



Published in final edited form as:

*Phys Med Biol.* 2008 July 21; 53(14): 3921–3942. doi:10.1088/0031-9155/53/14/013.

## Fast iterative image reconstruction methods for fully 3D multispectral bioluminescence tomography

Sangtae Ahn<sup>1</sup>, Abhijit J Chaudhari<sup>1,3</sup>, Felix Darvas<sup>1,4</sup>, Charles A Bouman<sup>2</sup>, and Richard M Leahy<sup>1</sup>

<sup>1</sup>Signal and Image Processing Institute, University of Southern California, Los Angeles, CA 90089, USA

<sup>2</sup>School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA

### Abstract

We investigate fast iterative image reconstruction methods for fully 3D multispectral bioluminescence tomography for applications in small animal imaging. Our forward model uses a diffusion approximation for optically inhomogeneous tissue, which we solve using a finite element method (FEM). We examine two approaches to incorporating the forward model into the solution of the inverse problem. In a conventional direct calculation approach one computes the full forward model by repeated solution of the FEM problem, once for each potential source location. We describe an alternative on-the-fly approach where one does not explicitly solve for the full forward model. Instead, the solution to the forward problem is included implicitly in the formulation of the inverse problem, and the FEM problem is solved at each iteration for the current image estimate. We evaluate the convergence speeds of several representative iterative algorithms. We compare the computation cost of those two approaches, concluding that the on-the-fly approach can lead to substantial reductions in total cost when combined with a rapidly converging iterative algorithm.

### 1. Introduction

Bioluminescence tomography is an *in vivo* imaging technique that localizes and quantifies bioluminescent sources in a small animal. It has recently gained a great deal of attention as a promising means for macroscopic *in vivo* imaging of gene expression and other molecular and cellular-level processes (Contag and Bachmann 2002, Ntziachristos *et al* 2005). The objective of bioluminescence tomography is to reconstruct the three-dimensional (3D) spatial distribution of bioluminescent sources inside the animal from images of the light emitted through the animal surface. Collecting measurement data in multiple spectral bands and from more than one view, so that the entire animal surface is imaged, facilitates tomographic reconstruction of the 3D optical source distribution (Chaudhari *et al* 2005).

© 2008 Institute of Physics and Engineering in Medicine

leahy@sipi.usc.edu

<sup>3</sup>Present address: Department of Biomedical Engineering, University of California, Davis, CA 95616, USA.

<sup>4</sup>Present address: Department of Neurological Surgery, University of Washington, Seattle, WA 98105, USA.

We have set up a bioluminescence tomographic imaging system using mirrors to collect four independent views (Chaudhari *et al* 2005). Multispectral data were acquired to improve localization (Chaudhari *et al* 2005, Dehghani *et al* 2006, Cong and Wang 2006, Wang *et al* 2006b, Kuo *et al* 2007, Allard *et al* 2007, Lv *et al* 2007). We then used an FEM solution of the diffusion equation to construct realistic forward models. We have successfully localized bioluminescent sources with this system, using the TOAST software to compute the forward model (Arridge *et al* 1993). However, TOAST is primarily intended to solve the diffuse optical tomography problem and is not well adapted for bioluminescence tomography applications; the forward model computation cost is high.

Fully 3D multispectral bioluminescence tomography is a computationally challenging inverse problem because (1) FEM-based forward models, which allow inhomogeneous tissue and realistic geometry, require the inversion of a very large matrix to solve the forward problem, (2) the use of multispectral data, which helps localize deep sources, increases the problem size by the number of spectral bins and (3) the inverse problem is intrinsically ill-posed due to the nature of the photon diffusion process and the limited information contained in data collected only on the animal surface.

In this paper we investigate various numerical techniques in order to minimize the cost of computing inverse solutions. The inverse problem is equivalent to a search for the source distribution which best predicts the measured data while also satisfying an *a priori* assumption regarding the smoothness of the source distribution. In order to iteratively solve the inverse problem, a forward model which maps a source to the data domain must be computed repeatedly. We explore two different approaches to incorporating the forward solutions into iterative algorithms for the inverse problem: (1) a direct calculation approach where the forward problem is solved in advance for every source location and then a forward solution is computed as a linear combination of these precomputed solutions, and (2) an on-the-fly approach where the FEM matrix inversion problem is solved as needed at each iteration using the current estimate of the source configuration. In this way, the full forward model need never be computed. We evaluate those approaches, combined with several different iterative algorithms, to determine the combination of forward and inverse algorithm that will minimize the computational cost.

Our goal in this paper is to introduce and evaluate methods that can rapidly reconstruct 3D bioluminescent images while still using a realistic FEM forward model. To the best of our knowledge, the on-the-fly approach is novel to this application and can substantially reduce the total reconstruction time, as shown in section 4. The iterative algorithms we consider in this paper include the class of incremental gradient or ordered subset methods, which have been widely explored for applications in nuclear medicine imaging (Hudson and Larkin 1994, Browne and De Pierro 1996, Ahn and Fessler 2003) but have not been investigated in the optical imaging literature. The algebraic reconstruction technique (ART) (Gordon *et al* 1970), which is a special case of the incremental gradient algorithms, was used for diffuse optical tomography (Arridge and Schweiger 1998). Our comparison of reconstruction times for the different methods described above should help guide the selection of reconstruction and forward modeling algorithms for developers of 3D bioluminescent and fluorescent imaging systems when using FEM-based forward models.

In section 2 we formulate the forward problem based on the diffusion equation using FEM models and describe approaches to incorporating the FEM solvers into iterative algorithms. In section 3 we define the inverse problem using regularized least squares (RLS) and describe various iterative algorithms for computing the inverse solution. Finally, in section 4 we evaluate different reconstruction methods in terms of computation times using *in vivo* bioluminescent imaging data from a mouse with an implanted brain tumor.

## 2. Forward problem

In the forward problem one needs to predict the photon flux from the animal surface for a given bioluminescent source distribution using known optical scatter and absorption properties within the animal. Since iterative solutions to the inverse problem require multiple solutions to the forward problem, a computationally efficient forward solution is important.

### 2.1. Forward model

Photon propagation in turbid media can be described by the radiative transfer or Boltzmann transport equation (Arridge 1999). While this model can be used in bioluminescence tomography (Klose 2007), the high computation cost and detailed knowledge required of the media's optical properties limit its use in practice. In contrast, the diffusion approximation under the assumption of isotropic scattering has been used extensively (Arridge *et al* 1993, Schweiger *et al* 1995, Jiang 1998, Arridge 1999, Hielscher *et al* 1999, Gu *et al* 2004, Wang *et al* 2004, Cong *et al* 2005, Slavine *et al* 2006, Dehghani *et al* 2006, Comsa *et al* 2006, Soloviev 2007, Kuo *et al* 2007). The diffusion approximation is reasonably accurate in soft tissue in the near-infrared region where scattering dominates absorption (Arridge 1999, Shen *et al* 2007). We, therefore, focus on the forward model based on the following steady-state diffusion equation (Arridge 1999),

$$\{-\nabla \cdot \kappa(\mathbf{r}, \lambda) \nabla + \mu_a(\mathbf{r}, \lambda)\} \phi(\mathbf{r}, \lambda) = q(\mathbf{r}, \lambda), \quad \text{for } \mathbf{r} \in \Omega, \quad (1)$$

subject to a Robin boundary condition

$$\phi(\mathbf{r}, \lambda) + 2\kappa(\mathbf{r}, \lambda) G \hat{\mathbf{n}}(\mathbf{r}) \cdot \nabla \phi(\mathbf{r}, \lambda) = 0 \quad \text{for } \mathbf{r} \in \partial\Omega \quad (2)$$

where  $q$  represents the bioluminescent source distribution,  $\phi$  denotes the photon density,  $\mathbf{r} \in \mathbb{R}^3$  denotes the location vector,  $\Omega \subset \mathbb{R}^3$  is the animal volume,  $\partial\Omega$  is the animal surface,  $\hat{\mathbf{n}}$  is a unit vector pointed outwardly normal to the surface  $\partial\Omega$ ,  $\lambda$  is the wavelength and  $\kappa(\mathbf{r}, \lambda) = 1 / \left[ 3 \left\{ \mu_a(\mathbf{r}, \lambda) + \mu'_s(\mathbf{r}, \lambda) \right\} \right]$ , with  $\mu_a$  and  $\mu'_s$  being the absorption and reduced scattering coefficients, respectively. In the boundary condition (2),  $G$  is a parameter modeling internal reflection at the boundary and can be computed as  $G = (1 + \gamma) / (1 - \gamma)$ , where  $\gamma \approx -1.4399n_{\text{int}}^{-2} + 0.7099n_{\text{int}}^{-1} + 0.6681 + 0.0636n_{\text{int}}$  with  $n_{\text{int}}$  being the refractive index of tissue (Schweiger *et al* 1995). The measurable photon flux at  $\mathbf{r} \in \partial\Omega$  is given by  $m(\mathbf{r}, \lambda) = \phi(\mathbf{r}, \lambda) / (2G)$ . The mapping from the source distribution  $q$  in the volume to the photon flux  $m$  through the surface can be obtained by solving (1) for  $\phi$  given  $q$ .

Approaches to solving the diffusion equation (1) include analytical methods that assume simplified geometry and homogeneous tissue (Rice *et al* 2001) and the more general finite element method (FEM) (Arridge *et al* 1993, 2000, Schweiger and Arridge 1997, Jiang 1998, Arridge 1999, Roy and Sevick-Muraca 2001, Gu *et al* 2004, Cong *et al* 2005, Alexandrakis *et al* 2005, Chaudhari *et al* 2005, Lv *et al* 2006). Although the analytical method is more computationally efficient, simplifying assumptions about geometry and optical properties can produce inaccurate results (Chaudhari *et al* 2005). Here we focus on the FEM-based forward model.

