussion, we will only present the treatment of 11 data fidelity term in the optimization problem (5), and the 11 or TV regularization can be easily adapted to the problem (5) in the same way as what we have done for (4) by (8) in the previous subsection.

Following the standard interior-point method, we first transform the problem

$$\min_{\alpha} \|JQ - X^m\|_1 \tag{14}$$

into the following constrained convex optimization problem

$$\min_{\mathcal{Q}} \sum_{j=1}^{M} \mathcal{Y}_{j} \tag{15}$$

with constraints  $-y_j \leq (\sum_{i=1}^{N_i} J_{ij}q_i - X_j^m) \leq y_j, 1 \leq j \leq M$ .

Similarly, (15) can be recasted as a sequence of unconstrained convex optimization problems as

$$\Phi_t(Q, y) = \sum_{j=1}^M y_j + \frac{1}{t}\phi(Q, y).$$
(16)

by penalizing the constraints with the logarithmic barrier function  $\phi(Q, y) = -\sum_{j=1}^{M} (\log y_j^+ + \log y_j^-)$  with  $y_j^+ = y_j + \sum_{i=1}^{N_s} J_{ij}q_i - X_j^m$  and  $y_j^- = y_j - \sum_{i=1}^{N_s} J_{ij}q_i + X_j^m$ .

To find the search direction, we need to compute

$$\nabla^2 \Phi_t \begin{bmatrix} \Delta Q \\ \Delta y \end{bmatrix} = -\nabla \Phi_t, \qquad (17)$$

for which we have  $\nabla \Phi_t = -\begin{bmatrix} J^T \operatorname{diag}(\alpha^5) / t \\ [\alpha^6]_y / t - [1]_y \end{bmatrix}$  and the Hessian

$$\nabla^2 \Phi_t = \frac{1}{t} \begin{bmatrix} J^T \operatorname{diag}(\sigma^5) J & J^T \operatorname{diag}(\sigma^6) \\ \operatorname{diag}(\sigma^6) J & \operatorname{diag}(\sigma^5) \end{bmatrix} \text{ with } \alpha_j^5 = g^-(y_j^+, y_j^-), \ \alpha_j^6 = g^+(y_j^+, y_j^-), \ \sigma_j^5 = f^+(y_j^+, y_j^-) \text{ and } g^-(y_j^+, y_$$

 $\sigma_j^6 = f^-(y_j^+, y_j^-), \ 1 \le j \le M$ . Here  $[1]_y$  is the unit vector for y,  $[\alpha^6]_y$  is the constraint vector for y and diag(·) is the diagonal matrix with diagonal elements in the bracket.

To invert (17) efficiently, we use a similar PCG method with a sparse symmetric positive definite preconditioner  $P = \begin{bmatrix} \operatorname{diag}(J^T \operatorname{diag}(\sigma^5)J) & 0 \\ 0 & \operatorname{diag}(\sigma^5) \end{bmatrix}$ .

Last, we want to emphasize in our algorithm we actually minimize (5) with normalized Jacobian instead of (14), i.e.: we solve a sequence of subproblems coming from the combination of (9) and (16) except 12 data fidelity term; in the barrier method, we compute  $\nabla^2 \Phi_t \Delta x = -\nabla \Phi_t$  with  $\Delta x := [\Delta Q \quad \Delta u \quad \Delta v \quad \Delta y]^T$  for search direction.

#### 3. Simulations

For comparison purpose, besides (1) for l2 regularization, we also consider H1 regularization by a discrete version of, but not exactly, H1 seminorm  $|Q|_{H_1}^2 := \sum_{k=1}^{N_e} (q^{k,l} - q^{k,r})^2$ , with the same definition as in (7), i.e.,

$$\min_{Q} \|JQ - X^{m}\|_{2}^{2} + \frac{\lambda_{2}}{2} \|Q\|_{H_{1}}^{2}, \qquad (18)$$

or by both  $|Q|_{H_1}^2$  and the usual discrete l2 norm  $||Q||_2^2 := \sum_{i=1}^{N_s} q_i^2$ , i.e.,

$$\min_{Q} \|JQ - X^{m}\|_{2}^{2} + \lambda_{2}(\|Q\|_{2}^{2} + r_{2} |Q|_{H_{1}}^{2}), \qquad (19)$$

where  $r_2$  is the parameter to balance between 12 norm and H1 norm.