In the FEM framework, the 3D animal volume  $\Omega$  is discretized into tetrahedral elements connected at  $v$  vertex nodes. The source  $q$  and the photon density  $\phi$  are also discretized using finite element basis functions and are represented by  $\mathbf{q} \in \mathbb{R}^v$  and  $\phi \in \mathbb{R}^v$ , respectively. Then the problem of solving the diffusion equation reduces to

$$\mathbf{F}\phi = \mathbf{q} \quad (3)$$

where  $\mathbf{F} \in \mathbb{R}^{v \times v}$ . The FEM matrix  $\mathbf{F}$  is a function of the optical absorption and reduced scattering coefficients and refractive index. The optical properties can be specified using published measurements applied to a segmented MR or CT volume (Wang *et al* 2004, 2006a, Chaudhari *et al* 2005, Alexandrakis *et al* 2005, 2006, Lv *et al* 2006, Allard *et al* 2007). A detailed description of how to construct  $\mathbf{F}$  can be found in Arridge *et al* (2000). The FEM matrix  $\mathbf{F}$  is symmetric (Arridge *et al* (2000), equations (29)–(34)). It is also sparse with on the order of 10 nonzero elements per row.

## 2.2. Forward solution method

There are two standard methods for solving the inversion problem in (3) for  $\phi$ : (1) Cholesky factorization and substitution (CFS) and (2) preconditioned conjugate gradient (PCG) (Arridge *et al* 1993, Davies *et al* 1997, Schweiger and Arridge 1997, Arridge 1999). See Schweiger and Arridge (1997) for an overview of those methods and an analysis of their computational complexities. We chose CFS for our forward problem solver since CFS was faster than PCG in our experiments using a realistic FEM mesh; a comparison can be found in Chaudhari (2006) and Ahn *et al* (2007a). The reason that CFS was more efficient in our case is that if Cholesky factorization is done once, subsequent forward and back-substitutions can be inexpensively performed for multiple sources, as noted in Schweiger and Arridge (1997). In CFS,  $\mathbf{F}$  is first decomposed into Cholesky factors as  $\mathbf{F} = \mathbf{U}' \mathbf{U}$ , where  $\mathbf{U}$  is an upper triangular matrix and  $'$  represents matrix or vector transpose. Next one solves  $\mathbf{U}' \mathbf{c} = \mathbf{q}$  for  $\mathbf{c} \in \mathbb{R}^v$  by forward substitution and then solves  $\mathbf{U}\phi = \mathbf{c}$  for  $\phi$  by back-substitution. The Cholesky factorization to calculate  $\mathbf{U}$  is a one-time computation for a given FEM model  $\mathbf{F}$  and the forward and back-substitution can be repeatedly computed for different  $\mathbf{q}$ 's using the precomputed Cholesky factor  $\mathbf{U}$ .

## 2.3. System matrix

Let  $\mathbf{w}_i \in \mathbb{R}^v$  be the load vector for a unit source at voxel  $i$  and  $\mathbf{W} = [\mathbf{w}_1 \mathbf{w}_2 \dots \mathbf{w}_p] \in \mathbb{R}^{v \times p}$  be a matrix of the load vectors. Then the source representation  $\mathbf{x} \in \mathbb{R}^p$  using the load

vectors as a basis, and the corresponding source  $\mathbf{q} \in \mathbb{R}^v$  in tessellation nodes have the following relationship:

$$\mathbf{q} = \mathbf{W}\mathbf{x}. \quad (4)$$

The data  $\mathbf{y} \in \mathbb{R}^N$ , which denote the photon flux through the animal surface as measured by a CCD camera, are given by

$$\mathbf{y} = \mathbf{D}\phi \quad (5)$$

where  $\mathbf{D} \in \mathbb{R}^{N \times v}$  maps the photon density in the volume to the CCD camera measurements. Typically,  $p < v$  and  $N < v$ ; therefore,  $\mathbf{W}$  is a thin matrix and  $\mathbf{D}$  is wide.

Combining (3), (4) and (5), we construct a system matrix  $\mathbf{A} \in \mathbb{R}^{N \times p}$  that transforms the source  $\mathbf{x}$  into the measurement data  $\mathbf{y}$  such that

$$\mathbf{y} = \mathbf{A}\mathbf{x} = \mathbf{D}\mathbf{F}^{-1}\mathbf{W}\mathbf{x}. \quad (6)$$

The matrices  $\mathbf{D}$  and  $\mathbf{W}$  are sparse so the cost for multiplying a vector by  $\mathbf{D}$  or  $\mathbf{W}$  is negligible. Note that the system matrix  $\mathbf{A}$  is not sparse whereas  $\mathbf{F}$  is. To compute  $\mathbf{F}^{-1}\mathbf{q}$  for some  $\mathbf{q}$ , one needs to solve the forward problem (3), as discussed in section 2.2.

## 2.4. Forward and back-projector implementation approach

The two important operations involving the system matrix  $\mathbf{A}$  are 'forward projection' and 'back-projection,' referred to jointly below as F/B projection. This terminology is borrowed from computed tomography (Herman 1980) although a mapping from a source to measurement data analogous to the system matrix was called the projection operator and the data were called the projection data in the context of optical tomography in Arridge and Schweiger (1998). For forward projection one multiplies a vector by the system matrix, that is, one computes  $\mathbf{y} = \mathbf{A}\mathbf{x}$  for some  $\mathbf{x} \in \mathbb{R}^p$ . For back-projection one computes  $\mathbf{x} = \mathbf{A}'\mathbf{y}$  for some  $\mathbf{y} \in \mathbb{R}^N$ . One needs to compute F/B projection multiple times in iterative algorithms. Since the time required for F/B projection is a major contributor to the total reconstruction time, it is important to implement F/B projection operators efficiently. We investigate two approaches to implementing F/B projectors: a straightforward scheme described below we call the 'direct calculation' approach and a new 'on-the-fly' method.

**2.4.1. Direct calculation approach**—In the direct calculation approach, one precomputes the full system matrix  $\mathbf{A}$  and implements the F/B projector by a matrix-vector multiplication. Each column of the system matrix is computed by taking a unit vector corresponding to each source location as follows:

compute Cholesky factor  $U$   
 for  $j=1, \dots, p$   
 $\hat{q} = W e^j$   
 solve  $U' U \hat{\phi} = \hat{q}$  for  $\hat{\phi}$  by forward and back-substitution (7)  
 $\hat{y} = D \hat{\phi}$   
 store  $\hat{y}$  as the  $j$ th column of  $A$   
 end

where  $e^j = [0, \dots, 0, 1, 0, \dots, 0]' \in \mathbb{R}^p$  is the  $j$ th unit vector with only the  $j$ th entry being 1. The substitution step in (7) must be repeated for a total of  $p$  source locations. However, once  $A$  is precomputed, F/B projection can be implemented as direct matrix-vector multiplications,  $Ax$  and  $A'y$ .

**2.4.2. On-the-fly approach**—An alternative that avoids precomputing the full system matrix, is to calculate the F/B projection in an on-the-fly manner. The on-the-fly forward projector can be implemented as follows:

compute Cholesky factor  $U$   
 if “ $Ax$ ” needs to be computed for some  $x \in \mathbb{R}^p$   
 $\hat{q} = Wx$   
 solve  $U' U \hat{\phi} = \hat{q}$  for  $\hat{\phi}$  by forward and back-substitution (8)  
 $\hat{y} = D \hat{\phi}$ , which yields “ $Ax$ ”  
 end.

Similarly, in view of the symmetry of  $F$ , one can compute the back-projection on the fly as follows:

compute Cholesky factor  $U$   
 if “ $A'y$ ” needs to be computed for some  $y \in \mathbb{R}^N$   
 $\hat{\phi} = D'y$   
 solve  $U' U \hat{q} = \hat{\phi}$  for  $\hat{q}$  by forward and back-substitution (9)  
 $\hat{x} = W'\hat{q}$ , which yield “ $A'y$ ”  
 end.

Every time we compute a forward or back-projection in the on-the-fly approach, we need to perform the substitution step only once.

**2.4.3. Comparison**—In the direct calculation approach, it is very expensive to precompute the full system matrix  $A \in \mathbb{R}^{N \times p}$  because the forward problem must be solved  $p$  times where  $p$  is the number of source voxels. However, once the matrix has been precomputed, the F/B projection can be computed rapidly as a single matrix vector product. On the other hand, in the on-the-fly approach, whenever one needs to compute the forward or back-projection, one must solve the forward problem in (3), which is more expensive than computing the product of the system matrix and a vector, partly because the FEM matrix  $F \in \mathbb{R}^{v \times v}$  is larger than the system matrix  $A \in \mathbb{R}^{N \times p}$  (typically,  $N < v$  and  $p < v$ ) despite the sparsity of  $F$ . For example, in the multispectral bioluminescence tomography system we consider in section 4,  $v^2$  was larger than  $Np$  by a factor of  $>150$ .

Which of the direct calculation and on-the-fly approaches should be used? This depends on how many images are to be reconstructed for a given forward model, how many iterations per image are required for practical convergence of an iterative algorithm and how many F/B projections per iteration are needed. If one wishes to investigate many different iterative algorithms and regularization schemes for a given subject or if one assumes that the animal's anatomical structure and position do not change much over time in a longitudinal study, the direct calculation approach will be beneficial because once the expensive precomputation is performed, subsequent forward solutions can be rapidly computed by matrix vector multiplication. In contrast, if one is to compute relatively few images with a small number of iterations, the on-the-fly approach should be chosen. Comparisons of image reconstruction times for different approaches using real mouse bioluminescent data are made in section 4.

One of the advantages of the direct calculation approach is that direct access to any column or row in the system matrix is possible when the matrix is precomputed. Column or row access is required by certain iterative algorithms such as coordinate descent and incremental gradient algorithms, as discussed in section 3. Furthermore, the direct calculation approach may be more readily parallelizable than the on-the-fly approach.

Interestingly, the on-the-fly approach is closely connected with the adjoint or reverse differentiation method which is widely used for computing the gradient of a cost function in nonlinear diffuse optical tomography (Davies *et al* 1997, Hielscher *et al* 1999), as discussed in the appendix.

## 2.5. Multispectral system model

In multispectral imaging with  $K$  spectral bins, a multispectral system matrix  $\mathbf{A}_{\text{mul}}$  is constructed by concatenating monochromatic system matrices  $\mathbf{A}_1(\lambda_1), \dots, \mathbf{A}_K(\lambda_K)$  for wavelengths  $\lambda_1, \dots, \lambda_K$  (Chaudhari *et al* 2005):

$$\mathbf{y}_{\text{mul}} = \mathbf{A}_{\text{mul}} \mathbf{x} \quad (10)$$

where  $\mathbf{A}_{\text{mul}} = [s_1 \mathbf{A}_1(\lambda_1)' \dots s_K \mathbf{A}_K(\lambda_K)']' \in \mathbb{R}^{NK \times p}$  with  $s_1, \dots, s_K$  being known emission spectra of the bioluminescent source, and  $\mathbf{y}_{\text{mul}} = [\mathbf{y}_1(\lambda_1)' \dots \mathbf{y}_K(\lambda_K)']' \in \mathbb{R}^{NK}$  the multispectral data. It is straightforward to apply the F/B projection methods discussed in section 2.4 by combining individual F/B projections for each wavelength. That is, the forward projection is given by

$$\mathbf{y}_{\text{mul}} = \mathbf{A}_{\text{mul}} \mathbf{x} = [s_1 (\mathbf{A}_1(\lambda_1) \mathbf{x})' \dots s_K (\mathbf{A}_K(\lambda_K) \mathbf{x})']'$$

and the back-projection

$$\mathbf{x} = \mathbf{A}'_{\text{mul}} \mathbf{y}_{\text{mul}} = \sum_{k=1}^K s_k \mathbf{A}_k(\lambda_k)' \mathbf{y}_k(\lambda_k).$$

Henceforth, we omit the subscript in  $\mathbf{A}_{\text{mul}}$  and  $\mathbf{y}_{\text{mul}}$  for notational simplicity and consider both monochromatic and multispectral systems in a common framework.

### 3. Inverse problem

The ultimate goal of bioluminescence tomography is to estimate the source  $\mathbf{x}$  from the photon flux measurement  $\mathbf{y}$  based on the system model (6) or (10). In this section we formulate the inverse problem and discuss iterative algorithms for the inverse solution.

#### 3.1. Regularized least squares (RLS)

We focus on the following regularized least squares (RLS) estimate:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \geq 0} \phi(\mathbf{x}), \quad \phi(\mathbf{x}) = L(\mathbf{x}) + \beta R(\mathbf{x}). \quad (11)$$

The nonnegativity constraint,  $\mathbf{x} \geq \mathbf{0}$ , is imposed because the source strength is physically nonnegative (Jiang and Wang 2004, Jiang *et al* 2007). Although the problem size can be reduced by allowing sources only in a permissible region (Cong *et al* 2004, 2005, Jiang and Wang 2004, Wang *et al* 2006a, 2006b, Cong and Wang 2006, Jiang *et al* 2007, Lv *et al* 2007), we do not consider such constraints here. In (11),  $L$  is the data-fit function

$$L(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2, \quad (12)$$

$\beta$  is a regularization parameter, and  $R$  is a quadratic regularization function such that

$$R(\mathbf{x}) = \frac{1}{2} \mathbf{x}' \mathbf{R} \mathbf{x} \quad (13)$$

for a nonnegative definite matrix  $\mathbf{R} \in \mathbb{R}^{p \times p}$ . The regularizer  $R$  stabilizes the noise in reconstructed images yet usually introduces bias. In a Bayesian framework, if the source and the additive noise are assumed to be zero-mean Gaussian with covariance of  $\mathbf{K}$  and  $\sigma^2 \mathbf{I}$ , respectively, then the RLS estimate  $\hat{\mathbf{x}}$  obtained with choosing  $\beta = \sigma^2$  and  $\mathbf{R} = \mathbf{K}^{-1}$  can be viewed as a maximum *a posteriori* (MAP) estimate, where  $\mathbf{I}$  is the identity matrix (Tarantola 2005, chapter 3). Other noise models such as the shot-noise model (Ye *et al* 1999) can also be used; other types of regularization functions include generalized Gaussian Markov random field priors (Bouman and Sauer 1993), non-Gaussian Gibbs priors (Hebert and Leahy 1989), which have an edge-preserving property, and L1-norm regularizers, which are known to produce sparse reconstruction (Cao *et al* 2007). We focus here on the quadratic cost function using (12) and (13). However, we note that the optimization methods described below are general, and can, with certain restrictions on differentiability and continuity, be used in conjunction with the nonquadratic regularizers referred to above.

We use a weighted L2-norm regularizer with a positive-definite diagonal matrix  $\mathbf{R}$  (Lin *et al* 2006). If  $\mathbf{R}$  is the identity matrix, then  $R(\mathbf{x})$  becomes a conventional L2-norm regularizer (Cong and Wang 2006, Han *et al* 2006, Wang *et al* 2006a, Lv *et al* 2006, Han and Wang 2007, Soloviev 2007). In this case, as the regularization parameter  $\beta$  increases, the reconstructed image becomes smooth and less noisy. But the reconstructed sources also tend



to become superficial, consequently increasing localization errors particularly for deep sources. This occurs because deep sources are highly attenuated so need larger amplitudes to match the surface photon flux from a superficial source; these large amplitudes are in turn penalized by the L2-norm. Quadratic weighting schemes have been proposed to reduce the preference for superficial sources in the related problem of cortical current imaging in EEG and MEG (Lin *et al* 2006, Baillet *et al* 2001). Nonquadratic penalties can also have attractive properties in this respect.

We use the following sensitivity-dependent weight matrix, which through column normalization removes the increased penalty for deep sources, effectively removing the bias towards superficial ones:

$$\mathbf{R} = \text{diag}_j \{ \gamma_j^2 \} \quad (14)$$

where  $\text{diag}\{\cdot\}$  denotes a diagonal matrix,  $\gamma_j$  is the sensitivity of the  $j$ th source location defined as

$$\gamma_j = \sum_i a_{ij}, \quad (15)$$

and  $a_{ij}$  are the entries in the system matrix  $\mathbf{A}$ . One can assume  $\gamma_j > 0$  for all  $j$  without loss of generality since one can exclude any source  $x_j$  with  $\gamma_j = 0$  from the parameter space. Note that  $\gamma_j$ 's are readily calculated using an on-the-fly back-projection  $\mathbf{A}'\mathbf{1}$  of uniform data  $\mathbf{1} = [1, \dots, 1]'$ , without precomputing  $\mathbf{A}$ .

The cost function in (11) with a positive-definite matrix  $\mathbf{R}$  is strictly convex and so there exists a unique solution  $\hat{x}$  of the nonnegativity-constrained optimization problem in (11) and a local minimum is also the global minimum (Bertsekas 1999, p 685). In the continuous domain, the uniqueness of a solution to the diffusion equation under certain conditions on the source is discussed in Wang *et al* (2004) for the unregularized case and in Han *et al* (2006) and Han and Wang (2007) for the regularized one.

Most of the iterative algorithms that we study in section 3.2 use the gradient  $\nabla\Phi$  of the cost function  $\Phi$ . The gradient is given by

$$\nabla\Phi(\mathbf{x}) = (\mathbf{A}'\mathbf{A} + \beta\mathbf{R})\mathbf{x} - \mathbf{A}'\mathbf{y}, \quad (16)$$

calculation of which requires F/B projections. The Hessian  $\mathbf{H}$  of the cost function is given by

$$\mathbf{H} = \nabla^2\Phi(\mathbf{x}) = \mathbf{A}'\mathbf{A} + \beta\mathbf{R}. \quad (17)$$

The diagonal elements of the Hessian are used in certain preconditioners in section 3.2.5.

The quality of the reconstructed image  $\hat{x}$  in (11), as reflected in source localization error, quantitation error, resolution and noise characteristics, is determined by the choice of cost function  $\Phi$  as well as the physical properties and limitations of the imaging system.

Inaccuracies in the system model  $A$  will introduce a systematic bias into the image. An inaccurate noise model, for which we assume a zero-mean white Gaussian distribution, can lead to suboptimal noise characteristics as well as additional bias. Image quality will also be affected by choice of the regularization function  $R$ . However, we note that our goal here is to explore methods for solving the inverse problem once the cost function  $\Phi$  is specified, rather than to investigate the relative merits of different cost functions. In section 4 we use a single regularization function which produces reasonably good localization results for the data used in our evaluations. We then compare the convergence speeds of different reconstruction algorithms that all converge to the same optimal solution for the chosen cost. Since all algorithms converge to the same solution we do not present comparisons of image quality in this paper.

### 3.2. Iterative algorithms

We will investigate iterative solutions to the optimization problem (11) since the nonnegativity constraint precludes an analytical solution. We choose four representative iterative algorithms: a gradient projection method (GPM), preconditioned conjugate gradient (PCG), coordinate descent (CD) and an incremental gradient algorithm, OS-SPS. We compare their convergence speeds using real mouse bioluminescent data in section 4. Those algorithms are described in detail below.

A number of different algorithms for bioluminescence tomography have been used in the literature. The expectation maximization (EM) algorithm (Dempster *et al* 1977) which was originally developed to compute a maximum likelihood solution for linear Poisson inverse problems (Shepp and Vardi 1982) was used in Jiang and Wang (2004), Alexandrakis *et al* (2005, 2006) and Jiang *et al* (2007). The EM algorithm maximizes an unregularized Poisson likelihood function and its convergence speed is generally very slow (Meng and van Dyk 1997, Qi and Leahy 2006) so we do not consider it in this paper. A deblurring variant of EM was used in Slavine *et al* (2006). A regularized Newton's algorithm (Gu *et al* 2004) and a modified Newton's method with an active set strategy (Cong *et al* 2004, 2005, Lv *et al* 2006) have also been used. A major disadvantage of Newton's methods for constrained optimization is that it is computationally costly both to enforce constraints and guarantee convergence to the solution (Bertsekas 1999, p 231); for example, naive orthogonal projection onto the nonnegative constraint set does not necessarily guarantee a decrease in the cost function (Bertsekas 1999, p 245). A constrained Landweber method was used in Jiang and Wang (2004), which requires nontrivial stepsize selection for both global convergence and optimal convergence rates.

The nonnegativity constraint in (11) can be more easily enforced by the four algorithms (GPM, PCG, CD and OS-SPS) considered here than by using active set methods (Cong *et al* 2004, 2005, Lv *et al* 2006, 2007). In GPM and PCG, the nonnegativity constraint is enforced by orthogonal projection onto the constraint set and a subsequent bent line search (Mumcuo lu *et al* 1994, Qi *et al* 1998); in CD and OS-SPS, the optimization problem is reduced to a set of one-dimensional constrained problems, which makes it trivial to enforce the constraint. All the algorithms are aimed at minimizing the common RLS cost function in (11) and converge to the optimal point, with the exception of OS-SPS which approaches an

approximate solution as stated below. None of these algorithms has any free parameters, such as step size, so that we can directly compare convergence rates for solving the problem in (11).