(1) and (18) can be solved through Levenberg-Marquardt method [41-43], in which we iteratively solve

$$Q^{k+1} = Q^{k} + (J^{T}J + \lambda_{2,k}D)^{-1}(J^{T}(JQ^{k} - X^{m})).$$
(20) [41-43]

D in (20) comes from l2 or H1 regularization. For l2-regularized problem (1), simply D = I; for H1-regularized problem (18), D comes from H1 norm and can be assembled from the local matrix  ${}_{[D]_k} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$  for each edge. The detailed implementation can be found in Algorithm 3.16 in [43] with one modification of the gain predicted by linear model as  $\frac{1}{2}(Q_{k+1}-Q_k)^T[\lambda_{2,k}D(Q_{k+1}-Q_k)-J^TX^m]$  to guarantee a successful automatic adjust of  $\lambda_2$  adaptively. Also (20) can be inverted with a similar PCG method in the previous section.

In this section we use various numerical simulations to validate our novel 11+TV regularization strategy for BLT and compare 11 data fidelity with 12 data fidelity. In all figures

shown in this section, reconstructed images are plotted as what they originally are from reconstructions without further image processing, such as thresholding. As pointed out at the beginning, radiative transport equation (RTE) is used as the underlying model for both forward and inverse problem. Data for forward problem is generated on a fine mesh that is different from the fine mesh used for inverse problem. The fast forward solver developed in [34] is used to compute both the forward problem and the Green's functions for inverse problem. The methods for solving l2-regularized BLT can be found in [23]. In the following figures for l2 regularization except Fig. 14 and 15, we only plot results from (1) since l2 regularization (18) and (19) give the similar results. In particular we design different simulations to demonstrate the following points.

First, with *a priori* knowledge of the rough size information of bioluminescent source, one can choose the ratio r between TV norm and 11 norm to optimize reconstruction results. On the other hand, from our numerical experiments, it seems that the reconstruction images are not very sensitive to r in the sense that 11+TV regularization with the ratio around unity in general produces satisfactory images. See Fig. 1.

Second, with angular-averaged or angular-resolved data for non-sparse or sparse source, our 11+TV regularization reconstruct better than either 11 or TV regularization. Moreover, all of these perform better than the conventional l2 regularization. See Fig. 2-13.

Fourth, for the reconstruction of non-piecewise-constant images, with 12 data fidelity, 12 regularization (1) or H1 regularization (18) may give better results than 11+TV regularization (4). On the other hand, 11 data fidelity (5) with 11+TV regularization gives as good results as 12 data fidelity with 12 or H1 regularization. We show examples in Fig. 14 and 15.

Last, 11 data fidelity is capable of dealing with noise containing *outliers* while 12 data fidelity fails to handle it. See Fig. 16.

The following simulations are performed on a square domain with 20mm side length with anisotropic factor g = 0.9 or 0.0, homogeneous  $\mu_a = 0.01 mm^{-1}$  and  $\mu_s = 1mm^{-1}$ . The reconstruction mesh has 32 angular directions, 1397 nodes and 2672 elements. To avoid inverse crime, all forward data for various simulations are generated with 64 angular directions on the corresponding spatial mesh, which is different from reconstruction mesh. All the reconstruction results are either with 120 boundary angular-averaged data or 900 boundary angular-resolved data. For all simulations with 11+TV regularization,  $\lambda_1 = 0.01$  and r = 1 unless otherwise indicated.

The purpose of simulations in Fig. 1 is to demonstrate the dependence of the choice of the ratio r between TV norm and 11 norm on A/C. In Fig. 1, we put in the square domain three inclusions sequentially with the increasing side length of the ratio 3: a square inclusion (a) with side length 1.6mm and the forward mesh with 1401 nodes and 2680 elements, a square inclusion (b) with side length 4.8mm and the forward mesh with 1423 nodes and 2720 elements and a square inclusion (c) with side length 13.4mm and the forward mesh with 1355 nodes and 2584 elements. The simulations are performed in each case with 0.2, 1 and 5 for r with angular-averaged data and g=0.9. For the smallest inclusion (a), the reconstruction (d) with r = 0.2 gives the best result. For the inclusion (c), the reconstruction (l) with r = 5 gives the best result. In conclusion, the results are not very sensitive to r, however the optimal choice does need *a priori* knowledge.

In Fig. 2, Source I (a) is composed of three objects with the forward mesh (b) of 1445 nodes and 2768 elements; Source II (c) is composed of three letters with the forward mesh (d) of 1509 nodes and 2896 elements. Regarding the relative size of sources, Source I is non-sparse and Source II is sparse. We present below in Fig. 3-10, 12 and 13 that 11+TV regularization (4) in general gives the best results of 12 regularization (1), 11 regularization (2) or TV regularization (3) for both Source I and II.