**3.2.1. Gradient projection method (GPM)**—A gradient projection method (GPM) is an iterative algorithm for constrained problems analogous to a gradient descent method for unconstrained problems (Bertsekas 1999, p 223). The gradient projection method we focus on here can be summarized as follows: for  $n = 0, 1, \dots$ ,

$$\mathbf{g}^n = \nabla \Phi(\mathbf{x})^n \quad (18)$$

$$\mathbf{d}^n = -\mathbf{P}^n \mathbf{g}^n \quad (19)$$

$$\alpha_n = \arg \min_{\alpha} \Phi(\mathbf{x}^n + \alpha \mathbf{d}^n) = -\frac{\mathbf{d}^{n'} \mathbf{g}^n}{\mathbf{d}^{n'} \mathbf{H} \mathbf{d}^n} \quad (20)$$

$$\mathbf{z}^n = \mathbf{x}^n + \alpha_n \mathbf{d}^n$$

$$\begin{aligned} & \text{if } [\mathbf{z}^n]_+ \neq \mathbf{z}^n \\ & \quad \mathbf{d}^n = [\mathbf{z}^n]_+ - \mathbf{x}^n \\ & \quad \alpha_n = \arg \min_{\alpha \leq 1} \Phi(\mathbf{x}^n + \alpha \mathbf{d}^n) = \min \left( -\frac{\mathbf{d}^{n'} \mathbf{g}^n}{\mathbf{d}^{n'} \mathbf{H} \mathbf{d}^n}, 1 \right) \quad (21) \\ & \text{end} \end{aligned}$$

$$\mathbf{x}^{n+1} = \mathbf{x}^n + \alpha_n \mathbf{d}^n \quad (22)$$

where  $\mathbf{x}^n \in \mathbb{R}^p$  is the  $n$ th iterate,  $\mathbf{x}^0$  is an initial estimate,  $\mathbf{P}^n \in \mathbb{R}^{p \times p}$  is a preconditioner,  $\nabla \Phi$  and  $\mathbf{H}$  are given in (16) and (17), respectively, and  $[\cdot]_+$  denotes the orthogonal projection onto the nonnegative orthant, that is,  $\mathbf{q} = [\mathbf{t}]_+$  implies  $q_j = \min(t_j, 0)$  for all  $j$ . We describe preconditioners for this method in section 3.2.5. The above algorithm can be implemented by keeping  $\mathbf{Ax}$  as a state vector so that GPM requires one F/B projection pair per iteration unless  $\mathbf{z}^n$  leaves the constraint set. Calculating the gradient in (18) and (16) requires one back-projection if  $\mathbf{Ax}^n$  is retained as a state vector, and calculating the denominator in (20) needs one forward projection since  $\mathbf{d}^{n'} \mathbf{H} \mathbf{d}^n = \|\mathbf{A} \mathbf{d}^n\|^2 + \beta \mathbf{d}^{n'} \mathbf{R} \mathbf{d}^n$ . If  $[\mathbf{z}^n]_+ \neq \mathbf{z}^n$ , then an additional forward projection is needed in (22) because  $\mathbf{d}^n$  is updated in (21).

**3.2.2. Preconditioned conjugate gradient (PCG)**—A PCG method is a general purpose optimization algorithm using a conjugate direction as a search direction (Bertsekas 1999, p 138). To enforce the nonnegativity constraint, we use a constrained version of PCG where the search direction is bent when an iterate leaves the constraint set, as in section 3.2.1 (Mumcuo lu *et al* 1994, Qi *et al* 1998). The version we use can be summarized as follows: for  $n = 0, 1, \dots$ ,

$$\mathbf{g}^n = \nabla \Phi(\mathbf{x}^n)$$

$$\mathbf{r}^n = \mathbf{P}^n \mathbf{g}^n$$

$$\eta^n = \begin{cases} \frac{\mathbf{r}^{n'} (\mathbf{g}^n - \mathbf{g}^{n-1})}{\mathbf{r}^{n-1'} \mathbf{g}^{n-1}}, & n > 1 \\ 0, & n = 0 \end{cases}$$

$$\mathbf{d}^n = -\mathbf{r}^n + \eta^n \mathbf{d}^{n-1}$$

$$\text{if } \mathbf{d}^{n'} \mathbf{g}^n > 0, \text{ then } \mathbf{d}^n = -\mathbf{r}^n$$

$$\alpha_n = \arg \min_{\alpha} \Phi(\mathbf{x}^n + \alpha \mathbf{d}^n) = -\frac{\mathbf{d}^{n'} \mathbf{g}^n}{\mathbf{d}^{n'} \mathbf{H} \mathbf{d}^n}$$

$$\mathbf{z}^n = \mathbf{x}^n + \alpha_n \mathbf{d}^n$$

$$\text{if } [\mathbf{z}^n]_+ \neq \mathbf{z}^n$$

$$\mathbf{d}^n = [\mathbf{z}^n]_+ - \mathbf{x}^n$$

$$\alpha_n = \arg \min_{\alpha \leq 1} \Phi(\mathbf{x}^n + \alpha \mathbf{d}^n) = \min \left( -\frac{\mathbf{d}^{n'} \mathbf{g}^n}{\mathbf{d}^{n'} \mathbf{H} \mathbf{d}^n}, 1 \right)$$

end

$$\mathbf{x}^{n+1} = \mathbf{x}^n + \alpha_n \mathbf{d}^n. \quad (23)$$

The PCG method requires one F/B projection pair per iteration with an extra forward projection, depending on whether  $\mathbf{z}^n$  leaves the constraint set.

**3.2.3. Coordinate descent (CD)**—A coordinate descent (CD) method is an iterative algorithm that minimizes the cost function along one coordinate direction for each update (Bertsekas 1999, p 267). The method has been successfully applied to image reconstruction applications (Bouman and Sauer 1996, Ye *et al* 1999). The CD method can be implemented as follows:

$$\mathbf{r} = \mathbf{y} - \mathbf{A}\mathbf{x}^0 \text{ for initial image } \mathbf{x}^0 \quad (24)$$

```

for  $n=1, 2, \dots$ 
  for  $j=1, \dots, p$ 
     $x_j^{\text{old}}=x_j$ 
     $x_j := \left[ x_j - \frac{(\nabla\Phi(\mathbf{x}))' \mathbf{e}^j}{\mathbf{H}_{jj}} \right]_+$  (25)
     $\mathbf{r} := \mathbf{r} + (x_j^{\text{old}} - x_j) \mathbf{A}_{.j}$ 
  end
   $\mathbf{x}^n = \mathbf{x}$ 
end

```

where  $\mathbf{e}^j$  is the  $j$ th unit vector and  $\mathbf{A}_{.j}$  is the  $j$ th column of  $\mathbf{A}$ . In (24), the  $j$ th diagonal element  $H_{jj}$  of the Hessian  $\mathbf{H}$  in (17) is given by

$$\mathbf{H}_{jj} = \xi_j + \beta \mathbf{R}_{jj} \quad (26)$$

where

$$\xi_j = \sum_i a_{ij}^2, \quad \mathbf{R}_{jj} = \gamma_j^2 \quad (27)$$

from (14), (15), and (17). The CD method requires access to each column  $\mathbf{A}_{.j}$  of the system matrix; therefore, the on-the-fly approach is not suitable for CD.

**3.2.4. Incremental gradient methods**—Incremental gradient methods (Bertsekas 1999, p 108) or ordered subset (OS) algorithms (Hudson and Larkin 1994) are optimization techniques which use only a subset of the data for each update with the aim of reducing the cost for computing gradients. Various incremental gradient methods have been successfully applied to PET/SPECT image reconstruction to accelerate convergence speeds (Hudson and Larkin 1994, Qi and Leahy 2006). The applicability of an OS algorithm to bioluminescence tomography was mentioned in Jiang and Wang (2004). In this paper we consider an OS version of separable paraboloidal surrogates (OS-SPS) where a line search is not required (Erdo an and Fessler 1999). We derive the OS-SPS algorithm for our problem by following the derivation of additive updates in Erdo an and Fessler (1999):

```

for  $n=0, 1, 2, \dots$ ,
   $x^{n,1} = x^n$ 
  for  $m=1, 2, \dots, M$ 
     $x^{n,m+1} = \left[ x^{n,m} - M \mathbf{P}^{\text{SPS}} \nabla \Phi_m(x^{n,m}) \right]_+$ 
  end
   $x^{n+1} = x^{n,M+1}$ 
end

```

where

$$\mathbf{P}^{\text{SPS}} = \text{diag}_j \left\{ \frac{1}{\sum_i a_{ij} \sum_k a_{ik} + \beta \mathbf{R}_{jj}} \right\}$$

and

$$\Phi_m(x) = \frac{1}{2} \sum_{i \in S_m} (y_i - [Ax]_i)^2 + \frac{1}{M} R(x) \quad (28)$$

for  $m = 1, \dots, M$ , with  $M$  being the number of subsets  $S_m$  or subcost functions  $\Phi_m$ . We choose the subsets  $S_m$  in (28) by downsampling the data space as  $S_m = \{i : i \equiv m - 1 \pmod{M}, i = 1, \dots, KN\}$ .

Generally, OS-SPS does not converge to an optimal solution of the problem in (11) but to a limit cycle of points which are only approximate solutions (Ahn *et al* 2006). The OS type algorithms can be forced to converge to the optimal solution by using relaxation parameters (Ahn and Fessler 2003), switching to a convergent algorithm at some point (Li *et al* 2005), or rederiving the algorithm in a framework of incremental optimization transfer (Ahn *et al* 2006). However, we use the original form of OS-SPS as outlined above for simplicity because it works well enough for our purposes, as shown in section 4. Since OS algorithms, including OS-SPS, require access to each row of the system matrix  $\mathbf{A}$  to compute the gradient  $\nabla\Phi_m$  of the subcost function, the on-the-fly approach is not suitable.

**3.2.5. Preconditioners**—For ill-conditioned problems, gradient based methods suffer from slow convergence but can be accelerated using preconditioning. Newton's method using the inverse of the Hessian as a preconditioner converges very quickly (Bertsekas 1999, p 26). However, calculating the Hessian inverse is impractical for large-scale problems. Here we consider only simple diagonal preconditioners.

A popular choice is a diagonal approximation to Newton's method (Bertsekas 1999, p 27),

$$\mathbf{P}^N = \text{diag}_j \left\{ \frac{1}{\mathbf{H}_{jj}} \right\} = \text{diag}_j \left\{ \frac{1}{\xi_j + \beta \mathbf{R}_{jj}} \right\}, \quad (29)$$

with  $\xi_j$  and  $\mathbf{R}_{jj}$  defined in (27). Calculating the term  $\xi_j = \sum_i a_{ij}^2$  requires precomputation of the full system matrix  $\mathbf{A}$  so the on-the-fly approach is not suitable for this preconditioner.

However, one can estimate  $\xi_j$  from  $\gamma_j^2$  where  $\gamma_j \triangleq \sum_i a_{ij}$  since  $\xi_j$  and  $\gamma_j^2$  are usually highly correlated and  $\gamma_j$  can be readily computed on the fly by back-projecting a vector of ones as discussed in section 3.1. Figure 1 shows an example scatter plot of  $\xi_j$  versus  $\gamma_j^2$  for the multispectral imaging system model from section 4. As shown in the figure,  $\xi_j$  and  $\gamma_j^2$  are strongly correlated (Pearson's correlation coefficient was 0.922). A practical strategy for estimating  $\xi_j$ , which works for the on-the-fly approach, is as follows. First, one computes  $\gamma_j$  for  $j = 1, \dots, p$  by back-projection  $\mathbf{A}'\mathbf{1}$ . Next, one randomly chooses  $T$  source locations  $j_1, \dots, j_T$ , and calculates projections  $\mathbf{y}(t) = \mathbf{A}e^{j_t t}$  and subsequently  $\xi_{j_t} = \|\mathbf{y}(t)\|^2 = \sum_i a_{ij_t}^2$  for  $t = 1, \dots, T$ . Then one estimates the slope  $\tau$  in the scatter plot by fitting a line  $\xi = \tau\gamma^2$  to the  $T$  samples  $\left\{ \left( \gamma_{j_t}^2, \xi_{j_t} \right) \right\}_{t=1}^T$ . Finally, one approximates  $\xi_j$  as  $\hat{\xi}_j = \hat{\tau}\gamma_j^2$  using the slope estimate  $\hat{\tau}$  for  $j = 1, \dots, p$ . We refer to the resulting preconditioner as the estimated diagonal approximation to Newton's (EN) method,

$$\mathbf{P}^{EN} = \text{diag}_j \left\{ \frac{1}{\hat{\xi}_j + \beta \mathbf{R}_{jj}} \right\}, \quad (30)$$

which can be applied in both direct calculation and on-the-fly approaches. Calculating this preconditioner requires one back-projection for computing  $\gamma_j^2$ 's and  $T$  forward projections for estimating  $\tau$ . In section 4, we choose  $T = 10$  source locations for estimating the slope  $\tau$ . The coefficient of variation (CV) of  $\hat{\tau}$  due to randomly choosing 10 source locations was about 29%, which is stable enough for our purposes (a very accurate estimate of  $\tau$  is not necessary).

For PCG we also consider an EM-type preconditioner as follows:

$$\mathbf{P}^{EM}(x) = \text{diag}_j \left\{ \frac{x_j + \epsilon}{\gamma_j} \right\}, \quad (31)$$

where  $\epsilon$  is a small positive number, e.g.,  $\epsilon = 10^{-3} \max(1, \max_l x_l) > 0$ . This preconditioner originates from the EM algorithm for maximizing a Poisson likelihood (Shepp and Vardi 1982) and has been used successfully to accelerate PCG algorithms for emission tomography (Kaufman 1987). It has also been used with PCG for bioluminescence tomography (Chaudhari *et al* 2005).

**3.2.6. Comparison**—The on-the-fly approach is not suitable for CD and OS-SPS, as discussed above, since they require access to each row or column of the system matrix; neither is it suitable for the diagonally approximated Newton preconditioner  $\mathbf{P}^N$  in (29) since calculating  $\xi_j$  needs the precomputed full system matrix. For GPM and PCG, using  $\mathbf{P}^{EN}$  in (30) or  $\mathbf{P}^{EM}$  in (31), both direct calculation and on-the-fly approaches can be applied.

The computation cost for the iterative algorithms is dominated by F/B projections. Therefore, the computational complexity of each algorithm can be represented by the number of F/B projections required per iteration. A forward or back-projection amounts to matrix-vector multiplication for the direct calculation approach, and to one FEM forward solution for the on-the-fly approach. GPM and PCG require one F/B projection pair plus possibly an extra forward projection. The cost per iteration for CD and OS-SPS equals roughly that of one F/B projection pair. Therefore, the computation costs per iteration for those algorithms considered here do not differ substantially once the F/B projection implementation approach is chosen. Other important factors are the cost for precomputation and the convergence speed, that is, how many iterations are required for practical convergence.

Memory requirements for the iterative algorithms are modest. During iterations, one needs a memory space  $O(p + KN)$  proportional to the image size  $p$  and to the data size  $KN$ . Additionally, for the direct calculation approach, one needs to store the  $KN \times p$  system matrix. On the other hand, for the on-the-fly approach, one must retain the  $v \times v$  FEM matrix, which is sparse and has about  $10v$  nonzero elements, for each wavelength if PCG is used for the forward problem solver; and one needs to keep the  $v \times v$  Cholesky factor, which

is a sparse upper triangular matrix, for each wavelength if CFS is chosen for the FEM solver.

The cost of all of these methods is dominated by the cost of (possibly partial) forward and back-projections. Since these operations all involve large matrix-vector multiplications for the direct calculation approach, parallelization using multithreading or other methods is relatively straightforward. Among the four methods, the full gradient methods (PCG and GPM) in conjunction with the direct calculation approach benefit most readily since the full forward and back-projection operations can be easily distributed across processors. The CD method which updates one variable at a time will benefit least from parallelization.

## 4. Results

We applied iterative reconstruction methods to real bioluminescent data obtained from a mouse with a brain tumor implanted in the right cerebral hemisphere. Our goal is to compare computation cost for different reconstruction methods. All methods described should converge to the RLS image  $\hat{x}$  in (11) with the exception of OS-SPS which may enter a limit cycle as described above.

Multispectral bioluminescent data were obtained from top and two side views by the IVIS 200 imaging system for wavelengths 580 nm, 600 nm, 620 nm and 640 nm. Since the tumor was near the dorsal surface of the head, no signal was detected from the bottom view and those three views (top and two sides) had sufficient information for localizing the source (Chaudhari *et al* 2005).

3D CT scan data were used to segment the anatomical volume into skull and soft tissue and to assign standard optical properties (Cheong *et al* 1990). The specific values of the absorption and reduced scattering coefficients for the skull and soft tissue we used can be found in Chaudhari *et al* (2005). A regularization parameter value of  $\beta = 0.05$  was determined empirically. All image reconstructions were performed on an AMD Opteron 870 2.0 GHz computer. For all the iterative algorithms, a zero uniform image,  $\mathbf{0} = [0, \dots, 0]^T \in \mathbb{R}^P$ , was used as an initial estimate  $\mathbf{x}^0$ .

There is a trade-off between the number of FEM mesh nodes and the accuracy of FEM model;  $10^5$ – $10^6$  nodes were recommended for sufficient accuracy in 3D problems (Schweiger and Arridge 1997). Here we used a FEM-based forward model with  $v = 71256$  tessellation nodes, a source space with  $p = 10814$  source locations, and  $N = 3085$  surface measurement nodes for  $K = 4$  spectral bins. We did not attempt to investigate different FEM models in terms of accuracy and computation cost since our purpose is to compare the computation speeds of different reconstruction methods for a given model; however, we achieved reasonably good localization results within a reasonable time, as shown below. See Chaudhari *et al* (2005) for a study of the effects of the number of surface measurement points on reconstruction results.

First, we precomputed the full system matrix  $A$  by using the `mldivide` function in Matlab (v7.3; Mathworks, MA, USA). We computed 500 columns of the system matrix at a time;



that is, we prepared a  $71256 \times 500$  matrix by combining 500 load vectors and solved the forward problem in (3) by putting the matrix on the right-hand side. It took 12 708 s to compute the full system matrix, as shown in table 1. Then we reconstructed the bioluminescence image with the system matrix using the direct calculation approach. The iterative algorithms were implemented in the C language. Table 1 shows, in the columns named iteration cost, the computation time in seconds and the number of iterations, required for each algorithm to achieve the relative error  $\mathcal{E} = \|x^n - \hat{x}\| / \|\hat{x}\| < 10\%$ ,  $< 5\%$ ,  $< 5\%$ , and  $< 1\%$ , respectively, where the RLS image  $\hat{x}$  in (11) was estimated by 2000 iterations of GPM-N. Note that the relative error  $\mathcal{E}$  is not with respect to the true source distribution which is unknown but with respect to the RLS solution  $\hat{x}$  given in (11). Therefore,  $\mathcal{E}$  represents how close an image is to the converged solution and can be used for evaluating the convergence speed of an iterative algorithm. In the table, GPM-U denotes the unpreconditioned GPM; GPM-N and GPM-EN denote GPM with preconditioner  $\mathbf{P}^N$  and  $\mathbf{P}^{EN}$  given in (29) and (30), respectively; CG, PCG-N, PCG-EN, and PCG-EM denote PCG with no preconditioner and with preconditioners  $\mathbf{P}^N$ ,  $\mathbf{P}^{EN}$ , and  $\mathbf{P}^{EM}$ , respectively; and OS-SPS-2, OS-SPS-5, and OS-SPS-10 denote OS-SPS with two, five and ten subsets, respectively. In table 1, the total reconstruction time is the sum of the precomputation time and the iteration cost. We did not include the computation cost for constructing the FEM matrix  $\mathbf{F}$ , which is common in all the methods.

Next, we reconstructed images using the on-the-fly approach. For the on-the-fly approach, the precomputation time (110 s) in the table represents the time required for Cholesky factorization. As discussed in section 3, we excluded CD and OS-SPS, and preconditioner  $\mathbf{P}^N$  for this approach. For the on-the-fly approach we implemented the iterative algorithms in Matlab. For the on-the-fly F/B projectors, the one-time computation for Cholesky factorization was performed by the chol function, and repeated substitutions were carried out by the mldivide function in Matlab.