Fig. 1. 11+TV regularized reconstructions of different square inclusions. Figures (a), (b) and (c) are true sources. Figures (d), (g) and (j) are reconstruction results of (a) with r=0.2, 1 and 5 respectively. Figures (e), (h) and (k) are reconstruction results of (b) with r=0.2, 1 and 5 respectively. Figures (f), (i) and (l) are reconstruction results of (c) with r=0.2, 1 and 5 respectively.



Fig. 2. Sources and meshes for non-sparse and sparse reconstructions. Source I and II are in Figure (a), (c) respectively with the corresponding meshes in Figure (b) and (d).

For non-sparse source with g=0.9 as shown in Fig. 3 and 4, regularization by 11+TV norm indeed gives better results than that with 12, 11 or TV norm. In Fig. 3, regarding the reconstruction with angular-averaged data as a severely ill-posed inverse problem, 11+TV method (d) is still able to separate three objects and roughly preserve image boundaries with acceptable background noise, while 11 norm (b) shrinks the support significantly and produces high intensity, TV norm alone (c) gives the blurred image with some noise in between objects, and one can hardly see anything through 12 norm (a). In Fig. 4, with angular-resolved data, 11+TV norm (d) still gives the most accurate reconstruction of image edges, details and intensities. Similarly, reconstruction results with g=0.0 are shown in Fig. 5 and 6, which are in diffusive regime. Here 11+TV is even more impressive for angular-averaged data in Fig. 5.



Fig. 3. Reconstructions with angular-averaged data for Source I with g=0.9. Figures (a), (b), (c) and (d) are reconstructed images by 12, 11, TV and 11+TV regularization respectively.



Fig. 4. Reconstructions with angular-resolved data for Source I with g=0.9. Figures (a), (b), (c) and (d) are reconstructed images by 12, 11, TV and 11+TV regularization respectively.



Fig. 5. Reconstructions with angular-averaged data for Source I with g=0.0. Figures (a), (b), (c) and (d) are reconstructed images by 12, 11, TV and 11+TV regularization respectively.



Fig. 6. Reconstructions with angular-resolved data for Source I with g=0.0. Figures (a), (b), (c) and (d) are reconstructed images by 12, 11, TV and 11+TV regularization respectively.



Fig. 7. Reconstructions with angular-averaged data for Source II with g=0.9. Figures (a), (b), (c) and (d) are reconstructed images by 12, 11, TV and 11+TV regularization respectively.



Fig. 8. Reconstructions with angular-resolved data for Source II with g=0.9. Figures (a), (b), (c) and (d) are reconstructed images by 12, 11, TV and 11+TV regularization respectively.



Fig. 9. Reconstructions with angular-averaged data for Source II with g=0.0. Figures (a), (b), (c) and (d) are reconstructed images by 12, 11, TV and 11+TV regularization respectively.



Fig. 10. Reconstructions with angular-resolved data for Source II with g=0.0. Figures (a), (b), (c) and (d) are reconstructed images by l2, l1, TV and l1+TV regularization respectively.

Next, we evaluate the reconstruction performance for sparse sources with g=0.9 as shown in Fig. 7 and Fig. 8. Regarding the reconstruction with angular-averaged data in Fig. 7, the conclusion is still the same as that for non-sparse sources: 11+TV method (d) gives the best result. However, since in this case not only three letters are too small, but also the data are limited, even 11+TV method is unable to resolve the image details. On the other hand, with angular-resolved data in Fig. 8, 11 norm (b) gives the best details although with nonsmooth intensities and 11+TV (d) is the second best although it blurs the image a little bit. This makes sense since 11 is the most appropriate choice for sparse sources. Similarly, reconstruction results with g=0.0 are shown in Fig. 9 and Fig. 10.

Then, we compute an example of our multilevel approach for Source I ((a) in Fig. 2) with g=0.0 as shown in Fig. 11. The single-level reconstruction (a) is done on the reconstruction mesh specified at the beginning of this section with 360 angular-resolved measurements. With the same reconstruction mesh as the fine mesh, the two-level reconstruction with 180 angular-resolved measurements also has a coarse mesh with 365 nodes and 668 elements. Both use 11+TV regularization. The results in Fig. 11 show that our multilevel approach is more efficient than single-level reconstruction with saving of at least half of the computational time.