As shown in table 1, for both GPM and PCG, the use of preconditioners  $\mathbf{P}^N$ ,  $\mathbf{P}^{EN}$  and  $\mathbf{P}^{EM}$  significantly accelerates convergence. In other words, the preconditioned algorithms (GPM-N, GPM-EN and PCG-N, PCG-EN, PCG-EM) require fewer iterations and lower iteration costs in seconds than the unpreconditioned ones (GPM-U and CG). Preconditioners  $\mathbf{P}^{EN}$  using the estimated diagonal elements of the Hessian and  $\mathbf{P}^N$  using the exact diagonals showed similar convergence speeds. In the direct calculation approach, CD and OS-SPS-10 were much faster than GPM and PCG. In fact, OS-SPS-10 showed the fastest convergence speed; it required only 8 and 15 iterations (11 and 20 s) for  $\mathcal{E} < 10\%$  and  $\mathcal{E} < 5\%$ , respectively. However, OS-SPS-10 (as well as OS-SPS-5) could not reach a point where  $\varepsilon < 1\%$  because of the limit cycle behavior. This is a typical characteristic of incremental gradient or OS algorithms; as the number of subsets increases the initial convergence speed becomes faster but the limit point moves further from the optimal solution. Therefore, when it comes to the number of subsets for OS algorithms, there is a trade-off between convergence speed and accuracy. For  $\varepsilon < 1\%$ , CD was fastest among the iterative algorithms for the direct calculation approach. On the other hand, for the on-the-fly approach, GPM-EN showed the fastest convergence rate for all three relative error cases, and PCG-EN showed a similar yet little slower convergence rate.

The precomputation time (12 708 s) for the direct calculation approach is much larger than that (110 s) for the on-the-fly approach whereas the cost per iteration for the direct approach is smaller than that for the on-the-fly approach. If few images are to be reconstructed for a given system, then the on-the-fly approach can substantially reduce the total computation cost. For example, for the relative error  $<10\%$ , GPM-EN, the fastest among the on-the-fly algorithms in total reconstruction time, is faster than OS-SPS-10, the fastest of the direct calculation algorithms, by a factor of  $12719/798 \approx 15.9$ ; similarly, for  $<5\%$ , by a factor of  $12728/1091 \approx 11.7$ . For  $\varepsilon < 1\%$ , GPM-EN in the on-the-fly approach is faster than CD in the direct calculation approach by a factor of  $12773/1724 \approx 7.4$ . As the number of required iterations increases, the factor of computational saving of the on-the-fly approach compared to the direct approach decreases.

Figure 2 shows the RLS image  $\hat{x}$  and figure 3 shows an image reconstructed by 29 iterations, taking 798 s in total reconstruction time, of GPM-EN, which is the fastest among the on-the-fly algorithms. Although the relative error between the images is about 10%, they look similar and the localization is reasonably accurate.

## 5. Conclusion

We have explored computationally efficient methods for fully 3D multispectral bioluminescence image reconstruction. When it comes to incorporating F/B projectors into iterative algorithms, we investigated the straightforward direct calculation approach and an on-the-fly approach where one does not have to precompute the system matrix. We evaluated those approaches combined with various iterative algorithms by using real mouse bioluminescence data. We achieved a substantial speed-up using the proposed on-the-fly approach by a factor of up to 15.9 when a single image reconstruction is required. However, if multiple image reconstructions are to be performed for a given model, the speed-up factor decreases and in this case the direct calculation approach may be more efficient. Using proper preconditioners accelerates convergence speeds significantly; the estimated diagonal approximation to Newton's preconditioner (EN), which applies to both direct calculation and on-the-fly approaches, reduced the iteration cost by a factor of  $>7$  for both GPM and PCG. In the on-the-fly approach, GPM-EN yielded the fastest convergence speed and PCG-EN showed a little slower convergence rates. For the direct calculation approach, OS-SPS, an incremental gradient algorithm, was the fastest one although it converged to an approximated RLS solution; CD was the second fastest, which converges to the RLS solution.

We note that the emphasis in this paper is on fast methods for computing inverse solutions. With the exceptions described above, all methods implemented should converge to the same solution. For this reason we have not explored measures of image quality such as bias, resolution and variance. The methods presented here represent the simplest (quadratic) error models and regularization functions. However, most of the ideas discussed can be extended to incorporate more sophisticated noise models and nonquadratic regularization functions. We also restricted our attention here to bioluminescence imaging. However, the methods can be extended to the related problem of reconstructing images in 3D fluorescence optical tomography (Ahn *et al* 2007b). In this case, the models must be extended to also include the

propagation of the excitation light through the animal. Since this process can also be approximated using the diffusion equation, similar fast FEM and on-the-fly methods can be developed.

## Acknowledgments

This work was supported by the National Cancer Institute under Grant 1R01CA121783.

## Appendix. Relationship between adjoint or reverse differentiation and F/B projection approaches

Suppose that  $F = F(\theta)$  and  $q = q(\theta)$  where  $F$  and  $q$  are defined in (3) and

$\theta = [\theta_1, \dots, \theta_S]' \in \mathbb{R}^S$  is an unknown parameter vector. In diffuse optical tomography,  $\theta$  represents the optical absorption and reduced scattering coefficients and  $q$  is a known constant in  $\theta$ . In bioluminescence tomography,  $\theta$  is the source distribution and  $F$  is a known constant. From (3), one can show

$$\frac{\partial \phi}{\partial \theta_s} = F^{-1} \left( \frac{\partial q}{\partial \theta_s} - \frac{\partial F}{\partial \theta_s} \phi \right) \quad (\text{A.1})$$

for  $s = 1, \dots, S$ . Let a data-fit function be defined as

$$L(\theta) = \frac{1}{2} \|Y - D\phi\|^2 \quad (\text{A.2})$$

where  $D$  is defined as in (5). Here our goal is to compute the gradient  $\nabla L(\theta)$  efficiently.

Using (A.1), one obtains

$$\begin{aligned} \frac{\partial L}{\partial \theta_s} &= -(\mathbf{y} - D\phi)' D F^{-1} \left( \frac{\partial q}{\partial \theta_s} - \frac{\partial F}{\partial \theta_s} \phi \right) \\ &= -\left\langle D'(\mathbf{y} - D\phi), F^{-1} \left( \frac{\partial q}{\partial \theta_s} - \frac{\partial F}{\partial \theta_s} \phi \right) \right\rangle \end{aligned} \quad (\text{A.3})$$

where  $\langle \cdot, \cdot \rangle$  denotes the real inner product. If one is to use the above expression to compute  $\nabla L(\theta)$ , one needs to solve the forward problem in (3) to compute  $\phi$  and additionally  $S$  forward problems to calculate  $F^{-1} \left\{ \frac{\partial q}{\partial \theta_s} - \frac{\partial F}{\partial \theta_s} \phi \right\}$  for  $s = 1, \dots, S$  by solving  $F\mathbf{u} = \frac{\partial q}{\partial \theta_s} - \frac{\partial F}{\partial \theta_s} \phi$ . Now one can rewrite (A.3) as

$$\begin{aligned} \frac{\partial L}{\partial \theta_s} &= -\left\langle (F^{-1})^* D'(\mathbf{y} - D\phi), \frac{\partial q}{\partial \theta_s} - \frac{\partial F}{\partial \theta_s} \phi \right\rangle \\ &= -\left( F^{-1} D'(\mathbf{y} - D\phi) \right)' \left( \frac{\partial q}{\partial \theta_s} - \frac{\partial F}{\partial \theta_s} \phi \right) \end{aligned} \quad (\text{A.4})$$

where  $(F^{-1})^*$  is the adjoint of  $F^{-1}$  and (A.4) is due to the fact that  $F$  is a real symmetric matrix, in other words, a self-adjoint operator. Computing  $\nabla L(\theta)$  by (A.4) requires only two forward solutions, one for computing  $\phi$  and the other for computing the vector  $F^{-1} D'(\mathbf{y} - D\phi)$  by solving  $F\mathbf{u} = D'(\mathbf{y} - D\phi)$  for  $\mathbf{u}$ , hence a computational saving by a factor of  $(S + 1)/2$  is achieved if the cost for forward solutions is dominant in the gradient computation. Using (A.4) for computing the gradient is often referred to as adjoint or reverse

differentiation in diffuse optical tomography (Davies *et al* 1997, Hielscher *et al* 1999). A similar idea is also used in Arridge and Schweiger (1995, 1998).

Now we consider the special case of bioluminescence tomography, where  $\mathbf{F}$  is a known constant and  $\mathbf{q} = \mathbf{W}\boldsymbol{\theta}$  as in (4). Then (A.3) becomes

$$\frac{\partial \mathbf{L}}{\partial \theta_s} = - \left\langle \mathbf{y} - \mathbf{D}\phi, \mathbf{D} \mathbf{F}^{-1} \mathbf{W} \mathbf{e}^s \right\rangle$$

where  $\mathbf{e}^s$  represents the  $s$ th unit vector. Calculating the vector  $\mathbf{D}\mathbf{F}^{-1}\mathbf{W}\mathbf{e}^s$  in the above equation for  $s = 1, \dots, S$  is equivalent to computing the full system matrix  $\mathbf{A} = \mathbf{D}\mathbf{F}^{-1}\mathbf{W}$  as in the direct calculation approach described in section 2.4.1. On the other hand, (A.4) for the adjoint differentiation method reduces to

$$\frac{\partial \mathbf{L}}{\partial \theta_s} = - \left( \mathbf{W}' \mathbf{F}^{-1} \mathbf{D}' (\mathbf{y} - \mathbf{D}\phi) \right)' \mathbf{e}^s = - \left( \mathbf{A}' (\mathbf{y} - \mathbf{D}\phi) \right)' \mathbf{e}^s$$

where calculating  $\mathbf{A}'(\mathbf{y}-\mathbf{D}\phi)$  by a forward solution method without precomputing the system matrix  $\mathbf{A}$  corresponds to the on-the-fly implementation of F/B projectors.

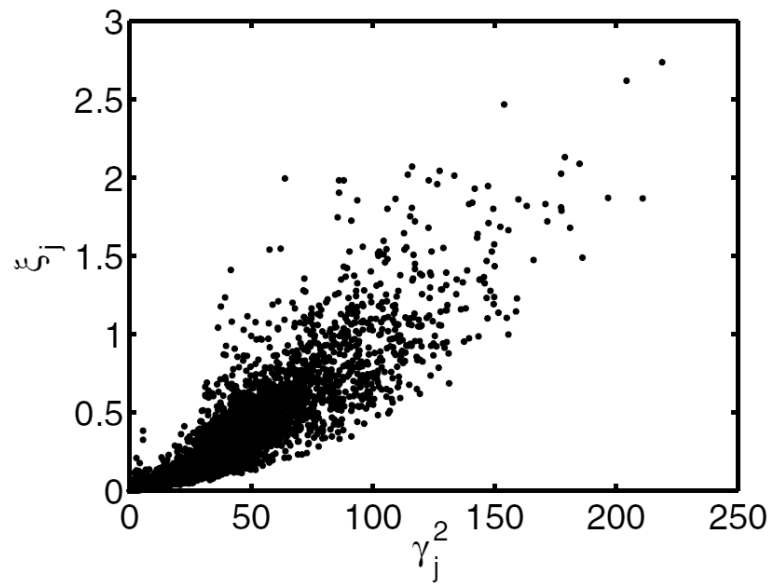
## References

- Allard H, Cote D, Davidson L, Dazai J, Henkelman RM. Combined magnetic resonance and bioluminescence imaging of live mice. *J. Biomed. Opt.* 2007; 12:034018. [PubMed: 17614726]
- Ahn S, Chaudhari AJ, Darvas F, Bouman CA, Leahy RM. Fast image reconstruction methods for fully 3D multispectral optical bioluminescence tomography. *Proc. IEEE Intl. Symp. Biomed. Imaging.* 2007a:229–32.
- Ahn S, Dutta J, Chaudhari AJ, Leahy RM. Computationally efficient image reconstruction methods for multispectral optical fluorescence tomography using FEM-based forward models. *Abstract Book Joint Molecular Imaging Conf.* 2007b:329–30.
- Ahn S, Fessler JA. Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms. *IEEE Trans. Med. Imaging.* 2003; 22:613–26. [PubMed: 12846430]
- Ahn S, Fessler JA, Blatt D, Hero AO. Convergent incremental optimization transfer algorithms: Application to tomography. *IEEE Trans. Med. Imaging.* 2006; 25:283–96. [PubMed: 16524085]
- Alexandrakis G, Rannou FR, Chatziioannou AF. Tomographic bioluminescence imaging by use of a combined optical-PET (OPET) system: a computer simulation feasibility study. *Phys. Med. Biol.* 2005; 50:4225–41. [PubMed: 16177541]
- Alexandrakis G, Rannou FR, Chatziioannou AF. Effect of optical property estimation accuracy on tomographic bioluminescence imaging: simulation of a combined optical-PET (OPET) system 2006. *Phys. Med. Biol.* 2006; 51:2045–53. [PubMed: 16585844]
- Arridge SR. Optical tomography in medical imaging. *Inverse Problems.* 1999; 15:41–93.
- Arridge SR, Dehghani H, Schweiger M, Okada E. The finite element model for the propagation of light in scattering media: a direct method for domains with nonscattering regions. *Med. Phys.* 2000; 27:252–64. [PubMed: 10659765]
- Arridge SR, Schweiger M. Photon-measurement density functions: part 2. Finite-element-method calculations. *Appl. Opt.* 1995; 34:8026–37. [PubMed: 21068901]
- Arridge SR, Schweiger M. A gradient-based optimisation scheme for optical tomography. *Opt. Express.* 1998; 2:213–26. [PubMed: 19377605]
- Arridge SR, Schweiger M, Hiraoka M, Delpy DT. A finite element approach for modeling photon transport in tissue. *Med. Phys.* 1993; 20:299–309. [PubMed: 8497214]

- Baillet S, Mosher JC, Leahy RM. Electromagnetic brain mapping. *IEEE Signal Process. Mag.* 2001; 18:14–30.
- Bertsekas, DP. *Nonlinear Programming*. 2nd edn. Athena Scientific; Belmont, MA: 1999.
- Bouman C, Sauer K. A generalized Gaussian image model for edge-preserving MAP estimation. *IEEE Trans. Image Process.* 1993; 2:296–310. [PubMed: 18296219]
- Bouman CA, Sauer K. A unified approach to statistical tomography using coordinate descent optimization. *IEEE Trans. Image Process.* 1996; 5:480–92. [PubMed: 18285133]
- Browne JA, De Pierro AR. A row-action alternative to the EM algorithms for maximizing likelihoods in emission tomography. *IEEE Trans. Med. Imaging.* 1996; 15:687–99. [PubMed: 18215950]
- Cao N, Nehorai A, Jacob M. Image reconstruction for diffuse optical tomography using sparsity regularization and expectation-maximization algorithm. *Opt. Express.* 2007; 15:13695–708. [PubMed: 19550641]
- Chaudhari, AJ. PhD Thesis. University of Southern California; Los Angeles, CA: 2006. Hyperspectral and multispectral optical bioluminescence and fluorescence tomography in small animal imaging.
- Chaudhari AJ, Darvas F, Bading JR, Moats RA, Conti PS, Smith DJ, Cherry SR, Leahy RM. Hyperspectral and multispectral bioluminescence optical tomography for small animal imaging. *Phys. Med. Biol.* 2005; 50:5421–41. [PubMed: 16306643]
- Cheong WF, Prahl SA, Welch AJ. A review of the optical properties of biological tissues. *IEEE J. Quantum Electron.* 1990; 26:2166–85.
- Comsa DC, Farrell TJ, Patterson MS. Quantification of bioluminescence images of point source objects using diffusion theory models. *Phys. Med. Biol.* 2006; 51:3733–46. [PubMed: 16861777]
- Cong AX, Wang G. Multispectral bioluminescence tomography: methodology and simulation. *Int. J. Biomed. Imaging.* 2006; 2006:57614. [PubMed: 23165040]
- Cong W, Kumar D, Liu Y, Cong A, Wang G. A practical method to determine the light source distribution in bioluminescent imaging. *Proc. SPIE 5535: Developments in X-Ray Tomography IV.* 2004:679–86.
- Cong W, et al. Practical reconstruction method for bioluminescence tomography. *Opt. Express.* 2005; 13:6756–71. [PubMed: 19498692]
- Contag CH, Bachmann MH. Advances in in vivo bioluminescence imaging of gene expression. *Annu. Rev. Biomed. Eng.* 2002; 4:235–60. [PubMed: 12117758]
- Davies AJ, Christianson DB, Dixon LCW, Roy R, van der Zee P. Reverse differentiation and the inverse diffusion problem. *Adv. Eng. Softw.* 1997; 28:217–21.
- Dehghani H, Davis SC, Jiang S, Pogue BW, Paulsen KD, Patterson MS. Spectrally resolved bioluminescence optical tomography. *Opt. Lett.* 2006; 31:365–7. [PubMed: 16480210]
- Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B.* 1977; 39:1–38.
- Erdo an H, Fessler JA. Ordered subsets algorithms for transmission tomography. *Phys. Med. Biol.* 1999; 44:2835–51. [PubMed: 10588288]
- Gordon R, Bender R, Herman GT. Algebraic reconstruction techniques (ART) for the three-dimensional electron microscopy and X-ray photography. *J. Theor. Biol.* 1970; 29:471–81. [PubMed: 5492997]
- Gu X, Zhang Q, Larcom L, Jiang H. Three-dimensional bioluminescence tomography with model-based reconstruction. *Opt. Express.* 2004; 12:3996–4000. [PubMed: 19483937]
- Han W, Cong W, Wang G. Mathematical theory and numerical analysis of bioluminescence tomography. *Inverse Problems.* 2006; 22:1659–75.
- Han W, Wang G. Theoretical and numerical analysis on multispectral bioluminescence tomography. *IMA J. Appl. Math.* 2007; 72:67–85.
- Hebert T, Leahy RM. A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors. *IEEE Trans. Med. Imaging.* 1989; 8:194–202. [PubMed: 18230517]
- Hebert T, Leahy RM, Singh M. Three-dimensional maximum-likelihood reconstruction for an electronically collimated single-photon-emission imaging system. *J. Opt. Soc. Am. A.* 1990; 7:1305–13. [PubMed: 2370591]

- Herman, GT. Image Reconstruction from Projections: The Fundamentals of Computerized Tomography. Academic; New York: 1980.
- Hielscher AH, Klose AD, Hanson KM. Gradient-based iterative image reconstruction scheme for time-resolved optical tomography. *IEEE Trans. Med. Imaging.* 1999; 18:262–71. [PubMed: 10363704]
- Hudson HM, Larkin RS. Accelerated image reconstruction using ordered subsets of projection data. *IEEE Trans. Med. Imaging.* 1994; 13:601–9. [PubMed: 18218538]
- Jiang H. Frequency-domain fluorescent diffusion tomography: a finite-element-based algorithm and simulations. *Appl. Opt.* 1998; 37:5337–43. [PubMed: 18286015]
- Jiang M, Wang G. Image reconstruction for bioluminescence tomography. *Proc. SPIE 5535: Developments in X-Ray Tomography IV.* 2004:335–51.
- Jiang M, Zhou T, Cheng J, Cong W, Wang G. Image reconstruction for bioluminescence tomography from partial measurement. *Opt. Express.* 2007; 15:11095–116. [PubMed: 19547465]
- Kaufman L. Implementing and accelerating the EM algorithm for positron emission tomography. *IEEE Trans. Med. Imaging.* 1987; 6:37–51. [PubMed: 18230425]
- Klose AD. Transport-theory-based stochastic image reconstruction of bioluminescent sources. *J. Opt. Soc. Am. A.* 2007; 24:1601–8.
- Kuo C, Coquoz O, Troy TL, Xu H, Rice BW. Three-dimensional reconstruction of *in vivo* bioluminescent sources based on multispectral imaging. *J. Biomed. Opt.* 2007; 12:024007. [PubMed: 17477722]
- Li Q, Ahn S, Leahy RM. Fast hybrid algorithms for PET image reconstruction. *Proc. IEEE Nucl. Sci. Symp. Medical Imaging Conf.* 2005; 4:1851–5.
- Lin FH, Witzel T, Ahlfors SP, Stufflebeam SM, Belliveau JW, Hamalainen MS. Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. *NeuroImage.* 2006; 31:160–71. [PubMed: 16520063]
- Lv Y, Tian J, Cong W, Wang G, Luo J, Yang W, Li H. A multilevel adaptive finite element algorithm for bioluminescence tomography. *Opt. Express.* 2006; 14:8211–23. [PubMed: 19529195]
- Lv Y, Tian J, Cong W, Wang G, Yang W, Qin C, Xu M. Spectrally resolved bioluminescence tomography with adaptive finite element analysis: methodology and simulation. *Phys. Med. Biol.* 2007; 52:4497–512. [PubMed: 17634646]
- Mumcuo lu EÜ, Leahy R, Cherry SR, Zhou Z. Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET images. *IEEE Trans. Med. Imaging.* 1994; 13:687–701. [PubMed: 18218547]
- Meng XL, van Dyk D. The EM algorithm - An old folk song sung to a fast new tune. *J. R. Stat. Soc. Ser. B.* 1997; 59:511–67.
- Ntziachristos V, Ripoll J, Wang LV, Weissleder R. Looking and listening to light: the evolution of whole-body photonic imaging. *Nat. Biotechnol.* 2005; 23:313–20. [PubMed: 15765087]
- Qi J, Leahy RM. Iterative reconstruction techniques in emission computed tomography. *Phys. Med. Biol.* 2006; 51:541–78.
- Qi J, Leahy RM, Hsu C, Farquhar TH, Cherry SR. Fully 3D Bayesian image reconstruction for the ECAT EXACT HR+ *IEEE Trans. Nucl. Sci.* 1998; 45:1096–103.
- Rice BW, Cable MD, Nelson MB. In vivo imaging of light-emitting probes. *J. Biomed. Opt.* 2001; 6:432–40. [PubMed: 11728202]
- Roy R, Sevick-Muraca EM. Three-dimensional unconstrained and constrained image-reconstruction techniques applied to fluorescence, frequency-domain photon migration. *Appl. Opt.* 2001; 40:2206–15. [PubMed: 18357229]
- Schweiger M, Arridge SR. The finite-element method for the propagation of light in scattering media: frequency domain case. *Med. Phys.* 1997; 24:895–902. [PubMed: 9198025]
- Schweiger M, Arridge SR, Hiraoka M, Delpy DT. The finite element method for the propagation of light in scattering media: boundary and source conditions. *Med. Phys.* 1995; 22:1779–92. [PubMed: 8587533]
- Shen H, Cong W, Qian X, Durairaj K, Wang G. Numerical study on the validity of the diffusion approximation for computational optical biopsy. *J. Opt. Soc. Am. A.* 2007; 24:423–29.