In our multilevel approach [23] with 11, TV or 11+TV regularization, we choose 11+TV regularization for the coarse reconstruction, since it localizes the source with proper boundary

and little background noise for both sparse and non-sparse sources. On the other hand, although 11 regularization is favorable for sparse sources, it is not necessary to use it on the coarse mesh since it is difficult to get high-resolution images through coarse reconstruction anyway. Next on the fine mesh, the element edges in (7) are restricted within the support of Q found through coarse reconstruction with the support boundary as the boundary for unknowns. Meanwhile, coarse reconstruction also helps to choose a proper ratio r in (8) to balance between 11 and TV norm for the fine reconstruction, i.e., more TV norm for non-sparse sources or more 11 norm for sparse sources.



Fig. 11. Single-level reconstruction and two-level reconstruction for Source I with g=0.0. Figures (a) is from single-level reconstruction with 360 angular-resolved measurements. Figures (b) is from two-level reconstruction with 180 angular-resolved measurements.



Fig. 12. Reconstructions with angular-resolved data for Source I with different intensities with g=0.9. Figures (a), (b), (c) and (d) are reconstructed images by 12, 11, TV and 11+TV regularization respectively.



Fig. 13. Reconstructions with angular-resolved data for Source I with different intensities with g=0.9. Figure (a) is the true source. Figure (b) is the reconstructed image by 11+TV regularization.

Last, we do the reconstruction for Source I with g=0.9 and varying intensities, i.e., 1 for hexagonal inclusion, 2 for square inclusion and 4 for triangular inclusion with reconstruction results in Fig. 12 and 13, which show that our algorithms can successfully reconstruct sources with different intensities.



Fig. 14. Reconstructions with angular-resolved data for non-piecewise-constant source with g=0.9. Figure (a) is the true source. Figure (b), (c), (d), (e) and (f) are for 12 data fidelity with 12, H1, 11, TV, 11+TV regularization respectively. Figure (g), (h) and (i) are for 11 data fidelity with 11, TV, 11+TV regularization respectively.



Fig. 15. Reconstructions with angular-resolved data for non-piecewise-constant source with g=0.0. Figure (a) is the true source. Figure (b), (c), (d), (e) and (f) are for 12 data fidelity with 12, H1, 11, TV, 11+TV regularization respectively. Figure (g), (h) and (i) are for 11 data fidelity with 11, TV, 11+TV regularization respectively.

In Fig. 14 and 15 we show an example with g=0.0 or 0.9 that 12 regularization gives better results than 11 or TV regularization for non-piecewise-constant source with angularresolved data. In the simulations, an elliptic source inclusion (a) is centered at (10mm, 10mm) with major radius 8mm and minor radius 4mm, the axis which is 45 degrees from x direction, intensity  $Q(x, y) = \begin{cases} e^{1-1/[1-(\frac{x'-10}{8})^2 - (\frac{y'-10}{4})^2]}, (x, y) \in E\\ 0, (x, y) \notin E \end{cases}$  where (x', y') is the coordinate of

(x, y) after 45 degree rotation. The forward mesh has 64 angular directions, 1421 nodes and 2720 elements.

From Fig. 14 and 15, with 12 data fidelity, H1 seminorm (c) gives the best result, while 11 regularization (d) significantly sparsifies the source distribution, TV regularization (e) and 11+TV regularization (f) tend to generate the piecewise-constant solution; on the other hand, 11 data fidelity (g), (h) and (i) tends to alleviate the tendency for the piecewise-constant solution and 11 data fidelity with 11+TV norm can reconstruct as good as 12 data fidelity with H1 norm.



Fig. 16. Comparison of 12 data fidelity with 11 data fidelity with 11 regularization. Figures (a), (c), (e) and (g) are reconstruction results with 12 data fidelity with 0% Gaussian noise, 10% Gausisan noise and 20% salt-and-pepper noise, 10% Gausisan noise and 33% salt-and-pepper noise respectively. Figures (b), (d), (f) and (h) are reconstruction results with 11 data fidelity with 0% Gaussian noise, 10% Gausisan noise, 10% Gausisan noise and 20% salt-and-pepper noise respectively.

In Fig. 16 we compare the performance of 11 data fidelity with that of 12 data fidelity with respect to Gaussian noise, salt-and-pepper noise or both. Five small inclusions with 1mm diameter are centered at (5mm, 5mm), (5mm, 10mm), (5mm, 15mm), (10mm, 5mm) and (15mm, 5mm) and the forward spatial mesh is of 1917 nodes and 3696 elements. The simulations are performed through 11 regularization with angular-resolved data with g=0.9 and  $\lambda_1 = 1$ . We add two different types of noise: one is Gaussian noise and the other is salt-and-pepper noise. For the salt-and-pepper noise, we randomly pick up certain number of measurements and set the value to a totally irrelevant one, e.g., 0.