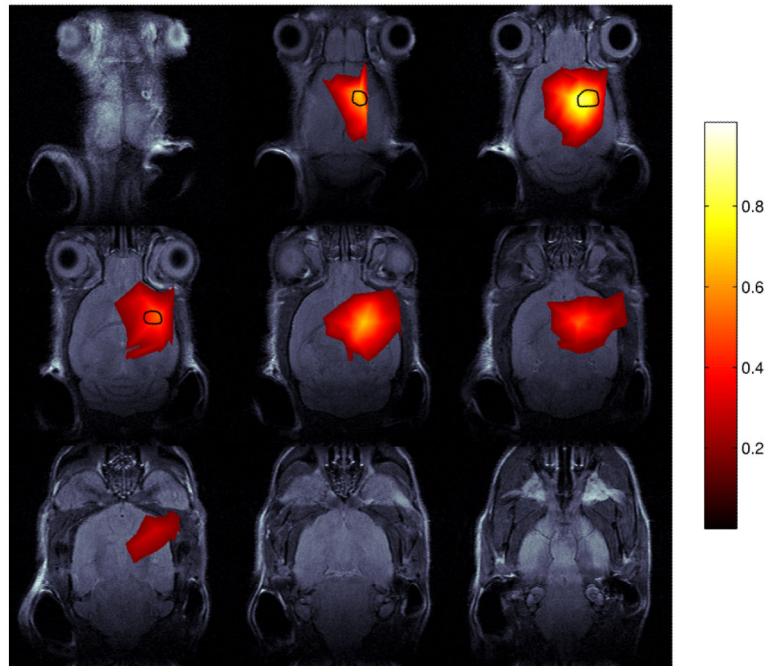
- Shepp LA, Vardi Y. Maximum likelihood reconstruction for emission tomography. *IEEE Trans. Med. Imaging.* 1982; 1:113–22. [PubMed: 18238264]
- Slavine NV, Lewis MA, Richer E, Antich PP. Iterative reconstruction method for light emitting sources based on the diffusion equation. *Med. Phys.* 2006; 33:61–8. [PubMed: 16485410]
- Soloviev VY. Tomographic bioluminescence imaging with varying boundary conditions. *Appl. Opt.* 2007; 46:2778–884. [PubMed: 17446927]
- Tarantola, A. *Inverse Problem Theory and Methods for Model Parameter Estimation.* SIAM; Philadelphia, PA: 2005.
- Wang G, Cong W, Durairaj K, Qian X, Shen H, Sinn P, Hoffman E, McLennan G, Henry M. *In vivo* mouse studies with bioluminescence tomography. *Opt. Express.* 2006a; 14:7801–9. [PubMed: 19529149]
- Wang G, Shen H, Durairaj K, Qian X, Cong W. The first bioluminescence tomography system for simultaneous acquisition of multiview and multispectral data. *Int. J. Biomed. Imaging.* 2006b; 2006:58601. [PubMed: 23165042]
- Wang G, Li Y, Jiang M. Uniqueness theorems in bioluminescence tomography. *Med. Phys.* 2004; 31:2289–99. [PubMed: 15377096]
- Ye JC, Webb KJ, Bouman CA, Millane RP. Optical diffusion tomography by iterative-coordinate-descent optimization in a Bayesian framework. *J. Opt. Soc. Am. A.* 1999; 16:2400–12.



**Figure 1.**

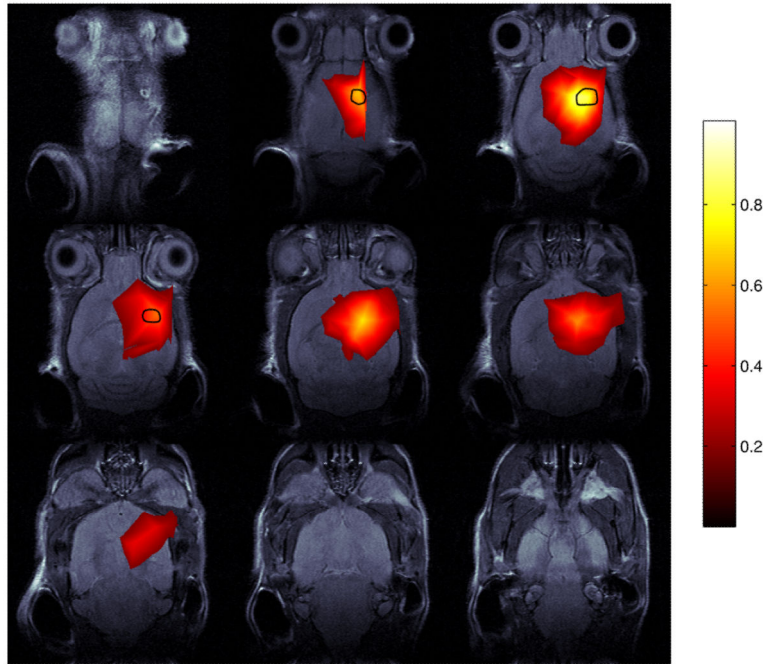
Example scatter plot of  $\xi_j = \sum_i a_{ij}^2$  versus  $\gamma_j^2 = \left(\sum_i a_{ij}\right)^2$  for the multispectral imaging system used in section 4.





**Figure 2.**

The RLS image  $\hat{x}$ , which minimizes the cost function  $\Phi$  given in (11), shown on nine horizontal planes. The reconstructed bioluminescence image was obtained by 2000 iterations of GPM-N and is overlaid on co-registered MR slices. Reconstructed values greater than or equal to 10% of the maximum value are displayed. The black contour denotes the boundary of the tumor obtained from the MR image. The color scale on the right shows relative intensity. These results are not quantitatively calibrated.



**Figure 3.** The same as figure 2 except that this image was reconstructed using 29 iterations of GPM-EN using the on-the-fly approach. The relative error  $\mathcal{E}$  is about 10%.

**Table 1**

Comparison of computation costs for different reconstruction methods when a single image reconstruction is performed.

F/B projection approach	Algorithm	$\mathcal{E} < 10\%$		$\mathcal{E} < 5\%$		$\mathcal{E} < 1\%$	
		Iteration cost (# iters)	Total recon. time	Iteration cost (# iters)	Total recon. time	Iteration cost (# iters)	Total recon. time
Direct calculation approach with precomputed system matrix (precomput. time: 12 708 s)	GPM-U	359 s (216)	13068 s	599 s (378)	13307 s	1444 s (879)	14152 s
	GPM-N	53 s (32)	12761 s	77 s (46)	12786 s	129 s (79)	12838 s
	GPM-EN	54 s (29)	12762 s	71 s (42)	12780 s	121 s (73)	12830 s
	CG	563 s (216)	13272 s	941 s (378)	13650 s	2376 s (878)	15084 s
	PCG-N	58 s (35)	12766 s	88 s (52)	12796 s	157 s (93)	12865 s
	PCG-EN	59 s (33)	12767 s	81 s (48)	12789 s	146s (86)	12854 s
	PCG-EM	81 s (45)	12789 s	111 s (66)	12819 s	268 s (151)	12976 s
	CD	37 s (20)	12746 s	46 s (26)	12754 s	64 s (38)	12773 s
	OS-SPS-2	48 s (36)	12756 s	80 s (61)	12788 s	178 s (136)	12887 s
	OS-SPS-5	20 s (15)	12729 s	35 s (26)	12743 s	–	–
On-the-fly approach (precomput. time: 110 s)	OS-SPS-10	11 s (8)	12719 s	20 s (15)	12728 s		
	GPM-U	4877 s (216)	4987 s	8523 s (378)	8633 s	16152 s (879)	16262 s
	GPM-EN	688 s (29)	798 s	981 s (42)	1091 s	1614 s (73)	1724 s
	CG	6531 s (216)	6641 s	11400 s (378)	11509 s	21557 s (878)	21666s
	PCG-EN	735 s (33)	845 s	1035 s (48)	1145 s	1797 s (86)	1907 s
	PCG-EM	1033 s (45)	1143 s	1480 s (66)	1590 s	3291 s (151)	3401 s