With 12 data fidelity, the result without noise is shown in (a) while the result with 10% Gaussian noise is shown in (c). Adding salt-and-pepper noise, we plot the results in (e) and (g) for randomly setting 180 and 300 measurements to zero respectively from 900 measurements with 10% Gaussian noise. Similarly, we plot images (b), (d), (f) and (h) with 11 data fidelity. From the reconstruction results, we conclude 11 data fidelity has good performance for both Gaussian noise and the pepper-and-salt noise while 12 data fidelity handles pepper-and-salt noise poorly. Therefore, 11 data fidelity (5) instead of 12 data fidelity (4) is strongly recommended in practice with *outliers*, such as salt-and-pepper noise.

Last we give the rough estimate of the computational storage and time for our algorithms on a typical reconstruction problem with  $N_a = 32$  directions,  $N_s = 2672$  elements and M = 900 measurements.

Regarding the memory storage, in our forward solver [34], the relaxation scheme needs two vectors of individual length  $N_a \cdot N_s \cdot 3$  for the residual and the solution respectively,

where 3 is from our spatial piecewise-linear DG discretization, and multigrid methods double the amount for one relaxation, which comes to a total of  $12N_a \cdot N_s$ ; on the other hand, the main storage for inverse problem is for  $B := J^T J$ , i.e.,  $N_s^2$ . Thus the total memory is around 70 megabytes if each value takes 8 bytes.

Regarding the computational time, our forward solver is roughly  $O(N_s \cdot N_a^2)$  for a fixed set of parameters. In the reconstruction, we need to compute B , which takes  $M \cdot N_s^2$ . Let n be either the number of subproblems (9) or iteration (20). With 12 data fidelity, each subproblem (9) for 11+TV regularization takes roughly  $O(N_s^2)$  and one iteration (20) for 12 or H1 regularization takes roughly  $O(N_s^2)$  as well, that both comes from the multiplication of B with another vector; as a result, the computation for 12 data fidelity is roughly  $O(N_s^2) \cdot n + M \cdot N_s^2$ . On the other hand, with 11 data fidelity, each subproblem (16) for optimization with or without 11+TV regularization takes  $O(N_s^2) + M \cdot N_s^2$  since one needs to compute B for each subproblem, and thus the total is roughly  $(O(N_s^2) + M \cdot N_s^2) \cdot n$ . Note each set of Gauss-Seidel iterations takes roughly  $O(N_s)$  due to the sparse and positivedefinite precondition matrix P. Finally, on a desktop with Intel CPU E6850 (3.00GHz), each forward solver takes 3 to 4 seconds and 900 measurements take around 50 minutes for adjoint method to compute Jacobian J; optimization with 12 data fidelity takes 3-5 minutes while optimization with 11 data fidelity takes 10-15 minutes. In summary, the major computational time is spent on the computation of Jacobian J while the optimization with 12 data fidelity takes usually less than ten percents of the computational time of J and that with 11 data fidelity takes around twenty percents. On the other hand, computation of the Jacobian can be easily parallelized to reduce the computation time significantly. Also there are some other potentially efficient methods for problems (4) or (5), such as [44-46], which may improve a lot on optimization time.

With our multilevel approach, we may reduce M and  $N_s$  substantially, especially for sparse sources, and thus to reduce both memory storage and computational time. However, note that the memory for forward solver does not reduce and the computational time of each forward solver does not decrease.

## 4. Conclusions and discussions

In this paper, we have introduced 11+TV regularization strategy for BLT and developed novel optimization algorithm for solving 11+TV-regularized BLT based on RTE. Simulations have validated that, for nearly piecewise-constant sources, of regularizations by 12, 11, TV or 11+TV norm, 11+TV gives the best results in the sense that it combines merits from both 11 and TV and is in general able to reconstruct image details, edges and intensities for both sparse and non-sparse images with boundary angular-averaged or angular-resolved data. We have also studied optimization with 11 data fidelity and developed the corresponding optimization algorithm. Simulations have shown that 11 data fidelity is preferred over 12 data fidelity with 000 the other hand, for non-piecewise-constant sources, 11 data fidelity with 11+TV regularization is able to reconstruct as good results as 12 data fidelity with H1 regularization, and both in general are better than other strategies. Besides, our optimization method with either 11+TV regularization or 11 data fidelity is quite general in the sense that it can be easily adapted to other problems similar to (4) and (5).

#### Acknowledgement

The work is partially supported by NSF grant DMS0811254 and ONR grant N00014-02-1-0090.